บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

การศึกษาเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ในการวิจัยครั้งนี้เป็นการพิจารณาหาวิธีการคัดเลือกสมการถดถอยที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด โดยวิธีการคัดเลือกสมการถดถอยที่นำมาศึกษามีดังนี้

1) วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC)) ซึ่งวิธีการนี้จะคำนวณค่า BIC สำหรับทุกตัวแบบที่เป็นไปได้ แล้วนำมาเปรียบเทียบกัน ตัวแบบใดที่มีค่า BIC ต่ำสุดจะเป็นตัวแบบที่ดีที่สุด
2) วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS)) เป็นวิธีการคัดเลือกตัวแปรโดยใช้หลักการค้นหาด้วยความน่าจะเป็น (Stochastic Search Variable Selection (SSVS)) ซึ่งใช้วิธีการสุ่มแบบกิบส์ (Gibbs Sampling)
3) วิธีการเฉลี่ยตัวแบบของเบส์ (Baysian Model Averaging Method (BMA)) เป็นการนำตัวแบบต่าง ๆ ที่เป็นไปได้มาเฉลี่ยกันโดยใช้หลักการของเบส์ ซึ่งในการวิจัยครั้งนี้จะทำการศึกษาวิธีการเฉลี่ยตัวแบบของเบส์ภายใต้ 2 แนวทาง คือ
   3.1) วิธีการเฉลี่ยตัวแบบของเบส์โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคมวินโดว์ (Occam 's Windows) (BMA_{OCC})
   3.2) วิธีการเฉลี่ยตัวแบบของเบส์ โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC3})
4) วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

เนื่องจากวิธีการเฉลี่ยตัวแบบของเบส์นั้นผู้วิจัยได้ทำการศึกษาภายใต้ 2 แนวทางดังกล่าวข้างต้น ดังนั้นวิธีการคัดเลือกสมการถดถอยที่นำมาศึกษาในการวิจัยครั้งนี้จะประกอบด้วย 5 วิธี คือ วิธี BIC BVS BMA_{OCC} BMA_{MC3} และ SR

การนำเสนอแนวคิดและทฤษฎีในบทนี้จะประกอบด้วยแนวคิดและทฤษฎีเกี่ยวกับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ ความน่าจะเป็นที่มีเงื่อนไข ความเป็นอิสระ ทฤษฎีของเบส์ และทฤษฎีของวิธีการคัดเลือกสมการถดถอยทั้ง 5 วิธีข้างต้น โดยมีรายละเอียดดังนี้

2.1 การวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Analysis)

ตัวแบบและข้อตกลงเบื้องต้นที่ใช้ในการวิจัยครั้งนี้ คือ ตัวแบบความถดถอยเชิงเส้นพหุคูณซึ่งมีรูปแบบทั่วไปดังนี้

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

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

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 สำหรับ $i \neq j$ มีการแจกแจงที่เหมือนกันและเป็นอิสระกัน ซึ่งจะทำให้ $Cov(\varepsilon_i, \varepsilon_j) = 0$ สำหรับ $i \neq j$

2) รูปแบบการถดถอยเป็นแบบเชิงเส้นตรงของพารามิเตอร์ (β_j)

3) ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่ และไม่มีพหุสัมพันธ์กัน

เนื่องจากตัวแบบการถดถอยที่ใช้ในการพิจารณาเป็นตัวแบบเชิงเส้น ดังนั้นในการประมาณค่าพารามิเตอร์เราสามารถใช้วิธีกำลังสองน้อยสุด และวิธีความควรจะเป็นสูงสุด เพื่อประมาณค่าเวกเตอร์พารามิเตอร์สัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}$ ของตัวแบบความถดถอยเชิงเส้นพหุคูณซึ่งมีรูปแบบตัวประมาณสำหรับแต่ละวิธีดังนี้

ตัวประมาณค่ากำลังสองน้อยสุด (Least Square Estimator : $\hat{\underset{\sim}{\beta}}$) คือ

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

เป็นตัวประมาณที่ใช้สำหรับวิธี BVS วิธี BMA_{MC3} และวิธี SR

ตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimator : $\hat{\underset{\sim}{\beta}}_{ML}$) คือ

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

เป็นตัวประมาณที่ใช้สำหรับวิธี BIC และวิธี BMA_{OCC}

2.2 **ความน่าจะเป็นที่มีเงื่อนไข** (Conditionnal Probability)[1]

ความน่าจะเป็นที่มีเงื่อนไขเป็นการศึกษาความน่าจะเป็นเมื่อทราบข้อมูลเพิ่มเติมเกี่ยวกับเหตุการณ์ที่สนใจศึกษา

<u>นิยามที่ 1</u>

ให้ A และ B เป็นเหตุการณ์ซึ่ง $P(A) > 0$ เราเรียก $P(B \mid A)$ ว่าเป็น "ความน่าจะเป็น**ที่มีเงื่อนไขของ** B **เมื่อกำหนด** A (the conditional probability of B given A)" ถ้า

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

<u>ทฤษฎีบทที่ 1</u> ทฤษฎีการคูณ (Multiplication Theorem)

ถ้าในการทดลองชนิดหนึ่งเหตุการณ์ A และ B สามารถเกิดขึ้นได้พร้อมกัน จะได้ว่า

$$P(B \cap A) = P(B \mid A) \,.\, P(A)$$

<u>ทฤษฎีบทที่ 2</u>

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

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

---

[1] ธีระพร วีระถาวร, **ความน่าจะเป็นกับการประยุกต์**, พิมพ์ครั้งที่ 2 (กรุงเทพมหานคร : วิทยพัฒน์ ,2539), หน้า 100-106.

2.3 **ทฤษฎีบทของเบส์** (Bayes' Theorem) [1]

<u>นิยามที่ 2</u>

เหตุการณ์ B_1, B_2 , ... จะแทนผลแบ่งกั้นของปริภูมิตัวอย่าง S ถ้า

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

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

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

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

<u>ทฤษฎีบทที่ 3</u>

ให้ B_1, B_2, ... แทนผลแบ่งกั้นของปริภูมิตัวอย่าง S และถ้า A เป็นเหตุการณ์ซึ่ง $P(A) > 0$ จะได้ว่า

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{\infty} P(B_i).P(A \mid B_i)}, \forall k = 1,2,3,\ldots$$

2.4 **ความเป็นอิสระ** (Independence) [2]

<u>นิยามที่ 3</u>

เหตุการณ์ 2 เหตุการณ์ A และ B เรียกว่า "เป็นอิสระต่อกัน (independence)" ก็ต่อเมื่อ

$P(B \mid A) = P(B)$ และ $P(A \mid B) = P(A)$ กล่าวคือ

$$P(A \cap B) = P(A) \,.\, P(B)$$

---

[1] ธีระพร วีระถาวร, **ความน่าจะเป็นกับการประยุกต์**, พิมพ์ครั้งที่ 2 (กรุงเทพมหานคร : วิทยพัฒน์ ,2539), หน้า 107-108.

[2] ธีระพร วีระถาวร, **ความน่าจะเป็นกับการประยุกต์**, พิมพ์ครั้งที่ 2 (กรุงเทพมหานคร : วิทยพัฒน์ ,2539), หน้า 114-116.

<u>นิยามที่ 4</u>

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

1) แต่ละคู่ของเหตุการณ์เป็นอิสระต่อกัน (pairwise independent) กล่าวคือ

$$P(A \cap B) = P(A) \,.\, P(B)$$

$$P(A \cap C) = P(A) \,.\, P(C)$$

$$P(B \cap C) = P(B) \,.\, P(C)$$

2) เหตุการณ์หนึ่งเหตุการณ์ใดเป็นอิสระจากสองเหตุการณ์ใด ๆ กล่าวคือ

$$P(A \cap B \cap C) = P(A) \,.\, P(B) \,.\, P(C)$$

2.5 วิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))

ในปี ค.ศ.1978 ชวาร์ซ (Schwarz) ได้เสนอเกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ เพื่อใช้เป็นเครื่องมือในการหาตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด นั่นคือ วิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC)) ซึ่งเป็นวิธีพื้นฐานในการหาเกณฑ์การคัดเลือกตัวแบบภายใต้แนวทางของเบส์ โดยเป็นการอนุมานด้วยความน่าจะเป็นภายหลัง (posterior probability) ซึ่งเกิดจากความน่าจะเป็นก่อน (prior probability) รวมกับฟังก์ชันความควรจะเป็น (likelihood function) สำหรับข้อกำหนดเบื้องต้นคือตัวประมาณได้มาจากการประมาณค่าด้วยวิธีการความควรจะเป็นสูงสุด และการแจกแจงก่อน (prior distribution) เป็นการแจกแจงแบบสม่ำเสมอ

เกณฑ์การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC)) คือ

$$BIC = -2\sum_{i=1}^{n} \ln f(x_i \mid \hat{\underset{\sim}{\beta}}_{ML}) + p\ln(n) \quad \text{หรือ}$$

$$BIC_k = -2\ln\left(ML_k\right) + p_k \ln(n) \qquad (2.1)$$

เมื่อ ML_k เป็นฟังก์ชันความควรจะเป็นสูงสุด (maximum likelihood function) ของตัวแบบที่ k

p_k เป็นจำนวนพารามิเตอร์ของตัวแบบที่ k

และ n เป็นขนาดตัวอย่าง

ซึ่งตัวแบบที่ให้ค่า BIC ต่ำสุด จะเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด

ในการวิจัยครั้งนี้พิจารณาเฉพาะกรณีที่ความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติ ดังนั้นฟังก์ชันความควรจะเป็นสำหรับตัวแบบที่มีรูปแบบดังสมการ (1.1) จะเป็นดังนี้

$$
\begin{aligned}
ML_k &= L(\underset{\sim}{\beta}_{ML},\sigma^2;y) \\
&= f(\underset{\sim}{y}\mid\underset{\sim}{\beta}_{ML},\sigma^2\mathrm{I}_n) \\
&= (2\pi\sigma^2)^{\frac{-n}{2}}\exp\left\{\frac{-1}{2\sigma^2}(\underset{\sim}{y}-X\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-X\underset{\sim}{\beta}_{ML})\right\}
\end{aligned}
$$

$$
\begin{aligned}
\therefore\ \ln ML_k &= \frac{-n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}(\underset{\sim}{y}-X\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-X\underset{\sim}{\beta}_{ML}) \\
&= \frac{-n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}(\underset{\sim}{y}-X\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-X\underset{\sim}{\beta}_{ML}) \\
&= \frac{-n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{n}{2} \\
&= \frac{-n}{2}\left(\ln(2\pi)+1\right)-\frac{n}{2}\ln(\hat{\sigma}^2)
\end{aligned}
$$

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

และ $\hat{\sigma}^2=\frac{1}{n}(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}_{ML})^T(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}_{ML})$

ดังนั้น BIC สำหรับการแจกแจงแบบปกติ คือ

$$
\begin{aligned}
BIC &= -2\ln L(\underset{\sim}{\beta}_{ML},\sigma^2;\underset{\sim}{y})+p\ln(n) \\
&= -2\left(-\frac{n}{2}\left(\ln(2\pi)+1\right)-\frac{n}{2}\ln(\hat{\sigma}^2)\right)+p\ln(n) \\
&= n(\ln(2\pi)+1)+n\ln(\hat{\sigma}^2)+p\ln(n)
\end{aligned}
$$

