การทดสอบอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด

นางสาวพัชรี วงศ์พงศาลี

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A TEST OF THE CONVERGENCE RATE OF A HIT-AND-RUN SAMPLER

ON SOME BOUNDED CONVEX REGIONS

Miss Patcharee Wongpongsalee

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2009

Copyright of Chulalongkorn University

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หัวข้อวิทยานิพนธ์	การทดสอบอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด
โดย	นางสาวพัชรี วงศ์พงศาลี
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์)

.... กรรมการ

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... กรรมการ

(อาจารย์ ดร.อรุณี กำลัง)

.... กรรมการภายนอกมหาวิทยาลัย

(รองศาสตราจารย์ ดร.สำรวม จงเจริญ)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พัชรี วงศ์พงศาลี : การทดสอบอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด. (A TEST OF THE CONVERGENCE RATE OF A HIT-AND-RUN SAMPLER ON SOME BOUNDED CONVEX REGIONS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร.เสกสรร เกียรติสุไพบูลย์, 68 หน้า.

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณที่มีขอบเขตจำกัด โดยกำหนดบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด 2 บริเวณ ได้แก่ บริเวณสี่เหลี่ยม S_{rect} และบนบริเวณสามเหลี่ยม S_{tri} ในมิติ (p) ต่าง ๆ สำหรับบริเวณสี่เหลี่ยม S_{rect} กำหนดให้ขนาดตัวอย่างเฉลี่ย (n_{av}) ที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อนที่ยอมรับได้ (e) เป็นค่าวัดอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน ในขณะที่บริเวณสามเหลี่ยม S_{tri} กำหนดให้ขนาดตัวอย่างเฉลี่ย (n_{av}) ที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่งค่าผิดพลาดมาตรฐานของตัวประมาณน้อยกว่า e เป็นค่าวัดอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน โดยกำหนดความคลาดเคลื่อน (e) เท่ากับ 0.1 0.05 และ 0.01

จากการศึกษาพบว่า ขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ของบริเวณที่ศึกษามีความสัมพันธ์ไปในทิศทางเดียวกัน คือ เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นอย่างรวดเร็ว ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นฟังก์ชันพหุนามอันดับที่ 2 ยกเว้นกรณีการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} เมื่อ e เท่ากับ 0.01 เพียงกรณีเดียวเท่านั้นที่ตัวแบบความสัมพันธ์ของ n_{av} กับ p เป็นฟังก์ชันพหุนามอันดับที่ 3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา สถิติ ลายมือชื่อนิสิต พัชรี วงศ์พงศาลี

สาขาวิชา สถิติ ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก เสกสรร เกียรติสุไพบูลย์

ปีการศึกษา 2552

5081851426 : MAJOR STATISTICS

KEYWORDS : MARKOV CHAIN MONTE CARLO / HIT-AND-RUN PROCESS / RATE OF CONVERGENCE

PATCHAREE WONGPONGSALEE: A TEST OF THE CONVERGENCE RATE OF A HIT-AND-RUN SAMPLER ON SOME BOUNDED CONVEX REGIONS. THESIS ADVISOR: ASST. PROF. SEKSAN KIATSUPAIBUL, Ph.D. 68 pp.

The purpose of this research is to investigate the convergence rate of a Hit-and-run sampler on some bounded convex regions. The bounded convex regions under investigation are hyper-rectangular region S_{rect} and hyper-triangular region S_{tri} in various dimensions (p). For the case of hyper-rectangular region S_{rect}, the convergence rate is defined as the average number of samples (n_{av}) that are drawn from the Hit-and-run sampler until the average samples is within the error (e). For the case of hyper-triangular region S_{tri}, the convergence rate is defined as the average number of samples (n_{av}) that are drawn from the Hit-and-run sampler until the standard error of the sample average is less than the error (e). The errors are set at 0.1, 0.05 and 0.01.

The results from all cases, except that from the case of hyper-triangular region S_{tri} when error is 0.01, show that the sample size average (n_{av}) is a quadratic function of the number of dimensions (p). In the case of hyper-triangular region S_{tri} when error is 0.01, the sample size average (n_{av}) is a polynomial function of degree 3 of the number of dimensions (p).

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Department : Statistics	Student's Signature
Field of Study : Statistics	Advisor's Signature
Academic Year : 2009	

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือ คำแนะนำและข้อเสนอแนะ รวมถึงการเอาใจใส่อย่างใกล้ชิดจากผู้ช่วยศาสตราจารย์ ดร .เสกสรร เกียรติสุไพบูลย์ ผู้เขียนขอน้อมกราบขอบพระคุณต่อท่านอาจารย์เป็นอย่างสูง ตลอดทั้งอาจารย์ทุกท่านที่ได้สอนผู้เขียนมา ณ โอกาสนี้ด้วย

ผู้เขียนขอบกราบขอบพระคุณ รองศาสตราจารย์ ดร .ธีระพร วีระถาวร ประธานกรรมการการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ ดร.อรุณี กำลัง และรองศาสตราจารย์ ดร.สำรวม จงเจริญ กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำอันเป็นประโยชน์ในการแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

ผู้เขียนขอขอบคุณเจ้าหน้าที่ห้องสมุด คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้เอื้ออำนวยความสะดวกในด้านตำราซึ่งใช้ประ กอบการทำวิทยานิพนธ์ และขอขอบคุณ รุ่นพี่ และเพื่อน ๆ ทุกคนที่ให้ความร่วมมือและเป็นกำลังใจแก่ผู้เขียน

สุดท้ายนี้ ผู้เขียนใคร่ขอกราบขอบพระคุณบิดา มารดา ที่ช่วยส่งเสริมและสนับสนุนให้ผู้เขียนได้มีโอกาสทางการศึกษาเสมอมาจนสำเร็จการศึกษา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย...	ง
บทคัดย่อภาษาอังกฤษ...	จ
กิตติกรรมประกาศ..	ฉ
สารบัญ...	ช
สารบัญตาราง..	ฌ
สารบัญภาพ...	ฎ
บทที่ 1 บทนำ...	1
1.1 ความเป็นมาและความสำคัญของปัญหา...	1
1.2 วัตถุประสงค์ของการวิจัย..	2
1.3 ขอบเขตของการวิจัย...	2
1.4 ข้อตกลงเบื้องต้น..	3
1.5 คำจำกัดความที่ใช้ในการวิจัย..	3
1.6 ประโยชน์ที่คาดว่าจะได้รับ..	3
1.7 วิธีดำเนินการวิจัย...	4
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง...	5
2.1 แนวคิดและทฤษฎี...	5
2.1.1 ลูกโซ่มาร์คอฟ	5
2.1.2 ความน่าจะเป็นในการเปลี่ยนสถานะ..	5
2.1.3 คุณสมบัติสถานะ..	6
2.1.4 ทฤษฎีลูกโซ่มาร์คอฟสำหรับสเปซทั่วไป..	8
2.1.5 ทฤษฎีบทการลู่เข้าของลูกโซ่มาร์คอฟในกรณี total variation..........	9
2.1.6 Reversibility	11
2.1.7 การสุ่มตัวอย่างแบบฮิตแอนด์รัน...	11
2.1.8 การวิเคราะห์สถานะเสถียรภาพของการจำลอง	12
2.1.9 วิธีค่าเฉลี่ยแบบกลุ่ม	13
2.2 เอกสารที่เกี่ยวข้องกับงานวิจัย...	13

หน้า

บทที่ 3 วิธีดำเนินการวิจัย.. 15
3.1 การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม.................................. 15
3.2 การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม............................... 18
3.3 การศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ยกับจำนวนมิติของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด ณ ระดับความคลาดเคลื่อนต่าง ๆ.................. 21
บทที่ 4 ผลการวิจัย.. 23
4.1 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม... 23
4.2 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม... 43
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ... 63
5.1 สรุปผลการวิจัย.. 63
5.2 อภิปรายผล... 66
5.3 ข้อเสนอแนะ.. 66
รายการอ้างอิง 67
ประวัติผู้เขียนวิทยานิพนธ์.. 68

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
4.1	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6…………………………	24
4.2	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15………………	25
4.3	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50…………..	27
4.4	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6………………………..	30
4.5	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15……………	32
4.6	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50……….	33
4.7	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6………………………..	37
4.8	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15……………	38
4.9	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50……….	40

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่		หน้า
4.10	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6........	43
4.11	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15........	45
4.12	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50........	46
4.13	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6........	50
4.14	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15........	51
4.15	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50........	53
4.16	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6........	56
4.17	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15........	58
4.18	แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50........	59

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพที่ หน้า

3.1 แสดงขั้นตอนการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยมS_{rect} 17

3.2 แสดงขั้นตอนการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมS_{tri} 20

4.1 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] กับจำนวนมิติ (p).. 28

4.2 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] กับจำนวนมิติ (p).. 35

4.3 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] กับจำนวนมิติ (p).. 41

4.4 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 กับจำนวนมิติ (p)........................... 48

4.5 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 กับจำนวนมิติ (p)........................ 54

4.6 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 กับจำนวนมิติ (p)........................ 61

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการอนุมานสถิติแบบเบส์มีความจำเป็นต้องทำการอินทิเกรตฟังก์ชันภายหลัง (posterior function) ซึ่งเป็นฟังก์ชันหลายตัวแปร ซึ่งโดยทั่วไปแล้วจะทำได้ยาก วิธีแก้ไขปัญหาวิธีหนึ่งที่นักสถิติใช้แก้ปัญหา ดังกล่าว คือ การใช้วิธีมอนติคาร์โลอินทิเกรชั่น (Monte Carlo Integration) ซึ่งเป็นการสร้างเวกเตอร์สุ่มที่มีการแจกแจงเป็นไปตามการแจกแจงภายหลัง วิธีมอนติคาร์โลอินทิเกรชั่นที่จัดว่ามีประสิทธิภาพวิธีหนึ่ง ได้แก่ วิธีในกลุ่มที่เรียกว่า เทคนิคลูกโซ่มาร์คอฟมอนติคาร์โล (Marko Chain Monte Carlo หรือ MCMC) เทคนิคลูกโซ่มาร์คอฟมอนติคาร์โลที่มีประสิทธิภาพและนิยมใช้ในการอนุมานสถิติแบบเบส์มีหลายวิธี อาทิ การสุ่มแบบกิบส์ (Gibbs Sampling) วิธีเมโทรโปรลิสแฮสติ้ง (Metropolis - Hasting) และระเบียบวิธีฮิตแอนด์รัน (Hit-and-Run algorithm)

ระเบียบวิธีฮิตแอนด์รัน (Hit-and-Run algorithm) ถูกคิดค้นโดย โรเบิร์ต แอล สมิธ (Robert L. Smith) ในปี 1984 เป็นวิธีที่เหมาะสำหรับการทำลูกโซ่มาร์คอฟมอนติคาร์โลบนบริเวณรูปร่างคอนเวกซ์ (convex body) จากการศึกษาพบว่า ตัวอย่างที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันนี้ จะลู่เข้าสู่การแจกแจงเป้าหมาย (Target distribution) บนบริเวณรูปร่างคอนเวกซ์ ได้เร็วกว่าการเดินแบบสุ่ม (Random walks) อื่น ๆ อาทิ การเดินแบบสุ่มในลูกบอล (Random ball walks) ในปี 1999 ลาสซ์โล ลอวอสซ์ (Laszlo Lovasz) ได้นำวิธีการสุ่มตัว อย่างแบบฮิตแอนด์รัน มาจำลองจุดตัวอย่างบนบริเวณรูปร่างคอนเวกซ์ใน p มิติ สามารถพิสูจน์ว่าความซับซ้อนในการคำนวณ (Computational complexity) อยู่ในพหุนามอันดับที่ 3 หรือ $O^*(p^3)$ ซึ่งจัดว่าเป็นวิธีที่เร็วที่สุดที่รู้จักกันในขณะนี้สำหรับการจำลองเวกเตอร์สุ่มอย่าง สม่ำเสมอบนบริเวณคอนเวกซ์หลายมิติ อย่างไรก็ตาม ผลดังกล่าวเป็นผลทางทฤษฎีซึ่งพิจารณาประสิทธิภาพของการลู่เข้าแบบ total variation ซึ่งอาจกล่าวได้ว่าเป็นการประเมินประสิทธิภาพในกรณีที่แย่ที่สุดที่เป็นไปได้ (worst case)

ในทางปฏิบัติหากไม่ได้พิจารณาประสิทธิภาพของก ารลู่เข้าแบบ total variation วิธีฮิตแอนด์รันอาจจะมีการลู่เข้าในลักษณะอื่นที่เร็วกว่าในทางทฤษฎี การวิจัยนี้จึงทำการศึกษาอัตราการลู่เข้าของค่าเฉลี่ยของตัวอย่างสุ่มแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ในหลายมิติที่มีขอบเขตจำกัด ได้แก่ บริเวณสี่เหลี่ยมในหลายมิติที่มี ขอบเขตอยู่ในช่วง [-1,1] และบริเวณสามเหลี่ยมในหลายมิติที่มีขอบเขตจำกัด ที่ระดับความคลาดเคลื่อน (error) ต่าง ๆ

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยมในหลายมิติที่มีขอบเขตจำกัด
2. เพื่อศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมในหลายมิติที่มีขอบเขตจำกัด

1.3 ขอบเขตของการวิจัย

1. สมมติฐานในการวิจัย คือ ความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มด้วยวิธีฮิตแอนด์รันจนกระทั่งลู่เข้าสู่ค่าเฉลี่ยของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด กับจำนวนมิติ (p) ของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด เป็นตัวแบบความถดถอยพหุนามที่มีอันดับไม่เกินอันดับที่ 3 ดังอสมการ

$$n_{av} \le b_0 + b_1 p + b_2 p^2 + b_3 p^3$$

เมื่อ

n_{av} แทน ขนาดตัวอย่างเฉลี่ยที่สุ่มด้วยวิธีฮิตแอนด์ รันจนกระทั่งลู่เข้าสู่ค่าเฉลี่ยของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด

p แทน จำนวนมิติของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด

2. ศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยมในหลายมิติที่มีขอบเขตจำกัด (S_{rect}) โดยที่

$$S_{rect} = \{X = (x_1, x_2, ..., x_p) \mid x_j \in [-1,1],\ j = 1,2,..., p\}$$

เมื่อ p แทน จำนวนมิติของบริเวณสี่เหลี่ยม S_{rect} ซึ่ง p ได้แก่ 2ถึง10 15 20 25 30 40 และ 50 ที่ระดับความคลาดเคลื่อน (error) ต่าง ๆ ได้แก่ 0.1 0.05 และ 0.01

3. ศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมในหลายมิติที่มีขอบเขตจำกัด (S_{tri}) โดยที่

$$S_{tri} = \{X = (x_1, x_2, ..., x_p) \mid x_1 \ge x_2 \ge ... \ge x_p,\ x_j \in [-1,1],\ j = 1,2,..., p\}$$

เมื่อ p แทน จำนวนมิติของบริเวณสามเหลี่ยม S_{tri} ซึ่ง p ได้แก่ 2ถึง10 15 20 25 30 40 และ 50 ที่ระดับความคลาดเคลื่อน (error) ต่าง ๆ ได้แก่ 0.1 0.05 และ 0.01

1.4 ข้อตกลงเบื้องต้น

บริเวณคอนเวกซ์ที่มีขอบเขตจำกัดในการวิจัยนี้มี 2 บริเวณ ได้แก่

1. บริเวณสี่เหลี่ยมหลายมิติที่มีขอบเขตจำกัด S_{rect} ที่มีค่าเฉลี่ ยอยู่ที่จุดกำเนิด การวิจัยนี้จึงศึกษาการลู่เข้าของการสุ่มตัวอย่างด้วยวิธีฮิตแอนด์รันบนบริเวณดังกล่าวสู่จุดกำเนิดที่ระดับความคลาดเคลื่อนต่างๆ

2. บริเวณสามเหลี่ยมหลายมิติที่มีขอบเขตจำกัด S_{tri} เนื่องจากในงานวิจัยนี้ไม่ทราบค่าเฉลี่ยของบริเวณสามเห ลี่ยมหลายมิติ การวิจัยนี้จึงศึกษาการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณดังกล่าวจนกระทั่ง

$$SE^* < e$$

เมื่อ $SE^* = \dfrac{t_{\frac{\alpha}{2},(m-1)} S_y}{\sqrt{m}}$

S_y แทน ค่าส่วนเบี่ยงเบนมาตรฐานของวิธีค่าเฉลี่ยแบบกลุ่ม (Batch Means)

m แทน จำนวนกลุ่ม (batch) ในการวิจัยนี้กำหนดให้ m เท่ากับ 30

1.5 คำจำกัดความที่ใช้ในการวิจัย

การลู่เข้า (Convergence) หมายถึง ค่าเฉลี่ยของตัวอย่างมีแนวโน้มที่จะลู่เข้าสู่ค่าเฉลี่ยของประชากรเมื่อขนาดตัวอย่างเพิ่มขึ้นโดยให้ $X_1, X_2, \ldots$ เป็นลำดับ (Sequence) ของตัวแปรสุ่ม $\{X_n\}$ จะมีการลู่เข้าสู่ X ในความน่าจะเป็น ก็ต่อเมื่อ

$$\lim_{n\to\infty} P(|X_n - X| < e) = 1 \quad \text{สำหรับทุกค่า } \varepsilon > 0$$

