บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ผู้วิจัยมุ่งที่จะศึกษาถึงการจำลองตัวแบบความถดถอยเชิงลำดับชั้น เมื่อการแจกแจงความคลาดเคลื่อนสุ่มมีการแจกแจงแบบไม่ปกติ โดยวิธีประมาณที่เลือกใช้ในการศึกษามีอยู่ 2 วิธีคือ วิธี Iterative Generalized Least Square และ วิธี Restricted Iterative Generalized Least Square และศึกษาว่าตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งของฮูเบอร์ไวท์ จะสามารถใช้แก้ปัญหาการประมาณค่าคลาดเคลื่อนมาตรฐานผิดพลาดในกรณีที่ค่าคลาดเคลื่อนสุ่มไม่ได้แจกแจงแบบปกติ โดยการเปรียบเทียบด้วยค่าความเอนเอียงสัมพัทธ์ของตัวประมาณพารามิเตอร์ และค่าอัตราส่วนค่าสัมบูรณ์ของผลต่างระหว่างค่าเฉลี่ยค่าคลาดเคลื่อนมาตรฐานและค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลสัมพัทธ์ และค่าอัตราส่วนค่าสัมบูรณ์ของผลต่างระหว่างค่าเฉลี่ยค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งและค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลสัมพัทธ์

2.1 ตัวแบบเชิงเส้นแบบ 2 ระดับ (Two-Level Linear Model)

สมมติว่าข้อมูลตัวอย่างถูกสุ่มจากประชากรที่มีโครงสร้างเป็นลำดับขั้น 2 ระดับ โดยตัวอย่างในระดับที่ 2 มีทั้งหมด J กลุ่ม และในแต่ละกลุ่ม j ประกอบด้วยตัวอย่างในระดับที่ 1 n_j หน่วย

ตัวแบบระดับที่ 1 (Level-1 Model)

(1.7) $$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$$

ตัวแบบระดับที่ 2 (Level-2 Model)

(1.8 ก.) $$\beta_{0j} = \gamma_{00} + u_{0j}$$

(1.8 ข.) $$\beta_{1j} = \gamma_{01} + u_{1j}$$

ตัวแบบรวม (Combined Model)

(1.9) $$y_{ij} = \gamma_{00} + \gamma_{01}x_{ij} + u_{0j} + u_{1j}x_{ij} + r_{ij}$$

จาก (1.8) จะเรียกส่วน $\gamma_{00} + \gamma_{01}x_{ij}$ ว่า ส่วนอิทธิพลคงที่ (Fixed Part) และเรียกส่วน $u_{0j} + u_{1j}x_{ij} + r_{ij}$ ว่าส่วนอิทธิพลสุ่ม (Random part)

ข้อสมมติเบื้องต้น

1. ค่าคลาดเคลื่อนสุ่มในระดับที่ 1 (r_{ij}) มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย 0 และความแปรปรวน σ^2 กล่าวคือ $r_{ij} \overset{iid}{\sim} N\ (0,\sigma^2)$
2. ค่าคลาดเคลื่อนสุ่มในระดับที่ 2 ($\underline{u}_j$) มีการแจกแจงแบบปกติหลายตัวแปร ที่มีค่าเฉลี่ย $\underline{0}$ และเมตริกซ์ความแปรปรวนร่วม T โดยที่ $T = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}$

   เมื่อ $Var(u_{0j}) = \tau_{00}$, $Var(u_{1j}) = \tau_{11}$ และ $Cov(u_{0j}, u_{1j}) = \tau_{01}$

   กล่าวคือ $\underline{u}_j = (u_{0j}, u_{1j})' \overset{iid}{\sim} N_2(\underline{0}, T)$
3. ค่าคลาดเคลื่อนสุ่มในทั้ง 2 ระดับเป็นอิสระต่อกัน กล่าวคือ $Cov(r_{ij}, u_{qj}) = 0$, $q = 0,1$
4. ตัวแปรอิสระในระดับที่ 1 (X_{ij}) เป็นอิสระกับค่าคลาดเคลื่อนทั้ง 2 ระดับ กล่าวคือ $Cov(x_{ij}, r_{ij}) = 0$ และ $Cov(x_{ij}, \underline{u}_j) = 0$

