บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

วิธีการวิเคราะห์ข้อมูลด้วยเบส์เซียนเป็นกระบวนการกำหนดตัวแบบความน่าจะเป็นให้กับชุดของข้อมูลและหาข้อสรุปโดยอาศัยการแจกแจงความน่าจะเป็นของพารามิเตอร์ของตัวแบบและปริมาณที่ไม่ทราบค่าเช่นค่าพยากรณ์ของค่าสังเกตชุดใหม่ ดังนั้นทฤษฎีที่เกี่ยวข้องส่วนใหญ่จะเป็นเรื่องเกี่ยวกับความน่าจะเป็นแบบมีเงื่อนไข

ความน่าจะเป็นที่มีเงื่อนไข (Conditional Probability)[1]

เป็นการศึกษาความน่าจะเป็นเมื่อทราบข้อมูลเพิ่มเติมเกี่ยวกับเหตุการณ์ที่สนใจศึกษา

<u>นิยามที่ 1</u>

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

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

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

ถ้าในการทดลองชนิดหนึ่งเหตุการณ์ A และ B สามารถเกิดขึ้นได้พร้อมกัน จะได้ว่า

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

<u>ทฤษฎีบทที่ 2</u>

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

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

---

[1] ธีระพร วีระถาวร. "ความน่าจะเป็นกับการประยุกต์."(กรุงเทพมหานคร : อักษรกราฟฟิค, 2537), หน้า 100-106

ทฤษฎีของเบส์ (Bayes' Theorem)[1]

นิยามที่ 2

เหตุการณ์ $B_1, B_2, \ldots$ จะแทนผลแบ่งกั้น (partition) ของปริภูมิตัวอย่าง S ถ้า

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

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

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

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

ทฤษฎีบทที่ 3 ทฤษฎีของเบส์ (Bayes' Theorem)

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

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

ความเป็นอิสระ (Independence)[2]

นิยามที่ 3

เหตุการณ์ 2 เหตุการณ์ A และ B เรียกว่า "เป็นอิสระต่อกัน (independent)" 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

ก) แต่ละคู่ของเหตุการณ์เป็นอิสระต่อกัน (pairwise independent) กล่าวคือ

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

ข) เหตุการณ์หนึ่งเหตุการณ์ใดเป็นอิสระจากสองเหตุการณ์ใด ๆ กล่าวคือ

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

---

[1] ธีระพร วีระถาวร. "ความน่าจะเป็นกับการประยุกต์."(กรุงเทพมหานคร : อักษรกราฟฟิค, 2537), หน้า 107-108

[2] ธีระพร วีระถาวร. "ความน่าจะเป็นกับการประยุกต์."(กรุงเทพมหานคร : อักษรกราฟฟิค, 2537), หน้า 114-1116

ตัวแบบของการถดถอยพหุนามแบบลำดับขั้น

(Hierachical Polynomial Interaction Regression Models)

ในปี ค.ศ.1994 ชิปแมน (Chipman H.) ได้เสนอวิธีการสำหรับการคัดเลือกตัวแบบที่เหมาะสมในการพยากรณ์เมื่อพิจารณาความสัมพันธ์ใดๆ ระหว่างตัวแปรอิสระโดยอาศัยความรู้เกี่ยวกับความน่าจะเป็น ดังรายละเอียดของทฤษฎีที่เกี่ยวข้องต่อไปนี้

การศึกษาการถดถอยแบบพหุนามแบบลำดับขั้นเพื่อให้การสร้างฟังก์ชันความหนาแน่นร่วมของแต่ละตัวแบบทำได้ง่ายขึ้น จะมีข้อตกลงเบื้องต้นเกี่ยวกับความสัมพันธ์ของพจน์ที่มีระดับของปัจจัยแตกต่างกัน ดังนี้

1. ปัจจัยหลัก (main effect) จะเป็นพจน์ที่มีอันดับ (order) ต่ำที่สุด
2. ปัจจัยอันตรกิริยา (interaction effect) จะเป็นพจน์ที่มีอันดับสูงกว่าปัจจัยหลัก

เมื่ออันดับของพจน์กำหนดโดยผลรวมของเลขชี้กำลัง (exponent) ของทุกปัจจัยในแต่ละพจน์เช่น AB^2, A^2B จะมีอันดับเท่ากัน ซึ่งเป็นพจน์ที่มีอันดับน้อยกว่า A^2B^2

นิยามที่ 5

Parent คือ พจน์ที่มีอันดับต่ำกว่า

นิยามที่ 6

Inherit หมายถึง พจน์ที่ได้รับการถ่ายทอดจากกลุ่มของพจน์ที่มีอันดับต่ำกว่าหรือเป็นผลคูณของพจน์ที่มีอันดับต่ำกว่า

นิยามที่ 7

Immediate Inherit หมายถึง Inherit ที่ถูกถ่ายทอดจากกลุ่มของ parent ที่เป็น Inherit ของพจน์ที่มีอันดับต่ำที่สุด เช่น AB^2 immediate inherit จาก AB และ B^2, และ AB inherit จาก A และ B

ทฤษฎีบทที่ 4

หลักการความอิสระอย่างมีเงื่อนไข (Conditional Independent Principle)

ปัจจัยที่อยู่ในอันดับเดียวกันจะเป็นอิสระต่อกัน แต่จะขึ้นอยู่กับ parent ที่สอดคล้องกัน

ทฤษฎีบทที่ 5

หลักการการถ่ายทอด (Inheritance Principle)

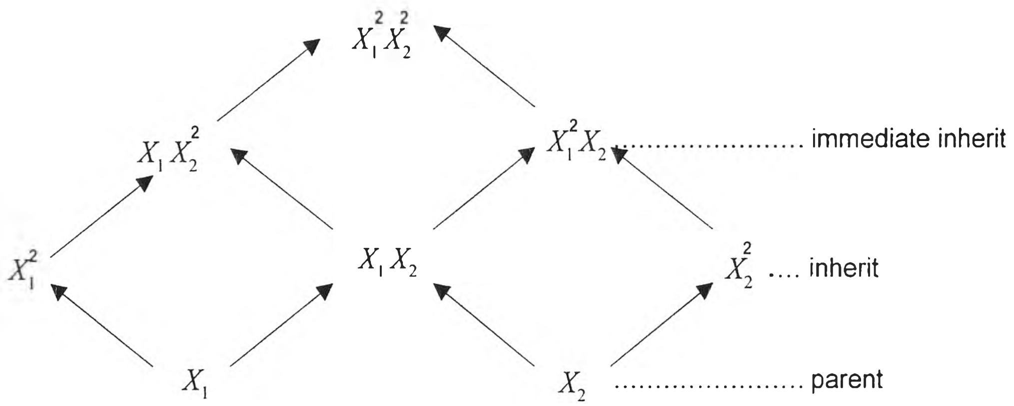
พจน์ที่มีอันดับสูงกว่าจะขึ้นอยู่กับ parent ซึ่งใช้ในการสร้างพจน์ที่มีอันดับสูงกว่านั้น

ทฤษฎีบทที่ 6

หลักการการถ่ายทอดระหว่างกลาง (Immediate Inheritance Principle)

เมื่อกำหนดพจน์ที่มีอันดับต่ำกว่า (parent) แล้ว inherit ซึ่งได้จาก parent นั้น ๆ จะเป็นอิสระจากพจน์อื่น ๆ ทั้งหมด