เมื่อ e แทน ความคลาดเคลื่อน

วิธีค่าเฉลี่ยแ บบกลุ่ม (Batch Means) เป็นวิธีการหนึ่งในการประมาณค่าเฉลี่ยของประชากร (μ) ด้วยค่าเฉลี่ยของกลุ่ม $(\bar{Y})$ เมื่อลำดับ $\{X_n\}$ มีความสัมพันธ์กัน

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยมที่มีขอบเขตจำกัดเมื่อจำนวนมิติของบริเวณดังกล่าวเพิ่มขึ้น

2. เพื่อเป็นแนวทางในการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมที่มีขอบเขตจำกัดเมื่อจำนวนมิติของบริเวณดังกล่าวเพิ่มขึ้น

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

1.7 วิธีดำเนินการวิจัย

1. ศึกษาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect}
2. ศึกษาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri}
3. ศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด ณ ระดับความคลาดเคลื่อน (e) ต่าง ๆ
4. สรุปผลการวิจัย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎี

2.1.1 ลูกโซ่มาร์คอฟ (Markov Chain)

พิจารณากระบวนการเฟ้นสุ่มแบบ ไม่ต่อเนื่อง (discrete-time stochastic process) $\{X_n; n = 0, 1, 2, \ldots\}$ กระบวนการนี้จะเป็นลูกโซ่มาร์คอฟแบบไม่ต่อเนื่อง (discrete-time Markov chain) ถ้า $m, n \geq 0$ และสถานะ $i, j \in S, x_u$ และ $0 \leq u < m$ เป็นจำนวนเต็มที่ไม่เป็นล บ (nonnegative integers)

$$P\left(X_{n+m} = j \mid X_m = i, X_u = x_u, 0 \leq u < m\right) = P\left(X_{n+m} = j \mid X_m = i\right)$$

ลูกโซ่มาร์คอฟแบบไม่ต่อเนื่อง หรือ ลูกโซ่มาร์คอฟ เป็นกระบวนการเฟ้นสุ่มที่มีคุณสมบัติมาร์คอฟ (Markovian property) คือ ความน่าจะเป็นแบบมีเงื่อนไขของอนาคต X_{n+m} เมื่อกำหนดปัจจุบัน X_m และอดีต X_u โดยที่ $0 \leq u < m$ จะขึ้นอยู่กับปัจจุบัน X_m เท่านั้น และเป็นอิสระกับอดีต X_u เรียก S ว่า สเปซของสถานะ (state space)

2.1.2 ความน่าจะเป็นในการเปลี่ยนสถานะ (Transition Probability)

กำหนดให้ สเปซ S เป็น เซตจำกัด (finite set) และกำหนดให้

$$p_{ij}^{n} = P\left(X_{n+m} = j \mid X_m = i\right) \text{ สำหรับ } m, n \geq 0 \text{ และ } i, j \geq 0$$

เป็นความน่าจะเป็นที่กระบวนการซึ่งปัจจุบันอยู่ในสถานะ i จะเปลี่ยนไปเป็นสถานะ j ในขั้นที่ n ที่เรียกว่า ความน่าจะเป็นในการเปลี่ยนสถานะขั้นที่ n (n-step transition probability) โดยที่

$$p_{ij}^{0} = \begin{cases} 1 & ; j = i \\ 0 & ; \end{cases}$$

และ $$p_{ij}^{1} = p_{ij}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เป็นคว ามน่าจะเป็นในการเปลี่ยนสถานะขั้นที่ 1 (One-step transition probability)

นิยาม สมการแชปแมน-โคลโมโกรอฟ (Chapman-Kolmogorov equation)

$$p_{ij}^{n+m} = \sum_{k=0}^{\infty} p_{ik}^{n} p_{kj}^{m} \quad \text{สำหรับทุกค่า } m,n \geq 0 \text{ และ } i,j \geq 0$$

บทพิสูจน์ จะได้ว่า

$$\begin{aligned} p_{ij}^{n+m} &= P\left(X_{n+m} = j \mid X_0 = i\right) \\ &= \sum_{k=0}^{\infty} P\left(X_{n+m} = j, X_n = k \mid X_0 = i\right) \\ &= \sum_{k=0}^{\infty} P\left(X_{n+m} = j \mid X_n = k, X_0 = i\right) \cdot P\left(X_n = k \mid X_0 = i\right) \\ &= \sum_{k=0}^{\infty} P\left(X_{n+m} = j \mid X_n = k\right) \cdot P\left(X_n = k \mid X_0 = i\right) \\ &= \sum_{k=0}^{\infty} p_{kj}^{m} \cdot p_{ik}^{n} \end{aligned}$$

ในทำนองเดียวกัน ถ้ากำหนดให้ $\mathrm{P}^{(n)}$ เป็นเมตริกซ์ของความน่าจะเป็นในการเปลี่ยนสถานะ p_{ij}^{n} จากสมการแชปแมน-โคลโมโกรอฟ จะได้ว่า

$$\mathrm{P}^{(n+m)} = \mathrm{P}^{(n)}\mathrm{P}^{(m)}$$

2.1.3 คุณสมบัติของสถานะ

คุณสมบัติของสถานะในลูกโซ่มาร์คอฟ ได้แก่

1. Irreducible

สถานะ i สามารถเข้าถึง (Accessible) จากสถานะ j ถ้า $p_{ij}^{n} > 0$ สำหรับบางค่า $n \geq 0$ ซึ่งสามารถเขียนแทนด้วย $j \leftarrow i$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สถานะ i และสถานะ j สามารถไปถึงกันได้ (Communicate) ก็ต่อเมื่อ $j \leftarrow i$ และ $i \leftarrow j$ ซึ่งสามารถเขียนแทนด้วย $i \leftrightarrow j$

<u>นิยาม</u> ลูกโซ่มาร์คอฟที่เพียงคลาส (Class) เดียว โดยที่ทุก ๆ สถานะสามารถเข้าถึงกันได้กับสถานะอื่น ๆ เราจะเรียกคุณสมบัติของสถานะในลูกโซ่มาร์คอฟนี้ว่า irreducible

2. Recurrent และ Transient

<u>นิยาม</u> กำหนดให้ f_i เป็น ความน่าจะเป็ นที่กระบวนการเริ่มต้นที่สถานะ i แล้วกลับมายังสถานะ i อีกครั้ง ดังนั้น ถ้า $f_i = 1$ แสดงว่า สถานะ i เป็น recurrent แต่ถ้า $f_i < 1$ แสดงว่า สถานะ i เป็น transient

ให้สถานะ i เป็น recurrent และ μ_{ii} เป็น ค่าคาดหวังของเวลาที่เริ่มต้นจากสถานะ i แล้วกระบวนการจะกลับมาที่สถานะ i อีกครั้ง จะได้ว่า ถ้า μ_{ii} หาค่าได้ (finite) แล้วเราจะเรียกสถานะ i ว่า positive recurrent แต่ถ้า μ_{ii} หาค่าไม่ได้ (infinite) แล้วเราจะเรียกสถานะ i ว่า null recurrent

สำหรับลูกโซ่มาร์คอฟที่มีสถานะจำนวนจำกัด (finite-state Markov chain) สถานะในลูกโซ่มาร์คอฟนี้จะเป็น positive recurrent เท่านั้น

จากบทนิยามข้างต้น มีข้อสรุปบางประการเกี่ยวกับสถานที่เป็น recurrent และ transient ได้แก่

- ถ้าสถานะ i เป็น recurrent และ $i \leftrightarrow j$ แล้ว สถานะ j จะเป็น recurrent
- ถ้าสถานะ i เป็น transient และ $i \leftrightarrow j$ แล้ว สถานะ j จะเป็น transient
- สำหรับลูกโซ่มาร์คอฟที่มีสถานะจำนวนจำกัดจะมีสถานะอย่างน้อย 1 สถานะที่เป็น recurrent
- ลูกโซ่มาร์คอฟที่มีสถานะจำนวนจำกัดและเป็น irreducible แล้ว ทุกสถานะในลูกโซ่มาร์คอฟนี้จะเป็น recurrent

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3. Aperiodic และ Periodic

<u>นิยาม</u> ให้ d เป็น ห.ร.ม.ของจำนวนเวลาที่กระบวนการจะกลับมาอยู่ในสถานะ i อีกครั้ง ถ้า $d=1$ แสดงว่า สถานะ i เป็น aperiodic แต่ถ้า $d>1$ แสดงว่า สถานะ i เป็น periodic

สำหรับลูกโซ่มาร์คอฟที่มีสถานะเป็น positive recurrent และ aperiodic แล้ว จะเรียกสถานะนั้นว่า เป็น ergodic

2.1.4 ทฤษฎีลูกโซ่มาร์คอฟสำหรับสเปซทั่วไป

สำหรับลูกโซ่มาร์คอฟบนสเปซ $(S,\Im)$ และ $\{X_n; n=0,1,2,...\}$ เป็น ลูกโซ่มาร์คอฟ ถ้า

$$\begin{aligned} P\left(X_{n+1}\in A \mid X_n, X_{n-1},..., X_0\right) &= P\left(X_{n+1}\in A \mid X_n\right) \\ &= P(X_n, A) \qquad \text{สำหรับ } A\in\Im \end{aligned}$$

ซึ่ง $P(X_n, A)$ มีคุณสมบัติดังนี้

1. สำหรับแต่ละ $x\in\Im$ ฟังก์ชัน $P(x,\cdot)$ เมเชอร์ความน่าจะเป็น (measure probability) บน $(S,\Im)$
2. สำหรับแต่ละ $A\in\Im$ ฟังก์ชัน $P(\cdot,A)$ สามารถวัดได้ (measurable)

<u>นิยาม</u> ทฤษฎีบทจำกัด

$$\mu P(A)=\int_S P(x,A)\mu(dx)$$

เมื่อกำหนดความน่าจะเป็นในการเปลี่ยนสถานะขั้นที่ 1 คือ

$$P^1(x,A)=P(x,A)$$

และ ความน่าจะเป็นในการเปลี่ยนสถานะขั้นที่ n คือ

$$P^n(x,A)=\int_S P^{n-1}(y,A)P(x,dy)$$

ดังนั้น

$$P(X_n\in A)=\mu P^n(A)$$

เมเชอร์ความน่าจะเป็น π บน $(S,\Im)$ เป็นการแจกแจงคงตัว (stationary distribution) ของ P ถ้า

$$\pi P=P$$

ให้ φ เป็น $\sigma-$เมเชอร์ที่หาค่าได้ (finite measure) บน $(S,\Im)$ P เป็น φ ที่ irreducible ถ้า n หาค่าได้แล้ว $P^n(x,A)>0$ สำหรับทุกค่า $x\in S$ และทุก $A\subset S$ โดยที่ $\varphi(A)>0$

P เป็น periodic ถ้าจำนวนเต็ม $d\geq 2$ และลำดับ $\{E_0,E_1,...,E_{d-1}\}$ ของ d ที่เป็นสับเซตที่ไม่ว่าง (nonempty subsets) ของ S จะได้ว่า สำหรับทุกค่า $i=0,1,...,d-1$ และทุกค่า $x\in E_i$ แล้ว

$$P(x,E_j)=1 \text{ เมื่อ } j=(i+1)\bmod d$$

ถ้านอกเหนือจากกรณีดังกล่าว P เป็น aperiodic

2.1.5 ทฤษฎีบทการลู่เข้าของลูกโซ่มาร์คอฟในกรณี total variation (Markov Chain Convergence in Total Variation)

ให้ P เป็น φ ที่ irreducible สำหรับบางค่า $\sigma-$เมเชอร์ที่หาค่าได้ φ บน $(S,\Im)$ ถ้า P เป็น aperiodic แล้ว

$$\lim_{n\to\infty}P\left(X_n\in A\,\middle|\,X_0=x\right)=\pi(A) \text{ สำหรับ ทุก } A\in\Im$$

เมื่อ π เป็น การแจกแจงคงตัว (stationary distribution) ของ P ดังนั้น

$$\pi(\cdot)=\pi P(\cdot)=\int_S P(x,\cdot)\pi(dx)$$

1. The First Lyapunov Condition (FLC)

สำหรับเซตไม่ว่าง $B \subset S$ สเกลาร์ที่เป็นบวก (positive scalars) $a < 1$ b, δ และจำนวนเต็ม $m \geq 1$ การแจกแจงความน่าจะเป็น φ บน S และฟังก์ชัน $V : S \to [1, \infty)$ จะได้ว่า

1. $P\left(X_m \in \cdot \mid X_0 \in z\right) \geq \delta\varphi(\cdot)$ สำหรับทุก $z \in B$

2. $E\left(V(X_1) \mid X_0 = z\right) \leq aV(z) + bI(z \in B)$ สำหรับทุก $z \in S$

หมายความว่า ถ้าลูกโซ่มาร์คอฟเป็นไปตามเงื่อนไข FLC และมีคุณสมบัติ aperiodic แล้วลูกโซ่มาร์คอฟนี้เป็น ergodic อย่างสม่ำเสมอ (Uniformly ergodic)

2. Markov Chain Strong Law of Large Numbers (MCSLLN)

ให้ $\{X_n; n = 0, 1, 2, ...\}$ เป็นลูกโซ่มาร์คอฟที่ ergodic อย่างสม่ำเสมอ (uniformly ergodic Markov chain) บนสเปซ S ที่มี π เป็นการแจกแจงความน่าจะเป็นคงตัว (stationary probability distribution) และฟังก์ชัน $h : S \to \Re$ ถ้า $\pi|h| = E_\pi\left[|h(X_0)|\right] = \int_S |h(x)|\pi(dx) < \infty$ แล้ว

$$\frac{1}{n}\sum_{i=0}^{n-1} h(X_i) \to \pi h = \int_S h(x)\pi(dx) \quad \text{เมื่อ } n \to \infty$$

3. Markov Chain Central Limit Theorem (MCCLT)

ให้ $\{X_n; n = 0, 1, 2, ...\}$ เป็นลูกโซ่มาร์คอฟที่ ergodic อย่างสม่ำเสมอ (uniformly ergodic Markov chain) บนสเปซ S และฟังก์ชัน $h : S \to \Re$ ด้วยค่า $h^2(z) \leq cV(z)$ สำหรับบางค่า $c > 0$ และทุกค่า z จะได้ว่า

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=0}^{n-1} h(X_i) - \pi h\right) \Rightarrow \sigma N(0,1)$$

เมื่อ π เป็นการแจกแจงความน่าจะเป็นคงตัว (stationary probability distribution) ของ X และ

$$\sigma^2 = Var_\pi\left[h(X_0)\right] + 2\sum_{k=1}^{\infty}\text{cov}_\pi\left[h(X_0), h(X_k)\right]$$

2.1.6 Reversibility

ถ้าลูกโซ่มาร์คอฟใดมีคุณสมบัติ reversibility แล้ว จะสามารถออกแบบเมตริกซ์ของความน่าจะเป็นในการเปลี่ยนสถานะ (transition probability matrix) ได้ง่ายขึ้น

<u>ทฤษฎีบท</u> ถ้าความน่าจะเป็นในการเปลี่ยนสถานะ P มีคุณสมบัติ reversibility ตามความน่าจะเป็น π แล้ว π จะเป็นการแจกแจงคงตัวของ P

ให้ $A \subset S$ แล้ว

$$\begin{aligned} \pi P(A) &= \int_S P(x, A)\pi(dx) \\ &= \int_A P(x, S)\pi(dx) \\ &= \int_A 1\,\pi(dx) \\ &= \pi(A) \end{aligned}$$

2.1.7 การสุ่มตัวอย่างแบบฮิตแอนด์รัน (Hit-and-run Sampler)

ลูกโซ่มาร์คอฟที่ได้จากการสร้างขึ้น (generate) ด้วยวิธีฮิตแอนด์รัน จะลู่เข้าสู่การแจกแจงแบบสม่ำเสมอ (uniform distribution) บนบริเวณ S

ให้บริเวณ S เป็นสับเซตของ $\Re^n$

$n = 0$

และ $X_n = x_0 \in S$

มีระเบียบวิธี (algorithm) ดังนี้

1. ให้ $X_n = x \in S$ สุ่มทิศทาง d บนพื้นผิวของ S สร้างส่วนของเส้นตรง L ให้ผ่านจุด x และอินเตอร์เซคกับ S โดยที่

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

$$L_n = \{ l \mid l = x + \lambda d, \quad \lambda \in \Re, l \in S \}$$

2. สุ่มจุด X_{n+1} อย่างสม่ำเสมอบนส่วนของเส้นตรง L

3. ให้ $n = n + 1$ แล้วกลับไปดำเนินการตามข้อ 1

2.1.8 การวิเคราะห์สถานะเสถียรภาพของการจำลอง (Steady State Analysis of a Simulation)

