บทที่ 1
บทนำ

ที่มาและความสำคัญของปัญหา

โดยทั่วไปงานวิจัยที่ใช้ตัวแบบการถดถอยเชิงเส้นเชิงเดียว (simple linear regression model) เพื่อใช้ในการพยากรณ์ จะใช้วิธีกำลังสองน้อยสุดในการประมาณค่าพารามิเตอร์ในตัวแบบ วิธีดังกล่าวเป็นวิธีการที่ใช้เพียงข้อมูลปัจจุบันในการประมาณค่าพารามิเตอร์ซึ่งจะทำให้เกิดความคลาดเคลื่อนของการประมาณระดับหนึ่ง ดังนั้นในการที่จะลดความคลาดเคลื่อนให้ต่ำลงจึงควรใช้ข้อมูลในอดีตของพารามิเตอร์มาช่วยในการประมาณค่า กล่าวคือเราควรใช้การแจกแจงก่อน (prior distribution) ของพารามิเตอร์มาพิจารณาด้วย โดยแนวความคิดนี้เป็นการใช้การวิเคราะห์เชิงเบส์ การวิเคราะห์ดังกล่าวมีส่วนประกอบที่สำคัญในการประมาณค่าพารามิเตอร์ 3 ส่วน คือ ข้อมูลปัจจุบันหรือฟังก์ชันความควรจะเป็น (likelihood function) เป็นข้อมูลที่ได้จากการทดลอง ข้อมูลในอดีตหรือการแจกแจงก่อน (prior distribution) เป็นข้อมูลเพิ่มเติมเกี่ยวกับพารามิเตอร์ที่ต้องการประมาณซึ่งเป็นตัวแปรสุ่มของการแจกแจง ซึ่งนักสถิติบางส่วนก็ให้ข้อวิจารณ์ว่าการกำหนดความรู้เกี่ยวกับพารามิเตอร์ในอดีตอาจเป็นความคิดแบบจิตวิสัย (subjective opinion) และข้อมูลในอนาคตหรือการแจกแจงภายหลัง (posterior distribution) เป็นข้อมูลที่เป็นผลจากการทราบข้อมูลปัจจุบันและข้อมูลในอดีต โดยใช้ความสัมพันธ์ที่ว่าการแจกแจงภายหลังแปรผันกับผลคูณของฟังก์ชันความควรจะเป็นกับการแจกแจงก่อน

งานวิจัยนี้ผู้วิจัยให้ความสำคัญในการศึกษาการแจกแจงก่อน เนื่องจากลักษณะการกระจายของการแจกแจงก่อนที่แตกต่างกันจะให้ลักษณะการแจกแจงที่แตกต่างกันของการแจกแจงภายหลังด้วย โดยที่การแจกแจงก่อนแบ่งออกได้เป็น 2 แบบ คือ การแจกแจงก่อนที่ให้ข้อมูล (informative prior distribution) ซึ่งเป็นการแจกแจงที่ให้ข้อมูลเกี่ยวกับตัวแปรสุ่มอย่างแน่ชัด และการแจกแจงก่อนที่ไม่ให้ข้อมูล (noninformative prior distribution) ซึ่งให้ข้อมูลเกี่ยวกับตัวแปรสุ่มคลุมเครือ (vague) หรือให้ข้อมูลน้อยเมื่อเปรียบเทียบกับข้อมูลที่ได้จากการทดลอง โดยผู้วิจัยจะศึกษาการแจกแจงก่อนทั้งสองแบบดังกล่าวและการแจกแจงก่อนของเจฟฟรีส์ (Jeffreys's prior distribution) โดยที่การแจกแจงก่อนของเจฟฟรีส์มีคุณสมบัติที่ไม่แปรเปลี่ยนภายใต้การแปลงพารามิเตอร์ (parameterization invariance) และให้การแจกแจงภายหลังที่มีรูปแบบที่แท้เสมอ (proper posterior distribution) กล่าวคือ ค่าอินทิกรัลของฟังก์ชันการแจกแจงหาค่าได้และมีฟังก์ชันที่ให้โมเมนต์ (moment generating function)

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบค่าประมาณพารามิเตอร์ที่ได้จากการใช้วิธีการต่างๆ ดังนี้

1.1 การวิเคราะห์เชิงเบส์ (Bayesian approach) เมื่อกำหนดการแจกแจงก่อนแบบต่างๆ ดังนี้

- การแจกแจงก่อนที่ไม่ให้ข้อมูล (noninformative prior distribution)
- การแจกแจงก่อนที่ให้ข้อมูล (informative prior distribution)
- การแจกแจงก่อนของเจฟฟรีส์ (Jeffreys's prior distribution)