เราสามารถอธิบายความสัมพันธ์ดังกล่าวได้ตามรูปที่ 1



<u>รูปที่ 1</u> แสดงความสัมพันธ์ระหว่าง Parent และ Inherit

กำหนดให้

δ แทน เวกเตอร์ที่ไม่ทราบค่า มีสมาชิกเป็น **0** และ **1** ซึ่งมีขนาด $k \times 1$

h แทน อันดับสูงสุด

O_i แทน ทุกพจน์ที่มีอันดับเป็น i

$F(\delta_i)$ แทน วงศ์ (family) ของ δ_i ซึ่งประกอบด้วยพจน์ที่มีอันดับต่ำกว่า (parent) δ_i ทั้งหมด

$P(\delta_i)$ แทน parent ของ δ_i เมื่อ δ_i ที่เป็น immediate inherit

k แทน จำนวนสมาชิกใน δ

ดังนั้นความหนาแน่นร่วมของ δ ประกอบด้วยกลุ่มของพจน์ที่มีอันดับต่าง ๆ กัน เมื่อกำหนดกลุ่มของพจน์ที่มีอันดับต่ำกว่า ดังนี้

$$\mathrm{P}(\delta) \;=\; \prod_{i=1}^{h} \mathrm{P}(O_i \mid O_j \;;\; j = 1,..,i-1)$$

ความหนาแน่นของ O_i สามารถแยกออกเป็นความหนาแน่นของสมาชิกแต่ละตัว (δ') ในเซตของ O_i ได้โดยหลักการความอิสระอย่างมีเงื่อนไข

$$\mathrm{P}(\delta) \;=\; \prod_{i=1}^{h} \prod_{\delta' \in O_i} \mathrm{P}(\delta' \mid O_j \;;\; j = 1,..,i-1)$$

เราสามารถลดจำนวนเทอมที่มีอันดับต่ำกว่า δ ได้โดยหลักการการถ่ายทอด ทำให้มีเพียงพจน์ที่มีอันดับต่ำกว่า (parent) ซึ่งใช้ในการสร้าง $_i$ โดยเขียนให้อยู่ในรูปวงศ์ของ $_i$ ($F(\delta_i)$) ได้ดังนี้

$$\mathrm{P}(\delta) \;=\; \prod_{i=1}^{h} \prod_{\delta' \in O_i} \mathrm{P}\big(\delta' \mid F(\delta')\big)$$

$$= \prod_{i=1}^{k} \mathrm{P}\big(\delta_i \mid F(\delta_i)\big)$$

และโดยหลักการการถ่ายทอดระหว่างกลาง เราสามารถลดวงศ์ $F(\delta_i)$ ให้อยู่ในรูปกลุ่ม parent ของ $_i$ ($P(\delta_i)$) ดังนี้

$$\mathrm{P}(\delta) = \prod_{i=1}^{k} \mathrm{P}\left(\delta_i \middle| P(\delta_i)\right)$$
$$= \prod_{i=1}^{k} p_i^{\delta_i}(1-p_i)^{1-\delta_i}$$

เมื่อ $p_i = \mathrm{P}\left(\delta_i = 1 \mid P(\delta_i)\right)$ แทนความน่าจะเป็นโดยหลักเกณฑ์ของ $_i$ ซึ่งจะได้กล่าวถึงต่อไป

พิจารณาความสัมพันธ์ในรูปที่ 1 เมื่อกำหนด

ปัจจัยหลัก : X_1 และ X_2

ปัจจัยอันตรกิริยา : X_1X_2 , X_1^2 และ X_2^2

ปัจจัยพหุนามเชิงเส้น : $X_1^2X_2$ และ $X_1X_2^2$

เราสามารถหาความหนาแน่นร่วมของ δ ได้จากพจน์ที่มีอันดับต่าง ๆ กัน เมื่อกำหนดกลุ่มของพจน์ที่มีอันดับต่ำกว่า ดังนี้

$$\begin{aligned}\mathrm{P}(\delta) &= \mathrm{P}(\delta_{X_1}, \delta_{X_2}, \delta_{X_1X_2}, \delta_{X_1^2}, \delta_{X_2^2}, \delta_{X_1^2X_2}, \delta_{X_1X_2^2}) \\ &= \mathrm{P}(\delta_{X_1}, \delta_{X_2})\mathrm{P}(\delta_{X_1X_2}, \delta_{X_1^2}, \delta_{X_2^2} \middle| \delta_{X_1}, \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2X_2}, \delta_{X_1X_2^2} \middle| \delta_{X_1}, \delta_{X_2}, \delta_{X_1X_2}, \delta_{X_1^2}, \delta_{X_2^2})\end{aligned}$$

เราสามารถแยกออกเป็นฟังก์ชันความหนาแน่นของสมาชิกแต่ละตัวในกลุ่มของพจน์ที่มีอันดับเท่ากันได้โดยหลักการความอิสระอย่างมีเงื่อนไข ดังนี้

$$\begin{aligned}\mathrm{P}(\delta) &= \mathrm{P}(\delta_{X_1})\mathrm{P}(\delta_{X_2})\,\mathrm{P}(\delta_{X_1X_2} \middle| \delta_{X_1}, \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2} \middle| \delta_{X_1}, \delta_{X_2})\,\mathrm{P}(\delta_{X_2^2} \middle| \delta_{X_1}, \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2X_2} \middle| \delta_{X_1}, \delta_{X_2}, \delta_{X_1X_2}, \delta_{X_1^2}, \delta_{X_2^2}) \\ &\quad \times \mathrm{P}(\delta_{X_1X_2^2} \middle| \delta_{X_1}, \delta_{X_2}, \delta_{X_1X_2}, \delta_{X_1^2}, \delta_{X_2^2})\end{aligned}$$

เมื่อพิจารณาหลักการการถ่ายทอดแต่ละพจน์จะขึ้นอยู่กับ parent ทุกอันดับที่พจน์นั้น inherit มา ดังนี้

$$\begin{aligned}\mathrm{P}(\delta) &= \mathrm{P}(\delta_{X_1})\mathrm{P}(\delta_{X_2})\mathrm{P}(\delta_{X_1X_2} \middle| \delta_{X_1}, \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2} \middle| \delta_{X_1})\,P(\delta_{X_2^2} \middle| \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2X_2} \middle| \delta_{X_1}, \delta_{X_2}, \delta_{X_1X_2}, \delta_{X_1^2}) \\ &\quad \times \mathrm{P}(\delta_{X_1X_2^2} \middle| \delta_{X_1}, \delta_{X_2}, \delta_{X_1X_2}, \delta_{X_2^2})\end{aligned}$$

และโดยหลักการการถ่ายทอดระหว่างกลางแต่ละพจน์จะขึ้นอยู่กับ parent ที่มีอันดับต่ำกว่าเพียงอันดับเดียว ดังนี้

$$\begin{aligned}\mathrm{P}(\delta) &= \mathrm{P}(\delta_{X_1})\mathrm{P}(\delta_{X_2})\,\mathrm{P}(\delta_{X_1X_2} \middle| \delta_{X_1}, \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2} \middle| \delta_{X_1})\,\mathrm{P}(\delta_{X_2^2} \middle| \delta_{X_2}) \\ &\quad \times \mathrm{P}(\delta_{X_1^2X_2} \middle| \delta_{X_1X_2}, \delta_{X_1^2})\mathrm{P}(\delta_{X_1X_2^2} \middle| \delta_{X_1X_2}, \delta_{X_2^2})\end{aligned}$$

<u>ทฤษฎีบทที่ 7</u>

ตัวแบบการถดถอยพหุนามแบบลำดับขั้นใด ๆ จะเรียกว่าเป็นตัวแบบที่มีหลักเกณฑ์ดี ก็ต่อเมื่อถ้ามีตัวพยากรณ์ใด ๆ ในตัวแบบแล้ว ตัวพยากรณ์ทุกตัวที่เป็น parent ของตัวพยากรณ์นั้นทุกอันดับต้องอยู่ในตัวแบบด้วย

การแจกแจงความน่าจะเป็น (Probability Distribution)