จาก (1.9) จะได้

(1.11)
$$E(y_{ij} \mid x_{ij}) = E(\gamma_{00} + \gamma_{01}x_{ij} + u_{0j} + u_{1j}x_{ij} + r_{ij}) = \gamma_{00} + \gamma_{01}x_{ij}$$

(1.12)
$$\begin{aligned} Var(y_{ij} \mid x_{ij}) &= Var(\gamma_{00} + \gamma_{01}x_{ij} + u_{0j} + u_{1j}x_{ij} + r_{ij}) \\ &= Var(u_{0j} + u_{1j}x_{ij} + r_{ij}) \\ &= \tau_{00} + \tau_{11}(x_{ij})^2 + 2\tau_{01}x_{ij} + \sigma^2 \end{aligned}$$

และสำหรับค่าสังเกตที่แตกต่างกัน 2 ค่าภายในกลุ่มเดียวกัน

(1.13)
$$Cov(y_{ij}, y_{i'j} \mid x_{ij}, x_{i'j}) = \tau_0^2 + \tau_{01}(x_{ij} + x_{i'j}) + \tau_1^2 x_{ij}x_{i'j} \qquad ;i \neq i'$$

วิธีการประมาณที่นิยมใช้กันในการประมาณพารามิเตอร์ในตัวแบบ (1.9) คือการประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด ในงานวิจัยนี้จะใช้วิธี Iterative Generalized Least Square และวิธี Restricted Iterative Generalized Least Square ในการหาค่าประมาณพารามิเตอร์

2.2 วิธี Iterative Generalized Least Square

พิจารณาตัวแบบใน (1.9) จะสามารถเขียนในรูปเมตริกซ์ได้ดังนี้

(1.14)
$$\underline{Y} = X\underline{\gamma} + W\underline{u} + \underline{r}$$

เมื่อ
$$\underline{Y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_J \end{bmatrix}_{\sum_{j=1}^{J} n_j \times 1} , \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_J \end{bmatrix}_{\sum_{j=1}^{J} n_j \times 2} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n_J J} \end{bmatrix}_{\sum_{j=1}^{J} n_j \times 2}$$

$$,\underline{\gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \end{bmatrix}_{2\times 1} \quad ,W = \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ 0 & 0 & X_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & X_J \end{bmatrix}_{\sum_{j=1}^{J} n_j \times (2.J)} \quad ,\underline{u} = \begin{bmatrix} \underline{u}_1 \\ \underline{u}_2 \\ \vdots \\ \underline{u}_J \end{bmatrix}_{(2.J)\times 1}$$

และ $$\underline{r} = \begin{bmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \vdots \\ \underline{r}_J \end{bmatrix}_{\sum_{j=1}^{J} n_j \times 1}$$

(1.15) $$Cov(\underline{Y} \mid X\underline{\gamma}) = Cov(\underline{E}) = V$$

ถ้าทราบค่าส่วนประกอบความแปรปรวน V จะสามารถประมาณพารามิเตอร์ของอิทธิพลคงที่ได้โดยใช้วิธีกำลังสองน้อยสุดนัยทั่วไป (Generalized Least Squares) ดังนี้

(1.16) $$\underline{\hat{\gamma}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \underline{Y}$$

และเมตริกซ์ความแปรปรวนร่วมของ $\underline{\hat{\gamma}}$ คือ

(1.17) $$Cov(\underline{\hat{\gamma}}) = (X^T V^{-1} X)^{-1} X^T V^{-1} Cov(\underline{Y}) V^{-1} X (X^T V^{-1} X)^{-1} = (X^T V^{-1} X)^{-1}$$