1.2 วิธีกำลังสองน้อยสุด (Least Square method)

2. เพื่อศึกษาตัวประมาณเชิงเบส์เมื่อใช้การแจกแจงก่อนที่ให้ข้อมูลซึ่งมีผลกระทบจากการเปลี่ยนค่าพารามิเตอร์ของการแจกแจง

สมมติฐานของการวิจัย

ค่าประมาณเบส์ที่ได้จากการใช้การแจกแจงก่อนของเจฟฟรีส์จะมีค่าใกล้เคียงค่าพารา-มิเตอร์มากที่สุดภายใต้ขนาดตัวอย่าง เปอร์เซ็นต์ของสัมประสิทธิ์ความแปรผันของตัวแปรอิสระและส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามในทุกกรณี รองลงมาคือค่าประมาณเบส์ที่ได้จากการใช้การแจกแจงก่อนที่ให้ข้อมูล ค่าประมาณเบส์ที่ได้จากการแจกแจงก่อนที่ไม่ให้ข้อมูล และวิธีกำลังสองน้อยสุด ตามลำดับ

ขอบเขตของการวิจัย

1. ตัวแบบที่ใช้ในการวิจัย คือ

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \;, \quad i = 1, 2, \ldots, n$$

เมื่อ y_i คือ ค่าสังเกต

β_0, β_1 คือ สัมประสิทธิ์การถดถอยซึ่งเป็นพารามิเตอร์ในตัวแบบ

x_i คือ ค่าตัวแปรอิสระ

ε_i คือ ค่าความคลาดเคลื่อน

และ n คือ ขนาดตัวอย่าง

2. จากข้อ1 เมื่อกำหนดตัวแบบเป็นตัวแบบการถดถอยเชิงเส้นเชิงเดียว (simple linear

regression model) จะได้ว่า $y_1,\ldots,y_n$ เป็นค่าสังเกตที่เป็นอิสระต่อกันและมีการแจกแจงปกติเหมือนกัน จะได้ว่า

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

เมื่อ $\underset{\sim}{y} = (y_1,\ldots,y_n)'$

$\underset{\sim}{\beta} = (\beta_0, \beta_1)'$

และ $X = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \end{bmatrix}'$

3. การแจกแจงของความคลาดเคลื่อนเป็นการแจกแจงปกติ ซึ่งในงานวิจัยนี้จะศึกษาเมื่อค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.3 0.5 0.7 และ 0.9

4. ตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มมาจากการแจกแจงปกติด้วยค่าเฉลี่ยเท่ากับ 1 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.15 0.2 0.25 0.3 0.5 0.7 และ 0.9 โดยมีเปอร์เซ็นต์ของสัมประสิทธิ์ความแปรผันเท่ากับ 10% 15% 20% 25% 30% 50% 70% และ 90% ตามลำดับ

5. การแจกแจงก่อนที่ใช้มีดังต่อไปนี้

5.1 การแจกแจงก่อนที่ไม่ให้ข้อมูล (noninformative prior distribution)

เป็นการแจกแจงที่ให้ข้อมูลในอดีตเกี่ยวกับเรื่องที่สนใจคลุมเครือ (vague) และเป็นการแจกแจงที่มักจะถูกนำมาใช้ เนื่องจากข้อมูลในอดีตเกี่ยวกับพารามิเตอร์ไม่ได้ให้ข้อมูลเลยว่าพารามิเตอร์ควรจะมีค่าที่แท้จริงเท่าใด เราทราบเพียงแต่ว่าแต่ละค่าของพารามิเตอร์มีโอกาสเกิดขึ้นเท่าๆ กัน ซึ่งการแจกแจงดังกล่าวมีรูปแบบดังนี้

$$p(\underset{\sim}{\beta}) \propto k \quad , k = \text{ค่าคงที่}$$

5.2 การแจกแจงก่อนที่ให้ข้อมูล (informative prior distribution)

เป็นการแจกแจงที่ให้ข้อมูลเกี่ยวกับพารามิเตอร์อย่างแน่ชัด ซึ่งตัวแบบที่ใช้ในงานวิจัยนี้เป็นตัวแบบการถดถอยเชิงเส้นปกติเชิงเดียว ดังนั้น การแจกแจงก่อนที่ให้ข้อมูลที่มีคุณสมบัติเป็นการแจกแจงก่อนสังยุค (conjugate prior distribution) คือ การแจกแจงปกติ 2 ตัวแปร ดังนี้

$$p(\underset{\sim}{\beta}) \;=\; (2\pi)^{-1}(\det\Sigma)^{-1/2}\exp\left\{-\frac{1}{2}(\underset{\sim}{\beta}-\underset{\sim}{\overline{\beta}})'\,\Sigma^{-1}(\underset{\sim}{\beta}-\underset{\sim}{\overline{\beta}})\right\}$$