1. <u>การแจกแจงความน่าจะเป็นโดยหลักเกณฑ์</u> (Prior Probability Distribution)

จากตัวแบบในสมการ (1) กำหนดให้ δ เป็นเวกเตอร์ที่ไม่ทราบค่า มีสมาชิกเป็น เป็น **0** และ **1** ขนาด $k \times 1$

สมาชิกใน δ จะสอดคล้องกับตัวแปรถดถอยในตำแหน่งตรงกัน ดังนี้

(2) $$\delta_i = \begin{cases} 0 \;;\; \beta_i \text{ มีอิทธิพลน้อย (แสดงว่าตัวพยากรณ์ตัวนี้ไม่อยู่ในตัวแบบ)} \\ 1 \;;\; \beta_i \text{ มีอิทธิพลมาก (แสดงว่าตัวพยากรณ์ตัวนี้อยู่ในตัวแบบ)} \end{cases}$$

จะได้ว่า $\delta_i \sim Ber(p_i)$, $i = 1,2,..,k$

ถ้าการแจกแจงความน่าจะเป็นโดยหลักเกณฑ์เป็นอิสระต่อกัน เราจะหาฟังก์ชันความหนาแน่นร่วมของ δ ได้ดังนี้

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

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

สำหรับตัวแบบพหุนามแบบลำดับขั้นเราจะกำหนดการแจกแจงความน่าจะเป็นโดยหลักเกณฑ์แบบผสม (mixture prior probability distribution) ของ β_i ได้ดังนี้

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

(4) นั่นคือ $$\beta_i|\delta_i \sim (1-\delta_i)N(0,\tau_i^2) + \delta_i N(0,(c_i\tau_i)^2)$$

จากสมการที่ (4) ถ้ากำหนด $\tau_i > 0$ และ $c_i >$ จะได้ว่า

ถ้า $_i = 0$ แล้ว $\beta_i \sim N(0,\tau_i^2)$ กล่าวคือ β_i จะมีความเป็นไปได้น้อยที่จะอยู่ในตัวแบบจึงกำหนดให้มีค่าประมาณ β_i เป็นศูนย์

ถ้า $_i =$ แล้ว $\beta_i \sim N(0,(c_i\tau_i)^2)$ กล่าวคือ β_i จะมีความเป็นไปได้มากที่จะอยู่ในตัวแบบ ดังนั้นค่าประมาณ β_i จะมีค่าไม่เป็นศูนย์ เมื่อ p_i เป็นค่าความน่าจะเป็นโดยหลักเกณฑ์ของ β_i ที่มีค่าประมาณไม่เป็นศูนย์ ซึ่งจะสอดคล้องกับตัวพยากรณ์ที่ i ควรจะอยู่ในตัวแบบ

จากสมการที่ (4) เป็นหลักเกณฑ์เกี่ยวกับ $\beta_i|\delta_i$ ที่มีการแจกแจงพหุแบบปกติ (Multivariate Normal Prior) ดังนี้

(5) $$\beta | \delta \sim N_k(\mathbf{0}, D_\delta R D_\delta)$$

เมื่อ $\delta = (\delta_1, \ldots, \delta_k)'$

R เป็นเมทริกซ์สหสัมพันธ์โดยหลักเกณฑ์ (Prior Correlation Matrix)

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

และ D_δ แสดงขนาดของหลักเกณฑ์เกี่ยวกับเมทริกซ์ความแปรปรวนร่วมที่สอดคล้องกับสมการที่ (4) โดยที่

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

สำหรับความแปรปรวนของความคลาดเคลื่อน (Residual Variance) σ^2 มีหลักเกณฑ์สังยุคเป็นแบบแกมมาผกผัน (Inverse Gamma Conjugate Prior) ดังนี้

(6) $$\beta \mid \delta \ \sim \ IG\left(\frac{\upsilon_\delta}{2},\frac{\upsilon_\delta\lambda_\delta}{2}\right)$$

ซึ่งจะสมมูลกับ $\frac{\upsilon_\delta\lambda_\delta}{\sigma^2} \ \sim \ \chi^2_{\upsilon_\delta}$ โดยที่ υ_δ และ λ_δ จะขึ้นอยู่กับ δ ที่เป็นไปตามความสัมพันธ์ระหว่าง β และ σ^2 กล่าวคือค่าของ σ^2 จะลดลงเมื่อมิติของ β (คือ จำนวนเทอมที่สมาชิกใน δ ไม่เป็นศูนย์) เพิ่มขึ้น

2. <u>การแจกแจงความน่าจะเป็นโดยประสบการณ์</u> (Posterior Probability Distribution)

การแจกแจงความน่าจะเป็นส่วนริมโดยประสบการณ์ (Marginal posterior distribution) แปรผันตามผลคูณระหว่างภาวะน่าจะเป็นของข้อมูล (Likelihood) และการแจกแจงความน่าจะเป็นโดยหลักเกณฑ์ (Prior distribution) ดังนี้

(7) $$f(\delta \mid y) \ \propto \ f(y \mid \delta)f(\delta)$$

เมื่อกำหนดข้อมูล y ความน่าจะเป็นโดยประสบการณ์จะปรับปรุงความน่าจะเป็นโดยหลักเกณฑ์เมื่อกำหนดแต่ละค่า δ จากค่าที่เป็นไปได้ทั้งหมด 2^k ค่า δ ที่มีความน่าจะเป็นโดยประสบการณ์ $f(\delta \mid y)$ สูง ๆ จะสอดคล้องกับตัวแบบที่เหมาะสมในการพยากรณ์

ปัจจัยต่างๆ ของที่มีผลต่อความน่าจะเป็นโดยประสบการณ์ได้แก่

1) หลักเกณฑ์เกี่ยวกับ $f(\delta)$

2) ค่าคงที่ $_1,\ldots,\tau_k$ และ $c_1,\ldots,c_k$ ของเมทริกซ์ D_δ

3) เมทริกซ์สหสัมพันธ์โดยหลักเกณฑ์ (prior correlation matrix) R ดังสมการที่ (5)

4) ค่าระดับความเป็นอิสระ υ_δ และ λ_δ ดังสมการที่ (6)

ดังรายละเอียดต่อไปนี้

2.1 <u>การกำหนดหลักเกณฑ์เกี่ยวกับ $f(\delta)$</u>

การแจกแจงส่วนริมของ δ เมื่อตัวแปรอิสระเป็นอิสระซึ่งกันและกันจะมีรูปแบบดังสมการที่ (3) ถ้ากำหนดให้มีหลักเกณฑ์แบบสม่ำเสมอ (uniform prior) แล้วตัวแปรอิสระแต่ละตัวจะมีโอกาสในการเข้าสู่ตัวแบบเท่า ๆ กัน ดังนั้น $f(\delta) = 2^{-k}$ เนื่องจากในงานวิจัยนี้เราทำการศึกษาตัวพยากรณ์ที่มีความสัมพันธ์พหุนามแบบลำดับขั้น ดังนั้นตัวพยากรณ์แต่ละตัวจะไม่มีอิสระ

กันทั้งหมด การกำหนด p_i ให้เป็นหลักเกณฑ์ที่ไม่ทราบข้อมูลมาก่อน (non informative prior) จำเป็นต้องมีข้อตกลงเบื้องต้นโดยอาศัยทฤษฎีบทที่ 4, 5 และ 6 ดังนี้

ก) การเข้าหรือออกจากตัวแบบของตัวพยากรณ์ที่มีอันดับต่ำสุดแต่ละตัวไม่มีผลกระทบกับตัวพยากรณ์ที่มีอันดับต่ำสุดตัวอื่น

