บทที่ 2

ทฤษฎีและสถิติที่ใช้ในการวิจัย

การศึกษาเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์ในการวิจัยครั้งนี้เป็นการพิจารณาหาวิธีการคัดเลือกสมการถดถอยที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด โดยวิธีการคัดเลือกสมการถดถอยที่นำมาศึกษามีดังนี้

1) วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่เมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ(BMA_{SVT})

2) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด(OPM)

3) วิธีการถดถอยแบบขั้นบันได(SR)

การนำเสนอทฤษฎีและสถิติที่เกี่ยวข้องจะประกอบด้วยแนวคิดและทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ความถดถอยเชิงพหุคูณ ความน่าจะเป็นที่มีเงื่อนไข ทฤษฎีบทของเบส์ ความเป็นอิสระ และทฤษฎีของวิธีการคัดเลือกสมการถดถอยทั้ง 3 วิธีข้างต้น โดยมีรายละเอียดดังนี้

2.1 การวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ[1]

ตัวแบบและข้อตกลงเบื้องต้นที่ใช้ในการวิจัยครั้งนี้คือตัวแบบความถดถอยเชิงเส้นพหุคูณซึ่งมีรูปแบบทั่วไปดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i \quad ; i = 1,2,...,n$$

เมื่อ p เป็นจำนวนตัวแปรอิสระ (ใช้สัญลักษณ์ $x_1, x_2, ...x_p$)

β_0 เป็นค่าคงที่ของการถดถอย

และ β_j เป็นสัมประสิทธิ์การถดถอย ; $j = 1,2,..., p$

ข้อตกลงเบื้องต้นของการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณมีดังนี้

1. ความคลาดเคลื่อนสุ่ม(ε_i) มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น σ^2 หรือเขียนได้เป็น ε_i จ.ม.อ. (i.i.d) $N(0, \sigma^2)$ กล่าวคือ ε_i และ ε_j สำหรับ $i \neq j$ มีการแจกแจงที่เหมือนกันและเป็นอิสระซึ่งกันและกัน เพราะฉะนั้น $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ สำหรับ $i \neq j$

2. รูปแบบการถดถอยเป็นแบบเชิงเส้นที่มีพารามิเตอร์ (β_j)

---

[1] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546) หน้า 8-9.

3. ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่และไม่มีพหุสัมพันธ์กัน

เนื่องจากตัวแบบการถดถอยที่ใช้ในการพิจารณาเป็นตัวแบบเชิงเส้น ดังนั้นในการประมาณค่าพารามิเตอร์เราสามารถใช้วิธีกำลังสองน้อยสุดเพื่อประมาณค่าเวกเตอร์พารามิเตอร์สัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}$ ของตัวแบบความถดถอยเชิงเส้นพหุคูณซึ่งมีตัวประมาณค่ากำลังสองน้อยสุด(Least Square estimator($\underset{\sim}{\hat{\beta}}$))อยู่ในรูปของ

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

2.2 ความน่าจะเป็นที่มีเงื่อนไข(Conditional Probability)[2]

ความน่าจะเป็นที่มีเงื่อนไข (Conditional Probability) เป็นการศึกษาหาความน่าจะเป็นของเหตุการณ์ที่สนใจในกรณีที่มีข้อมูลเพิ่มเติมเกี่ยวกับเหตุการณ์นั้น ซึ่งความน่าจะเป็นที่มีเงื่อนไขจะมีค่ามากขึ้นหรือน้อยลงก็ขึ้นอยู่กับข้อมูลที่ได้รับเพิ่มเติมว่าเกี่ยวข้องกับเหตุการณ์ที่สนใจมากหรือน้อยอย่างไร โดยข้อมูลที่ได้รับเพิ่มเติมอาจไม่มีอิทธิพลกับเหตุการณ์ที่สนใจ ถ้าข้อมูลนั้นเป็นอิสระต่อกันกับเหตุการณ์ที่สนใจ สามารถให้นิยามของความน่าจะเป็นที่มีเงื่อนไขดังนี้

<u>นิยามที่ 1</u> ให้ A และ B เป็นเหตุการณ์ซึ่ง $P(A) > 0$ เราเรียก $P(B \mid A)$ ว่าเป็นความน่าจะเป็นที่มีเงื่อนไขของ B เมื่อกำหนด A (the conditional probability of B given A) ถ้า

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \ ; \ P(A) > 0$$

<u>ทฤษฎีบทที่ 1</u> ทฤษฎีการคูณ (multiplication Theorem) ถ้าในการทดลองชนิดหนึ่งเหตุการณ์ A และ B สามารถเกิดขึ้นได้พร้อมกัน จะได้ว่า $P(B \cap A) = P(B \mid A) \cdot P(A)$

<u>ทฤษฎีบทที่ 2</u>

ถ้าในการทดลองชนิดหนึ่งซึ่งเหตุการณ์ $A_1, A_2, \ldots, A_n$ สามารถเกิดขึ้นพร้อมกันได้ จะได้ว่า

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1, A_2) \ldots P(A_n \mid A_1, A_2, \ldots, A_{n-1})$$

2.3 ทฤษฎีบทของเบส์ (Bayes' Theorem)[3]

ทฤษฎีบทของเบส์ (Bayesian Theorem) เป็นการศึกษาเกี่ยวกับความน่าจะเป็นของเหตุการณ์ที่สนใจในกรณีที่มีข้อมูลเพิ่มเติม ซึ่งสามารถให้นิยามของทฤษฎีบทของเบส์ ดังนี้

<u>นิยามที่ 2</u> เหตุการณ์ $B_1, B_2, \ldots$ จะแทนผลแบ่งกั้น (partition) ของปริภูมิตัวอย่าง S ถ้า

1) $B_i \cap B_j = \phi, \forall i \neq j$

---

[2] ธีระพร วีระถาวร, <u>ความน่าจะเป็นกับการประยุกต์</u>, พิมพ์ครั้งที่2 (กรุงเทพมหานคร: วิทยพัฒน์, 2539), หน้า 100-106.

[3] ธีระพร วีระถาวร, <u>ความน่าจะเป็นกับการประยุกต์</u>, พิมพ์ครั้งที่2 (กรุงเทพมหานคร: วิทยพัฒน์, 2539), หน้า 107-108.

