บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อทำการเปรียบเทียบความเร็วในการลู่เข้าของลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการอนุมานเชิงเบส์ เมื่อการแจกแจงภายหลังเป็นแบบปกติที่ถูกตัดหาง ระหว่างการสุ่มตัวอย่างแบบฮิตแอนด์รันกับการสุ่มตัวอย่างแบบกิบส์ ซึ่งข้อมูลที่นำมาใช้ในงานวิจัยนี้ ได้มาจากการจำลอง(Simulation)ด้วยโปรแกรม R ในบทนี้จะเป็นการนำเสนอวิธีดำเนินการวิจัยซึ่งผู้วิจัยได้ดำเนินการตามขั้นตอนต่อไปนี้

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับการวิจัย
2. จำลองข้อมูลด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รันให้เป็นไปตามขอบเขตที่กำหนด
3. จำลองข้อมูลด้วยการสุ่มตัวอย่างแบบกิบส์ให้เป็นไปตามขอบเขตที่กำหนด
4. คำนวณค่าครึ่งช่วงความเชื่อมั่นจากวิธีค่าเฉลี่ยกลุ่ม
5. คำนวณค่า MPSRF ของ บรูกซ์-เกลแมน
6. วิเคราะห์ และสรุปผลการวิจัย

โดยในขั้นตอนที่ 2-6 มีรายละเอียดของการดำเนินการดังนี้

3.1 การจำลองข้อมูลด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รัน

สำหรับขั้นตอนนี้จะทำการจำลองข้อมูลของพารามิเตอร์ซึ่งเป็นค่าคาดหวังของผลตอบแทน $\underset{\sim}{\alpha} = \left(\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(k)}\right)$ โดยกำหนดให้

$$\underset{\sim}{\alpha} \sim N_k\left(\underset{\sim}{0}, \Sigma\right)$$

k แทนจำนวนมิติของ $\underset{\sim}{\alpha}$

n แทนจำนวนรอบที่ทำการจำลองข้อมูล

ขั้นตอนมีดังนี้

1.) กำหนดจำนวนมิติ k และจำนวนรอบที่ทำการจำลอง n

2.) จำลองข้อมูลด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รัน

2.1) สุ่มเลือกจุดตัวอย่าง $\underset{\sim}{\alpha}_i = \underset{\sim}{\alpha}_1 = \left(\alpha_1^{(1)},\ldots,\alpha_1^{(k)}\right)$ ซึ่ง $\alpha_1^{(1)} \geq \alpha_1^{(2)} \geq \ldots \geq \alpha_1^{(k)}$

2.2) สุ่มเลือกทิศทาง $\underset{\sim}{d}$ บนพื้นผิวทรงกลมใน $\Re^k$

2.3) สุ่มเลือก λ จากการแจกแจงแบบปกติซึ่งมีค่าเฉลี่ย $\underset{\sim}{d}^T \Sigma^{-1}\left(-\underset{\sim}{\alpha}_i\right) / \underset{\sim}{d}^T \Sigma^{-1} \underset{\sim}{d}$ และส่วนเบี่ยงเบนมาตรฐาน $\sqrt{1/\underset{\sim}{d}^T \Sigma^{-1} \underset{\sim}{d}}$ ซึ่งถูกจำกัดอยู่ในบริเวณ $[\inf val, \sup val]$ โดยที่

$$\inf val = \max\left\{\lambda_{bound} \ (\lambda_{bound} < 0), -\infty\right\}$$

$$\sup val = \min\left\{\lambda_{bound} \ (\lambda_{bound} > 0), \infty\right\}$$

เมื่อ

$$\lambda_{bound} = \alpha_i^{(1)} - \alpha_i^{(2)} / d^{(2)} - d^{(1)}, \ldots, \alpha_i^{(k-1)} - \alpha_i^{(k)} / d^{(k)} - d^{(k-1)}$$

2.4) สุ่มเลือกจุด $\underset{\sim}{\alpha}_{i+1} = \underset{\sim}{\alpha}_i + \lambda \underset{\sim}{d}$

2.5) เพิ่ม i เป็น $i+1$ และกลับไป 2.2) จนกระทั่งจำนวนจุดตัวอย่างเท่ากับ n

3.2 การจำลองข้อมูลด้วยการสุ่มตัวอย่างแบบกิบส์

สำหรับขั้นตอนนี้จะทำการจำลองข้อมูลของพารามิเตอร์ซึ่งเป็นค่าคาดหวังของผลตอบแทน $\underset{\sim}{\alpha} = \left(\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(k)}\right)$ โดยกำหนดให้

$\underset{\sim}{\alpha} \sim N_k\left(\underset{\sim}{0}, \Sigma\right)$

k แทนจำนวนมิติของ $\underset{\sim}{\alpha}$

n แทนจำนวนรอบที่ทำการจำลองข้อมูล

ขั้นตอนมีดังนี้

1.) กำหนดจำนวนมิติ k และจำนวนรอบที่ทำการจำลอง n

2.) จำลองข้อมูลด้วยการสุ่มตัวอย่างแบบกิบส์

2.1) สุ่มจุดตัวอย่าง $\underset{\sim}{\alpha}_i = \underset{\sim}{\alpha}_1 = \left(\alpha_1^{(1)},\ldots,\alpha_1^{(j)},\ldots,\alpha_1^{(k)}\right)$ ซึ่ง $\alpha_1^{(1)} \geq \ldots \geq \alpha_1^{(j)} \geq \ldots \geq \alpha_1^{(k)}$

2.2) สุ่มเลือก j แจกแจงแบบสม่ำเสมอ บน $\{1,\ldots, k\}$

2.3) สุ่ม $\alpha_{i+1}^{(j)}$ จากการแจกแจงแบบปกติซึ่งมีค่าเฉลี่ย $-\left(\sum_{\substack{q=1 \\ q \neq p}}^{k} \alpha_i^{(q)} \rho_{p,q}\right) / \rho_{p,p}$ และส่วนเบี่ยงเบนมาตรฐาน $\sqrt{1/\rho_{p,p}}$ ซึ่งถูกจำกัดอยู่ในบริเวณ $[\inf\ val, \sup\ val]$ โดยที่ $\rho_{p,q}$ เป็นสมาชิกแถวที่ p หลักที่ q ใน Σ^{-1}

$$\inf\ val = \begin{cases} -\infty, & p = k \\ \alpha_i^{(p+1)}, & p \neq k \end{cases}$$

$$\sup\ val = \begin{cases} \infty, & p = 1 \\ \alpha_i^{(p-1)}, & p \neq 1 \end{cases}$$

2.4) ให้ $\underset{\sim}{\alpha}_{i+1} = \left(\alpha_i^{(1)},\ldots,\alpha_{i+1}^{(j)},\ldots,\alpha_i^{(k)}\right)$

2.5) เพิ่ม i เป็น $i+1$ และกลับไป 2.2) จนกระทั่งจำนวนจุดตัวอย่างเท่ากับ n

3.3 การคำนวณค่าครึ่งช่วงความเชื่อมั่นจากวิธีค่าเฉลี่ยกลุ่ม

หลังจากทำการจำลองข้อมูลด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รันและการสุ่มตัวอย่างแบบกิบส์ตามขั้นตอน 3.1 และ 3.2 แล้ว จะได้จุดตัวอย่าง $\left\{\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2,\ldots,\underset{\sim}{\alpha}_n\right\}$ จำนวน n จุด จากนั้นทำตามขั้นตอนดังนี้

1.) ทำการตัดข้อมูล (burn in) 10 เปอร์เซนต์แรกทิ้งไป และนำ 90 เปอร์เซนต์ที่เหลือมาแบ่งเป็น 30 กลุ่ม กลุ่มละเท่าๆ กัน

2.) หาค่าเฉลี่ยของตัวอย่างในแต่ละกลุ่ม และนำค่าเฉลี่ยของตัวอย่างในแต่ละกลุ่มมาหาค่าเฉลี่ยอีกครั้งหนึ่ง

3.) คำนวณค่าประมาณของ Monte Carlo standard error ซึ่งเท่ากับ $\frac{\hat{\sigma}_g}{\sqrt{(0.9)n}}$

4.) คำนวณค่าครึ่งช่วงความเชื่อมั่น ซึ่งเท่ากับ $t_{0.975,(29)} \dfrac{\hat{\sigma}_g}{\sqrt{(0.9)n}}$

3.4 การคำนวณค่า MPSRF ของ บรูกซ์-เกลแมน

หลังจากการจำลองข้อมูลด้วยการสุ่มตัวอย่างแบบฮิตแอนด์รันและการสุ่มตัวอย่างแบบกิบส์ตามขั้นตอน 3.1 และ 3.2 โดยจำลองข้อมูลเป็นลูกโซ่ที่อิสระกัน 5 ลูกโซ่ แล้ว จะได้จุดตัวอย่างของแต่ละลูกโซ่ $\left\{\underset{\sim}{\alpha}_{1j}, \underset{\sim}{\alpha}_{2j}, \ldots, \underset{\sim}{\alpha}_{nj}\right\} : j = 1,\ldots,5$ จำนวนลูกโซ่ละ n จุด จากนั้นทำตามขั้นตอนต่อไปนี้

1.) ทำการตัดข้อมูล (burn in) ครึ่งแรกของทุกๆ ลูกโซ่ทิ้งไป

2.) วิเคราะห์การลู่เข้าด้วย package coda ในโปรแกรม R

3.) คำนวณค่า MPSRF โดยใช้ฟังก์ชัน gelman.diag() ใน package coda

3.5 การวิเคราะห์และสรุปผลการวิจัย

ขั้นตอนมีดังนี้

1.) นำค่าครึ่งช่วงความเชื่อมั่นจากการสุ่มตัวอย่างแบบฮิตแอนด์รัน และค่าครึ่งช่วงความเชื่อมั่นจากการสุ่มตัวอย่างแบบกิบส์ ในกรณีที่จำนวนมิติเท่ากัน มาเปรียบเทียบกัน โดยใช้ค่าครึ่งช่วงความเชื่อมั่นของมิติที่ 1 ($\alpha^{(1)}$) มาเป็นตัวแทนในการเปรียบเทียบ ถ้าค่าครึ่งช่วงความเชื่อมั่นมีค่าน้อยกว่าจะมีประสิทธิภาพมากกว่า

2.) นำค่า MPSRF จากการสุ่มตัวอย่างแบบฮิตแอนด์รัน และค่า MPSRF จากการสุ่มตัวอย่างแบบกิบส์ ในกรณีที่จำนวนมิติเท่ากัน มาเปรียบเทียบกัน โดยถ้าค่า MPSRF มีค่าน้อยกว่าจะมีประสิทธิภาพมากกว่า

3.) วิเคราะห์ผลการเปรียบเทียบประสิทธิภาพของการสุ่มตัวอย่างแบบฮิตแอนด์รันกับการสุ่มตัวอย่างแบบกิบส์ ในกรณีที่มีมิติต่างๆ ที่กำหนด โดยใช้เกณฑ์ในการเปรียบเทียบประสิทธิภาพทั้งสองเกณฑ์ และทำการสรุปผล