ข) ตัวพยากรณ์ที่มีอันดับต่ำสุดมีโอกาสที่จะเข้าหรือออกจากตัวแบบไม่แตกต่างกัน นั่นคือกำหนดให้ตัวพยากรณ์ที่มีอันดับต่ำสุดมีการแจกแจงความน่าจะเป็นที่แบบสม่ำเสมอและเป็นอิสระต่อกัน ดังนั้นจะได้ว่า

$$p_i = \mathrm{P}(X_i = 1) = \frac{1}{2}, \quad \forall i = 1, 2, \ldots, 6$$

และโดยทฤษฎีบทที่ 2, 3 และ 4 ช่วยให้เราจะกำหนดหลักเกณฑ์ของตัวพยากรณ์ที่มีอันดับสูงขึ้นไปได้ดังนี้

ตัวพยากรณ์ที่เป็นพจน์พหุนาม

$$\mathrm{P}(X_i^2 = 1 | X_i) = \frac{\mathrm{P}(X_i^2 = 1)\mathrm{P}(X_i | X_i^2 = 1)}{\mathrm{P}(X_i^2 = 1)\mathrm{P}(X_i | X_i^2 = 1) + \mathrm{P}(X_i^2 = 0)\mathrm{P}(X_i | X_i^2 = 0)}$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2} + \left(1-\frac{1}{2}\right)\cdot\left(1-\frac{1}{2}\right)} = \frac{1}{2}$$

ตัวพยากรณ์ที่เป็นพจน์อันตรกิริยา

$$\mathrm{P}(X_i X_j = 1 | X_i, X_j) = \frac{a}{a+b}$$

เมื่อ $a = \mathrm{P}(X_i X_j = 1)\mathrm{P}(X_i, X_j | X_i X_j = 1)$

และ $b = \mathrm{P}(X_i X_j = 0)\mathrm{P}(X_i, X_j | X_i X_j = 0)$

จากนิยามที่ 3 และ 4 จะได้ว่า

$$\mathrm{P}(X_i X_j = 1 | X_i, X_j) = \frac{a}{a+b}$$

เมื่อ $a = \mathrm{P}(X_i = 1)\mathrm{P}(X_j = 1)\mathrm{P}(X_i | X_i X_j = 1)\mathrm{P}(X_j | X_i X_j = 1)$

$b = (1 - \mathrm{P}(X_i = 1)\mathrm{P}(X_j = 1))\mathrm{P}(X_i | X_i X_j = 0)\mathrm{P}(X_j | X_i X_j = 0)$

$$\mathrm{P}(X_i X_j = 1 | X_i, X_j) = \frac{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \left(1-\frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\frac{1}{2}}$$

$$= \frac{\frac{1}{16}}{\frac{1}{16}+\frac{3}{16}} = \frac{1}{4}$$

ดังนั้นเราจะกำหนดความน่าจะเป็นโดยหลักเกณฑ์แบบสม่ำเสมอภายใต้ความสัมพันธ์พหุนามแบบลำดับขั้นได้ดังนี้

(8)
$$P(X_i = 1) = \frac{1}{2} \ , \ \forall i = 1,2,...,6$$
$$P(X_i^2 = 1) = \frac{1}{2} \ , \ \forall i = 1,2,...,6$$
$$P(X_i X_j = 1) = \frac{1}{4} \ , \ \forall i,j = 1,2,...,6 \ ; \ i \neq j$$

โดยทฤษฎีของเบส์เมื่อมีข้อมูล (information) เกี่ยวกับพารามิเตอร์เพิ่มขึ้น เราสามารถปรับปรุงหลักเกณฑ์ใหม่ด้วยความน่าจะเป็นโดยประสบการณ์ที่มีอยู่ก่อน เช่น เมื่อทราบว่า δ มีการแจกแจงอื่นนอกจากการแจกแจงแบบสม่ำเสมอ เราสามารถสร้างความน่าจะเป็นโดยประสบการณ์ใหม่ เมื่อกำหนดให้ความน่าจะเป็นโดยประสบการณ์เดิมทำหน้าที่เป็นหลักเกณฑ์ใหม่ ดังนี้

ความน่าจะเป็นโดยประสบการณ์$_1 \propto$ ความน่าจะเป็นโดยหลักเกณฑ์$_1 \times$ ภาวะน่าจะเป็น$_1$

ความน่าจะเป็นโดยประสบการณ์$_2 \propto$ ความน่าจะเป็นโดยประสบการณ์$_1 \times$ ภาวะน่าจะเป็น$_2$

2.2 การกำหนดค่าคงที่ $\tau_1,\ldots,\tau_k$ และ $c_1,\ldots,c_k$

การเลือกค่า $_1,\ldots,\tau_k$ และ $c_1,\ldots,c_k$ ที่เหมาะสมจะใช้การลู่เข้าแบบกึ่งอัตโนมัติ (Semiautomatic Approach) โดยพิจารณาจุดตัด (intersection point) และความสูงสัมพัทธ์ (relative height) ณ ตำแหน่งศูนย์ของความหนาแน่นส่วนริม (marginal densities) ของ

$$(\hat{\beta}_i | \sigma_{\beta_i}, \delta_i = 0) \sim N(0, \sigma_{\beta_i}^2 + \tau_i^2) \quad \text{และ} \quad (\hat{\beta}_i | \sigma_{\beta_i}, \delta_i = 1) \sim N(0, \sigma_{\beta_i}^2 + c_i^2 \tau_i^2)$$

กำหนดให้ $t_i \sigma_{\beta_i}$ แทนจุดตัด

Z_i แทนตัวพยากรณ์ตัวที่ i ; i=1,2,...k

และให้ $\sigma_{\beta_i}^2$ แทนความแปรปรวนของตัวประมาณค่าโดยวิธีกำลังสองน้อยสุด $\hat{\beta}_i$

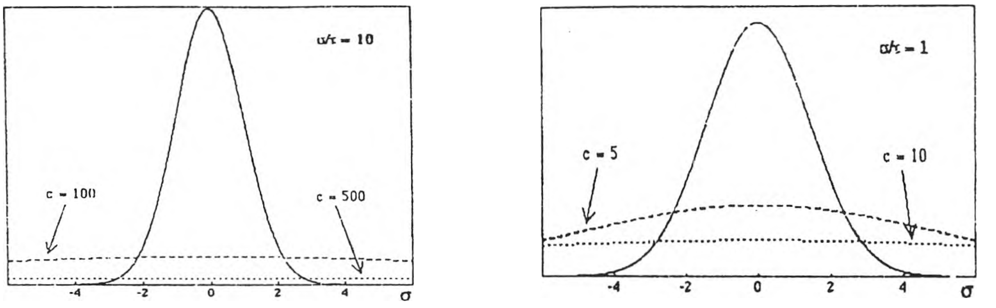
(9) เนื่องจาก $P(\delta_i = 1 | \hat{\beta}_i, \sigma_{\beta_i}) > p_i$ ก็ต่อเมื่อ $\dfrac{\hat{\beta}_i}{\sigma_{\beta_i}} > t_i$

ดังนั้นจุด t_i จะลู่เข้าสู่การแจกแจงแบบที (t-distribution) ซึ่งจะสอดคล้องกับการเพิ่มขึ้นของความน่าจะเป็นส่วนริม นั่นคือ Z_i ควรจะอยู่ในตัวแบบ ดังนั้น t_i ที่มีค่ามาก ๆ จะเป็นตัวแบบที่เหมาะสม

ความสูงสัมพัทธ์ของความหนาแน่นส่วนริมของ $\hat{\beta}_i$ ณ จุดศูนย์ จะเป็นดังนี้

(10)
$$r_i \equiv \sqrt{\frac{\sigma_{\beta_i}^2 / \tau_i^2 + c_i^2}{\sigma_{\beta_i}^2 / \tau_i^2 + 1}}$$

