บทที่ 3

การดำเนินการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการเลือกแบนวิดธ์จำนวน 6 วิธี คือ การเลือกแบนวิดธ์ด้วยวิธี Direct plug-in วิธี Rules of thumb วิธี Least squares cross-validation วิธี Silverman's rules of thumb วิธี Biased cross-validation และวิธี Solve the equation plug-in ในการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นด้วยวิธีของ Zhang และคณะ (1999) ซึ่งเป็นวิธีที่ใช้ในการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นเพื่อแก้ไขปัญหาการประมาณความหนาแน่นบริเวณขอบเขต ในการวิจัยจะสร้างตัวแปรสุ่มจากฟังก์ชันการแจกแจงตัวแปรสุ่ม โดยสร้างเลขคล้ายสุ่มด้วยเทคนิคการจำลองมอนติคาร์โล (Monte carlo simulation) แต่ละวิธีการเลือกแบนวิดธ์จะกระทำภายใต้สถานการณ์ต่าง ๆ ซึ่งมีการทำซ้ำในสถานการณ์เดียวกันจำนวน 1,000 รอบ ทำการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดธ์สำหรับการประมาณความหนาแน่นด้วยวิธีของ Zhang และคณะ (1999) สำหรับฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวแปรสุ่มทั้ง 2 ฟังก์ชัน โดยเปรียบเทียบจากค่าความคลาดเคลื่อนกำลังสองรวมเฉลี่ย $(MISE)$ ขนาดตัวอย่างในการวิจัย 5 ขนาดตัวอย่าง คือ 30 , 50 , 100 , 150 และ 200

3.1 ขั้นตอนการดำเนินการวิจัย

ผู้วิจัยทำการทดลองโดยการสร้างข้อมูลที่มีการแจกแจงตามที่กำหนด และทำการวิเคราะห์ผลการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นด้วยโปรแกรมคอมพิวเตอร์ R version 2.10.1 การวิจัยแบ่งเป็น 4 ขั้นตอนดังนี้

1. สร้างข้อมูลของตัวแปรสุ่ม X จากฟังก์ชันที่กำหนดไว้ 2 การแจกแจง คือ

   1.1 ตัวแปรสุ่ม X ถูกสร้างจากฟังก์ชันการแจกแจงตัวแปรสุ่ม $f(x) = 5e^{-5x}$

   1.2 ตัวแปรสุ่ม X ถูกสร้างจากฟังก์ชันการแจกแจงตัวแปรสุ่ม

$$f(x) = \frac{5}{4}(1+15x)e^{-5x}$$

จำนวนข้อมูลที่สร้างขึ้นมี 5 ขนาด คือ 30 , 50 ,100 , 150 และ 200

2. ประมาณค่าแบนวิดธ์จากวิธีการเลือกแบนวิดธ์ 6 วิธี

2.1 การเลือกแบนวิดธ์ด้วยวิธี Direct plug-in

$$h_{DPI} = \left[\frac{R(K)}{\mu_2(K)^2 \hat{\psi}_4(g) n} \right]^{\frac{1}{5}}$$

2.2 การเลือกแบนวิดธ์ด้วยวิธี Rules of thumb

$$h_{ROT} = 1.06\hat{\sigma}(n^{-\frac{1}{5}})$$

2.3 การเลือกแบนวิดธ์ด้วยวิธี Least squares cross-validation

$$h_{LSCV} = \min_h [LSCV(h)]$$

2.4 การเลือกแบนวิดธ์ด้วยวิธี Silverman's rules of thumb

$$h_{SROT} = 0.9A(n^{-\frac{1}{5}})$$

2.5 การเลือกแบนวิดธ์ด้วยวิธี Biased cross-validation

$$h_{BCV} = \min_h [BCV(h)]$$

2.6 การเลือกแบนวิดธ์ด้วยวิธี Solve the equation plug-in

$$h_{STE} = \left[\frac{R(K)}{\mu_2(K)^2 \hat{S}_D(\hat{\alpha}_2(h_{STE})) n} \right]^{\frac{1}{5}}$$

3. ประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นด้วยวิธีของ Zhang และคณะ (1999) โดยใช้ค่าแบนวิดธ์ที่ได้จากวิธีการเลือกแบนวิดธ์ 6 วิธี

3.1 $$\tilde{f}_{DPI}(x)=\frac{1}{nh_{DPI}}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h_{DPI}}\right)+K\left(\frac{x+g(X_i)}{h_{DPI}}\right)\right\}$$

3.2 $$\tilde{f}_{ROT}(x)=\frac{1}{nh_{ROT}}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h_{ROT}}\right)+K\left(\frac{x+g(X_i)}{h_{ROT}}\right)\right\}$$

3.3 $$\tilde{f}_{LSCV}(x)=\frac{1}{nh_{LSCV}}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h_{LSCV}}\right)+K\left(\frac{x+g(X_i)}{h_{LSCV}}\right)\right\}$$

3.4 $$\tilde{f}_{SROT}(x)=\frac{1}{nh_{SROT}}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h_{SROT}}\right)+K\left(\frac{x+g(X_i)}{h_{SROT}}\right)\right\}$$

3.5 $$\tilde{f}_{BCV}(x)=\frac{1}{nh_{BCV}}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h_{BCV}}\right)+K\left(\frac{x+g(X_i)}{h_{BCV}}\right)\right\}$$

3.6 $$\tilde{f}_{STE}(x)=\frac{1}{nh_{STE}}\sum_{i=1}^{n}\left\{K\left(\frac{x-X_i}{h_{STE}}\right)+K\left(\frac{x+g(X_i)}{h_{STE}}\right)\right\}$$