2) $\bigcup_{i=1}^{\infty} B_i = S$

3) $P(B_i) > 0, \forall i$

กล่าวคือเมื่อเราทำการทดลอง E จะได้ว่าเหตุการณ์ $B_i, i = 1,2,...,n$ จะไม่เกิดขึ้นพร้อม ๆ กัน

ทฤษฎีบทที่ 3 ทฤษฎีบทของเบส์ (Bayesian Theorem)

ให้ $B_1, B_2,...$ จะแทนผลแบ่งกั้นของปริภูมิตัวอย่าง S และถ้า A เป็นเหตุการณ์ซึ่ง $P(A) > 0$ จะได้ว่า $P(B_k \mid A) = \dfrac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{\infty} P(B_i) \cdot P(A \mid B_i)}, \forall k = 1,2,3,...$

เมื่อ (1) $P(B_k)$ อาจจะแทนความน่าจะเป็นของเหตุการณ์ที่เป็นไปได้ก่อนทราบข้อมูลซึ่งเรียกว่า ความน่าจะเป็นโดยหลักเกณฑ์ หรือความน่าจะเป็นก่อน(prior probability)

(2) $P(B_k \mid A)$ อาจจะแทนความน่าจะเป็นของเหตุการณ์ที่เป็นไปได้หลังทราบข้อมูล ซึ่งเรียกว่า ความน่าจะเป็นโดยประสบการณ์ หรือความน่าจะเป็นภายหลัง (posterior probability)

และ (3) $P(A \mid B_i)$ อาจจะแทนความน่าจะเป็นของเหตุการณ์ A ภายใต้ข้อสมมติว่าส่วนย่อย B_i เกิดขึ้น

2.4 ความเป็นอิสระ (Independence)[4]

ความเป็นอิสระ (Independence) เป็นการศึกษาเกี่ยวกับความเป็นอิสระของเหตุการณ์ตั้งแต่สองเหตุการณ์ขึ้นไป ซึ่งทำให้การหาความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้นพร้อมกันนั้นง่ายขึ้นถ้ามีคุณสมบัติดังกล่าว ซึ่งสามารถให้นิยามของความเป็นอิสระได้ดังนี้

นิยามที่ 3 เหตุการณ์ 2 เหตุการณ์ A และ B เรียกว่าเป็นอิสระต่อกัน (independence) ก็ต่อเมื่อ $P(B \mid A) = P(B)$ และ $P(A \mid B) = P(A)$ กล่าวคือ $P(A \cap B) = P(A) \cdot P(B)$

นิยามที่ 4 เหตุการณ์ 3 เหตุการณ์ A, B และ C เป็นอิสระซึ่งกันและกัน(mutually independent) ก็ต่อเมื่อ

(1) แต่ละคู่ของเหตุการณ์เป็นอิสระต่อกัน(pairwise independent)กล่าวคือ

$$P(A \cap B) = P(A) \cdot P(B)$$
$$P(A \cap C) = P(A) \cdot P(C)$$
$$P(B \cap C) = P(B) \cdot P(C)$$

(2) เหตุการณ์หนึ่งเหตุการณ์ใดเป็นอิสระจากสองเหตุการณ์ใด ๆ กล่าวคือ

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

[4] ธีระพร วีระถาวร, ความน่าจะเป็นกับการประยุกต์, พิมพ์ครั้งที่2 (กรุงเทพมหานคร: วิทยพัฒน์, 2539), หน้า 114-116.

2.5 วิธีการเฉลี่ยตัวแบบของเบส์ โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟเมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปร[5] (BMA_{SVT})

ในการนำเสนอแนวคิดและทฤษฎีของวิธี BMA_{SVT} จำเป็นต้องกล่าวถึงวิธีการดังต่อไปนี้

1. วิธีการเฉลี่ยตัวแบบของเบส์(BMA)
2. ขั้นตอนการคัดเลือกตัวแปรของวิธี BMA_{MC3}
3. ขั้นตอนของการคัดเลือกตัวแปรและการแปลงไปพร้อมๆ กันของวิธี BMA_{SVT}

โดยมีรายละเอียดดังนี้

1. วิธีการเฉลี่ยตัวแบบของเบส์ (BMA)[6]

วิธี BMA สำหรับตัวแบบการถดถอยเชิงเส้นเสนอโดย ราฟเทอร์รี(Raftery) เมดิแกน (Madigan) และโฮเอ็ททิง(Hoeting)เป็นวิธีการพิจารณาหาค่าความน่าจะเป็นภายหลัง (posterior probability) สำหรับทุก ๆ ตัวแบบที่เราสนใจ และนำตัวแบบทุกตัวแบบที่เราสนใจมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนักเพื่อหาค่าพยากรณ์ที่เหมาะสม

วิธีการเฉลี่ยตัวแบบของเบส์นี้เป็นการพิจารณาโดยคำนึงถึงความไม่แน่นอนของตัวแบบ (model uncertainty) และมีแนวคิดว่าการคัดเลือกตัวแบบที่ดีที่สุดเพียงตัวแบบเดียวจากทุกตัวแบบที่เป็นไปได้ถือเป็นการละเลยตัวแบบอื่น ๆ ซึ่งบ่อยครั้งเราจะพบว่ามีตัวแบบหลายรูปแบบที่มีประสิทธิภาพในการพยากรณ์ใกล้เคียงกัน ดังนั้นเมื่อพิจารณาถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบแล้ววิธีการเฉลี่ยตัวแบบของเบส์น่าจะทำให้ได้ค่าพยากรณ์ที่มีประสิทธิภาพมากกว่าค่าพยากรณ์ที่ได้จากตัวแบบเดียว

การแจกแจงความน่าจะเป็นภายหลังของสิ่งที่สนใจ เมื่อมีข้อมูลจะเป็นดังนี้

(2.1)
$$p(\Delta \mid D) = \sum_{k=1}^{K} p(\Delta \mid M_k, D) \cdot p(M_k \mid D)$$

เมื่อ Δ เป็นปริมาณของสิ่งที่สนใจ เช่น ค่าพยากรณ์ที่สนใจ

D เป็นข้อมูลของตัวแบบที่สนใจ

และ $M_1, M_2, \ldots, M_k$ เป็นตัวแบบที่พิจารณา

สมการ (2.1) เป็นการเฉลี่ยการแจกแจงความน่าจะเป็นภายหลังซึ่งถ่วงน้ำหนักด้วยความน่าจะเป็นภายหลังของแต่ละตัวแบบ โดยเรียกแนวคิดนี้ว่า“การเฉลี่ยตัวแบบของเบส์ (BMA)”

---

[5] จิตติมา ผสมญาติ, “การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ,” (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546).

[6] Adrian E.Raftery, Jennifer A.Hoeting and David Madigan, “Bayesian Model Averaging for Linear Regression Models,” Journal of the American Statistical Association (April 1998).

ความน่าจะเป็นภายหลังของตัวแบบ M_k คือ

$$p(M_k \mid D) = \frac{p(D \mid M_k) \cdot p(M_k)}{\sum_{l=1}^{K} p(D \mid M_l) \cdot p(M_l)} \tag{2.2}$$

(2.3) เมื่อ $p(D \mid M_k) = \int p(D \mid \underset{\sim}{\theta}_k, M_k) \cdot p(\underset{\sim}{\theta}_k \mid M_k) d\underset{\sim}{\theta}_k$