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



<u>รูปที่ 2</u> แสดงความหนาแน่นส่วนริมของการแจกแจง $N(0,\sigma^2+\tau^2)$ และ $N(0,\sigma^2+c^2\tau^2)$ เมื่อ $(\sigma/\tau, c) = (1,5),(1,10),(10,100),(10,500)$

เราจะเห็นว่า t_i และ r_i เป็นฟังก์ชันของ σ_{β_i}/τ_i และ c_i ดังนั้นวิธีการหนึ่งที่จะทำให้ได้ค่า t_i และ r_i ที่เหมาะสมก็คือ การกำหนดให้ σ_{β_i}/τ_i และ c_i เป็นค่าคงที่ วิธีการนี้จะมีผลให้การแปลงขนาด (rescaling) ของ Z_i มีความคงที่มาก นอกจากนี้แล้วการกำหนดดังกล่าวยังให้ผลลัพพ์ได้ง่ายและรวดเร็วขึ้น และยังคงเป็นผลลัพพ์ที่ดีเหมือนกับการกำหนดด้วยการลู่เข้าแบบกึ่งอัตโนมัติ เพื่อให้ได้ข้อสนเทศมากขึ้นจะพิจารณากำหนดให้มีค่าต่าง ๆ โดยกำหนดให้ $(\sigma_{\beta_i}/\tau_i, c_i) = (1,5)\ (1,10)\ (10,100)\ (10,500)$ เนื่องจากค่า $(\sigma_{\beta_i}/\tau_i, c_i) = (1,5)$ จะให้การแจกแจงปกติที่มีการกระจายของพารามิเตอร์แคบเกินไปเมื่อเปรียบเทียบกับ $(\sigma_{\beta_i}/\tau_i, c_i) = (1,10)$ และค่า $(\sigma_{\beta_i}/\tau_i, c_i) = (10,500)$ จะให้การแจกแจงปกติที่มีการกระจายของพารามิเตอร์ที่กว้างเกินไปเมื่อเปรียบเทียบกับค่า $(\sigma_{\beta_i}/\tau_i, c_i)$=(10,500) ค่าของ (t_i, r_i) ที่เหมาะสมคือ (2.4,1.7) (2.7,2.3) (2.1, 3.2), (2.8,6.8)[1] ฟังก์ชันความหนาแน่นส่วนริม (marginal density function) ที่สอดคล้องกับค่าที่กำหนดแสดงดังรูปที่ 2 จะเห็นว่าเมื่อค่า σ_{β_i}/τ_i และ c_i มีค่าเพิ่มขึ้น จะทำให้การแจกแจงของพารามิเตอร์ β_i เมื่อ $\beta_i \sim N(0,\tau_i^2)$ และ $\beta_i \sim N(0,(c_i\tau_i)^2)$ จะแยกออกจากกันอย่างเห็นได้ชัด

2.3 <u>การกำหนดหลักเกณฑ์ของเมทริกซ์สหสัมพันธ์</u> (Prior Correlation Matrix : R)

เมทริกซ์ R เป็นเมทริกซ์สหสัมพันธ์ที่เป็นความรู้เดิมเกี่ยวกับ δ เมื่อกำหนด δ การเลือก R ที่แตกต่างกันจะมีผลกระทบต่อเมทริกซ์ความแปรปรวนโดยประสบการณ์ (Posterior covariance matrix) ของ β ภายใต้การแจกแจง $f(\beta \mid y,\sigma,\delta)$ ดังนี้

$$\left(\sigma^{-2}Z'Z + D_\delta^{-1}R^{-1}D_\delta^{-1}\right)^{-1}$$

ถ้า $R = I$ สมาชิกใน β จะเป็นอิสระต่อกันภายใต้การแจกแจง $f(\beta \mid \delta)$ ซึ่งสหสัมพันธ์โดยประสบการณ์ (Posterior correlation) จะน้อยกว่ากรณีที่ $R \propto (Z'Z)^{-1}$ ในงานวิจัยนี้

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

กำหนดให้ $R \propto (Z'Z)^{-1}$ จะมีผลให้สหสัมพันธ์โดยประสบการณ์ จะมีลักษณะเป็นสหสัมพันธ์อย่างมีแผนแบบ(design correlation)

2.4 <u>การกำหนดค่าระดับความเป็นอิสระ υ_δ และ λ_δ</u>

ค่า υ_δ และ λ_δ สำหรับการแจกแจงโดยหลักเกณฑ์แบบแกมมาผกผันดังสมการที่ (6) โดยแนวคิดของการวางแผนการทดลอง (experimental design) การกำหนดหลักเกณฑ์ที่ไม่ถูกต้อง (เช่น υ_δ=0) จะทำให้ค่าของ σ^2 ที่ได้ไม่เหมาะสม เนื่องจากค่าของ σ^2 จะอยู่รอบ ๆ ศูนย์ ดังนั้นเราจึงกำหนดค่า υ_δ ให้สอดคล้องกับหลักเกณฑ์ที่ไม่ทราบข้อมูลมาก่อน (non informative prior) โดยให้ υ_δ มีค่าใกล้เคียงกับ 2[2] ค่าของ λ_δ เป็นฟังก์ชันลดของจำนวนสมาชิกใน δ ที่มีค่าไม่เป็นศูนย์ และ $\left(\frac{\upsilon_\delta}{\upsilon_\delta - 2}\right)\lambda_\delta$ เป็นการประมาณโดยหลักเกณฑ์ของ σ^2

ในงานวิจัยนี้กำหนดให้ υ_δ มีค่า 1.99 และ λ_δ เป็นค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบที่ได้ ซึ่งค่า λ_δ นี้จะมีค่าเปลี่ยนแปลงไปขึ้นอยู่กับตัวแบบที่ได้ เนื่องจากค่า λ_δ จะขึ้นอยู่กับขนาดของค่าพยากรณ์ที่ได้

การคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น
(Stochastic Search Variable Selection)

การคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น ซึ่งต่อไปนี้จะแทนด้วย SSVS จะใช้วิธีการสุ่มแบบกิบส์ (Gibbs Sampling หรือ Gibbs Sampler) ซึ่งเป็นกรณีหนึ่งของโครงสร้างแบบลูกโซ่มาร์คอฟ (Markov Chain Structure) การสร้างตัวเลขสุ่มโดยอาศัยโครงสร้างดังกล่าวเรียกว่า การสร้างเลขสุ่มแบบลูกโซ่มาร์คอฟ (Markov Chain Simulation)

แนวคิดของการสร้างเลขสุ่มแบบลูกโซ่มาร์คอฟ คือ สร้างแนวเดินเชิงสุ่ม (random walk) ในปริภูมิพารามิเตอร์ของ θ ที่ลู่เข้าสู่การแจกแจงที่คงที่ (stationary distribution) ซึ่งก็คือ การแจกแจงโดยประสบการณ์ร่วม (Joint Posterior Distribution) $f(\theta|y)$ การสร้างเลขสุ่มวิธีนี้เป็นการสร้างกระบวนการมาร์คอฟ (Markov Process) ของการแจกแจงที่คงที่คือ $f(\theta|y)$ ที่สนใจ และมีรัน (run) ในการสร้างเลขสุ่มให้ยาวเพียงพอเพื่อให้การแจกแจงของเลขสุ่มในปัจจุบันจะเข้าใกล้กับการแจกแจงที่คงที่ได้อย่างเพียงพอ