ในการจำลองกระบวนการเฟ้นสุ่มที่เป็นลูกโซ่มาร์คอฟ $\{X_n; n = 0, 1, 2, ...\}$ เมื่อ X_n เป็นเวกเตอร์ของตัวแปรสุ่ม ซึ่ง $X_n \in S$ และ S เป็นสเปซของสถานะ (state space) โดยเงื่อนไขทั่วไป การแจกแจงของ X_n จะลู่เข้าสู่การแจกแจงคงตัว π และเป็นอิสระกับจุดเริ่มต้น x_0 ดังนี้

$$P(X_n \in A \mid X_0 = x_0) \to \pi(A)$$

จาก MCSLLN จะได้ว่า ค่าเฉลี่ยในระยะยาว (long run average) จะสามารถประมาณค่า πh ดังนี้

$$\frac{1}{N}\sum_{i=0}^{N-1} h(X_i) \to \pi h \quad \text{สำหรับ } N \text{ ที่มีค่ามาก}$$

จาก MCCLT จะได้ว่า การแจกแจงของค่าเฉลี่ยในระยะยาว (distribution of long run average) จะประมาณ (approximate) ด้วยการแจกแจงแบบปกติ (normal distribution) ดังนี้

$$\frac{1}{N}\sum_{i=0}^{N-1} h(X_i) \sim N(\pi h, \sigma^2) \quad \text{สำหรับ } N \text{ ที่มีค่ามาก}$$

เมื่อ $\sigma^2 = Var_\pi[h(X_0)] + 2\sum_{k=1}^{\infty} \mathrm{cov}_\pi[h(X_0), h(X_k)]$

จากการลู่เข้าของลูกโซ่มาร์คอฟจึงมีค่าความแปรปรวนร่วม (covariance) เพราะ X_i แต่ละตัวไม่เป็นอิสระกัน ดังนั้น งานวิจัยนี้จึงใช้ Batch Means สำหรับแก้ปัญหาค่าความแปรปรวนร่วม

2.1.9 วิธีค่าเฉลี่ยแบบกลุ่ม (Batch Means)

เป็นวิธีการหนึ่งในการประมาณค่าเฉลี่ยของประชากร (μ) ด้วยค่าเฉลี่ยแบบกลุ่ม $(\bar{Y})$ เมื่อลำดับ $\{X_n\}$ มีความสัมพันธ์กัน

มีขั้นตอนดังต่อไปนี้

1) กำหนดขนาดตัวอย่างในแต่ละกลุ่ม (b)

2) หาจำนวนกลุ่ม (m) จาก $m = \frac{n}{b}$

3) หาค่าเฉลี่ยของตัวอย่างในกลุ่มที่ k ที่มีขนาดของกลุ่มเท่ากับ b จาก

$$Y_k = \frac{\sum_{i=(k-1)b+1}^{kb} h(X_i)}{b}$$

เมื่อ ตัวอย่างของกลุ่มที่ k คือ $X_{(k-1)b+1},\ldots, X_{kb}$

4) จะได้ $\bar{Y} = \frac{\sum_{i=1}^{m} Y_i}{m}$ และ $S_Y^2 = \frac{\sum_{i=1}^{m} (Y_i - \bar{Y})^2}{m-1}$

5) เนื่องจากในงานวิจัยนี้ไม่ทราบค่าเฉลี่ยของบริเวณสามเหลี่ยม จึงศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันของบริเวณสามเหลี่ยมจนกระทั่ง $t_{\frac{\alpha}{2},(m-1)} \, {S_Y}/{\sqrt{m}}$ ที่แทนด้วย SE^* มีค่าน้อยกว่าความคลาดเคลื่อนที่กำหนด (e)

วิธีค่าเฉลี่ยแบบกลุ่มนี้จะประมาณช่ว งความเชื่อมั่นของ πh ด้วย $\bar{Y} \pm t_{\frac{\alpha}{2},(m-1)} \, {S_Y}/{\sqrt{m}}$ จากทฤษฎี MCCLT จะได้ว่า เมื่อ $n \to \infty$ และจำนวนกลุ่มของข้อมูล (m) เป็นค่าคงที่แล้ว ค่าเฉลี่ยแบบกลุ่มทั้งหมดจะมีการแจกแจงแบบปกติเมื่อใกล้อนันต์ (asymptotically normally distribution) ด้วยค่าเฉลี่ย πh และค่าความแปรปรวน ${m\sigma^2}/{n}$

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

ลาสซ์โล ลอวอสซ์ (Laszlo Lovasz, 1999) ได้ศึกษาการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณรูปร่างคอนเวกซ์ใน p มิติ ซึ่งเป็นการสุ่มตัวอย่างแบบฮิตแอนด์รันในกรณีที่แย่ที่สุดเท่าที่จะเป็นไปได้ (total variation) พบว่า ความซับซ้อนในการคำนวณเป็นฟังก์ชันพหุนามอันดับที่ 3 หรือ $O^*(p^3)$ โดยมีค่าสัมประสิทธิ์ในพหุนามอันดับที่ 3 คือ 4×10^8

ลาสซ์โล ลอวอสซ์ และซานโตส เวมพาลา (Laszlo Lovasz and Santosh Vempala, 2003) ได้ศึกษาการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ จัดว่าเป็นวิธีการที่เร็วที่สุดในการจำลองจุดตัวอย่างบนบริเวณคอนเวกซ์หลายมิติ โดยการสุ่มตัวอย่างแบบฮิตแอนด์รันนี้จะจำลองจุดตัว อย่างมีการแจกแจงที่ถูกต้องโดยประมาณ (approximately the right distribution) ด้วยจำนวน $O^*(p^3)$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ที่มี ขอบเขตจำกัด ในบทนี้จึงเป็นการนำเสนอวิธีดำเนินการวิจัยซึ่งผู้วิจัยได้ดำเนินการตามขั้นตอนต่อไปนี้

1. ศึกษาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยมS_{rect}
2. ศึกษาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมS_{tri}
3. ศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย(n_{av}) กับจำนวนมิติ(p) ของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด ณ ระดับความคลาดเคลื่อน(e) ต่าง ๆ
4. สรุปผลการวิจัย

โดยขั้นตอนที่ 1-3 มีรายละเอียดของการดำเนินการดังนี้

3.1 การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect}

เป็นกรณีที่ทราบค่าเฉลี่ยของประชากร คือ จุดกำเนิด ซึ่งกำหนดให้

k แทน จำนวนรอบ

n_k แทน ขนาดตัวอย่างที่สุ่มด้วยวิธีฮิตแอนด์รันในรอบที่k

p แทน จำนวนมิติของบริเวณสี่เหลี่ยมS_{rect}

e แทน ความคลาดเคลื่อน

X_0 แทน จุดเริ่มต้นในการสุ่มตัวอย่างแบบฮิตแอนด์รันซึ่งเป็นจุดใดๆ บนบริเวณสี่เหลี่ยม S_{rect}

n_r แทน ขนาดตัวอย่างที่สุ่มด้วยวิธีฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน $[-e,e]$ ของ S_{rect} ใน p มิติของข้อมูลชุดที่ r

r แทน ลำดับของชุดข้อมูลของ n_r ซึ่ง $r = 1,2,...,30$

n_{av} แทน ขนาดตัวอย่างเฉลี่ยจากการเก็บค่าn_r จำนวน 30 ชุด

มีขั้นตอนดังต่อไปนี้

1) กำหนด p , e และ X_0
2) ให้ $r = 1$
3) ให้ $k = 1$ กับ $n_k = 1,900$ และสุ่มด้วยวิธีฮิตแอนด์รันขนาด n_k
4) ให้ $n = n_k$ คำนวณค่าเฉลี่ย

$$\bar{X}_{.j} = \frac{\sum_{i=1,001}^{n} X_{ij}}{n} \quad \text{เมื่อ} \quad j = 1,2,..., p$$

5) ถ้า $\bar{X}_{.j} \in [-e,e]$ สำหรับทุกค่า $j = 1,2,..., p$ แล้ว หยุดการสุ่มตัวอย่างและไปดำเนินการตามขั้นตอนที่ 6)

ในทางกลับกัน ถ้า $\bar{X}_{.j} \notin [-e,e]$ สำหรับบางค่า $j = 1,2,..., p$ แล้ว ให้ดำเนินการดังนี้

5.1) ให้ $k = k+1$ และ $n_k = 900$

5.2) สุ่มด้วยวิธีฮิตแอนด์รันขนาด n_k

5.3) ให้ $n = \sum_{l=1}^{k} n_l$ คำนวณค่าเฉลี่ย

$$\bar{X}_{.j} = \frac{\sum_{i=1,001}^{n} X_{ij}}{n} \quad \text{เมื่อ} \quad j = 1,2,..., p$$

5.4) ตรวจสอบว่า $\bar{X}_{.j} \notin [-e,e]$ สำหรับบางค่า $j = 1,2,..., p$ หรือไม่ ถ้าเป็นเช่นนั้นให้กลับไปดำเนินการตามขั้นตอนที่ 5.1) จนกระทั่ง $\bar{X}_{.j} \in [-e,e]$ สำหรับทุกค่า $j = 1,2,..., p$ จึงหยุดการสุ่มตัวอย่างและไปดำเนินการตามขั้นตอนที่ 6)

6) ให้ $n_r = n$ และ $r = r+1$ แล้วดำเนินการตามขั้นตอนที่ 2) จนกระทั่ง $r = 30$

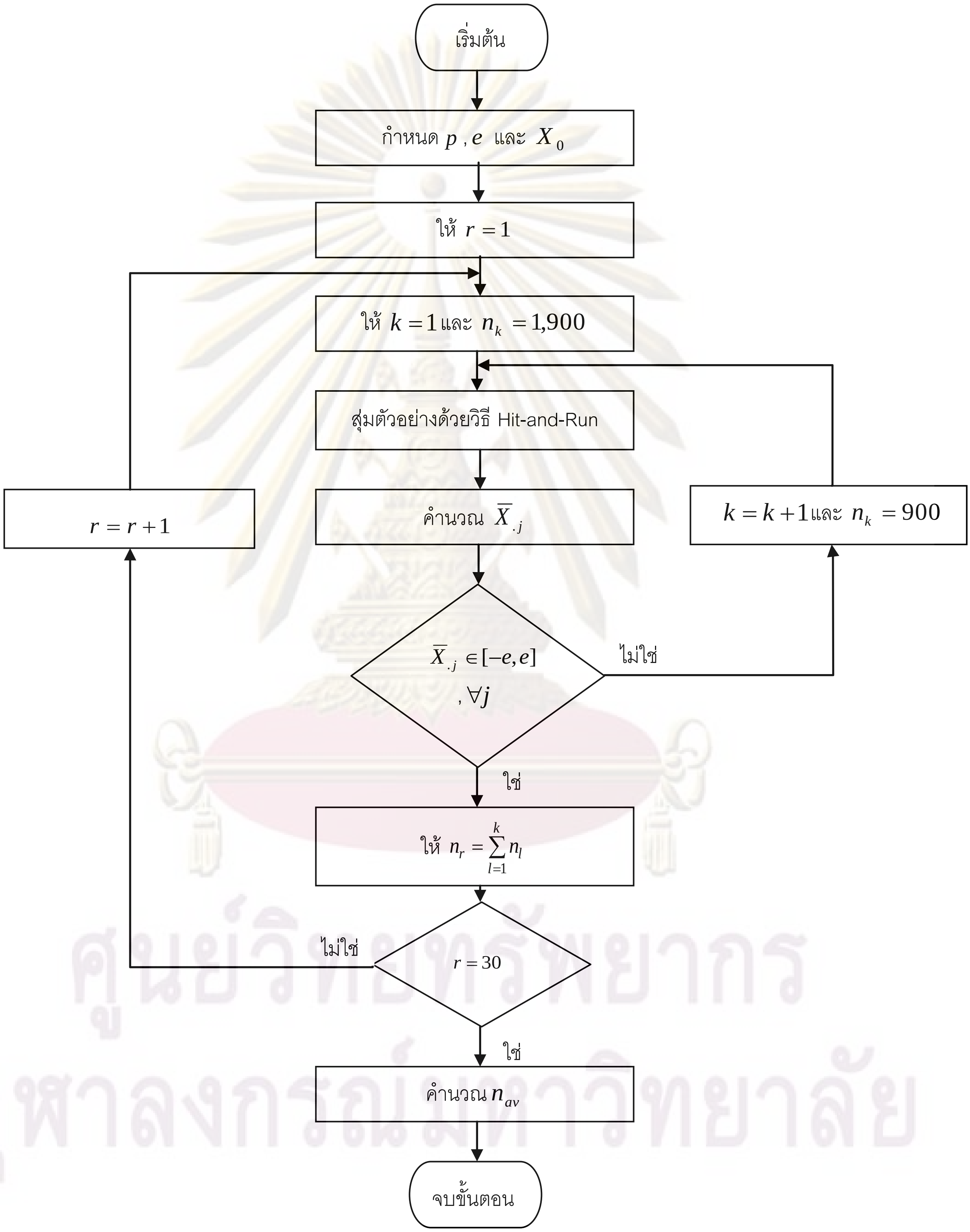
7) คำนวณขนาดตัวอย่างเฉลี่ย

$$n_{av} = \frac{\sum_{r=1}^{30} n_r}{30}$$

ขั้นตอนดังกล่าวสามารถในรูปที่ 3.1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 3.1 แสดงขั้นตอนการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect}



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.2 การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมS_{tri}

เป็นกรณีที่ไม่ทราบค่าเฉลี่ยของประชากร จึงศึกษาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันของบริเวณสามเหลี่ยม S_{tri} จนกระทั่ง $SE^* < e$ ซึ่งกำหนดให้

k แทน จำนวนรอบ

n_k แทน ขนาดตัวอย่างที่สุ่มด้วยวิธีฮิตแอนด์รันในรอบที่k

p แทน จำนวนมิติของบริเวณสามเหลี่ยมS_{tri}

e แทน ความคลาดเคลื่อน

X_0 แทน จุดเริ่มต้นในการสุ่มตัวอย่างแบบฮิตแอนด์รันซึ่งเป็นจุดใดๆ บนบริเวณสามเหลี่ยมS_{rect}

n_r แทน ขนาดตัวอย่างที่สุ่มด้วยวิธีฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน $[-e,e]$ ของ S_{rect} ใน p มิติของข้อมูลชุดที่ r

r แทน ลำดับของชุดข้อมูลของ n_r ซึ่ง $r = 1,2,...,30$

n_{av} แทน ขนาดตัวอย่างเฉลี่ยจากการเก็บค่าn_r จำนวน 30 ชุด

Y_i แทน ค่าเฉลี่ยของตัวอย่างใน batch ที่ i ซึ่ง $i = 1,2,..., m$

$$Y_i = \frac{\sum_{1+(i-1)b}^{ib} X_i}{b}$$

เมื่อ b แทน ขนาดตัวอย่างสุ่มในแต่ละ batch

$\hat{\rho}$ แทน ค่าสัมประสิทธิ์สหสัมพันธ์ของ Y

$$\hat{\rho} = \frac{\sum_{i=1}^{m-1}(Y_i - \bar{Y})(Y_{i+1} - \bar{Y})}{\sum_{i=1}^{m}(Y_i - \bar{Y})^2}$$

มีขั้นตอนดังต่อไปนี้

1) กำหนด p , e , m และ X_0
2) ให้ $r = 1$
3) ให้ $k = 1$ กับ $n_k = 1,900$ และสุ่มด้วยวิธีฮิตแอนด์รันขนาดn_k
4) คำนวณค่า$\hat{\rho}$
5) ทดสอบสมมติฐาน $H_0 : \rho = 0$ และ$H_1 : \rho \neq 0$ ถ้าไม่สามารถปฏิเสธ H_0 ให้ไปดำเนินการตามขั้นตอนที่ 6) ในทางกลับกันถ้าปฏิเสธH_0 ให้ดำเนินการดังนี้

    5.1) ให้ $k = k+1$ และ$n_k = 900$

5.2) สุ่มด้วยวิธีฮิตแอนด์รันขนาด n_k

5.3) คำนวณค่า $\hat{\rho}$

5.4) ทดสอบสมมติฐาน $H_0 : \rho = 0$ และ $H_1 : \rho \neq 0$ ถ้าไม่สามารถปฏิเสธ H_0 ให้ไปดำเนินการตามขั้นตอนที่ 6) ในทางกลับกันถ้าปฏิเสธ H_0 ให้ดำเนินการตามขั้นตอนที่ 5.1) จนกระทั่งไม่สามารถปฏิเสธ H_0

6) ให้ $n_1 = \sum_{l=1}^{k} n_l$, $k^* = 1$ กับ $n_{k^*} = n_1 - 1{,}000$ และสุ่มด้วยวิธีฮิตแอนด์รันขนาด n_1

7) คำนวณค่า SE^*

8) ตรวจสอบว่า $SE^* < e$ หรือไม่ ถ้าเป็นเช่นนั้นจึงหยุดการสุ่มตัวอย่างและไปดำเนินการตามขั้นตอนที่ 9) แต่ถ้า $SE^* \geq e$ ให้ดำเนินการดังนี้