เมื่อ $\underset{\sim}{\overline{\beta}}$ เป็นเวกเตอร์ค่าเฉลี่ย และ Σ เป็นเมทริกซ์ความแปรปรวนร่วมซึ่งเป็นเมทริกซ์บวกแน่นอน (positive definite matrix) และการแจกแจงดังกล่าวจะเป็นการแจกแจงก่อนที่ไม่ให้ข้อมูลเมื่อค่าในตำแหน่งที่ i ของเส้นทแยงมุมของเมทริกซ์ Σ มีค่าเข้าใกล้อนันต์

5.3 *การแจกแจงก่อนของเจฟฟรีส์* (Jeffreys's prior distribution)

เป็นการแจกแจงที่ได้จากการใช้กฎของเจฟฟรีส์ (Jeffreys's rule) ซึ่งการแจกแจงนี้เป็นสัดส่วนกับรากที่สองของดีเทอร์มิแนนท์ของเมทริกซ์ข้อสนเทศของฟิชเชอร์ (Fisher Information Matrix) ดังนี้

$$p(\underset{\sim}{\beta}) \;\propto\; \left|I(\underset{\sim}{\beta})\right|^{1/2}$$

เมื่อ $I(\underset{\sim}{\beta})$ คือ เมทริกซ์ข้อสนเทศของฟิชเชอร์

6. การใช้การวิเคราะห์เชิงเบส์ในการหาค่าประมาณพารามิเตอร์ เราจะได้ว่า

$$p(\underset{\sim}{\beta},\sigma|\underset{\sim}{y})=\begin{cases}\dfrac{p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma)\,p(\underset{\sim}{\beta},\sigma)}{\sum\limits_{\underset{\sim}{\beta}} p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma)\,p(\underset{\sim}{\beta},\sigma)} & \text{,กรณีไม่ต่อเนื่อง}\\[2ex] \dfrac{p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma)\,p(\underset{\sim}{\beta},\sigma)}{\int_{\Theta} p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma)\,p(\underset{\sim}{\beta},\sigma)\;d\underset{\sim}{\beta}\,d\sigma} & \text{,กรณีต่อเนื่อง}\end{cases}$$

เมื่อ $p(\underset{\sim}{y}|\underset{\sim}{\beta},\sigma)$ คือ ฟังก์ชันความหนาแน่นร่วม (probability density function) ของ $\underset{\sim}{y}$

$p(\underset{\sim}{\beta},\sigma)$ คือ การแจกแจงก่อนร่วม (joint prior distribution) สำหรับ $(\underset{\sim}{\beta},\sigma)$

$p(\underset{\sim}{\beta},\sigma|\underset{\sim}{y})$ คือ การแจกแจงภายหลังร่วม (joint posterior distribution) สำหรับ $(\underset{\sim}{\beta},\sigma)$

และ Θ คือ ปริภูมิพารามิเตอร์

7. ขนาดตัวอย่างที่ใช้คือ 10 30 50 และ 100

เกณฑ์การตัดสินใจ

การเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (average mean square of error : *AMSE*) ของค่าประมาณเบส์เมื่อกำหนดการแจกแจงก่อนที่ให้ข้อมูล การแจกแจงก่อนที่ไม่ให้ข้อมูลและการแจกแจงก่อนของเจฟฟรีส์ และค่าประมาณกำลังสองน้อยสุด โดยค่าประมาณที่ดีที่สุดจะให้ค่า *AMSE* ต่ำที่สุด

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อสามารถใช้การวิเคราะห์เชิงเบส์ในตัวแบบการถดถอยเชิงเส้นเชิงเดียว ซึ่งมีความแม่นยำในการหาค่าประมาณมากขึ้นและมีความแปรปรวนของค่าประมาณสัมประสิทธิ์การถดถอยน้อยลง
2. เพื่อใช้เป็นแนวทางการศึกษาตัวแบบเชิงเส้นที่ซับซ้อนขึ้นโดยใช้การวิเคราะห์เชิงเบส์

ตัวสถิติที่ใช้ในการวิจัย

1. การหาค่าประมาณสัมประสิทธิ์การถดถอยในตัวแบบด้วยการวิเคราะห์เชิงเบส์จะหาได้จากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จากฟังก์ชันที่ให้โมเมนต์ (moment generating function) ของการแจกแจงภายหลังขอบ (marginal posterior distribution) โดยที่ฟังก์ชันที่ให้โมเมนต์และการแจกแจงขอบ (marginal distribution) จะหาได้จากทฤษฎีบทและนิยามดังต่อไปนี้