ถ้าทราบค่าสัมประสิทธิ์ถดถอย $\underline{\gamma}$

กำหนดให้

(1.18) $$E^{*} = \underline{E} \cdot \underline{E}^T \quad \text{เมื่อ } \underline{E} = \underline{Y} - X\underline{\gamma}$$

โดยที่ค่าคาดหวังของ E^{*} มีค่าเท่ากับ V ดังนี้

(1.19) $$E(E^{*}) = E(\underline{E} \cdot \underline{E}^T) = V$$

และกำหนดให้

(1.20) $$E^{**} = vec(E^{*})$$

ซึ่งเป็นการเรียงเมตริกซ์ E^{*} ให้อยู่ในรูปเวกเตอร์โดยเรียงคอลัมม์ทุกคอลัมม์ในอยู่ในคอลัมม์แรกเพียงคอลัมม์เดียว

เราสามารถเขียนตัวแบบเชิงเส้นของ E^{**} ได้ดังนี้

(1.21) $$E^{**} = Z\underline{\theta} + \underline{R}$$

เมื่อ Z เป็น Design Matrix ของพารามิเตอร์ผลกระทบสุ่ม $\underline{\theta}$

$\underline{\theta}$ เป็นพารามิเตอร์ผลกระทบสุ่ม หรือส่วนประกอบความแปรปรวน กรณีใน (1.9)

$\underline{\theta} = (\tau_{00}, \tau_{11}, \tau_{01}, \sigma^2)'$

ดังนั้นจะสามารถประมาณส่วนประกอบความแปรปรวน V โดยวิธีสองน้อยสุดนัยทั่วไปเช่นเดียวกัน ดังนี้

(1.22) $\quad \hat{\underline{\theta}} = (Z^{*T}V^{*-1}Z^{*})^{-1}Z^{*T}V^{*-1}E^{**} \qquad$ โดยที่ $V^{*} = V \otimes V$

เมื่อ $\otimes$ คือ Kronecker Product หรือ Direct Product

และเมตริกซ์ความแปรปรวนร่วมของ $\hat{\underline{\theta}}$ คือ

(1.23)
$$Cov(\hat{\theta}) = (Z^{*T}V^{*-1}Z^{*})^{-1}Z^{*T}V^{*-1}Cov(E^{**})V^{*-1}Z^{*}(Z^{*T}V^{*-1}Z^{*})^{-1}$$
$$= 2(Z^{*T}V^{*-1}Z^{*})^{-1} \qquad \text{โดยที่} \quad Cov(E^{**}) = (V \otimes V)(I + S_N)$$

เมื่อ S_N คือ Vector Operator บน Permutation Matrix

วิธี Iterative Generalized Least Square เป็นอัลกอริทึมที่จะประมาณค่ากลับไปกลับมาระหว่าง (1.16) และ (1.21) ตัวประมาณที่ได้มาในแต่ละรอบ จะพัฒนาขึ้นลู่เข้าสู่ค่าที่เหมาะสม

ในการเลือกค่าเริ่มต้นของการประมาณ ผู้วิเคราะห์อาจเลือกค่าเริ่มต้นโดยใช้ตัวประมาณกำลังสองน้อยสุดแบบสามัญ (Ordinary Least Square Estimator) โดยในรอบแรกจะกำหนดให้ $V = \sigma^2 I$

ตัวประมาณที่ได้จากวิธีการนี้จะเป็นตัวประมาณภาวะน่าจะเป็นสูงสุด โดยเป็นตัวประมาณที่คงเส้นคงวา (Consistent) ไม่เอนเอียงเมื่อเข้าใกล้อนันต์ (Asymptotically unbiased) และมีประสิทธิภาพเมื่อเข้าใกล้อนันต์ (Asymptotically efficient) (Goldstein, 1986)

