



การศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์
การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงเส้นแบบพหุ
โดยใช้ขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟกึ่งวาระโตดย้อนกลับได้

โดย

นางสาวชลนรรจ์ สุขท้วญาติ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

สาขาวิชาคณิตศาสตร์และสถิติ

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2561

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

การศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์
การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงเส้นแบบพหุ
โดยใช้ขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟกึ่งวาระโดดย้อนกลับได้

โดย

นางสาวชลนรรจ์ สุขท้วญาติ



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

สาขาวิชาคณิตศาสตร์และสถิติ

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2561

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

A STUDY OF THE PERFORMANCE FOR VARIABLE AND MODEL
SELECTION OF ORDINAL REGRESSION ANALYSIS UNDER
MULTICOLLINEARITY USING REVERSIBLE JUMP
MARKOV CHAIN MONTE CARLO ALGORITHM

BY

MISS CHONLANAN SUKTHUAYAT



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF MATHEMATICS AND STATISTICS
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2018
COPYRIGHT OF THAMMASAT UNIVERSITY

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นางสาวชลนรรจ์ สุขทัญญาติ

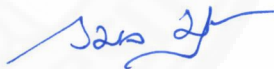
เรื่อง

การศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับเมื่อตัวแปร
มีความสัมพันธ์เชิงเส้นแบบพหุโดยใช้ขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟพ้าวาระโตดย้อนกลับได้

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

เมื่อ วันที่ 11 ธันวาคม พ.ศ. 2561

ประธานกรรมการสอบวิทยานิพนธ์



(รองศาสตราจารย์ ดร.กมล บุขบา)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์



(ผู้ช่วยศาสตราจารย์ ดร.แสงหล้า ชัยมงคล)

กรรมการสอบวิทยานิพนธ์



(รองศาสตราจารย์ ดร.วราฤทธิ์ พานิชกิจโกศลกุล)

กรรมการสอบวิทยานิพนธ์



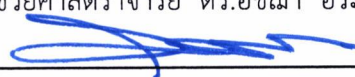
(อาจารย์ ดร.มณฑิรา ดวงสาพล)

กรรมการสอบวิทยานิพนธ์



(ผู้ช่วยศาสตราจารย์ ดร.อัชฌา อระวีพร)

คณบดี



(รองศาสตราจารย์ ดร.สมชาย ชคตระการ)

หัวข้อวิทยานิพนธ์	การศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงเส้นแบบพหุโดยใช้ขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้
ชื่อผู้เขียน	นางสาวชลนรรจ์ สุขทัญญาติ
ชื่อปริญญา	วิทยาศาสตร์มหาบัณฑิต (สถิติประยุกต์)
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร.แสงหล้า ชัยมงคล
ปีการศึกษา	2561

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงเส้นแบบพหุโดยใช้ขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้โดยศึกษา 3 ตัวแบบ คือ ตัวแบบบอดส์เชิงสัดส่วน (PO) ตัวแบบบอดส์ที่ไม่เป็นตามสัดส่วน (NPO) และตัวแบบบอดส์เชิงสัดส่วนบางส่วน (PPO) การจำลองข้อมูลภายใต้สถานการณ์ที่ศึกษาดังนี้ จำลองข้อมูล $n_{sim} = 100$ ชุด ในแต่ละชุดมีขนาดตัวอย่าง $n = 50, 100, 500$ และ $1,000$ กำหนดตัวแปรตอบสนอง 3 ระดับ ($Y = 1, 2, 3$) และมีจำนวนตัวแปรอธิบาย $k = 3, 5$ และ 10 ตัวแปร ที่มีการแจกแจงปกติหลายตัวแปรมีเวกเตอร์ค่าเฉลี่ยเป็น 0 และเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมเป็น Σ เมื่อกำหนดรูปแบบความสัมพันธ์ 4 ลักษณะ คือมีความสัมพันธ์แบบเป็นอิสระกัน, แบบคงที่, แบบ Toeplitz ให้ $r = 0, 0.5, 0.8$ และแบบ Hub Toeplitz ให้ $r_{min} = 0.3, r_{max} = 0.8$ และสร้างพารามิเตอร์ θ และ β จากการแจกแจงปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวน σ^2 คัดเลือกตัวแปรด้วยวิธี RJ-MCMC ใช้ลูกโซ่ 2 โซ่ จำนวนรอบซ้ำ 20,000 รอบ และตัดทิ้ง 10,000 รอบแรกทิ้ง (burn-in) เกณฑ์ที่ใช้ในการพิจารณาคือร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกผิดโดยเฉลี่ย ผลการศึกษาพบว่า

สมรรถนะของการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงอันดับด้วยขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้ขึ้นอยู่กับขนาดตัวอย่าง เมื่อขนาดตัวอย่างเพิ่มขึ้นจะทำให้สมรรถนะของการคัดเลือกตัวแปรอธิบายเพิ่มขึ้นตาม และขนาดตัวอย่างของตัวแบบ PO ควรมี

อย่างน้อย 10 เท่าของจำนวนตัวแปรอธิบาย แต่ตัวแบบ NPO ควรมีจำนวนตัวแปรสูงกว่า 20 เท่าของจำนวนตัวแปรอธิบาย

สมรรถนะของการคัดเลือกตัวแปรยังขึ้นอยู่กับระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย โดย $r = 0.5$ มีสมรรถนะของการคัดเลือกตัวแปรสูงกว่า $r = 0.8$ เมื่อตัวแปรอธิบายมีรูปแบบความสัมพันธ์เป็นแบบคงที่, แบบ Toeplitz แต่สมรรถนะการคัดเลือกตัวแปรไม่ขึ้นอยู่กับรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและจำนวนตัวแปรอธิบายยกเว้นในกรณีตัวแบบ NPO เมื่อจำนวนตัวแปรอธิบายเพิ่มขึ้นสมรรถนะการคัดเลือกตัวแปรจะลดลง

การคัดเลือกตัวแปรของตัวแบบ PO มีสมรรถนะสูงกว่าตัวแบบ NPO และวิธีมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้เป็นวิธีที่ไม่คัดเลือกตัวแปรที่ไม่มีความสัมพันธ์กับตัวแบบเข้ามาในตัวแบบ

คำสำคัญ: ตัวแบบออกัสเชิงสัดส่วน, ตัวแบบออกัสที่ไม่เป็นตามสัดส่วน, ตัวแบบออกัสเชิงสัดส่วนบางส่วน, มอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้

Thesis Title	A Study of The Performance for Variable and Model Selection of Ordinal Regression Analysis Under Multicollinearity Using Reversible Jump Markov Chain Monte Carlo Algorithm
Author	Miss Chonlanan Sukthuayat
Degree	Master of Science (Applied Statistics)
Department/Faculty/University	Mathematics and Statistics Faculty of Science and Technology Thammasat University
Thesis Advisor	Assistant Professor Saengla Chaimongkol, Ph.D.
Academic Year	2018

ABSTRACT

This research aims to study the performance for variable and model selection of ordinal regression analysis under multicollinearity using reversible jump Markov chain Monte Carlo algorithm. Three models used in this research are Proportional Odds model (PO), Non-Proportional Odds model (NPO) and Partial Proportional Odds model (PPO). We simulate the data under the following situation, $n_{sim} = 100$ data sets, each containing sample size (n) are 50, 100, 500 and 1,000, three sampled levels response variables and the number of explanatory variables are 3, 5 and 10 from multivariate normal distribution with an average vector to 0 and variance - covariance matrix to Σ . We specify four correlation patterns: independent, constant correlation structure, Toeplitz correlation structure with $r = 0, 0.5, 0.8$, and Hub Toeplitz correlation structure with $r_{min} = 0.3$ and $r_{max} = 0.8$. The threshold parameters θ , and regression parameters β were sampled from normal distribution with mean equal to μ and standard deviations equal to σ . For the RJ-MCMC routines, we used two chains were run 20,000 updates and the first 10,000 updates were discarded as burn-in. The criteria of this study are percentage of average variable and model selection and percentage of average misclassification. The results of this research can be summarized as follows:

The performance of variable and model selection for ordinal regression analysis using a reversible jump Markov chain Monte Carlo algorithm is corresponding to sample size. When the sample size increased, the performance of variable selection also increased. And, the sample size for the PO model should be at least 10 times of the number of explanatory variables and 20 times for the NPO model.

The performance of variable selection depends on the level of correlation between the explanatory variables. For the explanatory variables with constant correlation structure and Toeplitz correlation structure, the results show that $r = 0.5$ was higher performance than $r = 0.8$. Meanwhile, the performance of variable selection does not depend on the relationship between the explanatory variables and the number of explanatory variables, except for the NPO model. When the number of explanatory variables is increased, the performance of variable selection has decrease.

The performance of variable selection of PO models was higher than NPO models and this method does not select variables which are not related to the model.

Keywords: Proportional odds model (PO), Non-Proportional odds model (NPO), Partial proportional odds model (PPO), Reversible Jump Markov Chain Monte Carlo (RJ-MCMC)

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดีด้วยความช่วยเหลือจาก ผู้ช่วยศาสตราจารย์ ดร.แสงหล้า ชัยมงคล อาจารย์ที่ปรึกษาวิทยานิพนธ์ซึ่งกรุณาสละเวลาให้คำปรึกษาในทุกด้านและมอบความรู้พร้อมทั้งติดตามความคืบหน้าในทุกขั้นตอนการทำวิทยานิพนธ์และตรวจสอบข้อบกพร่องและให้คำแนะนำเพื่อใช้ในการปรับปรุงแก้ไข ซึ่งเป็นประโยชน์ต่อการจัดทำวิทยานิพนธ์ตลอดระยะเวลาการจัดทำวิทยานิพนธ์จนสำเร็จ

ขอขอบพระคุณท่านคณะกรรมการสอบวิทยานิพนธ์ อันได้แก่ รองศาสตราจารย์ ดร.กมล บุชบา รองศาสตราจารย์ ดร.วราฤทธิ์ พานิชกิจโกศลกุล อาจารย์ ดร.มณฑิรา ดวงสาพล ผู้ช่วยศาสตราจารย์ ดร.อชฌา อระวีพร ที่ให้แนวคิดและความรู้เพิ่มเติมอันเป็นประโยชน์ต่องานวิจัย พร้อมทั้งตรวจสอบข้อผิดพลาดให้มีความสมบูรณ์เพิ่มมากขึ้น ซึ่งเป็นประโยชน์ต่อการทำวิจัยเป็นอย่างมาก

สุดท้ายนี้ขอขอบพระคุณ คณาจารย์สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ที่เป็นผู้เริ่มให้ความรู้ทางสถิติแก่ผู้วิจัยเพื่อมาประยุกต์ใช้ในวิทยานิพนธ์ รวมทั้งเจ้าหน้าที่สาขาวิชาคณิตศาสตร์และสถิติที่อำนวยความสะดวกในการจัดทำวิทยานิพนธ์ครั้งนี้เป็นอย่างดีและขอขอบคุณสมาชิกในครอบครัวทุกท่านที่ให้อกำลังใจผู้ทำวิจัยตลอดมาจนสำเร็จการศึกษา

นางสาวชลนรรจ์ สุขท้วญาติ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(3)
กิตติกรรมประกาศ	(5)
สารบัญตาราง	(9)
สารบัญภาพ	(10)
รายการสัญลักษณ์และคำย่อ	(11)
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการศึกษา	4
1.3 ขอบเขตของการศึกษา	4
1.4 เกณฑ์ที่ใช้ในการพิจารณา	5
1.5 ประโยชน์ที่คาดว่าจะได้รับ	6
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	7
2.1 ทฤษฎีที่เกี่ยวข้อง	7
2.1.1 ตัวแบบเชิงเส้นน้อยทั่วไป	7
2.1.1.1 ส่วนประกอบเชิงสุ่ม	7
2.1.1.2 ส่วนประกอบเชิงระบบ	8
2.1.1.3 ส่วนประกอบฟังก์ชันเชื่อมโยง	8

2.1.2 การอนุมานแบบเบสและวิธีมอนติคาร์โลโซมาร์คอฟ	10
2.1.2.1 ขั้นตอนวิธีการของเมโทรโพลิส-เฮสติ้ง	11
2.1.3 ตัวแบบลอจิสต์สำหรับตัวแปรตอบสนองแบบอันดับ	12
2.1.3.1 ตัวแบบลอจิสต์อดส์เชิงสัดส่วน	13
2.1.3.2 ตัวแบบลอจิสต์อดส์ที่ไม่เป็นตามสัดส่วน	15
2.1.3.3 ตัวแบบลอจิสต์อดส์บางส่วน	16
2.1.4 วิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้ในการคัดเลือกตัวแบบและคัดเลือกตัวแปร	16
2.1.4.1 การคัดเลือกตัวแบบแบบเบสด้วยวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้	16
2.1.4.2 การคัดเลือกตัวแปรในตัวแบบการถดถอยลอจิสต์เชิงอันดับโดยใช้วิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้	18
(1) การคัดตัวแปรเข้าหรือออกจากตัวแบบ	18
(2) การปรับปรุงตัวแบบกรณีที่มีการคัดเลือกตัวแปรเข้ามาเพิ่มในตัวแบบ	19
(3) การปรับ	21
2.1.5 รูปแบบเมทริกซ์ความสัมพันธ์ของตัวแปรอธิบาย	22
2.1.5.1 รูปแบบความสัมพันธ์แบบคงที่	22
2.1.5.2 รูปแบบความสัมพันธ์แบบ Toeplitz	22
2.1.5.3 รูปแบบความสัมพันธ์แบบ Hub Toeplitz	23
2.2 งานวิจัยที่เกี่ยวข้อง	25
บทที่ 3 วิธีการวิจัย	27
3.1 วิธีการมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้	27
3.2 การดำเนินการวิจัย	27
3.3 แผนการจำลองข้อมูล	30

บทที่ 4 ผลการวิจัยและอภิปรายผล	31
4.1 ผลการศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบ กรณีตัวแบบ PO	31
4.2 ผลการศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบ กรณีตัวแบบ NPO	37
4.3 ผลการศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบ กรณีตัวแบบ PPO	44
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	54
5.1 สรุปผลการวิจัย	54
5.1.1 ผลการศึกษาสมรรถนะของวิธีมอนติคาร์โลโซ่มาร์คอฟกึ่งวาระโดด ย้อนกลับได้จากการจำลองตัวแบบ PO, NPO และ PPO	54
5.1.2 เปรียบเทียบสมรรถนะของวิธีมอนติคาร์โลโซ่มาร์คอฟกึ่งวาระโดด ย้อนกลับได้ระหว่างตัวแบบ PO, NPO และ PPO	55
5.2 ข้อเสนอแนะ	56
รายการอ้างอิง	57
ภาคผนวก	
ภาคผนวก ก	61
ประวัติผู้เขียน	70

สารบัญตาราง

ตารางที่	หน้า
4.1 ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50, 100, 500 และ 1,000	32
4.2 ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ NPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 100, 500 และ 1,000	38
4.3 ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50, 100, 500 และ 1,000	45

สารบัญภาพ

ภาพที่

หน้า

3.1 แผนผังขั้นตอนในการดำเนินการวิจัย

30



รายการสัญลักษณ์และคำย่อ

สัญลักษณ์/คำย่อ	คำเต็ม/คำจำกัดความ
X	เมทริกซ์ของตัวแปรอธิบาย
X_k	เวกเตอร์ของตัวแปรอธิบายของกลุ่มที่ k
X_{ik}	ตัวแปรอธิบายของหน่วยตัวอย่างที่ i และกลุ่มของตัวแปรอธิบายที่ k
Y	เวกเตอร์ของตัวแปรตอบสนองที่เป็นตัวแปรจำแนกประเภทแบบมีอันดับ
Y_j	ตัวแปรตอบสนองระดับที่ j โดย $j=1,2,3$
Y_i	ตัวแปรตอบสนองหน่วยตัวอย่างที่ i โดย $i=1,\dots,n$
J	จำนวนระดับของตัวแปรตอบสนอง
k	จำนวนตัวแปรอธิบาย
n	ขนาดตัวอย่าง
β	เวกเตอร์ของพารามิเตอร์สัมประสิทธิ์การถดถอย
β_k	พารามิเตอร์สัมประสิทธิ์การถดถอยกลุ่มตัวแปรอธิบายที่ k
β_{kj}	พารามิเตอร์สัมประสิทธิ์การถดถอยกลุ่มตัวแปรอธิบายที่ k และระดับของตัวแปรตอบสนองที่ j
PO	ตัวแบบออดส์เชิงสัดส่วน
NPO	ตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน
PPO	ตัวแบบออดส์เชิงสัดส่วนบางส่วน
PPA	ความน่าจะเป็นภายหลังของความเกี่ยวพัน
RJ-MCMC	วิธีมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้
BE	ปัจจัยเบส์ (Bayes factor)
P_j	ความน่าจะเป็นที่แต่ละหน่วยตัวอย่างจะอยู่ในระดับที่ j

p_{ij}	ความน่าจะเป็นของตัวแบบหน่วยตัวอย่างที่ i และระดับของตัวแปรตอบสนองที่ j
γ_{ij}	ความน่าจะเป็นสะสมของตัวแบบหน่วยตัวอย่างที่ i และระดับของตัวแปรตอบสนองที่ j
θ_j	พารามิเตอร์จุดตัดระดับของตัวแปรตอบสนองที่ j
$\Sigma_{k \times k}$	เมทริกซ์สหสัมพันธ์ขนาด $k \times k$
ρ_k	ความสัมพันธ์ขององค์ประกอบภายใน k กลุ่ม โดยมีค่าอยู่ระหว่าง $0 \leq \rho_k \leq 1$
$\rho_{\max}$	ความสัมพันธ์ขององค์ประกอบภายใน k กลุ่ม ที่มีค่ามากที่สุด
$\rho_{\min}$	ความสัมพันธ์ขององค์ประกอบภายใน k กลุ่ม ที่มีค่าน้อยที่สุด
r	ค่าสหสัมพันธ์บางส่วนของตัวแปรอธิบาย 2 ตัวใด ๆ

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ตัวแบบการถดถอยที่ใช้ศึกษาอิทธิพลของตัวแปรอธิบาย (Explanatory variable) หรือปัจจัยที่ส่งผลกระทบต่อตัวแปรตอบสนอง (Response variable) แบบจำแนกประเภท คือ ตัวแบบลอจิสต์ (Logit Model) ที่ประกอบด้วยตัวแปรตอบสนองแบบจำแนกประเภทตั้งแต่ 2 ระดับหรือมากกว่าทั้งแบบไม่มีอันดับ (Nominal responses) และแบบมีอันดับ (Ordinal responses) ส่วนตัวแปรอธิบายโดยทั่วไปเป็นแบบจำแนกประเภททั้งหมดหรือแบบต่อเนื่องทั้งหมดหรือเป็นแบบผสม รูปแบบของตัวแบบลอจิสต์นิยมแปลงจากรูปแบบไม่เชิงเส้นเป็นรูปแบบเชิงเส้น โดยใช้ฟังก์ชันเชื่อมโยงแบบลอจิสต์ กรณีพิเศษเมื่อตัวแบบลอจิสต์มีตัวแปรอธิบายต่อเนื่องอย่างน้อย 1 ตัวและตัวแปรตอบสนองจำแนกประเภทเป็นแบบไม่มีอันดับ 2 กลุ่ม (Binary) นิยมเรียกตัวแบบนี้ว่าตัวแบบการถดถอยลอจิสติก (Logistic regression model) โดยทั่วไปตัวแบบลอจิสต์หมายถึงตัวแบบลอจิสต์ที่ตัวแปรอธิบายเป็นแบบจำแนกประเภททั้งหมด ตัวแบบลอจิสต์เทียบกับกลุ่มฐาน (Baseline category logit) ตัวแบบลอจิสต์สะสม (Cumulative logit model) ซึ่งเป็นกลุ่มของตัวแบบเชิงเส้นน้อยทั่วไป (Generalized linear models) ที่นิยมใช้ในปัจจุบัน โดยหลายตัวแบบอาจมีชื่อเรียกตามเงื่อนไขเพิ่มเติม เช่น ตัวแบบออดส์เชิงสัดส่วน (Proportional Odds model) ตัวแบบออดส์เชิงไม่สัดส่วน (Non-Proportional Odds model) ตัวแบบออดส์เชิงสัดส่วนบางส่วน (Partial Proportional Odds model)

ตัวแบบการถดถอยลอจิสติกแบบอันดับ (Ordinal logistic regression model) หรือตัวแบบการถดถอยแบบอันดับ (Ordinal regression model) เป็นตัวแบบการถดถอยที่มีตัวแปรตอบสนองที่เป็นตัวแปรจำแนกประเภทแบบอันดับเมื่อมีตัวแปรอธิบายอย่างน้อย 1 ตัว และเป็นกรณีพิเศษของตัวแบบเชิงเส้นน้อยทั่วไปโดยตัวแบบที่นิยมใช้คือ ตัวแบบออดส์เชิงสัดส่วน

ตัวแบบออดส์เชิงสัดส่วน (Proportional Odds model: PO) เป็นตัวแบบลอจิสต์ที่นิยมใช้ในกรณีที่ตัวแปรตอบสนองเป็นแบบอันดับที่มีระดับตั้งแต่ 2 ระดับขึ้นไป ซึ่งแต่เดิมถูกเรียกว่าตัวแบบลอจิสต์สะสม (Cumulative logit model) ถูกเสนอโดย Walker และ Duncan ในปี 1967 ต่อมาได้ถูกนำมาปรับใช้และเรียกว่าตัวแบบออดส์เชิงสัดส่วน (Proportional Odds model) โดย McCullagh (1980) ตัวแบบนี้อยู่ภายใต้ข้อตกลงเบื้องต้นที่ว่าเวกเตอร์ความชัน (β) ในตัวแบบที่ส่งผลต่อแต่ละระดับหรือกลุ่มของตัวแปรตอบสนอง (Y_j) ; $j = 1, \dots, J-1$ เท่ากัน เมื่อ J คือ

จำนวนระดับของตัวแปรตอบสนอง กล่าวคือความสัมพันธ์ของตัวแปรอธิบาย (X) และตัวแปรตอบสนอง (Y) จะไม่เปลี่ยนแปลงเมื่อกลุ่มหรือระดับของ Y เปลี่ยนทำให้ β ไม่ขึ้นกับ j หรือจุดที่แบ่ง Y ออกเป็นกลุ่ม (McCullagh, 1980) โดยเรียกข้อสมมติของความเท่ากันของ log odds ในทุก ๆ จุดตัดว่า proportional odds assumption ภายใต้ข้อจำกัดการเรียงลำดับสุ่มของตัวแปรตอบสนอง ดังนั้นการสร้างตัวแบบจะต้องแน่ใจว่าความน่าจะเป็นสะสมของกลุ่มที่ j จะน้อยกว่าหรือเท่ากับความน่าจะเป็นสะสมของกลุ่มที่ $j+1$; $j=1,2,...,J-1$

ตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน (Non-Proportional odds model: NPO) ใช้ฟังก์ชันของการแจกแจงสะสมลอจิต (cumulative logit) ในการสร้างตัวแบบโดยที่ให้ความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองเปลี่ยนไปตามกลุ่มของตัวแปรตอบสนอง นั่นคือสัมประสิทธิ์การถดถอยจะแตกต่างกันในแต่ละกลุ่มของตัวแปรตอบสนองหรือ β ในแต่ละกลุ่มของตัวแปรตอบสนองไม่เท่ากัน (Erkan Ari และ Zeki Yildiz, 2014) ซึ่งตัวแบบ NPO จะมีความยืดหยุ่นกว่าตัวแบบ PO แต่การใช้ตัวแบบ NPO ยังไม่แพร่หลายมากนักเนื่องจากเงื่อนไขลำดับสุ่มอาจเป็นจริงเฉพาะช่วงใดช่วงหนึ่งของตัวแปรอธิบาย (Agresti, 2010) เป็นผลทำให้วิธีการประมาณค่าพารามิเตอร์ของตัวแบบที่ใช้กันโดยทั่วไปไม่สามารถทำได้และในกรณีที่ข้อสมมติของตัวแบบ PO ไม่เป็นจริง การประมาณค่าพารามิเตอร์ของตัวแบบ NPO อาจทำได้ยาก ดังนั้นทางเลือกของการสร้างตัวแบบทำได้โดยการสร้างตัวแบบการถดถอยลอจิสติกทวิภาค จำนวน $J-1$ สมการแยกกัน แต่การทำแบบนี้อาจทำให้ได้ค่าประมาณพารามิเตอร์ของตัวแบบที่แตกต่างไปจากการใช้ตัวแบบ NPO

ตัวแบบออดส์เชิงสัดส่วนบางส่วน (Partial Proportional Odds model; PPO) เป็นตัวแบบผสมของ PO และ NPO (Beredis Peterson et al., 1990) ภายใต้การทดสอบ proportional odds assumption ของตัวแบบ PPO ตัวแปรอธิบายบางตัวมีความสัมพันธ์กับตัวแปรตอบสนองแบบ PO ในขณะที่ตัวแปรที่เหลือจะมีความสัมพันธ์แบบ NPO และยังทำการทดสอบสมมติฐานว่างเมื่อกำหนดให้สัมประสิทธิ์การถดถอยในแต่ละกลุ่มของตัวแปรตอบสนองที่ไม่เท่ากันทั้ง $J-1$ มีค่าเท่ากับ 0 ทุก ๆ ค่าของ β ในแต่ละกลุ่มของตัวแปรตอบสนอง ถ้าสมมติฐานเป็นจริงแล้วจึงสามารถลดตัวแบบให้กลายเป็นตัวแบบ PO