8.1) ให้ $k^* = k^* + 1$ และสุ่มด้วยวิธีฮิตแอนด์รันขนาด n_{k^*}

8.2) คำนวณค่า SE^*

8.3) ตรวจสอบว่า $SE^* < e$ หรือไม่ ถ้าเป็นเช่นนั้นจึงหยุดการสุ่มตัวอย่างและไปดำเนินการตามขั้นตอนที่ 9) แต่ถ้า $SE^* \geq e$ ให้ดำเนินการตามขั้นตอนที่ 8.1) จนกระทั่ง $SE^* < e$

9) ให้ $n_r = n_1 + \sum_{l=1}^{k^*} n_l$ และ และ $r = r + 1$ แล้วดำเนินการตามขั้นตอนที่ 2) จนกระทั่ง $r = 30$

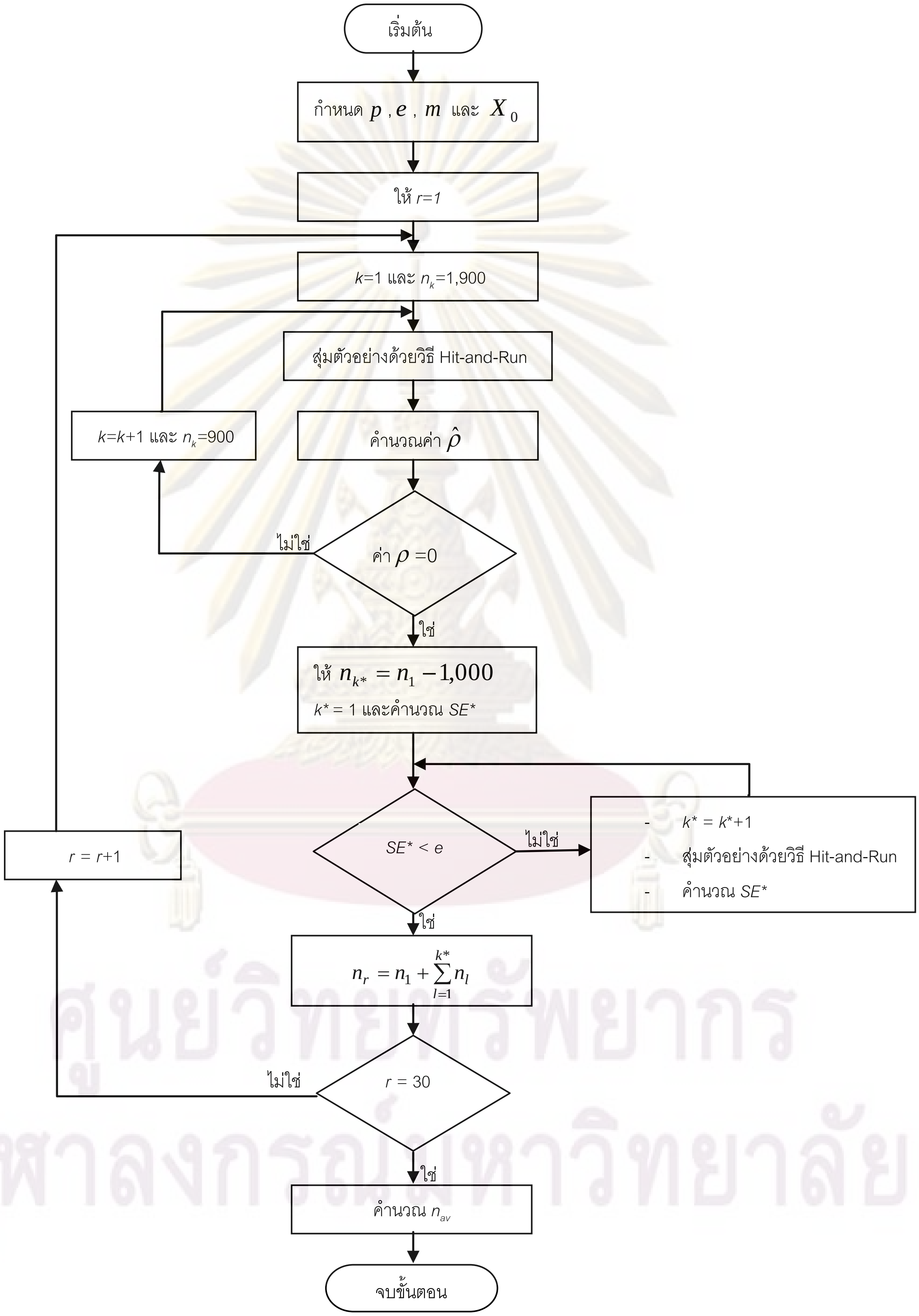
10) คำนวณขนาดตัวอย่างเฉลี่ย

$$n_{av} = \frac{\sum_{r=1}^{30} n_r}{30}$$

ขั้นตอนดังกล่าวสามารถในรูปที่ 3.2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 3.2 แสดงขั้นตอนการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยมS_{tri}



3.3 ศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย(n_{av}) กับจำนวนมิติ(p) ของบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด ณ ระดับความคลาดเคลื่อน(e) ต่าง ๆ

ในการวิจัยนี้ได้ศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) กับ จำนวนมิติ (p) ของบริเวณสี่เหลี่ยม S_{rect} และบริเวณสามเหลี่ยม S_{tri} ณ ระดับความคลาดเคลื่อน (e) ได้แก่ 0.1 0.05 และ 0.01 เมื่อกำหนดให้ ขนาดตัวอย่างเฉลี่ย (n_{av}) เป็น ตัวแปรตาม (dependent variable) และจำนวนมิติ (p) เป็น ตัวแปรอิสระ (independent variable) ซึ่งมีขั้นตอนดังต่อไปนี้

1. ศึกษาความสัมพันธ์ของ n_{av} กับ p ด้วยตัวแบบความถดถอยพหุนาม (Polynomial Regression Model) อันดับที่ 1 ถึง 3

2. ทดสอบสมมติฐานเชิงสถิติเพื่อคัดเลือกตัวแบบความถดถอยเชิงเส้นแบบติดกลุ่ม เมื่อสมมติให้ตัวแบบความถดถอยแบบติดกลุ่มทั้งสองตัวแบบเป็นดังนี้

$$M_1: \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_p X_{pi} + \varepsilon_i$$
$$M_2: \quad Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_g X_{gi} + \varepsilon_i \quad ; g < 3$$

ตัวแบบความถดถอย M_2 ติดกลุ่มอยู่ในตัวแบบความถดถอย M_1 โดยเรียกตัวแบบ M_1 ว่า ตัวแบบความถดถอยเต็มรูป (Full Model) และเรียกตัวแบบ M_2 ว่า ตัวแบบความถดถอยลดรูป (Reduced Model) และการทดสอบสมมติฐานเชิงสถิติทำได้ดังนี้

สมมติฐานคือ

$$H_0: \ \beta_{(g+1)} = \beta_{(g+2)} = ... = \beta_p = 0$$
$$H_1: \ \text{not } H_0$$

ตัวสถิติทดสอบ คือ

$$F_c = \frac{(SSE_R - SSE_F)/(p-g)}{SSE_F/[n-(p+1)]} \sim F_{(p-g),[n-(p+1)]}$$

โดยที่

SSE_R แทน ผลบวกกำลังสองของความคลาดเคลื่อนของตัวแบบความถดถอยลดรูป

SSE_F แทน ผลบวกกำลังสองของความคลาดเคลื่อนของตัวแบบความถดถอยเต็มรูป

g แทน จำนวนตัวแปรอิสระของตัวแบบความถดถอยลดรูป

p แทน จำนวนตัวแปรอิสระของตัวแบบความถดถอยเต็มรูป ซึ่งในที่นี้เท่ากับ 3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

n แทน จำนวนรายการของข้อมูลทั้งหมด ซึ่งในที่นี้เท่ากับ 15

ในการวิเคราะห์หาตัวแบบความถดถอยพหุนามสำหรับความสัมพันธ์ของ n_{av} กับ p นี้ด้วยโปรแกรม R

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิจัย

ในงานวิจัยนี้ได้ศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} และการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} พร้อมทั้งศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ในแต่ละบริเวณ ณ ระดับความคลาดเคลื่อน (e) เท่ากับ 0.1 0.05 และ 0.01 ผลการวิเคราะห์มีรายละเอียด ดังนี้

4.1 ผลการศึกษาอัตราการ ลู่เข้า ของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม

ในการวิจัยนี้ได้กำหนดให้การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} เป็นการสุ่มตัวอย่างในกรณีที่ทราบค่าเฉลี่ยของประชากร ซึ่งในที่นี้คือ จุดกำเนิด โดยกำหนดให้ขนาดตัวอย่างเฉลี่ย (n_{av}) จนกระทั่งกระบวนการลู่เข้าสู่จุดกำเ นิดภายในความคลาดเคลื่อน $[-e,e]$ เป็นค่าวัดอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน สามารถแสดงผลการศึกษาโดยจำแนกตามระดับความคลาดเคลื่อนได้ดังนี้

4.1.1 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} ณ ระดับความคลาดเคลื่อนเท่ากับ 0.1 (e = 0.1)

สำหรับการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} โดยวัดจากขนาดตัวอย่างที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่งกระบวนการ ลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ต่าง ๆ ดังแสดงในตารางที่ 4.1 – 4.3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.1 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
1	1,900	1,900	2,800	1,900	4,600
2	1,900	1,900	1,900	1,900	3,700
3	1,900	1,900	1,900	1,900	4,600
4	1,900	1,900	1,900	1,900	5,500
5	1,900	1,900	1,900	2,800	1,900
6	1,900	1,900	1,900	1,900	2,800
7	1,900	1,900	1,900	1,900	4,600
8	1,900	1,900	1,900	2,800	1,900
9	1,900	1,900	2,800	1,900	3,700
10	1,900	1,900	1,900	2,800	1,900
11	1,900	1,900	1,900	1,900	2,800
12	1,900	1,900	1,900	1,900	2,800
13	1,900	1,900	2,800	1,900	3,700
14	1,900	1,900	3,700	5,500	2,800
15	1,900	1,900	1,900	1,900	5,500
16	1,900	1,900	1,900	2,800	3,700
17	1,900	1,900	1,900	2,800	1,900
18	1,900	1,900	1,900	2,800	2,800
19	1,900	1,900	1,900	2,800	3,700
20	1,900	1,900	2,800	3,700	1,900
21	1,900	1,900	2,800	2,800	1,900
22	1,900	1,900	1,900	1,900	2,800
23	1,900	1,900	1,900	1,900	1,900
24	1,900	1,900	1,900	1,900	1,900
25	1,900	1,900	1,900	2,800	1,900

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.1 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
26	1,900	1,900	2,800	2,800	4,600
27	1,900	1,900	2,800	3,700	2,800
28	1,900	1,900	3,700	4,600	1,900
29	1,900	1,900	1,900	1,900	1,900
30	1,900	1,900	1,900	2,800	1,900
n_{av}	**1,900**	**1,900**	**2,230**	**2,560**	**3,010**

ตารางที่ 4.2 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
1	5,500	1,900	4,600	4,600	9,100
2	2,800	3,700	6,400	2,800	14,500
3	3,700	3,700	4,600	4,600	5,500
4	3,700	7,300	5,500	6,400	18,100
5	4,600	5,500	2,800	1,900	17,200
6	1,900	1,900	5,500	10,000	17,200
7	3,700	2,800	4,600	4,600	9,100
8	2,800	3,700	4,600	5,500	7,300
9	9,100	6,400	1,900	7,300	10,900
10	3,700	1,900	2,800	6,400	17,200
11	3,700	2,800	2,800	6,400	9,100
12	4,600	2,800	4,600	4,600	6,400
13	2,800	6,400	4,600	3,700	10,900

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.2 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
14	2,800	1,900	1,900	2,800	5,500
15	5,500	4,600	1,900	6,400	12,700
16	2,800	4,600	6,400	8,200	10,000
17	5,500	2,800	3,700	4,600	13,600
18	3,700	5,500	3,700	7,300	21,700
19	3,700	4,600	4,600	3,700	13,600
20	2,800	1,900	6,400	2,800	11,800
21	5,500	2,800	4,600	4,600	16,300
22	5,500	6,400	2,800	9,100	10,900
23	4,600	5,500	5,500	8,200	9,100
24	2,800	6,400	8,200	2,800	10,900
25	5,500	2,800	6,400	2,800	6,400
26	2,800	5,500	2,800	1,900	11,800
27	2,800	5,500	3,700	10,000	5,500
28	3,700	1,900	3,700	10,000	18,100
29	3,700	4,600	2,800	8,200	10,000
30	1,900	3,700	3,700	11,800	13,600
n_{av}	**3,910**	**4,060**	**4,300**	**5,890**	**11,980**

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.3 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
1	9,100	23,500	38,800	76,600	84,700
2	19,000	27,100	63,100	63,100	94,600
3	11,800	21,700	46,900	67,600	85,600
4	24,400	11,800	35,200	98,200	91,900
5	28,900	58,600	19,900	88,300	73,900
6	27,100	19,900	35,200	118,000	137,800
7	11,800	35,200	62,200	75,700	73,000
8	36,100	30,700	46,000	59,500	70,300
9	11,800	28,900	77,500	61,300	118,900
10	33,400	23,500	43,300	96,400	73,000
11	10,000	31,600	46,000	81,100	115,300
12	19,900	22,600	49,600	82,000	174,700
13	27,100	41,500	50,500	59,500	133,300
14	23,500	28,000	50,500	100,900	144,100
15	17,200	41,500	64,900	46,900	105,400
16	14,500	21,700	28,000	50,500	162,100
17	13,600	51,400	22,600	68,500	144,100
18	19,000	32,500	35,200	53,200	120,700
19	17,200	26,200	70,300	84,700	119,800
20	34,300	67,600	31,600	89,200	102,700
21	9,100	25,300	30,700	55,000	209,800
22	16,300	35,200	23,500	91,000	206,200
23	8,200	24,400	44,200	91,000	67,600
24	14,500	16,300	44,200	65,800	127,000
25	19,000	34,300	39,700	58,600	90,100

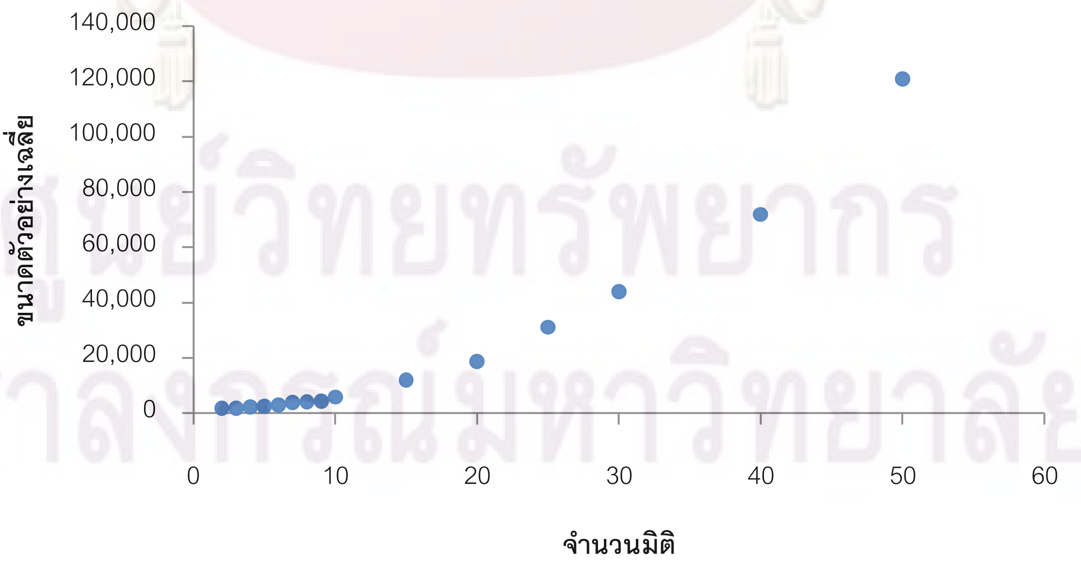
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.3 (ต่อ) แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
26	14,500	28,000	39,700	44,200	92,800
27	18,100	25,300	23,500	60,400	204,400
28	10,900	27,100	37,900	55,000	86,500
29	27,100	40,600	38,800	51,400	181,900
30	16,300	31,600	77,500	63,100	134,200
n_{av}	**18,790**	**31,120**	**43,900**	**71,890**	**120,880**

จากตารางที่ 4.1 - 4.3 พบว่า ขนาดตัวอย่างที่สุ่มได้จนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] มีขนาดที่เพิ่มขึ้น เมื่อจำนวนมิติ (p) ที่ต้องการสุ่มแบบฮิตแอนด์รันมีจำนวนเพิ่มขึ้น ด้วยเหตุนี้จึงศึกษาความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ทำให้ได้ผลดังแสดงในรูปที่ 4.1

รูปที่ 4.1 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] กับจำนวนมิติ (p)