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

---

[2] 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

ขั้นตอนของการสร้างเลขสุ่มโดยวิธีการสุ่มแบบกิบส์ เป็นดังนี้

1. นิยามเวกเตอร์พารามิเตอร์ θ ที่ถูกแบ่งออกเป็นเวกเตอร์ย่อยจำนวน d เวกเตอร์ ดังนี้

$$\theta_m = (\theta_{1m_1}, \theta_{2m_2}, \ldots, \theta_{dm_d})'$$

2. ในแต่ละรอบของการสุ่มแบบกิบส์ รอบที่ t ใด ๆ จะได้ค่าของเวกเตอร์ย่อยทั้งหมด d เวกเตอร์ย่อยตามลำดับ

$$\theta^t = (\theta_{1m_1}, \theta_{2m_2}, \ldots, \theta_{dm_d})' \qquad \text{เมื่อ } t = 1,2,\ldots$$

3. ค่าสุ่มของแต่ละเวกเตอร์ย่อยจะถูกกำหนดเงื่อนไขค่าสุ่มของเวกเตอร์ย่อยอื่น ๆ ทั้งหมด นั่นคือ ค่าของเวกเตอร์ย่อยปัจจุบันถูกสุ่มจากการแจกแจงแบบมีเงื่อนไขเมื่อกำหนดเวกเตอร์ย่อยอื่นทั้งหมดของ θ

$$f(\theta_j \mid \theta_{-j}^{t-1}, y)$$

เมื่อ θ_{-j}^{t-1} แทนเวกเตอร์ย่อยทั้งหมดของ θ ยกเว้น θ_j ซึ่งเป็นค่าสุ่มปัจจุบัน กล่าวคือ

$$\theta_{-j}^{t-1} = (\theta_1^t, \ldots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \ldots, \theta_d^{t-1})$$

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

การเลือกสมการการถดถอยที่ดีที่สุด (Selection the Best Regression Equation)

การสร้างตัวแบบจากวิธีการเลือกสมการการถดถอยที่ดีที่สุด สำหรับงานวิจัยนี้มี 3 วิธีดังนี้

1. <u>วิธีการกำจัดตัวแปรแบบถอยหลัง (Backward Elimination)</u>[1]

วิธีการนี้เริ่มต้นด้วยสมการถดถอยแบบเต็มรูป คือ ประกอบด้วยตัวแปรอิสระทุกตัวที่ใช้พิจารณาแล้วจะคัดตัวแปรอิสระออกจากสมการครั้งละ 1 ตัวแปร โดยเลือกตัวแปรที่มีความสัมพันธ์กับ y น้อยที่สุด (เมื่อตัวแปรอื่น ๆ คงที่) และค่าสัมประสิทธิ์การถดถอยของตัวแปรนั้นไม่มีนัยสำคัญ จากนั้นคำนวณหาสมการถดถอยสำหรับตัวแปรอิสระที่เหลือ และคัดตัวแปรอิสระที่มีนัยสำคัญน้อยที่สุดออก จนกระทั่งไม่มีตัวแปรอิสระใดถูกคัดออกแล้วตัวแปรอิสระที่เหลืออยู่จะอยู่ในสมการถดถอยทุกตัว ซึ่งเป็นอันสิ้นสุดของวิธีการนี้ สามารถแสดงขั้นตอนได้ดังนี้

1) สร้างสมการถดถอยเต็มรูปแบบ (Full Model) ที่ประกอบด้วยตัวแปรอิสระทุกตัว

2) คำนวณหาค่า Partial F ของตัวแปรอิสระทุกตัว เสมือนว่าตัวแปรอิสระนั้นเข้าสู่สมการเป็นตัวสุดท้าย

---

[1] นพมาศ อัครจันทโชติ. “การเปรียบเทียบวิธีที่ใช้ในการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีที่มี 2 ตัวแปรอิสระ.” (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย,2537) หน้า 11-13

3) เลือกตัวแปรอิสระที่มีค่า Partial F น้อยที่สุด (F_L) นำค่า F_L เปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_0) ถ้าพบว่า

ก) $F_L < F_0$ จะตัดตัวแปรอิสระที่มีค่า Partial F น้อยที่สุดนั้นออกจากสมการ และคำนวณหาค่าสัมประสิทธิ์ของสมการถดถอยเมื่อตัดตัวแปรอิสระนั้นออกแล้ว จากนั้นกลับไปทำในขั้นตอนที่ 2

ข) $F_L > F_0$ จะหยุดกระบวนการคัดเลือกตัวแปรอิสระ และได้สมการถดถอยที่เหมาะสม

2. <u>วิธีการถดถอยแบบขั้นบันได (Stepwise Regression)</u>

วิธีการนี้เป็นวิธีการคัดเลือกตัวแปรอิสระแบบเลือกเข้า โดยในแต่ละขั้นตอนของการเลือกตัวแปรอิสระเข้าแต่ละตัว จะตรวจสอบตัวแปรอิสระที่มีอยู่ในสมการการถดถอยก่อนที่ตัวแปรอิสระตัวล่าสุดจะเพิ่มเข้าในสมการโดยวิธีการกำจัดตัวแปรแบบถอยหลังเสียก่อน สามารถแสดงเป็นขั้นตอนได้ดังนี้

1) สร้างสมการถดถอยโดยเริ่มจากไม่มีตัวแปรอิสระอยู่ในสมการ

2) เลือกตัวแปรอิสระที่มีสัมประสิทธิ์สหสัมพันธ์ (Coefficient Correlation) กับ y สูงสุดเข้าสมการเป็นตัวแรก และทดสอบว่าตัวแปรอิสระดังกล่าวมีนัยสำคัญ หากตัวแปรอิสระดังกล่าวไม่มีนัยสำคัญ จะได้สมการถดถอยที่เหมาะสมคือ $y = \bar{y}$)

3) คำนวณค่า Partial F สำหรับตัวแปรอิสระทุกตัวที่ไม่อยู่ในสมการ

4) เลือกค่า Partial F ที่มากที่สุด (F_U) นำไปเปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_0)

ก) ถ้า $F_U < F_0$ จะไม่นำตัวแปรที่มีค่า Partial F มากที่สุดเข้าสมการ

ข) ถ้า $F_U > F_0$ จะนำตัวแปรอิสระที่มีค่า Partial F มากที่สุดเข้าสมการ และคำนวณค่าสัมประสิทธิ์ของสมการถดถอยเมื่อนำตัวแปรนั้นเข้าสมการ

5) คำนวณค่า Partial F ของตัวแปรอิสระทุกตัวที่อยู่ในสมการถดถอย

6) นำค่า Partial F ทุกค่า เปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_0)

ก) ถ้า $F < F_0$ จะตัดตัวแปรที่มีค่า Partial F นั้นออกจากสมการ และคำนวณค่าสัมประสิทธิ์ของสมการถดถอยเมื่อตัดตัวแปรอิสระนั้นแล้ว

ข) ถ้า $F > F_0$ จะไม่ตัดตัวแปรที่มีค่า Partial F นั้นออกจากสมการ

7) ถ้าไม่มีตัวแปรอิสระใดเข้าและออกจากสมการแล้วจะได้สมการที่เหมาะสม แต่ถ้ายังมีตัวแปรอิสระใดที่เป็นไปตามเงื่อนไขของการเข้าหรือออกจากสมการ ให้กลับไปทำในขั้นตอนที่ 3

3. <u>วิธีการคัดเลือกตัวแปรของเบส์เซียน (Bayesian Variable Selection)</u>

วิธีการคัดเลือกตัวแปรของเบส์เซียนเป็นกระบวนการกำหนดความน่าจะเป็นให้กับตัวแบบโดยใช้หลักการค้นหาตัวแปรด้วยความน่าจะเป็น เพื่อกำหนดตัวแบบที่เหมาะสมกับข้อมูลเมื่อทราบการแจกแจงความน่าจะเป็นของพารามิเตอร์ในตัวแบบ

วิธีการนี้เริ่มต้นด้วยสมการถดถอยแบบเต็มรูป คือ ประกอบด้วยตัวแปรอิสระทุกตัวที่ใช้พิจารณา แล้วทำการสุ่มแบบกิบส์ เพื่อสร้างลำดับ