2.6 วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS))

ในปี ค.ศ.1993 จอร์จ (George) และ แมคคัลลอช (McCulloch) ได้เสนอวิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS)) สำหรับตัวแบบการถดถอยเชิงเส้นพหุคูณ โดยใช้หลักการคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น (Stochastic

Search Variable Selection (SSVS)) ซึ่งใช้วิธีการสุ่มแบบกิบส์ (Gibbs Sampling) วิธีการนี้จะมีการกำหนดความน่าจะเป็นให้กับตัวแบบโดยใช้หลักการคัดเลือกตัวแปรเพื่อกำหนดตัวแบบที่เหมาะสมกับข้อมูล เมื่อทราบการแจกแจงความน่าจะเป็นของพารามิเตอร์ในตัวแบบ และนอกจากนั้นในปี ค.ศ.1996 จอร์จ (George) และ แมคคัลลอช (McCulloch) ยังได้มีการนำเสนอแนวคิดวิธีการคัดเลือกตัวแปรของเบส์สำหรับตัวแบบเชิงเส้นนัยทั่วไป (generalized linear models) ซึ่งยังคงใช้หลักการเดิมอีกด้วย

การนำเสนอแนวคิดและทฤษฎีของวิธีการคัดเลือกตัวแปรของเบส์จำเป็นต้องกล่าวถึงการแจกแจงความน่าจะเป็นก่อน (prior probability) การแจกแจงความน่าจะเป็นภายหลัง (posterior probability) และการคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น ก่อนที่จะกล่าวถึงขั้นตอนของวิธีการคัดเลือกตัวแปรของเบส์ โดยมีรายละเอียดดังนี้

2.6.1 การแจกแจงความน่าจะเป็นก่อน (Prior Probability Distribution)

จากตัวแบบการถดถอยในสมการ (1.1) กำหนดให้ $\underset{\sim}{\delta}$ เป็นเวกเตอร์ที่ไม่ทราบค่าขนาด $k \times 1$ มีสมาชิกเป็น 0 และ 1 โดยที่สมาชิกใน $\underset{\sim}{\delta}$ จะสอดคล้องกับตัวแปรอิสระในสมการถดถอยในตำแหน่งที่ตรงกันดังนี้

(2.2) $$\delta_i = \begin{cases} 0 & , \beta_i \text{ มีอิทธิพลน้อย แสดงว่าตัวแปรอิสระนี้ไม่อยู่ในตัวแบบ} \\ 1 & , \beta_i \text{ มีอิทธิพลมาก แสดงว่าตัวแปรอิสระนี้อยู่ในตัวแบบ} \end{cases}$$

จะได้ว่า $\delta_i \sim Ber(p_i) \qquad ; i = 1, 2, \ldots, k$

ถ้าการแจกแจงความน่าจะเป็นก่อนเป็นอิสระกัน จะได้ว่าฟังก์ชันความหนาแน่นร่วมของ $\underset{\sim}{\delta}$ คือ

(2.3) $$f(\underset{\sim}{\delta}) = \prod_{i=1}^{k} p_i^{\delta_i} (1 - p_i)^{1-\delta_i}$$

เมื่อ $p_i = P(\delta_i = 1) = 1 - P(\delta_i = 0)$

ในตัวแบบการถดถอยเชิงเส้นพหุคูณ เราจะกำหนดการแจกแจงความน่าจะเป็นก่อนแบบผสม (mixture prior probability distribution) ของ β_i ได้ดังนี้

$$f(\beta_i \mid \delta_i) \sim \begin{cases} N(0,\tau_i^2) & \text{ถ้า } \delta_i = 0 \\ N(0,(c_i\tau_i)^2) & \text{ถ้า } \delta_i = 1 \end{cases}$$

นั่นคือ

(2.4) $$\beta_i \mid \delta_i \sim (1-\delta_i)\, N(0,\tau_i^2) + \delta_i\, N\left(0,(c_i\tau_i)^2\right)$$

จากสมการ (2.4) ถ้ากำหนด $\tau_i > 0$ และ $c_i > 1$ จะได้ว่า

ถ้า $\delta_i = 0$ แล้ว $\beta_i \sim N(0,\tau_i^2)$ กล่าวคือตัวแปรอิสระ X_i มีความเป็นไปได้น้อยที่จะอยู่ในตัวแบบ จึงกำหนดให้ β_i มีค่าประมาณเป็นศูนย์

ถ้า $\delta_i = 1$ แล้ว $\beta_i \sim N(0,(c_i\tau_i)^2)$ กล่าวคือตัวแปรอิสระ X_i มีความเป็นไปได้สูงที่จะอยู่ในตัวแบบ ดังนั้นค่าประมาณของ β_i จึงไม่เท่ากับศูนย์ เมื่อ p_i เป็นความน่าจะเป็นก่อนของ β_i ที่มีค่าประมาณไม่เท่ากับศูนย์ ซึ่งจะสอดคล้องกับตัวแปรอิสระ X_i ที่ควรจะอยู่ในตัวแบบ

จากสมการ (2.4) จะได้หลักเกณฑ์เกี่ยวกับ $\beta_i \mid \delta_i$ ที่มีการแจกแจงพหุแบบปกติ (multivariate normal prior) ดังนี้

(2.5) $$\underset{\sim}{\beta} \mid \underset{\sim}{\delta} \sim N_k(\underset{\sim}{0}, D_\delta R D_\delta)$$

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

R **เป็นเมทริกซ์สหสัมพันธ์ก่อน** (prior correlation matrix)

และ $D_\delta \equiv diag[a_1\tau_1,\ldots,a_k\tau_k]$

โดยที่

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

ความแปรปรวนของความคลาดเคลื่อน (residual variance) : σ^2 มีหลักเกณฑ์สังยุคแบบแกมมาผกผัน (inverse gamma conjugate prior) ดังนี้

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

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

2.6.2 การแจกแจงความน่าจะเป็นภายหลัง (Posterior Probability Distribution)

การแจกแจงความน่าจะเป็นภายหลังขอบ (marginal posterior distribution) แปรผันตามผลคูณระหว่างความควรจะเป็นของข้อมูล(likelihood) และการแจกแจงความน่าจะเป็นก่อน (prior distribution) ดังนี้

(2.7) $$f(\underset{\sim}{\delta} \mid \underset{\sim}{y}) \propto f(\underset{\sim}{y} \mid \underset{\sim}{\delta}) \,.\, f(\underset{\sim}{\delta})$$

เมื่อกำหนดข้อมูล $\underset{\sim}{y}$ ความน่าจะเป็นภายหลังจะปรับปรุงความน่าจะเป็นก่อนเมื่อกำหนดแต่ละค่าของ $\underset{\sim}{\delta}$ จากค่าที่เป็นไปได้ 2^k ค่า $\underset{\sim}{\delta}$ ที่มีความน่าจะเป็นภายหลัง $f(\underset{\sim}{\delta} \mid \underset{\sim}{y})$ สูง ๆ จะสอดคล้องกับตัวแบบที่เหมาะสมในการพยากรณ์

ปัจจัยต่าง ๆ ที่มีผลต่อความน่าจะเป็นภายหลัง ได้แก่

1) หลักเกณฑ์เกี่ยวกับ $f(\underset{\sim}{\delta})$

2) ค่าคงที่ $\tau_1, \ldots, \tau_k$ และ $c_1, \ldots, c_k$ ของเมทริกซ์ D_δ

3) เมทริกซ์สหสัมพันธ์ก่อน (prior correlation matrix) : R จากสมการ (2.5)

4) ค่าระดับขั้นความเสรี (degree of freedom) υ_δ และ λ_δ จากสมการ (2.6) ดังรายละเอียดต่อไปนี้

2.6.2.1 การกำหนดหลักเกณฑ์เกี่ยวกับ $f(\underset{\sim}{\delta})$

การแจกแจงขอบของ $\underset{\sim}{\delta}$ เมื่อตัวแปรอิสระเป็นอิสระซึ่งกันและกัน จะมีรูปแบบดังสมการ (2.3) ถ้ากำหนดให้มีหลักเกณฑ์แบบสม่ำเสมอ (uniform prior) แล้วตัวแปรอิสระแต่ละตัวแปรจะมีโอกาสเข้าสู่ตัวแบบเท่า ๆ กัน ดังนั้น $f(\underset{\sim}{\delta}) = 2^{-k}$ ซึ่งเกิดจากการกำหนด p_i ให้เป็นหลักเกณฑ์ที่ไม่ทราบข้อมูลมาก่อน (non informative prior) กล่าวคือ

$$p_i = P(\delta_i = 1) = \frac{1}{2} \qquad , \forall i = 1,2,\ldots,k$$

จากทฤษฎีของเบส์ เมื่อมีข้อมูล (information) เกี่ยวกับพารามิเตอร์เพิ่มขึ้น จะสามารถปรับปรุงหลักเกณฑ์ใหม่ด้วยความน่าจะเป็นภายหลังที่มีอยู่ก่อน เช่น เมื่อทราบว่า $\underset{\sim}{\delta}$ มีการแจกแจงอื่นนอกจากการแจกแจงแบบสม่ำเสมอ เราสามารถสร้างความน่าจะเป็นภายหลังใหม่ เมื่อกำหนดความน่าจะเป็นภายหลังเดิมทำหน้าที่เป็นหลักเกณฑ์ใหม่ ดังนี้

ความน่าจะเป็นภายหลัง$_1$ α ความน่าจะเป็นก่อน$_1$ x ความควรจะเป็น$_1$

ความน่าจะเป็นภายหลัง$_2$ α ความน่าจะเป็นภายหลัง$_1$ x ความควรจะเป็น$_2$

2.6.2.2 การกำหนดค่าคงที่ $\tau_1, \ldots, \tau_k$ และ $c_1, \ldots, c_k$ ของเมทริกซ์ D_δ

การเลือกค่าคงที่ $\tau_1, \ldots, \tau_k$ และ $c_1, \ldots, c_k$ นั้นควรเลือกให้เป็นค่าคงที่ที่สามารถแบ่งแยกการแจกแจงของพารามิเตอร์ $\beta_i \sim N(0,\tau_i^2)$ และ $\beta_i \sim N(0,(c_i\tau_i)^2)$ ได้อย่างชัดเจน วิธีการหนึ่งที่ใช้ในการเลือกค่าคงที่ τ_i และ c_i คือวิธีการสู่เข้ากึ่งอัตโนมัติ (Semiautomatic Approach) โดยพิจารณาจุดตัด (intersection point) และความสูงสัมพัทธ์ (relative height) ณ ตำแหน่งศูนย์ของความหนาแน่นขอบ (marginal densities) ของ

$$(\hat{\beta}_i \mid \sigma_{\beta_i}, \delta_i = 0) \sim N(0, \sigma^2_{\beta_i} + \tau_i^2) \qquad \text{และ}$$

$$(\hat{\beta}_i \mid \sigma_{\beta_i}, \delta_i = 0) \sim N(0, \sigma^2_{\beta_i} + c_i^2\tau_i^2)$$

กำหนดให้ $t_i\sigma_{\beta_i}$ แทนจุดตัด

X_i แทนตัวแปรอิสระตัวที่ i , i = 1,2, … , k

และ $\sigma^2_{\beta_i}$ แทนความแปรปรวนของตัวประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\hat{\beta}_i$ โดยวิธีกำลังสองน้อยสุด

