บทที่ 2

ทฤษฎีเกี่ยวกับริดจ์รีเกรสชัน

ในกรณีที่ตัวแปรอิสระในตัวแบบมีสภาพไม่เหมาะสม (ill-condition) ซึ่งในที่นี้หมายถึงตัวแปรอิสระมีค่าสหสัมพันธ์สูง การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองต่ำสุดจะไม่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด Hoerl และ kennard ได้เสนอวิธีการที่ใช้การบวกค่าคงที่ที่มากกว่าศูนย์กับสมาชิกทุกตัวบนเส้นทแยงมุมของเมตริกซ์ $X'X$ วิธีการดังกล่าวนี้จะให้ค่าประมาณของพารามิเตอร์ที่เอนเอียง และมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยที่สุด ซึ่งต่อมาในปีพ.ศ.2518 Richard F.Gunst และ Robert L. Mason ได้นำวิธีการของ Horel และ kennard ไปตรวจสอบซึ่งให้ผลดังที่กล่าวไว้ โดยที่รายละเอียดต่างๆมีดังนี้

2.1 ตัวแบบทั่วไป

ตัวแบบทั่วไปของการวิเคราะห์การถดถอยพหุเชิงเส้นเขียนได้ดังนี้

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

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$ โดยที่ n คือจำนวนค่าสังเกต

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$ โดยที่ p คือจำนวนตัวแปรอิสระ

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด $(p+1) \times 1$

และ $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของค่าความคลาดเคลื่อนที่เกิดขึ้นซึ่งมีขนาด $n \times 1$ และมีการแจกแจงแบบปกติ โดยที่ $E(\underset{\sim}{\varepsilon}) = 0$ และ $Cov(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

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

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

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด

ถ้าตัวแปรอิสระมีค่าสหสัมพันธ์สูงกล่าวคือสมาชิกของเมตริกซ์สหสัมพันธ์ของตัวแปรอิสระที่ไม่อยู่บนเส้นทแยงมุมมีค่าเข้าใกล้ 1 (ไม่อยู่ในรูปเมตริกซ์เอกลักษณ์การคูณ) การประมาณค่าด้วยวิธีกำลังสองน้อยที่สุดจะมีผลทำให้เกิดข้อผิดพลาดของตัวประมาณ แต่ในทางปฏิบัติสหสัมพันธ์ของตัวแปรอิสระที่มีลักษณะดังกล่าวข้างต้นนี้อาจหลีกเลี่ยงไม่ได้ในบางกรณี

ดังนั้นจึงควรพิจารณาวิธีการประมาณค่า β ที่เหมาะสมกว่านี้ ซึ่งในที่นี้ควรพิจารณาวิธีริดจ์รีเกรสชันแทนวิธีกำลังสองน้อยที่สุด

การคำนวณโดยวิธีริดจ์รีเกรสชัน จะให้ตัวประมาณ β ดังนี้ ตัวประมาณค่าพารามิเตอร์ β

(2.1.3) $\beta^* = (X'X + kI)^{-1} X'y \quad ; \quad k > 0$

จากสมการ (2.1.3) ใช้หลักการคำนวณเหมือนกับสมการ (2.1.2) โดยที่วิธีริดจ์รีเกรสชัน มีส่วนที่แตกต่างจากวิธีกำลังสองน้อยที่สุด คือ จะมีค่า k เข้ามาปรับค่า $(X'X)^{-1}$ โดยการนำค่า k บวกกับสมาชิกทุกตัวบนเส้นทแยงมุมของเมตริกซ์ $X'X$ เพื่อให้ตัวประมาณ β^* มีคุณสมบัติที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดต่ำลง จากสมการ (2.1.3) กำหนดให้ $k > 0$ เราไม่อาจหาค่า k ที่แน่นอนได้นอกจากจะต้องมีการทดลองให้ k มีค่าเพิ่มจากศูนย์ทีละน้อยและแทนค่า k ในสมการ (2.1.3) จนกว่าจะได้ค่า k ที่เหมาะสมจะเห็นได้ว่าค่าตัวประมาณ β^* จะได้จากการคำนวณหลายรอบ (iterative) เพราะขึ้นกับค่า k ที่เปลี่ยนแปลงไปเรื่อยๆทุกครั้งที่กำหนดค่า k เพื่อคำนวณค่าตัวประมาณจะนำค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดขึ้นจากวิธีนี้ และ วิธีกำลังสองน้อยที่สุดมาเปรียบเทียบกันไปเรื่อยๆ จนกว่าจะพบว่า ค่า k ใดที่ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่า β ด้วยวิธีริดจ์รีเกรสชันให้ค่าน้อยกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่า β ด้วยวิธีกำลังสองน้อยที่สุด เราจะถือว่าค่า k นั้นเริ่มจะเป็นค่า k ที่เหมาะสมแล้วจะค่อยๆเปลี่ยนค่า k ไปจนกว่าจะทำให้หาค่า k ที่เหมาะสมที่สุด ซึ่งจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด ดังนั้นอาจกล่าวได้ว่าวิธีการริดจ์รีเกรสชันจะไม่สามารถกำหนดค่า k ที่แน่นอนได้

2.2 คุณสมบัติของตัวประมาณค่า ด้วยวิธีกำลังสองน้อยที่สุด

จากวิธีกำลังสองน้อยที่สุด ซึ่งให้ตัวประมาณค่าเป็น

$$\hat{\beta} = (X'X + kI)^{-1} X'y$$

โดยตัวประมาณ $\hat{\beta}$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดในกลุ่มของตัวประมาณค่า β ที่ไม่เอนเอียง ดังนั้นค่าความผิดพลาดเขียนได้ในรูปของ

$$e = y - X\beta$$

ซึ่งค่าผลบวกของความคลาดเคลื่อนกำลังสองมีค่าเท่ากับ $e'e$ ถ้าให้ $\phi(\hat{\beta})$ เป็น ผลบวกความคลาดเคลื่อนกำลังสองเมื่อใช้ตัวประมาณ $\hat{\beta}$ จะได้