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

ลำดับในสมการที่ (10) จะลู่เข้าสู่การแจกแจงของ $\delta \sim f(\delta \mid y)$ ซึ่งเป็นลำดับที่มีความน่าจะเป็นค่อนข้างสูง เพราะบรรจุข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปร

เราใช้วิธีการสุ่มแบบกิบส์เพื่อสร้างเวกเตอร์พารามิเตอร์ดังนี้

(12) $$\beta^0, \sigma^0, \delta^0, \beta^1, \sigma^1, \delta^1, \ldots, \beta^j, \sigma^j, \delta^j, \ldots$$

เมื่อ β^0, σ^0 เป็นค่าเริ่มต้นที่ได้จากการประมาณด้วยวิธีกำลังสองน้อยสุดของสมการที่ (1)
และ δ^0 ถูกกำหนดค่าเริ่มต้นให้เป็น $\delta^0 \equiv (1,1,\ldots,1)$

ในแต่ละรอบของการสุ่มค่าของลำดับ $\beta^j, \sigma^j, \delta^j$ ได้จากการสร้างค่าที่สอดคล้องกับขั้นตอนต่อไปนี้

1) สุ่มเวกเตอร์ของสัมประสิทธิ์ β^j จากการแจกแจง

(13) $$f(\beta^j \mid y, \sigma^{j-1}, \delta^{j-1}) = N_p\left(A_{\delta^{j-1}} (\sigma^{j-1})^{-2} Z'Z\hat{\beta}\ , A_{\delta^{j-1}}\right)$$

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

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

2) สุ่มค่าความแปรปรวนของ σ^j จากการแจกแจง

(14) $$f(\sigma^j \mid y, \beta^j, \delta^{j-1}) = IG\left(\frac{n + \upsilon_{\delta^{j-1}}}{2}, \frac{\left|Y - Z\ \ ^j\right| + \upsilon_{\delta^{j-1}} \lambda_{\delta^{j-1}}}{2}\right)$$

3) ค่าเวกเตอร์ δ^j ได้จากการแยกองค์ประกอบ (componentwise) โดยการสุ่มอย่างต่อเนื่องกัน (sampling consecutively) จากการแจกแจงแบบมีเงื่อนไขดังนี้

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

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

จะเห็นว่าการแจกแจงในสมการที่ (15) ไม่ขึ้นอยู่กับค่าของ y

ในสมการที่ (15) δ_i^j แต่ละตัวมีการแจกแจงแบบแบร์นูลลี ด้วยความน่าจะเป็น

(16 ก) $$\mathrm{P}(\delta_i^j = 1 \mid \beta^j, \sigma^j, \delta_{-i}^j) = \frac{a}{a+b}$$

(16 ข) เมื่อ $a = f(\beta^j \mid \delta_{-i}^j, \delta_i^j = 1) \times f(\sigma^j \mid \delta_{-i}^j, \delta_i^j = 1) \times f(\delta_{-i}^j, \delta_i^j = 1)$

(16 ค) และ $b = f(\beta^j \mid \delta_{-i}^j, \delta_i^j = 0) \times f(\sigma^j \mid \delta_{-i}^j, \delta_i^j = 0) \times f(\delta_{-i}^j, \delta_i^j = 0)$

ความยาวของสมการที่ (11) จะเพิ่มขึ้นเรื่อย ๆ ทำให้การแจกแจงของค่าที่แท้จริงของ δ จะลู่เข้าหาการแจกแจงโดยประสบการณ์ $f(\delta \mid y)$ การลู่เข้านี้จะเป็นไปอย่างรวดเร็วถ้า $f(\delta \mid y)$ เป็นค่าสูงสุด ตัวแบบที่มีน้ำหนักมาก ๆ มีจำนวนไม่มาก ตัวแบบเหล่านี้จะมีความแม่นยำสูงเมื่อ $f(\delta \mid y)$ บรรจุข้อมูลส่วนใหญ่ของการคัดเลือกตัวแปร

ข้อมูลที่สอดคล้องกับการคัดเลือกตัวแปรจะอยู่ในลำดับของสมการที่ (12) ค่าของ δ ที่สอดคล้องกับกลุ่มย่อยของ $X_1, X_2, \ldots, X_k$ จะมีความถี่สูงสุด เพราะว่าลำดับย่อยนี้จะเหมาะสมกับค่าที่มีความน่าจะเป็นสูงสุดภายใต้ $f(\delta \mid y)$

การประมาณค่าสัมประสิทธิ์การถดถอยพหุนามเชิงเส้นโดยวิธีกำลังสองน้อยสุด

การประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นมีรากฐานจากทฤษฎีการประมาณเชิงเส้นที่คิดขึ้นโดยนักคณิตศาสตร์ชาวเยอรมันคือ คาร์ล เฟรดริก เกาส์ (Carl Friedrich Gauss) ในปี ค.ศ. 1777-1855 และนักคณิตศาสตร์ชาวรัสเซีย คือ อังเดร แอนดรีวิช มาร์คอฟ (Andrei Andeevich Markov) ในปี ค.ศ. 1856-1922 โดยมีหลักในการประมาณค่าสัมประสิทธ์ คือ หาค่าประมาณที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (Sum Square of Error : SSE) มีค่าน้อยที่สุด

<u>นิยามที่ 8</u>

กำหนดให้ $y = X\beta + \varepsilon$ โดยที่ $\varepsilon \sim N(0, \sigma^2 I_n)$ จะได้ตัวประมาณกำลังสองน้อยสุดของ β คือ $\hat{\beta}$ ที่จะทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด โดยที่ค่าประมาณของ $\hat{\beta}$ คือ $\hat{\beta} = (X'X)^{-1}X'y$ และเมทริกซ์ความแปรปรวนร่วมของตัวประมาณ $\hat{\beta}$ คือ $Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} I_n$

การทดสอบเอฟบางส่วน (Partial F-test)[1]

การทดสอบเอฟบางส่วนเป็นการทดสอบที่ใช้ตรวจสอบนัยสำคัญของ β_j เพื่อตัดสินใจว่าตัวแปรอิสระใดควรอยู่ในสมการหรือตัวแปรอิสระใดควรตัดออกจากสมการ โดยที่ β_j จะปรากฏอยู่ ณ ตำแหน่งใดในแบบจำลองก็ได้ แต่ในทางปฏิบัติจะทำโดยถือว่าตัวแปรอิสระนั้นเข้าสู่สมการเป็นตัวสุดท้าย

ให้ z_j คือฟังก์ชันของตัวแปรอิสระ x_j และสมการแสดงความสัมพันธ์ของ y กับ z เป็นดังนี้

(16) $$y = \beta_0^* + \beta_1 z_1 + \ldots + \beta_q z_q + \varepsilon$$

โดยที่ $\beta_0^* = \beta_0 + \beta_1 \bar{z}_1 + \ldots + \beta_q \bar{z}_q$

จากสมการ (16) เราสามารถหาค่าตัวประมาณ ${}_1\hat{\beta}' = \left({}_1\hat{\beta}_1, {}_1\hat{\beta}_2, \ldots, {}_1\hat{\beta}_q\right)$ โดยวิธีกำลังสองน้อยสุดและค่าผลบวกกำลังสองที่เกี่ยวข้องได้ดังนี้

---

[1] นพมาศ อัครจันทโชติ. “การเปรียบเทียบวิธีที่ใช้ในการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีที่มี 2 ตัวแปรอิสระ.” (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539) หน้า 15-18

ก) $_1\hat{\beta}' = ({}_1Z'{}_1Z)^{-1}{}_1Z'y$ เมื่อ $_1Z$ คือเมทริกซ์ขนาด $n\times q$