เนื่องจาก

$$P(\delta_i = 1 \mid \hat{\beta}_i, \sigma_{\beta_i}) > p_i (= P(\delta_i = 1)) \quad \text{ก็ต่อเมื่อ} \quad \frac{\hat{\beta}_i}{\sigma_{\beta_i}} > t_i \tag{2.8}$$

ดังนั้นจุด t_i จะลู่เข้าสู่การแจกแจงแบบที (t – distribution) ซึ่งจะสอดคล้องกับการเพิ่มขึ้นของความน่าจะเป็นขอบ นั่นคือ X_i ควรจะอยู่ในตัวแบบ ดังนั้น t_i ที่มีค่ามาก ๆ จะเป็นตัวแบบที่เหมาะสม

ความสูงสัมพัทธ์ของความหนาแน่นส่วนริม $\hat{\beta}_i$ ณ จุดศูนย์ คือ

$$r_i \equiv \sqrt{\frac{\sigma^2_{\beta_i}/\tau^2_i + c^2_i}{\sigma^2_{\beta_i}/\tau^2_i + 1}} \qquad (2.9)$$

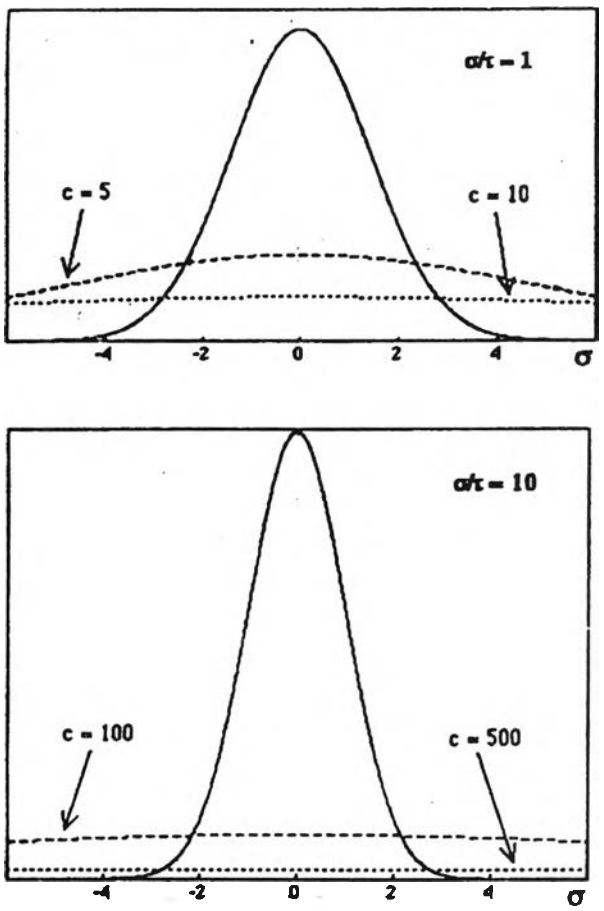
เมื่อ r_i แทนความน่าจะเป็นภายหลังขอบ (marginal posterior probability) ของ X_i ที่อยู่ในตัวแบบทั้ง ๆ ที่ค่า $\hat{\beta}_i = 0$

แม้ว่าการเลือกค่าคงที่ $\tau_1, \ldots, \tau_k$ และ $c_1, \ldots, c_k$ ด้วยวิธีการลู่เข้ากึ่งอัตโนมัติจะทำให้ได้ค่า t_i และ r_i ที่เหมาะสม แต่ขั้นตอนต่าง ๆ ค่อนข้างยุ่งยากและเสียเวลาในการคำนวณมาก ดังนั้นในการวิจัยครั้งนี้จะใช้วิธีการกำหนดให้ σ_{β_i}/τ_i และ c_i เป็นค่าคงที่ ซึ่งจากงานวิจัยของจอร์จ และแมคคัลลอช (George and McCulloch ,1993) ได้ชี้ให้เห็นว่าการกำหนดให้ σ_{β_i}/τ_i และ c_i เป็นค่าคงที่เป็นวิธีการหนึ่งที่จะทำให้ได้ค่า t_i และ r_i ที่เหมาะสม นอกจากการกำหนดดังกล่าวจะทำให้ได้ผลลัพธ์ง่ายและรวดเร็วแล้ว ยังจะมีผลทำให้การแปลงขนาด (rescaling) ของ X_i มีความคงที่มาก และยังเป็นผลลัพธ์ที่ดีเหมือนกับการกำหนดด้วยวิธีการลู่เข้ากึ่งอัตโนมัติ เพื่อให้ได้ข้อสนเทศมากขึ้นจะพิจารณากำหนดให้ σ_{β_i}/τ_i และ c_i มีค่าต่าง ๆ ดังนี้ $(\sigma_{\beta_i}/\tau_i, c_i)$ = (1,5) (1,10) (10,100) และ (10,500) เนื่องจากค่า $(\sigma_{\beta_i}/\tau_i, c_i)$ = (1,5) จะให้การแจกแจงแบบปกติที่มีการกระจายของพารามิเตอร์แคบเกินไปเมื่อเปรียบเทียบกับ $(\sigma_{\beta_i}/\tau_i, c_i)$ = (1,10) และค่า $(\sigma_{\beta_i}/\tau_i, c_i)$ = (10,500) จะให้การแจกแจงแบบปกติที่มีการกระจายของพารามิเตอร์กว้างเกินไปเมื่อเปรียบเทียบกับ $(\sigma_{\beta_i}/\tau_i, c_i)$ = (10,100) ซึ่งการกำหนดค่าคงที่ $(\sigma_{\beta_i}/\tau_i, c_i)$ ดังกล่าวจะทำให้ได้ค่า (t_i, r_i) ที่เหมาะสมคือ (2.4,1.7) (2.7,2.3) (2.1,3.2) และ (2.8,6.8) ตามลำดับ [1] ฟังก์ชันความหนาแน่นขอบ (marginal density function) ที่สอดคล้องกับการกำหนดค่าคงที่ $(\sigma_{\beta_i}/\tau_i, c_i)$ ดังกล่าวแสดงในรูปที่ 2.1 ซึ่งจะเห็นได้ว่า

---

[1] George, E.I. and McCulloch, R.E., "Variable Selection Via Gibbs Sampling," **Journal of American Statistical Association** 88 (September 1993): 881-889.

เมื่อค่า $\sigma_{\beta_i} / \tau_i$ และ c_i มีค่าเพิ่มขึ้น จะทำให้การแจกแจงของพารามิเตอร์ β_i เมื่อ $\beta_i \sim N(0,\tau_i^2)$ และ $\beta_i \sim N(0,(c_i\tau_i)^2)$ จะแยกออกจากกันอย่างชัดเจน



รูปที่ 2.1 แสดงความหนาแน่นขอบของการแจกแจง $N(0, \sigma_{\beta_i}^2 + \tau_i^2)$ และ $N(0, \sigma_{\beta_i}^2 + c_i^2\tau_i^2)$ เมื่อ $(\sigma_{\beta_i} / \tau_i, c_i)$ = (1,5) (1,10) (10,100) (10,500)

2.6.2.3 **การกำหนดหลักเกณฑ์ของเมทริกซ์สหสัมพันธ์ก่อน** (Prior Correlation Matrix : R)

เมทริกซ์ R เป็นเมทริกซ์สหสัมพันธ์ที่เป็นความรู้เดิมเกี่ยวกับ $\underset{\sim}{\beta}$ เมื่อกำหนด $\underset{\sim}{\delta}$ การเลือกเมทริกซ์ R ที่แตกต่างกันจะมีผลกระทบต่อเมทริกซ์ความแปรปรวนร่วมภายหลัง (posterior covariance matrix) ของ $\underset{\sim}{\beta}$ ภายใต้การแจกแจง $f(\underset{\sim}{\beta} \mid \underset{\sim}{y}, \sigma, \underset{\sim}{\delta})$ ดังนี้

$$f(\underset{\sim}{\beta} \mid \underset{\sim}{y}, \sigma, \underset{\sim}{\delta}) \;\alpha\; (\sigma^{-2} X'X + D_{\delta}^{-1} R^{-1} D_{\delta}^{-1})^{-1}$$

ถ้า $R = \mathrm{I}$ สมาชิกใน $\underset{\sim}{\beta}$ จะเป็นอิสระต่อกันภายใต้การแจกแจง $f(\underset{\sim}{\beta} \mid \underset{\sim}{\delta})$ ซึ่งสหสัมพันธ์ภายหลัง (posterior correlation) จะน้อยกว่ากรณีที่ $R \propto (X'X)^{-1}$ สำหรับการวิจัยครั้งนี้จะกำหนดให้ $R \propto (X'X)^{-1}$ ซึ่งจะมีผลให้สหสัมพันธ์ภายหลังจะมีลักษณะเป็นสหสัมพันธ์อย่างมีแบบแผน (design correlation)

2.6.2.4 **การกำหนดค่าระดับชั้นความเสรี υ_δ และ λ_δ**

การกำหนดค่า υ_δ และ λ_δ สำหรับการแจกแจงก่อนแบบแกมมาผกผันดังสมการ (2.6) จะใช้แนวคิดของการวางแผนการทดลอง (experimental design) ซึ่งการกำหนดหลักเกณฑ์ที่ไม่ถูกต้อง (เช่น $\upsilon_\delta = 0$) จะทำให้ได้ค่า σ^2 ที่ไม่เหมาะสมเนื่องจากค่าของ σ^2 จะอยู่รอบ ๆ ศูนย์ ดังนั้นเราจึงกำหนดค่า υ_δ ให้สอดคล้องกับ**หลักเกณฑ์ที่ไม่ทราบข้อมูลมาก่อน** (non informative prior) โดยให้ υ_δ มีค่าใกล้เคียง 2[1] ส่วนค่าของ λ_δ เป็นฟังก์ชันลดของจำนวนสมาชิกใน $\underset{\sim}{\delta}$ ที่มีค่าไม่เป็นศูนย์ และ $\left(\frac{\upsilon_\delta}{\upsilon_\delta - 2}\right)\lambda_\delta$ เป็นการประมาณโดยหลักเกณฑ์ของ σ^2

ในการวิจัยครั้งนี้จะกำหนดให้ υ_δ มีค่า 1.99 และ λ_δ เป็นค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบที่ได้ ซึ่งค่า λ_δ นี้จะมีการเปลี่ยนแปลงโดยขึ้นอยู่กับตัวแบบที่ได้ เนื่องจากค่า λ_δ จะขึ้นอยู่กับขนาดของค่าพยากรณ์ที่ได้

---

[1] Chipman, H., Hamada, M. and Wu, C.F.J. "A Bayesian Variable Selection approach for Analyzing Designed Experiments with Complex Aliasing," University of Chicago, 1996.