การคัดเลือกตัวแบบที่เหมาะสมตามวิธีการทางสถิติแบบดั้งเดิม (Classical statistical framework) จะใช้เกณฑ์ในการเปรียบเทียบตัวแบบ เช่น คัดเลือกตัวแบบด้วย AIC (Akaike's Information Criterion) หรือใช้การทดสอบอัตราส่วนภาวะน่าจะเป็น ซึ่งเป็นวิธีการที่ใช้ข้อมูลจากค่าประมาณพารามิเตอร์เพียงจุดเดียวและไม่สนใจความคลาดเคลื่อน นอกจากนั้นในการอนุมานจะทำตามเงื่อนไขในตัวแบบที่เลือกและไม่คำนึงถึงตัวแบบอื่น ๆ ที่เป็นไปได้ ดังนั้นวิธีการนี้อาจทำให้ตัวแปรอธิบายที่มีความสัมพันธ์กับตัวแปรตอบสนองในหลาย ๆ ตัวแปร ไม่ถูกเลือกเข้ามาในตัวแบบสุดท้าย ดังนั้น McKinley, T. J. และคณะ (2015) ได้เสนอแนวทางของเบสในการคัดเลือกตัวแบบโดยการใช้

ความน่าจะเป็นภายหลังของความเกี่ยวพัน (Posterior Probability of Association: PPA) ซึ่งหลักการของวิธีนี้เป็น การหาระดับความสัมพันธ์ของตัวแปรอธิบายที่มีต่อตัวแปรตอบสนองของตัวแบบหนึ่ง ๆ โดยเฉลี่ยจากตัวแบบทั้งหมดที่เป็นไปได้ แต่ตัวแบบ PO และ NPO มีจำนวนพารามิเตอร์ที่แตกต่างกัน คือตัวแบบ PO มีพารามิเตอร์ความชันแค่ 1 ตัว ในขณะที่ NPO มีจำนวน $j-1$ ตัว McKinley, T. J. และคณะจึงใช้วิธีมอนติคาร์โลโซมาร์คอฟก้าวกะโดดย้อนกลับได้ (Reversible-Jump Markov Chain Monte Carlo: RJ-MCMC) ในการสร้างตัวแบบลอจิตโดยที่ไม่จำเป็นต้องรู้รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองล่วงหน้าว่ามีลักษณะแบบ PO หรือ NPO

การคัดเลือกตัวแบบตามแนวทางของเบส์จะอยู่ในรูปของการประมาณค่าพารามิเตอร์มากกว่าจะเป็นการค้นหาตัวแบบที่เหมาะสมเพียงตัวแบบเดียว โดยแนวทางของเบส์จะพยายามประมาณความน่าจะเป็นภายหลังของทุก ๆ ตัวแบบที่เป็นไปได้ ซึ่งในหลายสถานการณ์จะประมาณความน่าจะเป็นภายหลังตามขอบ (marginal posterior probability) ของตัวแปรที่ควรจะมีอยู่ในตัวแบบ ดังนั้นวิธีการ RJ-MCMC ที่ McKinley, T. J. และคณะได้เสนอไว้จึงสามารถใช้ในการคัดเลือกตัวแปรของตัวแบบการถดถอยแบบอันดับได้อีกด้วย โดยทั่วไปการคัดเลือกตัวแปรในตัวแบบการถดถอยอาจแบ่งออกเป็น 2 กรณีคือ กรณีแรกการคัดเลือกตัวแปรตามความรู้หรือความคาดหวังของผู้วิจัยในเรื่องนั้น ๆ กรณีนี้ควรจะสร้างตัวแบบการถดถอยแบบเต็มที่มีตัวแปรอธิบายทุกตัวในตัวแบบแล้วอธิบายค่าประมาณการแจกแจงภายหลังในเทอมของสำคัญของตัวแปรนั้น ๆ ตรงข้ามกับกรณีที่สอง การสร้างตัวแบบการถดถอยที่มีวัตถุประสงค์เพื่ออธิบายตัวแปรตอบสนองด้วยชุดของตัวแปรอธิบายจำนวนหนึ่งจากจำนวนตัวแปรอธิบายที่มีขนาดใหญ่ แต่การคัดตัวแปรเข้าหรือตัดตัวแปรที่อยู่ในตัวแบบการถดถอยออกอาจขึ้นอยู่กับรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย เมื่อมีจำนวนตัวแปรอธิบายมากกว่า 1 ตัวแปร ตัวแปรอธิบายเหล่านั้นอาจมีความสัมพันธ์กันสูงระหว่างตัวแปรอธิบายแต่ละตัวซึ่งจะก่อให้เกิดปัญหาความสัมพันธ์เชิงเส้นแบบพหุ (Multicollinearity) และอาจส่งผลต่อการคัดเลือกตัวแปรในแต่ละรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายได้ ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงสนใจศึกษาสมรรถนะของวิธีการ RJ-MCMC ในการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับโดยศึกษาเพิ่มเติมจาก McKinley, T. J. และคณะ (2015) กำหนดตัวแปรอธิบายให้มีรูปแบบความสัมพันธ์แบบอิสระมีระดับสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0 ส่วนในรูปแบบความสัมพันธ์แบบคงที่และแบบ Toeplitz กำหนดระดับสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5 และ 0.8 และในรูปแบบความสัมพันธ์แบบ Hub Toeplitz มีระดับสัมประสิทธิ์สหสัมพันธ์ $r_{\min} = 0.3$ และ $r_{\max} = 0.8$

1.2 วัตถุประสงค์ของการศึกษา

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาสมรรถนะของขั้นตอนวิธีมอนติคาร์โลไฮมาร์คอฟก๊าวกระโดดย้อนกลับได้ในการคัดเลือกตัวแปรและตัวแบบการถดถอยเชิงอันดับเมื่อตัวแปรอธิบายมีความสัมพันธ์เชิงเส้นพหุ

1.3 ขอบเขตของการศึกษา

ขอบเขตการศึกษาเป็นการจำลองสถานการณ์ดังต่อไปนี้

1. ขนาดตัวอย่าง $n = 50, 100, 500, 1000$
2. จำนวนตัวแปรอธิบาย $k = 3, 5, 10$
3. สร้างตัวแปรอธิบายเป็นตัวแปรสุ่มแบบต่อเนื่องจากรูปแบบความสัมพันธ์ 4 รูปแบบ

คือ

3.1 รูปแบบความสัมพันธ์แบบอิสระต่อกัน โดยกำหนดระดับสัมประสิทธิ์สหสัมพันธ์ $r = 0$

3.2 รูปแบบความสัมพันธ์แบบคงที่ (Constant) กำหนดระดับสัมประสิทธิ์สหสัมพันธ์ $r = 0.5, 0.8$

3.3 รูปแบบความสัมพันธ์แบบ Toeplitz กำหนดระดับสัมประสิทธิ์สหสัมพันธ์ $r = 0.5, 0.8$

3.4 รูปแบบความสัมพันธ์แบบ Hub Toeplitz กำหนดระดับสัมประสิทธิ์สหสัมพันธ์ $r_{\min} = 0.3$ และ $r_{\max} = 0.8$

4. รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแต่ละตัวกับตัวแปรตอบสนองแบบอันดับ 3 รูปแบบ คือ ตัวแบบ PO, NPO และไม่มีความสัมพันธ์ ดังนั้นตัวแบบที่เป็นไปได้ทั้งหมดมีจำนวน 3^k (k เท่ากับจำนวนตัวแปรอธิบาย)

4.1 กรณีที่เป็นตัวแบบ PO สุ่มตัวแปรอธิบาย 1 ตัวที่ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง Y นั่นคือ ($\beta = 0$) แล้วกำหนดตัวแปรอธิบายที่เหลือให้มีความสัมพันธ์แบบ PO

4.2 กรณีที่เป็นตัวแบบ NPO กำหนดเช่นเดียวกับตัวแบบ PO แต่ใช้ความสัมพันธ์เป็น NPO

4.3 กรณีที่เป็นตัวแบบ PPO สุ่มตัวแปรอธิบาย 1 ตัวให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ในกรณีที่มีจำนวนตัวแปรอธิบาย 3 ตัวแปร สุ่มตัวแปรอย่างละ 1 ตัวให้มีความสัมพันธ์เป็น NPO และ PO เมื่อมีจำนวนตัวแปรอธิบาย 5 ตัว สุ่มตัวแปรอธิบายอย่างละ 2 ตัวให้มี

ความสัมพันธ์เป็น NPO และ PO และเมื่อมีจำนวนตัวแปรอธิบาย 10 ตัว สุ่มตัวแปรอธิบาย 4 ตัวให้มีความสัมพันธ์เป็น NPO และตัวแปรอธิบายที่เหลือกำหนดเป็น PO

5. ตัวแปรตอบสนอง Y กำหนดให้เป็นตัวแปรเชิงอันดับที่มี 3 อันดับ ($Y = 1, 2, 3$)

6. วิธีการคัดเลือกตัวแปรด้วยวิธี RJ-MCMC ใช้จำนวนโซ่ 2 โซ่ จำนวนรอบซ้ำ 20,000 รอบ และตัดทิ้ง 10,000 รอบแรกทิ้ง (burn-in) และใช้ชุดคำสั่ง BayesOrd จาก <https://github.com/tjmckinley/BayesOrd>

1.4 เกณฑ์ที่ใช้ในการพิจารณา

เกณฑ์ที่ใช้ในการพิจารณาสมรรถนะ คือ

1. ร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย จะพิจารณาค่าความน่าจะเป็นภายหลังของความเกี่ยวพัน (Posterior Probability of Association: PPA) ที่ได้ในแต่ละตัวแปรอธิบายที่ถูกคัดเลือกเข้าในตัวแบบในลักษณะ PO, NPO หรือถูกคัดออกจากตัวแบบ เมื่อค่า PPA ในกลุ่มใดมีมากที่สุดตัวแปรจะถูกคัดเลือกเป็นลักษณะตัวแบบนั้น ๆ เพียงอย่างเดียวอย่างหนึ่งใน 3 รูปแบบ โดยค่า PPA ของตัวแปรในแต่ละตัวแบบสามารถคำนวณได้ดังนี้

$$PPA = \frac{1}{1 + BF_{01} \times \text{prior odds}} = \frac{BF_{10} \times \text{prior odds}(H_1)}{1 + BF_{10} \times \text{prior odds}(H_1)}$$

โดย

$$\text{prior odds}(H_1) = \frac{\Pr(H_1)}{\Pr(H_0)}$$

$$BF_{01} = \frac{\Pr(x|H_0)}{\Pr(x|H_1)}$$

และ

$$BF_{10} = \frac{\Pr(x|H_1)}{\Pr(x|H_0)}$$

เมื่อ H_0 : ตัวแปรอธิบาย (X) มีรูปแบบความสัมพันธ์กับตัวแปรตอบสนอง (Y) ตามที่กำหนด

H_1 : ตัวแปรอธิบาย (X) มีรูปแบบความสัมพันธ์กับตัวแปรตอบสนอง (Y) ไม่เป็นไปตามที่กำหนด

การอธิบายระดับของค่า PPA คือ

ถ้า PPA มีค่า < 0.50 แสดงว่าตัวแปรอธิบาย (X) ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง (Y) หรือตัวแปรนี้ไม่อยู่ในตัวแบบ

PPA มีค่า $0.50-0.74$ แสดงว่าตัวแปรอธิบาย (X) มีความสัมพันธ์กับตัวแปรตอบสนอง (Y) ค่อนข้างน้อย

PPA มีค่า $0.75-0.94$ แสดงว่าตัวแปรอธิบาย (X) มีความสัมพันธ์กับตัวแปรตอบสนอง (Y) ปานกลาง

PPA มีค่า $0.95-0.99$ แสดงว่าตัวแปรอธิบาย (X) มีความสัมพันธ์กับตัวแปรตอบสนอง (Y) มาก

PPA มีค่า ≥ 0.99 แสดงว่าตัวแปรอธิบาย (X) มีความสัมพันธ์กับตัวแปรตอบสนอง (Y) แน่นนอน

2. ร้อยละการจำแนกตัวแบบผิด (misclassification) โดยเฉลี่ย ซึ่งคำนวณจากค่าเฉลี่ยของร้อยละของการคัดเลือกตัวแปรและตัวแบบในกรณีที่ตัวแปรอธิบายถูกจำแนกผิดเป็นตัวแบบอื่น โดยพิจารณาจากตัวแปรอธิบายที่มีจำนวนตัวแปร, ขนาดตัวอย่าง, รูปแบบความสัมพันธ์และระดับความสัมพันธ์ระหว่างตัวแปรอธิบายเดียวกัน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

เป็นแนวทางการเลือกใช้วิธีมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้ในการศึกษาการคัดเลือกตัวแปรอธิบายและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับ เมื่อตัวแปรอธิบายมีรูปแบบความสัมพันธ์แบบต่าง ๆ

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องสำหรับการศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงเส้นแบบพหุโดยขั้นตอนวิธีมอนติคาร์โลโซ่มาร์คอฟกึ่งวาระโดดย้อนกลับได้ ประกอบด้วยทฤษฎีเกี่ยวกับตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Models: GLMs) การอนุมานแบบเบย์และวิธีมอนติคาร์โลโซ่มาร์คอฟกึ่งวาระโดดย้อนกลับได้ในการคัดเลือกตัวแปรและตัวแบบตัวแบบลอจิสต์สำหรับตัวแปรตอบสนองแบบอันดับ และรูปแบบเมทริกซ์ความสัมพันธ์ของตัวแปรอธิบาย

2.1.1 ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Models :GLMs)

วีรานันท์ (2555) ได้กล่าวว่าตัวแบบเชิงเส้นนัยทั่วไปเป็นตัวแบบที่ขยายจากตัวแบบเชิงเส้นแบบคลาสสิก (Classical linear model) โดยขยายใน 3 ส่วน คือ (1) ขยายส่วนประกอบเชิงสุ่ม (Random component) ของตัวแปรตอบสนองจากเดิมที่ใช้สำหรับการแจกแจงปกติไปสู่การใช้การแจกแจงในวงส์เลขชี้กำลัง (2) ขยายส่วนประกอบเชิงระบบ (Systematic component) จากเดิมที่เป็นผลรวมเชิงเส้นบนเทอมพารามิเตอร์ของตัวแปรอธิบายแบบต่อเนื่องให้ใช้ได้สำหรับตัวแปรอธิบายแบบจำแนกประเภททั้งหมด หรือเป็นแบบผสมทั้งแบบจำแนกประเภทหรือเชิงกลุ่มและแบบต่อเนื่อง และ (3) ขยายส่วนประกอบฟังก์ชันเชื่อมโยง (Link functions) ที่เดิมเคยใช้เฉพาะฟังก์ชันเชื่อมโยงเอกลักษณ์ให้ใช้ได้กับฟังก์ชันเชื่อมโยงแบบอื่นที่เป็นฟังก์ชันทางเดียวที่หาอนุพันธ์ได้ (Monotonic differentiable function) มีรายละเอียดของแต่ละส่วนประกอบดังนี้

2.1.1.1 ส่วนประกอบเชิงสุ่ม (Random component)

ส่วนประกอบเชิงสุ่มเป็นส่วนของลักษณะการแจกแจงของตัวแปรตอบสนองเชิงสุ่ม (Y) โดยมีค่าสังเกตของ Y ขนาด n หน่วย คือ $Y_i, i=1, \dots, n$ ที่เป็นอิสระต่อกันและ $Y = (y_1, \dots, y_n)$ มีการแจกแจงในวงส์เลขชี้กำลัง ที่มีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f_y(y; \theta, \phi) = \exp\left[\{y\theta - b(\theta)\} / a(\phi) + c(y, \phi)\right] \quad (2.1)$$

เมื่อ $a(\cdot)$, $b(\cdot)$ และ $c(\cdot)$ แทนฟังก์ชันต่าง ๆ และ (θ, ϕ) แทนพารามิเตอร์ของการแจกแจง

ถ้าทราบ ϕ แล้วสมการ (2.1) คือ การแจกแจงหนึ่งในวงศ์เลขชี้กำลังที่มีพารามิเตอร์ θ แต่ถ้าไม่ทราบ ϕ สมการ (2.1) อาจเป็นหรือไม่เป็นการแจกแจงหนึ่งในวงศ์เลขชี้กำลัง ที่มีพารามิเตอร์ 2 ตัว (θ, ϕ) สำหรับพารามิเตอร์ θ เรียกว่า Natural parameter ส่วน ϕ เรียกว่า Dispersion parameter

2.1.1.2 ส่วนประกอบเชิงระบบ (Systematic component)

ส่วนประกอบเชิงระบบของเซตของตัวแปรอธิบายที่มีระบบหรือรูปแบบเชิงเส้นในเทอมของพารามิเตอร์ หรือผลรวมเชิงเส้น $X\beta$ ซึ่งทำหน้าที่เชื่อมกับเวกเตอร์ η และเรียก $X\beta$ ว่าส่วนประกอบเชิงระบบ ดังนี้

$$\eta = X\beta$$

$$\eta_i = \sum_j \beta_j X_{ij}, i=1, \dots, n, j=1, \dots, k$$

เมื่อ η แทน ตัวพยากรณ์เชิงเส้น (Linear predictor) หรือ $\eta = (\eta_1, \dots, \eta_n)'$

X แทน เมทริกซ์ของตัวแปรอธิบาย อาจเรียก X ว่า Design matrix ที่มีขนาด $(n \times k)$

β แทน เวกเตอร์ของพารามิเตอร์ $(\beta_1, \dots, \beta_k)'$

เช่น ส่วนประกอบเชิงระบบ $X\beta = \sum_j \beta_j X_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, i=1, \dots, n, j=1, 2, 3$ ที่มีตัวแปรอธิบายจำนวน 3 ตัว คือ x_1, x_2 และ x_3 ดังนั้น η มีขนาด $(n \times 1)$, X ขนาด $(n \times 3)$, β มีขนาด (3×1)

2.1.1.3 ส่วนประกอบฟังก์ชันเชื่อมโยง (Link functions)

ฟังก์ชันเชื่อมโยง หมายถึง ฟังก์ชันของค่าเฉลี่ยของตัวแปรตอบสนองเชิงสุ่ม Y_i หรือ μ_i ที่ใช้เป็นฟังก์ชันเชื่อมโยงระหว่างส่วนประกอบเชิงสุ่มและส่วนประกอบเชิงระบบเข้าด้วยกัน ดังนี้

ให้ $\mu_i = E(Y_i), i=1, \dots, n$ โดย μ_i เกี่ยวพันกับ η_i ในรูปฟังก์ชันของค่าเฉลี่ย ดังนี้

$$\eta_i = g(\mu_i) = \sum_j \beta_j X_{ij}, i=1, \dots, n, j=1, \dots, k$$

เมื่อ g แทน ฟังก์ชันทางเดียวที่หาอนุพันธ์ได้ (Monotonic differentiable function)

และ k แทน จำนวนตัวแปรอธิบาย

ดังนั้น ตัวแบบที่ต้องการเชื่อมโยงระหว่างฟังก์ชันค่าเฉลี่ยของ Y กับเซตของตัวแปรอธิบาย X 's มีรูปแบบทั่วไป คือ

$$g(\mu_i) = \sum_j \beta_j X_{ij}, i = 1, \dots, n$$

ถ้า $g(\mu) = \mu$ จะได้ว่า $\eta_i = \mu_i$ คือ ฟังก์ชันเชื่อมโยงเอกลักษณ์ (Identity link) และอาจเรียกฟังก์ชันเชื่อมโยง อีกอย่างหนึ่งได้ว่าฟังก์ชันเชื่อมโยงคานอนิคัล (Canonical link) โดยมีการแปลงค่าเฉลี่ยให้อยู่ในเทอมของพารามิเตอร์ นั่นคือ

$$g(\mu_i) = Q(\theta_i)$$

และ

$$Q(\theta_i) = \sum_j \beta_j X_{ij}, i = 1, \dots, n, j = 1, \dots, k$$

โดยฟังก์ชันทางเดียวที่หาอนุพันธ์ได้ (Monotonic differentiable) มีรูปแบบคือ $\eta_i = g_i(\mu_i) = \beta' X_i$ ซึ่งโดยทั่วไปนิยมใช้ฟังก์ชันแบบเดียวกันสำหรับทุกค่าสังเกตของ $Y_i, i = 1, \dots, n$ นั่นคือ $\eta = g(\mu) = \beta' x$ หรือ $x\beta$

ตัวอย่างของฟังก์ชันเชื่อมโยงแบบต่าง ๆ ที่ใช้บ่อยมีดังนี้

1. $g(\mu) = \mu$ (Identity link function)
2. $g(\mu) = \frac{1}{\mu^2}$ (Quadratic inverse link function)
3. $g(\mu) = \log(\mu)$ (The logarithmic link function)
4. $g(p) = \log\left(\frac{p}{1-p}\right)$ (The logit link)
5. $g(p) = \log\{-\log(1-p)\}$ (The complementary log-log link)
6. $g(p) = -\log\{-\log(p)\}$ (The negative log-log link)
7. $g(p) = \tan\{\pi(p-1/2)\}$ (The Cauchit link or the inverse Cauchy link)
8. $g(p) = \Phi^{-1}(p)$ (The probit link)

เมื่อ Φ แทน ฟังก์ชันของการแจกแจงสะสมแบบปรกติมาตรฐาน และ p แทน ความน่าจะเป็นสำเร็จของตัวตอบสนองหรือ $p = P(Y=1|x)$ และ $g(p)$ แทนฟังก์ชันของค่าเฉลี่ยของตัวแปรตอบสนอง

2.1.2 การอนุมานแบบเบย์และวิธีมอนติคาร์โลโซ่มาร์คอฟ (Baysain inference and MCMC)

การอนุมานเชิงสถิติด้วยวิธีการของเบย์มีแนวคิด คือ พิจารณาพารามิเตอร์โดยการอนุมานเป็นตัวแปรสุ่ม ดังนั้นพารามิเตอร์จึงมีการแจกแจงความน่าจะเป็น ซึ่งอาจกำหนดขึ้นโดยประสบการณ์ ความเชื่อ หรือข้อมูลข้างเคียงการแจกแจงที่กำหนดนี้อาจไม่เหมาะสมที่จะใช้ประโยชน์ จึงใช้ข้อมูลหรือค่าสังเกตใหม่ที่ทำให้การปรับแก้การแจกแจงแรกเริ่มข้างต้น ได้การแจกแจงหลังและนำมาใช้ประโยชน์ต่อการอนุมานเชิงสถิติ

Congdon (2006) ได้อธิบายตัวแบบเบย์ไว้ดังนี้ ให้ $Y_1, Y_2, \dots, Y_n$ เป็นตัวอย่างสุ่มขนาด n มีการแจกแจงเป็น $f(y|\theta)$ โดยที่ θ เป็นค่าของตัวแปรสุ่ม Θ และ Θ มีการแจกแจงเป็น $g(\theta)$ เมื่อ $X_1, X_2, \dots, X_k$ เป็นตัวแปรอธิบาย เรียก $g(\theta)$ ว่าการแจกแจงก่อน (Prior distribution)

ฟังก์ชันภาวะน่าจะเป็น (Likelihood function) คือ

$$f(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^n f(y_i | \theta)$$

การแจกแจงภายหลัง (Posterior distribution) คือ

$$f(\theta | y_1, y_2, \dots, y_n) = \frac{\prod_{i=1}^n f(y_i | \theta) g(\theta)}{\int \prod_{i=1}^n f(y_i | \theta) g(\theta) d\theta}$$

โดย $\int \prod_{i=1}^n f(y_i | \theta) g(\theta) d\theta$ เรียกว่า การแจกแจงพยากรณ์ก่อน (Prior predictive distribution)

$$\text{และ } f(\theta | y_1, y_2, \dots, y_n) \propto \prod_{i=1}^n f(y_i | \theta) g(\theta)$$

การประมาณค่าพารามิเตอร์ในการแจกแจงภายหลัง (Posterior distribution) คือ การหาค่าของพารามิเตอร์แบบจุด (point estimate) ที่ทำให้การแจกแจงภายหลังมีค่าสูงสุด ถ้าการแจกแจงภายหลังมีมิติขนาดใหญ่จะทำให้มีค่า local maximum จำนวนมาก การหาจุดที่ทำให้การแจกแจงภายหลังมีค่าสูงสุดทำได้ยาก จนมีการค้นพบวิธีการหาของมอนติคาร์โลสำหรับการประมาณค่า ทำให้มีการนำการแจกแจงภายหลังไปใช้อย่างกว้างขวาง

วิธีมอนติคาร์โลใช้หลักการคือ ถ้าต้องการรู้ค่าพารามิเตอร์ของการแจกแจงใด จะใช้การสุ่มตัวอย่างจากการแจกแจงนั้นซ้ำกันหลาย ๆ ครั้ง แล้วดูผลที่เกิดขึ้นในการสุ่มตัวอย่างแต่ละครั้ง เช่น ถ้าต้องการคำนวณค่าคาดหวังภายหลังสามารถคำนวณได้จากสูตร

$$E(\theta|y_1, y_2, \dots, y_n) = \int \theta f(\theta|y_1, y_2, \dots, y_n) d\theta$$

เมื่อสร้างลำดับการสุ่มตัวอย่างแบบสุ่ม จำนวน G ครั้ง กำหนดให้ $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(G)}$ จาก $f(\theta|y_1, y_2, \dots, y_n)$ ก็จะสามารถประมาณค่าคาดหวังภายหลังได้จาก

$$E(\theta|y_1, y_2, \dots, y_n) = \int \theta f(\theta|y_1, y_2, \dots, y_n) d\theta = \frac{1}{G} \sum_{g=1}^G \theta^{(g)}$$