ข) $SSR_1 = {}_1\hat{\beta}_1'Z'y$

ค) $SSE = y'y - {}_1\hat{\beta}_1'Z'y$

และ $MSE_1 = \sigma_1^2$

$$= \frac{1}{n-q}(y'y - {}_1\hat{\beta}'{}_1Z'y)$$

เมื่อ SSR หมายถึง Sum Squares of Regression

(17) กำหนดให้ $y = \beta_0^* + \beta_1 z_1 + \ldots + \beta_q z_q + \beta_{q+1} z_{q+1} + \ldots + \beta_p z_p + \varepsilon$

เมื่อ $\beta_0^* = \beta_0 + \beta_1\bar{z}_1 + \ldots + \beta_q\bar{z}_q + \beta_{q+1}\bar{z}_{q+1} + \ldots + \beta_p\bar{z}_p$ เป็นสมการแสดงความสัมพันธ์ของ y กับ z โดยที่ $p>q$ จากสมการ (17) สามารถหาค่าตัวประมาณ $_2\hat{\beta}' = ({}_2\hat{\beta}_1, {}_2\hat{\beta}_2, \ldots, {}_2\hat{\beta}_q, {}_2\hat{\beta}_{q+1}, \ldots, {}_2\hat{\beta}_p)$ โดยวิธีกำลังสองน้อยสุดและค่าผลบวกกำลังสอง (Sum of Squares) ที่เกี่ยวข้องได้ดังนี้

ก) $_2\hat{\beta} = ({}_2Z'Z)^{-1}{}_2Z'y$ เมื่อ $_2Z$ คือเมทริกซ์ขนาด $n\times q$

ข) $SSR_2 = {}_2\hat{\beta}'{}_2Z'y$

ค) $SSE_2 = {}_2\hat{\beta}'{}_2Z'y$

และ $MSE_2 = \sigma_2^2$

$$= \frac{1}{n-p}(y'y - {}_2\hat{\beta}'{}_2Z'y)$$

จากผลลัพธ์ข้างต้นจะพบว่า Extra Sum of Squares Regression คือ $ESSR$ โดยที่

$$ESSR = SSR_2 - SSR_1$$
$$= {}_2\hat{\beta}'{}_2Z'y - {}_1\hat{\beta}'{}_1Z'y$$

ซึ่ง Extra Sum of Squares นี้เป็นค่าผลบวกกำลังสองของตัวแปรอิสระ $z_{q+1}, z_{q+2}, \ldots, z_p$ ที่เพิ่มขึ้นมาจากสมการ (16)

โดยอาศัยความรู้เรื่อง Distribution of Quadratic Form เราสามารถพิสูจน์ได้ว่า

$$\frac{ESSR}{\sigma^2} \sim \chi^2_{(p-q)} \quad \text{และ} \quad \frac{SSE_2}{\sigma^2} \sim \chi^2_{(n-p)}$$

ดังนั้นค่าทดสอบเอฟบางส่วน $= \dfrac{ESSR/(p-q)\sigma^2}{SSE_2/(n-p)\sigma^2}$

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

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

จากความรู้ในเรื่องการทดสอบเอฟบางส่วนนี้ เราสามารถนำมาประยุกต์ใช้กับวิธีการหาสมการการถดถอยที่ดีที่สุดได้ เช่น วิธีกำจัดตัวแปรย้อนหลัง วิธีการถดถอยขั้นบันได เป็นต้น โดยการนำไปใช้ทำได้ดังนี้

จากสมการ $y = \beta_0 + \beta_1 z_1 + ... + \beta_p z_p + \varepsilon$ เราจะหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบ (Mean Squares Error) และผลบวกกำลังสองของความถดถอย (Sum of Squares Regression) ของเฉพาะ z_j จากสมการต่อไปนี้

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

$$\begin{aligned} SS&\left(z_j \middle| z_1, z_2, ..., z_{j-1}, z_{j+1}, ..., z_p\right) \\ &= SS\left(z_1, z_2, ..., z_p\right) - SS\left(z_1, z_2, ..., z_{j-1}, z_{j+1}, ..., z_p\right) \end{aligned}$$

ดังนั้น ค่าสถิติเอฟบางส่วน คือ

$$F_c = \frac{SS(z_j | z_1, z_2, ..., z_{j-1}, z_{j+1}, ..., z_p)}{\hat{\sigma}^2}$$

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

สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

สัมประสิทธิ์สหสัมพันธ์เป็นตัวสถิติที่ใช้วัดความสัมพันธ์เชิงเส้นของตัวแปรว่ามีความสัมพันธ์กันมากน้อยเพียงใด ซึ่งค่าสัมประสิทธิ์นั้นจะมีค่าอยู่ระหว่าง -1 และ 1 ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ -1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันมากในทางตรงกันข้าม ถ้าเป็น 1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันมากในทิศทางเดียวกัน ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นศูนย์หรือเข้าใกล้ศูนย์ ถือว่าตัวแปรทั้งสองไม่มีความสัมพันธ์เชิงเส้นกันหรือมีความสัมพันธ์เชิงเส้นกันน้อย

1. สัมประสิทธิ์สหสัมพันธ์เชิงเดียว (Simple Correlation Coefficient) เป็นตัวสถิติที่ใช้วัดความสัมพันธ์ของตัวแปร 2 ตัวใด 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 แทนขนาดตัวอย่าง

y_i แทนตัวแปรตามที่ i

และ x_i แทนตัวแปรอิสระที่ i

2. สัมประสิทธิ์สหสัมพันธ์บางส่วน (Coefficient of Partial Correlation) เป็นตัวสถิติที่ใช้เป็นค่าวัดระดับและทิศทางความสัมพันธ์เชิงเส้นระหว่างตัวแปรสองตัว โดยที่ควบคุมให้ ตัวแปรอื่น ๆ ที่เหลือไม่ให้มีการเปลี่ยนแปลงหรือมีค่าคงที่ เช่น ถ้าตัวแปรอิสระคือ x_1, x_2, x_3 และตัวแปร y และต้องการหาสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรตามกับตัวแปรอิสระสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง y กับ x_1 โดยควบคุมให้ x_2 และ x_3 ไม่มีการเปลี่ยนแปลงจะแทนสัญลักษณ์ $r_{y1.23}$ และสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง y กับ x_2

และ x_3 คือ $r_{y2.13}$ และ $r_{y3.12}$ ตามลำดับ สำหรับการคำนวณหาค่าของสัมประสิทธิ์สหสัมพันธ์บางส่วนคู่ใด ๆ นั้น ทำได้ดังนี้

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

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

เมื่อ $r_{ij.1,2,3,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,k}$ แทนค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรที่ i และ j โดยที่ตัวแปรอื่นคงที่

r_{ij} แทนค่าสัมประสิทธิ์สหสัมพันธ์เชิงเดียวระหว่างตัวแปรที่ i และ j

k แทนจำนวนตัวแปรทั้งหมด

และ a_{ij} แทนสมาชิกแถวที่ i แนวตั้งที่ j ของเมทริกซ์ผกผันของเมทริกซ์สหสัมพันธ์ (R)

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & & r_{2k} \\ \vdots & & & \vdots \\ r_{k1} & & \cdots & 1 \end{bmatrix}$$

$$A = R^{-1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & & & a_{2k} \\ \vdots & & & \vdots \\ a_{k1} & & \cdots & a_{kk} \end{bmatrix}$$

โดยที่ R แทนเมทริกซ์สหสัมพันธ์

และ A แทนเมทริกซ์ผกผันของเมทริกซ์สหสัมพันธ์

ส่วนเครื่องหมายกำหนดตามเครื่องหมายของสัมประสิทธิ์การถดถอย