2.6.3 **การคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น** (Stochastic Search Variable Selection) [1]

การคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น (Stochastic Search Variable Selection (SSVS)) จะใช้วิธีการสุ่มแบบกิบส์ (Gibbs Sampling) ซึ่งเป็นกรณีหนึ่งของโครงสร้างแบบลูกโซ่มาร์คอฟ (Markov Chain Structure) การสร้างเลขสุ่มโดยอาศัยโครงสร้างดังกล่าวเรียกว่า การสร้างเลขสุ่มแบบลูกโซ่มาร์คอฟ (Markov Chain Simulation)

แนวคิดของการสร้างเลขสุ่มแบบลูกโซ่มาร์คอฟ คือ การสร้างแนวเดินเชิงสุ่ม (random walk) ในปริภูมิของพารามิเตอร์ $\underset{\sim}{\theta}$ ที่ลู่เข้าสู่การแจกแจงที่คงที่ (stationary distribution) ซึ่งก็คือการแจกแจงภายหลังร่วม (joint posterior distribution) $f(\underset{\sim}{\theta} \mid \underset{\sim}{y})$ การสร้างเลขสุ่มวิธีนี้เป็นการสร้างกระบวนการมาร์คอฟ (Markov Process) ของการแจกแจงที่คงที่คือ $f(\underset{\sim}{\theta} \mid \underset{\sim}{y})$ ที่สนใจ และมีรัน (run) ในการสร้างเลขสุ่มให้มีความยาวเพียงพอเพื่อให้การแจกแจงของเลขสุ่มในปัจจุบันเข้าใกล้กับการแจกแจงที่คงที่ได้อย่างเพียงพอ

วิธีการสุ่มแบบกิบส์หรือวิธีการสุ่มที่มีเงื่อนไขอื่น (Alternative Conditional Sampling) เป็นโครงสร้างของลูกโซ่มาร์คอฟเมื่อกำหนดให้เป็นการสุ่มแบบมีเงื่อนไข ซึ่งเป็นประโยชน์อย่างมากในการศึกษาปัญหาแบบหลายมิติ (Multidimentional problem)

ขั้นตอนของการสร้างเลขสุ่มโดยวิธีการสุ่มแบบกิบส์ เป็นดังนี้

1) นิยามเวกเตอร์พารามิเตอร์ $\underset{\sim}{\theta}$ ที่ถูกแบ่งออกเป็นเวกเตอร์ย่อยจำนวน d เวกเตอร์ ดังนี้

$$\underset{\sim}{\theta}_m = (\underset{\sim}{\theta}_{1m_1}, \underset{\sim}{\theta}_{2m_2}, \ldots, \underset{\sim}{\theta}_{dm_d})'$$

2) ในแต่ละรอบของการสุ่มแบบกิบส์ รอบที่ t ใด ๆ จะได้ค่าของเวกเตอร์ย่อยทั้งหมด d เวกเตอร์ย่อยตามลำดับ

$$\underset{\sim}{\theta}^t_m = (\underset{\sim}{\theta}^t_{1m_1}, \underset{\sim}{\theta}^t_{2m_2}, \ldots, \underset{\sim}{\theta}^t_{dm_d})' \quad \text{เมื่อ } t = 1, 2, \ldots$$

---

[1] นุชรินทร์ ทิพยวรรณากร, "การเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบที่คัดเลือกตัวแปรด้วยวิธีเบส์เซียน วิธีการกำจัดตัวแปรแบบถอยหลัง และวิธีการถดถอยแบบขั้นบันได ในการวิเคราะห์ความถดถอยพหุนามแบบลำดับขั้น," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2540), หน้า 18-19.

3) ค่าสุ่มของแต่ละเวกเตอร์ย่อยจะถูกกำหนดเงื่อนไขค่าสุ่มของเวกเตอร์ย่อยอื่นๆ ทั้งหมด นั่นคือ ค่าของเวกเตอร์ย่อยปัจจุบันถูกสุ่มจากการแจกแจงแบบมีเงื่อนไขเมื่อกำหนดเวกเตอร์ย่อยอื่นทั้งหมดของ $\underset{\sim}{\theta}$

$$f(\underset{\sim}{\theta}_i \mid \underset{\sim}{\theta}_{-i}^{t-1}, \underset{\sim}{y})$$

เมื่อ $\underset{\sim}{\theta}_{-i}^{t-1}$ แทนเวกเตอร์ย่อยทั้งหมดของ $\underset{\sim}{\theta}$ ยกเว้น $\underset{\sim}{\theta}_i$ ซึ่งเป็นค่าสุ่มปัจจุบัน กล่าวคือ

$$\underset{\sim}{\theta}_{-j}^{t-1} = (\underset{\sim}{\theta}_1^t, \ldots, \underset{\sim}{\theta}_{j-1}^t, \underset{\sim}{\theta}_{j+1}^{t-1}, \ldots, \underset{\sim}{\theta}_d^{t-1})'$$

ดังนั้นแต่ละเวกเตอร์ย่อย θ_j จะถูกปรับเงื่อนไขในค่าของ θ ค่าล่าสุด นั่นคือค่าเวกเตอร์ย่อยที่ถูกสุ่มมาในรอบที่ t จะถูกปรับค่าด้วยค่าของเวกเตอร์ย่อยที่ปรับแล้วในรอบที่ t และค่าของเวกเตอร์ทั้งหมดที่มีอยู่ในรอบที่ $t-1$

2.6.4 **ขั้นตอนของวิธีการคัดเลือกตัวแปรของเบส์** (Bayesian Variable Selection Method (BVS))

ขั้นตอนแรกของวิธีการนี้เริ่มต้นจากสมการถดถอยแบบเต็มรูป คือ ประกอบด้วยตัวแปรอิสระครบทุกตัว แล้วทำการสุ่มแบบกิบส์ เพื่อสร้างลำดับ

(2.10) $$\underset{\sim}{\delta}^1, \underset{\sim}{\delta}^2, \ldots, \underset{\sim}{\delta}^j, \ldots$$

ลำดับในสมการ (2.10) จะลู่เข้าสู่การแจกแจงของ $\underset{\sim}{\delta} \sim f(\underset{\sim}{\delta} \mid \underset{\sim}{y})$ ซึ่งเป็นลำดับที่มีความน่าจะเป็นค่อนข้างสูง เพราะบรรจุข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปร โดย $\underset{\sim}{\delta}^j$ ใด ๆ ที่มีความน่าจะเป็นสูงสุด ทำให้ง่ายต่อการตัดสินใจในการเลือกตัวแบบที่ดีที่สุด

เราใช้วิธีการสุ่มแบบกิบส์เพื่อสร้างเวกเตอร์เป็นลำดับแบบกิบส์ ดังนี้

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

เมื่อ $\underset{\sim}{\beta}^0, \sigma^0$ เป็นค่าเริ่มต้นที่ได้จากการประมาณค่าด้วยวิธีกำลังสองน้อยสุดของสมการ (1.1)

และ $\underset{\sim}{\delta}^0$ ถูกกำหนดค่าเริ่มต้นให้เป็น $\underset{\sim}{\delta}^0 \equiv (1,1,\ldots,1)'$

ในแต่ละรอบของการสุ่มค่าของลำดับ $\underset{\sim}{\beta}^{i}$, σ^{i}, $\underset{\sim}{\delta}^{i}$ ได้จากการสร้างค่าที่สอดคล้องกับขั้นตอนต่อไปนี้

1) สุ่มเวกเตอร์ของสัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}^{i}$ จากการแจกแจง

(2.12) $$f(\underset{\sim}{\beta}^{i} \mid \underset{\sim}{y}, \sigma^{i-1}, \underset{\sim}{\delta}^{i-1}) = N_{p}\left(A_{\delta^{i-1}}\left(\sigma^{i-1}\right)^{-2} X'X\underset{\sim}{\beta}_{LS}, A_{\delta^{i-1}}\right)$$

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

และ $$D_{\delta^{i-1}}^{-1} = diag[(a_1\tau_1)^{-1}, (a_2\tau_2)^{-1}, \ldots, (a_k\tau_k)^{-1}]$$

2) สุ่มค่าความแปรปรวน σ^{i} จากการแจกแจง

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

3) ค่าเวกเตอร์ $\underset{\sim}{\delta}^{i}$ ได้จากการแยกองค์ประกอบ (componentwise) โดยการสุ่มอย่างต่อเนื่องกัน (sampling consecutively) จากการแจกแจงแบบมีเงื่อนไขดังนี้

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

เมื่อ $$\delta_{-i}^{j} = \left(\delta_{i}^{j}, \ldots, \delta_{i-1}^{j}, \delta_{i+1}^{j}, \ldots, \delta_{k}^{j}\right)$$

จะเห็นได้ว่าการแจกแจงในสมการ (2.14) จะไม่ขึ้นอยู่กับค่าของ $\underset{\sim}{y}$

ในสมการ (2.14) δ_{i}^{j} แต่ละตัวจะมีการแจกแจงแบบแบร์นูลลี(Bernoulli distribution) ด้วยความน่าจะเป็น

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

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

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

ความยาวของลำดับในสมการ (2.10) จะเพิ่มขึ้นเรื่อย ๆ ทำให้การแจกแจงของค่าที่แท้จริงของ $\underset{\sim}{\delta}$ จะลู่เข้าสู่การแจกแจงภายหลัง $f(\underset{\sim}{\delta} \mid \underset{\sim}{y})$ การลู่เข้านี้จะเป็นไปอย่างรวดเร็ว ถ้า $f(\underset{\sim}{\delta} \mid \underset{\sim}{y})$ เป็นค่าสูงสุด ตัวแบบที่มีน้ำหนักมาก ๆ มีจำนวนไม่มาก ตัวแบบเหล่านี้จะมีความแม่นยำสูง เมื่อ $f(\underset{\sim}{\delta} \mid \underset{\sim}{y})$ บรรจุข้อมูลส่วนใหญ่ของการคัดเลือกตัวแปร

ข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปรจะอยู่ในลำดับของสมการ (2.11) ค่าของ $\underset{\sim}{\delta}$ ที่สอดคล้องกับกลุ่มย่อยของ $X_1, X_2, \ldots, X_k$ จะมีความถี่สูงสุด เพราะว่าลำดับย่อยนี้จะเหมาะสมกับตัวแบบที่มีค่าความน่าจะเป็นภายหลังสูงสุดภายใต้ $f(\underset{\sim}{\delta} \mid \underset{\sim}{y})$ กล่าวคือ ตัวแบบที่เหมาะสมจะเป็นตัวแบบที่มีความถี่สูงสุดและจะมีค่าความน่าจะเป็นภายหลังสูงสุดด้วย

2.7 วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA))

วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) สำหรับตัวแบบการถดถอยเชิงเส้น เสนอโดยราฟเทอร์รี่ เมดิแกน และ โฮเอ็ททิง (Raftery Madigan and Hoeting ,1997) เป็นวิธีการพิจารณาหาค่าความน่าจะเป็นภายหลัง (posterior probability) สำหรับทุก ๆ ตัวแบบที่เราสนใจ และนำตัวแบบทุกตัวแบบที่เราสนใจมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสม

วิธีการเฉลี่ยตัวแบบของเบส์นี้เป็นการพิจารณาโดยคำนึงถึงความไม่แน่นอนของตัวแบบ (Model Uncertainly) และมีแนวคิดว่าการคัดเลือกตัวแบบที่ดีที่สุดเพียงตัวแบบเดียวจากทุกตัวแบบที่เป็นไปได้ ถือเป็นการละเลยตัวแบบอื่น ๆ ซึ่งบ่อยครั้งเราจะพบว่ามีตัวแบบหลายรูปแบบที่มีประสิทธิภาพในการพยากรณ์ใกล้เคียงกัน แต่วิธีการคัดเลือกตัวแบบ และวิธีการคัดเลือกตัวแปรอิสระที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน เป็นการนำตัวแบบเพียงรูปแบบเดียวมาใช้ในการพยากรณ์เท่านั้น ดังนั้นเมื่อพิจารณาถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบแล้ว วิธีการเฉลี่ยตัวแบบของเบส์น่าจะทำให้ได้ค่าพยากรณ์ที่มีประสิทธิภาพมากกว่าค่าพยากรณ์ที่ได้จากตัวแบบเดียว ซึ่งจากงานวิจัยของนักสถิติหลายท่าน เช่น เมดิแกน และ ราฟเทอร์รี่ (Madigan and Raftery ,1994) เมดิแกน และ ยอร์ค (Madigan and York ,1995) และ ราฟเทอร์รี่ (Raftery ,1996) เป็นต้น ได้ชี้ให้เห็นว่าความไม่แน่นอนของตัวแบบเป็นสิ่งที่สำคัญและควรนำมาพิจารณาในการคัดเลือกตัวแบบสำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ

การแจกแจงความน่าจะเป็นภายหลังของสิ่งที่สนใจ เมื่อมีข้อมูลจะเป็นดังนี้

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

เมื่อ Δ เป็นปริมาณของสิ่งที่สนใจ เช่น ค่าพยากรณ์ที่สนใจ

D เป็นข้อมูลของตัวแบบที่สนใจ

และ M_1 , ... M_k เป็นตัวแบบที่พิจารณา

ซึ่งสมการ (2.18) **เป็นการเฉลี่ยการแจกแจงความน่าจะเป็นภายหลัง ถ่วงน้ำหนักด้วยความน่าจะเป็นภายหลังของแต่ละตัวแบบ โดยเรียกแนวคิดนี้ว่า** "การเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging (BMA))"

ความน่าจะเป็นภายหลังของตัวแบบ M_k **คือ**

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

โดยที่

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

ซึ่งเป็นความควรจะเป็นขอบ (marginal likelihood) **ของตัวแบบ** M_k

เมื่อ $\underset{\sim}{\theta}_k$ เป็นเวกเตอร์ของพารามิเตอร์ สำหรับการวิเคราะห์ความถดถอย 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

และ $p(M_k)$ เป็นความน่าจะเป็นก่อน สำหรับตัวแบบ M_k

ค่าเฉลี่ยและความแปรปรวนภายหลังของสิ่งที่สนใจ Δ เป็นดังนี้

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

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

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

นั่นคือจากสมการ (2.21) เราสามารถสรุปได้ว่า ถ้าสิ่งที่เราสนใจ Δ เป็นค่าพยากรณ์ $\hat{Y}$ ค่าคาดหมายของค่าพยากรณ์ $\hat{Y}$ เมื่อมีข้อมูล คือ การเฉลี่ยค่าพยากรณ์ของแต่ละตัวแบบถ่วงน้ำหนักด้วยความน่าจะเป็นภายหลังของแต่ละตัวแบบ M_k จะทำให้เราได้ค่าพยากรณ์ $\hat{Y}$ ตามที่ต้องการ

นอกจากนี้ในงานวิจัยของ เมดิแกน และราฟเทอรี่ (Madigan and Raftery ,1994) ยังแสดงให้เห็นว่าการเฉลี่ยตัวแบบจะทำให้การพยากรณ์มีประสิทธิภาพสูงกว่าการใช้ตัวแบบเดียวในการพยากรณ์ โดยใช้กฎของคะแนนลอการิทึม และแนวคิดเกี่ยวกับรูปแบบข้อสนเทศของคูลล์แบ็ค-ไลเบอร์ (Kullback-Leibler ,1951) จะได้ว่า

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

ซึ่งเป็นการยืนยันว่าวิธีการเฉลี่ยตัวแบบจะมีประสิทธิภาพในการพยากรณ์สูงกว่าการใช้ตัวแบบเดียวในการพยากรณ์

2.7.1 ความน่าจะเป็นก่อน (prior probability) และความน่าจะเป็นภายหลังขอบ (marginal posterior probability) ของตัวแบบ

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

2.7.1.1 ความน่าจะเป็นก่อน (prior probability) ของตัวแบบ

ความน่าจะเป็นก่อนของแต่ละตัวแบบจะอยู่ในรูปแบบดังนี้

$$p(M_i) = \prod_{j=1}^{k} p_j^{\delta_{ij}} \left(1 - p_j\right)^{1-\delta_{ij}} \tag{2.23}$$

เมื่อ p_j เป็นความน่าจะเป็นที่ตัวแปรอิสระที่ j จะรวมอยู่ในตัวแบบ M_i

δ_{ij} เป็นตัวบ่งชี้ที่มีค่า 0 หรือ 1

โดย 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 $(\frac{1}{2})$ กล่าวคือเป็นหลักเกณฑ์แบบสม่ำเสมอ (uniform prior) หรือตัวแปรอิสระทุกตัวจะมีโอกาสอยู่ในตัวแบบหรือไม่อยู่ในตัวแบบเท่า ๆ กันนั่นเอง

2.7.1.2 **ความน่าจะเป็นภายหลังขอบ** (marginal posterior probability) **ของตัวแบบ**

จากสมการ (2.20) การหาค่าความน่าจะเป็นภายหลังขอบของตัวแบบ M_k คือ $p(D\,|\,M_k)$ จะทำได้ดังนี้

เนื่องจากความน่าจะเป็นภายหลังขอบของตัวแบบ M ใด ๆ จะ แปรผันตามฟังก์ชันความควรจะเป็นคูณกับการแจกแจงก่อนของพารามิเตอร์ กล่าวคือ

$$p(D\,|\,M) \propto likelihood \times prior$$

ดังนั้นในขั้นแรกต้องทำการกำหนดการแจกแจงก่อน (prior distribution) ของพารามิเตอร์ซึ่งมีรายละเอียดดังนี้ [1]

การกำหนดการแจกแจงก่อนของพารามิเตอร์ได้ใช้แนวคิด Maximal Data Informative Prior (Zellner ,1971) ซึ่งเป็นแบบไม่มีข้อสนเทศ (non informative prior) กล่าวคือ เราไม่ทราบลักษณะพื้นฐานของข้อมูล หรือข้อมูลมีลักษณะคลุมเครือ ไม่ชัดเจน ซึ่งเหตุผลในการใช้การแจกแจงก่อนของพารามิเตอร์เป็นแบบไม่มีข้อสนเทศ เนื่องมาจาก

1) เป็นการใช้ข้อมูลที่มีอยู่อธิบายลักษณะของข้อมูลเอง

2) บ่อยครั้งเราไม่สามารถขจัดการแจกแจงก่อนที่เกิดจากความเชื่อของแต่ละบุคคลหรือที่เรียกว่า จิตวิสัย (subjective prior) ได้ เนื่องจากข้อจำกัดด้านเวลาและค่าใช้จ่าย หรือเนื่องจากความไม่ชำนาญของผู้กำหนด ดังนั้นการแจกแจงก่อนของ $\underset{\sim}{\beta}$ และ $log(\sigma)$ จะเป็นการแจกแจงแบบสม่ำเสมอ (uniform distribution) และมีความเป็นอิสระซึ่งกันและกัน โดยมีรูปแบบดังนี้

$$p(\underset{\sim}{\beta},\sigma) \propto \frac{1}{\sigma} \qquad ; -\infty < \underset{\sim}{\beta} < \infty \quad , \quad 0 < \sigma < \infty$$

---

[1] พจนา แววสวัสดิ์, "การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบความถดถอยพหุนามแบบติดกลุ่ม," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์-มหาวิทยาลัย, 2543), หน้า 24.

ฟังก์ชันความหนาแน่นน่าจะเป็นร่วมภายหลังสำหรับตัวแบบ M คือ

$$p(D|M) \propto likelihood \times prior$$

$$p(D|M) \propto l(\underset{\sim}{\beta}_{ML},\sigma|\underset{\sim}{y}) \times p(\underset{\sim}{\beta}_{ML},\sigma)$$

$$p(D|M) \propto \frac{1}{\sigma^{n+1}}\exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})\right]$$

$$p(D|M) \propto \frac{1}{\sigma^{n+1}}\exp\left[-\frac{1}{2\sigma^2}\left[\upsilon s^2+(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML})'x'x(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML})\right]\right]$$

เมื่อ $\hat{\underset{\sim}{\beta}}_{ML} = (x'x)^{-1}x'\underset{\sim}{y}$

$$s^2 = \frac{(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})}{\upsilon}$$

และ $\upsilon = n - p$

โดยที่ υ เป็นระดับขั้นความเสรี

n เป็นขนาดตัวอย่าง

p เป็นจำนวนพารามิเตอร์ในตัวแบบ

และ M เป็นตัวแบบที่นำมาพิจารณา

การแจกแจงความน่าจะเป็นภายหลังขอบของตัวแบบความถดถอย M คือ

$$p(D|M) \propto \int_0^\infty likelihood \times prior \; d\sigma$$

$$\propto \int_0^\infty l(\underset{\sim}{\beta}_{ML}|\sigma,M) \times p(\underset{\sim}{\beta}_{ML},\sigma) \; d\sigma$$

$$\propto \int_0^\infty \frac{1}{\sigma^n}\exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})\right]\bullet\left(\frac{1}{\sigma}\right) d\sigma$$

$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}}\exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})'(\underset{\sim}{y}-x\underset{\sim}{\beta}_{ML})\right] d\sigma$$

$$p(D\,|\,M) \propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\left(\underset{\sim}{y}-x\hat{\underset{\sim}{\beta}}_{ML}-x\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)\right)'\left(\underset{\sim}{y}-x\hat{\underset{\sim}{\beta}}_{ML}-x\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)\right)\right] d\sigma$$

$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\left(\left(\underset{\sim}{y}-x\hat{\underset{\sim}{\beta}}_{ML}\right)'\left(\underset{\sim}{y}-x\hat{\underset{\sim}{\beta}}_{ML}\right)+\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)' x'x\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)\right)\right] d\sigma$$

(เนื่องจากพจน์ $\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)' x'\left(\underset{\sim}{y}-x\hat{\underset{\sim}{\beta}}_{ML}\right)=\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)'\left[x'\underset{\sim}{y}-x'x(x'x)^{-1}x'\underset{\sim}{y}\right]=0$)

$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\left(\upsilon s^2+\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)' x'x\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)\right)\right] d\sigma$$

ใช้เทคนิคการอินทิเกรตของแกมมา

กำหนดให้ $u=\left(\upsilon s^2+\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)' x'x\left(\underset{\sim}{\beta}_{ML}-\hat{\underset{\sim}{\beta}}_{ML}\right)\right)$ จะได้ว่า

$$p(D\,|\,M) \propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{u}{2\sigma^2}\right] d\sigma$$