<u>นิยามที่ 1</u>

ให้ X เป็นตัวแปรสุ่ม เราเรียก μ_k ว่าเป็นโมเมนต์ที่ k รอบจุดศูนย์กลาง ถ้า

$$\mu_k = E\left(X^k\right), k=1,2,3,\ldots$$
$$= \int_{-\infty}^{\infty} x^k dF(X)$$

เมื่อ $E\left(\left|X^k\right|\right)<\infty$

<u>นิยามที่ 2</u>

ให้ X เป็นตัวแปรสุ่ม เราเรียก $M_X(t)$ ว่าเป็นฟังก์ชันที่ให้โมเมนต์ ถ้า

$$M_X(t)=\begin{cases}\sum_{\beta} e^{tx} f(x) & ,X \text{ ไม่ต่อเนื่อง} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx & ,X \text{ ต่อเนื่อง}\end{cases}$$

เมื่อ $t \in (a,b)$

<u>ทฤษฎีบทที่ 1</u>

ให้ X เป็นตัวแปรสุ่มที่มีฟังก์ชันที่ให้โมเมนต์ซึ่งมีค่าจำกัดบนช่วงเปิดใดๆ ที่คลุมศูนย์ กล่าวคือ $M_X(t) < \infty,\ t \in (-h,h)$ เมื่อ $h > 0$ จะได้ว่า

$$\mu_k = M_X^{(k)}(0),\ k = 1,2,3,\ldots$$

เมื่อ $M_X^{(k)}(0)$ คือ ค่าอนุพันธ์ของฟังก์ชันที่ให้โมเมนต์ ณ จุดศูนย์

<u>บทแทรก</u>

ให้ X เป็นตัวแปรสุ่มซึ่งมีฟังก์ชันที่ให้โมเมนต์ $M_X(t)$ และกำหนดให้ $\varphi_X(t) = \ln M_X(t)$ จะได้ว่า

ค่าเฉลี่ย $= \mu = E(X) = \varphi_X'(0)$

และความแปรปรวน $= Var(X) = \varphi_X''(0)$

<u>ทฤษฎีบทที่ 2</u>

ถ้า $\underset{\sim}{x} \sim N_m(\underset{\sim}{\mu}, \Sigma)$, $B(k \times m)$ เมื่อ $k < m$ เป็นเมทริกซ์ที่เจาะจง และ $\underset{\sim}{b}(k \times 1)$ เป็นเวกเตอร์ที่เจาะจง จะได้ว่า $B\underset{\sim}{x} + \underset{\sim}{b} \sim N_k(B\underset{\sim}{\mu} + \underset{\sim}{b}, B \Sigma B')$

พิสูจน์ (ธีระพร วีระถาวร, 2536 : 319)

<u>ทฤษฎีบทที่ 3</u>

ถ้า $\underset{\sim}{x} \sim N_m(\underset{\sim}{\mu}, \Sigma)$ จะได้ว่าการแจกแจงขอบ (marginal distribution) ของเวกเตอร์ย่อยใดๆ ของ $\underset{\sim}{x}$ ซึ่งมีสมาชิก k ตัว จะมีการแจกแจงปกติของตัวแปร k ตัว โดยเราสามารถแสดงได้ดังนี้

พิสูจน์ (ธีระพร วีระถาวร, 2536 : 319)

ให้ $\underset{\sim}{x} = \begin{bmatrix} \underset{\sim 1}{x} \\ \underset{\sim 2}{x} \end{bmatrix} \begin{matrix} k \\ m-k \end{matrix}$, $\underset{\sim}{\mu} = \begin{bmatrix} \underset{\sim 1}{\mu} \\ \underset{\sim 2}{\mu} \end{bmatrix} \begin{matrix} k \\ m-k \end{matrix}$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} \begin{matrix} k \\ m-k \end{matrix}$

$B = \begin{bmatrix} I_k & 0 \end{bmatrix}_{m-k}$, และ $\underset{\sim}{b} = \underset{\sim}{0}$

และจากทฤษฎีบทที่ 2 เราจะได้ว่า

$$B\underset{\sim}{x} + \underset{\sim}{b} = \underset{\sim 1}{x} \sim N_k(\underset{\sim 1}{\mu}, \Sigma_{11})$$

2. ค่าประมาณกำลังสองน้อยสุด ซึ่งหาได้จากการทำให้ผลรวมกำลังสองของความแตก-ต่างของค่าสังเกตกับค่าคาดหวังของค่าสังเกตมีค่าน้อยที่สุด