จากรูปที่ 4.1 พบว่า ลักษณะความสัมพันธ์ระหว่าง n_{av} กับ p เป็นเส้นโค้งที่ n_{av} มีขนาดเพิ่มขึ้นอย่างรวดเร็ว เมื่อ p มีจำนวนเพิ่มขึ้น ดังนั้น ตัวแบบความถดถอยที่จะอธิบายความสัมพันธ์ดังกล่าว คือ ตัวแบบความถดถอยพหุนาม เมื่อวิเคราะห์ความถดถอยพหุนามอันดับที่ 1 ถึง 3 ของความสัมพันธ์ระหว่าง n_{av} กับ p ด้วยโปรแกรม R ได้ผลดังนี้

- ตัวแบบที่ 1 คือ $n_{av} = -12{,}908.40 + 2{,}230.50p$ ซึ่งมีค่าสถิติ $F = 153.82$ ค่า $p-value = 1.4106 \times 10^{-8}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9221$
- ตัวแบบที่ 2 คือ $n_{av} = 2{,}068.67 - 130.95p + 49.43p^2$ ซึ่งมีค่าสถิติ $F = 3{,}947$ ค่า $p-value < 2.2 \times 10^{-16}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9985$
- ตัวแบบที่ 3 คือ $n_{av} = 517.12 + 258.76p + 29.40p^2 + 0.26p^3$ ซึ่งมีค่าสถิติ $F = 3{,}069$ ค่า $p-value = 2.36 \times 10^{-16}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9988$

จากตัวแบบทั้งสามข้างต้น สามารถทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 1 กับตัวแบบที่ 2 ดังนี้

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 704.35$ ด้วยค่า $p-value = 2.52 \times 10^{-11}$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 2

เมื่อทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 2 กับตัวแบบที่ 3 ดังนี้

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 2.99$ ด้วยค่า $p-value = 0.1117$ จึงยอมรับสมมติฐาน H_0

ดังนั้น ตัวแบบความถดถอยพหุนามที่เหมาะสมสำหรับอธิบายความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.1, 0.1] กับจำนวนมิติ (p) คือ

$$n_{av} = 2{,}068.67 - 130.95p + 49.43p^2 \text{ และ } R^2 = 0.9985$$

4.1.2 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} ณ ระดับความคลาดเคลื่อนเท่ากับ 0.05 (e = 0.05)

สำหรับการศึกษาอัตราการลู่ เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} โดยวัดจากขนาดตัวอย่างที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ต่าง ๆ ดังแสดงในตารางที่ 4.4 – 4.6

ตารางที่ 4.4 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
1	1,900	4,600	5,500	3,700	5,500
2	3,700	1,900	4,600	8,200	5,500
3	1,900	4,600	3,700	8,200	11,800
4	1,900	3,700	3,700	8,200	6,400
5	1,900	1,900	9,100	6,400	2,800
6	1,900	4,600	8,200	4,600	8,200
7	1,900	1,900	5,500	5,500	10,900
8	2,800	2,800	3,700	3,700	6,400
9	1,900	1,900	3,700	3,700	7,300
10	1,900	1,900	7,300	7,300	3,700

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.4 (ต่อ) แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
11	1,900	2,800	2,800	5,500	2,800
12	1,900	1,900	3,700	6,400	11,800
13	1,900	3,700	2,800	6,400	7,300
14	1,900	1,900	4,600	1,900	10,000
15	1,900	3,700	1,900	1,900	2,800
16	4,600	2,800	3,700	5,500	8,200
17	1,900	4,600	6,400	4,600	5,500
18	3,700	1,900	2,800	18,100	10,000
19	1,900	1,900	1,900	2,800	5,500
20	2,800	2,800	1,900	5,500	7,300
21	1,900	6,400	2,800	4,600	6,400
22	1,900	2,800	5,500	10,900	8,200
23	2,800	3,700	4,600	2,800	7,300
24	1,900	3,700	5,500	9,100	3,700
25	1,900	4,600	3,700	4,600	3,700
26	2,800	1,900	3,700	8,200	7,300
27	1,900	1,900	2,800	10,900	10,000
28	1,900	3,700	10,900	8,200	14,500
29	3,700	4,600	3,700	3,700	14,500
30	1,900	2,800	3,700	8,200	4,600
n_{av}	**2,290**	**3,130**	**4,480**	**6,310**	**7,330**

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.5 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่ เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
1	9,100	10,900	14,500	10,900	35,200
2	18,100	8,200	10,000	18,100	26,200
3	10,000	9,100	15,400	19,000	31,600
4	3,700	8,200	12,700	10,900	28,000
5	20,800	10,000	13,600	10,900	29,800
6	5,500	11,800	20,800	10,000	32,500
7	19,900	17,200	6,400	9,100	39,700
8	12,700	10,900	6,400	21,700	48,700
9	9,100	15,400	12,700	25,300	28,900
10	19,000	14,500	16,300	20,800	16,300
11	10,000	10,900	9,100	37,000	23,500
12	16,300	13,600	10,000	19,000	43,300
13	8,200	5,500	6,400	22,600	39,700
14	8,200	19,900	18,100	5,500	55,000
15	9,100	10,000	19,000	12,700	77,500
16	9,100	11,800	17,200	20,800	47,800
17	4,600	10,000	12,700	23,500	37,900
18	14,500	6,400	8,200	13,600	40,600
19	5,500	5,500	11,800	13,600	27,100
20	4,600	10,000	8,200	16,300	41,500
21	11,800	7,300	14,500	23,500	42,400
22	12,700	8,200	7,300	15,400	18,100
23	6,400	12,700	12,700	28,000	22,600
24	4,600	16,300	13,600	12,700	46,900
25	5,500	15,400	16,300	10,900	37,900

ตารางที่ 4.5 (ต่อ) แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
26	10,000	4,600	11,800	18,100	48,700
27	10,900	6,400	15,400	14,500	27,100
28	8,200	10,900	13,600	29,800	26,200
29	10,000	33,400	13,600	10,900	19,900
30	5,500	14,500	9,100	25,300	37,000
n_{av}	**10,120**	**11,650**	**12,580**	**17,680**	**35,920**

ตารางที่ 4.6 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
1	77,500	60,400	157,600	360,100	765,100
2	52,300	87,400	166,600	511,300	397,000
3	41,500	109,000	271,900	181,000	471,700
4	55,900	91,900	188,200	290,800	490,600
5	43,300	106,300	151,300	255,700	682,300
6	49,600	117,100	152,200	381,700	460,900
7	61,300	127,900	95,500	227,800	393,400
8	64,900	154,900	118,900	451,000	418,600
9	53,200	63,100	106,300	341,200	532,000
10	87,400	91,900	83,800	224,200	494,200

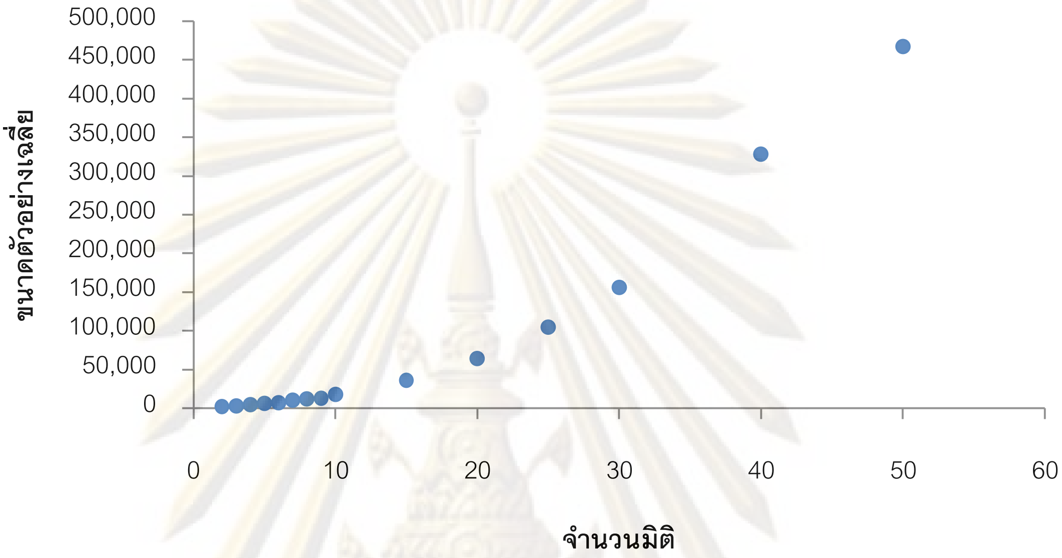
ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.6 (ต่อ) แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.05, 0.05] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
11	61,300	91,000	99,100	365,500	674,200
12	93,700	64,000	190,000	356,500	334,000
13	90,100	245,800	184,600	279,100	435,700
14	45,100	114,400	230,500	415,000	405,100
15	52,300	136,000	152,200	499,600	415,900
16	73,000	90,100	138,700	166,600	460,900
17	22,600	136,000	203,500	224,200	378,100
18	31,600	113,500	164,800	500,500	393,400
19	25,300	79,300	136,900	386,200	451,900
20	127,900	48,700	114,400	245,800	299,800
21	82,000	28,000	156,700	218,800	378,100
22	78,400	55,000	163,000	363,700	334,000
23	69,400	107,200	70,300	350,200	561,700
24	58,600	136,900	141,400	262,900	322,300
25	55,000	155,300	125,200	564,400	399,700
26	122,500	107,200	126,100	223,300	594,100
27	54,100	106,300	129,700	228,700	375,400
28	37,000	154,900	132,400	393,400	613,900
29	93,700	70,300	127,000	306,100	570,700
30	67,600	82,000	393,400	269,200	514,000
n_{av}	**64,270**	**104,393**	**155,740**	**328,150**	**467,290**

จากตารางที่ 4.4 - 4.6 พบว่า ขนาดตัวอย่างที่สุ่มได้จนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] มีขนาดที่เพิ่มขึ้น เมื่อจำนวนมิติ (p) ที่ต้องการสุ่มแบบฮิตแอนด์รันมีจำนวนเพิ่มขึ้น ด้วยเหตุนี้จึงศึกษาความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ทำให้ได้ผลดังแสดงในรูปที่ 4.2

รูปที่ 4.2 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] กับจำนวนมิติ (p)



จากรูปที่ 4.2 พบว่า ลักษณะความสัมพันธ์ระหว่าง n_{av} กับ p เป็นเส้นโค้งที่ n_{av} มีขนาดเพิ่มขึ้นอย่างรวดเร็ว เมื่อ p มีจำนวนเพิ่มขึ้น ดังนั้น ตัวแบบความถดถอยที่จะอธิบายความสัมพันธ์ดังกล่าว คือ ตัวแบบความถดถอยพหุนาม เมื่อวิเคราะห์ความถดถอยพหุนามอันดับที่ 1 ถึง 3 ของความสัมพันธ์ระหว่าง n_{av} กับ p ด้วยโปรแกรม R ได้ผลดังนี้

- ตัวแบบที่ 1 คือ $n_{av} = -58{,}818.40 + 9{,}032.50p$ ซึ่งมีค่าสถิติ $F = 137.86$ ค่า $p-value = 2.713\times10^{-8}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9138$

- ตัวแบบที่ 2 คือ $n_{av} = 4{,}382.59 - 932.62p + 208.60p^2$ ซึ่งมีค่าสถิติ $F = 1{,}517$ ค่า $p-value = 3.731\times10^{-15}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9961$

- ตัวแบบที่ 3 คือ $n_{av} = 14{,}047.43 - 3{,}360.24p + 333.39p^2 - 1.65p^3$ ซึ่งมีค่าสถิติ $F = 1{,}150$ ค่า $p-value = 5.146\times10^{-14}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9968$

จากตัวแบบทั้งสามข้างต้น สามารถทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 1 กับตัวแบบที่ 2 ดังนี้

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 284.6734$ ด้วยค่า $p-value = 3.285 \times 10^{-9}$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 2

เมื่อทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 2 กับตัวแบบที่ 3 ดังนี้

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 2.6335$ ด้วยค่า $p-value = 0.1329$ จึงยอมรับสมมติฐาน H_0

ดังนั้น ตัวแบบความถดถอยพหุนามที่เหมาะสมสำหรับอธิบายความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.05, 0.05] กับจำนวนมิติ (p) คือ

$$n_{av} = 4{,}382.59 - 932.62p + 208.60p^2 \text{ และ } R^2 = 0.9961$$

4.1.3 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน บนบริเวณสี่เหลี่ยม S_{rect} ณ ระดับความคลาดเคลื่อนเท่ากับ 0.01 (e = 0.01)

สำหรับการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} โดยวัดจากขนาดตัวอย่างที่ได้จากการสุ่มตัวอย่า งแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ต่าง ๆ ดังแสดงในตารางที่ 4.7 – 4.9

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.7 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
1	3,700	46,000	44,200	16,300	91,000
2	10,000	44,200	39,700	55,000	236,800
3	3,700	11,800	3,700	240,400	220,600
4	12,700	29,800	85,600	46,000	179,200
5	22,600	1,900	6,400	133,300	137,800
6	6,400	32,500	23,500	80,200	143,200
7	4,600	24,400	44,200	45,100	95,500
8	1,900	38,800	23,500	165,700	112,600
9	5,500	92,800	62,200	23,500	121,600
10	7,300	17,200	24,400	91,900	72,100
11	2,800	52,300	27,100	136,900	50,500
12	7,300	55,900	69,400	74,800	20,800
13	15,400	90,100	48,700	213,400	141,400
14	18,100	60,400	12,700	4,600	200,800
15	2,800	42,400	46,000	136,000	93,700
16	6,400	10,900	7,300	29,800	65,800
17	5,500	33,400	72,100	82,000	287,200
18	26,200	84,700	11,800	33,400	26,200
19	8,200	36,100	18,100	66,700	32,500
20	5,500	34,300	37,900	87,400	26,200
21	22,600	23,500	56,800	91,000	127,900
22	19,900	6,400	83,800	22,600	237,700
23	2,800	107,200	47,800	12,700	93,700
24	3,700	42,400	21,700	49,600	73,900
25	2,800	20,800	77,500	52,300	57,700

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.7 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่ เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
26	7,300	4,600	31,600	46,000	172,000
27	9,100	10,000	19,000	112,600	124,300
28	10,900	24,400	15,400	19,000	55,000
29	3,700	29,800	36,100	30,700	33,400
30	3,700	23,500	63,100	86,500	47,800
n_{av}	**8,770**	**37,750**	**38,710**	**76,180**	**112,630**

ตารางที่ 4.8 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
1	190,900	396,100	132,400	324,100	442,900
2	264,700	109,900	214,300	55,000	794,800
3	121,600	151,300	130,600	250,300	643,600
4	43,300	54,100	235,900	521,200	860,500
5	147,700	395,200	280,900	301,600	409,600
6	268,300	161,200	144,100	281,800	714,700
7	46,000	194,500	159,400	888,400	987,400
8	162,100	197,200	208,900	350,200	743,500
9	135,100	81,100	70,300	184,600	843,400
10	56,800	190,900	249,400	305,200	811,900

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.8 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
11	93,700	77,500	334,900	264,700	1,446,400
12	152,200	129,700	325,900	120,700	845,200
13	54,100	183,700	209,800	356,500	516,700
14	91,900	69,400	93,700	152,200	1,202,500
15	152,200	167,500	301,600	849,700	1,324,900
16	103,600	118,000	411,400	377,200	433,000
17	227,800	123,400	108,100	214,300	1,106,200
18	178,300	31,600	120,700	221,500	943,300
19	259,300	76,600	192,700	181,000	487,000
20	91,900	104,500	116,200	233,200	475,300
21	175,600	257,500	123,400	257,500	955,000
22	32,500	73,000	180,100	330,400	489,700
23	94,600	191,800	353,800	125,200	1,144,000
24	79,300	74,800	271,000	270,100	710,200
25	71,200	143,200	124,300	715,600	391,600
26	199,000	130,600	235,000	280,900	710,200
27	379,000	152,200	172,000	102,700	943,300
28	109,900	118,900	202,600	208,000	1,117,000
29	203,500	914,500	79,300	1,015,300	615,700
30	84,700	231,400	53,200	282,700	262,900
n_{av}	**142,360**	**176,710**	**194,530**	**334,060**	**779,080**

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.9 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
1	774,100	1,274,500	5,239,900	4,365,100	11,933,200
2	656,200	3,534,400	7,767,100	5,778,100	10,036,900
3	1,189,000	4,016,800	2,755,000	4,381,300	8,439,400
4	971,200	3,300,400	3,051,100	6,223,600	12,597,400
5	523,900	2,306,800	3,277,900	4,584,700	9,886,600
6	1,115,200	3,354,400	4,279,600	5,415,400	12,617,200
7	1,665,100	1,558,000	2,454,400	14,163,400	9,156,700
8	1,489,600	1,851,400	3,556,900	6,978,700	8,684,200
9	1,477,000	2,520,100	2,834,200	6,166,900	15,616,900
10	2,533,600	1,730,800	1,865,800	8,928,100	13,838,500
11	3,812,500	1,082,800	3,895,300	8,122,600	8,936,200
12	1,972,000	2,568,700	2,258,200	13,438,000	10,898,200
13	1,300,600	2,258,200	3,349,900	5,056,300	8,913,700
14	1,060,300	1,279,000	2,820,700	10,624,600	8,637,400
15	1,172,800	2,599,300	5,302,000	4,582,000	13,116,700
16	1,986,400	2,661,400	2,981,800	4,840,300	12,031,300
17	811,900	3,196,900	4,896,100	5,556,700	14,856,400
18	1,745,200	1,815,400	4,439,800	6,270,400	9,539,200
19	1,202,500	1,480,600	4,933,900	7,363,900	7,736,500
20	1,575,100	2,375,200	2,413,000	7,404,400	15,076,900
21	2,143,900	2,884,600	5,337,100	11,581,300	8,410,600
22	1,734,400	1,852,300	2,651,500	3,113,200	5,690,800
23	720,100	3,537,100	4,969,000	4,833,100	12,235,600
24	2,108,800	3,867,400	4,319,200	5,959,900	9,674,200
25	646,300	1,694,800	4,003,300	5,996,800	16,807,600