จากนั้นกำหนดให้ $v=\dfrac{u}{2\sigma^2}$ จะได้ว่า

$$p(D\,|\,M) \propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp[-v] \quad d\sigma$$

$$p(D\,|\,M) \propto \int_0^\infty \left(\frac{u}{2v}\right)^{-\frac{1}{2}(n+1)} \exp[-v]\left(\frac{u}{2}\right)^{\frac{1}{2}} v^{-\frac{3}{2}} dv$$

$$\propto \int_0^\infty \left(\frac{u}{2}\right)^{-\frac{1}{2}(n+1)} \bullet v^{\frac{n}{2}+\frac{1}{2}} \exp[-v]\left(\frac{u}{2}\right)^{\frac{1}{2}} \frac{v^{-\frac{3}{2}}}{2} dv$$

$$\propto \int_0^\infty \left(\frac{u}{2}\right)^{-\frac{n}{2}} \bullet v^{\frac{n}{2}-1} \exp[-v] \bullet \frac{1}{2} dv$$

$$\propto \frac{2^{\left(\frac{n-2}{2}\right)}}{u^{\frac{n}{2}}} \int_0^\infty v^{\left(\frac{n-2}{2}\right)} e^{-v} dv$$

$$p(D|M) \propto \frac{2^{\left(\frac{n-2}{2}\right)}\Gamma\left(\frac{n}{2}\right)}{\upsilon^{\frac{n}{2}}}$$

$$p(D|M) \propto 2^{\left(\frac{n-2}{2}\right)}\Gamma\left(\frac{n}{2}\right)\left(\upsilon s^2 + \left(\underset{\sim}{\beta}_{ML} - \hat{\underset{\sim}{\beta}}_{ML}\right)' x'x\left(\underset{\sim}{\beta}_{ML} - \hat{\underset{\sim}{\beta}}_{ML}\right)\right)^{-\frac{n}{2}}$$

2.7.2 การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window)

เนื่องจากวิธีการเฉลี่ยตัวแบบของเบส์จะพิจารณาหาค่าความน่าจะเป็นภายหลังสำหรับทุก ๆ ตัวแบบที่เราสนใจ ในบางครั้งปริภูมิตัวแบบที่เป็นไปได้มีขนาดใหญ่ ทำให้เสียเวลาในการคำนวณมากถึงแม้ว่าในปัจจุบันความก้าวหน้าทางเทคโนโลยีคอมพิวเตอร์จะช่วยลดปัญหานี้ได้ก็ตาม วิธีการหนึ่งที่นำมาใช้ในการค้นหาปริภูมิตัวแบบ ทำให้ปริภูมิตัวแบบที่ได้มีขนาดเล็กลงก็คือการค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) เสนอโดย เมดิแกน และราฟเทอร์รี (Madigan and Raftery ,1994) ซึ่งนำมาใช้สำหรับตัวแบบการถดถอย วิธีการนี้ถ้าตัวแบบมีความน่าจะเป็นภายหลังห่างจากตัวแบบที่ดีที่สุดตัวแบบนั้นจะไม่น่าเชื่อถือ และไม่นำมาพิจารณาอีกต่อไป

(2.24) $$A = \left\{ M_k \,\middle|\, \frac{\max_l \{p(M_l | D)\}}{p(M_k | D)} \le a \right\}$$

ดังนั้นตัวแบบที่ไม่เป็นไปตามสมการ (2.24) จะถูกตัดออกจากปริภูมิตัวแบบ ในสมการ (2.24) ค่า a เป็นค่าคงที่ที่มีขนาดใหญ่ และเป็นค่าที่ถูกเลือกโดยผู้วิจัย การเลือกค่าคงที่ a จะมีผลต่อจำนวนตัวแบบในปริภูมิตัวแบบ โดยจำนวนตัวแบบในปริภูมิตัวแบบจะแปรผันตามค่าคงที่ a ซึ่งจากงานวิจัยของเมดิแกน และราฟเทอร์รี (Madigan and Raftery ,1994) แนะนำว่าค่าคงที่ a ควรอยู่ระหว่าง 10 ถึง 100 ในการวิจัยครั้งนี้ผู้วิจัยกำหนดให้ a = 20 สำหรับการเปรียบเทียบกรณีที่ระดับนัยสำคัญ เท่ากับ 0.05 และ a = 100 สำหรับการเปรียบเทียบกรณีที่ระดับนัยสำคัญเท่ากับ 0.01 (Jeffreys ,1961 : Appendix B.)

สำหรับตัวแบบที่ได้รับการสนับสนุนน้อยจากข้อมูลเมื่อเทียบกับตัวแบบอย่างง่ายกว่า ก็จะถูกตัดออกจากปริภูมิตัวแบบดังสมการ (2.25)

(2.25) $$B = \left\{ M_k \,\middle|\, \exists M_l \in A, M_l \subset M_k, \frac{p(M_l | D)}{p(M_k | D)} > 1 \right\}$$

ซึ่งจะทำให้การเฉลี่ยตัวแบบของเบส์ในสมการ (2.18) เปลี่ยนเป็น

$$p(\Delta \mid D) = \sum_{M_k \in S} p(\Delta \mid M_k, D) \bullet p(M_k \mid D) \tag{2.26}$$

เมื่อ $S = A - B$

นั่นคือการเฉลี่ยตัวแบบจะอยู่ภายใต้ปริภูมิตัวแบบที่ผ่านการค้นหาปริภูมิตัวแบบด้วยวิธีออกแคมส์วินโดว์ แทนที่การเฉลี่ยตัวแบบภายใต้ตัวแบบที่เป็นไปได้ทั้งหมด

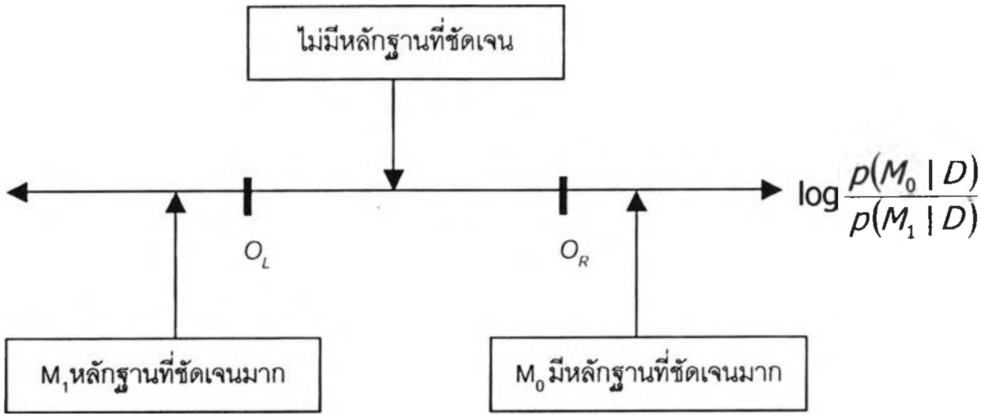
หลักการของออกส์แคมวินโดว์ (Occam 's Window) จะเป็นการยืนยันการแปลผลอัตราส่วนความน่าจะเป็นภายหลังของตัวแบบ $\dfrac{p(M_0 \mid D)}{p(M_1 \mid D)}$ เมื่อตัวแบบ M_0 เป็นตัวแบบที่มีตัวแปรอิสระน้อยกว่าตัวแบบ M_1 อย่างน้อย 1 ตัว สำหรับหลักการในการพิจารณาอัตราส่วนความน่าจะเป็นภายหลังของการเปรียบเทียบ 2 ตัวแบบ มีดังนี้

1) ถ้าล็อกออดส์ภายหลัง (Log Posterior Odds) มีค่าเป็นบวก แสดงว่าข้อมูลมีความชัดเจนในการสนับสนุนตัวแบบอย่างง่าย ดังนั้นเราจะปฏิเสธ M_1 และมาพิจารณา M_0 เราสามารถกำหนดล็อกออดส์ภายหลัง (Log Posterior Odds) เป็นค่าคงที่บวก $O_R = a$ เมื่อ a เป็นค่าคงที่ที่กำหนดจากสมการ (2.24) ก่อนที่จะปฏิเสธ M_1

2) ถ้าล็อกออดส์ภายหลัง (Log Posterior Odds)มีค่าน้อยและเป็นลบ แสดงว่าข้อมูลที่มีอยู่ไม่มีความชัดเจนในการสนับสนุนตัวแบบอย่างง่าย ดังนั้นเราจะเก็บตัวแบบทั้งสองนี้เอาไว้ก่อน

3) ถ้าล็อกออดส์ภายหลัง (Log Posterior Odds) มีค่ามากและเป็นลบ นั่นคือมีค่าน้อยกว่า $O_L = -\log(a)$ เมื่อ a เป็นค่าคงที่ที่กำหนดจากสมการ (2.24) ดังนั้นเราจะปฏิเสธ M_0 และจะพิจารณา M_1

การเปรียบเทียบตัวแบบจากหลักการข้างต้นสามารถพิจารณาได้จากรูปที่ 2.2



รูปที่ 2.2 แสดงการแปลผลของล็อกออดส์ภายหลัง(Log Psterior Odds) สำหรับ วิธีออกส์แคม วินโดว์ (Occam 's Window)

วิธีการที่จะใช้หาปริภูมิตัวแบบตามหลักการของออกส์แคม วินโดว์ (Occam 's Window) ข้างต้น เรียกว่า "กระบวนการ Up - Down" ซึ่งเป็นการใช้กระบวนการ Up และกระบวนการ Down ในการค้นหาตัวแบบ โดยกระบวนการ Up เป็นการค้นหาตัวแบบโดยการเพิ่มตัวแปร ส่วนกระบวนการ Down เป็นการค้นหาตัวแบบโดยการลดตัวแปร ซึ่งมีรายละเอียดดังนี้

<u>กระบวนการ Up – Down</u>

ขั้นตอนแรกของการเลือกตัวแบบเริ่มต้น ในทางปฏิบัติเราจะเลือกตัวแบบเริ่มต้นจากตัวแบบที่เป็นไปได้ทั้งหมด ซึ่งวิธีการในการค้นหาตัวแบบเริ่มต้นจะมี 2 วิธีนั่นคือ

"Up" เป็นกระบวนการในการค้นหาตัวแบบเริ่มต้นโดยการเพิ่มตัวแปร และ "Down" เป็นกระบวนการในการค้นหาตัวแบบเริ่มต้นโดยการลดตัวแปร เมื่อเริ่มจากตัวแบบที่ได้จากการกำหนดตัวแปร ขั้นตอนแรกจะใช้กระบวนการ Up และจะทำการกำหนดตัวแบบเริ่มต้นจากกระบวนการ Down

กำหนดให้ A และ C เป็นสมาชิกของปริภูมิตัวแบบ M เมื่อ A แทนตัวแบบที่ได้รับการยอมรับ และ C แทนตัวแบบภายใต้การพิจารณา สำหรับทั้งสองกระบวนการจะเริ่มจากการกำหนด $A = \phi$ และ C เท่ากับตัวแบบเริ่มต้นที่ได้จากการกำหนดโดยหลักการของออกแคมส์วินโดว์ (Occam 's Window) ข้างต้น

<u>กระบวนการ Down สำหรับการคัดเลือกตัวแบบ</u>

1) เลือกตัวแบบ M จาก C

2) $C \leftarrow C - M$ และ $A \leftarrow A + M$

3) เลือกตัวแบบย่อย M_0 จาก M โดยการลดตัวแปรออกจาก M

4) คำนวณค่า $B = \log \frac{p(M_0 \mid D)}{p(M \mid D)}$

5) ถ้า $B > O_R$ แล้ว $A \leftarrow A - M$ และถ้า $M_0 \notin C$ แล้ว $C \leftarrow C + M_0$

6) ถ้า $O_L \leq B \leq O_R$ แล้วถ้า $M_0 \notin C$ ดังนั้น $C \leftarrow C + M_0$