2.3 วิธี Restricted Iterative Generalized Least Square

จากที่กล่าวไว้ในหัวข้อ 2.2 ว่าตัวประมาณที่ได้จากวิธี Iterative Generalized Least Square เป็นตัวประมาณภาวะน่าจะเป็นสูงสุด ซึ่งเป็นตัวประมาณที่มีความเอนเอียง ดังนั้นค่าประมาณจะไม่เหมาะสมเมื่อขนาดตัวอย่างมีขนาดเล็ก ในปีค.ศ. 1989 ฮาร์วี โกลสเตน จึงได้พัฒนาอัลกอริทึมที่ใช้ในการประมาณพารามิเตอร์ในตัวแบบเรียกว่าวิธี Restricted Iterative Generalized Least Square ซึ่งเป็นตัวประมาณภาวะน่าจะเป็นสูงสุดแบบมีข้อจำกัด (Restricted Maximum Likelihood) โดยที่อัลกอริทึมในการประมาณยังเหมือนกับวิธี Iterative Generalized Least Square แต่ได้มีการปรับสูตรในการประมาณดังนี้

เนื่องจาก $\hat{\underline{\gamma}}$ ซึ่งเป็นตัวประมาณกำลังสองน้อยสุดนัยทั่วไปของ $\underline{\gamma}$ และถ้าทราบค่าส่วนประกอบความแปรปรวน V ดังนั้นจะได้

(1.24) $\underline{E}_R = \underline{Y} - X\hat{\underline{\gamma}}$

และจะได้

(1.25) $\underline{E}_R \underline{E}_R^T = (\underline{Y} - X\hat{\underline{\gamma}})(\underline{Y} - X\hat{\underline{\gamma}})^T$

โดยที่

(1.26) $E(\underline{E}_R \underline{E}_R^T) = V - X(X^T V^{-1} X)^{-1} X^T$

เรียกค่า $X(X^T V^{-1} X)^{-1} X^T$ ว่าเป็นค่าปรับความเอนเอียง (Bias Correction)

โดยในการประมาณค่าในแต่ละรอบจะใช้

(1.27) $(Y - X\hat{\underline{\gamma}})(Y - X\hat{\underline{\gamma}})^T + X(X^T \hat{V}^{-1} X)^{-1} X^T$

ในการประมาณส่วนประกอบความแปรปรวน V ซึ่งพัฒนามาจากค่าประมาณของส่วนประกอบความแปรปรวนในรอบที่แล้ว $\hat{V}$

เนื่องจากเมตริกซ์ $(X^T \hat{V}^{-1} X)^{-1}$ สามารถเขียนได้ในรูป $U^{-1}(U^{-1})^T$ เมื่อ $(X^T \hat{V}^{-1} X) = U^T U$ และ U เป็นเมตริกซ์สามเหลี่ยมบน

กำหนดให้ $Z = XU^{-1}$ ดังนั้น (1.27) จะสามารถเขียนได้เป็น

(1.28) $\underline{E}_R \underline{E}_R^T + ZZ^T$

2.4 ตัวประมาณค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่ง (Robust Standard Error Estimator)

อาจเรียกว่า ตัวประมาณฮูเบอร์ไวท์ (Huber/White Estimator) โดยจากที่กล่าวไว้ในข้างต้น สมการ (1.16) จะสามารถประมาณพารามิเตอร์ของอิทธิพลคงที่ ได้โดยใช้ตัวประมาณกำลังสองน้อยสุดนัยทั่วไปดังนี้

(1.16) $\hat{\underline{\gamma}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \underline{Y}$

และมีเมตริกซ์ความแปรปรวนร่วมดังนี้

(1.17) $Cov(\hat{\underline{\gamma}}) = (X^T V^{-1} X)^{-1}$

โดยที่ $Cov(\underline{Y} \mid X\underline{\gamma}) = V$ เป็นพารามิเตอร์ที่ไม่ทราบค่า ดังนั้นในทางปฏิบัติจึงนิยมที่จะใช้ค่าประมาณซึ่งก็คือ $\hat{V}$ ในการหาค่าประมาณของค่าคลาดเคลื่อนมาตรฐาน แต่เมื่อการแจกแจงของค่าคลาดเคลื่อนสุ่มไม่ได้เป็นไปตามข้อสมมติเบื้องต้น ค่าประมาณของค่าคลาดเคลื่อนมาตรฐานที่ได้จาก (1.16) จะมีความผิดพลาด ซึ่งจะทำให้การประมาณค่าแบบช่วง และการทดสอบสมมติฐานผิดพลาด