(2.2.2) $\phi(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta})$

เนื่องจากข้อมูลตัวอย่างที่นำมาศึกษาการถดถอยพหุเชิงเส้นส่วนมากมีตัวแปรอิสระมีความสัมพันธ์ภายในต่อกัน โดยเฉพาะในอัตราที่สูง จะมีผลทำให้เมตริกซ์ $X'X$ เป็น singular matrix หรือมีลักษณะของ Near Singularity[1] เพราะฉะนั้นถ้าประมาณ β ด้วย $\hat{\beta}$ จากวิธีกำลังสองน้อยที่สุดอาจมีผลทำให้เมตริกซ์ของค่าความแปรปรวนและค่าความแปรปรวนร่วมของตัวประมาณ $\hat{\beta}$ ไม่มีขนาดเล็กที่สุด เราจึงควรพิจารณาคุณสมบัติของตัวประมาณค่า β 2 กรณี คือ เมตริกซ์ค่าความแปรปรวนและค่าความแปรปรวนร่วมของตัวประมาณ $\hat{\beta}$ และค่าเฉลี่ยของกำลังสองระยะทางจาก $\hat{\beta}$ ไปยัง β ซึ่งสามารถเขียนอยู่ในรูปฟังก์ชันของ $X'X$ และ σ^2 ได้ดังต่อไปนี้

กำหนดให้ความแปรปรวนของเวกเตอร์ตัวประมาณ $\hat{\beta}$ มีค่าเป็น $VAR(\hat{\beta})$ ดังนั้น

(2.2.3) $\quad Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

และให้ระยะทางจาก $\hat{\beta}$ ถึง β มีค่าเป็น L_1 ดังนั้น

(2.2.4) $\quad L^2_1 = (\hat{\beta} - \beta)' (\hat{\beta} - \beta)$

ค่าเฉลี่ยของกำลังสองระยะทางจาก β ไปยัง $\hat{\beta}$ เป็นดังนี้

(2.2.5) $\quad E[L^2_1] = \sigma^2 \text{ trace } (X'X)^{-1}$

$E[L^2_1]$ จะมีค่าสมมูล (Equivalent) กับ $E[\hat{\beta}' \hat{\beta}]$

(2.2.6) $\quad E[\hat{\beta}'\hat{\beta}] = \beta'\beta + \sigma^2 \text{ trace } (X'X)^{-1}$

เมื่อ ε มีการแจกแจงแบบปกติจะได้

(2.2.7) $\quad VAR(L^2_1) = 2\sigma^4 \text{ trace } (X'X)^{-2}$

จะเห็นได้ว่าจากสมการ (2.2.3),(2.2.5) และ (2.2.6) ค่าความแปรปรวนของเวกเตอร์ตัวประมาณ $\hat{\beta}$ ($Var(\hat{\beta})$), ค่าเฉลี่ยของกำลังสองระยะทางจาก β ไปยัง $\hat{\beta}$ ($E(L^2_1)$) และ ค่าความแปรปรวนของกำลังสองระยะทางจาก β ไปยัง $\hat{\beta}$ ($V(L^2_1)$) เหล่านี้เป็นฟังก์ชันของ $\hat{\beta}$ เพื่อความสะดวกในการเปรียบเทียบการประมาณค่า β ระหว่างสองวิธีนี้ควรแปลงเมตริกซ์ $(X'X)^{-1}$ ให้อยู่ในรูปฟังก์ชันของค่า eigenvalue ของเมตริกซ์ $X'X$ โดยใช้คุณสมบัติที่สำคัญข้อหนึ่งของค่า eigenvalue ของ $X'X$ กล่าวคือถ้า λ_1 เป็นค่า

---

1 ถ้า column vector ที่ประกอบกันขึ้นเป็นเมตริกซ์ X มีลักษณะของ linearly dependent เมตริกซ์ X จะไม่มี full rank หมายความว่าเมตริกซ์ $X'X$ เป็น singular matrix คือ $|X'X|=0$ อันมีผลทำให้ $(X'X)^{-1}$ ไม่ปรากฏค่า แต่ถ้า column vector ไม่ถึงกับเป็น linearly dependent คือ $|X'X|$ มีค่าไม่เท่ากับ 0 แต่ใกล้ 0 เรายังคงหา $(X'X)^{-1}$ ได้แต่สมาชิกของ $(X'X)^{-1}$ จะมีค่าสูง

eigenvalue ของ X′X เมื่อ i = 1,2,..,p เมื่อ p คือ จำนวนตัวแปรอิสระ แล้ว $\sum \lambda i = trace(X'X)$

สมมติให้ค่าไอเก็น (eigenvalue) ของ X′X มีค่าเป็น

(2.2.8) $(\lambda_{max} = \lambda_1) > \lambda_2 > \dots (\lambda_p = \lambda_{min}) > 0$

จากสมการ (2.2.5) ค่าเฉลี่ยความผิดพลาดยกกำลังสองระหว่าง $\hat{\beta}$ กับ β (risk function) อาจเขียนอยู่ ในรูปฟังก์ชันของค่า eigenvalue ของ X′X ดังนี้

(2.2.9) $E[L^2_1] = \sigma^2 \sum (1/\lambda_i)$

และจากสมการ (2.2.7) ค่า $VAR[L^2_1]$ อาจเขียนอยู่ในรูปฟังก์ชันของค่า eigenvalue ของ X′X ดังนี้