7) ถ้ามีตัวแบบย่อย M ให้กลับไปทำในขั้นตอนที่ 3

8) ถ้า $C \neq \phi$ ให้กลับไปทำในขั้นตอนที่ 1

<u>กระบวนการ Up สำหรับการคัดเลือกตัวแบบ</u>

1) เลือกตัวแบบ M จาก C

2) $C \leftarrow C - M$ และ $A \leftarrow A + M$

3) เลือกตัวแบบ M_1 จาก M โดยการเพิ่มตัวแปรเข้าไปใน M

4) คำนวณค่า $B = \log \frac{p(M \mid D)}{p(M_1 \mid D)}$

5) ถ้า $B > O_L$ แล้ว $A \leftarrow A - M$ และถ้า $M_1 \notin C$ แล้ว $C \leftarrow C + M_1$

6) ถ้า $O_L \leq B \leq O_R$ แล้วถ้า $M_1 \notin C$ ดังนั้น $C \leftarrow C + M_1$

7) ถ้ามีตัวแบบย่อย M ให้กลับไปทำในขั้นตอนที่ 3

8) ถ้า $C \neq \phi$ ให้กลับไปทำในขั้นตอนที่ 1

2.7.3 วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3))

วิธีการเฉลี่ยตัวแบบของเบส์นั้นมีสิ่งสำคัญ 2 ประการ คือ การหาค่าความน่าจะเป็นภายหลัง และการหาปริภูมิตัวแบบ สำหรับวิธีการหาปริภูมิตัวแบบนั้นนอกจากวิธีออกส์แคมวินโดว์ (Occam 's Window) ที่ได้กล่าวมาแล้วข้างต้น ราฟเทอร์รี่ เมดิแกน และ โฮเอ็ททิง (Raftery Madigan and Hoeting ,1997) ยังได้แนะนำการใช้วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (BMA_{MC3}) สำหรับการหาค่าความน่าจะเป็นภายหลัง และการหาปริภูมิตัวแบบสำหรับการเฉลี่ยตัวแบบ ซึ่งเป็นการ

นำเทคนิคมอนติคาร์โล โดยอาศัยโครงสร้างของลูกโซ่มาร์คอฟมาประยุกต์ใช้เพื่อหาองค์ประกอบของตัวแบบ หลักการของวิธี BMA_{MC3} จะเหมือนกับหลักการของวิธีการคัดเลือกตัวแปรของเบส์ (BVS) ดังนั้นจะไม่ขอกล่าวถึงแนวคิดและทฤษฎีต่าง ๆอีก (สามารถดูรายละเอียดเกี่ยวกับแนวคิดและทฤษฎีได้ในหัวข้อ 2.6) แต่วิธี BVS นั้นจะคัดเลือกตัวแบบเพียงรูปแบบเดียวที่มีความถี่ในการถูกเลือกสูงสุด ซึ่งจะเป็นตัวแบบที่มีค่าความน่าจะเป็นภายหลังสูงสุดด้วย ในขณะที่วิธีการ BMA_{MC3} จะนำตัวแบบในทุกรูปแบบที่ปรากฏอยู่ในปริภูมิตัวแบบมาเฉลี่ยกัน โดยใช้ความน่าจะเป็นภายหลังเป็นตัวถ่วงน้ำหนัก เพื่อความเข้าใจที่ชัดเจน ผู้วิจัยขอยกตัวอย่างประกอบเพื่อเปรียบเทียบความแตกต่างระหว่างวิธี BVS และวิธี BMA_{MC3} ดังนี้

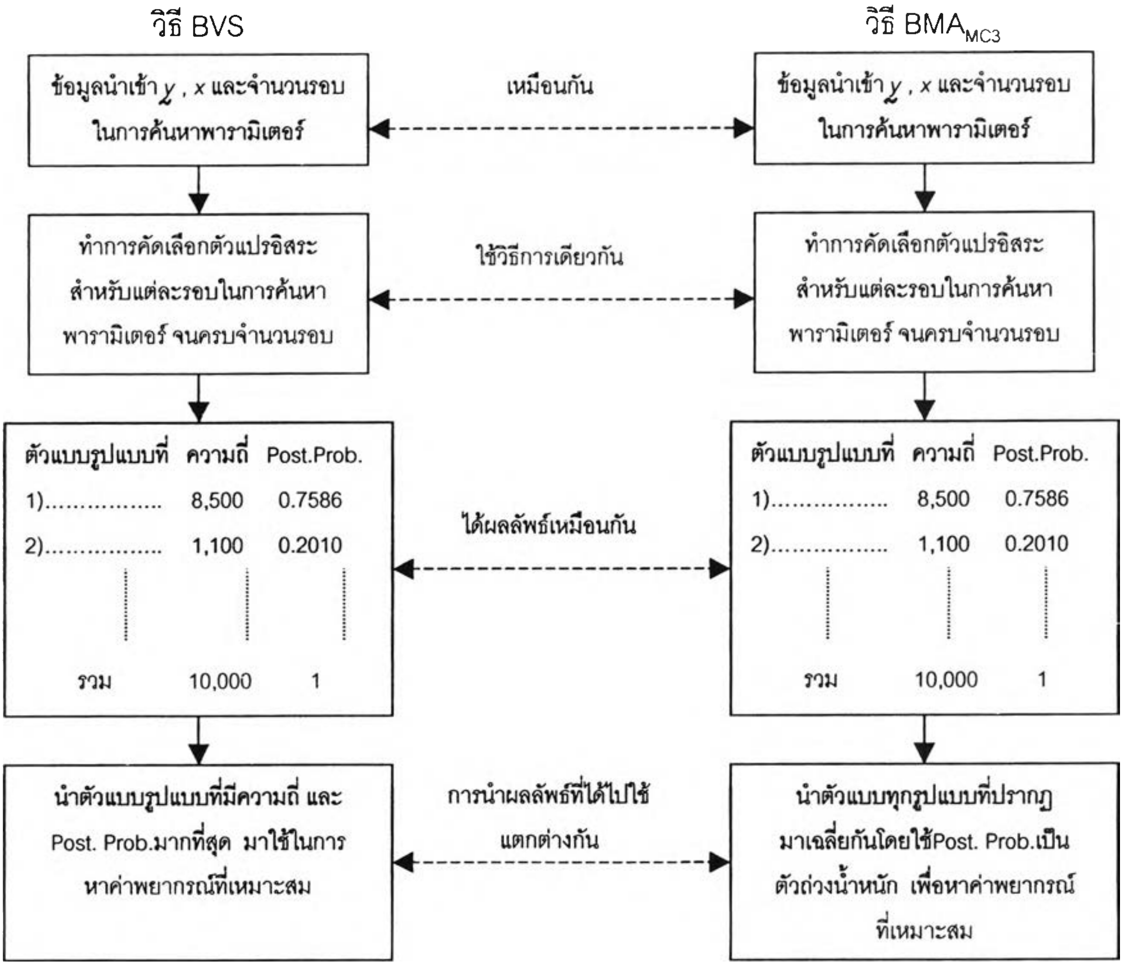
สมมติว่าใช้ข้อมูลนำเข้าชุดเดียวกัน โดยที่กำหนด

1) $\underset{\sim}{y}$ เป็นเวกเตอร์ที่บรรจุข้อมูลตัวแปรตาม

2) X เป็นเมทริกซ์ของตัวแปรอิสระ

3) จำนวนรอบในการค้นหาพารามิเตอร์ หรือความยาวของลูกโซ่มาร์คอฟ สมมติให้เท่ากับ 10,000

การเปรียบเทียบวิธี BVS และวิธี BMA_{MC3} ดังแสดงในรูปที่ 2.3



รูปที่ 2.3 แสดงการเปรียบเทียบวิธีการคัดเลือกตัวแปรของเบส์กับวิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ

จากแผนภาพการเปรียบเทียบจะเห็นได้ว่าวิธี BVS กับวิธี BMA$_{MC3}$ มีความแตกต่างกันในขั้นตอนสุดท้ายที่จะนำผลลัพธ์ไปประยุกต์ใช้เท่านั้น ซึ่งจะเป็นการเปรียบเทียบกันได้อย่างชัดเจนว่าวิธี BMA$_{MC3}$ ที่คำนึงถึงความไม่นอนของตัวแบบ จะมีประสิทธิภาพดีกว่าวิธี BVS ที่ใช้ตัวแบบเพียงรูปแบบเดียว โดยไม่คำนึงถึงความไม่แน่นอนของตัวแบบหรือไม่

2.8 วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

วิธีการนี้เป็นวิธีการคัดเลือกตัวแปรอิสระแบบเลือกเข้า โดยในแต่ละขั้นตอนของการเลือกตัวแปรอิสระเข้าสู่สมการถดถอยจะตรวจสอบตัวแปรอิสระที่มีในสมการถดถอยก่อนที่ตัวแปรอิสระตัวล่าสุดจะเพิ่มเข้าในสมการโดยวิธีการกำจัดตัวแปรอิสระแบบถอยหลังเสียก่อน ซึ่งสามารถแสดงขั้นตอนของวิธีการถดถอยแบบขั้นบันไดได้ดังนี้ [1]

1) สร้างสมการถดถอยโดยเริ่มจากไม่มีตัวแปรอิสระอยู่ในสมการ

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

3) คำนวณค่าสถิติเอฟบางส่วนสำหรับทุกตัวแปรอิสระที่ไม่อยู่ในสมการ

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

- ถ้า $F_U < F_O$ จะไม่นำตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนมากที่สุดนั้นเข้าสู่สมการถดถอย
- ถ้า $F_U > F_O$ จะนำตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนมากที่สุดนั้นเข้าสู่สมการถดถอย และคำนวณค่าสัมประสิทธิ์ของสมการถดถอยเมื่อนำตัวแปรอิสระนั้นเข้าสู่สมการ ด้วยวิธีกำลังสองน้อยสุด

5) คำนวณค่าสถิติเอฟบางส่วนของทุกตัวแปรอิสระที่อยู่ในสมการถดถอย

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

- ถ้า $F_L < F_O$ จะตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนน้อยที่สุดนั้นออกจากสมการถดถอย และคำนวณค่าสัมประสิทธิ์ของสมการถดถอย เมื่อตัดตัวแปรอิสระนั้นแล้ว

---

[1] นุชรินทร์ ทิพยวรรณากร, "การเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบที่คัดเลือกตัวแปรด้วยวิธีเบส์เซียน วิธีการกำจัดตัวแปรแบบถอยหลัง และวิธีการถดถอยแบบขั้นบันได ในการวิเคราะห์ความถดถอยพหุนามแบบลำดับขั้น," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2540), หน้า 20.