ซึ่งเป็นความควรจะเป็นขอบ (marginal likelihood) ของตัวแบบ M_k

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

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

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

และ $p(M_k)$ เป็นความน่าจะเป็นก่อน (prior probability) สำหรับตัวแบบ M_k

ส่วนค่าเฉลี่ยและความแปรปรวนภายหลังของสิ่งที่สนใจ Δ เป็นดังนี้

$$E[\Delta \mid D] = \sum_{k=0}^{K} \hat{\Delta}_k p(M_k \mid D) \tag{2.4}$$

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

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

นั่นคือจากสมการ(2.4) เราสามารถสรุปได้ว่าถ้าสิ่งที่เราสนใจ Δ เป็นค่าพยากรณ์ $\hat{y}$ ดังนั้นค่าคาดหมายของค่าพยากรณ์ $\hat{y}$ เมื่อมีข้อมูลคือการเฉลี่ยค่าพยากรณ์ของแต่ละตัวแบบถ่วงน้ำหนักด้วยความน่าจะเป็นภายหลังของแต่ละตัวแบบ M_k จะทำให้เราได้ค่าพยากรณ์ $\hat{y}$ ตามที่ต้องการ

นอกจากนี้ ในงานวิจัยของ เมดิแกน(Madigan) และ ราฟเทอร์รี่(Raftery) ยังแสดงให้เห็นว่า การเฉลี่ยตัวแบบจะทำให้การพยากรณ์มีประสิทธิภาพสูงกว่าการใช้ตัวแบบเดียวในการพยากรณ์ โดยใช้กฎของคะแนนลอการิทึม และแนวคิดเกี่ยวกับรูปแบบข้อสนเทศของ คูลล์แบ็ค-ไลเบอร์ (Kullback-Leibler,1951) จะได้ว่า

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

ซึ่งเป็นการยืนยันว่าวิธีการเฉลี่ยตัวแบบจะมีประสิทธิภาพในการพยากรณ์สูงกว่าการใช้ตัวแบบเดียวในการพยากรณ์

ในการหาการแจกแจงความน่าจะเป็นภายหลัง (posterior probability distribution) ของแต่ละตัวแบบจำเป็นต้องทราบถึงความน่าจะเป็นก่อน (prior probability distribution) ของตัวแบบ ซึ่งมีรายละเอียดดังนี้

<u>การแจกแจงความน่าจะเป็นก่อน</u>(prior probability distribution)<u>ของตัวแบบ</u> จะอยู่ในรูปแบบดังนี้

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

เมื่อ p_j เป็นความน่าจะเป็นที่ตัวแปรอิสระที่ j จะรวมอยู่ในตัวแบบ M_i

และ δ_{ij} เป็นตัวบ่งชี้ที่มีค่า 0 หรือ 1

โดยที่ δ_{ij} เป็น 0 เมื่อ β_j มีอิทธิพลน้อย แสดงว่าตัวแปรอิสระที่ j ไม่อยู่ในตัวแบบ M_i

และ δ_{ij} เป็น 1 เมื่อ β_j มีอิทธิพลมาก แสดงว่าตัวแปรอิสระที่ j อยู่ในตัวแบบ M_i

จะได้ว่า $\delta_j \sim Ber\left(p_j\right)$ เมื่อ $j=1,2,...,k$

การวิจัยครั้งนี้จะกำหนด p_j สำหรับทุก ๆ ตัวแปรอิสระเป็น $\frac{1}{2}$ หรือกล่าวได้ว่าตัวแปรอิสระทุกตัวมีโอกาสที่จะอยู่ในตัวแบบหรือไม่อยู่ในตัวแบบเท่า ๆ กัน

ในตัวแบบการถดถอยเชิงเส้นพหุคูณ เราจะกำหนดการแจกแจงความน่าจะเป็นก่อนแบบผสม[7] *(mixture prior probability distribution) ของ β_j ได้ดังนี้*

(2.7) $$\beta_j\,|\,\delta_j \sim (1-\delta_j)N(0,\tau_j^2)+\delta_j N(0,(c_j\tau_j)^2)$$

จากสมการ(2.7) ถ้ากำหนด[8] $\tau_j>0$ *และ* $c_j>1$ *จะได้ว่า*

ถ้า $\delta_j=0$ แล้ว $\beta_j \sim N(0,\tau_j^2)$ กล่าวคือตัวแปรอิสระ X_j มีความเป็นไปได้น้อยที่จะอยู่ในตัวแบบ จึงกำหนดให้ β_j มีค่าประมาณเป็นศูนย์

ถ้า $\delta_j=1$ แล้ว $\beta_j \sim N(0,(c_j\tau_j)^2)$ กล่าวคือตัวแปรอิสระ X_j มีความเป็นไปได้สูงที่จะอยู่ในตัวแบบ ดังนั้นค่าประมาณของ β_j จึงไม่เท่ากับศูนย์ เมื่อ p_j เป็นความน่าจะเป็นก่อนของ β_j ที่มีค่าประมาณไม่เท่ากับศูนย์ ซึ่งจะสอดคล้องกับตัวแปรอิสระ X_j ที่ควรจะอยู่ในตัวแบบ

จากสมการ (2.7)จะได้หลักเกณฑ์เกี่ยวกับ $\beta_j\,|\,\delta_j$ ที่มีการแจกแจงพหุแบบปกติ (multivariate normal prior) ดังนี้

(2.8) $$\underset{\sim}{\beta}\,|\,\underset{\sim}{\delta} \sim N_k(0,D_\delta^2)$$

เมื่อ $\underset{\sim}{\delta}=(\delta_1,...,\delta_k)'$

และ $D_\delta \equiv diag\left[a_1\tau_1,...,a_k\tau_k\right]$

โดยที่

$$a_i=\begin{cases}1 & ,\delta_i=0\\ c_i & ,\delta_i=1\end{cases}$$

---

[7] นิทัสน์ สุขสุวรรณ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2545) หน้า 16-18 และหน้า 28.

[8] การเลือกที่เหมาะสมและนำไปสู่ประสิทธิภาพต่อการคำนวณของ c_j จะมีค่าอยู่ที่ระหว่าง 10 และ 100

ในการกำหนดค่าคงที่[9] $\tau_1,...,\tau_k$ *และ* $c_1,...,c_k$ *ของเมทริกซ์* D_δ *นั้นควรเลือกค่าคงที่ที่สามารถแบ่งแยกการแจกแจงของพารามิเตอร์* $\beta_j \sim N(0,\tau_j^2)$ *และ* $\beta_j \sim N(0,(c_j\tau_j^2))$ *ได้อย่างชัดเจน วิธีการหนึ่งที่ใช้ในการเลือกค่าคงที่* τ_j *และ* c_j *คือวิธีการลู่เข้ากึ่งอัตโนมัติ (semiautomatic approach) โดยพิจารณาจุดตัด (intersection point) และความสูงสัมพัทธ์ (relative height) ณ ตำแหน่งศูนย์ของความหนาแน่นขอบ (marginal densities) ของ*

$$(\hat{\beta}_j \mid \sigma_{\beta_j}, \delta_j = 0) \sim N(0, \sigma_{\beta_j}^2 + \tau_j^2)$$

และ

$$(\hat{\beta}_j \mid \sigma_{\beta_j}, \delta_j = 0) \sim N(0, \sigma_{\beta_j}^2 + c_j^2\tau_j^2)$$

กำหนดให้ $t_j\sigma_{\beta_j}$ *แทนจุดตัด*

X_j *แทนตัวแปรอิสระตัวที่* j , $j = 1,2,...,k$

และ $\sigma_{\beta_j}^2$ *แทนความแปรปรวนของตัวประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอย* $\hat{\beta}_j$ *โดยวิธีกำลังสองน้อยสุด*