ความถูกต้องของค่าประมาณขึ้นอยู่กับขั้นตอนวิธีการที่ใช้สำหรับสุ่มตัวอย่างและจำนวนครั้งของการสุ่มขั้นตอนวิธีการที่ใช้แก้ปัญหาในงานของเบส์ที่ใช้กันมากคือ วิธีการ Gibbs Sampling และวิธีการของเมโทรโพลิส-เฮสติ้ง (Metropolis-Hastings) โดยทั้ง 2 วิธีการของเป็นวิธีของมอนติคาร์โลโซมาร์คอฟ (MCMC) คือ ลำดับของการสุ่มตัวอย่างเป็นอิสระกัน และครั้งที่ของการสุ่มตัวอย่างขึ้นอยู่กับสุ่มตัวอย่างครั้งก่อนหน้าในที่นี้จะกล่าวรายละเอียดเฉพาะขั้นตอนวิธีการของเมโทรโพลิส-เฮสติ้ง

2.1.2.1 ขั้นตอนวิธีการของเมโทรโพลิส-เฮสติ้ง (Metropolis-Hastings

Algorithm: MH)

ขั้นตอนวิธีการของเมโทรโพลิส-เฮสติ้งโดยทั่วไป มีวัตถุประสงค์เพื่อต้องการจำลองตัวอย่างจากการแจกแจงของ $\pi(\theta)$ และมีแนวคิดจากการผันกลับของโซมาร์คอฟสามารถสรุปหลักการของเมโทรโพลิส-เฮสติ้งได้ดังนี้

1. กำหนดให้ $t = 0$ และจุดเริ่มต้น θ_0 โดยที่ $\pi(\theta_0) > 0$

2. สุ่มตัวแคนดิเดต (Candidate) θ' จากการแจกแจงก่อนหน้า $Q(\theta': \theta_t)$ คำนวณหาความน่าจะเป็นในการยอมรับ (Acceptance Probability) โดยจะเป็นอัตราส่วนของฟังก์ชันความหนาแน่นระหว่างตัวอย่างปัจจุบันกับตัวแคนดิเดต

$$\alpha(\theta_t, \theta') = \begin{cases} \min\left(\frac{\pi(\theta')Q(\theta_t: \theta')}{\pi(\theta_t)Q(\theta': \theta_t)}, 1\right) & \text{if } \pi(\theta_t)Q(\theta': \theta_t) > 0 \\ 1 & \text{if } \pi(\theta_t)Q(\theta': \theta_t) = 0 \end{cases}$$

3. สร้างเลขสุ่มของการยอมรับ $U \sim \text{uniform}(0,1)$

โดย $\theta_{t+1} = \theta'$ เมื่อ $u < \alpha(\theta_t, \theta')$ ด้วยความน่าจะเป็น $\alpha(\theta_t, \theta')$

$\theta_{t+1} = \theta_t$ เมื่อ $u \geq \alpha(\theta_t, \theta')$ ด้วยความน่าจะเป็น $1 - \alpha(\theta_t, \theta')$

หมายเหตุ การกำหนดการแจกแจงก่อน $Q(*:*)$ เป็นแบบใดนั้นไม่มีผลต่อการแจกแจงที่ต้องการ $\pi(\cdot)$ แต่จะมีผลกับอัตราการลู่เข้าสู่การแจกแจงที่ต้องการ $\pi(\cdot)$ โดยถ้ากำหนดการแจกแจงก่อนเหมาะสมกับการแจกแจงที่ต้องการ $\pi(\cdot)$ แล้วจะทำให้อัตราการลู่เข้าสูงในการกำหนดการแจกแจงก่อนสามารถแบ่งได้หลายประเภท ดังนี้

1. โซ่แบบสมมาตร (Symmetric chain) โดยจะกำหนดให้

$$Q(\theta':\theta_t) = Q(\theta_t:\theta')$$

ดังนั้น

$$\alpha(\theta_t, \theta') = \min\left(\frac{\pi(\theta')}{\pi(\theta_t)}, 1\right)$$

2. โซ่แบบเดินสุ่ม (Random-Walk chain) โดยกำหนดให้

$$Q(\theta':\theta_t) = Q(\theta_t:\theta')$$

ดังนั้น

$$\alpha(\theta_t, \theta') = \min\left(\frac{\pi(\theta')}{\pi(\theta_t)}, 1\right)$$

3. โซ่แบบอิสระ (Independence chain) โดยกำหนดให้

$$Q(\theta':\theta_t) = Q(\theta')$$

ดังนั้น

$$\alpha(\theta_t, \theta') = \min\left(\frac{w(\theta')}{w(\theta_t)}, 1\right) \quad \text{เมื่อ } w(\theta_t) = \pi(\theta_t)/Q(\theta_t)$$

2.1.3 ตัวแบบลอจิตสำหรับตัวแปรตอบสนองแบบอันดับ

ให้ $Y = (Y_1, \dots, Y_J)$ เป็นจำนวนตัวอย่างในแต่ละกลุ่มที่เรียงลำดับ ($j = 1, 2, \dots, J$) ดังนั้น Y มีการแจกแจงแบบอนอกนาม (Multinomial Distribution) คือ

$$Y \sim \text{Mult}(n, p) \quad (2.2)$$

เมื่อ n คือจำนวนตัวอย่างทั้งหมด และ $p = (p_1, \dots, p_J)$ คือ ความน่าจะเป็นที่แต่ละหน่วยตัวอย่างจะอยู่ในกลุ่มที่ j โดยที่ $\sum_{j=1}^J p_j = 1$ ในกรณีที่มีจำนวนตัวแปรอธิบายจำนวน k ตัว

$X_i = (X_{i1}, \dots, X_{iJ})$ สำหรับกลุ่มตัวอย่างจำนวน i กลุ่ม ($i = 1, \dots, I$) โดยที่ จำนวนตัวอย่างทั้งหมด $n = n_1 + \dots + n_I$ แล้ว จะได้ว่า

$$Y_i \sim \text{Mult}(n_i, p_i) \quad (2.3)$$

เมื่อ $p_i = (p_{i1}, \dots, p_{iJ})$ และ $\sum_{j=1}^J p_{ij} = 1$

ดังนั้นถ้า $I = n$ แล้วจะได้ว่า $n_i = 1$ สำหรับทุก i กำหนดให้ C_i คือกลุ่มที่หน่วยตัวอย่างที่ i อยู่ (C_i มีค่าเป็น $1, 2, \dots, J$) แล้วจะได้ความน่าจะเป็นสะสม $P(C_i \leq j) = \gamma_{ij}$

การสร้างตัวแบบของ γ_{ij} โดยเลือกฟังก์ชันเชื่อมโยงที่มีทิศทางเดียวและสามารถหาอนุพันธ์ได้ดังนี้

$$h(\gamma_{ij}) = \theta_j - \mu_i \quad (2.4)$$

เมื่อ $\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$

และ $\beta = (\beta_0, \dots, \beta_k)$ จึงมี $k+1$ พารามิเตอร์ถดถอย

θ_j คือ จุดตัด (cut point) ที่เป็นตัวแปรต่อเนื่องที่ไม่ทราบค่า

โดยที่ $-\infty < \theta_1 < \dots < \theta_{J-1} < \infty$ ในที่นี้เพื่อให้ตัวแบบสามารถประมาณค่าได้จะกำหนดให้ $\beta_0 = 0$ และความน่าจะเป็นของแต่ละกลุ่มเป็นดังนี้

$$p_{ij} = \begin{cases} \gamma_{ij} & \text{for } j = 1, \\ \gamma_{ij} - \gamma_{i(j-1)} & \text{for } j = 2, \dots, J-1, \\ 1 - \gamma_{i(J-1)} & \text{for } j = J \end{cases} \quad (2.5)$$

ในตัวแบบ (2.4) อิทธิพลของตัวแปรอธิบายจะเป็นอิสระกับกลุ่มหรือระดับของตัวแปรตอบสนอง ดังนั้นตัวแบบนี้จะแสดงการมีลำดับสุ่มอย่างแท้จริง นั้นหมายความว่าจากเงื่อนไข $-\infty < \theta_1 < \dots < \theta_{J-1} < \infty$ ความน่าจะเป็นสะสมของตัวแบบจะเป็นเรียงลำดับ $0 < \gamma_{i1} < \gamma_{i2} < \dots < \gamma_{i(J-1)} < 1$

แต่ในกรณีทั่วไปอิทธิพลของตัวแปรอธิบายอาจจะแตกต่างระหว่างระดับหรือกลุ่ม ดังนี้

$$h(\gamma_{ij}) = \theta_j - \mu_{ij} \quad (2.6)$$

ในกรณีตัวแบบมีลักษณะลำดับสุ่มเพียงบางช่วงของตัวแปรอธิบายเท่านั้น โดยตัวแบบทั้ง 2 นี้จะกล่าวในรายละเอียดในหัวข้อถัดไป โดยจะเน้นเฉพาะการใช้ฟังก์ชันเชื่อมโยงแบบลอจิสติกเท่านั้น

2.1.3.1 ตัวแบบลอจิสติกเชิงสัดส่วน (Proportional odds logit model:

PO)

ตัวแบบลอจิสติกเชิงสัดส่วน มีรูปแบบดังนี้

$$\begin{aligned} h(\gamma_{ij}) &= \log\left(\frac{\gamma_{ij}}{1 - \gamma_{ij}}\right) = \theta_j - \mu_i \\ h(\gamma_{ij}) &= \log\left(\frac{\gamma_{ij}}{1 - \gamma_{ij}}\right) = \theta_j - (\beta_1 x_{i1} + \dots + \beta_k x_{ik}) \end{aligned} \quad (2.7)$$

ตัวแบบลอจิสติกเชิงสัดส่วนมีเงื่อนไขของเวกเตอร์ความชันที่ส่งผลต่อระดับของ Y_j นั้นคงที่ กล่าวคือ $\beta_j = \beta$ คงที่ในแต่ละระดับของ Y_j เมื่อ $j = 1, \dots, J-1$ ดังนั้นรูปแบบของแต่ละลอจิสติกใน

รูปแบบเชิงเส้นจะขนานกันทั้ง $J-1$ เส้น และอัตราส่วนลิกออดส์สะสมสำหรับตัวแปรอธิบายสองชุด (X_1 และ X_2) คือ

$$\begin{aligned}\text{logit}(\gamma_{1j}) - \text{logit}(\gamma_{2j}) &= \theta_j - \beta^T X_1 - \theta_j + \beta^T X_2 \\ &= \beta^T (X_2 - X_1)\end{aligned}\quad (2.8)$$

เมื่อ θ_j แทน พารามิเตอร์จุดตัด (cut-point) ที่ระดับของตัวแปรตอบสนองที่ j

β แทน เวกเตอร์ความชัน (Slope)

จะเห็นว่าอัตราส่วนลิกออดส์สะสมเป็นสัดส่วนระหว่างระยะห่างของตัวแปรอธิบาย X_1 และ X_2 ในกรณีที่มีการแจกแจงก่อนของพารามิเตอร์ θ และ β เป็นอิสระต่อกัน ดังนั้นการแจกแจงก่อนหน้าร่วม (joint prior distribution) สามารถเขียนได้ดังนี้ $f(\beta, \theta) = f(\theta) f(\beta)$ เมื่อ

$$f(\beta) = \prod_{k=1}^K f(\beta_k) \text{ และ } f(\theta) = f(\theta_1) \prod_{j=2}^{J-1} f(\theta_j | \theta_{j-1}) \quad (2.9)$$

โดย $f(\theta_1)$ จะกำหนดในช่วง $(-\infty, \infty)$ และ $f(\theta_j | \theta_{j-1})$ จะกำหนดในช่วง (θ_{j-1}, ∞) สำหรับ $j = 2, \dots, J-1$ เพื่อให้เงื่อนไขของการเรียงลำดับสุ่มเป็นจริงสำหรับทุกค่าของ β จึงกำหนดให้ $\theta_1 \sim N(0, \sigma_\theta^2)$, $\beta_k \sim N(0, \sigma_\beta^2)$ (สำหรับ $k = 1, \dots, K$) และ

$$\theta_j | \theta_{j-1} \sim N(0, \sigma_\theta^2) I(T_{j-1}, \infty) \text{ สำหรับ } j = 2, \dots, J-1 \quad (2.10)$$

โดย $I(T_{j-1}, \infty)$ จะหมายถึงการแจกแจงจะถูกตัดทอนในขอบเขต (T_{j-1}, ∞) เพื่อ $T_{j-1} = \theta_{j-1}$ ในงานวิจัยนี้ β_k แต่ละตัวจะได้มาจากการแจกแจงการเดินสุ่มปกติ (Normal Random Walk) นั่นคือ

$$\beta_k' | \beta_k^{(i)} \sim N(\beta_k^{(i)}, \sigma_{p\beta}^2) \quad (2.11)$$

โดย $\sigma_{p\beta}^2$ คือ ความแปรปรวน สำหรับพารามิเตอร์จุดตัด θ_j สร้างจากการเดินสุ่มของการแจกแจงเอกรูปแบบตัดทอน นั่นคือ

$$\theta_j' | \theta^{(i)} = \begin{cases} U(\theta_j^{(i)} - \tau_\theta, \min[\theta_j^{(i)} + \tau_\theta, \theta_{j+1}^{(i)}]) & \text{if } j = 1, \\ U(\max[\theta_j^{(i)} - \tau_\theta, \theta_{j-1}^{(i)}], \min[\theta_j^{(i)} + \tau_\theta, \theta_{j+1}^{(i)}]) & \text{if } j = 2, \dots, J-2, \\ U(\max[\theta_j^{(i)} - \tau_\theta, \theta_{j-1}^{(i)}], \theta_j^{(i)} + \tau_\theta) & \text{if } j = J-1 \end{cases} \quad (2.12)$$

2.1.3.2 ตัวแบบลอจิสติกที่ไม่เป็นตามสัดส่วน (Non-proportional odds

logit model: NPO)

ตัวแบบลอจิสติกที่ไม่เป็นตามสัดส่วนมีรูปแบบดังนี้

$$\begin{aligned}\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) &= \theta_j - \mu_{ij} \\ \log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) &= \theta_j - (\beta_{1j}X_{i1} + \dots + \beta_{Kj}X_{iK})\end{aligned}\quad (2.13)$$

ในตัวแบบนี้พารามิเตอร์ถดถอยจะแตกต่างกันตามระดับของกลุ่ม j สิ่งสำคัญคือเงื่อนไขของลำดับสุ่มนั้นคือ

$$-\infty < \theta_1 - \beta_1 X < \theta_2 - \beta_2 X < \dots < \theta_{J-1} - \beta_{J-1} X < \infty \quad (2.14)$$

สำหรับทุก X ในงานวิจัยนี้จะกำหนดให้ $\beta_{0j} = 0$ และถ้ามีตัวแปรอธิบาย K ตัว จะได้ว่า สำหรับ $j = 1, \dots, J-2$

$$\theta_j - \theta_{j+1} < \sum_{k=1}^K (\beta_{kj} - \beta_{k(j+1)}) X_k \quad (2.15)$$

จะต้องเป็นจริงสำหรับทุก ๆ ค่าของ X_k ดังนั้น ถ้าทราบขอบเขตล่างและขอบเขตบนที่เป็นไปได้ของ X_k แล้วค่า

$$\theta_j - \theta_{j+1} < \sum_{k=1}^K \min(\beta_{kj} - \beta_{k(j+1)}) X_k^m, [\beta_{kj} - \beta_{k(j+1)}] X_k^M \quad (2.16)$$

เมื่อ X_k^m และ X_k^M คือ ค่าต่ำสุดและค่าสูงสุดของ X_k ตามลำดับ หรือกำหนดให้

$$C_{kj} = \min[X_k^m (\beta_{kj} - \beta_{k(j+1)}), X_k^M (\beta_{kj} - \beta_{k(j+1)})] \text{ และ } C_j = \sum_{k=1}^K C_{kj} \quad (2.17)$$

การแจกแจงก่อนหน้าร่วมของ β และ θ คือ

$$\begin{aligned}f(\beta, \theta) &= f(\theta|\beta) f(\beta) \\ &= f(\theta_1) \left[\prod_{j=2}^{J-1} f(\theta_j | \theta_{j-1}, \beta_j, \beta_{j-1}) \right] \prod_{k=1}^K f(\beta_k)\end{aligned}\quad (2.18)$$

โดย $f(\beta_k) (k=1, \dots, K)$ และ $f(\theta_1)$ กำหนดให้ได้เช่นเดียวกับของตัวแบบ PO ในสมการ (2.9) และ $f(\theta_j | \theta_{j-1}, \beta_j, \beta_{j-1})$ คือ ฟังก์ชันความหนาแน่นความน่าจะเป็นสำหรับการแจกแจงปกติแบบตัดทอน $N(0, \sigma_\theta^2) I(T_{j-1}, \infty)$ เมื่อ $T_{j-1} = \theta_{j-1} - C_{j-1}$ ในงานวิจัยนี้จะกำหนดให้

$$\beta_{kj}' | \beta_{kj}^{(i)} \sim U(\beta_{kj}^{(i)} - \tau_\beta, \beta_{kj}^{(i)} + \tau_\beta) \quad (2.19)$$

โดย $\tau_\beta > 0$ จะควบคุมขนาดความแตกต่างสูงสุดของค่าที่ได้กับค่าปัจจุบันในแต่ละการวนซ้ำ สำหรับพารามิเตอร์จุดตัด θ_j เลือกการแจกแจงจากการเดินสุ่มเอกรูปแบบตัดทอน โดยที่

$$\theta_j^* | \theta^{(i)} = \begin{cases} U(\theta_j^{(i)} - \tau_\theta, \min[\theta_j^{(i)} + \tau_\theta, \theta_{j+1}^{(i)} + C_j^{(i)}]) & \text{if } j = 1, \\ U(\max[\theta_j^{(i)} - \tau_\theta, \theta_{j-1}^{(i)} - C_j^{(i)}], \min[\theta_j^{(i)} + \tau_\theta, \theta_{j+1}^{(i)} + C_j^{(i)}]) & \text{if } j = 2, \dots, J-2, \\ U(\max[\theta_j^{(i)} - \tau_\theta, \theta_{j-1}^{(i)} - C_j^{(i)}], \theta_j^{(i)} + \tau_\theta) & \text{if } j = J-1 \end{cases} \quad (2.20)$$

2.1.3.3 ตัวแบบลอจิสติกบางส่วน (Partial proportional odds logit model: PPO)

ตัวแบบลอจิสติกบางส่วน นำเสนอโดย Peterson และ Harrell (1990) เป็นตัวแบบผสมระหว่างตัวแบบลอจิสติกเชิงสัดส่วน (Proportional odds: PO) และตัวแบบลอจิสติกที่ไม่เป็นตามสัดส่วน (Non-proportional odds model: NPO) จะอยู่ในรูปแบบสมการดังนี้

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j - \beta^T X_i - \eta_j^T U_i \quad (2.21)$$

โดย $i=1, \dots, n$ และ $j=1, \dots, J-1$ ซึ่ง β เป็นพารามิเตอร์ถดถอยที่สอดคล้องกับตัวแปรอธิบาย X_i ที่มีความสัมพันธ์เป็นตัวแบบออดส์เชิงสัดส่วน และ η_j พารามิเตอร์ถดถอยที่สอดคล้องกับตัวแปรอธิบาย U_i ที่มีความสัมพันธ์เป็นตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน ตามที่ได้อธิบายในหัวข้อ (2.1.3.1) และ (2.1.3.2) โดยที่ตัวแปร PO และ NPO (X_i และ U_i) ต้องทราบล่วงหน้า

2.1.4 วิธีมอนติคาร์โลโซมาร์คอฟกึ่งการกระโดดย้อนกลับได้ในการคัดเลือกตัวแบบและคัดเลือกตัวแปร

2.1.4.1 การคัดเลือกตัวแบบแบบเบสด้วยวิธีมอนติคาร์โลโซมาร์คอฟกึ่งการกระโดดย้อนกลับได้

สมมติว่ามีตัวแบบแข่งขันที่ต้องการเลือกจำนวน V ตัวแบบ เมื่อ V คือจำนวนตัวแบบ แนวทางการคัดเลือกตัวแบบแบบเบสจะทำได้โดยการประมาณความน่าจะเป็นภายหลังที่ถือว่าตัวแบบ (M_v) เป็นตัวแบบที่ถูกคัดเลือกจากตัวแบบที่เป็นไปได้ทั้งหมด ($M_1, \dots, M_v$) โดยมีสูตรความน่าจะเป็นภายหลังดังนี้

$$P(M_v | D) = \frac{f(D | M_v) P(M_v)}{\sum_{u=1}^V f(D | M_u) P(M_u)} \quad (2.22)$$

โดย $P(M_v)$ คือ ความน่าจะเป็นก่อนของตัวแบบ M_v (prior probability)

และฟังก์ชันภาวะน่าจะเป็นคือ

$$f(D|M_v) = \int_{\Omega_v} f(D|\omega_v, M_u) f(\omega_v|M_u) d\omega_v \quad (2.23)$$

โดยที่ D คือ ข้อมูล

Ω_v คือ ปริภูมิพารามิเตอร์ (หลายมิติ)

ω_v คือ เวกเตอร์พารามิเตอร์ของตัวแบบ M_u

สมการ (2.22) บางครั้งจะเรียกว่าความน่าจะเป็นภายหลังของความสัมพันธ์ (posterior probability of association, PPA)

การคัดเลือกตัวแบบโดยใช้วิธีมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้ (Reversible-Jump Markov chain Monte Carlo: RJ-MCMC) จะทำได้โดย กำหนดให้ $v=i$ เป็นตัวบ่งชี้ของตัวแบบสำหรับการวนซ้ำในรอบนั้น ๆ โซ่จะอยู่ในตัวแบบ M_i และให้ $P(M_i \rightarrow M_j)$ คือ ความน่าจะเป็นที่จะกระโดดจากตัวแบบ M_i ไปตัวแบบ M_j โดย M_j คาดว่าจะเป็นตัวแบบที่จะถูกเลือก ซึ่งตัวแบบทั้ง 2 นี้อาจมีจำนวนตัวแปรอธิบายในตัวแบบหรือจำนวนพารามิเตอร์หรือมิติแตกต่างกัน ดังนั้นเพื่อให้การข้ามระหว่างตัวแบบของมิติที่แตกต่างกันได้ พารามิเตอร์ในตัวแบบ i หรือ ω_i จะถูกจับคู่กับชุดพารามิเตอร์ของตัวแบบ j หรือ ω_j โดยการสร้างชุดพารามิเตอร์ดัมมี่ u_i และ u_j ขึ้นมาเพื่อให้ได้ว่า มิติ(ω_i, u_i) = มิติ(ω_j, u_j) หลังจากนั้นจะใช้ฟังก์ชัน g_{ij} (bijective function) เพื่อจะเชื่อมระหว่างคู่ลำดับของ (ω_i, u_i) ไป (ω_j, u_j) นั่นคือ $g_{ij}(\omega_i, u_i) = (\omega_j, u_j)$ และในทางกลับกัน การเชื่อมระหว่างคู่ลำดับของ (ω_j, u_j) ไป (ω_i, u_i) ก็ใช้ฟังก์ชัน $g_{ji}(\omega_j, u_j) = (\omega_i, u_i)$ โดยความน่าจะเป็นในการยอมรับของการย้ายจะกำหนดได้ดังนี้

$$\alpha = \min \left[1, \frac{f(D|\omega_j, M_j)}{f(D|\omega_i, M_i)} \times \frac{f(\omega_j|M_j)}{f(\omega_i|M_i)} \times \frac{P(M_j)}{P(M_i)} \times \frac{P(M_j \rightarrow M_i)q_u(u_j)}{P(M_i \rightarrow M_j)q_u(u_i)} \times \left| \frac{\partial(\omega_j, u_j)}{\partial(\omega_i, u_i)} \right| \right] \quad (2.24)$$

โดย $q_u(u_i)$ และ $q_u(u_j)$ คือฟังก์ชันความหนาแน่นน่าจะเป็นที่เสนอสำหรับพารามิเตอร์ดัมมี่ u_i

และ u_j ตามลำดับ และ $\left| \frac{\partial(\omega_j, u_j)}{\partial(\omega_i, u_i)} \right|$ คือค่าดีเทอร์มิแนนของเมทริกซ์จาโคเบียน

2.1.4.2 การคัดเลือกตัวแปรในตัวแบบการถดถอยลอจิสติกอันดับโดยใช้วิธีมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้

ในงานวิจัยนี้วิธีการมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้ที่ใช้ในการคัดเลือกตัวแปรในตัวแบบการถดถอยลอจิสติกแบบอันดับ นั้นจะใช้ตามวิธีการที่เสนอโดย MicKinley และคณะ (2015) โดยมีรายละเอียดดังนี้