(2.2.10) $VAR[L^2_1] = 2\sigma^4 \sum(1/\lambda_i)^2$

จากสมการ (2.2.9) และสมการ (2.2.10) จะได้ขีดจำกัดล่าง (lower bound) ของค่าเฉลี่ยกำลังสองของระยะทางจาก $\hat{\beta}$ ไปยัง β หรือ $E[L^2_1]$ และค่าความแปรปรวนของกำลังสองของระยะทางจาก $\hat{\beta}$ ไปยัง β หรือ $VAR[L^2_1]$ มีค่าเป็น σ^2/λ_{min} และ $2\sigma^4/\lambda^2_{min}$ ตามลำดับ

แต่ในกรณีที่ตัวแปรอิสระเกิดมีสภาพที่ไม่เหมาะสม(เกิดความสัมพันธ์ระหว่างตัวแปรอิสระในอัตราที่สูง) ค่า eigenvalue ของเมตริกซ์ X′X จะมีค่าต่ำมาก (เนื่องจาก |X′X| มีค่าเท่ากับผลคูณของค่า eigenvalue และมีค่าต่ำเมื่อระดับความสัมพันธ์ของตัวแปรอิสระอยู่ในอัตราที่สูง) ซึ่งมีผลทำให้ระยะทางจาก $\hat{\beta}$ ไปยัง β มีค่ามาก พิจารณาค่า $E[L^2_1]$ และ $VAR[L^2_1]$ จะเห็นว่าค่าทั้งสองสูงขึ้น สาเหตุที่เกิดเช่นนี้อาจมีผลมาจากการทำให้ค่าต่ำสุดของค่าผลบวกของความคลาดเคลื่อนกำลังสอง $\phi(\beta)$

2.3 การกำหนดค่า β โดยวิธีริดจ์รีเกรสชัน

A.E. Hoerl (1963) ได้ใช้วิธีทางคณิตศาสตร์มาควบคุมคุณสมบัติของค่าความแปรปรวนของตัวประมาณค่า β ที่คำนวณได้โดยวิธีกำลังสองน้อยที่สุดให้มีค่าลดต่ำลงได้ดังนี้ คือ กำหนดให้ค่าประมาณ β ด้วยวิธีริดจ์รีเกรสชันและวิธีกำลังสองน้อยที่สุดมีค่าเท่ากับ β^* และ β ตามลำดับ จากสมการ (2.1.3) ค่า $\beta^* = [X'X + kI]^{-1} X'Y$ โดยที่ $k > 0$ สามารถแปลง β^* ให้อยู่ในรูปฟังก์ชันใหม่โดยกำหนดให้ $[X'X + kI]^{-1} = W$ ดังนี้

(2.3.1) $\beta^* = W X' Y$

เนื่องจาก k ไม่ใช่ค่าคงที่จะต้องกำหนดค่า k หลายค่า ดังนั้นจะเห็นได้ว่า β^* มีค่าหลายชุดจากตัวประมาณค่า β^* สามารถเขียนให้อยู่ในรูปความสัมพันธ์กับ β เพื่อความ

สะดวกในการเปรียบเทียบวิธีการประมาณค่า β ทั้งสองวิธีดังนี้

(2.3.3) $\hat{\beta}^* = [I + k(X'X)^{-1}]^{-1} \hat{\beta}$

(2.3.4) $= Z \hat{\beta}$

โดยที่ $Z = [I+k(X'X)^{-1}]^{-1}$ และ $W = [X'X + kI]^{-1}$ และค่า Z ,$\hat{\beta}^*$ และ W เหล่านี้จะนำไปพิจารณาเปรียบเทียบคุณสมบัติต่างๆ ของการประมาณค่าพารามิเตอร์ β ตลอดจนการพิสูจน์ทฤษฎีต่างๆ ที่สำคัญของการประมาณค่าสัมประสิทธิ์ความถดถอยของวิธีริดจ์รีเกรสชันต่อไป คุณสมบัติที่จำเป็นของ Z ,$\hat{\beta}^*$ และ W มีดังนี้

2.3.1 ให้ $E_i(W)$ และ $E_i(Z)$ เป็นค่า eigenvalue ของ W และ Z ตามลำดับซึ่งคำนวณได้จาก characteristic equation

$$| W - E_i I | = 0$$

และ $| Z - E_i I | = 0$

(2.3.5) $E_i(W) = 1 / (\lambda_i + k)$

(2.3.6) $E_i(Z) = \lambda_i / (\lambda_i + k)$

ค่า Z จากสมการ (2.3.2) อาจแสดงในรูปฟังก์ชันของ kW ดังนี้

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

(2.3.7) $= I-kW$

ค่าตัวประมาณ $\hat{\beta}^*$ จากสมการ (2.3.3) จะมีค่าน้อยกว่า $\hat{\beta}$ เมื่อ $k>0$

(2.3.8) $(\hat{\beta}^*)'(\hat{\beta}^*) < \hat{\beta}'\hat{\beta}$

<u>พิสูจน์</u> จากนิยาม $\hat{\beta}^* = Z \hat{\beta}$ และ $X'X$, Z ซึ่งมีคุณสมบัติ Symmetric positive define * จะได้ว่า

$$(\hat{\beta}^*)'(\hat{\beta}^*) = (Z \hat{\beta})'(Z \hat{\beta})$$

$$< \sum E^2_i(Z) \ \hat{\beta}'\hat{\beta}$$

(2.3.9) $< E^2_{i(max)}(Z) \ \hat{\beta}'\hat{\beta}$

---

* ถ้าเมตริกซ์สมมาตร A จะมีคุณสมบัติ positive definite (p.d.) ถ้า $\underset{\sim}{x}'A\underset{\sim}{x} > 0$ สำหรับทุกค่าของ $\underset{\sim}{x}$ โดยที่เมตริกซ์ A มีคุณสมบัติต่อไปนี้