(2.9) *เนื่องจาก* $P(\delta_j = 1 \mid \hat{\beta}_j, \sigma_{\beta_j}) > (p_j = P(\delta_j = 1))$ *ก็ต่อเมื่อ* $\dfrac{\hat{\beta}_j}{\sigma_{\beta_j}} > t_j$

ดังนั้นจุด t_j *จะลู่เข้าสู่การแจกแจงแบบที (t-distribution) ซึ่งจะสอดคล้องกับการเพิ่มขึ้นของความน่าจะเป็นขอบ กล่าวคือ* X_j *ควรจะอยู่ในตัวแบบ ซึ่ง* t_j *ที่มีค่ามากๆจะเป็นตัวแบบที่เหมาะสม*

ความสูงสัมพัทธ์ของความหนาแน่นส่วนริม $\hat{\beta}_j$ *ณ จุดศูนย์ คือ*

$$r_j \equiv \sqrt{\frac{\dfrac{\sigma_{\beta_j}^2}{\tau_j^2} + c_j^2}{\dfrac{\sigma_{\beta_j}^2}{\tau_j^2} + 1}} \qquad (2.10)$$

เมื่อ r_j *แทนความน่าจะเป็นภายหลังขอบ (marginal posterior probability) ของ* X_j *ที่อยู่ในตัวแบบทั้งๆที่ค่า* $\hat{\beta}_j = 0$

แม้ว่าการเลือกค่าคงที่ $\tau_1,...,\tau_k$ *และ* $c_1,...,c_k$ *ด้วยวิธีการลู่เข้ากึ่งอัตโนมัติจะทำให้ได้ค่า* t_j *และ* r_j *ที่เหมาะสม แต่ขั้นตอนต่างๆค่อนข้างยุ่งยากและเสียเวลาในการคำนวณมาก ดังนั้นในการวิจัยครั้งนี้จะใช้วิธีการกำหนด* $\dfrac{\sigma_{\beta_j}}{\tau_j}$ *และ* c_j *เป็นค่าคงที่ ซึ่งจากงานวิจัยของจอร์จ (George) และ*

---

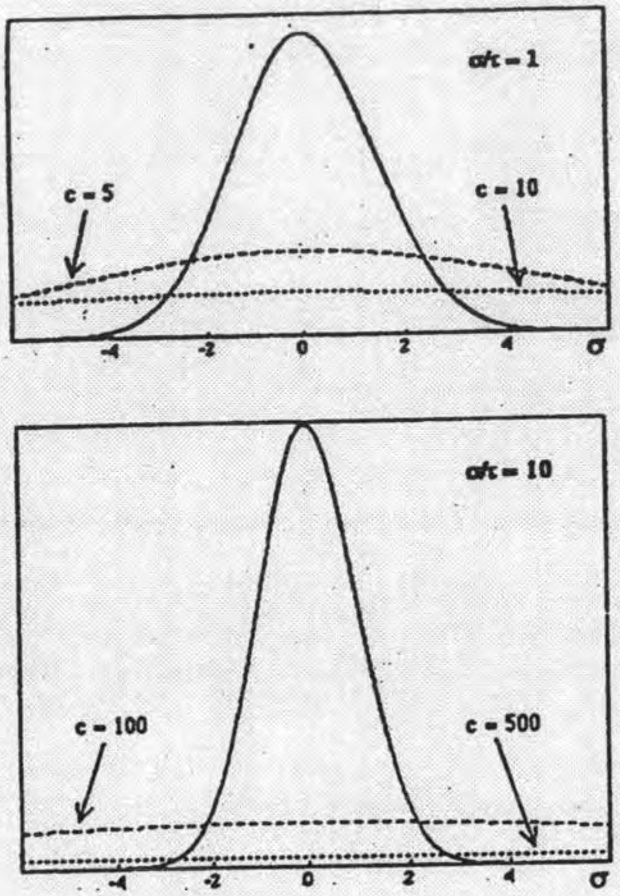
[9] นิทัสน์ สุขสุวรรณ, “การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ,” (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2545) หน้า19-21.

แมคคัลลอช (McColloch)[10] *ได้ชี้ให้เห็นว่า* การกำหนด $\frac{\sigma_{\beta_j}}{\tau_j}$ *และ* c_j *เป็นค่าคงที่เป็นวิธีการหนึ่งที่จะทำให้ได้ค่า* t_j *และ* r_j *ที่เหมาะสม นอกจากการกำหนดดังกล่าวจะทำให้หาผลลัพธ์ได้ง่ายและรวดเร็วแล้วยังจะมีผลทำให้การแปลงขนาด (rescaling) ของ* X_j *มีความคงที่มาก นอกจากนั้นยังเป็นผลลัพธ์ที่ดีเหมือนกับการกำหนดด้วยวิธีการลู่เข้ากึ่งอัตโนมัติเพื่อให้ได้ข้อสนเทศมากขึ้น พิจารณา*[11] *การกำหนดค่าคงที่* $\left(\frac{\sigma_{\beta_j}}{\tau_j}, c_j\right)$ = *(1,5) (1,10) (10,100) และ (10,500) ซึ่งการกำหนดค่าคงที่ดังกล่าวจะทำให้ได้ค่า* (t_j, r_j) *ที่เหมาะสมคือ (2.4 ,1.7) (2.7,2.3) (2.1,3.2) และ (2.8,6.8) ตามลำดับ ฟังก์ชันความหนาแน่นขอบ(marginal density function) ที่สอดคล้องกับการกำหนดค่าคงที่ดังกล่าวแสดงในรูปที่ 2.1ซึ่งจะเห็นได้ว่า เมื่อค่า* $\frac{\sigma_{\beta_j}}{\tau_j}$ *และ* c_j *มีค่าเพิ่มขึ้นจะทำให้การแจกแจงของพารามิเตอร์* β_j *กรณี* $\beta_j \sim N(0,\tau_j^2)$ *หรือ* $\beta_j \sim N(0,(c_j\tau_j^2))$ *แยกออกจากกันอย่างชัดเจน*

---

[10] George, E.I. and McCulloch, R.E., "Variable Selection via Gibbs Sampling ," Journal of American Statistical Association 88 (September 1993).