- ถ้า $F_L > F_O$ จะไม่ตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนน้อยที่สุดนั้นออกจากสมการถดถอย

7) ถ้าไม่มีตัวแปรใดเข้าและออกจากสมการถดถอยแล้ว จะได้สมการถดถอยที่เหมาะสม แต่ถ้ายังมีตัวแปรอิสระใดที่เป็นไปตามเงื่อนไขของการเข้าหรือออกจากสมการ ให้กลับไปทำในขั้นตอนที่ 3

จากขั้นตอนของวิธีการถดถอยแบบขั้นบันไดจะเห็นได้ว่าต้องอาศัยวิธีการทดสอบเอฟบางส่วน และการหาสัมประสิทธิ์สหสัมพันธ์ ซึ่งจะกล่าวถึงรายละเอียดดังนี้

2.8.1 **การทดสอบเอฟบางส่วน** (Partial F-test)[1]

การทดสอบเอฟบางส่วนเป็นการทดสอบที่ใช้ตรวจสอบนัยสำคัญของ β_j เพื่อตัดสินใจว่าตัวแปรอิสระใดควรอยู่ในสมการหรือตัวแปรอิสระใดควรตัดออกจากสมการถดถอย โดยที่ β_j จะปรากฏอยู่ ณ ตำแหน่งใดในแบบจำลองก็ได้ แต่ในทางปฏิบัติจะทำโดยถือว่าตัวแปรอิสระนั้นเข้าสู่สมการถดถอยเป็นตัวสุดท้าย

จากสมการถดถอยแสดงความสัมพันธ์ระหว่างตัวแปรตาม *(y)* กับตัวแปรอิสระ (x_j) คือ

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \tag{2.27}$$

เราสามารถหาค่าตัวประมาณ ${}_1\hat{\underset{\sim}{\beta}} = ({}_1\hat{\beta}_0, {}_1\hat{\beta}_1, {}_1\hat{\beta}_2, \ldots, {}_1\hat{\beta}_k)'$ โดยวิธีกำลังสองน้อยสุด และค่าผลบวกกำลังสองที่เกี่ยวข้องเป็นดังนี้

1) ${}_1\hat{\underset{\sim}{\beta}} = \left({}_1X'\,{}_1X\right)^{-1}{}_1X'\underset{\sim}{y}$

เมื่อ ${}_1X$ คือเมทริกซ์ของตัวแปรอิสระขนาด *n x (k+1)* ซึ่งรวมเทอมของค่าคงที่

2) $SSR_1 = {}_1\hat{\underset{\sim}{\beta}}'\,{}_1X'\underset{\sim}{y}$

3) $SSE_1 = \underset{\sim}{y}'\underset{\sim}{y} - {}_1\hat{\underset{\sim}{\beta}}'\,{}_1X'\underset{\sim}{y}$ และ

$$MSE_1 = \hat{\sigma}_1^2 = \frac{1}{n-(k+1)}\left(\underset{\sim}{y}'\underset{\sim}{y} - {}_1\hat{\underset{\sim}{\beta}}'\,{}_1X'\underset{\sim}{y}\right)$$

---

[1] นพมาศ อัครจันทโชติ, "การเปรียบเทียบวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีที่มี 2 ตัวแปรอิสระซึ่งเกิดอันตรกิริยา," (วิทยานิพนธ์ ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539), หน้า 15-17.

กำหนดให้

(2.28) $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \beta_{k+1} x_{k+1} + \ldots + \beta_p x_p + \varepsilon$$

เป็นสมการถดถอยแสดงความสัมพันธ์ระหว่างตัวแปรตาม (y) กับตัวแปรอิสระ (x_i) โดยที่ $p > k$ จากสมการ (2.28) สามารถหาค่าตัวประมาณ ${}_2\hat{\underset{\sim}{\beta}} = ({}_2\hat{\beta}_0, {}_2\hat{\beta}_1, \ldots, {}_2\hat{\beta}_k, {}_2\hat{\beta}_{k+1}, \ldots, {}_2\hat{\beta}_p)'$ โดยวิธีกำลังสองน้อยสุด และค่าผลบวกกำลังสองที่เกี่ยวข้องเป็นดังนี้

1) $${}_2\hat{\underset{\sim}{\beta}} = ({}_2X'{}_2X)^{-1}{}_2X'\underset{\sim}{y}$$

เมื่อ ${}_2X$ คือเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$ ซึ่งรวมเทอมของค่าคงที่

2) $$SSR_2 = {}_2\hat{\underset{\sim}{\beta}}'{}_2X'\underset{\sim}{y}$$

3) $$SSE_2 = \underset{\sim}{y}'\underset{\sim}{y} - {}_2\hat{\underset{\sim}{\beta}}'{}_2X'\underset{\sim}{y}$$ และ

$$MSE_2 = \hat{\sigma}_2^2 = \frac{1}{n-(p+1)}\left(\underset{\sim}{y}'\underset{\sim}{y} - {}_2\hat{\underset{\sim}{\beta}}'{}_2X'\underset{\sim}{y}\right)$$

จากผลลัพธ์ข้างต้นพบว่า Extra Sum of Squares Regression (ESSR) คือ

$$\begin{aligned} \text{ESSR} &= SSR_2 - SSR_1 \\ &= {}_2\hat{\underset{\sim}{\beta}}'{}_2X'\underset{\sim}{y} - {}_1\hat{\underset{\sim}{\beta}}'{}_1X'\underset{\sim}{y} \end{aligned}$$

ซึ่ง Extra Sum of Squares นี้เป็นค่าผลบวกกำลังสองของตัวแปรอิสระ $x_{k+1}, x_{k+2}, \ldots, x_p$ ที่เพิ่มขึ้นมาจากสมการ (2.27)

โดยอาศัยความรู้เรื่อง Distribution of Quadratic Form เราสามารถพิสูจน์ได้ว่า

$$\frac{ESSR}{\sigma^2} \sim \chi^2_{(p-k)}$$

$$\frac{SSE_2}{\sigma^2} \sim \chi^2_{(n-(p+1))}$$

ดังนั้น ค่าทดสอบเอฟบางส่วน $= \dfrac{ESSR/(p-k)\sigma^2}{SSE_2/(n-p-1)\sigma^2}$

$$= \frac{ESSR/(p-k)}{\hat{\sigma}_2^2}$$

จะมีการแจกแจงเอฟบางส่วน ณ ระดับขั้นความเสรี *(p-k , n-p-1)* [1] และเราจะปฏิเสธสมมติฐาน H_0 : $\beta_{k+1} = \beta_{k+2} = \ldots = \beta_p = 0$ ณ ระดับนัยสำคัญ α เมื่อค่าทดสอบเอฟบางส่วนมากกว่า $F_{1-\alpha,p-k,n-p-1}$

จากความรู้ในเรื่องการทดสอบเอฟบางส่วนนี้ เราสามารถนำมาประยุกต์ใช้กับวิธีการหาสมการถดถอยที่ดีที่สุด โดยวิธีการถดถอยแบบขั้นบันไดได้ดังนี้

จากสมการถดถอย $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$ เราจะหาค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squares Error (MSE)) และผลบวกกำลังสองของความถดถอย (Sum of Squares Regression) ของเฉพาะ X_j ได้จากสมการต่อไปนี้

$$MSE = \hat{\sigma}^2 = \frac{1}{(n-p-1)}\left(\underset{\sim}{y}'\underset{\sim}{y} - \hat{\underset{\sim}{\beta}}'X'\underset{\sim}{y}\right)$$

เมื่อ $SS\left(x_j \mid x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p\right) = SS\left(x_1, x_2, \ldots, x_p\right) - SS\left(x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p\right)$

ดังนั้นค่าสถิติเอฟบางส่วน คือ

$$F_c = \frac{SS\left(x_j \mid x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p\right)}{\hat{\sigma}^2}$$

โดยจะปฏิเสธสมมติฐาน H_0 : $\beta_j = 0$ ณ ระดับนัยสำคัญ α เมื่อ $F_c > F_{1-\alpha,p-k,n-p-1}$

---

[1] $U = \dfrac{\chi_1^2/\upsilon_1}{\chi_2^2/\upsilon_2} \sim F_{(\upsilon_1,\upsilon_2)}$

2.8.2 สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) [1]

สัมประสิทธิ์สหสัมพันธ์เป็นตัวสถิติที่ใช้วัดความสัมพันธ์เชิงเส้นของตัวแปรว่ามีความสัมพันธ์กันมากน้อยเพียงใด ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์จะเป็นค่าที่บอกทั้งทิศทางและขนาดของสหสัมพันธ์ โดยจะมีค่าอยู่ระหว่าง –1 ถึง 1 ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ –1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันมากในทางตรงกันข้าม ถ้าเข้าใกล้ 1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันมากในทิศทางเดียวกัน แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นศูนย์หรือเข้าใกล้ศูนย์ แสดงว่าตัวแปรทั้งสองไม่มีความสัมพันธ์เชิงเส้นต่อกัน หรือมีความสัมพันธ์เชิงเส้นกันน้อย สำหรับสัมประสิทธิ์สหสัมพันธ์ที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรสองตัวสามารถจำแนกได้ดังนี้

1) สัมประสิทธิ์สหสัมพันธ์เชิงเดียว (Simple Correlation Coefficient) เป็นตัวสถิติที่ใช้วัดความสัมพันธ์ระหว่าง 2 ตัวแปรใด ๆ ซึ่งมีสูตรดังนี้

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

เมื่อ n เป็นขนาดตัวอย่าง

y_i เป็นตัวแปรตามที่ i

และ x_i เป็นตัวแปรอิสระที่ i

2) สัมประสิทธิ์สหสัมพันธ์บางส่วน (Coefficient of Partial Correlation) เป็นตัวสถิติที่ใช้เป็นค่าวัดระดับและทิศทางของความสัมพันธ์เชิงเส้นระหว่างตัวแปร 2 ตัวแปร โดยที่ควบคุมให้ตัวแปรอื่น ๆ คงที่ เช่น กรณีตัวแปรอิสระ 3 ตัวแปร ได้แก่ x_1 , x_2 , x_3 และตัวแปรตาม y ถ้าต้องการหาสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรตาม y กับตัวแปรอิสระ x_1 โดยควบคุมให้ x_2 และ x_3 คงที่ จะใช้สัญลักษณ์ $r_{y1.23}$ สำหรับการคำนวณหาค่าของสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรคู่ใด ๆ นั้นหาได้จาก

---

[1] นพมาศ อัครจันทโชติ, "การเปรียบเทียบวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีที่มี 2 ตัวแปรอิสระซึ่งเกิดอันตรกิริยา," (วิทยานิพนธ์ ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539), หน้า 17-18.

$$r^2_{ij.1,2,3,...,i-1,i+1,...,j-1,j+1,...k} = \left(\frac{a^2_{ij}}{a_{ii}a_{jj}} \right)$$

และ

$$r_{ij.1,2,3,...i-1,i+1,...j-1,j+1,...k} = sign\sqrt{r^2}$$

$$R = \begin{bmatrix} 1 & a_{12} & \cdots & r_{1k} \\ a_{21} & 1 & \cdots & r_{2k} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix}_{k \times k}$$

$$A = R^{-1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}_{k \times k}$$

เมื่อ $r_{ij.1,2,3,...,i-1,i+1,...,j-1,j+1,...k}$ เป็นค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรที่ i และ j โดยตัวแปรอื่นคงที่

r_{ij} เป็นสัมประสิทธิ์สหสัมพันธ์เชิงเดียวระหว่าตัวแปรที่ i และ j

k เป็นจำนวนตัวแปรทั้งหมด

a_{ij} เป็นสมาชิกแถวที่ i แนวตั้งที่ j ของเมทริกซ์ผกผันของเมทริกซ์สหสัมพันธ์ (R)

R เป็นเมทริกซ์สหสัมพันธ์

และ A เป็นเมทริกซ์ผกผันของเมทริกซ์สหสัมพันธ์

ส่วนเครื่องหมายของสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระใด ๆ นั้นกำหนดตามเครื่องหมายของสัมประสิทธิ์การถดถอยของตัวแปรอิสระนั้น