1. ค่า eigenvalue ของเมตริกซ์ A ทุกตัวจะมีค่าเป็นบวก
2. เมตริกซ์ A เป็น nonsinguluar
3. rank[CAC'] = rank C

เมื่อ $E_{(max)}(Z) = \lambda_{(max)} / (\lambda_{(max)}+k)$ เมื่อ $\lambda_{(max)}$ เป็นค่า eigenvalue ที่มีค่ามากที่สุดของ $X'X$

ดังนั้น $(\beta^*)'(\beta^*) < \beta'\beta$

เนื่องจากผลบวกของความคลาดเคลื่อนกำลังสองของค่าตัวประมาณ β^* เป็นดังนี้

(2.3.10) $\phi^*(k) = (Y-X\beta^*)'(Y-X\beta^*)$

จากการประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุดเราสามารถเขียน $\phi(\beta)$ ซึ่งเป็นผลบวกของความคลาดเคลื่อนกำลังสองของตัวประมาณ β เป็น $y'y-\beta'X'y$ ดังนั้น เพื่อให้เห็นความแตกต่างของ $\phi^*(k)$ และ $\phi(\beta)$ เราสามารถเขียน $\phi^*(k)$ ในเทอมของ $Y'Y$ และ $X'Y$ ได้ดังนี้

(2.3.11) $\phi^*(k) = Y'Y - (\beta^*)'X'Y - k(\beta^*)'(\beta^*)$

จาก $\phi(\beta)$ และ $\phi^*(k)$ สรุปได้ว่าผลบวกของความคลาดเคลื่อนกำลังสองของวิธีริดจ์รีเกรสชันจะให้ค่าน้อยกว่าผลบวกของความคลาดเคลื่อนกำลังสองที่คำนวณโดยวิธีกำลังสองน้อยที่สุด

2.4 คุณสมบัติของผลบวกของความคลาดเคลื่อนกำลังสองของริดจ์รีเกรสชัน

จากการประมาณค่าสัมประสิทธิ์การถดถอย β ด้วยวิธีกำลังสองน้อยที่สุดจะได้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองดังนี้

(2.4.1) $$MSE(\beta) = VAR(\beta) + BIAS^2(\beta) = \sigma^2(X'X)^{-1}$$

เนื่องจาก $E(\beta)=\beta$ ดังนั้นการประมาณค่าพารามิเตอร์ β ด้วยวิธีกำลังสองน้อยที่สุดจะได้ตัวประมาณค่า β ที่ไม่เอนเอียง แต่ในการประมาณค่าพารามิเตอร์ β^* ด้วยวิธีริดจ์รีเกรสชันจะเป็นตัวประมาณที่เอนเอียง กล่าวคือ $E(\beta^*) = Z\beta$

สำหรับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณสัมประสิทธิ์ความถดถอยด้วยวิธีริดจ์รีเกรสชัน ซึ่งมีค่าเป็น $E[L^2_1(k)]$ อาจพิจารณาได้ดังนี้

$$E[L^2_1(k)] = E[(\beta^*-\beta)'(\beta^*-\beta)]$$

(2.4.1) $$= E[(\beta-\beta)'Z'Z(\beta-\beta)] + (Z\beta-\beta)'(Z\beta-\beta)$$

(2.4.2) $$= \sigma^2 \, trace(X'X)Z'Z + \beta'(Z-I)'(Z-I)\beta$$

(2.4.3) $$= \sigma^2[trace(X'X+kI)^{-1} - k\, trace(X'X+kI)^{-2}] + k^2\beta'(X'X+kI)^{-2}\beta$$

(2.4.4) $= \sigma^2 \sum \lambda_i/(\lambda_i+k)^{-2}$

$+ k^2 \beta'(X'X + kI)^{-2}\beta$

(2.4.5) $= r_1(k) + r_2(k)$

จะเห็นว่า $E[L_1{}^2(k)]$ เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองซึ่งอยู่ในรูปผลบวกของฟังก์ชัน $r_1(k)$ และ $r_2(k)$

โดยที่ $r_1(k)$ เป็นค่าความแปรปรวนของตัวประมาณ β^* $(Var(\beta^*(k))$ เมื่อพิจารณาเทอม $r_2(k)$ ซึ่งเป็นระยะทางจาก $Z\beta$ ไปยัง β จะเห็นว่าเป็นศูนย์เมื่อ $k = 0$ โดยที่ Z เป็นเมตริกซ์เอกลักษณ์การคูณ ซึ่งสอดคล้องกับค่าความเอนเอียงกำลังสองของ β ที่มีค่าเป็นศูนย์สำหรับกรณีที่ $k>0$ อาจพิจารณา $r_2(k)$ ในเทอมของความเอนเอียงของ β กำลังสอง กล่าวคือ $r_2(k) = (BIAS^2(\beta^*))$

ส่วนเทอม $r_1(k)$ นั้นเป็นผลบวกของความแปรปรวนของตัวประมาณค่า β^* ซึ่งเขียนในเทอมของ y ได้ดังนี้

$\beta^* = Z\,\beta$

(2.4.6) $= Z\,(X'X)^{-1}\,X'y$

$v(\beta^*) = Z\,(X'X)^{-1}\,X'\,\sigma^2\,X(X'X)^{-1}\,Z'$

(2.4.7) $= \sigma^2$ trace $Z\,(X'X)^{-1}\,Z'$

เนื่องจากความแปรปรวนของตัวประมาณ β^* ทั้งหมด $v(\beta^*)$ คำนวณได้จากผลบวกของสมาชิกทุกตัวบนเส้นทแยงมุมของเมตริกซ์ $\sigma^2\,Z\,(X'X)^{-1}\,Z'$ เพื่อให้สอดคล้องกับ $r_1(k)$ ของสมการ (2.4.5) ซึ่งเป็นฟังก์ชันที่อยู่ในเทอมของ eigenvalue ของ $X'X$ และ k ดังนั้นเมตริกซ์ $X'X$ และ Z สามารถปรับให้เป็นเมตริกซ์อยู่ในรูปของ λ_i ซึ่งเป็นค่า eigenvalue ของ $X'X$ เมื่อ $i = 1,2,..,p$ และ p หมายถึงจำนวนตัวแปรอิสระ X