[11] การกำหนดให้ค่าคงที่ $\left(\frac{\sigma_{\beta_j}}{\tau_j}, c_j\right)$ เท่ากับ (1,5) (1,10) (10,100) และ(10,500) ตามลำดับ (ค่าคงที่(1,5) จะให้การแจกแจงแบบปกติที่มีการกระจายของพารามิเตอร์แคบเกินไปเมื่อเปรียบเทียบกับค่าคงที่ (1,10)และค่าคงที่ (10,500)จะให้การแจกแจงแบบปกติที่มีการกระจายของพารามิเตอร์กว้างเกินไปเมื่อเปรียบเทียบกับค่าคงที่ (10,100))



รูปที่ 2.1 แสดงความหนาแน่นขอบของการแจกแจง $N(0,\sigma_{\beta_j}^2+\tau_j^2)$ และ $N(0,\sigma_{\beta_j}^2+c_j^2\tau_j^2)$

เมื่อ $\left(\dfrac{\sigma_{\beta_j}}{\tau_j}, c_j\right)$ = (1,5) (1,10) (10,100) (10,500)

ความแปรปรวนของความคลาดเคลื่อน (residual variance): σ^2 มีสังยุคแบบแกมมาผกผัน (inverse gamma conjugate distribution) ดังนี้

(2.11)
$$\sigma^2 \mid \underset{\sim}{\delta} \sim IG\left(\frac{\upsilon_\delta}{2}, \frac{\upsilon_\delta \lambda_\delta}{2}\right)$$

ซึ่งจะสมมูลกับ $\dfrac{\upsilon_\delta \lambda_\delta}{2} \sim \chi^2_{\upsilon_\delta}$ โดยที่ υ_δ และ λ_δ จะขึ้นอยู่กับ $\underset{\sim}{\delta}$ ที่เป็นไปตามความสัมพันธ์ระหว่าง $\underset{\sim}{\beta}$ และ σ^2 กล่าวคือค่าของ σ^2 จะลดลงเมื่อมิติของ $\underset{\sim}{\beta}$ (คือ จำนวนเทอมที่สมาชิกใน $\underset{\sim}{\delta}$ ไม่เท่ากับศูนย์) เพิ่มขึ้น

<u>การแจกแจงความน่าจะเป็นภายหลังขอบ(marginal posterior probability)ของตัวแบบ</u>

เนื่องจากความน่าจะเป็นภายหลังขอบของตัวแบบ จะแปรผันตามฟังก์ชันความควรจะเป็นคูณกับการแจกแจงก่อนของพารามิเตอร์ กล่าวคือ

$$posterior \propto likelihood \times prior$$

<u>หมายเหตุ</u> การแจกแจงเบตา(Beta Distribution)[12]เป็นการแจกแจงที่สร้างมาจากการแจกแจงแกมมา และสามารถแสดงในทฤษฎีบทด้านล่างนี้

<u>ทฤษฎีบทที่ 4</u> ให้ X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน และ X~Gam(α_1,β) และ Y~Gam(α_2,β) จะได้ว่า U และV เป็นอิสระต่อกันด้วย

$$U=X+Y \sim Gam(\alpha_1+\alpha_2,\beta) \text{ ส่วน } V=\frac{X}{Y}$$

และจะมี $f(v)$ในรูปของ $f(v)=\dfrac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\dfrac{v^{\alpha_1-1}}{(1-v)^{\alpha_1+\alpha_2}}\quad ,v>0$

นอกจากนั้นW=$\dfrac{X}{X+Y}\sim Beta(\alpha_1,\alpha_2)$ซึ่งกรณีพิเศษของการแจกแจงเบตาที่ $\alpha=\beta=1$ คือการแจกแจงสม่ำเสมอ (uniform distribution)

<u>นิยามที่ 5</u> ตัวแปรสุ่ม X เรียกว่ามี การแจกแจงเบตาซึ่งมีพารามิเตอร์ $\alpha>0,\beta>0$ ถ้า X มีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\quad x^{\alpha-1}(1-x)^{\beta-1}\quad ,0<x<1$$

และเราเขียนได้เป็น $X\sim Beta(\alpha,\beta)$

<u>หมายเหตุ</u> 1. จากสมการข้างต้น จะได้ว่า Beta(1,1) = U(0,1)

2. เราสามารถแสดงได้ว่า สมการข้างต้น เป็นฟังก์ชันความหนาแน่นโดยใช้

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)},\quad \alpha,\beta>0$$

---

[12] ธีระพร วีระถาวร, <u>ความน่าจะเป็นกับการประยุกต์</u>, พิมพ์ครั้งที่2 (กรุงเทพมหานคร: วิทยพัฒน์, 2539), หน้า 199.

<u>นิยามที่ 6</u> ตัวแปรสุ่ม X เรียกว่า ตัวแปรสุ่มแบร์นูลลี ถ้า

$$X = \begin{cases} 0, & \text{ผลการทดลองคือล้มเหลว} \\ 1, & \text{ผลการทดลองคือสำเร็จ} \end{cases}$$

เมื่อ $P(X=1) = p, \quad 0 < p < 1$

และ $P(X=0) = 1-p$

อาจเขียนแทนด้วย $X \sim Ber(p)$ ซึ่งฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

$$P(X=x) = p^x(1-p)^{1-x}, \quad x = 0,1$$

<u>นิยามที่ 7</u> ตัวแปรสุ่ม X เรียกว่าเป็นตัวแปรสุ่มทวินาม ถ้า $X =$ จำนวนผลสำเร็จทั้งหมดในการทดลองแบบแบร์นูลลีที่เป็นอิสระซึ่งกันและกัน n ครั้ง และเขียนแทนด้วย $X \sim B(n,p)$ ซึ่งฟังก์ชันความน่าจะเป็นจะอยู่ในรูปของ

$$P(X=x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0,1,\ldots,n$$

เมื่อ $p \in (0,1)$

ในการหาฟังก์ชันความหนาแน่นที่มีเงื่อนไขในกรณีที่ตัวแปรสุ่มตัวหนึ่งเป็นตัวแปรสุ่มต่อเนื่อง และตัวแปรสุ่มอีกตัวหนึ่งเป็นตัวแปรสุ่มไม่ต่อเนื่อง หรือกล่าวได้ว่ามีการแจกแจงผสม (Mixed Distribution)[13]

<u>นิยามที่ 8</u> ให้ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่องและต่อเนื่องตามลำดับ และกำหนด $D = \{x \mid g(x) > 0\}, E = \{y \mid h(y) > 0\}$ เราเรียก $H(y \mid x)$ ว่าเป็นฟังก์ชันการแจกแจงที่มีเงื่อนไขของ Y เมื่อกำหนด X = x (the conditional distribution of Y given X = x)

$$H(y \mid x) = P(Y \le y \mid X = x) = \frac{P(Y \le y, X = x)}{P(X = x)}$$

เพราะฉะนั้น $h(y \mid x) = H'(y \mid x)$

---

[13] ธีระพร วีระถาวร, <u>ความน่าจะเป็นกับการประยุกต์</u>, พิมพ์ครั้งที่2 (กรุงเทพมหานคร: วิทยพัฒน์, 2539), หน้า 361-363.