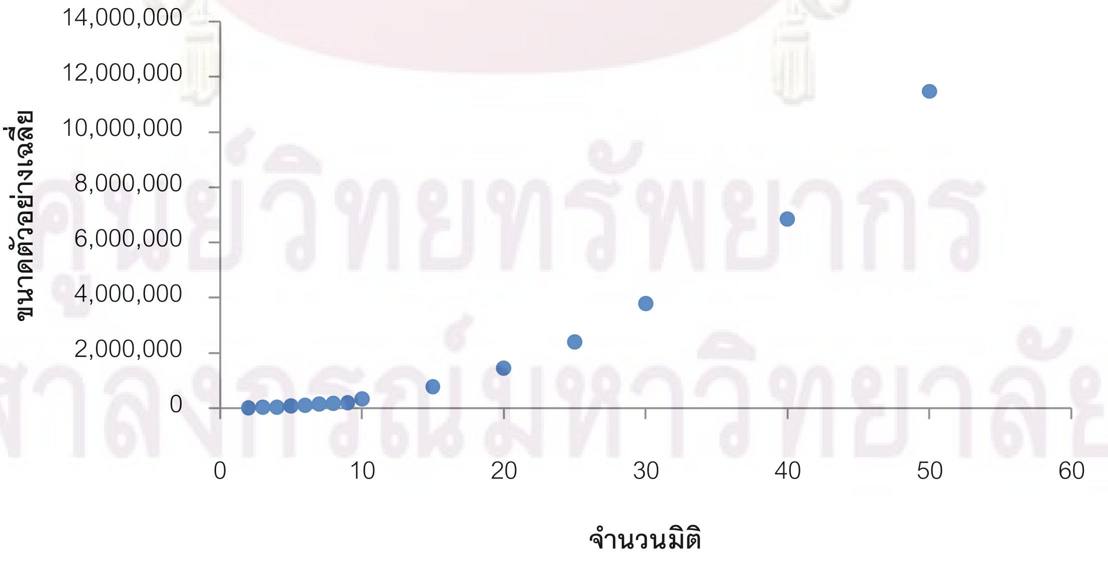
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.9 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนกา รลู่เข้าสู่จุดกำเนิดภายในความ คลาดเคลื่อน [-0.01, 0.01] เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
26	1,138,600	1,542,700	2,180,800	6,141,700	24,747,400
27	1,098,100	1,774,000	3,434,500	10,063,000	15,365,800
28	1,788,400	2,827,000	3,291,400	5,995,000	14,230,900
29	1,308,700	1,683,100	4,524,400	4,870,900	5,018,500
30	1,549,900	3,373,300	4,448,800	6,907,600	9,494,200
n_{av}	1,442,380	2,393,380	3,784,420	6,856,900	11,474,170

จากตารางที่ 4.7 – 4.9 พบว่า ขนาดตัวอย่างที่สุ่มได้จนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] มีขนาดที่เพิ่มขึ้น เมื่อจำนวนมิติ (p) ที่ต้องการสุ่มแบบฮิตแอนด์รันมีจำนวนเพิ่มขึ้น ด้วยเหตุนี้จึงศึกษาความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ทำให้ได้ผลดังแสดงในรูปที่ 4.3

รูปที่ 4.3 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] กับจำนวนมิติ (p)



จากรูปที่ 4.3 พบว่า ลักษณะความสัมพันธ์ระหว่าง n_{av} กับ p เป็นเส้นโค้งที่ n_{av} มีขนาดเพิ่มขึ้นอย่างรวดเร็ว เมื่อ p มีจำนวนเพิ่มขึ้น ดังนั้น ตัวแบบความถดถอยที่จะอธิบายความสัมพันธ์ดังกล่าว คือ ตัวแบบความถดถอยพหุนาม เมื่อวิเคราะห์ความถดถอยพหุนามอันดับที่ 1 ถึง 3 ของความสัมพันธ์ระหว่าง n_{av} กับ p ด้วยโปรแกรม R ได้ผลดังนี้

- ตัวแบบที่ 1 คือ $n_{av} = -1{,}469{,}819 + 213{,}245p$ ซึ่งมีค่าสถิติ $F = 123.83$ ค่า $p - value = 5.138 \times 10^{-8}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9050$
- ตัวแบบที่ 2 คือ $n_{av} = 138{,}535.75 - 40{,}350.58p + 5{,}308.42p^2$ ซึ่งมีค่าสถิติ $F = 15{,}650$ ค่า $p - value < 2.2 \times 10^{-16}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9996$
- ตัวแบบที่ 3 คือ $n_{av} = 29{,}008.41 - 12{,}839.52p + 3{,}894.20p^2 + 18.65p^3$ ซึ่งมีค่าสถิติ $F = 17{,}480$ ค่า $p - value < 2.2 \times 10^{-16}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9998$

จากตัวแบบทั้งสามข้างต้น สามารถทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 1 กับตัวแบบที่ 2 ดังนี้

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 4{,}964.4503$ ด้วยค่า $p - value = 5.846 \times 10^{-16}$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 2

เมื่อทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 2 กับตัวแบบที่ 3 ดังนี้

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 9.1075$ ด้วยค่า $p - value = 0.0117$ จึงยอมรับสมมติฐาน H_0

ดังนั้น ตัวแบบความถดถอยพหุนามที่เหมาะสมสำหรับอธิบายความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน [-0.01, 0.01] กับจำนวนมิติ (p) คือ

$$n_{av} = 138{,}535.75 - 40{,}350.58p + 5{,}308.42p^2 \text{ และ } R^2 = 0.9996$$

4.2 ผลการศึกษาอัตราการ ลู่เข้า ของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม

ในการวิจัยนี้ได้กำหนดให้การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{tri} เป็นการสุ่มตัวอย่างในกรณีที่ไม่ทราบค่าเฉลี่ยของประชากร จึงศึกษาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันของบริเวณสามเหลี่ยม S_{tri} จนกระทั่ง $SE^* < e$ โดยมีขนาดตัวอย่างเฉลี่ย (n_{av}) จนกระทั่ง $SE^* < e$ เป็นค่าวัดอัตราการลู่เข้ าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน สามารถแสดงผลการศึกษาโดยจำแนกตามระดับความคลาดเคลื่อนได้ดังนี้

4.2.1 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} ณ ระดับความคลาดเคลื่อนเท่ากับ 0.1 (e = 0.1)

สำหรับการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} โดยวัดจากขนาดตัวอย่างที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ต่าง ๆ ดังแสดงในตารางที่ 4.10 – 4.12

ตารางที่ 4.10 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
1	1,900	2,800	3,700	3,700	6,400
2	1,900	1,900	3,700	2,800	5,500
3	1,900	2,800	4,600	7,300	11,800
4	1,900	2,800	3,700	6,400	9,100
5	1,900	2,800	3,700	9,100	10,000
6	1,900	1,900	2,800	3,700	6,400

ตารางที่ 4.10 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
7	1,900	3,700	3,700	11,800	9,100
8	1,900	2,800	3,700	5,500	3,700
9	2,800	2,800	2,800	4,600	7,300
10	1,900	4,600	4,600	3,700	11,800
11	1,900	1,900	5,500	9,100	6,400
12	1,900	2,800	3,700	4,600	7,300
13	1,900	2,800	4,600	9,100	11,800
14	1,900	2,800	2,800	4,600	4,600
15	1,900	2,800	4,600	2,800	5,500
16	1,900	2,800	2,800	8,200	3,700
17	1,900	2,800	2,800	3,700	4,600
18	1,900	3,700	3,700	4,600	9,100
19	1,900	3,700	2,800	4,600	10,900
20	1,900	1,900	4,600	4,600	4,600
21	1,900	2,800	4,600	5,500	10,000
22	1,900	3,700	1,900	3,700	7,300
23	1,900	1,900	4,600	3,700	9,100
24	3,700	1,900	6,400	5,500	8,200
25	1,900	1,900	3,700	8,200	10,000
26	1,900	2,800	4,600	7,300	5,500
27	1,900	1,900	5,500	3,700	3,700
28	1,900	1,900	2,800	7,300	6,400
29	1,900	1,900	2,800	4,600	5,500
30	1,900	1,900	2,800	4,600	10,900
n_{av}	1,990	2,650	3,820	5,620	7,540

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.11 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
1	11,800	6,400	18,100	10,000	16,300
2	8,200	10,900	8,200	14,500	21,700
3	4,600	5,500	19,000	9,100	15,400
4	4,600	11,800	8,200	9,100	17,200
5	5,500	10,900	13,600	10,000	38,800
6	13,600	10,000	6,400	26,200	17,200
7	8,200	12,700	23,500	17,200	34,300
8	4,600	4,600	6,400	14,500	19,900
9	10,900	18,100	8,200	14,500	37,900
10	4,600	10,900	12,700	10,000	19,000
11	8,200	11,800	10,000	21,700	14,500
12	11,800	4,600	7,300	18,100	20,800
13	4,600	6,400	12,700	25,300	27,100
14	9,100	10,900	19,900	10,900	23,500
15	16,300	5,500	8,200	10,000	25,300
16	8,200	4,600	10,900	19,000	27,100
17	5,500	8,200	13,600	24,400	28,900
18	10,000	19,900	8,200	15,400	20,800
19	4,600	12,700	12,700	24,400	18,100
20	14,500	14,500	7,300	19,900	27,100
21	4,600	19,900	7,300	11,800	26,200
22	10,900	10,900	19,900	13,600	19,000
23	5,500	5,500	8,200	11,800	20,800
24	13,600	11,800	18,100	9,100	20,800
25	5,500	4,600	8,200	12,700	20,800

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.11 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
26	8,200	4,600	10,000	10,000	21,700
27	11,800	19,900	7,300	21,700	29,800
28	7,300	10,900	17,200	25,300	37,900
29	6,400	5,500	11,800	10,900	34,300
30	15,400	19,000	13,600	19,000	28,900
n_{av}	**8,620**	**10,450**	**11,890**	**15,670**	**24,370**

ตารางที่ 4.12 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
1	49,600	110,800	114,400	267,400	470,800
2	44,200	67,600	136,900	217,000	482,500
3	46,000	123,400	106,300	204,400	422,200
4	37,000	50,500	137,800	237,700	426,700
5	39,700	129,700	178,300	208,000	406,900
6	55,000	78,400	118,900	231,400	396,100
7	39,700	30,700	164,800	225,100	361,000
8	43,300	42,400	132,400	287,200	343,900
9	40,600	37,300	138,700	287,200	364,600
10	38,800	98,200	147,700	222,400	514,000
11	22,600	128,800	169,300	222,400	395,200
12	64,900	95,500	104,500	289,000	381,700
13	38,800	118,000	119,800	244,900	417,700

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

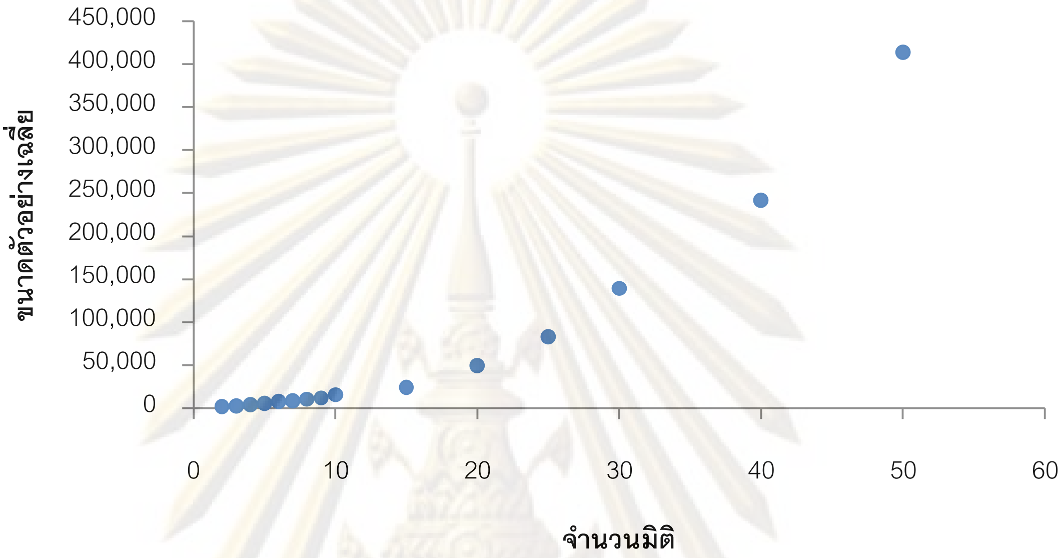
ตารางที่ 4.12 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
14	44,200	68,500	166,600	237,700	317,800
15	43,300	109,000	167,500	239,500	460,000
16	56,800	79,300	115,300	186,400	365,500
17	48,700	93,700	167,500	226,000	468,100
18	30,700	89,200	136,000	198,100	461,800
19	55,900	64,000	172,000	215,200	397,900
20	63,100	122,500	171,100	294,400	409,600
21	63,100	80,200	117,100	202,600	395,200
22	54,100	37,900	123,400	260,200	486,100
23	51,400	105,400	125,200	289,000	519,400
24	60,400	77,500	116,200	242,200	536,500
25	32,500	46,000	109,000	253,000	377,200
26	37,000	62,200	143,200	269,200	357,400
27	63,100	45,100	109,900	194,500	470,800
28	46,900	40,600	123,400	244,900	321,400
29	62,200	122,500	169,300	262,900	352,000
30	104,500	80,200	177,400	289,000	343,900
n_{av}	**49,270**	**81,170**	**139,330**	**241,630**	**414,130**

จากตารางที่ 4.10 – 4.12 พบว่า ขนาดตัวอย่างที่สุ่มแบบฮิตแอนด์รั นจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 มีขนาดที่เพิ่มขึ้น เมื่อจำนวนมิติ (p) มีจำนวนเพิ่มขึ้ น ด้วยเหตุนี้จึงศึกษาความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ทำให้ได้ผลดังแสดงในรูปที่ 4.4

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.4 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 กับจำนวนมิติ (p)



จากรูปที่ 4.4 พบว่า ลักษณะความสัมพันธ์ระหว่าง n_{av} กับ p เป็นเส้นโค้งที่ n_{av} มีขนาดเพิ่มขึ้นอย่างรวดเร็ว เมื่อ p มีจำนวนเพิ่มขึ้น ดังนั้น ตัวแบบความถดถอยที่จะอธิบายความสัมพันธ์ดังกล่าว คือ ตัวแบบความถดถอยพหุนาม เมื่อวิเคราะห์ความถดถอยพหุนามอันดับที่ 1 ถึง 3 ของความสัมพันธ์ระหว่าง n_{av} กับ p ด้วยโปรแกรม R ได้ผลดังนี้

- ตัวแบบที่ 1 คือ $n_{av} = -50{,}028.7 + 7{,}566.6p$ ซึ่งมีค่าสถิติ $F = 112.59$ ค่า $p-value = 9.005\times10^{-8}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.8965$

- ตัวแบบที่ 2 คือ $n_{av} = 9{,}528.93 - 1{,}824.10p + 196.57p^2$ ซึ่งมีค่าสถิติ $F = 4{,}207$ ค่า $p-value < 2.2\times10^{-16}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9986$

- ตัวแบบที่ 3 คือ $n_{av} = 3{,}673.90 - 353.43p + 120.97p^2 + p^3$ ซึ่งมีค่าสถิติ $F = 3{,}543$ ค่า $p-value < 2.2\times10^{-16}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9999$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตัวแบบทั้งสามข้างต้น สามารถทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 1 กับตัวแบบที่ 2 ดังนี้

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 1{,}086.2201$ ด้วยค่า $p - value = 2.395 \times 10^{-12}$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 2

เมื่อทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 2 กับตัวแบบที่ 3 ดังนี้

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 4.1528$ ด้วยค่า $p - value = 0.06635$ จึงยอมรับสมมติฐาน H_0

ดังนั้น ตัวแบบความถดถอยพหุนามที่เหมาะสมสำหรับอธิบายความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย(n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.1 กับจำนวนมิติ (p) คือ

$$n_{av} = 9{,}528.93 - 1{,}824.10p + 196.57p^2 \text{ และ } R^2 = 0.9986$$

4.2.2 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} ณ ระดับความคลาดเคลื่อนเท่ากับ 0.1 (e = 0.05)

สำหรับการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} โดยวัดจากขนาดตัวอย่างที่ได้ จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ต่าง ๆ ดังแสดงในตารางที่ 4.13 – 4.15