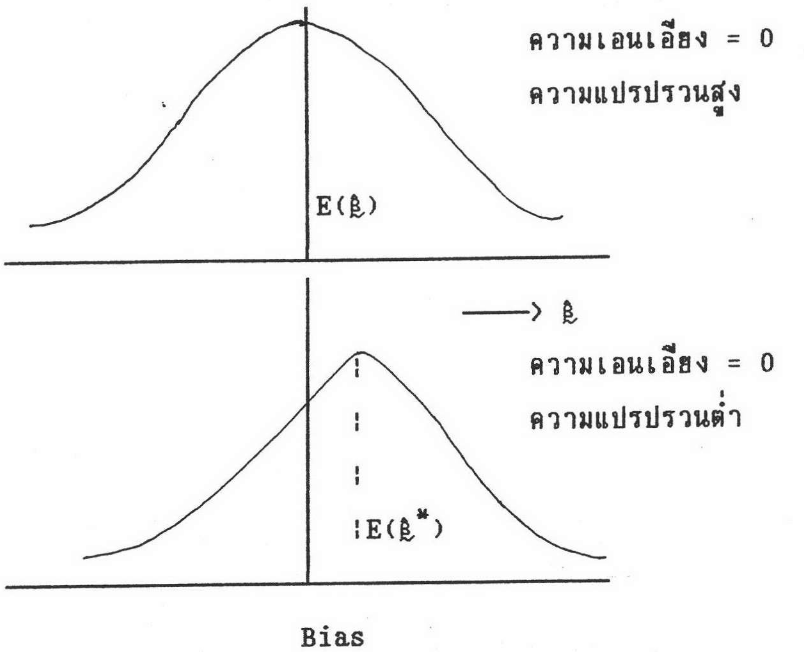
(2.4.8) $\text{trace}(X'X)^{-1} = \sum (1/\lambda_i)$

(2.4.9) $\text{trace}(Z) = \sum \lambda_i/(\lambda_i+k)$

จากสมการ (2.4.7) จะได้ว่า

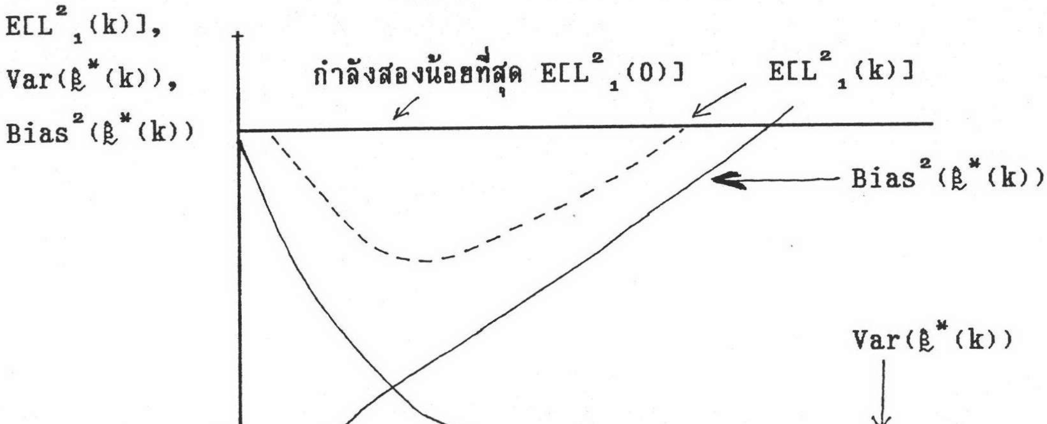
σ^2 trace $Z\,(X'X)^{-1}\,Z' = \sigma^2 \sum \lambda_i/(\lambda_i+k)^2$

ดังนั้น $E[L^2_1(k)] = Var(\beta^*(k)) + Bias^2(\beta^*(k))$



<u>รูปที่ 2.4.1</u>

แสดงการกระจายของตัวประมาณพารามิเตอร์ β ของกำลังสองน้อยที่สุดและริดจ์รีเกรสชัน



<u>รูปที่ 2.4.2</u>

แสดงกราฟของ $E[L^2_1(k)]$ $(MSE(\hat{\beta}^*(k))$,

$Var(\hat{\beta}^*(k))$, $Bias^2(\hat{\beta}^*(k))$

จากรูปที่ 2 เป็นกราฟแสดงความสัมพันธ์ระหว่างความแปรปรวนของตัวประมาณ β^* ความเอนเอียงกำลังสองกับค่า k เมื่อกำหนดให้ $k \in (0,1]$ กล่าวคือความแปร-ปรวนจะมีค่าลดลงเรื่อยๆ เมื่อ k มีค่าเพิ่มขึ้น และ ความเอนเอียงกำลังสองจะมีค่าแปรผันตาม k รูปเส้นไข่ปลาเป็นกราฟที่แสดง ค่า $E[L^2_1(k)]$ ซึ่งเป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดจากตัวประมาณค่า β^* เมื่อ $k>0$

กราฟของฟังก์ชัน $r_1(k)$ จะมีลักษณะเป็นฟังก์ชันที่ลดลง (monotonically decreasing) ในรูปฟังก์ชันของ k และกราฟของฟังก์ชัน $r_2(k)$ มีลักษณะเป็น monotonically increasingในรูปฟังก์ชันของ k จากรูปที่ 2 จะเห็นว่า $E[L^2_1(k)]$ มีค่าน้อยกว่า $E[L^2_1(0)]$ โดยที่ $E[L^2_1(0)]$ เป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่เกิดจากการประมาณค่าพารามิเตอร์ β โดยกำลังสองน้อยที่สุด

จากรูปที่ 2 แสดงให้เห็นการเปลี่ยนแปลงของค่าต่างๆ ที่กล่าวมาแล้วข้างต้น ซึ่งสามารถคำนวณหาขีดจำกัด (Limit) ของความแปรปรวนของตัวประมาณ β^* และความเอนเอียงกำลังสองของตัวประมาณ β^* ได้จากอนุพันธ์ของแต่ละฟังก์ชันดังนี้

(2.4.10) $\lim_{k \to 0} (\partial r_1/\partial k) = -2\sigma^2 \sum (1/\lambda_i^2)$

(2.4.11) $\lim_{k \to 0} (\partial r_2/\partial k) = 0$

ดังนั้นอนุพันธ์ของความแปรปรวนของตัวประมาณ β^* จะมีค่าเป็นลบและจะลู่เข้าสู่ $-2\sigma^2\sum (1/\lambda_i^2)$ เมื่อ $k \to 0$ โดยที่ผลคูณของเมตริกซ์ของตัวแปรอิสระ $(X'X)$ เป็นออร์ธอโกนอลและ $r_1(k)$ มีค่าเข้าสู่ $-\infty$ ในกรณีที่เมตริกซ์ของตัวแปรอิสระมีสภาพไม่เหมาะสมนั้น ค่า eigenvalue $\lambda_p \to 0$

ส่วนอนุพันธ์ของความเอนเอียงกำลังสองมีค่อยๆลดลงและเป็นศูนย์ที่จุดกำเนิดเมื่อ $k \to 0$ คุณสมบัติของสมการ (2.4.10) และ (2.4.11) จะเป็นจริงเมื่อ $k>0$ ซึ่งสรุปได้ว่าค่าความเอนเอียงเพิ่มขึ้นเรื่อยๆ ขณะที่ $Var(\beta^*(k))$ ลดลง ซึ่งจะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลงด้วย

2.4.1 **ทฤษฎีต่างๆที่สำคัญของฟังก์ชันค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง**

ก) ความแปรปรวนทั้งหมด $r_1(k)$ เป็นฟังก์ชันต่อเนื่องและมีลักษณะเป็น monotonically decreasing ฟังก์ชันในรูปของ k

พิสูจน์ จาก (2.4.4) $r_1(k) = \sigma^2 \sum \lambda_i/(\lambda_i+k)^{-2}$

หาอนุพันธ์ของ $r_1(k)$ เทียบกับ k

(2.4.12) $\frac{\partial r_1(k)}{\partial k} = \frac{\partial}{\partial k}(\sigma^2 \sum \partial \lambda_i/(\lambda_i+k)^{-2})$

$= \sigma^2 \sum \lambda_i(-2)(\lambda_i+k)^{-3}$

(2.4.13) $r_1'(k) = -2\sigma^2 \sum \lambda_i/(\lambda_i+k)^3$

$r_1(k)$ จะเป็นฟังก์ชันต่อเนื่องของ k และ $r_1'(k)$ เป็นฟังก์ชันที่มีค่าเป็นลบ ดังนั้น $\lim_{k\to 0} r_1'(k)$ จะมีลักษณะเป็น monotonically decreasing ฟังก์ชันในรูปของ k

ดังนั้นอนุพันธ์อันดับหนึ่งของ $r_1(k)$ เมื่อเทียบกับ k มีค่าเป็น $r_1'(k)$ จะลู่เข้าสู่ $-\infty$ เมื่อ $k \to 0$ และ $\lambda_p \to 0$

นั่นคือ $\lim_{k\to 0} r_1'(k) = -\infty$

ข) ความเอนเอียงกำลังสองของ $r_2(k)$ เป็นฟังก์ชันต่อเนื่องและ monotonically increasing ฟังก์ชันในรูปของ k

พิสูจน์ จาก (2.4.4) $r_2(k) = k^2 \beta'(X'X + kI)^{-2} \beta$

ถ้า λ เป็นเมตริกซ์ของ eigenvalue $X'X$ และ p เป็น orthogonal เมตริกซ์ที่มีคุณสมบัติ $X'X = P\lambda P$ ดังนั้น $r_2(k)$ อาจเขียนใหม่ได้ดังนี้

(2.4.14) $r_2(k) = k^2 \sum \alpha_i^2 /(\lambda_i+k)^2$

(2.4.15) เมื่อ $\alpha = P\beta$

เนื่องจาก $\lambda_i > 0$ ทุกค่าของ i และ $k>0$ ดังนั้น แต่ละเทอมของ (λ_i+k) จะเป็นบวก จาก (2.4.14) จะได้ $r_2(0)=0$ เมื่อ $k=0$ แล้วจะได้ $r_2(k)$ เป็นฟังก์ชันต่อเนื่องเมื่อ $k>0$ อาจเขียน $r_2(k)$ ใหม่ดังนี้

(2.4.16) $r_2(k) = \sum \alpha_i^2 /[1+(\lambda_i/k)]^2$

ทุกค่าของ $i, \lambda_i>0$ ดังนั้น λ_i/k จะเป็นฟังก์ชันที่ลดลง(monotonically decreasing) เมื่อ k มีค่าเพิ่มขึ้นและแต่ละเทอมของ $r_2(k)$ เป็นฟังก์ชันที่เพิ่มขึ้น (monotonically increasing) ดังนั้น $r_2(k)$ เป็นฟังก์ชันต่อเนื่องและ monotonically increasing ฟังก์ชันในรูปของ k