ตัวประมาณค่าคลาดเคลื่อนที่มีความแกร่งจะปรับสูตรในการประมาณ $Cov(\hat{\underline{\gamma}})$ โดยใช้เมตริกซ์ทแยงมุม (Block Diagonal) ของ $E^* = \underline{E} \cdot \underline{E}^T$ ซึ่งเป็นตัวประมาณที่คงเส้นคงวาของ V

แทน $Cov(\hat{\gamma})$ และสำหรับในกรณีของพารามิเตอร์ของอิทธิพลสุ่ม หรือส่วนประกอบความแปรปรวน ก็สามารถกระทำได้ในทำนองเดียวกันกับข้างต้น

นอกจากนี้ค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งยังสามารถใช้ในการตรวจสอบความถูกต้องของตัวแบบได้อีกด้วย การที่ค่าคลาดเคลื่อนมาตรฐานที่ได้จากตัวแบบ มีความแตกต่างจากค่าคลาดเคลื่อนมาตรฐานที่มีความแกร่งมาก อาจเป็นการบ่งบอกว่าผู้วิจัยเลือกใช้ตัวแบบผิดพลาด (Goldstein; 1995, Cheong, Fotiu and Raudenbush; 2001, Raudenbush and Bryk; 2002)

2.4 ค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม (Intraclass Correlation: ICC)

ค่า ICC เป็นปัจจัยหนึ่งที่ส่งผลกระทบต่อการประมาณค่าพารามิเตอร์และค่าคลาดเคลื่อนมาตรฐานในตัวแบบเชิงลำดับชั้น โดยในตัวแบบเชิงลำดับชั้นที่มี 2 ระดับเราสามารถประมาณค่า ICC ได้จากตัวแบบว่าง (Null Model) ดังนี้

(1.18) $$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

จากตัวแบบ (1.18) เราจะสามารถประมาณค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มได้จากสูตร

(1.19) $$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

จะเห็นว่าจาก (1.19) สูตรการหาค่า ICC ในตัวแบบเชิงลำดับชั้นที่มี 2 ระดับสามารถหาได้จากสัดส่วนของความผันแปรระหว่างกลุ่ม (Between Groups Variation) ต่อความแปรผันรวม (Total Variation) ซึ่งค่า ICC เป็นเครื่องมือที่ใช้ในการชี้วัดถึงระดับความสัมพันธ์กันเองของค่าสังเกตภายในกลุ่มเดียวกัน

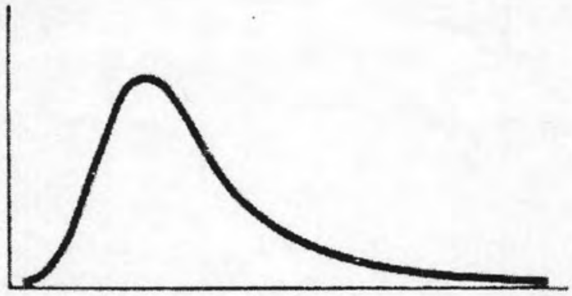
2.5 ความเบ้และความโด่ง (Skewness and Kurtosis)

2.5.1 ความเบ้ (Skewness)

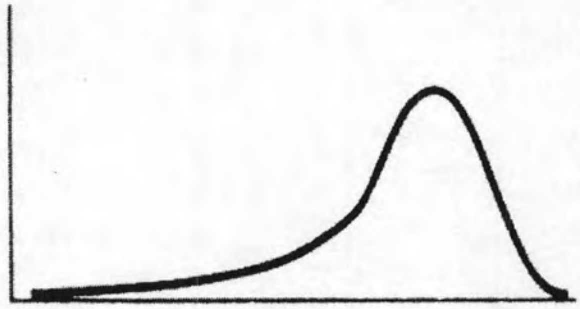
ประชากรที่มีการแจกแจงสมมาตรนั้น ตัวแปรสุ่ม X จะมีการแจกแจงสมมาตรรอบจุดใดจุดหนึ่ง ในที่นี้จะขอกำหนดให้ชื่อจุด θ กล่าวคือตัวแปรสุ่ม X จะสมมาตรรอบจุด θ ถ้า

$$P(X \geq \theta - x) = P(X \leq \theta + x) \qquad \text{สำหรับทุกค่า } x$$

แต่ถ้าประชากรมีการแจกแจงแบบไม่สมมาตร จะมีลักษณะของกราฟเบ้ไปข้างใดข้างหนึ่งดังรูป



(ก)



(ข)

ภาพประกอบที่ 2.1 แสดงตัวอย่างเส้นโค้งที่มีลักษณะเบ้ขวา และเบ้ซ้าย

จะเห็นว่าในรูปที่ 2.1 (ก) ประชากรมีการแจกแจงแบบเบ้ไปทางขวาเพราะพื้นที่ใต้โค้งทางด้านขวาของฐานนิยมมากกว่าพื้นที่ใต้โค้งทางด้านซ้าย และในรูปที่ 2.1 (ข) ประชากรมีการแจกแจงแบบเบ้ไปทางซ้ายเพราะพื้นที่ใต้เส้นโค้งทางด้านซ้ายของฐานนิยมมากกว่าพื้นที่ใต้เส้นโค้งด้านขวา

การวัดความเบ้จะใช้การวัดโดยวิธีโมเมนต์ สูตรสำหรับหาค่าสัดสมมาตร หรือสัมประสิทธิ์ความเบ้คือ $\gamma_3 = \dfrac{E((x-\mu)^3)}{(Var(x))^{3/2}}$

การวัดความเบ้ด้วยโมเมนต์ศูนย์กลางอันดับที่ 3 จะให้ผลต่างๆ ดังนี้

1. ถ้าการแจกแจงสมมาตร ค่าสัมประสิทธิ์ความเบ้จะเท่ากับศูนย์
2. ถ้าการแจกแจงเบ้ขวา ค่าสัมประสิทธิ์ความเบ้จะเป็นบวก
3. ถ้าการแจกแจงเบ้ซ้าย ค่าสัมประสิทธิ์ความเบ้จะเป็นลบ

2.5.2 **ความโด่ง** (Kurtosis)

ความโด่งของการแจกแจงมี 3 ลักษณะดังนี้

1. เส้นโค้งมีความโด่งเป็นปกติ (Mesokurtic)
2. เส้นโค้งที่แบนราบกว่าปกติ (Platykurtic)

3. เส้นโค้งที่โด่งกว่าปกติ (Leptokurtic)

การวัดความโด่ง หรือการหาสัมประสิทธิ์ความโด่ง มีสูตรดังนี้ $\gamma_4 = \dfrac{E((x-\mu)^4)}{(Var(x))^2}$

การวัดความโด่งด้วยโมเมนต์ศูนย์กลางอันดับที่ 4 จะให้ผลต่างๆดังนี้

1. ถ้าเส้นโค้งมีความโด่งเป็นปกติ ค่าสัมประสิทธิ์ความโด่งจะเท่ากับ 3
2. ถ้าเส้นโค้งมีความแบนราบกว่าปกติ ค่าสัมประสิทธิ์ความโด่งจะน้อยกว่า 3
3. ถ้าเส้นโค้งมีความโด่งกว่าปกติ ค่าสัมประสิทธิ์ความโด่งจะมากกว่า 3

2.6 การแจกแจงเอกซ์โพเนนเชียลพาวเวอร์แบบไม่สมมาตร (Asymmetric Exponential Power Distribution)

นิยาม กำหนดให้ $\alpha > 0$, $\theta \in \Re$, $\sigma > 0$ และ $\kappa > 0$

ตัวแปรสุ่ม X จะมีการแจกแจงแบบเอกซ์โพเนนเชียลพาวเวอร์แบบไม่สมมาตร (Asymmetric Exponential Power Distribution : EP) ก็ต่อเมื่อ

$$f(x) = \frac{\alpha}{\sigma \cdot \Gamma(\frac{1}{\alpha})} \frac{\kappa}{1+\kappa^2} \exp\left(-\frac{\kappa^\alpha}{\sigma^\alpha} [(x-\theta)^+]^\alpha - \frac{1}{\sigma^\alpha \cdot \kappa^\alpha} [(x-\theta)^-]^\alpha \right)$$

เมื่อ $u^+ = \begin{cases} u & ;u \ge 0 \\ 0 & ;u < 0 \end{cases}$ และ $u^- = \begin{cases} -u & ;u \ge 0 \\ 0 & ;u < 0 \end{cases}$

และเขียนแทนด้วย $X \sim EP(\theta, \sigma, \kappa, \alpha)$

โดยที่ θ คือ พารามิเตอร์ตำแหน่ง (location parameter)

σ คือ พารามิเตอร์สเกล (scale parameter)

κ คือ พารามิเตอร์ความเบ้ (skew parameter)

α คือ พารามิเตอร์ความโด่ง (shape parameter)

ฟังก์ชันการแจกแจงความน่าจะเป็น (distribution function) ของ $X \sim EP(\theta, \sigma, \kappa, \alpha)$ เป็นดังนี้

$$F(x) = P(X \le x) = \begin{cases} \dfrac{\kappa^2}{1+\kappa^2} \dfrac{1}{\Gamma\left(\frac{1}{\alpha}\right)} \Gamma\left(\dfrac{1}{\alpha}, [\dfrac{\theta - x}{\sigma\kappa}]^\alpha \right) & ;x < \theta \\ 1 - \dfrac{1}{1+\kappa^2} \dfrac{1}{\Gamma\left(\frac{1}{\alpha}\right)} \Gamma\left(\dfrac{1}{\alpha}, \dfrac{\kappa^\alpha}{\sigma^\alpha} (x-\theta)^\alpha \right) & ;x \ge \theta \end{cases}$$

เมื่อ $$\Gamma(a,x) = \int_x^{\infty} t^{a-1} \cdot \exp(-t)dt, \quad a > 0, x > 0$$

ในกรณีที่ $\kappa = 1$ การแจกแจงดังกล่าวจะเป็นการแจกแจงเอกโพเนเชียลพาวเวอร์ที่สมมาตรรอบ θ

ถ้า $\alpha = 1$ การแจกแจงดังกล่าวจะเป็นการแจกแจงลาปลาส (laplace distribution)

ถ้า $\alpha = 2$ การแจกแจงดังกล่าวจะเป็นการแจกแจงปกติ (normal distribution)

ในกรณีที่ $\kappa \neq 1$ การแจกแจงดังกล่าวจะเป็นการแจกแจงเอกโพเนเชียลพาวเวอร์แบบไม่สมมาตร

ถ้า $\alpha = 1$ การแจกแจงดังกล่าวจะเป็นการแจกแจงลาปลาสแบบไม่สมมาตร (skew laplace distribution)

ถ้า $\alpha = 2$ การแจกแจงดังกล่าวจะเป็นการแจกแจงปกติแบบไม่สมมาตร (skew normal distribution) (Ayebo and KoZubowski, 2003)

ทฤษฎีบทที่ 2.5.1

ถ้า $X \sim EP(\theta, \sigma, \kappa, \alpha)$ ดังนั้น

ค่าเฉลี่ย และความแปรปรวน จะมีสูตรดังนี้

$$\mu_x = E(X) = \theta + \sigma\left(\frac{1}{\kappa} - \kappa\right)\frac{\Gamma\left(\frac{2}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}$$

$$\sigma_x^2 = Var(X) = \sigma^2 \frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}\frac{1+\kappa^6}{\kappa^2(1+\kappa^2)} - \sigma^2 \frac{\Gamma^2\left(\frac{2}{\alpha}\right)}{\Gamma^2\left(\frac{1}{\alpha}\right)}\frac{(1-\kappa^2)^2}{\kappa^2}$$

โมเมนต์ที่ 3 รอบค่าเฉลี่ย หรือสัมประสิทธิ์ความเบ้ จะมีสูตรดังนี้

$$\gamma_3 = \frac{(1-\kappa^8)\Gamma^2\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{4}{\alpha}\right) - 3(1-\kappa^2)(1+\kappa^6)\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right) + 2(1-\kappa^2)^3(1+\kappa^2)\Gamma^3\left(\frac{2}{\alpha}\right)}{(1+\kappa^2)\left(\sqrt{\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)\frac{1+\kappa^6}{1+\kappa^2} - \Gamma^2\left(\frac{2}{\alpha}\right)(1-\kappa^2)^2}\right)^3}$$

และ โมเมนต์ที่ 4 รอบค่าเฉลี่ย หรือสัมประสิทธิ์ความโด่ง จะมีสูตรดังนี้

$$\gamma_4 = \frac{1+\kappa^{2}}{\left[\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)(1+\kappa^{6}) - \Gamma^{2}\left(\frac{2}{\alpha}\right)(1-\kappa^{2})(1+\kappa^{2})\right]^{2}} \times$$

$$\{\Gamma^{3}\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{5}{\alpha}\right)(1+\kappa^{10}) - 4\Gamma^{2}\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{2}{\alpha}\right)\Gamma\left(\frac{4}{\alpha}\right)(1-\kappa^{2})(1-\kappa^{8}) + 6\Gamma\left(\frac{1}{\alpha}\right)\Gamma^{2}\left(\frac{2}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)(1-\kappa^{2})(1+\kappa^{6})$$

$$-3\Gamma^{4}\left(\frac{2}{\alpha}\right)(1-\kappa^{2})(1+\kappa^{2})\}$$

ทฤษฎีบทที่ 2.5.2

ถ้า $X \sim EP(\theta,\sigma,\kappa,\alpha)$ จะได้ว่า

$$X = \theta + \sigma I W^{1/\alpha}$$

เมื่อ W เป็นตัวแปรสุ่มที่มีการแจกแจงแบบแกมมา ที่มีฟังก์ชันความหนาแน่น

$$f(w) = \frac{1}{\Gamma\left(\frac{1}{\alpha}\right)} w^{\frac{1}{\alpha}-1} \exp(-x), \qquad w > 0$$

และ $I = \begin{cases} -\kappa & ; P(I=-\kappa) = \dfrac{\kappa^2}{1+\kappa^2} \\ \dfrac{1}{\kappa} & ; P(I=\dfrac{1}{\kappa}) = \dfrac{1}{1+\kappa^2} \end{cases}$

ทฤษฎีบทที่ 2.5.3

ในการสร้างตัวแปรสุ่มที่มีการแจกแจงแบบเอกโพเนนเชียลพาวเวอร์แบบไม่สมมาตร สามารถทำได้ดังอัลกอริทึมต่อไปนี้

1. สร้างตัวแปรสุ่ม W ที่มีการแจกแจงแบบแกมมา ($W \sim Gamma(\frac{1}{\alpha},1)$)
2. สร้างตัวแปรสุ่ม U ที่มีการแจกแจงแบบยูนิฟอร์ม (0,1) ($Uniform(0,1)$)
3. ถ้า $U < \frac{1}{1+\kappa^2}$ จะให้ $I = \frac{1}{\kappa}$

   ถ้า $U \geq \frac{1}{1+\kappa^2}$ จะให้ $I = -\kappa$
4. สร้าง $X = \theta + \sigma I W^{1/\alpha}$