ตารางที่ 4.13 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
1	2,800	3,700	15,400	19,000	18,100
2	3,700	5,500	9,100	10,900	18,100
3	2,800	10,000	10,900	25,300	28,900
4	4,600	6,400	8,200	13,600	7,300
5	3,700	6,400	5,500	8,200	23,500
6	4,600	4,600	19,000	16,300	10,900
7	2,800	9,100	6,400	13,600	27,100
8	2,800	8,200	8,200	10,900	7,300
9	3,700	10,000	8,200	10,900	43,300
10	3,700	9,100	6,400	6,400	12,700
11	3,700	13,600	6,400	7,300	18,100
12	2,800	5,500	19,900	14,500	8,200
13	2,800	9,100	9,100	9,100	27,100
14	4,600	5,500	8,200	27,100	23,500
15	2,800	9,100	10,000	10,000	12,700
16	2,800	10,000	5,500	14,500	12,700
17	4,600	7,300	9,100	9,100	20,800
18	4,600	3,700	15,400	14,500	8,200
19	2,800	5,500	8,200	19,000	19,000
20	4,600	4,600	9,100	15,400	7,300
21	3,700	6,400	9,100	13,600	23,500
22	4,600	7,300	11,800	21,700	19,900
23	3,700	6,400	10,900	18,100	21,700
24	2,800	6,400	11,800	13,600	18,100
25	2,800	9,100	6,400	9,100	19,900

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.13 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
26	3,700	5,500	10,900	19,000	31,600
27	3,700	6,400	6,400	18,100	20,800
28	3,700	9,100	5,500	10,900	10,000
29	4,600	8,200	11,800	11,800	33,400
30	4,600	5,500	15,400	13,600	20,800
n_{av}	**3,640**	**7,240**	**9,940**	**14,170**	**19,150**

ตารางที่ 4.14 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
1	17,200	18,100	31,600	33,400	49,600
2	22,600	17,200	33,400	47,800	45,100
3	25,300	16,300	28,000	28,000	52,300
4	16,300	38,800	32,500	47,800	38,800
5	28,000	23,500	36,100	36,100	50,500
6	25,300	19,900	31,600	28,900	47,800
7	20,800	15,400	29,800	32,500	68,500
8	23,500	14,500	29,800	47,800	57,700
9	23,500	14,500	28,900	55,900	46,000
10	21,700	31,600	30,700	18,100	43,300
11	19,900	44,200	28,900	29,800	54,100
12	21,700	24,400	28,000	28,900	58,600
13	21,700	37,900	33,400	48,700	49,600
14	22,600	35,200	25,300	37,900	47,800
15	38,800	35,200	47,800	46,900	46,000

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.14 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
16	22,600	17,200	28,000	31,600	49,600
17	19,900	17,200	28,000	25,300	59,500
18	23,500	25,300	28,000	29,800	47,800
19	17,200	45,100	46,000	25,300	50,500
20	19,000	40,600	34,300	51,400	53,200
21	19,900	17,200	28,000	29,800	55,900
22	19,000	28,000	35,200	30,700	39,700
23	23,500	16,300	28,900	55,900	70,300
24	22,600	27,100	31,600	32,500	44,200
25	24,400	25,300	28,000	34,300	66,700
26	24,400	25,300	27,100	37,000	68,500
27	16,300	26,200	32,500	23,500	46,900
28	42,400	32,500	33,400	47,800	57,700
29	21,700	26,200	31,600	44,200	61,300
30	19,900	39,700	25,300	46,000	60,400
n_{av}	**22,840**	**26,530**	**31,390**	**37,120**	**52,930**

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.15 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
1	99,100	182,800	309,700	514,900	870,400
2	123,400	195,400	325,000	532,900	1,054,000
3	116,200	208,900	312,400	601,300	965,800
4	110,800	200,800	307,000	496,900	938,800
5	107,200	177,400	315,100	551,800	856,900
6	115,300	216,100	324,100	595,900	957,700
7	108,100	192,700	315,100	500,500	863,200
8	130,600	203,500	304,300	528,400	1,063,900
9	138,700	214,300	342,100	566,200	1,067,500
10	95,500	183,700	321,400	541,900	782,200
11	117,100	167,500	281,800	523,000	928,900
12	108,100	202,600	316,000	543,700	876,700
13	112,600	188,200	275,500	518,500	809,200
14	90,100	162,100	262,000	455,500	836,200
15	122,500	185,500	258,400	478,900	835,300
16	137,800	217,900	331,300	523,000	900,100
17	98,200	182,800	288,100	579,700	878,500
18	114,400	211,600	345,700	509,500	889,300
19	103,600	201,700	308,800	532,900	793,900
20	112,600	193,600	321,400	514,900	802,000
21	140,500	191,800	327,700	507,700	1,006,300
22	114,400	185,500	296,200	540,100	886,600
23	110,800	213,400	326,800	527,500	1,015,300
24	112,600	169,300	281,800	502,300	974,800
25	107,200	199,900	328,600	528,400	997,300

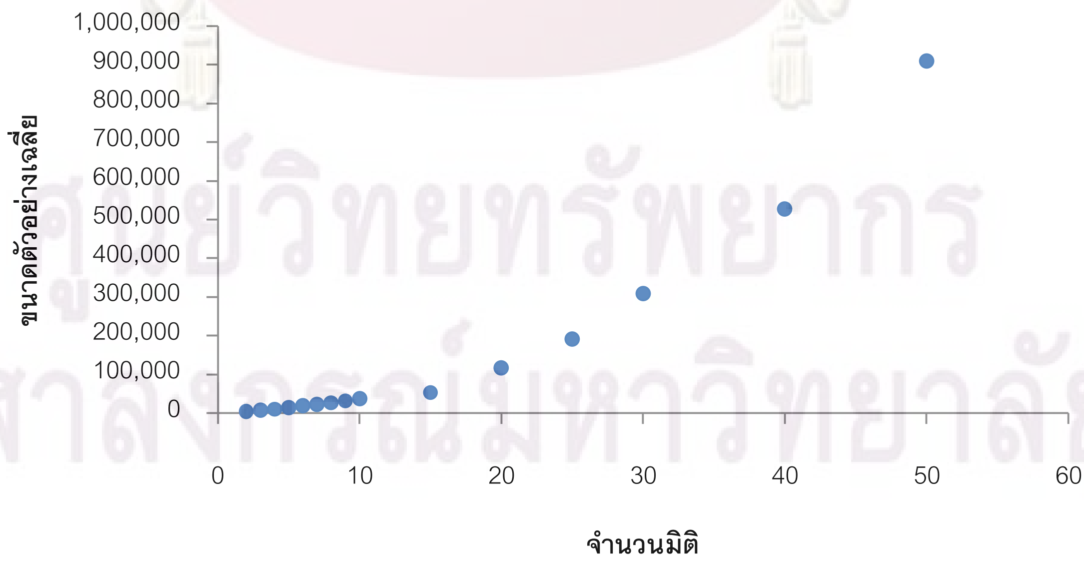
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.15 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
26	134,200	179,200	349,300	529,300	951,400
27	134,200	178,300	311,500	545,500	802,000
28	116,200	173,800	309,700	532,000	924,400
29	119,800	172,900	273,700	520,300	935,200
30	129,700	168,400	275,500	478,900	834,400
n_{av}	**116,050**	**190,720**	**308,200**	**527,410**	**909,940**

จากตารางที่ 4.13 – 4.15 พบว่า ขนาดตัวอย่างที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 มีขนาดที่เพิ่มขึ้น เมื่อจำนวนมิติ (p) มีจำนวนเพิ่มขึ้ น ด้วยเหตุนี้จึงศึกษาความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ทำให้ได้ผลดังแสดงในรูปที่ 4.5

รูปที่ 4.5 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 กับจำนวนมิติ (p)



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากรูปที่ 4.5 พบว่า ลักษณะความสัมพันธ์ระหว่าง n_{av} กับ p เป็นเส้นโค้งที่ n_{av} มีขนาดเพิ่มขึ้นอย่างรวดเร็ว เมื่อ p มีจำนวนเพิ่มขึ้น ดังนั้น ตัวแบบความถดถอยที่จะอธิบายความสัมพันธ์ดังกล่าว คือ ตัวแบบความถดถอยพหุนาม เมื่อวิเคราะห์ความถดถอยพหุนามอันดับที่ 1 ถึง 3 ของความสัมพันธ์ระหว่าง n_{av} กับ p ด้วยโปรแกรม R ได้ผลดังนี้

- ตัวแบบที่ 1 คือ $n_{av} = -106{,}878 + 16{,}583p$ ซึ่งมีค่าสถิติ $F = 116.08$ ค่า $p - value = 7.525 \times 10^{-8}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.8993$
- ตัวแบบที่ 2 คือ $n_{av} = 21{,}413.56 - 3{,}645.15p + 423.43p^2$ ซึ่งมีค่าสถิติ $F = 3{,}357$ ค่า $p - value < 2.2 \times 10^{-16}$ และค่าสัมประสิทธิ์ การตัดสินใจ $R^2 = 0.9982$
- ตัวแบบที่ 3 คือ $n_{av} = 5{,}712.60 + 298.61p + 2220.70p^2 + 2.67p^3$ ซึ่งมีค่าสถิติ $F = 3{,}057$ ค่า $p - value = 2.408 \times 10^{-16}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9988$

จากตัวแบบทั้งสามข้างต้น สามารถทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 1 กับตัวแบบที่ 2 ดังนี้

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 908.416$ ด้วยค่า $p - value = 6.338 \times 10^{-12}$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 2

เมื่อทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 2 กับตัวแบบที่ 3 ดังนี้

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 5.353$ ด้วยค่า $p - value = 0.04058$ จึงยอมรับสมมติฐาน H_0

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ดังนั้น ตัวแบบความถดถอยพหุนามที่ เหมาะสมสำหรับอธิบายความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย(n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.05 กับจำนวนมิติ (p) คือ

$$n_{av} = 21{,}413.56 - 3{,}645.15p + 423.43p^2 \text{ และ } R^2 = 0.9982$$

4.2.3 ผลการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} ณ ระดับความคลาดเคลื่อนเท่ากับ 0.1 (e = 0.01)

สำหรับการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} โดยวัดจากขนาดตัวอย่างที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ต่าง ๆ ดังแสดงในตารางที่ 4.16 – 4.18

ตารางที่ 4.16 แสดงขนาดตัวอย่ างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
1	13,600	44,200	68,500	104,500	219,700
2	10,900	40,600	69,400	103,600	204,400
3	9,100	51,400	73,900	113,500	240,400
4	12,700	41,500	71,200	101,800	217,000
5	10,000	48,700	78,400	104,500	217,000
6	12,700	54,100	60,400	103,600	184,600
7	11,800	40,600	73,900	121,600	223,300
8	12,700	37,900	73,900	112,600	207,100
9	10,000	44,200	66,700	109,000	252,100
10	10,900	43,300	73,000	103,600	203,500
11	9,100	44,200	66,700	95,500	255,700
12	10,000	43,300	74,800	114,400	250,300
13	12,700	51,400	67,600	105,400	198,100
14	9,100	46,000	63,100	114,400	229,600

ตารางที่ 4.16 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 2 ถึง 6

ชุดที่	p=2	p=3	p=4	p=5	p=6
15	11,800	44,200	67,600	116,200	216,100
16	9,100	45,100	75,700	92,800	221,500
17	12,700	50,500	68,500	107,200	215,200
18	10,000	38,800	64,900	119,800	269,200
19	10,900	51,400	73,000	113,500	214,300
20	10,000	39,700	59,500	120,700	208,000
21	11,800	37,900	65,800	94,600	231,400
22	10,000	42,400	63,100	113,500	217,000
23	10,900	48,700	70,300	117,100	218,800
24	10,900	41,500	67,600	107,200	215,200
25	12,700	50,500	64,900	93,700	259,300
26	12,700	38,800	72,100	110,800	236,800
27	10,900	46,900	64,900	100,900	211,600
28	10,900	44,200	61,300	105,400	224,200
29	10,000	48,700	61,300	122,500	243,100
30	11,800	42,400	68,500	118,900	216,100
n_{av}	11,080	44,770	68,350	108,760	224,020

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.17 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
1	225,100	371,800	514,000	765,100	1,324,000
2	289,000	320,500	422,200	892,900	1,257,400
3	325,900	433,900	452,800	714,700	1,540,000
4	263,800	307,900	368,200	792,100	1,332,100
5	302,500	365,500	485,200	968,500	1,293,400
6	283,600	305,200	518,500	846,100	1,498,600
7	284,500	350,200	467,200	775,900	1,378,000
8	240,400	290,800	549,100	901,000	1,337,500
9	273,700	333,100	419,500	714,700	1,308,700
10	289,900	351,100	421,300	802,000	1,396,900
11	324,100	379,000	440,200	885,700	1,340,200
12	306,100	436,600	590,500	748,000	1,481,500
13	305,200	386,200	566,200	910,000	1,208,800
14	313,300	248,500	386,200	806,500	1,601,200
15	241,300	334,900	617,500	684,100	1,582,300
16	281,800	288,100	487,900	802,000	1,307,800
17	223,300	387,100	427,600	807,400	1,261,900
18	247,600	301,600	488,800	835,300	1,478,800
19	292,600	360,100	426,700	905,500	1,477,900
20	289,900	363,700	559,900	804,700	1,431,100
21	319,600	433,900	402,400	845,200	1,328,500
22	313,300	341,200	443,800	764,200	1,171,000
23	208,900	400,600	496,900	725,500	1,673,200
24	216,100	291,700	607,600	730,900	1,300,600
25	294,400	396,100	467,200	679,600	1,550,800

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.17 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 7 8 9 10 และ 15

ชุดที่	p=7	p=8	p=9	p=10	p=15
26	253,000	412,300	532,900	888,400	1,380,700
27	278,200	301,600	523,900	887,500	1,587,700
28	226,000	383,500	471,700	918,100	1,476,100
29	267,400	379,000	597,700	662,500	1,341,100
30	220,600	410,500	509,500	835,300	1,539,100
n_{av}	**273,370**	**355,540**	**488,770**	**809,980**	**1,406,230**

ตารางที่ 4.18 แสดงขนาดตัวอย่าง จำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
1	3,218,500	7,354,000	11,279,800	21,251,800	27,775,000
2	3,206,800	6,756,400	12,336,400	19,099,000	30,070,000
3	3,048,400	7,012,900	11,769,400	20,226,700	26,004,700
4	2,941,300	6,853,600	12,314,800	19,245,700	29,816,200
5	3,025,000	6,616,900	11,745,100	21,624,400	27,561,700
6	2,935,000	6,535,000	11,514,700	21,825,100	25,473,700
7	3,115,000	6,649,300	11,952,100	21,988,900	24,580,900
8	3,063,700	6,336,100	12,761,200	17,878,600	28,735,300
9	3,194,200	6,247,900	11,389,600	19,530,100	31,398,400
10	2,652,400	6,877,900	11,129,500	20,580,400	35,859,700
11	2,937,700	7,088,500	12,491,200	23,572,000	30,385,900
12	2,822,500	7,377,400	12,125,800	23,883,400	33,784,300
13	3,049,300	6,308,200	12,510,100	21,030,400	34,237,000

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

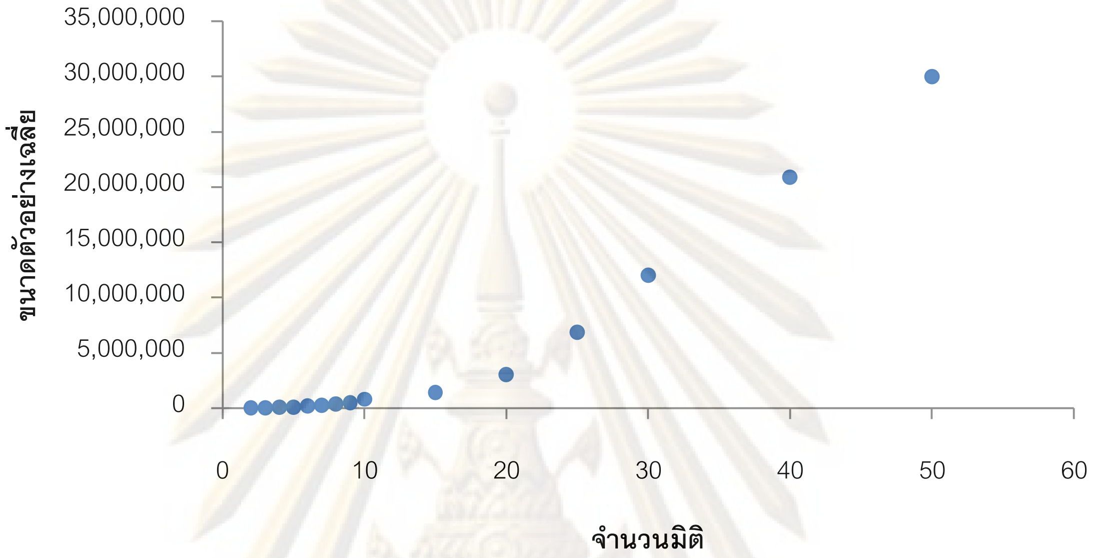
ตารางที่ 4.18 (ต่อ) แสดงขนาดตัวอย่างจำนวน 30 ชุดและขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 เมื่อกำหนดจำนวนมิติ (p) ตั้งแต่ 20 25 30 40 และ 50