<u>หมายเหตุ</u>

1. $g(x \mid y) = \dfrac{h(y \mid x) \cdot g(x)}{h(y)} \propto h(y \mid x) \cdot g(x)$

เมื่อ $h(y) = \sum_{x \in D} h(y \mid x) g(x), y \in E$

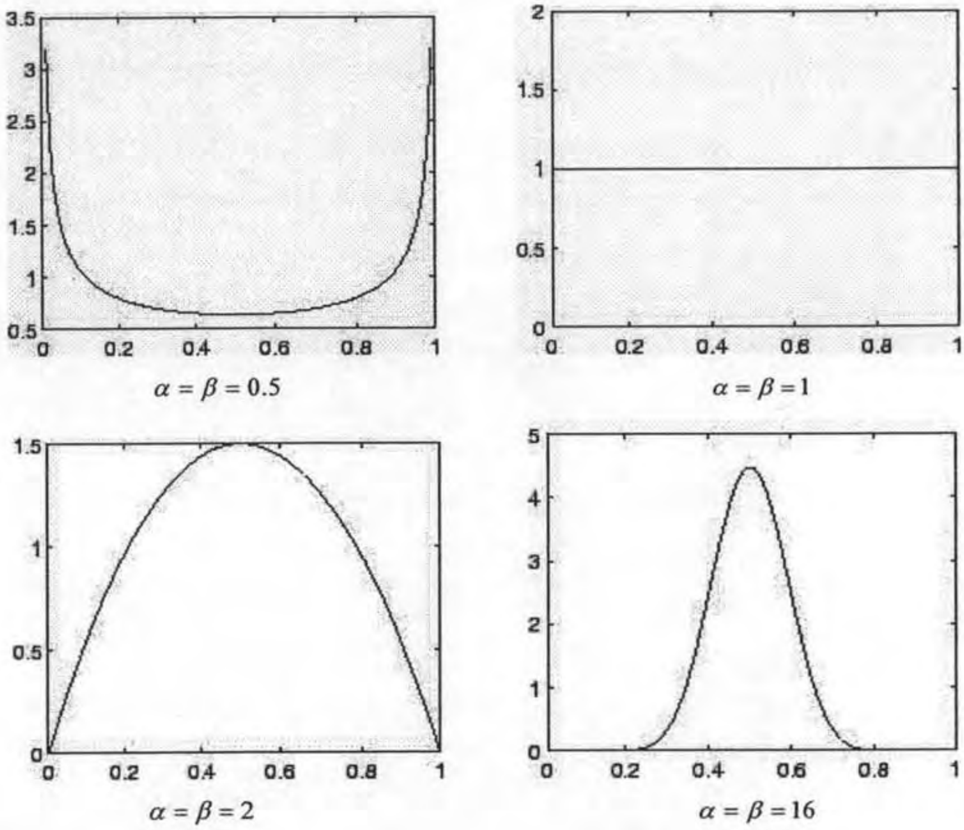
กล่าวคือ การแจกแจงของ X ในกรณีที่ทราบข้อมูลเกี่ยวกับ y แปรผันตามการแจกแจงของข้อมูลเกี่ยวกับ Y ภายใต้ข้อกำหนด X = x คูณด้วยการแจกแจงของ x

2. $h(y \mid x) = \dfrac{g(x \mid y) \cdot h(y)}{g(x)} \propto g(x \mid y) \cdot h(y)$

เมื่อ $g(x) = \int_E g(x \mid y) h(y) dy, y \in E$

จากนิยามที่ 5 สามารถแสดงได้ว่าการแจกแจงเบตามีคุณสมบัติเป็นการแจกแจงก่อนคู่สังยุคเมื่อฟังก์ชันความควรจะเป็นมีการแจกแจงเบอร์นูลลี เพราะการแจกแจงก่อนที่เป็นการแจกแจงเบตา (beta prior distribution) จะทำให้การแจกแจงภายหลัง (posterior distribution) มีการแจกแจงเบตาเหมือนกัน ซึ่งผู้วิจัยจะแสดงตัวอย่างการหาการแจกแจงภายหลังข้างต้นดังในภาคผนวก หน้า 250

ในงานวิจัยนี้ใช้ค่า $\alpha = \beta = 2$ เพราะการแจกแจงเบตาหลายตัวแปร(multivariate beta distribution) มีรูปแบบสมมาตรและมีรูปแบบเข้าใกล้การแจกแจงปกติที่สุด ซึ่งการวิเคราะห์การถดถอย(regression analysis)โดยทั่วไป จะวิเคราะห์ภายใต้ข้อสมมุติของการแจกแจงปกติ และวิธีการประมาณที่นิยมใช้กันคือวิธีกำลังสองน้อยสุด (ordinary least squares :OLS) แสดงดังรูปที่ 2.1



รูปที่ 2.1 แสดงการเปรียบเทียบแจกแจงเบตาหลายตัวแปรที่มีค่าอยู่ในช่วงระหว่าง 0 ถึง 1 หรือ [0,1]

2. ขั้นตอนการคัดเลือกตัวแปรของวิธี BMA_{MC3}[14]

การคัดเลือกตัวแปรของวิธี BMA_{MC3} มีขั้นตอนเหมือนวิธีการคัดเลือกตัวแปรของเบส์เซียนแต่วิธี BMA_{MC3}จะนำตัวแบบในทุกรูปแบบที่ปรากฏอยู่ในปริภูมิตัวแบบมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก โดยรายละเอียดของขั้นตอนการคัดเลือกตัวแปรของเบส์เซียน[15]มีดังนี้

ขั้นตอนแรกของวิธีการนี้เริ่มต้นจากสมการถดถอยเต็มรูปประกอบด้วยตัวแปรอิสระครบทุกตัว แล้วทำการสุ่มแบบกิบส์เพื่อสร้างลำดับ

(2.12) $$\delta^1, \delta^2, \ldots$$

ลำดับในสมการ (2.12) จะลู่เข้าสู่การแจกแจงภายหลัง ซึ่งเป็นลำดับที่มีความน่าจะเป็นค่อนข้างสูง เพราะบรรจุข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปร โดย $\underset{\sim}{\delta}$ ใดๆที่มีความน่าจะเป็นสูงสุดซึ่งทำให้ง่ายต่อการตัดสินใจในการเลือกตัวแบบที่ดีที่สุด

เราใช้วิธีการสุ่มแบบกิบส์เพื่อสร้างเวกเตอร์เป็นลำดับแบบกิบส์ ดังนี้

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

เมื่อ $\underset{\sim}{\beta}^0, \sigma^0$ เป็นค่าเริ่มต้นที่ได้จากการประมาณค่าด้วยวิธีกำลังสองน้อยสุดของสมการ (1.1) และ $\underset{\sim}{\delta}^0$ ถูกกำหนดค่าเริ่มต้นให้เป็น $\underset{\sim}{\delta}^0 \equiv (1,1,\ldots,1)'$

ในแต่ละรอบของการสุ่มค่าของ $\underset{\sim}{\beta}$, σ^2 และ $\underset{\sim}{\delta}$ ได้จากการสร้างค่าที่สอดคล้องกับขั้นตอนต่อไปนี้

ก) สุ่มเวกเตอร์ของสัมประสิทธิ์การถดถอย ($\underset{\sim}{\beta}$)จากการแจกแจง

(2.14) $$p\left(\underset{\sim}{\beta} \mid \sigma^2, \underset{\sim}{\delta}, \underset{\sim}{y}\right) = N_p\left(\left(X'X + \sigma^2 D_\delta^{-2}\right)^{-1} X'Y, \sigma^2\left(X'X + \sigma^2 D_\delta^{-2}\right)^{-1}\right)$$

ข) สุ่มค่าความแปรปรวน (σ^2) จากการแจกแจง

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

---

[14] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546).

[15] Wei Chen and Debashis Ghosh , A Bayesian method for finding interaction in genomic studies, (University of Michigan School of Public Health, 2004), p.1-30.

ค) ค่าของเวกเตอร์ $\underset{\sim}{\delta}$ ได้จากการแยกองค์ประกอบ (component wise) โดยการสุ่มอย่างต่อเนื่องกัน (sampling consecutively) จากการแจกแจงแบบมีเงื่อนไขดังนี้

$$\delta_i \sim p\left(\delta_i \mid \underset{\sim}{y}, \underset{\sim}{\beta}, \sigma^2, \delta_{-i}\right) = p\left(\delta_i \mid \underset{\sim}{\beta}, \sigma^2, \delta_{-i}\right) \qquad (2.16)$$

โดย $\delta_{-i} = \left(\delta_i, \ldots, \delta_{i-1}, \delta_{i+1}, \ldots, \delta_k\right)$

เมื่อ δ_{-i} แทนเวกเตอร์ย่อยทั้งหมดของ δ ยกเว้น γ_i ซึ่งเป็นค่าสุ่มปัจจุบัน

จะเห็นได้ว่าการแจกแจงในสมการ (2.16) จะไม่ขึ้นอยู่กับค่าของ $\underset{\sim}{y}$

ในสมการ (2.16) δ_i แต่ละตัวจะมีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution)ด้วยความน่าจะเป็น

$$\begin{aligned} P\left(\delta_i = 1 \mid \underset{\sim}{\beta}, \sigma^2, \delta_{-i}\right) &= P\left(\delta_i = 1 \mid \underset{\sim}{\beta}, \delta_{-i}\right) \\ &= \frac{P\left(\delta_i = 1, \underset{\sim}{\beta}, \delta_{-i}\right)}{P\left(\delta_i = 1, \underset{\sim}{\beta}, \delta_{-i}\right) + P\left(\delta_i = 0, \underset{\sim}{\beta}, \delta_{-i}\right)} \end{aligned} \qquad (2.17)$$

ความยาวของลำดับในสมการ (2.12)จะเพิ่มขึ้นเรื่อยๆทำให้การแจกแจงของค่าที่แท้จริงของ $\underset{\sim}{\delta}$ จะลู่เข้าสู่การแจกแจงภายหลังการลู่เข้านี้จะเป็นไปอย่างรวดเร็วถ้าการแจกแจงภายหลังเป็นค่าสูงสุด ตัวแบบที่มีน้ำหนักมาก ๆ มีจำนวนไม่มาก ตัวแบบเหล่านี้จะมีความแม่นยำสูง เมื่อการแจกแจงภายหลังบรรจุข้อมูลส่วนใหญ่ของการคัดเลือกตัวแปร

3. <u>ขั้นตอนของการคัดเลือกตัวแปรและการแปลงไปพร้อมๆ กันของวิธี BMA_{SVT}</u> [16]

วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟเมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ(BMA_{SVT}) เสนอโดย ราฟเทอร์รี(Raftery) เมดิแกน(Madigan) และโฮเอ็ททิง(Hoeting) ซึ่งสามารถเพิ่มประสิทธิภาพของวิธีการเฉลี่ยตัวแบบของเบส์ให้มากขึ้น ซึ่งมีขั้นตอนดังนี้

(1) นำตัวแปรอิสระเข้าสู่กระบวนการ Alternating Conditional Expectation Algorithm (ACE)

(2) พิจารณารูปกราฟแสดงความสัมพันธ์ระหว่างตัวแปรอิสระที่ทำการแปลงและตัวแปรอิสระไม่ได้ทำการแปลง

---

[16] Adrian E.Raftery, Jennifer A.Hoeting and David Madigan, <u>Bayesian Simultaneous Variable and Transformation Selection in Linear Regression</u> (June 1995).

(3) ถ้าตัวแปรตาม(Y) ในกรณีที่ทำการแปลงและตัวแปรตามดั้งเดิมมีความสัมพันธ์กันแบบไม่เป็นเชิงเส้น(non-linear of feature) แสดงว่าควรทำการแปลงตัวแปรตามด้วย หลังจากนั้นนำข้อมูลเข้าสู่กระบวนการการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟด้วยกำลังการแปลงที่เหมาะสม 4 กรณี คือ $p=-1,\ 0,\ 0.5,\ 1$ แต่ถ้าหากความสัมพันธ์ดังกล่าวเป็นเชิงเส้น ไม่จำเป็นต้องทำการแปลงค่าตัวแปรตาม

(4) ในกรณีตัวแปรอิสระ (X) ที่ทำการแปลงและตัวแปรอิสระเดิมมีความสัมพันธ์กันแบบไม่เป็นเชิงเส้น (non-linear of feature) แสดงว่าควรทำการแปลงตัวแปรอิสระนั้นๆโดยแต่ละตัวแปรอิสระ จะทำการพิจารณากราฟของการแปลงที่แนะนำโดยกระบวนการ ACE และพิจารณาการแปลงที่เหมาะสมด้วยปัจจัยเบส์ (Bayes factor) เพื่อตัดสินใจว่าตัวแปรอิสระนั้นๆสมควรทำการแปลงหรือไม่

(5) นำตัวแปรอิสระทุกตัวและตัวแปรอิสระที่ควรทำการแปลงเข้าสู่กระบวนการ MC^3 พร้อมกันเพื่อหาตัวแบบที่เหมาะสมที่สุดในการพยากรณ์

2.6 วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด(Optimal Predictive Model (OPM))[17]

วิธีการนี้เสนอโดยบาร์บีรี(Barbieri) และเบอเกอร์(Berger) ซึ่งมีแนวคิดว่าตัวแบบที่เหมาะสมที่สุดในการพยากรณ์อาจไม่ใช่ตัวแบบที่มีความน่าจะเป็นภายหลังสูงสุด (the model with highest posterior probability) แต่กลับเป็นตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐาน (median probability model) ซึ่งใกล้เคียงกับวิธีการเฉลี่ยตัวแบบของเบส์เมื่อพิจารณาความไม่แน่นอนของตัวแบบ และในการเลือกตัวแบบที่เหมาะสมที่สุดจากกลุ่มของตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐานจะใช้ค่าความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสอง (square error loss)โดยพิจารณาเลือกตัวแบบที่มีค่าความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสองต่ำสุด

1) <u>การค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐาน</u> (median probability model)

การค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐาน จะทำโดยใช้วิธีการค้นหาตัวแบบด้วยเทคนิคลูกโซ่มาร์คอฟเชนมอนติคาร์โล (ordinary MCMC model search schemes) วิธีการนี้ได้พัฒนาลูกโซ่มาร์คอฟให้เคลื่อนไประหว่างตัวแบบด้วยค่าความน่าจะเป็นภายหลังของตัวแบบนั้นๆซึ่งคำนวณค่าความน่าจะเป็นภายหลังได้จากสัดส่วนของจำนวนครั้งที่ตัวแบบอยู่ในลูกโซ่ การค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐานจะพิจารณาค่าความน่าจะเป็นภายหลังสำหรับ

---

[17] Maria M.Barbieri and James O.Berger "Optimal predictive model selection," <u>Technical Report 02-02</u> (April 2002).

ตัวแปรอิสระใด ๆ เป็นเกณฑ์ในการคัดเลือก กล่าวคือเมื่อจบสิ้นกระบวนการ MCMC เราจะทำการเลือกตัวแปรอิสระใดๆที่มีค่าสัดส่วนของจำนวนครั้งที่ตัวแปรอิสระดังกล่าวอยู่ในตัวแบบมากกว่า $\frac{1}{2}$

กำหนดสมการ (1.1) เป็นสมการของตัวแบบเต็ม (full models) เราจะได้สมการของตัวแบบย่อย (submodels) ที่นำมาพิจารณาเป็นดังสมการ (2.18)

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

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

หรือ $l_j = \begin{cases} 1 & , X_j \text{ อยู่ในตัวแบบ} \\ 0 & , \text{อื่นๆ} \end{cases}$

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

และค่าพยากรณ์ที่ได้จะอยู่ในรูป

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

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

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

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

<u>นิยามที่ 9</u> ผลรวมของความน่าจะเป็นภายหลัง[18]สำหรับตัวแปรอิสระที่ i (posterior inclusion probability for variable i)

(2.21) $$p_i \equiv \sum_{l: l_i = 1} P(M_l \mid \underset{\sim}{y})$$

ดังนั้น ตัวแปรอิสระที่ i จะอยู่ในตัวแบบก็ต่อเมื่อ $p_i \geq \frac{1}{2}$ และจะได้ตัวแบบที่มีความน่าจะเป็นมัธยฐาน(median probability model : M_{l^*}) เป็นตัวแบบที่ประกอบด้วยตัวแปรอิสระที่มี p_i อย่างน้อยเท่ากับ $\frac{1}{2}$ หรือ $l_i^* = \begin{cases} 1, & \text{ถ้า } p_i \geq \frac{1}{2} \\ 0, & \text{อื่นๆ} \end{cases}$

---

[18] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546).

2) การคัดเลือกตัวแบบที่เหมาะสมที่สุด

จากการค้นหาตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐาน เราจะทำการคัดเลือกตัวแบบที่เหมาะสมที่สุดและและใช้วิธีการเลือกตัวแบบที่เหมาะสมที่สุดจากกลุ่มของตัวแบบที่มีความน่าจะเป็นภายหลังมัธยฐาน เป็นความสูญเสียอันเกิดจากความผิดพลาดยกกำลังสอง(square error loss) โดยจะเลือกตัวแบบที่มีค่าดังกล่าวต่ำสุด ซึ่งรายละเอียดของการคัดเลือกตัวแบบที่เหมาะสมที่สุดมีดังนี้

สมมติว่า $Q = X'X$ เป็นเมทริกซ์ทแยงมุม (diagonal matrix) เราจะได้ค่าเฉลี่ยภายหลัง $\tilde{\beta}_l$ (the posterior means) เป็นดังสมการ (2.22)

(2.22) $$\tilde{\beta}_l = H_l \tilde{\beta}$$

เมื่อ $\tilde{\beta}$ เป็นค่าเฉลี่ยภายหลังของตัวแบบเต็ม(posterior mean in the full model)

$\tilde{\beta}_l$ เป็นค่าประมาณกำลังสองน้อยสุด (the least squares estimates)

และ H_l เป็นเมทริกซ์ขนาด $p \times p_l$ ของคู่ลำดับ (j,k) ที่มีค่าเป็น 1

2.7 วิธีการถดถอยแบบขั้นบันได[19] (stepwise regression method (SR))

วิธีการถดถอยแบบขั้นบันไดเป็นวิธีการคัดเลือกตัวแปรอิสระแบบเลือกเข้า โดยในแต่ละขั้นตอนของการเลือกตัวแปรอิสระเข้าสู่สมการถดถอยจะตรวจสอบตัวแปรอิสระที่มีในสมการถดถอยก่อนที่ตัวแปรอิสระตัวล่าสุดจะเพิ่มเข้าในสมการ โดยวิธีการกำจัดตัวแปรอิสระแบบถอยหลัง ซึ่งมีขั้นตอนดังนี้

1) สร้างสมการถดถอยที่ไม่มีตัวแปรอิสระอยู่ในสมการ

2) เลือกตัวแปรอิสระที่มีสัมประสิทธิ์สหสมพันธ์(correlation coefficient)กับตัวแปรตาม (y) สูงสุดเข้าสู่สมการเป็นตัวแปรแรก และทดสอบว่าตัวแปรอิสระดังกล่าวมีนัยสำคัญหรือไม่ หากตัวแปรอิสระดังกล่าวไม่มีนัยสำคัญ จะได้สมการถดถอยที่เหมาะสมคือ $y = \bar{y}$ แต่ถ้าตัวแปรอิสระดังกล่าวมีนัยสำคัญให้ทำขั้นตอนต่อไป

3) คำนวณค่าสถิติเอฟบางส่วนสำหรับตัวแปรอิสระที่ไม่อยู่ในสมการ

4) ค่าสถิติเอฟบางส่วนที่มากที่สุด (F_U) จะถูกเปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_0)

---

[19] จิตติมา ผสมญาติ, "การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงก่อนแบบคู่สังยุคปกติ," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย,2546).

- ถ้า $F_U < F_0$ จะไม่นำตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนที่มากที่สุดเข้าสู่สมการถดถอย

- ถ้า $F_U > F_0$ จะนำตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนที่มากที่สุดเข้าสู่สมการถดถอย และคำนวณค่าสัมประสิทธิ์ของสมการถดถอย เมื่อนำตัวแปรอิสระนั้นเข้าสู่สมการด้วยวิธีกำลังสองน้อยสุด

5) คำนวณค่าสถิติเอฟบางส่วนของทุกตัวแปรอิสระที่อยู่ในสมการถดถอย

6) ค่าสถิติเอฟบางส่วนที่น้อยที่สุด (F_L) จะถูกเปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_0)

- ถ้า $F_L < F_0$ จะตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนที่น้อยที่สุดนั้นออกจากสมการถดถอย และคำนวณค่าสัมประสิทธิ์ของสมการถดถอย เมื่อตัดตัวแปรอิสระนั้นแล้ว

- ถ้า $F_L > F_0$ จะไม่ตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนที่น้อยที่สุดนั้นออกจากสมการถดถอย

7) ถ้าไม่มีตัวแปรใดเข้าและออกจากสมการถดถอยแล้ว จะได้สมการถดถอยที่เหมาะสมแต่ถ้ายังมีตัวแปรอิสระใดที่เป็นไปตามเงื่อนไขของการเข้าหรือออกจากสมการ ให้กลับไปทำในขั้นตอนที่ 3

จากขั้นตอนของวิธีการถดถอยขั้นบันไดจะเห็นได้ว่าต้องอาศัยวิธีการทดสอบเอฟบางส่วนและการหาสัมประสิทธิ์สหสัมพันธ์ ซึ่งได้กล่าวถึงรายละเอียดไว้ในภาคผนวก