สมมติให้มีตัวแปรอธิบายจำนวน K ตัว แต่ละตัวจะมีรูปแบบความสัมพันธ์กับตัวแปรตอบสนอง Y ที่เป็นไปได้โดยวิธีใดวิธีหนึ่งใน 3 รูปแบบ คือมีความสัมพันธ์ในลักษณะตัวแบบ PO, NPO หรือไม่มีความสัมพันธ์กับตัว Y ดังนั้นจำนวนตัวแบบที่เป็นไปได้ทั้งหมดมีเท่ากับ 3^K ตัวแบบ และการสร้างตัวแบบจะทำได้โดยการสร้างตัวแปรบ่งชี้ (indicator variable) สำหรับตัวแปรแต่ละตัว คือ S_k สำหรับ $k=1,...,K$ โดยกำหนดให้

$$S_k = \begin{cases} 0 & \text{ถ้าตัวแปรอธิบาย } X_k \text{ มีความสัมพันธ์เป็น PO} \\ 1 & \text{ถ้าตัวแปรอธิบาย } X_k \text{ มีความสัมพันธ์เป็น NPO} \\ 2 & \text{ถ้าตัวแปรอธิบาย } X_k \text{ ถูกคัดออกจากตัวแบบ} \end{cases} \quad (2.25)$$

ทั้ง 3 กรณีนี้อาจพิจารณาว่าเป็นกรณีพิเศษของตัวแบบ NPO โดยถ้าตัวแปรมีความสัมพันธ์กับ Y ในตัวแบบ PO แล้วจะกำหนดให้ $\beta_{kj} = \beta_k$ สำหรับ $j=1,...,J-1$ โดยการแจกแจงก่อนจะอยู่ในรูปจุดที่เป็นอิสระกัน (independent point-mass prior) $\beta_{k2},..., \beta_{k(J-1)}$ เพื่อให้ $f(\beta_{kj} = \beta_k) = 1$ และในกรณีที่ตัวแปรไม่มีความสัมพันธ์หรือตัวแปรถูกคัดออกจากตัวแบบนั้น ทำได้โดยการกำหนดให้ $\beta_{kj} = 0$ ที่มีการแจกแจงก่อนในรูปจุด $f(\beta_{kj} = 0) = 1$

(1) การคัดตัวแปรเข้าหรือออกจากตัวแบบ

การคัดตัวแปรที่ขณะนี้ไม่อยู่ในตัวแบบเข้ามาเพิ่มในตัวแบบ หรือตัดตัวแปรที่ขณะนี้อยู่ในตัวแบบออกจากตัวแบบจะทำได้ด้วยความน่าจะเป็น P_{jump} และจะไม่มีการนำเข้าหรือคัดออกด้วยความน่าจะเป็น $1-P_{jump}$

ในกรณีที่มีการคัดตัวแปรเพิ่มเข้าในตัวแบบ จะกำหนดให้ความน่าจะเป็นในการคัดตัวแปรเข้าในตัวแบบด้วยความสัมพันธ์แบบ PO และ NPO คือ p_{PO} และ $p_{NPO} = 1 - p_{PO}$ ตามลำดับ และกำหนดฟังก์ชันการเพิ่มตัวแปรด้วยความสัมพันธ์แบบ PO คือ

$$g_{0 \rightarrow PO}(u_1) = u_1 = \beta_k \quad (2.26)$$

เมื่อ u_1 ถูกสุ่มจากการแจกแจงที่มีฟังก์ชันความหนาแน่นน่าจะเป็น $q_u(\cdot)$ และการเพิ่มตัวแปรในตัวแบบโดยความสัมพันธ์แบบ NPO จะกำหนดฟังก์ชันดังนี้

$$g_{0 \rightarrow NPO}(u_1, ..., u_{J-1}) = (u_1, ..., u_{J-1}) = (\beta_{k1}, ..., \beta_{k(J-1)}) \quad (2.27)$$

โดย $u_1, ..., u_{J-1}$ เป็นอิสระกันและมีการแจกแจงเหมือนกัน ถูกสุ่มมาจากการแจกแจงที่มีฟังก์ชันความหนาแน่นน่าจะเป็น $q_u(\cdot)$

ความน่าจะเป็นการยอมรับการเพิ่มตัวแปรด้วยตัวแบบ PO หรือตัวแบบจะเปลี่ยนจากตัวแบบที่เป็นอยู่ปัจจุบันเป็นตัวแบบที่มีตัวแปรเพิ่มขึ้นด้วยความสัมพันธ์แบบ PO หรือ $0 \rightarrow PO$ คือ

$$\alpha = \min \left[1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{f(\beta'_k)}{1} \times \frac{1}{q_u(u_1)} \times \frac{1}{p_{PO}} \right] \quad (2.28)$$

สำหรับการเพิ่มตัวแปรด้วยตัวแบบ NPO หรือจากตัวแบบ $0 \rightarrow \text{NPO}$ ความน่าจะเป็นการยอมรับคือ

$$\alpha = \min \left[1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{\left[\prod_{j=1}^{J-1} f(\beta'_{kj}) \right] \left[\prod_{j=2}^{J-1} f(\theta_j^{(i)} | \theta_{j-1}^{(i)}, \beta'_j, \beta'_{j-1}) \right]}{\left[\prod_{j=2}^{J-1} f(\theta_j^{(i)} | \theta_{j-1}^{(i)}, \beta_j^{(i)}, \beta_{j-1}^{(i)}) \right]} \times \frac{1}{q_u(u_1, \dots, u_{J-1})} \times \frac{1}{1 - p_{PO}} \right] \quad (2.29)$$

กำหนด u_j เป็นตัวแปรสุ่มที่อิสระกันและมีการแจกแจงเหมือนกันที่มีการแจกแจงปกติค่าเฉลี่ยเป็นศูนย์ และ $\sigma_{p\beta}^2$ คือ ความแปรปรวนที่นำเสนอ เมื่อกำหนด $\sigma_{p\beta}^2 = 1$ หรือเขียนได้เป็น $u_j \sim N(0,1)$

การคัดตัวแปรออกจากตัวแบบ ทำได้โดยย้อนกลับกระบวนการการคัดตัวแปรเข้า โดยเปลี่ยนความน่าจะเป็นที่ยอมรับได้ให้เป็นความน่าจะเป็นของการคัดตัวแปรออก

(2) การปรับปรุงตัวแบบกรณีที่มีการคัดเลือกตัวแปรเข้ามาเพิ่มในตัวแบบ

ในกรณีที่ตัวแปรตัวใหม่เพิ่มเข้ามาในตัวแบบ RJ-MCMC จะทำการปรับปรุงตัวแบบ 2 ประการคือ ประการแรกจะทำการปรับปรุงค่าพารามิเตอร์ของตัวแปรที่ละตัว (สมมติ ตัวแปร K) โดยลำดับการปรับปรุงจะทำแบบสุ่ม ซึ่งค่าพารามิเตอร์ค่าใหม่จะถูกแทนค่าเดิมด้วยความน่าจะเป็น P_{move} และประการที่สอง จะทำการสลับความสัมพันธ์จากตัวแบบ $PO \rightarrow \text{NPO}$ (ถ้าตัวแปร K มีความสัมพันธ์แบบ PO) หรือ $\text{NPO} \rightarrow PO$ (ถ้าตัวแปร K มีความสัมพันธ์แบบ NPO) ด้วยความน่าจะเป็น $1 - P_{move}$

กรณีที่ตัวแปร K มีความสัมพันธ์แบบ PO การปรับค่าของ β_k ด้วยค่าใหม่ที่ได้จากการแจกแจงที่เสนอที่มีฟังก์ชันความหนาแน่นความน่าจะเป็น $q_\beta(\beta_k^{(i)})$ ขั้นตอนนี้จะใช้วิธีการมาตรฐานของเมโทรโพลิส-เฮสติงส์ เช่นเดียวกันกับ β_{kj} ($j=1, \dots, J-1$) ถ้าตัวแปร k มีความสัมพันธ์แบบ NPO

สำหรับขั้นตอนการสลับลักษณะความสัมพันธ์จะอาศัยขั้นตอนวิธีมอนติคาร์โลไฮมาร์คอฟก้าวกระโดดย้อนกลับได้ โดยการสลับตัวแบบ $\text{NPO} \rightarrow PO$ หรือเชื่อมโยง $(\beta_{k1}, \dots, \beta_{k(J-1)}) \rightarrow \beta_k$ ทำได้โดยการกำหนดฟังก์ชันดังนี้

$$\begin{aligned} g_{NPO \rightarrow PO}(\beta_{k1}, \dots, \beta_{k(J-1)}) &= (\bar{\beta}_{k\cdot}, \bar{\beta}_{k\cdot} - 2\beta_{k2}, \dots, \bar{\beta}_{k\cdot} - 2\beta_{k(J-1)}) \\ &= (\beta_k, u_1, \dots, u_{J-2}) \end{aligned} \quad (2.30)$$

โดย $\bar{\beta}_{k\cdot} = (J-1)^{-1} \sum_{j=1}^{J-1} \beta_{kj}$ และการสลับกลับจะใช้ฟังก์ชันผกผันนั้นคือ

$$\begin{aligned} g_{PO \rightarrow NPO}(\beta_k, u_1, \dots, u_{J-2}) &= \frac{J-1}{2} \left(\beta_k (4-J) + \sum_{j=1}^{J-2} u_j, (\beta_k - u_1), \dots, (\beta_k - u_{J-2}) \right) \\ &= (\beta_{k1}, \dots, \beta_{k(J-1)}) \end{aligned} \quad (2.31)$$

ความน่าจะเป็นในการยอมรับสำหรับการสลับตัวแบบ $PO \rightarrow NPO$ คือ

$$\begin{aligned} \alpha = \min & \left[1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{\prod_{j=1}^{J-1} f(\beta'_{kj})}{f(\beta_k^{(i)})} \times \frac{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta'_j, \beta'_{j-1})}{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta_j^{(i)}, \beta'_{j-1})} \right. \\ & \left. \times \frac{1}{q_u(u_1, \dots, u_{J-2})} \times (J-1) \left(\frac{J-1}{2} \right)^{J-2} \right] \end{aligned} \quad (2.32)$$

โดยในเทอมสุดท้ายคือค่าสมบูรณ์ของดีเทอร์มิแนนของเมทริกซ์จาโคเบียน เช่นเดียวกับความน่าจะเป็นในการยอมรับในการสลับตัวแบบ $NPO \rightarrow PO$ คือ

$$\begin{aligned} \alpha = \min & \left[1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{f(\beta'_k)}{\prod_{j=1}^{J-1} f(\beta'_{kj})} \times \frac{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta'_j, \beta'_{j-1})}{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta_j^{(i)}, \beta'_{j-1})} \right. \\ & \left. \times \frac{q_u(u_1, \dots, u_{J-2})}{1} \times \left(\frac{1}{J-1} \right) \left(\frac{2}{J-1} \right)^{J-2} \right] \end{aligned} \quad (2.33)$$

สำหรับการปรับค่าจุดตัด $\theta^{(i)}$ มีวิธีการเช่นเดียวกับที่อธิบายไว้ในส่วนของการสร้างตัวแบบ PPO ที่ 2.1.3.3

(3) การปรับ (Tuning)

ในบางกรณีของการคัดเลือกตัวแปรด้วยวิธีที่กล่าวข้างต้น อาจไม่สามารถประมาณค่าพารามิเตอร์การถดถอยรวมถึงตัวบ่งชี้ของพารามิเตอร์การถดถอยนั้นๆ ได้ ซึ่งปัญหาอาจเกิดจากฟังก์ชันภาวะน่าจะเป็นไม่มีการเปลี่ยนเมื่อมีการตัดตัวแปรออกบางตัว และเมื่อพารามิเตอร์ของตัวแปรที่อยู่ในตัวแบบมีค่าเข้าใกล้ศูนย์ วิธีการหนึ่งที่ใช้แก้ปัญหานี้ ทำได้โดยใช้ hyperprior

ของความแปรปรวนสำหรับ การแจกแจงก่อนของพารามิเตอร์การถดถอย β ซึ่งสามารถทำได้หลายวิธี แต่สำหรับความสัมพันธ์แบบ PO ในที่นี่จะกำหนดให้

$$\beta_k \sim N(0, \sigma_{k\beta}^2) \text{ โดยที่ } \sigma_{k\beta} \sim U(0, \xi) \quad (2.34)$$

และ ξ คือค่าสูงสุดของการแจกแจงก่อนของ $\sigma_{k\beta}$ และสำหรับความสัมพันธ์แบบ NPO กำหนดให้

$$\beta_{kj} \sim N(0, \sigma_{kj\beta}^2) \text{ ที่ } \sigma_{kj\beta} \sim U(0, \xi) \quad (2.35)$$

การทำเช่นนี้เป็นการเพิ่มความซับซ้อนของตัวแบบเนื่องจากมีพารามิเตอร์เพิ่มในระหว่างขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟกึ่งการโคดย้อนกลับได้ เช่น การย้าย $0 \rightarrow PO$ จะประกอบด้วยการย้ายจาก $0 \rightarrow (\beta_k, \sigma_{k\beta})$ ในทำนองเดียวกันการกระโดดจาก $PO \rightarrow NPO$ จะประกอบด้วยการย้ายจาก $(\beta_k, \sigma_{k\beta}) \rightarrow (\beta_{k1\beta}, \sigma_{k1\beta}, \dots, \beta_{k(J-1)\beta}, \sigma_{k(J-1)\beta})$ และอื่นๆ เมื่อต้องการทำเช่นนั้นการอัปเดตแต่ละพารามิเตอร์ของ β และพารามิเตอร์ σ_β ที่สอดคล้องกันจะทำพร้อมกันโดยใช้การแจกแจงที่เสนอที่เป็นอิสระกัน

ความซับซ้อนอีกอย่างหนึ่งนั่นคือส่วนเบี่ยงเบนมาตรฐานต้องมีค่าเป็นบวก ดังนั้นสำหรับการย้าย $0 \rightarrow PO$ หรือ $0 \rightarrow NPO$ (หรือการย้อนกลับ) ฟังก์ชันการย้ายยังคงใช้เช่น เดียวกันที่อธิบายในหัวข้อ (1) ยกเว้นตัวแปรตามมีสำหรับส่วนเบี่ยงเบนมาตรฐานที่เป็นอิสระกันและมีการแจกแจงเหมือนกันถูกสุ่มมาจากการแจกแจง $U(0, \xi)$ ความน่าจะเป็นที่ยอมรับได้จะถูกปรับดังนี้ สำหรับ $PO \rightarrow NPO$ ค่าของตัวแปร $u_1, \dots, u_{J-2}$ ที่เป็นอิสระกันและมีการแจกแจงเหมือนกันจะกำหนดให้ $U(\max[0, \sigma_k - \tau_\sigma], \min[\sigma_k + \tau_\sigma, \xi])$ เมื่อ $\tau_\sigma = 1$ และกำหนดฟังก์ชันของการย้ายเป็น

$$\begin{aligned} g_{PO \rightarrow NPO}(\sigma_k, u_1, \dots, u_{J-2}) &= (\sigma_k, \sigma_k + u_1, \dots, \sigma_k + u_{J-2}) \\ &= (\sigma_{k1}, \dots, \sigma_{k(J-1)}) \end{aligned} \quad (2.36)$$

การย้ายกลับไม่จำเป็นต้องเสนอค่าใหม่สามารถใช้ฟังก์ชันผกผัน ดังนี้

$$\begin{aligned} g_{NPO \rightarrow PO}(\sigma_{k1}, \dots, \sigma_{k(J-1)}) &= (\sigma_{k1}, \sigma_{k2} - \sigma_{k1}, \dots, \sigma_{k(J-1)} - \sigma_{k1}) \\ &= (\sigma_k, u_1, \dots, u_{J-2}) \end{aligned} \quad (2.37)$$

ความน่าจะเป็นของการยอมรับมีการปรับปรุงตาม ซึ่งการเพิ่มเทอมของส่วนเบี่ยงเบนมาตรฐานไม่ทำให้เทอมจาโคเบียนใน (2.32) หรือ (2.33) เปลี่ยนแปลง

2.1.5 รูปแบบเมทริกซ์ความสัมพันธ์ของตัวแปรอธิบาย

2.1.5.1 รูปแบบความสัมพันธ์แบบคงที่ (Constant correlation model)

รูปแบบความสัมพันธ์แบบคงที่ เป็นกรณีที่ความสัมพันธ์ของตัวแปรในแต่ละกลุ่มและระหว่างกลุ่มมีความสัมพันธ์คงที่ โดยรูปแบบของความสัมพันธ์แบบคงที่มีรูปแบบดังนี้

$$\Sigma_{k \times k} = \begin{bmatrix} 1 & \rho_k & \dots & \rho_k \\ \rho_k & 1 & \dots & \rho_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_k & \dots & 1 \end{bmatrix}$$

โดยที่ $\Sigma_{k \times k}$ คือ เมทริกซ์สหสัมพันธ์ขนาด $(k \times k)$

k คือ จำนวนกลุ่มการจำแนกหรือจำนวนตัวแปรอธิบาย $(k > 0)$

ρ_k คือ ความสัมพันธ์ขององค์ประกอบภายใน k กลุ่ม โดยมีค่าอยู่ระหว่าง $0 \leq \rho_k \leq 1$

ตัวอย่างรูปแบบความสัมพันธ์ที่เมื่อตัวแปรอธิบาย 5 ตัว และ $\rho_k = 0.50$ มีรูปแบบดังนี้

$$\Sigma = \begin{bmatrix} 1 & 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & 1 & 0.50 & 0.50 & 0.50 \\ 0.50 & 0.50 & 1 & 0.50 & 0.50 \\ 0.50 & 0.50 & 0.50 & 1 & 0.50 \\ 0.50 & 0.50 & 0.50 & 0.50 & 1 \end{bmatrix}$$

2.1.5.2 รูปแบบความสัมพันธ์แบบ Toeplitz correlation model

รูปแบบความสัมพันธ์แบบ Toeplitz เป็นตัวแบบที่มีความสัมพันธ์สูงสำหรับกลุ่มตัวแปรอธิบายที่อยู่ใกล้กันในเมทริกซ์ความสัมพันธ์ และตัวแบบจะมีค่าความสัมพันธ์ลดลงสำหรับกลุ่มตัวแปรอธิบายที่ห่างออกไป Guo et al. (2007) ได้อธิบายรูปแบบของความสัมพันธ์ของประชากรแบบ Toeplitz ดังนี้

$$\Sigma_{k \times k} = \begin{bmatrix} 1 & \rho_k & \rho_k^2 & \rho_k^3 & \dots & \rho_k^{g_k-1} \\ \rho_k & 1 & \rho_k & \rho_k^2 & \dots & \rho_k^{g_k-2} \\ \rho_k^2 & \rho_k & 1 & \rho_k & \dots & \rho_k^{g_k-3} \\ \rho_k^3 & \rho_k^2 & \rho_k & 1 & \dots & \rho_k^{g_k-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_k^{g_k-1} & \rho_k^{g_k-2} & \rho_k^{g_k-3} & \rho_k^{g_k-4} & \dots & 1 \end{bmatrix}$$

โดยที่ $\Sigma_{k \times k}$ คือ เมทริกซ์สหสัมพันธ์ขนาด $(k \times k)$

k คือ จำนวนกลุ่มการจำแนกหรือจำนวนตัวแปรอธิบาย $(k > 0)$

g_k คือ ขนาดของจำนวนกลุ่มการจำแนกหรือขนาดตัวอย่าง $(g_k > 0)$

ρ_k คือ ความสัมพันธ์ขององค์ประกอบภายใน k กลุ่ม โดยมีค่าอยู่ระหว่าง $0 \leq \rho_k \leq 1$

ตัวอย่างมีจำนวนตัวแปรอธิบายทั้งหมด 5 ตัวแปร และกำหนดให้ $\rho_k = 0.5$

$$\Sigma = \begin{bmatrix} 1 & 0.5^{|1-2|} & 0.5^{|1-3|} & 0.5^{|1-4|} & 0.5^{|1-5|} \\ 0.5^{|2-1|} & 1 & 0.5^{|2-3|} & 0.5^{|2-4|} & 0.5^{|2-5|} \\ 0.5^{|3-1|} & 0.5^{|3-2|} & 1 & 0.5^{|3-4|} & 0.5^{|3-5|} \\ 0.5^{|4-1|} & 0.5^{|4-2|} & 0.5^{|4-3|} & 1 & 0.5^{|4-5|} \\ 0.5^{|5-1|} & 0.5^{|5-2|} & 0.5^{|5-3|} & 0.5^{|5-4|} & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1 & 0.5 & 0.025 & 0.125 & 0.0625 \\ 0.5 & 1 & 0.5 & 0.025 & 0.125 \\ 0.025 & 0.5 & 1 & 0.5 & 0.025 \\ 0.125 & 0.025 & 0.5 & 1 & 0.5 \\ 0.0625 & 0.125 & 0.025 & 0.5 & 1 \end{bmatrix}$$

รูปแบบความสัมพันธ์ Toeplitz นี้ได้ถูกนำไปใช้ในตัวแบบอนุกรมเวลา (Time Series) ที่มีการจำลองเมทริกซ์ความสัมพันธ์รูปแบบ Toeplitz ถูกใช้อย่างแพร่หลายในการวิเคราะห์การจำแนกประเภท (Discriminant Analysis) ความสัมพันธ์ในแต่ละกลุ่ม (Group Correlation) ความสัมพันธ์ระหว่างค่าสังเกตตัวที่ l และ m คือ $|l-m|$ จะลดลงอย่างเอกซโพเนนเชียล (Exponential) นั่นคือ $\rho_k^{|l-m|}$

2.1.5.3 รูปแบบความสัมพันธ์แบบ Hub Toeplitz correlation mode

รูปแบบความสัมพันธ์แบบ Hub Toeplitz จะสมมติว่าเราทราบความสัมพันธ์ระหว่างจุดศูนย์รวมค่าสังเกต (Hub Observation) ซึ่งโดยปกติจะเป็นค่าสังเกตแรก และแต่ละค่าสังเกตอื่น ๆ ซึ่งความสัมพันธ์ระหว่างค่าสังเกตที่ 1 กับค่าสังเกตที่ m จะลดลงตามลำดับที่ m ซึ่ง ($m=2,3,\dots,g$)

เราสามารถอธิบายรูปแบบความสัมพันธ์ ได้ดังนี้สมมติว่าแถวแรกและหลักแรกของเมทริกซ์สหสัมพันธ์ A มีขนาด $g \times g$ กำหนดค่าได้ดังนี้

$$A_{11}=1, \quad A_{1m} = \rho_{\max} - (\rho_{\max} - \rho_{\min}) \left(\frac{m-2}{g-2} \right)^{\gamma}$$

ซึ่งค่าความสัมพันธ์จะลดลงจาก $A_{12} = \rho_{\max}$ จนถึง $A_{1g} = \rho_{\min}$ สำหรับ $2 \leq m \leq g$ เมื่อตัวแปรอธิบายอยู่ในรูป $\gamma=1$ ค่าสังเกตจะเป็นเส้นตรง และนำมาใช้อธิบายได้อย่างง่าย นอกจากนั้นแทนที่จะระบุ $\rho_{\max}$ และ $\rho_{\min}$ เราจะระบุเฉพาะ $\rho_{\max}$ และแทน $\tau = \frac{(\rho_{\max} - \rho_{\min})}{g-2}$ รูปแบบ

ความสัมพันธ์แบบ Hub Toeplitz เป็นดังนี้

$$\Sigma_{k \times k} = \begin{bmatrix} 1 & \alpha_{k,2} & \alpha_{k,3} & \alpha_{k,4} & \dots & \alpha_{k,g_k} \\ \alpha_{k,2} & 1 & \alpha_{k,2} & \alpha_{k,3} & \dots & \alpha_{k,g_k-1} \\ \alpha_{k,3} & \alpha_{k,2} & 1 & \alpha_{k,2} & \dots & \alpha_{k,g_k-2} \\ \alpha_{k,4} & \alpha_{k,3} & \alpha_{k,2} & 1 & \dots & \alpha_{k,g_k-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{k,g_k-1} & \alpha_{k,g_k-1} & \alpha_{k,g_k-2} & \alpha_{k,g_k-3} & \dots & 1 \end{bmatrix}$$

โดยที่ $\Sigma_{k \times k}$ คือ เมทริกซ์สหสัมพันธ์ขนาด $(k \times k)$

k คือ จำนวนกลุ่มการจำแนกหรือจำนวนตัวแปรอธิบาย ($k > 0$)

g_k คือ ขนาดของจำนวนกลุ่มการจำแนกหรือขนาดตัวอย่าง ($g_k > 0$)

ρ_k คือ ความสัมพันธ์ขององค์ประกอบภายใน k กลุ่ม โดยมีค่าอยู่ระหว่าง $0 \leq \rho_k \leq 1$

$$\alpha_{k,1} = 1 \quad \text{และ} \quad \alpha_{k,m} = \rho_{\max} - \tau_k (m-2) \quad \text{เมื่อ} \quad \tau_k = \frac{(\rho_{\max} - \rho_{\min})}{g_k - 2}$$

ตัวอย่างมีจำนวนตัวแปรอธิบายทั้งหมด 5 ตัวแปร และกำหนดให้ $\rho_{\max} = 0.95$ และ $\rho_{\min} = 0.5$

$$\text{นั่นคือ} \quad \tau_k = \frac{(\rho_{\max} - \rho_{\min})}{g_k - 2} = \left(\frac{0.95 - 0.5}{5 - 2} \right) = 0.15$$

$$\text{จะได้} \quad \alpha_{k,2} = 0.95$$

$$\alpha_{k,3} = \rho_{\max} - \tau_k (m-2) = 0.95 - 0.15(3-2) = 0.8$$

$$\alpha_{k,4} = 0.95 - 0.15(4-2) = 0.65$$

$$\text{และ} \quad \alpha_{k,5} = 0.5$$

$$\Sigma = \begin{bmatrix} 1 & 0.95 & 0.8 & 0.65 & 0.5 \\ 0.95 & 1 & 0.95 & 0.8 & 0.65 \\ 0.8 & 0.95 & 1 & 0.95 & 0.8 \\ 0.65 & 0.8 & 0.95 & 1 & 0.95 \\ 0.5 & 0.65 & 0.8 & 0.95 & 1 \end{bmatrix}$$

2.2 งานวิจัยที่เกี่ยวข้อง

Green (1995) เป็นคนแรก que คิดวิธีแก้การโคตย้อนกลับได้ ที่นำมาใช้ในวิธีมอนติคาร์โลโซมาร์คอฟ โดยได้ดัดแปลงและพัฒนาขึ้นจากการขยายขอบเขตของวิธีการ Metropolis-Hastings เพื่อให้สามารถนำมาประยุกต์ใช้กับปัญหาที่มีปริภูมิย่อยหลายอันซึ่งมีมิติแตกต่างกันมาก ๆ ดังนั้นจึง

จำเป็นต้องคิดการเคลื่อนย้ายระหว่างปริภูมีย่อยเพื่อให้สามารถเคลื่อนได้อย่างอิสระทั่วปริภูมิพารามิเตอร์ที่ถูกนำมารวมกัน ซึ่งทฤษฎีและการประยุกต์ใช้ที่นำเสนอในงานวิจัยนี้ได้แสดงให้เห็นว่าข้อดีของการคำนวณด้วยขั้นตอนวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้สามารถช่วยคัดเลือกว่าแปรในตัวแบบถดถอยและตัวแบบถดถอยที่ไม่ซับซ้อนได้ รวมทั้งการเลือกตัวแบบเบสระหว่างตัวแบบที่มีจำนวนพารามิเตอร์ต่างกัน

ต่อมาในปี 2001 Rasma Waagepetersen และ Deniel Sorensen ได้ให้รายละเอียดของวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้ที่สามารถเอาไปใช้ประยุกต์กับปัญหาได้หลากหลายมากกว่าที่ Green (1995) ได้เสนอไว้ เช่นในตัวแบบผสม (mixture model) ที่ไม่ทราบจำนวนองค์ประกอบของส่วนผสม การปรับให้เรียบของสถิติที่ไม่ใช่พารามิเตอร์แบบเบส ตัวแบบการถดถอยเชิงเส้นที่มีจำนวนตัวแปรอธิบายไม่คงที่โดยได้ใช้ตัวอย่างเกี่ยวกับตำแหน่งที่แสดงปริมาณลักษณะการสืบสายพันธุ์ (Quantitative Trait Loci: QTL) ที่สมมติว่าข้อมูลเป็นไปตัวแบบการถดถอยเชิงเส้นที่มีจำนวนตัวแปรอธิบายขึ้นอยู่กับจำนวนของ QTL ที่ไม่ทราบค่า

Brooks และ คณะ (2003) ได้เน้นการใช้วิธีการมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้ในการเลือกการแจกแจงเป้าหมายที่กำหนดไว้ล่วงหน้าให้มีประสิทธิภาพโดยไม่คำนึงถึงลักษณะของปัญหา ซึ่งมี 2 วิธีการคือ วิธีการแรกจะอาศัยการวิเคราะห์ความน่าจะเป็นของการยอมรับสำหรับวิธีการก้ำวกระโดดซึ่งวิธีการนี้จะเกี่ยวข้องกับการกระจายอนุกรมเทเลอร์ของความน่าจะเป็นของการยอมรับรอบๆ ก้ำวกระโดดได้ที่เป็นที่ยอมรับ (canonical jumps) วิธีที่สองจะขยายวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้โดยการใช้แนวคิดปริภูมิเต็มสมบูรณ์ (saturated space approach) นั่นคือการย้ายจากตัวแบบที่เล็กไปยังตัวแบบที่ใหญ่ขึ้น จะใช้ข้อมูลล่าสุดของการย้ายผันกลับ โดยได้ประยุกต์วิธีการเหล่านี้กับการวิเคราะห์อนุกรมเวลา (autoregressive time series) การสร้างตัวแบบเชิงกราฟ (graphical modeling) และการสร้างตัวแบบผสม (mixture modeling)

Hastie และ Green (2011) ได้ทบทวนวรรณกรรมที่เกี่ยวข้องกับวิธีมอนติคาร์โลโซมาร์คอฟก้ำวกระโดดย้อนกลับได้ที่ได้ตีพิมพ์ในฐานข้อมูล ISI (ISI Web of Knowledge) ในช่วงเวลาที่เขียนบทความ พบว่ามีบทความมากกว่า 500 บทความและได้รับการอ้างอิงมากกว่า 1400 ครั้ง โดยพบว่าส่วนใหญ่จะเกี่ยวข้องกับการนำไปใช้งานของวิธีก้ำวกระโดด โดยจะนำเสนอหลักการเชิงสถิติของวิธีการและการนำไปใช้ในการวิเคราะห์ข้อมูลชุดหนึ่งๆ ซึ่งการนำไปใช้นั้นจะมีความหลากหลายในด้านต่างๆ โดยถ้าใช้ในด้านสถิติและความน่าจะเป็นพบว่ามีบทความของด้านนี้ประมาณ ร้อยละ 45 ที่มีการอ้างอิงผลงานของ Green (1995) ร้อยละ 28 ที่เป็นผลงานทางด้านชีววิทยา พันธุกรรมและทางการแพทย์ ร้อยละ 20 ทางด้านคอมพิวเตอร์และวิศวกรรม และ ร้อยละ 15 เป็นด้านอื่นๆตั้งแต่วิทยาการจัดการไปจนถึงงานวิจัยด้านทรัพยากรน้ำ

McKinley และ คณะ (2015) ได้ใช้วิธีการมอนติคาร์โลโซมาร์คอฟกึ่งวาระโดยย้อนกลับได้ใน การเลือกระหว่างตัวแบบออดส์เชิงสัดส่วน (PO) และตัวแบบออดส์ไม่เชิงสัดส่วน (NPO) สำหรับตัว แปรอธิบายแต่ละตัว นอกจากนั้นยังได้แสดงวิธีการในการคัดเลือกตัวแปรโดยใช้วิธีกึ่งวาระโดยใน ตัวแบบการถดถอยแบบอันดับที่ใช้ฟังก์ชันเชื่อมโยงสะสม แต่อย่างไรก็ตามวิธีการที่เสนอนี้สามารถใช้ได้กับฟังก์ชันเชื่อมโยงอื่นๆที่มีลักษณะเพิ่มขึ้นทิศทางเดียว และได้ประยุกต์ใช้วิธีการนี้กับข้อมูลที่ เก็บตามช่วงเวลาหลายปี (longitudinal data) ของสุนัขในเมือง Zenzele ของแอฟริกาใต้ โดย ต้องการศึกษาปัจจัยทางด้านภูมิคุ้มกันและด้านประชากร ที่มีผลต่อความคุ้มครองของวัคซีนโรคพิษ สุนัขบ้าที่วัดออกมาเป็นคะแนนตามสภาพของร่างกาย ผลการศึกษาพบว่าเมื่อพิจารณาค่า PPA ปัจจัย ของเพศมีความสัมพันธ์กับคะแนนตามสภาพของร่างกายค่อนข้างน้อย ส่วนปัจจัยของสุนัขที่มีลักษณะ มีครรภ์และให้น้ำนม, อายุ, การกักขัง, เจ้าของรายงานอาการก่อนในช่วง 7 วัน และการสังเกตอาการ ด้วยผู้เชี่ยวชาญมีความสัมพันธ์กับคะแนนตามสภาพของร่างกายแน่นอน และสามารถระบุลักษณะ ของตัวแปรที่ถูกคัดเลือกเข้าในตัวแบบได้ดังนี้ ปัจจัยเพศ, ลักษณะมีครรภ์และให้น้ำนม, อายุและการ สังเกตอาการด้วยผู้เชี่ยวชาญมีลักษณะเป็น PO ส่วนปัจจัยเจ้าของรายงานอาการก่อนในช่วง 7 วัน มี ลักษณะเป็น NPO

บทที่ 3

วิธีการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาสมรรถนะของวิธีมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้ในการคัดเลือกตัวแปรและตัวแบบในการวิเคราะห์การถดถอยเชิงอันดับในกรณีที่ตัวแปรอธิบายมีความสัมพันธ์แบบคงที่, Toeplitz model และ Hub Toeplitz model

3.1 วิธีการมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้

การคัดเลือกตัวแปรและตัวแบบ PO, NPO และ PPO โดยวิธีมอนติคาร์โลโซ่มาร์คอฟก้ำวกระโดดย้อนกลับได้ สามารถเรียกใช้คำสั่ง BayesOrd จากโปรแกรม R ซึ่งได้ถูกพัฒนาโดย Trevelyan J. McKinley, Michelle Morters, and James L. N. Wood (2015) โดยสามารถใช้ชุดคำสั่งจาก <https://github.com/tjmckinley/BayesOrd>

3.2 การดำเนินการวิจัย

ทำการจำลองข้อมูลโดยใช้โปรแกรม R เวอร์ชัน 3.4.3

1. สร้างเมทริกซ์ตัวแปรอธิบายขนาด $n \times k$ โดยสุ่มจากการแจกแจงปรกติหลายตัวแปร (Multivariate Normal Distribution) ที่มีเวกเตอร์ค่าเฉลี่ย 0 และมีเมทริกซ์ความแปรปรวนและความแปรปรวนรวมเป็น Σ ดังนั้น $X \sim N(0, \Sigma)$ โดยที่เมทริกซ์ความแปรปรวนมีรูปแบบความสัมพันธ์ที่แตกต่างกันโดยแบ่งเป็นสถานการณ์ในการจำลองข้อมูลจาก Σ ได้ดังนี้

1.1 รูปแบบความสัมพันธ์แบบคงที่ (Constant correlation) ใน Σ จะกำหนดให้

$$\rho_k = r$$

1.2 รูปแบบความสัมพันธ์แบบ Toeplitz ใน Σ จะกำหนดให้ $\rho_k = r^{|l-m|}$

1.3 รูปแบบความสัมพันธ์แบบ Hub Toeplitz เมื่อ $\alpha_{k,m}$ ใน Σ คำนวณได้จาก

$$\alpha_{k,m} = \rho_{\max} - \tau_k (m-2) \quad \text{เมื่อ} \quad \tau_k = \frac{(\rho_{\max} - \rho_{\min})}{g_k - 2}$$

2. สร้างพารามิเตอร์ถดถอย β_{kj} , $k=1,2,\dots,K$ และ $j=1,2,\dots,J-1$ ในงานวิจัยนี้ศึกษาตัวแบบการถดถอยเชิงอันดับที่มีตัวแปรตอบสนอง 3 กลุ่ม ดังนั้น $j=1,2$ ซึ่งพารามิเตอร์ถดถอย

β_{kj} สุ่มจากการแจกแจงปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวน σ^2 นั่นคือ $\beta_{kj} \sim N(\mu, \sigma^2)$ ในกรณีนี้จะทำการสุ่มพารามิเตอร์ μ และ σ^2 เพื่อให้ได้ตัวแปรตอบสนองทั้ง 3 กลุ่มโดยแต่ละกลุ่มต้องมีจำนวนอย่างน้อย 5 % หากไม่เป็นตามเงื่อนไขจะทำการสุ่มใหม่ และในตัวแบบ PO นั้นจะกำหนดให้ $\beta_{kj} = \beta_{k1}$ สำหรับทุกกลุ่มของ j

3. สร้างพารามิเตอร์จุดตัด θ_j , $j=1,2$ โดยกำหนดให้พารามิเตอร์ θ_1 มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ 0.1^2 นั่นคือ $\theta_1 \sim N(\mu, 0.1^2)$ ซึ่งในแต่ละตัวแบบมีการกำหนดพารามิเตอร์ θ_2 แตกต่างกันดังนี้

3.1 ในตัวแบบ PO พารามิเตอร์ θ_2 จะขึ้นอยู่กับ θ_1 โดยมีเงื่อนไขคือ $\theta_1 < \theta_2$ นั่นคือ $\theta_2 \sim N(0, 0.1^2) + N(\mu, \sigma^2)$

3.2 ในตัวแบบ NPO พารามิเตอร์ θ_2 ถูกกำหนดตามตัวแบบ PO ในข้อ 3.1 และมีการตรวจสอบเงื่อนไขดังนี้

$$\theta_j - \theta_{j+1} < \sum_{k=1}^K \min(\beta_{kj} - \beta_{k(j+1)}) X_k^m, [\beta_{kj} - \beta_{k(j+1)}] X_k^M$$

โดย X_k^m และ X_k^M คือค่าต่ำสุดและค่าสูงสุดของ X_k ตามลำดับ

4. สร้างตัวแบบถดถอยเชิงอันดับตามสมการ เนื่องจากมีตัวแปรอธิบาย 3 ตัว จึงมีทั้งหมด 2 ลอจิต ดังนี้

4.1 ตัวแบบ PO

$$\text{logit 1 : } \log\left(\frac{\gamma_{i1}}{1-\gamma_{i1}}\right) = \theta_1 - (\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$$

$$\text{logit 2 : } \log\left(\frac{\gamma_{i2}}{1-\gamma_{i2}}\right) = \theta_2 - (\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$$

4.2 ตัวแบบ NPO

$$\text{logit 1 : } \log\left(\frac{\gamma_{i1}}{1-\gamma_{i1}}\right) = \theta_1 - (\beta_{11} X_{i1} + \beta_{21} X_{i2} + \beta_{31} X_{i3})$$

$$\text{logit 2 : } \log\left(\frac{\gamma_{i2}}{1-\gamma_{i2}}\right) = \theta_2 - (\beta_{12} X_{i1} + \beta_{22} X_{i2} + \beta_{32} X_{i3})$$

4.3 ตัวแบบ PPO

สมมติมีตัวแปรอธิบาย 3 ตัว โดยมีตัวแปรที่อยู่ในตัวแบบ PO จำนวน 2 ตัว และตัวแบบ NPO จำนวน 1 ตัว จะได้ลอจิสติกดังนี้

$$\text{logit 1 : } \log\left(\frac{\gamma_{i1}}{1-\gamma_{i1}}\right) = \theta_1 - (\beta_1 X_{i1} + \beta_2 X_{i2}) - (\beta_{31} X_{i3})$$

$$\text{logit 2 : } \log\left(\frac{\gamma_{i2}}{1-\gamma_{i2}}\right) = \theta_2 - (\beta_1 X_{i1} + \beta_2 X_{i2}) - (\beta_{32} X_{i3})$$

5. คำนวณความน่าจะเป็นของแต่ละกลุ่มจากความน่าจะเป็นสะสม

$$p_{ij} = \begin{cases} \gamma_{ij} & \text{for } j = 1, \\ \gamma_{ij} - \gamma_{i(j-1)} & \text{for } j = 2, \\ 1 - \gamma_{i(j-1)} & \text{for } j = 3 \end{cases}$$

ตัวอย่างความน่าจะเป็นของตัวแบบ PO เป็นดังนี้

$$p_{i1} = \frac{\exp(\theta_1 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}{1 + \exp(\theta_1 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}$$

$$p_{i2} = \frac{\exp(\theta_2 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}{1 + \exp(\theta_2 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))} - \frac{\exp(\theta_1 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}{1 + \exp(\theta_1 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}$$

$$p_{i3} = 1 - \frac{\exp(\theta_2 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}{1 + \exp(\theta_2 - (\beta_1 X_{i1} + \dots + \beta_K X_{iK}))}$$

6. สร้างตัวแปร Y_i โดยกำหนด u ดังนั้น $u \sim U(0,1)$ เป็นจุดตัด (cut-point) ในการแบ่งกลุ่มตัวแปรอธิบาย โดย $Y = 1$ ด้วยความน่าจะเป็น p_{i1} , $Y = 2$ ด้วยความน่าจะเป็น p_{i2} และ $Y = 3$ ด้วยความน่าจะเป็น p_{i3} ซึ่งจะเปรียบเทียบได้ดังนี้

ถ้า $u < p_{i1}$ แล้ว $Y = 1$

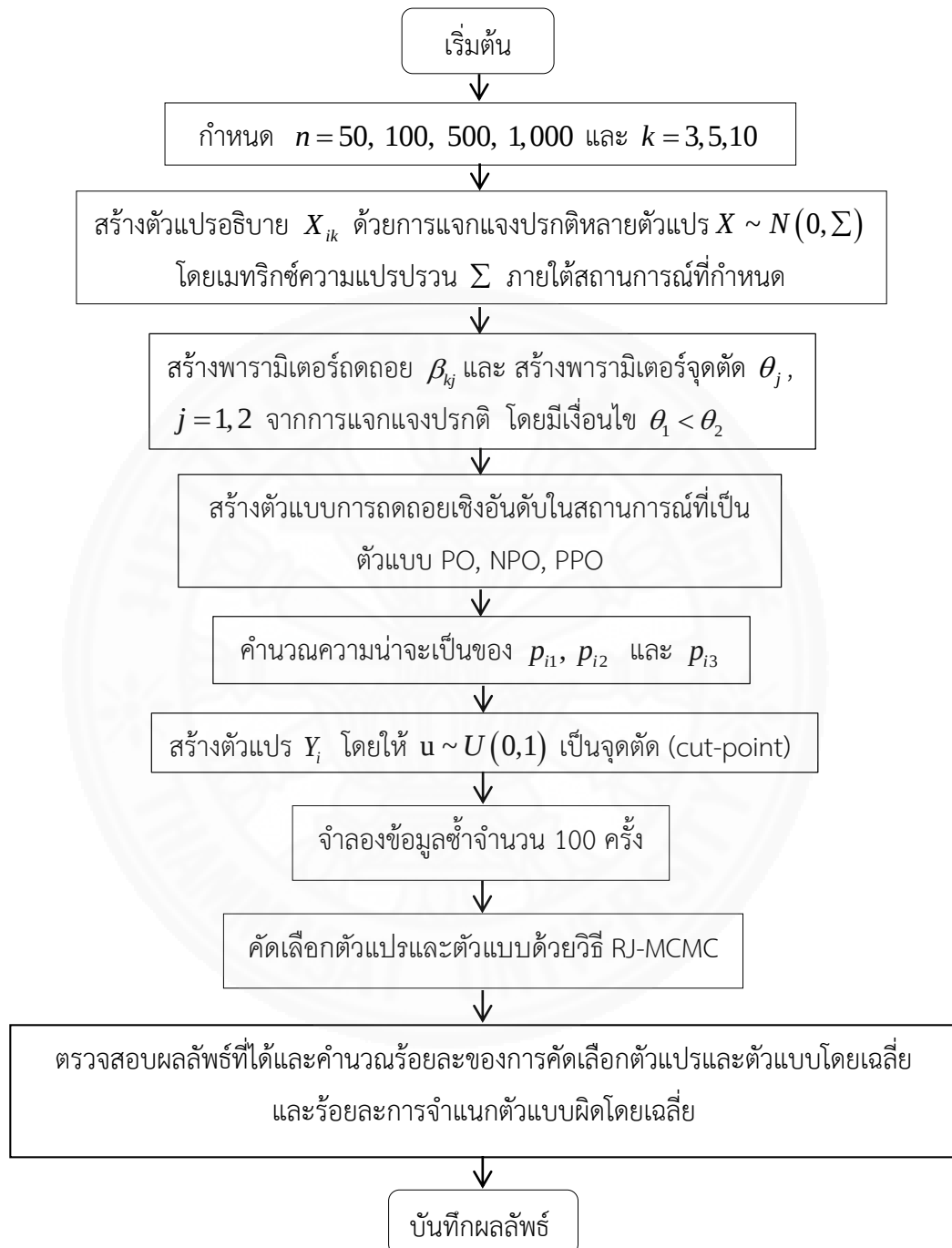
ถ้า $p_{i1} \leq u < p_{i1} + p_{i2}$ แล้ว $Y = 2$

ถ้า $u \geq p_{i1} + p_{i2}$ แล้ว $Y = 3$

7. เมื่อได้ค่าสังเกต Y_i, X_i แล้วทำซ้ำในขั้นตอนที่ 6 จำนวน 100 ครั้ง ใช้วิธีมอนติคาร์โลโซ่มาร์คอฟกึ่งการกระโดดย้อนกลับได้ โดยใช้ขั้นตอนวิธีตาม Trevelyan J. McKinley et al. (2015) ใช้คำสั่ง BayesOrd จาก <https://github.com/tjmckinley/BayesOrd> (รายละเอียดขั้นตอนตามบทที่ 2)

8. การคำนวณร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการแนกตัวแบบผิดโดยเฉลี่ยโดยการพิจารณาจากค่า PPA

3.3 แผนการจำลองข้อมูล



ภาพที่ 3.1 แผนผังขั้นตอนในการดำเนินการวิจัย

บทที่ 4

ผลการวิจัยและอภิปรายผล

ในงานวิจัยนี้เป็นการศึกษาสมรรถนะของวิธีมอนติคาร์โลโซมาร์คอฟกึ่งอิสระแบบย้อนกลับได้ในการคัดเลือกตัวแปรและตัวแบบการถดถอยเชิงอันดับกรณีที่ตัวแปรอธิบายมีความสัมพันธ์เชิงเส้นแบบพหุ ตัวแบบการถดถอยลอจิสติกที่ใช้ในการศึกษามี 3 ตัวแบบคือ ตัวแบบออกัสเชิงสัดส่วน (PO) ตัวแบบออกัสที่ไม่เป็นตามสัดส่วน (NPO) และตัวแบบออกัสเชิงสัดส่วนบางส่วน (PPO) ในแต่ละตัวแบบกำหนดตัวแปรอธิบายเป็น 3, 5 และ 10 ตัว โดยตัวแปรอธิบายมีความสัมพันธ์กันในรูปแบบความสัมพันธ์เป็นแบบอิสระ แบบคงที่ แบบ Toeplitz และแบบ Hub Toeplitz ที่กำหนดระดับสัมประสิทธิ์สหสัมพันธ์เป็น $r = 0.5, 0.8$ สำหรับรูปแบบความสัมพันธ์แบบคงที่และแบบ Toeplitz และสำหรับรูปแบบความสัมพันธ์แบบ Hub Toeplitz กำหนดสัมประสิทธิ์สหสัมพันธ์เป็น $r_{\min} = 0.3$ และ $r_{\max} = 0.8$

ในส่วนนี้จะนำเสนอผลการดำเนินการวิจัยด้วยร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย ซึ่งคำนวณได้จากค่า PPA ที่ได้ในแต่ละตัวแปรอธิบายที่ถูกคัดเลือกเข้าในตัวแบบ PO, NPO หรือถูกคัดออกจากตัวแบบ นอกจากนี้ยังเสนอร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของแต่ละกรณี

4.1 ผลการศึกษาศมรรถนะการคัดเลือกตัวแปรและตัวแบบ กรณีตัวแบบ PO

ตัวแบบ PO ในทุกกรณีศึกษากำหนดให้สุ่มตัวแปรอธิบายจำนวน 1 ตัว ให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ($\beta = 0$) ส่วนตัวแปรอื่น ๆ ที่เหลือมีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ PO ผลการศึกษาแสดงในตารางที่ 4.1 ที่จำแนกตามจำนวนตัวแปรอธิบาย ($k = 3, 5, 10$) และขนาดตัวอย่าง ($n = 50, 100, 500, 1,000$)

ตารางที่ 4.1 ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
50	อิสระกัน	r = 0	PO	94.78	0.00	5.22	4.48	95.31	0.00	4.69	3.75	-	-	-	-
			ไม่มีความสัมพันธ์	2.99	0.00	97.01		0.00	0.00	100.00		-	-	-	
	คงที่	r = 0.5	PO	94.93	0.00	5.07	4.35	93.23	0.00	6.77	5.83	-	-	-	-
			ไม่มีความสัมพันธ์	2.90	0.00	97.10		2.08	0.00	97.92		-	-	-	
		r = 0.8	PO	86.17	0.00	13.83	12.77	66.53	0.00	33.47	28.47	-	-	-	-
			ไม่มีความสัมพันธ์	10.64	0.00	89.36		8.47	0.00	91.53		-	-	-	
	Toeplitz	r = 0.5	PO	84.21	0.00	15.79	11.70	92.00	0.00	8.00	7.20	-	-	-	-
			ไม่มีความสัมพันธ์	3.51	0.00	96.49		4.00	0.00	96.00		-	-	-	
		r = 0.8	PO	62.26	0.00	37.74	26.42	78.57	0.00	21.43	20.00	-	-	-	-
			ไม่มีความสัมพันธ์	3.77	0.00	96.23		14.29	0.00	85.71		-	-	-	
	Hub Toeplitz	r _{min} = 0.3, r _{max} = 0.8	PO	64.58	0.00	35.42	23.61	84.72	0.69	14.58	15.56	-	-	-	-
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		16.67	0.00	83.33		-	-	-	

ตารางที่ 4.1 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
100	อิสระกัน	$r = 0$	PO	99.37	0.00	0.63	1.27	99.63	0.00	0.37	0.60	84.21	0.00	15.79	14.74
			ไม่มีความสัมพันธ์	2.53	0.00	97.47		1.49	0.00	98.51		5.26	0.00	94.74	
	คงที่	$r = 0.5$	PO	97.02	0.00	2.98	2.38	100.00	0.00	0.00	0.86	83.89	0.56	15.56	16.00
			ไม่มีความสัมพันธ์	1.19	0.00	98.81		4.29	0.00	95.71		15.00	0.00	85.00	
		$r = 0.8$	PO	82.95	0.00	17.05	14.39	96.99	0.60	2.41	4.82	76.09	0.00	23.91	22.12
			ไม่มีความสัมพันธ์	9.09	0.00	90.91		12.05	0.00	87.95		6.06	0.00	93.94	
	Toeplitz	$r = 0.5$	PO	94.83	1.15	4.02	5.36	99.65	0.35	0.00	0.83	68.52	0.00	31.48	30.00
			ไม่มีความสัมพันธ์	5.75	0.00	94.25		2.78	0.00	97.22		16.67	0.00	83.33	
		$r = 0.8$	PO	85.06	0.65	14.29	12.55	98.21	0.60	1.19	2.38	68.25	0.00	31.75	29.29
			ไม่มีความสัมพันธ์	7.79	0.00	92.21		4.76	0.00	95.24		7.14	0.00	92.86	
	Hub Toeplitz	$r_{\min} = 0.3, r_{\max} = 0.8$	PO	82.72	0.00	17.28	14.40	97.73	0.32	1.95	1.82	67.90	0.00	32.10	30.00
			ไม่มีความสัมพันธ์	8.64	0.00	91.36		0.00	0.00	100.00		11.11	0.00	88.89	

ตารางที่ 4.1 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
500	อิสระกัน	r = 0	PO	100.00	0.00	0.00	0.34	99.49	0.51	0.00	0.40	99.51	0.24	0.24	0.66
			ไม่มีความสัมพันธ์	0.00	1.01	98.99		0.00	0.00	100.00		2.20	0.00	97.80	
	คงที่	r = 0.5	PO	100.00	0.00	0.00	0.34	100.00	0.00	0.00	0.20	99.31	0.46	0.23	0.62
			ไม่มีความสัมพันธ์	1.02	0.00	98.98		1.00	0.00	99.00		0.00	0.00	100.00	
		r = 0.8	PO	100.00	0.00	0.00	0.67	100.00	0.00	0.00	0.20	96.54	0.12	3.35	3.12
			ไม่มีความสัมพันธ์	2.02	0.00	97.98		1.00	0.00	99.00		0.00	0.00	100.00	
	Toeplitz	r = 0.5	PO	100.00	0.00	0.00	0.34	99.75	0.25	0.00	0.20	99.52	0.12	0.36	0.65
			ไม่มีความสัมพันธ์	1.03	0.00	98.97		0.00	0.00	100.00		2.17	0.00	97.83	
		r = 0.8	PO	99.48	0.52	0.00	0.69	100.00	0.00	0.00	0.00	96.45	0.23	3.32	3.51
			ไม่มีความสัมพันธ์	0.00	1.04	98.96		0.00	0.00	100.00		3.09	0.00	96.91	
	Hub Toeplitz	r _{min} = 0.3, r _{max} = 0.8	PO	99.46	0.00	0.54	1.79	100.00	0.00	0.00	0.20	98.84	0.12	1.04	1.15
			ไม่มีความสัมพันธ์	4.30	0.00	95.70		1.00	0.00	99.00		1.04	0.00	98.96	

ตารางที่ 4.1 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
1000	อิสระกัน	r = 0	PO	100.00	0.00	0.00	0.33	99.00	1.00	0.00	0.40	100.00	0.00	0.00	0.20
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		0.25	0.00	99.75		1.00	1.00	98.00	
	คงที่	r = 0.5	PO	100.00	0.00	0.00	0.67	100.00	0.00	0.00	0.20	99.89	0.11	0.00	0.10
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		1.00	0.00	99.00		0.00	0.00	100.00	
		r = 0.8	PO	99.50	0.50	0.00	1.67	100.00	0.00	0.00	0.20	99.56	0.11	0.33	0.60
			ไม่มีความสัมพันธ์	4.00	0.00	96.00		1.00	0.00	99.00		2.00	0.00	98.00	
	Toeplitz	r = 0.5	PO	100.00	0.00	0.00	1.00	100.00	0.00	0.00	0.60	99.67	0.33	0.00	0.30
			ไม่มีความสัมพันธ์	3.00	0.00	97.00		3.00	0.00	97.00		0.00	0.00	100.00	
		r = 0.8	PO	100.00	0.00	0.00	0.67	100.00	0.00	0.00	0.20	99.56	0.11	0.33	0.60
			ไม่มีความสัมพันธ์	2.00	0.00	97.00		1.00	0.00	99.00		2.00	0.00	98.00	
	Hub Toeplitz	r _{min} = 0.3, r _{max} = 0.8	PO	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	99.78	0.22	0.00	0.30
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		0.00	0.00	100.00		1.00	0.00	99.00	

จากตารางที่ 4.1 เมื่อพิจารณาขนาดตัวอย่าง $n=50$ ในกรณีที่ตัวแปรอธิบาย $k=10$ วิธีมอนติคาร์โลโซมาร์คอฟกัวกระโดดย้อนกลับได้ไม่สามารถทำการคัดเลือกตัวแปรได้ และสำหรับกรณีที่ตัวแปรอธิบาย $k=3$ และ $k=5$ ผลลัพธ์ที่ได้มีแนวโน้มเช่นเดียวกันดังนี้

สำหรับตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ PO ส่วนใหญ่จะถูกคัดเลือกเข้าในตัวแบบ PO โดยมีร้อยละเกิน 90 เมื่อตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์เป็นแบบอิสระหรือมีความสัมพันธ์แบบคงที่ เมื่อ $r=0.5$ และจะลดลงเมื่อ $r=0.8$ และในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Toeplitz ร้อยละของการคัดเลือกเข้าในตัวแบบ PO เมื่อ $r=0.5$ สูงกว่าเมื่อ $r=0.8$ โดยรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Toeplitz ที่ $r=0.8$ ให้ค่าร้อยละของการคัดเลือกตัวแปรเข้าในตัวแบบ PO ต่ำที่สุด ซึ่งตัวแปรอธิบายเหล่านี้จะไม่ถูกคัดเลือกผิดเป็น NPO ในทุกกรณีของความสัมพันธ์ระหว่างตัวแปรอธิบายแต่จะถูกคัดเลือกผิดโดยการคัดออกจากสมการ

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกิน 80 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและร้อยละไม่เกิน 4 ที่คัดเลือกผิดโดยถูกคัดเลือกเข้าในตัวแบบ PO ยกเว้นกรณีที่ระดับความสัมพันธ์ $r=0.8$ และรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Hub Toeplitz ที่มีร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยเป็น 16.67

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยพบว่ามีความค่าค่อนข้างต่ำโดยเฉพาะในกรณีที่ตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์เป็นแบบอิสระหรือมีความสัมพันธ์ในระดับ $r=0.5$ และมีค่าเพิ่มขึ้นเมื่อ $r=0.8$ แต่เมื่อเปรียบเทียบระหว่างรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายจะเห็นว่าในกรณี $k=3$ รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Toeplitz ที่ $r=0.8$ ให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยสูงสุดเป็น 26.42 ในขณะที่กรณี $k=5$ รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่ ที่ $r=0.8$ ให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยสูงสุดเป็น 28.47

เมื่อพิจารณากรณีที่ขนาดตัวอย่างเพิ่มขึ้นเป็น $n=100,500$ และ $1,000$ พบว่า

สำหรับตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ PO ถูกคัดเลือกเข้าในตัวแบบ PO ร้อยละเกือบ 100 โดยเฉพาะถ้า $n=500$ และ $n=1,000$ ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและจำนวนตัวแปรอธิบายและมีร้อยละการคัดเลือกผิดที่ถูกเลือกเข้าไปในตัวแบบ NPO น้อยมากไม่เกินร้อยละ 1.15

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ส่วนใหญ่ถูกคัดออกจากตัวแบบเกินร้อยละ 83.33 เมื่อ $n=100$ และเกินร้อยละ 95.70 เมื่อ $n=500$ และ

$n=1,000$ ตัวแปรอธิบายจะถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO ไม่เกินร้อยละ 16.67 เมื่อ $n=100$ และเกินร้อยละ 4.30 เมื่อ $n=500$ และ $n=1,000$

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยพบว่ามีค่าน้อยมากในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและจำนวนตัวแปรอธิบายเมื่อ $n=500$ และ $n=1,000$ แต่เมื่อ $n=100$ และ $k=10$ ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยอยู่ระหว่างช่วงร้อยละ 14.74 - 30 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและจำนวนตัวแปรอธิบาย

4.2 ผลการศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบ กรณีตัวแบบ NPO

ตัวแบบ NPO ในทุกกรณีศึกษากำหนดให้สุ่มตัวแปรอธิบายจำนวน 1 ตัว ให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ($\beta=0$) ส่วนตัวแปรอธิบายอื่น ๆ ที่เหลือมีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ผลการศึกษาแสดงในตารางที่ 4.2 จำแนกตามจำนวนตัวแปรอธิบาย ($k=3, 5, 10$) และขนาดตัวอย่าง ($n=100, 500, 1,000$)

ตารางที่ 4.2 ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ NPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
100	อิสระกัน	r = 0	NPO	13.86	57.23	28.92	29.72	-	-	-	-	-	-	-	-
			ไม่มีความสัมพันธ์	2.41	1.20	96.39		-	-	-		-	-	-	
	คงที่	r = 0.5	NPO	16.43	56.43	27.14	30.48	-	-	-	-	-	-	-	-
			ไม่มีความสัมพันธ์	4.29	0.00	95.71		-	-	-		-	-	-	
		r = 0.8	NPO	21.51	54.07	24.42	32.95	-	-	-	-	-	-	-	-
			ไม่มีความสัมพันธ์	4.65	2.33	93.02		-	-	-		-	-	-	
	Toeplitz	r = 0.5	NPO	17.47	54.22	28.31	32.13	-	-	-	-	-	-	-	-
			ไม่มีความสัมพันธ์	4.82	0.00	95.18		-	-	-		-	-	-	
		r = 0.8	NPO	16.15	50.00	33.85	35.38	-	-	-	-	-	-	-	-
			ไม่มีความสัมพันธ์	4.62	1.54	93.85		-	-	-		-	-	-	
	Hub Toeplitz	r _{min} = 0.3, r _{max} = 0.8	NPO	16.67	52.67	30.67	32.44	-	-	-	-	-	-	-	-
			ไม่มีความสัมพันธ์	2.67	0.00	97.33		-	-	-		-	-	-	

หมายเหตุ - คือ วิธีมอนติคาร์โลโซมาร์คอฟกัววาระโตดย้อนกลับได้ไม่สามารถคัดเลือกตัวแปรได้

ตารางที่ 4.2 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ NPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
500	อิสระกัน	r = 0	NPO	4.55	95.45	0.00	3.70	30.00	70.00	0.00	24.00	54.17	45.83	0.00	50.00
			ไม่มีความสัมพันธ์	2.02	0.00	97.98		0.00	0.00	100.00		0.00	12.50	87.50	
	คงที่	r = 0.5	NPO	12.24	87.76	0.00	8.16	28.75	71.25	0.00	25.00	53.80	45.61	0.58	48.95
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		5.00	5.00	90.00		0.00	0.00	100.00	
		r = 0.8	NPO	14.50	84.00	1.50	11.00	40.38	59.62	0.00	35.38	62.22	33.33	4.44	60.00
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		0.00	15.38	84.62		0.00	0.00	100.00	
	Toeplitz	r = 0.5	NPO	8.08	91.41	0.51	6.40	29.55	70.45	0.00	27.27	57.41	42.59	0.00	51.67
			ไม่มีความสัมพันธ์	2.02	0.00	97.98		0.00	18.18	81.82		0.00	0.00	100.00	
		r = 0.8	NPO	15.00	85.00	0.00	10.67	37.88	61.36	0.76	33.64	56.11	38.89	5.00	55.00
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		0.00	13.64	86.36		0.00	0.00	100.00	
	Hub Toeplitz	r _{min} = 0.3, r _{max} = 0.8	NPO	12.50	87.00	0.50	9.00	35.45	64.55	0.00	31.34	44.44	44.44	11.11	50.00
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		2.99	11.94	85.07		0.00	0.00	100.00	

ตารางที่ 4.2 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ NPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
1000	อิสระกัน	$r = 0$	NPO	0.50	99.50	0.00	0.67	17.01	82.99	0.00	15.00	48.41	51.59	0.00	44.46
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		1.39	5.56	93.06		3.57	5.36	91.07	
	คงที่	$r = 0.5$	NPO	0.50	99.50	0.00	0.33	20.18	79.82	0.00	19.28	47.67	52.33	0.00	43.06
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		0.00	15.66	84.34		0.00	1.61	98.39	
		$r = 0.8$	NPO	2.50	97.50	0.00	1.67	28.75	71.25	0.00	28.25	48.44	50.67	0.89	44.40
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		0.00	26.25	73.75		0.00	0.00	100.00	
	Toeplitz	$r = 0.5$	NPO	0.00	100.00	0.00	0.33	20.48	79.52	0.00	19.15	44.73	55.27	0.00	41.03
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		0.00	13.83	86.17		1.28	6.41	92.31	
		$r = 0.8$	NPO	1.50	98.50	0.00	1.33	26.18	73.82	0.00	25.41	48.48	50.00	1.52	45.00
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		0.00	22.35	77.65		0.00	0.00	100.00	
	Hub Toeplitz	$r_{\min} = 0.3,$ $r_{\max} = 0.8$	NPO	2.50	97.50	0.00	2.33	26.54	73.46	0.00	26.42	47.11	52.22	0.67	43.80
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		0.00	25.93	74.07		0.00	8.00	92.00	

จากตารางที่ 4.2 เมื่อพิจารณาขนาดตัวอย่าง $n=100$ ในกรณีที่ตัวแปรอธิบาย $k=5$ และ $k=10$ วิธีมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้ไม่สามารถทำการคัดเลือกตัวแปรได้ แต่สำหรับกรณีที่ตัวแปรอธิบาย $k=3$ มีผลลัพธ์ดังนี้ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO จะถูกคัดเลือกเข้าในตัวแบบ NPO โดยมีร้อยละในช่วง 50 - 57.23 เมื่อตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์เป็นแบบอิสระจะถูกคัดเลือกเข้าในตัวแบบ NPO มีร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยที่สูงที่สุดแต่รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Toeplitz ที่ $r=0.8$ จะถูกคัดเลือกเข้าในตัวแบบ NPO มีร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยที่ต่ำที่สุด และรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบคงที่และ Toeplitz มีร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยที่ถูกคัดเลือกเข้าในตัวแบบ NPO เมื่อ $r=0.5$ มีค่าสูงกว่า $r=0.8$ นอกจากนี้ตัวแปรอธิบายในตัวแบบ NPO จะถูกคัดเลือกผิดเป็นตัวแบบ PO และถูกคัดออกจากสมการด้วยร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยไม่เกิน 21.51 และ 33.85 ตามลำดับ

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกิน 93.02 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO และ NPO ด้วยร้อยละไม่เกิน 4.82 และ 2.33 ตามลำดับ ยกเว้นกรณีที่ระดับความสัมพันธ์ $r=0.5$ และรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Hub Toeplitz ที่มีร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยเป็นตัวแบบ NPO เท่ากับ 0

เมื่อพิจารณาร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยพบว่ามีค่าอยู่ในช่วงร้อยละ 29.72 - 35.38 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยมีค่าต่ำที่สุดในกรณีที่ตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์เป็นแบบอิสระหรือมีความสัมพันธ์แบบคงที่ เมื่อ $r=0.5$ เมื่อเปรียบเทียบระดับความสัมพันธ์ $r=0.5$ ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยมีค่าต่ำกว่า $r=0.8$

เมื่อพิจารณากรณีที่ขนาดตัวอย่างเพิ่มขึ้นเป็น $n=500$ ในกรณีที่ตัวแปรอธิบาย $k=3$ พบว่าสำหรับตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ร้อยละมากกว่า 84 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายแต่มีการคัดเลือกตัวแปรและตัวแบบผิดที่ถูกเลือกเข้าไปในตัวแบบ PO ไม่เกินร้อยละ 15 และถูกคัดออกจากสมการน้อยมากไม่เกินร้อยละ 1.5 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกือบ 100 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย แต่ถูกคัดเลือกตัวแปรผิดโดยถูกคัดเข้าในตัวแบบ PO ไม่เกินร้อยละ 2.02

ในกรณีที่ตัวแปรอธิบาย $k=5$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO มีร้อยละมากกว่า 70 ถูกคัดเลือกเข้าในตัวแบบ NPO เมื่อตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์เป็นแบบอิสระหรือมีความสัมพันธ์แบบคงที่หรือแบบ Toeplitz เมื่อ $r=0.5$ และจะลดลงเมื่อ $r=0.8$ มีร้อยละการคัดเลือกตัวแปรผิดที่ถูกเลือกเข้าไปในตัวแบบ PO ไม่เกินร้อยละ 40.38 และถูกคัดออกจากสมการประมาณร้อยละ 0 ยกเว้นในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Toeplitz เมื่อ $r=0.8$

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองมีร้อยละเกิน 81.82 ถูกคัดออกจากตัวแบบในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย และไม่เกินร้อยละ 5 และ 18.18 ที่คัดเลือกผิดโดยถูกคัดเลือกเข้าในตัวแบบ PO และ NPO ตามลำดับ

ในกรณีที่ตัวแปรอธิบาย $k=10$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ด้วยร้อยละในช่วง 33.33 - 45.83 และมีร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ยที่เข้าในตัวแบบ NPO จะมีค่าลดลงถ้าระดับความสัมพันธ์ระหว่างตัวแปรอธิบายมีค่าเพิ่มขึ้นในกรณีของรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบคงที่และ Toeplitz นอกจากนั้นตัวแปรอธิบายมีการคัดเลือกผิดที่ถูกเลือกเข้าไปในตัวแบบ PO ด้วยร้อยละที่มากด้วยร้อยละเกิน 44.44 และถูกคัดออกจากสมการไม่เกินร้อยละ 11.11 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบร้อยละ 100 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายยกเว้นกรณีที่ตัวแปรอธิบายมีรูปแบบความสัมพันธ์แบบอิสระกัน (ร้อยละ 87.5) และไม่ถูกคัดเลือกผิดเป็นตัวแบบ PO และ NPO ยกเว้นกรณีที่ตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์แบบอิสระจะถูกคัดเลือกตัวแปรผิดเป็นตัวแบบ NPO ร้อยละ 12.5

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยมีค่าค่อนข้างต่ำในกรณี $k=3$ (ในช่วงร้อยละ 3.7 - 11) และจะมีค่าสูงขึ้นเมื่อจำนวนตัวแปรเพิ่มขึ้น ในกรณี $k=5$ และ $k=10$ ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยมีค่าอยู่ในช่วง 24 - 35.38 และอยู่ในช่วง 50 - 60 ตามลำดับ เมื่อเปรียบเทียบระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย $r=0.5$ มีร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยมีค่าต่ำกว่า $r=0.8$ ในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่และ Toeplitz

เมื่อพิจารณากรณีที่ขนาดตัวอย่างเป็น $n=1,000$ ในกรณีที่ตัวแปรอธิบาย $k=3$ พบว่าสำหรับตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ร้อยละเกือบ 100 และร้อยละของการคัดเลือกตัวแปรเข้าในตัวแบบ NPO เมื่อ $r=0.5$ จะมีค่ามากกว่าเมื่อ $r=0.8$ แต่มีร้อยละการคัดเลือกตัวแปรและตัวแบบผิดที่ถูกเลือกเข้าไปในตัวแบบ PO ไม่เกินร้อยละ 2.5 และตัวแปรอธิบายไม่ถูกคัดออกจากสมการในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกือบ 100 และถูกคัดเลือกตัวแปรผิดโดยถูกคัดเลือกเข้าในตัวแบบ PO ไม่เกินร้อยละ 2 แต่ไม่ถูกตัดเลือกผิดเป็นตัวแบบ NPO ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

ในกรณีที่ตัวแปรอธิบาย $k=5$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ร้อยละในช่วง 71.25 - 82.99 และที่ระดับความสัมพันธ์ของตัวแปรอธิบาย $r=0.5$ มีร้อยละของการคัดเลือกตัวแปรเข้าในตัวแบบ NPO มีค่ามากกว่าเมื่อ $r=0.8$ เช่นเดียวกับ $k=3$ และมีร้อยละการคัดเลือกตัวแปรผิดที่ถูกเลือกเข้าไปในตัวแบบ PO ช่วงร้อยละ 17.1-28.75 แต่ตัวแปรอธิบายไม่ถูกคัดออกจากสมการในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองมีร้อยละเกิน 77.75 ถูกคัดออกจากตัวแบบในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย และไม่ถูกคัดเลือกผิดเป็นตัวแบบ PO ยกเว้นรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบอิสระ และถูกคัดเลือกผิดเป็นตัวแบบ NPO ไม่เกินร้อยละ 26.25

กรณีที่ตัวแปรอธิบาย $k=10$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ด้วยร้อยละในช่วง 50 - 55.27 และมีร้อยละของการคัดเลือกเข้าในตัวแบบ NPO เมื่อ $r=0.5$ มีค่ามากกว่าเมื่อ $r=0.8$ มีการคัดเลือกตัวแปรผิดที่ถูกเลือกเข้าไปในตัวแบบ PO ด้วยร้อยละที่มากมีร้อยละในช่วง 44.73 - 48.48 และถูกคัดออกจากสมการไม่เกินร้อยละ 1.52 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองมีร้อยละมากกว่า 91.07 ถูกคัดออกจากตัวแบบในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย เมื่อ $r=0.5$ มีร้อยละของการคัดเลือกตัวแปรออกจากตัวแบบมากกว่าเมื่อ $r=0.8$ และถูกคัดเลือกผิดเป็นตัวแบบ PO และ NPO ไม่เกินร้อยละ 3.57 และ 8 ตามลำดับ

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยจะมีค่าสูงขึ้นเมื่อมีจำนวนตัวแปรอธิบายมากขึ้นเช่น กรณี $k=3$ มีช่วงร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย $0.33 - 2.33$, กรณี $k=5$ มีช่วงร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย $15 - 28.25$ และกรณี $k=10$ มีช่วงร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย $41.03 - 45$ เมื่อเปรียบเทียบระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย $r=0.5$ มีร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยมีค่าต่ำกว่า $r=0.8$ และถ้าขนาดตัวอย่างเพิ่มขึ้นจะทำให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยจะยิ่งลดลงในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและจำนวนตัวแปรอธิบาย

4.3 ผลการศึกษาสมรรถนะการคัดเลือกตัวแปรและตัวแบบ กรณีตัวแบบ PPO

ตัวแบบ PPO ในทุกกรณีศึกษาทำการสุ่มตัวแปรอธิบายจำนวน 1 ตัว ให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ($\beta=0$) และกำหนดให้ตัวแปรอธิบายอื่น ๆ มีตัวแบบแยกตามจำนวนตัวแปรดังนี้ เมื่อมีตัวแปรอธิบาย 3 ตัว สุ่มตัวแปรอธิบายเป็นตัวแบบ PO และ NPO อย่างละ 1 ตัวแปร เมื่อมีตัวแปรอธิบาย 5 ตัว สุ่มตัวแปรอธิบายเป็นตัวแบบ PO และ NPO อย่างละ 2 ตัวแปร และกรณีตัวแปรอธิบาย 10 ตัวแปร สุ่มตัวแปรอธิบาย 5 ตัวเป็นตัวแบบ PO และตัวแปรอธิบายที่เหลือกำหนดเป็นตัวแบบ NPO ผลการศึกษาแสดงในตารางที่ 4.3 จำแนกตามจำนวนตัวแปรอธิบาย ($k=3, 5, 10$) และขนาดตัวอย่าง ($n=50, 100, 500, 1,000$)

ตารางที่ 4.3 ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย		
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก			
k=3						k=5						k=10					

50	อิสระกัน	$r = 0$	PO	84.85	0.00	15.15	7.07	50.00	0.00	50.00	53.00	-	-	-	-
			NPO	0.00	100.00	0.00		32.50	25.00	42.50		-	-	-	
			ไม่มีความสัมพันธ์	6.06	0.00	93.94		10.00	5.00	85.00		-	-	-	
	คงที่	$r = 0.5$	PO	80.49	2.44	17.07	9.76	41.30	4.35	54.35	59.13	-	-	-	-
			NPO	0.00	95.12	4.88		39.13	13.04	47.83		-	-	-	
			ไม่มีความสัมพันธ์	4.88	0.00	95.12		0.00	4.35	95.65		-	-	-	
		$r = 0.8$	PO	72.41	6.90	20.69	13.79	18.42	7.89	73.68	73.68	-	-	-	-
			NPO	6.90	86.21	6.90		21.05	5.26	73.68		-	-	-	
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		10.53	5.26	84.21		-	-	-	
	Toeplitz	$r = 0.5$	PO	65.91	0.00	34.09	14.39	42.86	4.76	52.38	57.14	-	-	-	-
			NPO	0.00	93.18	6.82		35.71	16.67	47.62		-	-	-	
			ไม่มีความสัมพันธ์	2.27	0.00	97.73		4.76	0.00	95.24		-	-	-	
		$r = 0.8$	PO	45.71	8.57	45.71	27.62	33.93	1.79	64.29	60.71	-	-	-	-
			NPO	0.00	74.29	25.71		35.71	19.64	44.64		-	-	-	
			ไม่มีความสัมพันธ์	2.86	0.00	97.14		7.14	3.57	89.29		-	-	-	
	Hub Toeplitz	$r_{\min} = 0.3,$ $r_{\max} = 0.8$	PO	31.58	2.63	65.79	35.09	24.24	1.52	74.24	65.45	-	-	-	-
			NPO	2.63	78.95	18.42		31.82	12.12	56.06		-	-	-	
			ไม่มีความสัมพันธ์	10.53	5.26	84.21		0.00	0.00	100.00		-	-	-	

หมายเหตุ - คือ วิธีมอนติคาร์โลโซมาร์คอฟกัววาระโตดย้อนกลับได้ไม่สามารถคัดเลือกตัวแปรได้

ตารางที่ 4.3 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

n	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย		
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก			
k=3						k=5						k=10					
100	อิสระกัน	$r = 0$	PO	97.30	0.00	2.70	1.35	50.00	1.02	48.98	42.04	-	-	-	-		
			NPO	0.00	100.00	0.00		43.88	44.90	11.22		-	-	-			
			ไม่มีความสัมพันธ์	1.35	0.00	98.65		0.00	0.00	100.00		-	-	-			
	คงที่	$r = 0.5$	PO	98.55	0.00	1.45	1.93	49.21	2.38	48.41	41.90	-	-	-	-		
			NPO	0.00	100.00	0.00		42.86	47.62	9.52		-	-	-			
			ไม่มีความสัมพันธ์	4.35	0.00	95.65		3.17	0.00	96.83		-	-	-			
		$r = 0.8$	PO	85.87	0.00	14.13	7.97	36.57	5.22	58.21	51.04	-	-	-	-		
			NPO	0.00	100.00	0.00		24.63	37.31	38.06		-	-	-			
			ไม่มีความสัมพันธ์	9.78	0.00	90.22		2.99	0.00	97.01		-	-	-			
	Toeplitz	$r = 0.5$	PO	100.00	0.00	0.00	0.00	50.00	0.85	49.15	42.37	-	-	-	-		
			NPO	0.00	100.00	0.00		33.90	45.76	20.34		-	-	-			
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		1.69	1.69	96.61		-	-	-			
		$r = 0.8$	PO	84.85	3.03	12.12	6.57	44.74	3.95	51.32	48.16	-	-	-	-		
			NPO	1.52	98.48	0.00		33.55	38.82	27.63		-	-	-			
			ไม่มีความสัมพันธ์	3.03	0.00	96.97		7.89	0.00	92.11		-	-	-			
	Hub Toeplitz	$r_{\min} = 0.3,$ $r_{\max} = 0.8$	PO	94.44	1.39	4.17	3.70	45.24	4.17	50.60	46.67	-	-	-	-		
			NPO	1.39	98.61	0.00		29.76	43.45	26.79		-	-	-			
			ไม่มีความสัมพันธ์	4.17	0.00	95.83		10.71	0.00	89.29		-	-	-			

หมายเหตุ - คือ วิธีมอนติคาร์โลโซมาร์คอฟกัววาระโตดย้อนกลับได้ไม่สามารถคัดเลือกตัวแปรได้

ตารางที่ 4.3 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

η	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย		
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก			
k=3						k=5						k=10					

500	อิสระกัน	$r = 0$	PO	100.00	0.00	0.00	0.67	54.50	2.50	43.00	31.20	98.57	0.95	0.48	13.33
			NPO	0.00	100.00	0.00		32.00	68.00	0.00		30.36	69.64	0.00	
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		1.00	0.00	99.00		2.38	2.38	95.24	
	คงที่	$r = 0.5$	PO	100.00	0.00	0.00	0.00	57.00	0.00	43.00	33.60	96.67	0.00	3.33	22.22
			NPO	0.00	100.00	0.00		41.00	59.00	0.00		50.00	50.00	0.00	
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		0.00	0.00	100.00		5.56	0.00	94.44	
		$r = 0.8$	PO	100.00	0.00	0.00	0.67	51.53	0.51	47.96	30.20	82.50	2.50	15.00	33.75
			NPO	0.00	100.00	0.00		27.04	72.96	0.00		57.81	40.63	1.56	
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		0.00	0.00	100.00		6.25	6.25	87.50	
	Toeplitz	$r = 0.5$	PO	100.00	0.00	0.00	0.33	54.50	0.50	45.00	33.60	100.00	0.00	0.00	10.77
			NPO	0.00	100.00	0.00		37.50	62.50	0.00		26.92	73.08	0.00	
			ไม่มีความสัมพันธ์	1.00	0.00	99.00		2.00	0.00	98.00		0.00	0.00	100.00	
		$r = 0.8$	PO	99.00	1.00	0.00	0.33	55.00	0.00	45.00	34.00	81.14	8.00	10.86	27.92
			NPO	0.00	100.00	0.00		38.00	61.50	0.50		45.00	54.29	1.43	
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		3.00	0.00	97.00		0.00	0.00	100.00	
	Hub Toeplitz	$r_{\min} = 0.3,$ $r_{\max} = 0.8$	PO	99.00	1.00	0.00	1.00	54.50	0.50	45.00	32.40	90.67	8.00	2.67	29.80
			NPO	0.00	100.00	0.00		34.50	65.50	0.00		61.67	38.33	0.00	
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		2.00	0.00	98.00		0.00	0.00	100.00	

ตารางที่ 4.3 (ต่อ) ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบ PPO เมื่อกำหนดตัวแปรอธิบาย 3, 5 และ 10 ตัว และขนาดตัวอย่าง 50,100,500 และ 1,000

η	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย	สัมประสิทธิ์สหสัมพันธ์ (r)	รูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย	ร้อยละการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย			ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย
				PO	NPO	คัดออก		PO	NPO	คัดออก		PO	NPO	คัดออก	
				k=3				k=5				k=10			
1000	อิสระกัน	$r = 0$	PO	100.00	0.00	0.00	0.67	59.50	0.00	40.50	20.60	95.62	4.38	0.00	8.88
			NPO	0.00	100.00	0.00		10.50	89.50	0.00		15.41	84.59	0.00	
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		1.00	0.00	99.00		0.00	1.37	98.63	
	คงที่	$r = 0.5$	PO	100.00	0.00	0.00	0.00	59.50	1.50	39.00	22.20	96.76	2.94	0.29	5.00
			NPO	0.00	100.00	0.00		14.00	86.00	0.00		8.46	91.54	0.00	
			ไม่มีความสัมพันธ์	0.00	0.00	100.00		2.00	0.00	98.00		0.00	0.00	100.00	
		$r = 0.8$	PO	100.00	0.00	0.00	0.67	56.00	1.00	43.00	28.40	92.73	1.82	5.45	23.18
			NPO	0.00	100.00	0.00		26.50	73.50	0.00		48.86	51.14	0.00	
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		1.00	0.00	99.00		0.00	0.00	100.00	
	Toeplitz	$r = 0.5$	PO	100.00	0.00	0.00	0.67	54.00	0.50	45.50	25.80	95.63	3.91	0.46	6.90
			NPO	0.00	100.00	0.00		18.50	81.50	0.00		11.49	88.51	0.00	
			ไม่มีความสัมพันธ์	2.00	0.00	98.00		0.00	0.00	100.00		1.15	0.00	98.85	
		$r = 0.8$	PO	100.00	0.00	0.00	0.34	58.00	0.00	42.00	28.40	93.33	6.15	0.51	16.15
			NPO	0.00	100.00	0.00		27.50	72.50	0.00		31.41	68.59	0.00	
			ไม่มีความสัมพันธ์	1.02	0.00	98.98		3.00	0.00	97.00		2.56	0.00	97.44	
	Hub Toeplitz	$r_{\min} = 0.3,$ $r_{\max} = 0.8$	PO	100.00	0.00	0.00	0.33	57.50	0.00	42.50	28.20	94.87	4.05	1.08	17.16
			NPO	0.00	100.00	0.00		27.50	72.50	0.00		35.47	64.53	0.00	
			ไม่มีความสัมพันธ์	0.00	1.00	99.00		1.00	0.00	99.00		4.05	0.00	95.95	

จากตารางที่ 4.3 เมื่อพิจารณาขนาดตัวอย่าง $n=50$ ในกรณีที่ตัวแปรอธิบาย $k=10$ วิธีมอนติคาร์โลโซ่มาร์คอฟก้าวกระโดดย้อนกลับได้ไม่สามารถทำการคัดเลือกตัวแปรได้ และสำหรับกรณีที่ตัวแปรอธิบาย $k=3$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ PO มีร้อยละเกิน 80 ที่ถูกคัดเลือกเข้าในตัวแบบ PO เมื่อตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบความสัมพันธ์แบบอิสระหรือมีความสัมพันธ์แบบคงที่ เมื่อ $r=0.5$ และมีร้อยละของการคัดเลือกตัวแปรเข้าในตัวแบบ PO เมื่อ $r=0.5$ สูงกว่า $r=0.8$ เมื่อมีรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบคงที่และ Toeplitz ส่วนในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Hub Toeplitz ให้ค่าร้อยละของการคัดเลือกเข้าในตัวแบบ PO ต่ำที่สุด นอกจากนั้นตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ NPO ไม่เกินร้อยละ 8.57 และถูกคัดเลือกผิดโดยการคัดออกจากสมการด้วยในช่วงร้อยละ 15.15 - 65.79 โดยรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Hub Toeplitz ถูกคัดออกจากสมการด้วยค่าร้อยละสูงสุด

ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ด้วยร้อยละเกิน 74.29 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย และระดับความสัมพันธ์ $r=0.5$ มีร้อยละการคัดเลือกตัวแปรเข้าในตัวแบบ NPO สูงกว่า $r=0.8$ โดยรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบอิสระให้ค่าร้อยละของการคัดเลือกตัวแปรเข้าในตัวแบบ NPO สูงที่สุด ตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ PO ไม่เกินร้อยละ 6.90 และถูกคัดเลือกผิดโดยการคัดออกจากสมการโดยร้อยละที่ต่ำ ยกเว้นรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Toeplitz เมื่อ $r=0.8$ ให้ค่าร้อยละสูงสุดร้อยละ 25.71

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกือบ 100 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย ยกเว้นของรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Hub Toeplitz และมีร้อยละไม่เกิน 10.53 ที่คัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO แต่ไม่ถูกคัดเลือกผิดเป็นตัวแบบ NPO ยกเว้นรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Hub Toeplitz ที่มีร้อยละการคัดเลือกผิดเป็น 5.26

กรณีที่ตัวแปรอธิบาย $k=5$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ PO ถูกคัดเลือกเข้าในตัวแบบ PO ค่อนข้างต่ำอยู่ในช่วงร้อยละ 18.42 - 50 โดยร้อยละที่มีระดับความสัมพันธ์ของตัวแปรอธิบาย $r=0.5$ มีการคัดเลือกเข้าในตัวแบบ PO ร้อยละสูงกว่า $r=0.8$ และเมื่อตัวแปรอธิบายมีรูปแบบความสัมพันธ์แบบอิสระให้ค่าร้อยละของการคัดเลือกตัวแปรเข้าในตัวแบบ PO สูงที่สุด ตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ NPO ไม่เกิน

ร้อยละ 7.89 แต่ตัวแปรอธิบายส่วนใหญ่จะถูกคัดเลือกผิดโดยการคัดออกจากสมการมีร้อยละช่วง 50 – 74.24

สำหรับตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ค่อนข้างต่ำเช่นเดียวกันโดยมีร้อยละอยู่ในช่วง 5.26-25 ตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ PO ในช่วงร้อยละ 21.05-39.13 แต่ตัวแปรอธิบายส่วนใหญ่จะถูกคัดเลือกผิดโดยการคัดออกจากสมการอยู่ในช่วงร้อยละ 42.5 – 73.68

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกิน 84.21 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO ร้อยละไม่เกิน 10.53 และร้อยละค่อนข้างต่ำที่ถูกคัดเลือกผิดเป็นตัวแบบ NPO

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยพบว่าเมื่อจำนวนตัวแปรอธิบายมากขึ้นร้อยละการจำแนกผิดโดยเฉลี่ยจะเพิ่มขึ้นตาม กรณี $k = 3$ มีร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยอยู่ในช่วง 7.07-35.09 โดยรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Hub Toeplitz ให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยสูงสุดและรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบอิสระให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยต่ำที่สุด และกรณี $k = 5$ มีร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยอยู่ในช่วง 53-73.8 โดยรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่ เมื่อ $r = 0.8$ ให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยสูงสุดและรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบอิสระให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยต่ำที่สุด และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยจะมีค่าสูงขึ้นเมื่อระดับสหสัมพันธ์ระหว่างตัวแปรอธิบายเพิ่มขึ้น

เมื่อพิจารณาขนาดตัวอย่างเพิ่มเป็น $n = 100$ ในกรณีที่ตัวแปรอธิบาย $k = 10$ วิธึมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้ไม่สามารถทำการคัดเลือกตัวแปรได้เช่นเดียวกับ $n = 50$ และผลการศึกษาเป็นดังนี้ สำหรับกรณีที่ตัวแปรอธิบาย $k = 3$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ PO ที่ถูกคัดเลือกเข้าในตัวแบบ PO มีร้อยละเกิน 84.85 ในทุกรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย $r = 0.5$ มีร้อยละการคัดเลือกเข้าในตัวแบบ PO สูงกว่า $r = 0.8$ ตัวแปรอธิบายเหล่านี้ไม่ถูกคัดเลือกผิดเป็น NPO ยกเว้นรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็น Teplitz เมื่อ $r = 0.8$ และ Hub Teplitz เพียงร้อยละน้อยไม่เกิน 3.03 และ 1.39 ตามลำดับ และตัวแปรอธิบายถูกคัดออกจากตัวแบบไม่เกินร้อยละ 14.13

ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ด้วยร้อยละเกือบ 100 แต่ถูกคัดเลือกผิดเป็นตัวแบบ PO ไม่เกินร้อยละ

1.52 และตัวแปรอธิบายจะไม่ถูกคัดออกจากสมการในทุกกรณีของความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกิน 90.22 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและร้อยละไม่เกิน 9.78 ที่ถูกคัดเลือกตัวแปรผิดโดยถูกคัดเข้าในตัวแบบ PO แต่ไม่ถูกคัดเลือกผิดเป็นตัวแบบ NPO

กรณีที่ตัวแปรอธิบาย $k = 5$ ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ PO ถูกคัดเลือกเข้าในตัวแบบ PO ก่อนข้างต่ำในช่วงร้อยละ 36.57 – 50 โดยร้อยละที่มีระดับความสัมพันธ์ของตัวแปรอธิบาย $r = 0.5$ มีการคัดเลือกเข้าในตัวแบบ PO ร้อยละสูงกว่า $r = 0.8$ และตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ NPO ไม่เกินร้อยละ 5.22 แต่ตัวแปรอธิบายถูกคัดเลือกผิดโดยการคัดออกจากสมการก่อนข้างสูงมีร้อยละช่วง 48.41 – 58.21

สำหรับตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กันตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO มีร้อยละอยู่ในช่วง 37.31-47.62 และร้อยละที่มีระดับความสัมพันธ์ของตัวแปรอธิบาย $r = 0.5$ มีการคัดเลือกเข้าในตัวแบบ NPO มีร้อยละสูงกว่า $r = 0.8$ แต่ตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ PO ในช่วงร้อยละ 24.63-43.88 และมีตัวแปรอธิบายส่วนใหญ่จะถูกคัดเลือกผิดโดยการคัดออกจากสมการอยู่ในช่วงร้อยละ 9.52-38.06

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะถูกคัดออกจากตัวแบบด้วยร้อยละเกิน 89.29 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย และตัวแปรอธิบายถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO ร้อยละไม่เกิน 10.71 แต่ไม่ถูกคัดเลือกผิดเป็นตัวแบบ NPO ยกเว้นรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายเป็นแบบ Toeplitz เมื่อ $r = 0.5$ ด้วยร้อยละค่อนข้างต่ำ (ร้อยละ 1.69)

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยพบว่าเมื่อจำนวนตัวแปรอธิบายมากขึ้นทำให้อัตราการจำแนกตัวแบบผิดโดยเฉลี่ยเพิ่มขึ้นและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย $r = 0.5$ มีร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยต่ำกว่า เมื่อ $r = 0.8$ ในทุกกรณีรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบาย

เมื่อพิจารณากรณีที่ขนาดตัวอย่างเพิ่มขึ้นเป็น $n = 500$ และ 1,000 ผลลัพธ์ที่ได้มีแนวโน้มเดียวกัน กรณี $k = 3$ เมื่อตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ PO ถูกคัดเลือกเข้าในตัวแบบ PO ร้อยละเกือบ 100 โดยเฉพาะถ้า $n = 1,000$ ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรอธิบายไม่ถูกคัดเลือกผิดเป็นตัว

แบบ NPO และไม่ถูกคัดออกจากสมการยกเว้นขนาดตัวอย่าง $n = 500$ มีร้อยละการจำแนกผิดที่ถูกเลือกเข้าไปในตัวแบบ NPO น้อยมากไม่เกินร้อยละ 1

ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ทั้งหมด (ร้อยละ 100) ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ส่วนใหญ่ถูกคัดออกจากตัวแบบด้วยร้อยละเกือบ 100 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและไม่เกินร้อยละ 2 ถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO

กรณี $k = 5$ เมื่อตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ PO ร้อยละเกิน 51.53 ถูกคัดเลือกเข้าในตัวแบบ PO ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ NPO ด้วยร้อยละค่อนข้างต่ำ (ช่วงร้อยละ 0-2.5) แต่มีร้อยละค่อนข้างสูง (ช่วงร้อยละ 39-47.96) ถูกคัดออกจากสมการ

ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ร้อยละเกิน 59 เมื่อ $n = 500$ และร้อยละเกิน 72.50 เมื่อ $n = 1,000$ ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย มีตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ PO อยู่ในช่วงร้อยละ 10.50 - 41 แต่ตัวแปรอธิบายไม่ถูกคัดออกจากสมการยกเว้นที่ $n = 500$ ในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบ Toeplitz เมื่อ $r = 0.8$ (ร้อยละ 0.5)

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ส่วนใหญ่ถูกคัดออกจากตัวแบบด้วยร้อยละเกือบ 100 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและไม่เกินร้อยละ 3 ถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO และไม่ตัวแปรอธิบายถูกคัดเลือกเข้าในตัวแบบ NPO

กรณี $k = 10$ เมื่อตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ PO มีร้อยละเกิน 81.14 เมื่อ $n = 500$ และร้อยละเกิน 92.73 เมื่อ $n = 1,000$ ถูกคัดเลือกเข้าในตัวแบบ PO ซึ่งที่ระดับความสัมพันธ์ของตัวแปรอธิบาย $r = 0.5$ มีร้อยละมากกว่า $r = 0.8$ ในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่และ Toeplitz และตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ NPO ด้วยร้อยละไม่เกิน 8 แต่มีตัวแปรถูกคัดออกจากสมการไม่เกินร้อยละ 15 เมื่อ $n = 500$ และไม่เกินร้อยละ 5.45 เมื่อ $n = 1,000$

ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์กับตัวแปรตอบสนองในตัวแบบ NPO ถูกคัดเลือกเข้าในตัวแบบ NPO ร้อยละเกิน 40.63 เมื่อ $n = 500$ และร้อยละเกิน 51.14 เมื่อ

$n=1,000$ และระดับความสัมพันธ์ระหว่างตัวแปรอธิบายในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่และ Toeplitz เพิ่มขึ้นร้อยละการคัดเลือกเข้าในตัวแบบ NPO จะลดลง และมีตัวแปรอธิบายถูกคัดเลือกผิดเป็นตัวแบบ PO อยู่ในช่วงร้อยละ 8.46 - 61.67 แต่ตัวแปรอธิบายไม่ถูกตัดออกจากสมการยกเว้นที่ $n=500$ ในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่และ Toeplitz ที่ $r=0.8$ (ร้อยละ 1.56 และ 1.43 ตามลำดับ)

สำหรับตัวแปรอธิบายที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ส่วนใหญ่ถูกคัดออกจากตัวแบบด้วยร้อยละเกิน 87.50 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและไม่เกินร้อยละ 6.25 ถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ PO แต่ตัวแปรอธิบายไม่ถูกคัดเลือกผิดโดยถูกคัดเข้าในตัวแบบ NPO ยกเว้นรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบอิสระ เมื่อ $n=500$ และ $n=1,000$ และรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่ เมื่อ $r=0.8$ เมื่อ $n=500$

เมื่อเปรียบเทียบร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยพบว่ามีค่าน้อยมากในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบายในกรณี $k=3$ แต่ในกรณี $k=5$ และ $k=10$ มีร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยอยู่ในช่วงร้อยละ 5 - 34 ในทุกกรณีของรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย เมื่อเปรียบเทียบระดับความสัมพันธ์ระหว่างตัวแปรอธิบายในรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบคงที่และ Toeplitz เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอธิบายเพิ่มร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยจะเพิ่มขึ้นตาม นอกจากนั้นขนาดตัวอย่างที่เพิ่มขึ้น ($n=100,500,1000$) มีผลให้ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยลดลงตาม

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาสมรรถนะของวิธีมอนติคาร์โลโซมาร์คอฟก้าวกระโดดย้อนกลับได้ในการคัดเลือกตัวแปรและตัวแบบการถดถอยเชิงอันดับกรณีที่ตัวแปรอธิบายมีความสัมพันธ์เชิงเส้นแบบพหุ โดยทำการศึกษาตัวแบบการถดถอยลอจิสติกซึ่งแบ่งเป็น 3 ตัวแบบคือ ตัวแบบออดส์เชิงสัดส่วน (PO), ตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน (NPO) และตัวแบบออดส์เชิงสัดส่วนบางส่วน (PPO) ในแต่ละตัวแบบกำหนดตัวแปรอธิบายเป็น 3, 5 และ 10 ตัวแปร ซึ่งตัวแปรอธิบายมีความสัมพันธ์กันในรูปแบบความสัมพันธ์เป็นแบบอิสระ, แบบคงที่, แบบ Toeplitz กำหนดระดับสัมประสิทธิ์สหสัมพันธ์เป็น $r = 0, 0.5, 0.8$ และรูปแบบความสัมพันธ์แบบ Hub Toeplitz กำหนดสัมประสิทธิ์สหสัมพันธ์เป็น $r_{\min} = 0.3$ และ $r_{\max} = 0.8$ และในแต่ละกรณีกำหนดขนาดตัวอย่างเป็น $n = 50, 100, 500$ และ $1,000$ เกณฑ์ที่ใช้ในการพิจารณาสมรรถนะ คือ ร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย โดยมีการสรุปผลและข้อเสนอแนะดังนี้

5.1 สรุปผลการวิจัย

ผลการศึกษาสมรรถนะของวิธีมอนติคาร์โลโซมาร์คอฟก้าวกระโดดย้อนกลับได้จากการจำลองตัวแบบการถดถอยลอจิสติกทั้ง 3 ตัวแบบ โดยพิจารณาจากร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย

5.1.1 ผลการศึกษาสมรรถนะของวิธีมอนติคาร์โลโซมาร์คอฟก้าวกระโดดย้อนกลับได้จากการจำลองตัวแบบ PO, NPO และ PPO

ผลการดำเนินการวิจัยด้วยร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ยของตัวแบบในการจำลองตัวแบบ PO, NPO และ PPO

เมื่อพิจารณาขนาดตัวอย่างที่เหมาะสมสรุปได้ว่า ตัวแบบการถดถอยลอจิสติกทั้ง 3 ตัวแบบ ตัวแบบ PO เป็นตัวแบบที่ต้องการขนาดตัวอย่างน้อยกว่าตัวแบบ NPO และ PPO ส่วนตัวแบบ NPO เป็นตัวแบบที่ต้องการขนาดตัวอย่างมากที่สุด จากการวิจัยพบว่า ขนาดตัวอย่างที่เหมาะสมสำหรับตัวแบบ PO และ NPO ควรมีอย่างน้อย 10 และ 20 เท่าของจำนวนตัวแปรอธิบายตามลำดับ ส่วนตัวแบบ PPO ควรมีขนาดตัวอย่างอยู่ระหว่างตัวแบบ PO และ NPO

จากการพิจารณาจำนวนตัวแปรอธิบายที่อยู่ในตัวแบบเมื่อมีขนาดตัวอย่างที่เพียงพอ พบว่ากรณีตัวแบบ PO เมื่อ $k = 3, 5, 10$ ตัวแปรอธิบายจะถูกคัดเลือกเข้าในตัวแบบ PO มีสมรรถนะที่สูงในทุกกรณี โดยมีการคัดเลือกผิดเป็นตัวแบบ NPO น้อยมาก ส่วนกรณีตัวแบบ NPO เมื่อ $k = 3$ ตัวแปรอธิบายจะถูกคัดเลือกเข้าด้วยตัวแบบ NPO สูงจึงทำให้มีสมรรถนะในการคัดเลือกตัวแปรอธิบายสูง แต่ที่ $k = 5$ สมรรถนะในการคัดเลือกตัวแปรอธิบายจะลดลง และเมื่อจำนวนตัวแปรมากขึ้นเป็น $k = 10$ สมรรถนะในการคัดเลือกตัวแปรอธิบายจะลดลงมากโดยตัวแปรอธิบายจะถูกคัดเลือกเข้าด้วยตัวแบบ NPO เพียงร้อยละ 50 (โดยประมาณ) ส่วนที่เหลือจะถูกคัดเลือกผิดเป็นตัวแบบ PO และกรณีตัวแบบ PPO การคัดเลือกตัวแปรอธิบายของ $k = 3$ มีสมรรถนะมากกว่า $k = 5$ และ $k = 10$ ส่วนในกรณีที่ตัวแปรอธิบายไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ตัวแปรอธิบายจะถูกคัดเลือกออกจากตัวแบบในทุกกรณีโดยไม่ขึ้นอยู่กับตัวแบบ PO หรือ NPO แสดงว่าวิธีมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้เป็นวิธีที่ไม่คัดเลือกตัวแปรที่ไม่มีความสัมพันธ์กับตัวแบบเข้ามาในตัวแบบ

การคัดเลือกตัวแปรของตัวแบบ PO มีสมรรถนะมากกว่า NPO โดยตัวแปรอธิบายที่มีตัวแบบ NPO จะถูกคัดเลือกเข้าในตัวแบบ PO เนื่องจากตัวแบบ NPO มีจำนวนพารามิเตอร์ถดถอยและความซับซ้อนของตัวแบบมากกว่าตัวแบบ PO จึงมีความแตกต่างระหว่างมิติของพารามิเตอร์ทำให้การดำเนินงานของตัวแบบ PO ที่เป็นตัวแบบที่ง่ายกว่าจึงถูกเลือกมากกว่า

เมื่อพิจารณารูปแบบความสัมพันธ์และระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย โดยพิจารณาจากร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย พบว่าสมรรถนะการคัดเลือกตัวแปรไม่ขึ้นอยู่กับรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายแบบอิสระ, แบบคงที่, แบบ Toeplitz และแบบ Hub Toeplitz ในตัวแบบการถดถอยลอจิสติกทั้ง 3 ตัวแบบ แต่ระดับความสัมพันธ์ระหว่างตัวแปรอธิบายในรูปแบบความสัมพันธ์เป็นแบบคงที่และ Toeplitz อาจมีแนวโน้มได้ว่าเมื่อระดับสัมประสิทธิ์สหสัมพันธ์สูงจะทำให้สมรรถนะการคัดเลือกตัวแปรลดลงแต่ไม่สามารถสรุปได้แน่ชัด เนื่องจากเมื่อระดับสัมประสิทธิ์สหสัมพันธ์เป็น 0 สมรรถนะการคัดเลือกตัวแปรกลับลดลง

5.1.2 เปรียบเทียบสมรรถนะของวิธีมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับได้ระหว่างตัวแบบ PO, NPO และ PPO

ทำการเปรียบเทียบสมรรถนะของวิธีมอนติคาร์โลโซมาร์คอฟก๊าวกระโดดย้อนกลับโดยพิจารณาจากร้อยละการจำแนกผิดโดยเฉลี่ย เมื่อตัวแปรอธิบายมีจำนวนตัวแปร, ขนาดตัวอย่าง, รูปแบบความสัมพันธ์และระดับความสัมพันธ์ระหว่างตัวแปรอธิบายเดียวกัน พบว่าในกรณีที่ตัวแบบ PO ตัวแปรอธิบายถูกคัดเลือกถูกต้องร้อยละเกือบ 100 ในทุกกรณี แต่ในกรณีตัวแบบ

NPO มีการคัดเลือกตัวแปรผิดเป็น NPO ประมาณร้อยละ 50 และกรณีตัวแบบ PPO มีร้อยละการคัดเลือกตัวแปรผิดอยู่ระหว่างตัวแบบ PO และ NPO แสดงว่าสมรรถนะของวิธีมอนติคาร์โลโซ่มาร์คอฟก้าวกระโดดย้อนกลับได้ของตัวแบบ PO มีมากที่สุดรองลงมาเป็นตัวแบบ PPO และ NPO ตามลำดับ

5.2 ข้อเสนอแนะ

1. การจำลองตัวแบบการถดถอยลอจิสติกในแต่ละตัวแบบกำหนดตัวแปรอธิบายเพียง 1 ตัวแปรให้ไม่มีความสัมพันธ์กับตัวแปรตอบสนอง ($\beta = 0$) อาจทำการศึกษาในกรณีที่ตัวแปรไม่มีความสัมพันธ์กับตัวแปรตอบสนองมากกว่า 1 ตัว
2. ควรทำการศึกษาวิธีการโซ่มาร์คอฟก้าวกระโดดย้อนกลับได้เปรียบเทียบกับวิธีการคัดเลือกตัวแปรวิธีอื่น ๆ

รายการอ้างอิง

หนังสือและบทความในหนังสือ

- วีรานันท์ พงศาภักดี. (2541). *การวิเคราะห์ข้อมูลเชิงกลุ่ม: ทฤษฎีและการประยุกต์(กับ GLIM และ SPSS/FW)* (พิมพ์ครั้งที่ 2). นครปฐม: โรงพิมพ์มหาวิทยาลัยศิลปากร.
- วีรานันท์ พงศาภักดี. (2555). *การวิเคราะห์ข้อมูลจำแนกประเภท: ทฤษฎีและการประยุกต์ด้วย GLIM, SPSS, SAS และ MTB* (พิมพ์ครั้งที่ 3). นครปฐม: โรงพิมพ์มหาวิทยาลัยศิลปากร.
- Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed. Canada: John Wiley & Sons.
- Congdon, P. (2006). *Bayesian Statistical Modelling*. 2nd ed. New York: John Wiley & Sons.
- Gelman, A., Carlin, J. B., Stern, H. S., David B. Dunson, Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. 3rd ed. United States: CRC Press.

บทความวารสาร

- Bender, R., and Grouven, U. (1998). Using Binary Logistic Regression Models for Ordinal Data with Non-proportional Odds. *Journal of Clinical Epidemiology*, 51(10), 809–816.
- Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society Series B*, 65, 3–55. doi: 1369–7412/03/65003
- Dellaportas, P., Forster, J. J., and Ntzoufras, I. (1998). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 21(1), 27–36.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–32.
- Hardin, J., Garcia, S. R., and Golan, D. (2013). A Method for Generating Realistic Correlation Matrices. *The Annals of Applied Statistics*, 7(3), 1733–1762. doi: 10.1214/13-AOAS638.

รายการอ้างอิง (ต่อ)

- Hastie, D., and Green, P. J. (2011). Model choice using reversible jump Markov Chain Monte Carl. *Statistica Neerlandica*, 66(3), 309–338. doi: 10.1111/j.1467-9574.2012.00516.x.
- McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society Series B (Methodological)*. 42(2). 109-142.
- McKinley, T. J., Murcia, P. R., Gog, J. R., Varela M, Wood JLN (2011) A Bayesian Approach to Analyse Genetic Variation within RNA Viral Populations. *PLoS Comput Biol*, 7(3), e1002027. doi: 10.1371/journal.pcbi.1002027.
- McKinley, T. J., Morters, M., and Wood, J. L. N. (2015). Bayesian Model Choice in Cumulative Link Ordinal Regression Models. *International Society for Bayesian Analysis*, 10,1-30. doi: 10.1214/14-BA884.
- O'Hara, R. B. and Sillanpää, M. J. (2009). A Review of Baysian Variable Selection Methods: What, How and Which. *Bayesian Analysis*, 4, 85-118. doi: 10.1214/09-BA403.
- Peterson, B., Harrell, F. E., and Jr. (1990). Partial Proportional Odds Models for Ordinal Response Variables. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 39(2), 205-217.
- Richardson, S., and Green, P. J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society Series B*, 59(4), 731-792. doi: 0035–9246/97/59731.
- Stephens, M., and Balding ,D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10, 686- 690. doi: 10.1038/nrg2615.
- Tutz, G., and Scholz, T. (2003). *Ordinal regression modelling between proportional odds and non-proportional odds*. Technical report, Institute of Statistics, University of Munich.

รายการอ้างอิง (ต่อ)

- Waagepetersen, R., and Sorensen, D. (2001). A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping. *International Statistical Review*, 69(1), 49–61.
- Web, E. L. and Forster, J. J. (2007). Bayesian model determination for multivariate ordinal and binary data. *Computational Statistics & Data Analysis*, 52 (2008), 2632 – 2649. doi: 10.1016/j.csda.2007.09.008.

วิทยานิพนธ์

- ศิริลักษณ์ ชัยโชติชัย. (2549). การประมาณแบบเบย์สำหรับเมทริกซ์สหสัมพันธ์เมื่อข้อมูลเกิดการสูญหายแบบเท่ากันและการแจกแจงเริ่มแรกเป็นแบบสมมาตร. (วิทยานิพนธ์ปริญญาโท). จุฬาลงกรณ์มหาวิทยาลัย, คณะพาณิชยศาสตร์และการบัญชี.

ภาคผนวก



ภาคผนวก ก

โปรแกรมที่ใช้ในการสร้างและวิเคราะห์ข้อมูล

1. โปรแกรมที่ใช้ในการจำลองข้อมูล เมื่อกำหนดตัวแปรอธิบายเป็นตัวแบบ PO มีจำนวนตัวแปรอธิบาย 3 ตัวแปร

```
library(MASS)
library(devtools)
install_github("tjmckinley/BayesOrd")
library(BayesOrd) #install_github("tjmckinley/BayesOrd")
library(coda)

n <- 50      # จำนวนตัวอย่าง ( $n = 50, 100, 500, 1000$ )
k <- 3       # จำนวนตัวแปร ( $k = 3, 5, 10$ )
r <- 0       # ระดับสัมประสิทธิ์สหสัมพันธ์ ( $r = 0, 0.5, 0.8$ )
m <- 100     # จำนวนการจำลองข้อมูลซ้ำ

สุ่มพารามิเตอร์ของตัวแบบ PO เมื่อมีจำนวนตัวแปรอธิบาย 3 ตัวแปร
theta1 <- rnorm(n = 1, mean = -1, sd = 0.1)
theta2 <- rnorm(n = 1, mean = 1, sd = 0.1)+rnorm(n = 1, mean = 0, sd = 0.1)
beta1 <- 0
beta2 <- rnorm(n = 1, mean = 0, sd = 5)
beta3 <- rnorm(n = 1, mean = 0, sd = 5)
#theta1; theta2; beta1; beta2; beta3

สร้างข้อมูลตัวแปรอธิบาย ( $X$ ) และตัวแปรตอบสนอง ( $Y$ )
x <- c()      #initialize this storage vector
x1 <- c()
x2 <- c()
x3 <- c()
u <- c()

สร้างเมทริกซ์ความสัมพันธ์ภายใต้สถานการณ์ดังนี้
# กรณีสร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบคงที่
R <- matrix(rep(r,(k*k)),ncol=p,nrow = k)
```

```

diag(R) <- 1
cor <- matrix(c(R),ncol = k)
# กรณีสร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบ Toeplitz
cor<- outer(1:k, 1:, function(x,y) {r^abs(x-y)})
# กรณีสร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบ Hub Toeplitz
rmax<-0.8
rmin<-0.3
R <- outer(1:k, 1:k, function(x,y) {rmax-(((rmax-rmin)/(k-2))*(abs(x-y)-1))})
diag(R) <- 1
cor<- matrix(c(R),ncol = k)
สร้างข้อมูลตัวแปรอธิบาย ( $X$ ) 3 ตัวแปร
x <- mvrnorm(n, rep(0,k),cor)
x <- abs(x)
x <- round(x , digits = 4)
x1 <- x[1:n]
x2 <- x[(n+1):(2*n)]
x3 <- x[((2*n)+1):(3*n)]
กำหนดรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองเป็นแบบ PO
dn1 <- theta1-(beta1*x1+beta2*x2+beta3*x3)
dn2 <- theta2-(beta1*x1+beta2*x2+beta3*x3)
คำนวณความน่าจะเป็นของแต่ละกลุ่มตัวแปรตอบสนอง ( $Y$ )
pr1 <-exp(dn1)/(1+exp(dn1))
pr2 <-exp(dn2)/(1+exp(dn2))-exp(dn1)/(1+exp(dn1))
pr3 <-1-exp(dn2)/(1+exp(dn2))
สร้างข้อมูลตัวแปรตอบสนอง ( $Y$ ) แบ่งเป็น 3 อันดับ
#-----Cut-point Uniform-----#
y.new = matrix(0,n,m)
u = runif(n*m,0,1)
U = matrix(u,n,m,byrow = T)
จำลองตัวแปรตอบสนองชุดที่ 1

```

```

sink("1")
l <- length(pr1)
y1 = rep(0,l)
for(i in 1:l) {
  if(U[i,1] < pr1[i]) {y1[i]=1}
  if((pr1[i] <= U[i,1]) & (U[i,1] < pr1[i]+pr2[i])) {y1[i]=2}
  if(U[i,1] >= (pr1[i]+pr2[i])) {y1[i]=3}
}
y1 = as.factor(y1)
data <- data.frame(x1,x2,x3,y1)

```

ใช้ชุดคำสั่ง BayesOrd ในการคำนวณค่า PPA

```

bayesord.PO <- bayesord(y1 ~ x1 + x2 + x3, data = data, var.select=T, model.type =
c("both"), niter = 2e+04, noutputsum = 1e+04, nchains = 2, start = 5000, thin = 150)
bayesord.PO <- window(bayesord.PO, start = 5000, thin = 150)
res1 <- summary(bayesord.PO)
print(summary(bayesord.PO)$variable)
fitted.PO <- fitted(bayesord.PO)
print(fitted.PO)
result1<-data.frame(res1$variables)
write.csv(result1,file = "res_1.csv")
sink()

```

ทำการจำลองข้อมูลและใช้การวิเคราะห์ข้อมูลด้วยชุดคำสั่ง BayesOrd ซ้ำจนครบ 100 ชุด หลังจากนั้นนำผลลัพธ์ที่ได้มาคำนวณร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และ ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย

#####

2. โปรแกรมที่ใช้ในการจำลองข้อมูล เมื่อกำหนดตัวแปรอธิบายเป็นตัวแบบ NPO มีจำนวนตัวแปรอธิบาย 3 ตัวแปร

```
library(MASS)
library(devtools)
install_github("tjmckinley/BayesOrd")
library(BayesOrd) #install_github("tjmckinley/BayesOrd")
library(coda)

n <- 50      # จำนวนตัวอย่าง ( $n = 50, 100, 500, 1000$ )
k <- 3       # จำนวนตัวแปร ( $k = 3, 5, 10$ )
r <- 0       # ระดับสัมประสิทธิ์สหสัมพันธ์ ( $r = 0, 0.5, 0.8$ )
m <- 100     # จำนวนการจำลองข้อมูลซ้ำ

# สุ่มพารามิเตอร์ของตัวแบบ NPO เมื่อมีจำนวนตัวแปรอธิบาย 3 ตัวแปร
theta1 <- rnorm(n = 1, mean = 0, sd = 0.1)
theta2 <- rnorm(n = 1, mean = 1, sd = 0.1) + rnorm(n = 1, mean = 0, sd = 0.1)
beta11 <- 0
beta21 <- rnorm(n = 1, mean = 3, sd = 0.5)
beta31 <- rnorm(n = 1, mean = 2, sd = 0.2)
beta12 <- 0
beta22 <- rnorm(n = 1, mean = -2.5, sd = 0.2)
beta32 <- rnorm(n = 1, mean = 1, sd = 1)
theta1; theta2; beta11; beta21; beta31; beta12; beta22; beta32;

# สร้างข้อมูลตัวแปรอธิบาย ( $X$ ) และตัวแปรตอบสนอง ( $Y$ )
x <- c()      # initialize this storage vector
x1 <- c()
x2 <- c()
x3 <- c()
u <- c()

# สร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบคงที่
R <- matrix(rep(r, (k*k)), ncol=p, nrow = k)
```

```

diag(R) <- 1
cor <- matrix(c(R),ncol = k)
สร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบ Toeplitz
cor<- outer(1:k, 1:, function(x,y) {r^abs(x-y)})
สร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบ Hub Toeplitz
rmin<-0.3
rmax<-0.8
R <- outer(1:k, 1:k, function(x,y) {rmax-(((rmax-rmin)/(k-2))*(abs(x-y)-1))})
diag(R) <- 1
cor<- matrix(c(R),ncol = k)
สร้างข้อมูลตัวแปรอธิบาย (X) 3 ตัวแปร
x <- mvrnorm(n, rep(0,k),cor)
x <- abs(x)
x <- round(x , digits = 4)
x1 <- x[1:n]
x2 <- x[(n+1):(2*n)]
x3 <- x[((2*n)+1):(3*n)]
**ทดสอบเงื่อนไข  $\theta_j - \theta_{j+1} < \sum_{k=1}^K \min(\beta_{kj} - \beta_{k(j+1)}) X_k^m, [\beta_{kj} - \beta_{k(j+1)}] X_k^M$  ในตัวแบบ NPO
theta1-theta2 <- sum(min((beta11-beta12)*min(x1), (beta11-beta12)*max(x1)),
                     min((beta21-beta22)*min(x2), (beta21-beta22)*max(x2)),
                     min((beta31-beta32)*min(x3), (beta31-beta32)*max(x3)))
กำหนดรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองเป็นแบบ NPO
dn1 <- theta1-(beta11*x1+beta21*x2+beta31*x3)
dn2 <- theta2-(beta12*x1+beta22*x2+beta32*x3)
คำนวณความน่าจะเป็นของแต่ละกลุ่มตัวแปรตอบสนอง (Y)
pr1 <- exp(dn1)/(1+exp(dn1))
pr2 <- exp(dn2)/(1+exp(dn2))-exp(dn1)/(1+exp(dn1))
pr3 <- 1-exp(dn2)/(1+exp(dn2))
สร้างข้อมูลตัวแปรตอบสนอง (Y) แบ่งเป็น 3 อันดับ
#-----Cut-point Uniform-----#

```

```
y.new = matrix(0,n,m)
u = runif(n*m,0,1)
U = matrix(u,n,m,byrow = T)
```

จำลองตัวแปรตอบสนองของชุดที่ 1

```
sink("1")
l <- length(pr1)
y1 = rep(0,l)
for(i in 1:l) {
  if(U[i,1] < pr1[i]) {y1[i]=1}
  if((pr1[i] <= U[i,1]) & (U[i,1] < pr1[i]+pr2[i])) {y1[i]=2}
  if(U[i,1] >= (pr1[i]+pr2[i])) {y1[i]=3}
}
y1 = as.factor(y1)
data <- data.frame(x1,x2,x3,y1)
```

ใช้ชุดคำสั่ง BayesOrd ในการคำนวณค่า PPA

```
bayesord.NPO <- bayesord(y1~ x1+x2+x3, data = data, var.select=T, model.type =
c("both"), niter = 2e+04, noutputsum = 1e+04, nchains = 2, start = 5000, thin = 150)
bayesord.NPO <- window(bayesord.NPO, start = 5000, thin = 150)
res1 <- summary(bayesord.NPO)
print(summary(bayesord.NPO)$variable)
fitted.NPO <- fitted(bayesord.NPO)
print(fitted.NPO)
result1<-data.frame(res1$variables)
write.csv(result1,file = "res_1.csv")
sink()
```

ทำการจำลองข้อมูลและใช้การวิเคราะห์ข้อมูลด้วยชุดคำสั่ง BayesOrd ซ้ำจนครบ 100 ชุด หลังจากนั้นนำผลลัพธ์ที่ได้มาคำนวณร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และ ร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย

```
#####
```

3. โปรแกรมที่ใช้ในการจำลองข้อมูล เมื่อกำหนดตัวแปรอธิบายเป็นตัวแบบ PPO มีจำนวนตัวแปรอธิบาย 3 ตัวแปร

```
library(MASS)
library(devtools)
install_github("tjmckinley/BayesOrd")
library(BayesOrd) #install_github("tjmckinley/BayesOrd")
library(coda)

n <- 50      # จำนวนตัวอย่าง ( $n = 50, 100, 500, 1000$ )
k <- 3       # จำนวนตัวแปร ( $k = 3, 5, 10$ )
r <- 0       # ระดับสัมประสิทธิ์สหสัมพันธ์ ( $r = 0, 0.5, 0.8$ )
m <- 100     # จำนวนการจำลองข้อมูลซ้ำ

สุ่มพารามิเตอร์ของตัวแบบ PPO เมื่อมีจำนวนตัวแปรอธิบาย 3 ตัวแปร
theta1 <- rnorm(n = 1, mean = 0, sd = 0.1)
theta2 <- rnorm(n = 1, mean = 1, sd = 0.1) + rnorm(n = 1, mean = 0, sd = 0.1)
beta11 <- 0
beta21 <- rnorm(n = 1, mean = 3, sd = 0.5)
beta31 <- rnorm(n = 1, mean = 1, sd = 2)
beta12 <- 0
beta22 <- rnorm(n = 1, mean = -3, sd = 0.2)
theta1; theta2; beta11; beta21; beta31; beta12; beta22;

สร้างข้อมูลตัวแปรอธิบาย ( $X$ ) และตัวแปรตอบสนอง ( $Y$ )
x <- c()     #initialize this storage vector
x1 <- c()
x2 <- c()
x3 <- c()
u <- c()

สร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบคงที่
R <- matrix(rep(r, (k*k)), ncol=p, nrow = k)
diag(R) <- 1
cor <- matrix(c(R), ncol = k)
```

สร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบ Toeplitz

```
cor<- outer(1:k, 1:, function(x,y) {r^abs(x-y)})
```

สร้างเมทริกซ์รูปแบบความสัมพันธ์เป็นแบบ Hub Toeplitz

```
rmax<-0.8
```

```
rmin<-0.3
```

```
R <- outer(1:k, 1:k, function(x,y) {rmax-(((rmax-rmin)/(k-2))*(abs(x-y)-1))})
```

```
diag(R) <- 1
```

```
cor<- matrix(c(R),ncol = k)
```

สร้างข้อมูลตัวแปรอธิบาย (X) 3 ตัวแปร

```
x <- mvrnorm(n, rep(0,k),cor)
```

```
x <- abs(x)
```

```
x <- round(x , digits = 4)
```

```
x1 <- x[1:n]
```

```
x2 <- x[(n+1):(2*n)]
```

```
x3 <- x[((2*n)+1):(3*n)]
```

****ทดสอบเงื่อนไข $\theta_j - \theta_{j+1} < \sum_{k=1}^K \min(\beta_{kj} - \beta_{k(j+1)}) X_k^m, [\beta_{kj} - \beta_{k(j+1)}] X_k^M$ ในตัวแบบ PPO**

```
theta1-theta2 < sum(min((beta11-beta12)*min(x1), (beta11-beta12)*max(x1)),
                    min((beta21-beta22)*min(x2), (beta21-beta22)*max(x2)),
                    min((beta31-beta31)*min(x3), (beta31-beta31)*max(x3)))
```

กำหนดรูปแบบความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองเป็นแบบ PPO

```
dn1 <- theta1-(beta11*x1+beta21*x2+beta31*x3)
```

```
dn2 <- theta2-(beta12*x1+beta22*x2+beta31*x3)
```

คำนวณความน่าจะเป็นของแต่ละกลุ่มตัวแปรตอบสนอง (Y)

```
pr1 <-exp(dn1)/(1+exp(dn1))
```

```
pr2 <-exp(dn2)/(1+exp(dn2))-exp(dn1)/(1+exp(dn1))
```

```
pr3 <-1-exp(dn2)/(1+exp(dn2))
```

สร้างข้อมูลตัวแปรตอบสนอง (Y) แบ่งเป็น 3 อันดับ

```
#-----Cut-point Uniform-----#
```

```
y.new = matrix(0,n,m)
```

```
u = runif(n*m,0,1)
```

```
U = matrix(u,n,m,byrow = T)
```

จำลองตัวแปรตอบสนองชุดที่ 1

```
sink("1")
l <- length(pr1)
y1 = rep(0,l)
for(i in 1:l) {
  if(U[i,1] < pr1[i]) {y1[i]=1}
  if((pr1[i] <= U[i,1]) & (U[i,1] < pr1[i]+pr2[i])) {y1[i]=2}
  if(U[i,1] >= (pr1[i]+pr2[i])) {y1[i]=3}
}
y1 = as.factor(y1)
data <- data.frame(x1,x2,x3,y1)
```

ใช้ชุดคำสั่ง BayesOrd ในการคำนวณค่า PPA

```
bayesord.PPO <- bayesord(y1 ~ x1+x2+x3, data = data, var.select=T, model.type =
c("both"), niter = 2e+04, noutputsum = 1e+04, nchains = 2, start = 5000, thin = 150)
bayesord.PPO <- window(bayesord.PPO, start = 5000, thin = 150)
res1 <- summary(bayesord.PPO)
print(summary(bayesord.PPO)$variable)
fitted.PPO <- fitted(bayesord.PPO)
print(fitted.PPO)
result1<-data.frame(res1$variables)
write.csv(result1,file = "res_1.csv")
sink()
```

ทำการจำลองข้อมูลและใช้การวิเคราะห์ข้อมูลด้วยชุดคำสั่ง BayesOrd ซ้ำจนครบ 100 ชุด หลังจากนั้นนำผลลัพธ์ที่ได้มาคำนวณร้อยละของการคัดเลือกตัวแปรและตัวแบบโดยเฉลี่ย และร้อยละการจำแนกตัวแบบผิดโดยเฉลี่ย

```
#####
```

ประวัติผู้เขียน

ชื่อ	นางสาวชลนรรจ์ สุขทั่วญาติ
วันเดือนปีเกิด	3 มกราคม พ.ศ.2537
วุฒิการศึกษา	ปีการศึกษา 2558: วิทยาศาสตร์บัณฑิต สาขาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

ผลงานทางวิชาการ

ชลนรรจ์ สุขทั่วญาติ และ แสงหล้า ชัยมงคล. (ตุลาคม 2561). “การศึกษาสมรรถนะการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงเส้นพหุโดยใช้วิธีการก้าวกระโดดย้อนกลับได้.” การประชุมวิชาการสถิติประยุกต์และเทคโนโลยีสารสนเทศระดับชาติประจำปี 2561, สถาบันบัณฑิตพัฒนบริหารศาสตร์, กรุงเทพฯ, 30 ตุลาคม 2561