ชุดที่	p=20	p=25	p=30	p=40	p=50
14	2,919,700	7,000,300	12,131,200	22,934,800	33,896,800
15	2,949,400	7,553,800	12,019,600	22,596,400	30,812,500
16	3,222,100	7,405,300	13,297,600	17,917,300	29,527,300
17	2,967,400	7,179,400	11,593,000	20,937,700	29,741,500
18	2,801,800	7,481,800	11,239,300	22,670,200	29,336,500
19	2,858,500	6,553,900	11,404,900	20,628,100	33,338,800
20	3,358,000	6,364,900	11,710,000	22,467,700	31,727,800
21	3,142,000	6,722,200	11,008,000	19,980,100	26,314,300
22	2,692,900	6,304,600	12,639,700	21,388,600	32,857,300
23	2,802,700	6,225,400	11,721,700	20,312,200	29,150,200
24	2,753,200	6,846,400	12,694,600	18,786,700	26,403,400
25	3,113,200	6,932,800	12,664,900	20,376,100	28,765,000
26	3,050,200	6,986,800	12,691,900	22,844,800	30,393,100
27	3,332,800	7,028,200	11,359,900	21,556,000	27,037,900
28	3,077,200	6,730,300	11,661,400	20,111,500	35,330,500
29	3,174,400	7,555,600	11,890,900	18,388,000	29,937,700
30	3,165,400	6,881,500	12,877,300	21,523,600	28,593,100
n_{av}	**3,019,690**	**6,858,910**	**11,997,520**	**20,905,210**	**29,961,550**

จากตารางที่ 4.16 – 4.18 พบว่า ขนาดตัวอย่างที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 มีขนาดที่เพิ่มขึ้น เมื่อจำนวนมิติ (p) มีจำนวนเพิ่มขึ้ น ด้วยเหตุนี้จึงศึกษาความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ทำให้ได้ผลดังแสดงในรูปที่ 4.6

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.6 แสดงความสัมพันธ์ระหว่างขนาดตัวอย่างเฉลี่ย (n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 กับจำนวนมิติ (p)



จากรูปที่ 4.6 พบว่า ลักษณะความสัมพันธ์ระหว่าง n_{av} กับ p เป็นเส้นโค้งที่ n_{av} มีขนาดเพิ่มขึ้นอย่างรวดเร็ว เมื่อ p มีจำนวนเพิ่มขึ้น ดังนั้น ตัวแบบความถดถอยที่จะอธิบายความสัมพันธ์ดังกล่าว คือ ตัวแบบความถดถอยพหุนาม เมื่อวิเคราะห์ความถดถอยพหุนามอันดับที่ 1 ถึง 3 ของความสัมพันธ์ระหว่าง n_{av} กับ p ด้วยโปรแกรม R ได้ผลดังนี้

- ตัวแบบที่ 1 คือ $n_{av} = -4{,}167{,}083 + 594{,}188p$ ซึ่งมีค่าสถิติ $F = 145.55$ ค่า $p-value = 1.96 \times 10^{-8}$ และ ค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.918$

- ตัวแบบที่ 2 คือ $n_{av} = -200{,}870 - 31{,}180p + 13{,}091p^2$ ซึ่งมีค่าสถิติ $F = 874.6$ ค่า $p-value = 1.00 \times 10^{-13}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9932$

- ตัวแบบที่ 3 คือ $n_{av} = 1{,}242{,}366.2 - 393{,}692.1p + 31{,}725.6p^2 - 245.7p^3$ ซึ่งมีค่าสถิติ $F = 1{,}217$ ค่า $p-value = 2.97 \times 10^{-14}$ และค่าสัมประสิทธิ์การตัดสินใจ $R^2 = 0.9971$

จากตัวแบบทั้งสามข้างต้น สามารถทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 1 กับตัวแบบที่ 2 ดังนี้

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 287.599$ ด้วยค่า $p - value = 3.111 \times 10^{-9}$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 2

เมื่อทำการทดสอบสมมติฐานเพื่อเปรียบเทียบตัวแบบที่ 2 กับตัวแบบที่ 3 ดังนี้

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

ผลการทดสอบสมมติฐานพบว่า ที่ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ ค่าสถิติของการทดสอบ $F_c = 15.065$ ด้วยค่า $p - value = 0.002558$ จึงปฏิเสธสมมติฐาน H_0 หมายความว่า ตัวแบบความถดถอยที่เหมาะสมกว่าน่าจะเป็นตัวแบบที่ 3

ดังนั้น ตัวแบบความถดถอยพหุนามที่เหมาะสมสำหรับอธิบายความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย(n_{av}) ที่สุ่มแบบฮิตแอนด์รันจนกระทั่ง $SE^* < e$ โดยที่ e = 0.01 กับจำนวนมิติ (p) คือ

$$n_{av} = 1,242,366.2 - 393,692.1p + 31,725.6p^2 - 245.7p^3 \text{ และ } R^2 = 0.9971$$

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

งานวิจัยนี้ได้ศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} และบนบริเวณสามเหลี่ยม S_{tri} พร้อมทั้งศึกษาความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) ที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รันกับจำนวนมิติ (p) ของแต่ละบริเวณ ณ ระดับความคลาดเคลื่อน (e) เท่ากับ 0.1 0.05 และ 0.01 สามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

5.1.1 การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect}

งานวิจัยนี้กำหนดให้การศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} เป็นการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างในกรณีที่ทราบค่าเฉลี่ยของประชากร ในที่นี้คือจุดกำเนิด โดยมีขนาดตัวอย่างเฉลี่ย (n_{av}) ที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รัน จนกระทั่งกระบวนการลู่เข้าสู่จุดกำเนิดภายในความคลาดเคลื่อน $[-e,e]$ บนบริเวณสี่เหลี่ยม S_{rect} เป็นค่าวัดอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน จากการศึกษาพบความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ของบริเวณสี่เหลี่ยม S_{rect} สามารถสรุปผลการวิจัยโดยจำแนกตามระดับความคลาดเคลื่อน (e) ได้ดังนี้

1. การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} เมื่อ $e = 0.1$

จากการศึกษาพบว่า ความสัมพันธ์ของ n_{av} กับ p เป็นไปในทิศทางเดียวกัน เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นอย่างรวดเร็ว ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้ง จึงศึกษ าตัวแบบความถดถอยพหุนามอันดับที่ 1 ถึงอันดับที่ 3 ผลการศึกษาพบว่า ตัวแบบความถดถอยพหุนามอันดับที่ 2 เป็นตัวแบบความถดถอยพหุนามที่เหมาะสมกับความสัมพันธ์ n_{av} กับ p เมื่อ $e = 0.1$ ดังนี้

$$n_{av} = 2{,}068.67 - 130.95p + 49.43p^2 \text{ โดยมีค่า } R^2 = 0.9985$$

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

2. การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} เมื่อ $e = 0.05$

จากการศึกษาพบว่า ความสัมพันธ์ของ n_{av} กับ p เป็นไปในทิศทางเดียวกัน เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นอย่างรวดเร็ว ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้ง ซึ่งเป็นไปในทำนองเดียวกับความคลาดเคลื่อน $e = 0.1$ จึงศึกษา ตัว แบบความถดถอยพหุนามอันดับที่ 1 ถึงอันดับที่ 3 ผลการศึกษาพบว่า ตัวแบบความถดถอยพหุนามอันดับที่ 2 เป็นตัวแบบความถดถอยพหุนามที่เหมาะสมกับความสัมพันธ์ n_{av} กับ p เมื่อ $e = 0.05$ ดังนี้

$$n_{av} = 4{,}382.59 - 932.62p + 208.60p^2 \text{ โดยมีค่า } R^2 = 0.9961$$

3. การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสี่เหลี่ยม S_{rect} เมื่อ $e = 0.01$

จากการศึกษาพบว่า ความสัมพันธ์ของ n_{av} กับ p เป็นไปในทิศทางเดียวกัน เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นอย่างรวดเร็ว ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้ง ซึ่งเป็นไปในทำนองเดียวกับความคลาดเคลื่อน $e = 0.1$ และ $e = 0.05$ จึงศึกษาตัวแบบความถดถอยพหุนามอันดับที่ 1 ถึงอันดับที่ 3 ผลการศึกษาพบว่า ตัวแบบความถดถอยพหุนามอันดับที่ 2 เป็นตัวแบบความถดถอยพหุนามที่เหมาะสมกับความสัมพันธ์ n_{av} กับ p เมื่อ $e = 0.01$ ดังนี้

$$n_{av} = 138{,}535.75 - 40{,}350.58p + 5{,}308.42p^2 \text{ โดยมีค่า } R^2 = 0.9996$$

5.1.2 การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri}

งานวิจัยนี้กำหนดให้การศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} เป็นการศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างในกรณีที่ไม่ทราบค่าเฉลี่ยของประชากร จึงศึกษาอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} จนกระทั่ง $SE^* < e$ โดยมีขนาดตัวอย่างเฉลี่ย (n_{av}) ที่ได้จากการสุ่มตัวอย่างแบบฮิตแอนด์รัน จนกระทั่งกระบวนการลู่เข้าจนกระทั่ง $SE^* < e$ เป็นค่าวัดอัตราการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รัน จากการศึกษาพบความสัมพันธ์ของขนาดตัวอย่างเฉลี่ย (n_{av}) กับจำนวนมิติ (p) ของบริเวณสามเหลี่ยม S_{tti} สามารถสรุปผลการวิจัยโดยจำแนกตามระดับความคลาดเคลื่อน (e) ได้ดังนี้

1. การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} เมื่อ $e = 0.1$

จากการศึกษาพบว่ า ความสัมพันธ์ของ n_{av} กับ p เป็นไปในทิศทางเดียวกัน เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นอย่างรวดเร็ว ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้ง จึงศึกษาตัวแบบความถดถอยพหุนามอันดับที่ 1 ถึงอันดับที่ 3 ผลการศึกษาพบว่า ตัวแบบความถดถอยพหุนามอันดับที่ 2 เป็นตัวแบบความถดถอยพหุนามที่เหมาะสมกับความสัมพันธ์ n_{av} กับ p เมื่อ $e = 0.1$ ดังนี้

$$n_{av} = 9{,}528.93 - 1{,}824.10p + 196.57p^2 \text{ โดยมีค่า } R^2 = 0.9986$$

2. การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} เมื่อ $e = 0.05$

จากการศึกษาพบว่า ความสัมพันธ์ของ n_{av} กับ p เป็นไปในทิศทางเดียวกัน เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นอย่างรวดเร็ว ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้ง ซึ่งเป็นไปในทำนองเดียวกับความคลาดเคลื่อน $e = 0.1$ จึงศึกษา ตัวแบบความถดถอยพหุนามอันดับที่ 1 ถึงอันดับที่ 3 ผลการศึกษาพบว่า ตัวแบบความถดถอยพหุนามอันดับที่ 2 เป็นตัวแบบความถดถอยพหุนามที่เหมาะสมกับความสัมพันธ์ n_{av} กับ p เมื่อ $e = 0.05$ ดังนี้

$$n_{av} = 21{,}413.56 - 3{,}645.15p + 423.43p^2 \text{ โดยมีค่า } R^2 = 0.9982$$

3. การสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} เมื่อ $e = 0.01$

จากการศึกษาพบว่า ความสัมพันธ์ของ n_{av} กับ p เป็นไปในทิ ศทางเดียวกัน เมื่อ p มีจำนวนเพิ่มขึ้น n_{av} ก็มีขนาดเพิ่มขึ้นรวดเร็วมาก ลักษณะของความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้ง จึงศึกษาตัวแบบความถดถอยพหุนามอันดับที่ 1 ถึงอันดับที่ 3 ผลการศึกษาพบว่า ตัวแบบความถดถอยพหุนามอันดับที่ 3 เป็นตัวแบบความถดถอยพหุนามที่เหมาะสมกับความสัมพันธ์ n_{av} กับ p เมื่อ $e = 0.01$ คือ

$$n_{av} = 1{,}242{,}366.2 - 393{,}692.1p + 31{,}725.6p^2 - 245.7p^3 \text{ โดยมีค่า } R^2 = 0.9971$$

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

5.2 อภิปรายผล

จากผลการวิจัยจะพบว่า ในการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณคอนเวกซ์ที่มีขอบเขตจำกัด ซึ่งได้แก่ บนบริเวณสี่เหลี่ยม S_{rect} และบนบริเวณสามเหลี่ยม S_{tri} ลักษณะความสัมพันธ์ของ n_{av} กับ p เป็นเส้นโค้งแบบพหุนามที่ทุกกรณีสามารถอธิบายความสัมพันธ์ดังกล่าวด้วยตัวแบบความถดถอยพหุนามอันดับที่ 2 ยกเว้นกรณีที่มีการสุ่มตัวอย่างแบบฮิตแอนด์รันบนบริเวณสามเหลี่ยม S_{tri} เมื่อ $e = 0.01$ เพียงกรณีเดียวเท่านั้นที่ความสัมพันธ์ของ n_{av} กับ p เป็นตัวแบบความถดถอยพหุนามอันดับที่ 3 อย่างไรก็ตาม สัมประสิทธิของ p^3 มีค่าเป็นค่าลบ ดังนั้นแสดงว่าความซับซ้อนของการคำนวณไม่ถูกกระทบด้วยพจน์ p^3 และ หากเปรียบเทียบค่าสัมประสิ ทธิ์ของฟังก์ชันพหุนามที่ได้จากงานวิจัยนี้กับ ค่าสัมประสิทธิ์ในพหุนามอันดับที่ 3 ของลาสซ์โล ลอวอสซ์(1999) จะสังเกตได้ว่า ค่าสัมประสิทธิ์ในฟังก์ชันพหุนามในงานวิจัยนี้ต่ำกว่าค่าสัมประสิทธิ์ในพหุนามอันดับที่ 3 ของลาสซ์โล ลอวอสซ์อยู่มาก ดังนั้น ในทางปฏิบัติหากพิ จารณาการลู่เข้าของการสุ่มตัวอย่างแบบฮิตแอนด์รันโดยใช้ค่าเฉลี่ยของตัวอย่าง อัตราการลู่เข้าในทางปฏิบัติน่าจะมีอัตราการลู่เข้าได้เร็วกว่าอัตราการลู่เข้าในทางทฤษฎีที่ใช้ระยะทาง total variation

5.3 ข้อเสนอแนะ

1. ในการแก้ปัญหาความแปรปรวนร่วมของกระบวนการฮิตแอน ด์รันที่มีความสัมพันธ์กันนั้น อาจจะนำวิธี Regenerative Method จากหนังสือ Handbooks in Operations Research and Management Science Volume 13 Simulation มาใช้ในการแก้ไขปัญหาดังกล่าวได้

2. สำหรับกรณีที่กำหนด $e = 0.01$ หากเริ่มต้นกำหนดให้กระบวนการลู่เข้าสู่จุดที่ต้องการที่ระดับความคลาดเคลื่อน $e = 0.1$ ก่อน แล้วจากนั้นจึงปรับให้มีการลู่เข้าจุดที่ต้องการที่ระดับ $e = 0.05$ และ $e = 0.01$ ตามลำดับ อาจจะสามารถทำให้กระบวนการลู่เข้าสู่จุดที่ต้องการได้เร็วขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

เสกสรร เกียรติสุไพบูลย์. เอกสารประกอบการสอนวิชาจำลองเชิงสถิติ. กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย, 2551.

สุพล ดุรงค์วัฒนา. ตัวแบบและการวิเคราะห์ความถดถอย สำหรับการวิจัยชั้นสูง. กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย, 2549.

ภาษาอังกฤษ

Shane G. Henderson and Barry L. Nelson. Handbooks in Operations Research and Management Science. First Edition. Volume 13. Oxford: North-Holland, 2006.

Sheldon M. Ross. Introduction to Probability Models. 8th Edition. Volume 13. California: Academic, 2003.

Laszlo Lovasz. Hit-and-Run mixes fast. Math. Program Series A86(1999): 443-461.

Laszlo Lovasz and Santosh Vempala. Hit-and-Run is Fast and Fun. Technical Report MSR-TR-2003-05 (2003) : 1-27.

Robert L. Smith. Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions. Operations Research 32(1984): 1296-1308.

Robert L. Smith. The Hit-and-Run Sampler: A Globally Reaching Markov Chain Sampler for Generating Arbitrary Multivariate Distributions. Proceeding of the 1996 Winter Simulation Conference (1996): 260-264.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวพัชรี วงศ์พงศาลี เกิดวันที่ 14 ธันวาคม 2518 สำเร็จการศึกษาระดับปริญญาบัณฑิต หลักสูตรวิทยาศาสตรบัณฑิต (สถิติประยุกต์) จากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปีการศึกษา 2539 และเข้าศึกษาระดับปริญญามหาบัณฑิต หลักสูตรสถิติศาสตรมหาบัณฑิต ที่คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2550

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย