บทที่ 2

ทฤษฎีและสถิติที่ใช้ในการวิจัย

การศึกษาเปรียบเทียบวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยที่ดีที่สุดในการวิจัยครั้งนี้เป็นการศึกษาเพื่อดูแนวโน้มว่าวิธีคัดเลือกตัวแบบถดถอยที่ดีที่สุดเชิงเบส์จะมีความเหมาะสมมากกว่าวิธีพื้นฐานและให้ค่าพยากรณ์ที่ถูกต้องและแม่นยำมากที่สุด โดยทำการศึกษาเปรียบเทียบวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ 3 วิธี คือ

1. วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC^3})

2. วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (Optimal Predictive Model Selection : median probability model (OPM))

3. วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

การนำเสนอแนวคิดและทฤษฎีในบทนี้จะประกอบด้วยแนวคิดและทฤษฎีที่เกี่ยวกับการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ และทฤษฎีของวิธีการคัดเลือกตัวแบบทั้ง 3 วิธีข้างต้น โดยมีรายละเอียดดังนี้

2.1 การวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis)[1]

ตัวแบบและข้อตกลงเบื้องต้นที่ใช้ในการวิจัยครั้งนี้คือตัวแบบความถดถอยเชิงเส้นพหุคูณซึ่งมีรูปแบบทั่วไปดังนี้

(2.1) $$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i \qquad ; i = 1,2,\ldots,n$$

เมื่อ β_j เป็นสัมประสิทธิ์การถดถอยที่ j ; $j = 1,2,\ldots,p$

ข้อตกลงเบื้องต้นของการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณมีดังนี้

1. ε_i มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น σ^2 คือ ε_i จ.ม.อ. (i.i.d.) $N(0,\sigma^2)$ กล่าวคือ ε_i และ ε_j , $i \neq j$ มีการแจกแจงที่เหมือนกันและเป็นอิสระซึ่งกันและกัน ดังนั้น $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$

2. รูปแบบการถดถอยเป็นแบบเชิงเส้นของพารามิเตอร์

3. ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่และไม่มีพหุสัมพันธ์กัน

---

[1] นิทัศน์ สุขสุวรรณ, การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2545), หน้า 11.

เนื่องจากตัวแบบการถดถอยที่ใช้ในการพิจารณาเป็นตัวแบบเชิงเส้น ดังนั้นในการประมาณค่าพารามิเตอร์จะใช้วิธีกำลังสองน้อยสุดเพื่อประมาณค่าเวกเตอร์พารามิเตอร์สัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}$ ของตัวแบบความถดถอยเชิงเส้นพหุคูณ ซึ่งตัวประมาณค่ากำลังสองน้อยสุดเป็นตัวประมาณไม่เอนเอียงและเป็นวิธีที่ไม่จำเป็นต้องอาศัยข้อตกลงเกี่ยวกับการแจกแจงของตัวแปรสุ่ม ซึ่งมีรูปแบบตัวประมาณดังนี้

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

2.2 วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA))[2]

วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) สำหรับตัวแบบการถดถอยเชิงเส้น เสนอโดย ราฟเทอร์รี เมดิแกน และโฮเอ็ททิง (Raftery, Madigan and Hoeting, 1997) เป็นวิธีการพิจารณาหาค่าความน่าจะเป็นภายหลัง (posterior probability) สำหรับทุกๆ ตัวแบบที่เราสนใจ และนำตัวแบบทุกตัวแบบที่เราสนใจมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสม

วิธีการเฉลี่ยตัวแบบของเบส์นี้เป็นการพิจารณาโดยคำนึงถึงความไม่แน่นอนของตัวแบบ (model uncertainty) และมีแนวคิดว่าการคัดเลือกตัวแบบที่ดีที่สุดเพียงตัวแบบเดียวจากทุกตัวแบบที่เป็นไปได้ ถือเป็นการละเลยตัวแบบอื่นๆ ซึ่งบ่อยครั้งเราจะพบว่ามีตัวแบบหลายรูปแบบที่มีประสิทธิภาพในการพยากรณ์ใกล้เคียงกัน แต่วิธีการคัดเลือกตัวแบบและวิธีการคัดเลือกตัวแปรอิสระที่นิยมใช้กันอย่างแพร่หลายในปัจจุบันเป็นการนำตัวแบบเพียงรูปแบบเดียวมาใช้ในการพยากรณ์เท่านั้น ดังนั้นเมื่อพิจารณาถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบแล้ววิธีการเฉลี่ยตัวแบบของเบส์น่าจะทำให้ได้ค่าพยากรณ์ที่มีประสิทธิภาพมากกว่าค่าพยากรณ์ที่ได้จากตัวแบบเดียว ซึ่งจากงานวิจัยของนักสถิติหลายท่านได้ชี้ให้เห็นว่าความไม่แน่นอนของตัวแบบเป็นสิ่งที่สำคัญและควรนำมาพิจารณาในการคัดเลือกตัวแบบสำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ

การแจกแจงความน่าจะเป็นภายหลังของสิ่งที่สนใจ เมื่อมีข้อมูลจะเป็นดังนี้

$$p(\Delta \mid D) = \sum_{k=1}^{K} p(\Delta \mid M_k, D) \cdot p(M_k \mid D) \tag{2.2}$$

เมื่อ Δ เป็นปริมาณของสิ่งที่สนใจ เช่น ค่าพยากรณ์ที่สนใจ

---

[2] จิตติมา ผสมญาติ, <u>การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ</u>, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2546), หน้า 9-12.

D เป็นข้อมูลของตัวแบบที่สนใจ

และ $M_1, M_2, ..., M_k$ เป็นตัวแบบที่พิจารณา

สมการ (2.2) เป็นการเฉลี่ยการแจกแจงความน่าจะเป็นภายหลัง ซึ่งถ่วงน้ำหนักด้วยความน่าจะเป็นภายหลังของแต่ละตัวแบบ โดยเรียกแนวคิดนี้ว่า"การเฉลี่ยตัวแบบของเบส์"(Bayesian Model Averaging Method (BMA))

ความน่าจะเป็นภายหลังของตัวแบบ M_k คือ

(2.3) $$p(M_k \mid D) = \frac{p(D \mid M_k) \cdot p(M_k)}{\sum_{l=1}^{K} p(D \mid M_l) \cdot p(M_l)}$$

(2.4) เมื่อ $p(D \mid M_k) = \int p\left(D \mid \underset{\sim}{\theta}_k, M_k\right) \cdot p\left(\underset{\sim}{\theta}_k \mid M_k\right) d\underset{\sim}{\theta}_k$ เป็นความควรจะเป็นขอบ (marginal likelihood) ของตัวแบบ M_k

โดยที่ $\underset{\sim}{\theta}_k$ เป็นเวกเตอร์ของพารามิเตอร์สำหรับการวิเคราะห์ความถดถอย หมายถึง $\underset{\sim}{\beta}$ และ σ^2

$p(\underset{\sim}{\theta}_k \mid M_k)$ เป็นความหนาแน่นของ $\underset{\sim}{\theta}_k$ ภายใต้ตัวแบบ M_k

$p(D \mid \underset{\sim}{\theta}_k, M_k)$ เป็นฟังก์ชันความควรจะเป็น

และ $p(M_k)$ เป็นความน่าจะเป็นก่อนสำหรับตัวแบบ M_k

ค่าเฉลี่ยและความแปรปรวนภายหลังของสิ่งที่สนใจ Δ เป็นดังนี้

(2.5) $$E[\Delta \mid D] = \sum_{k=0}^{K} \hat{\Delta}_k p(M_k \mid D)$$

(2.6) $$Var[\Delta \mid D] = \sum_{k=0}^{K} \left(Var[\Delta \mid D, M_k] + \hat{\Delta}_k^2 \right) p(M_k \mid D) - E[\Delta \mid D]^2$$

เมื่อ $\hat{\Delta}_k = E[\Delta \mid D, M_k]$

นั่นคือจากสมการ (2.5) เราสามารถสรุปได้ว่า ถ้าสิ่งที่เราสนใจ Δ เป็นค่าพยากรณ์ $\hat{y}$ ค่าคาดหมายของค่าพยากรณ์ $\hat{y}$ เมื่อมีข้อมูลคือการเฉลี่ยค่าพยากรณ์ของแต่ละตัวแบบถ่วงน้ำหนักด้วยความน่าจะเป็นภายหลังของแต่ละตัวแบบ M_k จะทำให้เราได้ค่าพยากรณ์ $\hat{y}$ ตามที่ต้องการ

นอกจากนี้ในงานวิจัยของราฟเทอร์รีและเมดิแกน (Raftery and Madigan, 1994) ยังแสดงให้เห็นว่าการเฉลี่ยตัวแบบจะทำให้การพยากรณ์มีประสิทธิภาพสูงกว่าการใช้ตัวแบบเดียวในการ

พยากรณ์ โดยใช้กฎคะแนนลอการิทึมและแนวคิดเกี่ยวกับรูปแบบข้อสนเทศของคูลล์แบค-ไลเบอร์ (Kullback-leibler, 1951) จะได้ว่า

$$-E\left[\log\left\{\sum_{k=1}^{K} p(\Delta \mid M_k, D)\cdot p(M_k \mid D)\right\}\right] \le -E\left[\log\left\{p(\Delta \mid M_j, D)\right\}\right] \quad , j=1,2,\ldots,K$$

ซึ่งเป็นการยืนยันว่าวิธีการเฉลี่ยตัวแบบจะมีประสิทธิภาพในการพยากรณ์สูงกว่าการใช้ตัวแบบเดียวในการพยากรณ์

ในการหาความน่าจะเป็นภายหลัง (posterior probability) ของแต่ละตัวแบบ จำเป็นต้องทราบถึงความน่าจะเป็นก่อน (prior probability) และความน่าจะเป็นภายหลังขอบ (marginal posterior probability) ของตัวแบบ ซึ่งมีรายละเอียดดังนี้

2.2.1 <u>ความน่าจะเป็นก่อน (prior probability) ของตัวแบบ</u>

ความน่าจะเป็นก่อนของตัวแบบ จะอยู่ในรูปแบบดังนี้

(2.7) $$p(M_i) = \prod_{j=1}^{k} p_j^{\delta_{ij}} (1-p_j)^{1-\delta_{ij}}$$

เมื่อ p_j เป็นความน่าจะเป็นที่ตัวแปรอิสระที่ j จะรวมอยู่ในตัวแบบ M_i

และ δ_{ij} เป็นตัวบ่งชี้ที่มีค่า 0 หรือ 1

โดยที่ δ_{ij} เป็น 0 เมื่อตัวแปรอิสระที่ j ไม่อยู่ในตัวแบบ M_i

และ δ_{ij} เป็น 1 เมื่อตัวแปรอิสระที่ j อยู่ในตัวแบบ M_i

สำหรับการวิจัยครั้งนี้จะกำหนด p_j สำหรับทุก ๆ ตัวแปรอิสระเป็น $\frac{1}{2}$ นั่นคือตัวแปรอิสระทุกตัวมีโอกาสที่จะอยู่ในตัวแบบหรือไม่อยู่ในตัวแบบเท่า ๆ กัน

2.2.2 <u>ความน่าจะเป็นภายหลังขอบ (marginal posterior probability) ของตัวแบบ</u>

เนื่องจากความน่าจะเป็นภายหลังขอบของตัวแบบ M ใดๆ จะแปรผันตามฟังก์ชันความควรจะเป็นคูณกับการแจกแจงก่อนของพารามิเตอร์ กล่าวคือ

$$p(D \mid M) \;\alpha\; likelihood \times prior$$

ดังนั้นการวิจัยครั้งนี้ผู้วิจัยได้กำหนดการแจกแจงก่อนคู่สังยุค (conjugate prior distribution) สำหรับ $\underset{\sim}{\beta}$ เป็นการแจกแจงแบบแกมมา (gamma distribution) เมื่อ a คือพารามิเตอร์รูปร่าง (shape parameter) และ b คือพารามิเตอร์ขนาด (scale parameter) โดยกำหนดค่าพารามิเตอร์ $a = 2$ และ $b = \frac{1}{2}$ เพราะมีลักษณะการกระจายข้อมูลเป็นการแจกแจง

แบบแกมมา โดยที่ข้อมูลส่วนใหญ่กระจายตัวอยู่ใกล้ค่าศูนย์ ซึ่งแสดงถึงเหตุการณ์ที่ค่าที่มีค่าน้อยๆ หรือความผิดพลาดน้อยๆ จะเกิดขึ้นได้มาก ดังนั้นผู้วิจัยได้กำหนดการแจกแจงก่อนเป็นการแจกแจงแบบแกมมาและมีการแจกแจงความน่าจะเป็นภายหลังเป็นการแจกแจงแบบแกมมาจึงเป็นหลักเกณฑ์คู่สังยุคแบบแกมมา (conjugate gamma prior distribution)

เนื่องจากการแจกแจงความน่าจะเป็นภายหลังสำหรับตัวแบบความถดถอย M ใดๆ จะแปรผันตามฟังก์ชันความควรจะเป็นคูณกับการแจกแจงก่อนของพารามิเตอร์ในตัวแบบ กล่าวคือ

$$p(D \mid M) \quad \alpha \quad l\left(\underset{\sim}{\beta} \mid \underset{\sim}{y}\right) \times p(\underset{\sim}{\beta})$$

การวิจัยครั้งนี้ผู้วิจัยได้กำหนดการแจกแจงก่อนเป็นหลักเกณฑ์แบบคู่สังยุคแกมมา ดังนั้นการแจกแจงก่อนจะเป็นการแจกแจงแบบแกมมา คือ

$$\begin{aligned} p(\underset{\sim}{\beta}) \quad &\alpha \quad Gamma(a,b) \\ &\alpha \quad \underset{\sim}{\beta}^{a-1} \exp(-b\underset{\sim}{\beta}) \end{aligned}$$

เมื่อ a คือพารามิเตอร์รูปร่าง (shape parameter)

b คือพารามิเตอร์ขนาด (scale parameter)

ดังนั้นการแจกแจงความน่าจะเป็นภายหลังของตัวแบบความถดถอย M คือ

$$p(D \mid M) \quad \alpha \quad e^{-\underset{\sim}{\beta}(b+n)} \underset{\sim}{\beta}^{(a+\sum y_i - 1)}$$

$$p(D \mid M) \quad \alpha \quad Gamma\left(a + \sum_{i=1}^{n} y_i, b+n\right)$$

ซึ่งสามารถศึกษาการแจกแจงความน่าจะเป็นก่อน การแจกแจงความน่าจะเป็นภายหลัง การแจกแจงแบบคู่สังยุคได้จากทฤษฎีของเบส์ โดยรายละเอียดดูได้จากภาคผนวก

2.3 วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) เป็นการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC^3})[3]

การคัดเลือกตัวแปรของวิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลเมื่อใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition

---

[3] นิทัสน์ สุขสุวรรณ, <u>การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ</u>, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2545), หน้า 24-26.

$\left(MC^3\right)$) $\left(BMA_{MC^3}\right)$ มีขั้นตอนเหมือนวิธีการคัดเลือกตัวแปรของเบส์เซียนแต่วิธี BMA_{MC^3} จะนำตัวแบบในทุกรูปแบบที่ปรากฏอยู่ในปริภูมิตัวแบบมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก โดยรายละเอียดของขั้นตอนการคัดเลือกตัวแปรของเบส์เซียนมีดังนี้

ขั้นตอนแรกของวิธีการนี้เริ่มต้นจากสมการถดถอยเต็มรูป คือ ประกอบด้วยตัวแปรอิสระครบทุกตัว แล้วทำการสุ่มแบบกิบส์ซึ่งเป็นกรณีหนึ่งของโครงสร้างแบบลูกโซ่มาร์คอฟเพื่อสร้างลำดับ

(2.8) $$\delta^1, \delta^2, \ldots, \delta^j, \ldots$$

ลำดับในสมการ (2.8) จะลู่เข้าสู่การแจกแจงของ $\underset{\sim}{\delta} \sim f\left(\underset{\sim}{\delta} \mid \underset{\sim}{y}\right)$ ซึ่งเป็นลำดับที่มีความน่าจะเป็นค่อนข้างสูงเพราะบรรจุข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปร โดย $\underset{\sim}{\delta}^j$ ใดๆ ที่มีความน่าจะเป็นสูงสุด ซึ่งทำให้ง่ายต่อการตัดสินใจในการเลือกตัวแบบที่ดีที่สุด

เราใช้วิธีการสุ่มแบบกิบส์เพื่อสร้างเวกเตอร์เป็นลำดับแบบกิบส์ ดังนี้

(2.9) $$\underset{\sim}{\beta}^0, \sigma^0, \underset{\sim}{\delta}^0, \underset{\sim}{\beta}^1, \sigma^1, \underset{\sim}{\delta}^1, \ldots, \underset{\sim}{\beta}^j, \sigma^j, \underset{\sim}{\delta}^j, \ldots$$

เมื่อ $\underset{\sim}{\beta}^0, \sigma^0$ เป็นค่าเริ่มต้นที่ได้จากการประมาณค่าด้วยวิธีกำลังสองน้อยสุดของสมการ (2.1) และ $\underset{\sim}{\delta}^0$ ถูกกำหนดค่าเริ่มต้นให้เป็น $\underset{\sim}{\delta}^0 = (1,1,\ldots,1)'$

ในแต่ละรอบของการสุ่มค่าของลำดับ $\underset{\sim}{\beta}^j, \sigma^j, \underset{\sim}{\delta}^j$ ได้จากการสร้างค่าที่สอดคล้องกับขั้นตอนต่อไปนี้

1. สุ่มเวกเตอร์ของสัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}^j$ จากการแจกแจง

(2.10) $$\underset{\sim}{\beta}^j \mid \underset{\sim}{y}, \sigma^{j-1}, \underset{\sim}{\delta}^{j-1} \sim N_p\left(A_{\delta^{j-1}}\left(\sigma^{j-1}\right)^{-2} X'X \underset{\sim LS}{\beta}, A_{\delta^{j-1}}\right)$$

เมื่อ $A_{\delta^{j-1}} = \left(\left(\sigma^{j-1}\right)^{-2} X'X + D_{\delta^{j-1}}^{-1} R^{-1} D_{\delta^{j-1}}^{-1}\right)^{-1}$

และ $D_{\delta^{j-1}}^{-1} = diag[(a_1\tau_1)^{-1}, (a_2\tau_2)^{-1}, \ldots, (a_p\tau_p)^{-1}]$

2. สุ่มค่าความแปรปรวน σ^j จากการแจกแจง

(2.11) $$\sigma^j \mid \underset{\sim}{y}, \underset{\sim}{\beta}^j, \underset{\sim}{\delta}^{j-1} \sim IG\left(\frac{n+\upsilon_{\delta^{j-1}}}{2}, \frac{\left| \underset{\sim}{y} - X\underset{\sim}{\beta}^j \right|^2 + \upsilon_{\delta^{j-1}}\lambda_{\delta^{j-1}}}{2} \right)$$

3. ค่าของเวกเตอร์ $\underset{\sim}{\delta}^j$ ได้จากการแยกองค์ประกอบ (componentwise) โดยการสุ่มอย่างต่อเนื่องกัน (sampling consecutively) จากการแจกแจงแบบมีเงื่อนไขดังนี้

(2.12) $$\delta_i^j \sim f\left(\delta_i^j \mid \underset{\sim}{y}, \underset{\sim}{\beta}^j, \sigma^j, \delta_{-i}^j \right) = f\left(\delta_i^j \mid \underset{\sim}{\beta}^j, \sigma^j, \delta_{-i}^j \right)$$

เมื่อ $\delta_{-i}^j = \left(\delta_i^j, \dots, \delta_{i-1}^j, \delta_{i+1}^j, \dots, \delta_k^j \right)$

จะเห็นได้ว่าการแจกแจงในสมการ (2.13) จะไม่ขึ้นอยู่กับค่าของ $\underset{\sim}{y}$

ในสมการ (2.13) δ_i^j แต่ละตัวจะมีการแจกแจงแบบแบร์นูลลี (Bernoulli distribution) ด้วยความน่าจะเป็น

(2.13) $$P\left(\delta_i^j = 1 \mid \underset{\sim}{\beta}^j, \sigma^j, \delta_{-i}^j \right) = \frac{a}{a+b}$$

(2.14) เมื่อ $a = f\left(\underset{\sim}{\beta}^j \mid \delta_{-i}^j, \delta_i^j = 1 \right) \cdot f\left(\sigma^j \mid \delta_{-i}^j, \delta_i^j = 1 \right) \cdot f\left(\delta_{-i}^j, \delta_i^j = 1 \right)$

(2.15) และ $b = f\left(\underset{\sim}{\beta}^j \mid \delta_{-i}^j, \delta_i^j = 0 \right) \cdot f\left(\sigma^j \mid \delta_{-i}^j, \delta_i^j = 0 \right) \cdot f\left(\delta_{-i}^j, \delta_i^j = 0 \right)$

ความยาวของลำดับในสมการ (2.8) จะเพิ่มขึ้นเรื่อย ๆ ทำให้การแจกแจงของค่าที่แท้จริงของ $\underset{\sim}{\delta}$ จะลู่เข้าสู่การแจกแจงภายหลัง $f\left(\underset{\sim}{\delta} \mid \underset{\sim}{y} \right)$

การลู่เข้านี้จะเป็นไปอย่างรวดเร็วถ้า $f\left(\underset{\sim}{\delta} \mid \underset{\sim}{y} \right)$ เป็นค่าสูงสุด ตัวแบบที่มีน้ำหนักมาก ๆ มีจำนวนไม่มาก ตัวแบบเหล่านี้จะมีความแม่นยำสูง เมื่อ $f\left(\underset{\sim}{\delta} \mid \underset{\sim}{y} \right)$ บรรจุข้อมูลส่วนใหญ่ของการคัดเลือกตัวแปร

ข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปรจะอยู่ในลำดับของสมการ (2.9) ค่าของ $\underset{\sim}{\delta}$ ที่สอดคล้องกับกลุ่มย่อยของ $x_1, x_2, \dots, x_k$ จะมีความถี่สูงสุด เพราะว่าลำดับย่อยนี้จะเหมาะสมกับตัวแบบที่มีค่าความน่าจะเป็นภายหลังสูงสุดภายใต้ $f\left(\underset{\sim}{\delta} \mid \underset{\sim}{y} \right)$ กล่าวคือ ตัวแบบที่เหมาะสมที่ได้จากวิธีการคัดเลือกตัวแปรของเบส์เซียนจะเป็นตัวแบบที่มีความถี่สูงสุดและจะมีค่าความน่าจะเป็นภายหลังสูงสุดด้วย แต่วิธี BMA_{MC^3} จะนำตัวแบบในทุกรูปแบบที่ปรากฏอยู่ในปริภูมิตัวแบบมา

เฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสมต่อไป

2.4 วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (Optimal Predictive Model Selection : median probability model (OPM))[4]

วิธีการนี้เสนอโดยบาร์บีรีและเบอเกอร์ (Barbieri and Berger, 2002) ซึ่งมีแนวคิดว่าตัวแบบที่เหมาะสมที่สุดในการพยากรณ์อาจไม่ใช่ตัวแบบที่มีความน่าจะเป็นภายหลังสูงสุด (the model with highest posterior probability) แต่กลับเป็นตัวแบบที่มีความน่าจะเป็นภายหลังกลางๆ (the median probability model) ซึ่งใกล้เคียงกับวิธีการเฉลี่ยตัวแบบของเบส์เมื่อพิจารณาความไม่แน่นอนของตัวแบบ และเกณฑ์ที่ใช้ในการเลือกตัวแบบที่เหมาะสมที่สุดจากกลุ่มของตัวแบบที่มีความน่าจะเป็นภายหลังกลางๆ คือความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสอง (square error loss) โดยจะเลือกตัวแบบที่มีค่าความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสองต่ำสุด

2.4.1 <u>การค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังกลางๆ</u>

การค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังกลาง ๆ กระทำโดยใช้แผนการค้นหาตัวแบบด้วยเทคนิคลูกโซ่มาร์คอฟเชนมอนติคาร์โล (ordinary MCMC model search schemes) แผนการนี้ได้พัฒนาลูกโซ่มาร์คอฟให้เคลื่อนไประหว่างตัวแบบด้วยค่าความน่าจะเป็นภายหลังของตัวแบบนั้น ๆ ซึ่งคำนวณค่าความน่าจะเป็นภายหลังได้จากสัดส่วนของจำนวนครั้งที่ตัวแบบอยู่ในลูกโซ่ การค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังกลาง ๆ จะพิจารณาค่าความน่าจะเป็นภายหลังสำหรับตัวแปรอิสระใด ๆ เป็นเกณฑ์ในการคัดเลือก กล่าวคือ เมื่อจบสิ้นกระบวนการ MCMC เราจะทำการเลือกตัวแปรอิสระใด ๆ ที่มีค่าสัดส่วนของจำนวนครั้งที่ตัวแปรอิสระดังกล่าวอยู่ในตัวแบบมากกว่า $\frac{1}{2}$

หากกำหนดให้สมการ (2.1) เป็นสมการของตัวแบบเต็ม เราจะได้สมการของตัวแบบย่อย (submodels) ที่นำมาพิจารณาเป็นดังสมการ (2.16)

(2.16) $$M_{\underset{\sim}{l}} : \underset{\sim}{y} = X_{\underset{\sim}{l}} \underset{\sim}{\beta}_{\underset{\sim}{l}} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{l} = (l_1, l_2, ..., l_p)$ เป็นดัชนีของตัวแบบซึ่งบอกว่าตัวแปรอิสระใดควรอยู่ในตัวแบบ

---

[4]จิตติมา ผสมญาติ, <u>การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ</u>, (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2546), หน้า 16-19.

หรือ $l_j = \begin{cases} 1 & \beta_j \neq 0 \\ 0 & \beta_j = 0 \end{cases}$

และค่าพยากรณ์ที่ได้จะอยู่ในรูป

(2.17) $$\underset{\sim}{y}^* = \underset{\sim}{X}^* \beta + \underset{\sim}{\varepsilon}$$

เมื่อ $X^* = \left(X_1^*, ..., X_p^*\right)$ เป็นเวกเตอร์ของตัวแปรอิสระที่ใช้ในการพยากรณ์

ค่าพยากรณ์ที่ได้จะมีค่าความสูญเสียอันเกิดจากการพยากรณ์ $\underset{\sim}{y}^*$ ด้วย $\underset{\sim}{\hat{y}}^*$ เป็น

(2.18) $$L\left(\underset{\sim}{\hat{y}}^*, \underset{\sim}{y}^*\right) = \left(\underset{\sim}{\hat{y}}^* - \underset{\sim}{y}^*\right)^2$$

<u>นิยามที่ 1</u>

ผลรวมของความน่าจะเป็นภายหลังสำหรับตัวแปรอิสระที่ j (posterior inclusion probability for variable j)

(2.19) $$p_j = \sum_{\underset{\sim}{l}:l_j=1} P\left(M_{\underset{\sim}{l}} \mid \underset{\sim}{y}\right)$$

ดังนั้น ตัวแปรอิสระที่ j จะอยู่ในตัวแบบก็ต่อเมื่อ $p_j \geq \frac{1}{2}$ และจะได้ตัวแบบที่มีความน่าจะเป็นกลาง ๆ (median probability model : $M_{\underset{\sim}{l}}^*$) เป็นตัวแบบที่ประกอบด้วยตัวแปรอิสระที่มี p_j อย่างน้อยเท่ากับ $\frac{1}{2}$ หรือ

$$l_j^* = \begin{cases} 1, p_j \geq \dfrac{1}{2} \\ 0, otherwise \end{cases}$$

2.4.2 <u>การคัดเลือกตัวแบบที่เหมาะสมที่สุด</u>

การคัดเลือกตัวแบบที่เหมาะสมที่สุด ภายหลังการค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังกลาง ๆ เราจะทำการคัดเลือกตัวแบบที่เหมาะสมที่สุดและเกณฑ์ที่ใช้ในการคัดเลือกตัวแบบที่เหมาะสมที่สุดจากกลุ่มของตัวแบบที่มีความน่าจะเป็นภายหลังกลาง ๆ คือ ความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสอง (square error loss) โดยจะเลือกตัวแบบที่มีค่าดังกล่าวต่ำสุด ซึ่งรายละเอียดของการคัดเลือกตัวแบบที่เหมาะสมที่สุดมีดังนี้

สมมติว่า $Q = X'X$ เป็นเมทริกซ์ทแยงมุม เราจะได้ค่าเฉลี่ยภายหลัง $\widetilde{\beta}_{\underset{\sim}{l}}$ (the posterior means) เป็นดังนี้

(2.20) $$\widetilde{\beta}_{\underset{\sim}{l}} = H_{\underset{\sim}{l}} \widetilde{\beta}$$

เมื่อ $\widetilde{\beta}$ เป็นค่าเฉลี่ยภายหลังของตัวแบบเต็ม

และ H_l เป็นเมทริกซ์ขนาด $p \times p_l$ ของคู่ลำดับ (j,k) ที่มีค่าเป็น 1 ถ้า $l_j = 1$ และ $k = \sum_{r-1}^{i} l_r$ หรือมีค่าเป็น 0 ในกรณีอื่น ๆ

ดังนั้นถ้า Q เป็นเมทริกซ์ทแยงมุมซึ่งประกอบด้วยค่าในแนวทแยง $q_j > 0$ และสมการ (2.20) เป็นจริงแล้วเราสามารถหาค่าความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสองของตัวแบบย่อยใด ๆ ได้ดังสมการ

(2.21) $$R\left(M_l\right) = \sum_{j=1}^{p} \widetilde{\beta}_j^2 q_j \left(l_j - p_j\right)^2$$

เมื่อ p_j เป็นผลรวมของความน่าจะเป็นภายหลังสำหรับตัวแปรอิสระที่ j ดังสมการ (2.19)

2.5 วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))[5]

วิธีการนี้เป็นวิธีที่ปรับปรุงตัวแบบที่นำเอาทั้งการคัดเลือกตัวแปรอิสระแบบเลือกเข้าและวิธีการกำจัดตัวแปรอิสระแบบถอยหลังมาผสมผสานกัน กล่าวคือเมื่อตัวแปรอิสระได้รับการคัดเลือกและนำเข้าสู่สมการแล้วตัวแปรอิสระทุกตัวในสมการจะต้องผ่านการพิจารณาเพื่อคัดออก ตัวแปรที่ถูกคัดออกถือว่าเป็นตัวแปรส่วนเกินเพราะไม่มีนัยสำคัญ ขั้นตอนของวิธีการถดถอยแบบขั้นบันไดมีดังนี้

1. เริ่มนำตัวแปรอิสระที่มีสหสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสู่สมการแล้ววัดค่า R^2 และตรวจสอบนัยสำคัญด้วย t – test หรือ F - test

2. เพิ่มตัวแปรอิสระอื่นเข้าสู่สมการในข้อ 1 ครั้งละ 1 ตัว ถ้าพบว่าตัวแปรดังกล่าว มีคุณสมบัติสอดคล้องกับเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้

2.1 มีค่า Partial Correlation กับตัวแปรตามสูงที่สุด

2.2 เพิ่มค่า R^2 สูงกว่าตัวแปรอิสระอื่น ๆ

2.3 มีค่า t หรือ Partial F สูงกว่าตัวแปรอิสระอื่นโดยที่ t > ค่า t จากตาราง และ Partial F > $F_{1-\alpha,1,n-p}$

3. ในทุกครั้งที่เพิ่มตัวแปรอิสระเข้าสู่สมการอันมีผลให้ในสมการมีจำนวนตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไปให้พิจารณาคัดตัวแปรอิสระที่อาจเป็นส่วนเกินทิ้งไป

ตัวแปรที่จะต้องถูกคัดทิ้งคือตัวแปรที่มีคุณสมบัติสอดคล้องกับเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้

---

[5]มนตรี พิริยะกุล, เทคนิคการวิเคราะห์สมการถดถอย เล่ม 1, พิมพ์ครั้งที่ 2 (กรุงเทพมหานคร : ศรีเมืองการพิมพ์, 2532), หน้า 232 – 234.

3.1 มีค่า Partial Correlation กับตัวแปรตามต่ำที่สุด

3.2 ไม่ช่วยเพิ่มค่า R^2 ของสมการสูงขึ้น หรือถ้าคัดออกแล้วทำให้ค่า R^2 เดิมลดลงน้อยกว่าการคัดเลือกตัวแปรอิสระอื่น ๆ ทิ้ง

3.3 มีค่า t หรือ Partial F ต่ำกว่าตัวแปรอิสระอื่นโดยที่ t < ค่า t จากตาราง หรือ Partial F < $F_{1-\alpha,1,n-p}$

3.4 หยุดดำเนินการเมื่อพบว่าตัวแปรอิสระภายนอกสมการมิได้มีนัยสำคัญทางสถิติหรือมิได้เพิ่ม R^2

ผู้วิจัยขอนำเสนอรายละเอียดขั้นตอนของวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณทั้ง 3 วิธี พร้อมทั้งยกตัวอย่างเพิ่มเติมแสดงในภาคผนวก