บทแทรก 2.4.1.1 ความเอนเอียงกำลังสอง $r_2(k)$ จะมีค่าเข้าใกล้ $\beta'\beta$ ซึ่งเป็นขีดจำกัดบน

พิสูจน์ จาก (2.4.16)

$$\lim_{k \to 0} r_2(k) = \sum \alpha_i^2 = \alpha'\alpha = \beta' p' p\, \beta = \beta'\beta$$

บทแทรก 2.4.1.2 อนุพันธ์ $r_2'(k)$ จะมีค่าเข้าใกล้ 0 ขณะที่ $k \to 0$

พิสูจน์ จาก (2.4.14) $r_2(k) = k^2 \sum \alpha_i^2 / (\lambda_i+k)^2$

(2.4.17) $$\frac{\partial\, r_2(k)}{\partial k} = 2k \sum \lambda_i\, \alpha_i^2/(\lambda_i+k)^3$$

ผลบวกของแต่ละเทอม $2k\, \lambda_i\, \alpha_i^2/(\lambda_i+k)^3$ เมื่อ $i = 1,..,p$ เป็นฟังก์ชันต่อเนื่องและขีดจำกัด (limit) ของแต่ละเทอมจะเข้าใกล้ 0 ขณะที่ $k \to 0$ ดังนั้น $r_2'(k)$ จะเข้าใกล้ 0 ขณะที่ $k \to 0$

<u>ทฤษฎี 2.4.1.3</u> สำหรับทุกค่าของ $k > 0$

$E[L_1^2(k)] < E[L_1^2(0)]$ เมื่อ $E[L_1^2(0)] = \sigma^2 \sum (1/\lambda_i)$

<u>พิสูจน์</u> จาก (2.4.4) (2.4.14) (2.4.17) จะได้ว่า

$$\frac{\partial\, E[L_1^2(k)]}{\partial k} = \frac{\partial\, r_1(k)}{\partial k} + \frac{\partial\, r_2(k)}{\partial k}$$

(2.4.18) $$= -2\sigma^2 \sum \lambda_i/(\lambda_i+k)^3 + 2k \sum \lambda_i \alpha_i^2/(\lambda_i+k)^3$$

$$\frac{\partial\, E[L_1^2(0)]}{\partial k} = \frac{\partial r_1(k)}{\partial k} + \frac{\partial r_2(k)}{\partial k}$$

$$= \frac{\partial\, [\sigma^2 \sum (1/\lambda_i)]}{\partial k} + 0$$

(2.4.19) $$= 0$$

เนื่องจาก $k = 0$ จะได้ $r_1(0) = \sigma^2 \sum (1/\lambda_i)$ และ $r_2(0) = 0$

การที่แสดง $E[L_1^2(k)] < E[L_1^2(0)]$ จะต้องใช้สมการ

$$\frac{\partial E[L_1^2(k)]}{\partial k} < \frac{\partial E[L_1^2(0)]}{\partial k} \quad ; k > 0$$

จาก (2.4.19) $\frac{\partial E[L_1^2(0)]}{\partial k} = 0$ ดังนั้น $\frac{\partial E[L_1^2(k)]}{\partial k} < 0$

จึงจะสรุปได้ว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่คำนวณโดยวิธีริดจ์รีเกรสชันจะให้ค่าน้อยกว่าที่คำนวณโดยวิธีกำลังสองน้อยที่สุด

เมื่อแทนค่า $k < \sigma^2/\beta^2_{max}$ ลงในสมการ (2.5.18) จะได้ $\frac{\partial E[L_1^2(k)]}{\partial k} < 0$

นั่นคือ $E[L_1^2(k)] < E[L_1^2(0)]$

<u>ข้อสังเกต</u> การที่ $\frac{\partial E[L_1^2(k)]}{\partial k} < 0$ ค่า k ที่จะแทนค่าลงในสมการ (2.5.18) อาจมีค่ามากกว่า σ^2/β^2_{max} แต่จะต้องไม่เกิน σ^2/β^2_{min} ดังนั้นค่า k ที่เหมาะสมจะเป็นค่าที่อยู่ในช่วง $(\sigma^2/\beta^2_{max}, \sigma^2/\beta^2_{min})$ ที่ทำให้อนุพันธ์อันดับที่ 1 ของ $E[L_1^2(k)]$ เทียบกับ k มีค่าเป็น 0

2.5 สูตรและทฤษฎีที่เกี่ยวข้องในการประมาณค่า k ในวิธีริดจ์รีเกรสชัน

2.5.1 วิธี HKB (Hoerl-Kennard-Baldwin Method)

วิธีนี้เสนอโดย Hoerl, Kennard และ Baldwin (1975) เนื่องจากว่าวิธี Ridge Regression จะให้ค่า MSE ต่ำถ้า $k_i = \sigma^2/\beta_i^2$ เมื่อต้องการค่า k เพียงค่าเดียวจึงหาค่า k มาจากส่วนเฉลี่ย Harmonic Mean ของ k โดยที่

$$\begin{aligned} 1/k &= \sum (1/k_i) \,/\, p \\ &= \sum (\beta_i^2/\sigma^2) \,/\, p \end{aligned}$$

$$= \sum \hat{\beta}_i^2 / (p\,\hat{\sigma}^2)$$

$$= \hat{\beta}'\hat{\beta} / (p\,\hat{\sigma}^2)$$

จะได้ว่า $k = p\,\hat{\sigma}^2 / (\hat{\beta}'\hat{\beta})$

เมื่อ p เป็นจำนวนตัวแปรอิสระ โดยที่ $\hat{\beta} = (X'X)^{-1} X'y$

และ $\hat{\sigma}^2 = \dfrac{(y'y - \hat{\beta}'X'y)}{(n-p)}$

2.5.2 วิธี Tze-San-Lee (Tze-San Lee Method)

วิธีนี้เสนอโดย Tze-San-Lee (1986) โดยให้ k มีค่าเท่ากับ λ_p เมื่อ λ_p เป็นค่า eigen value ที่มีค่าต่ำที่สุดของ $(X'X)$

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \,tr(X'X)^{-1}$$

$$= \beta'\beta + \sigma^2 \sum(1/\lambda_i)$$

จะได้ว่า $E(\hat{\beta}'\hat{\beta}) > \beta'\beta + \sigma^2/\lambda_p$

เมื่อ λ_i เป็นค่าไอเก็นของ $(X'X)$ และ λ_p เป็นค่าไอเก็น ที่มีค่าน้อยที่สุด $(\lambda_p=\lambda_{min})$

ในการวิเคราะห์พิจารณาเฉพาะ λ_{min} ก็เพียงพอที่จะทำให้เห็นข้อเท็จจริงได้ และ λ_{min} ยิ่งมีค่าต่ำเพียงใด $E(\hat{\beta}'\hat{\beta})$ ก็ยิ่งมีความยาวโดยเฉลี่ยสูงกว่าความยาวของ $\beta'\beta$ มากเท่านั้น และเพราะว่า

$$MSE(\hat{\beta}^*(k)) = E[\hat{\beta}^*(k) - \beta]' [\hat{\beta}^*(k)-\beta]$$

$$MSE(\hat{\beta}^*(k)) = VAR(\hat{\beta}^*(k)) + BIAS^2(\hat{\beta}^*(k))$$

$$= \sigma^2 \sum \lambda_i/(\lambda_i+k)^2 + k^2 \hat{\beta}'(X'X + kI)^{-2} \hat{\beta}$$

จะได้ว่าค่า $Bias^2$ มีค่าเพิ่มขึ้นเมื่อ k มีค่าเพิ่มขึ้น ดังนั้นจึงควรเลือกค่า k ที่มีค่าน้อย เพื่อทำให้ $bias^2$ มีค่าน้อย กล่าวคือ

$$k = \lambda_p$$

$$= \lambda_{min}$$

2.5.3 วิธี HK (Hoerl & Kennard Method)

วิธีการนี้เสนอโดย Hoerl และ Kennard (1970) โดยใช้

$$k = \sigma^2 / \max(\hat{\beta}_i^2)$$

ซึ่งสามารถแสดงได้ดังนี้

(3.5.1) $$\hat{\beta}^* = [D + k]^{-1} X'y$$

เมื่อ $\hat{\beta}^*$ เป็นตัวประมาณค่าพารามิเตอร์ β ขนาด p X 1

D เป็นไดออร์ซอโกนอลเมตริกซ์ที่สมาชิกทุกตัวมีค่าเป็น λ_i โดยที่ λ_i เป็นค่าไอเก็นของ X'X เมื่อ i = 1,..,p และเมตริกซ์ D มีขนาด n X n

k เป็นค่าคงที่มากกว่าศูนย์

W.J. Hermerle และ T.F. Brantle (1978) ให้ข้อสังเกตไว้คือ การเลือก k ที่เหมาะสมนั้นได้จากการหาค่าต่ำสุดของค่าเฉลี่ยของระยะทางจาก $\hat{\beta}^*$ ไปยัง β กำลังสอง $E[(\hat{\beta}^*-\beta)'(\hat{\beta}^*-\beta)]$ ดังนี้คือ

จากสมการ (3.5.1) ค่าเฉลี่ยของระยะทางของตัวประมาณค่า $\hat{\beta}^*$ ห่างจากค่าพารามิเตอร์ β กำลังสองคือ

(2.5.2) $$E[(\hat{\beta}^*-\beta)'(\hat{\beta}^*-\beta)] = Q$$

และจากสมการ (2.5.2) William J. Hermerly ได้มีการพิสูจน์การหาค่า k ที่เหมาะสมดังนี้คือเทอม $E[(\hat{\beta}^*-\beta)'(\hat{\beta}^*-\beta)]$ สามารถกระจายให้อยู่ในรูป $\lambda_i, \hat{\beta}_i, k$ และ σ^2 ได้

(2.5.3) $$Q = \Sigma(\sigma^2 \lambda_i + \hat{\beta}_i^2 k_i^2)/(\lambda_i + k_i)^2$$

เพื่อคำนวณค่า k ที่เหมาะสมจะหาค่าต่ำสุดของสมการ (2.5.3) ได้โดยหาอนุพันธ์ Q เทียบ k และกำหนดให้มีค่าเท่ากับศูนย์ จะได้ว่า

$$\frac{(\lambda_i+k_i)(2k_i \hat{\beta}_i^2) - (\sigma^2\lambda_i+\hat{\beta}_i^2 k_i^2)(2\lambda_i+2k_i)}{(\lambda_i+k_i)^4} = 0$$

$$\frac{2\lambda_i(\lambda_i + k_i)(k_i\hat{\beta}_i^2 - \sigma^2)}{(\lambda_i+k_i)^4} = 0$$

$$k_i = \sigma^2 / \hat{\beta}_i^2 \quad ; i = 1,..,P$$

จะได้ว่าค่า k เป็นฟังก์ชันที่ขึ้นกับค่า σ^2 และ $\hat{\beta}_i$ เนื่องจาก $\hat{\beta}_i$ และ k_i มีหลายค่าเพื่อความสะดวกในการคำนวณค่า k เราอาจกำหนดค่า k เพียงค่าเดียวที่สามารถปรับเมตริกซ์ X′X ที่มีผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุด คือ

$$k_i = \sigma^2 / \max(\hat{\beta}_i^2)$$