4. คำนวณค่าความคลาดเคลื่อนกำลังสองรวมเฉลี่ย $(MISE)$

$$\frac{\sum_{r=1}^{1000}MISE_r(\tilde{f}(x))}{1000}$$

โดยที่ $$MISE_r(\tilde{f}(x))=\frac{\sum_{j=1}^{k}(\tilde{f}(y_j)-f(y_j))^2}{k}$$

เมื่อ y_j คือ ลำดับเลขคณิต ที่มีผลต่างร่วม คือ 0.005
เมื่อ $j = 1,2,...,k$ โดยที่
$y_1 = 0, y_2 = 0.005, ..., \min_k y_k > \max_{1 \le i \le n}(X_i) + 0.05$

$\max(X)$ คือ ค่ามากที่สุดของตัวแปรสุ่ม

r คือ การทำซ้ำรอบที่ r เมื่อ $r = 1,2,...,1000$

$MISE_r(\tilde{f}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r

$MISE_r(\tilde{f}_{DPI}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r โดยใช้แบนวิดธ์ที่ได้จากวิธี Direct plug-in

$MISE_r(\tilde{f}_{ROT}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r โดยใช้แบนวิดธ์ที่ได้จากวิธี Rules of thumb

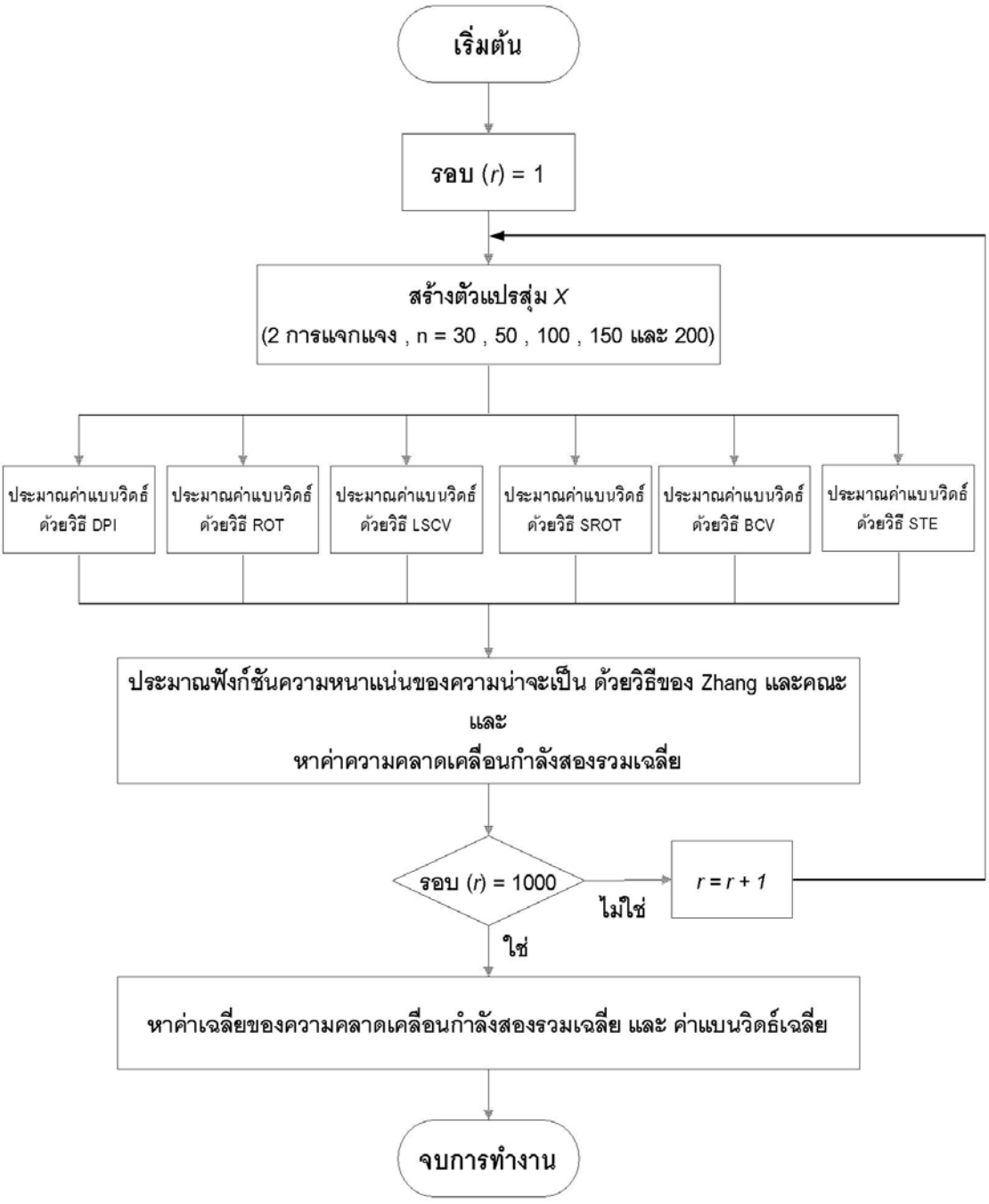
$MISE_r(\tilde{f}_{LSCV}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r โดยใช้แบนวิดธ์ที่ได้จากวิธี Least squares cross-validation

$MISE_r(\tilde{f}_{SROT}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r โดยใช้แบนวิดธ์ที่ได้จากวิธี Silverman's rules of thumb

$MISE_r(\tilde{f}_{BCV}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r โดยใช้แบนวิดธ์ที่ได้จากวิธี Biased cross-validation

$MISE_r(\tilde{f}_{STE}(x))$ คือ ค่าประมาณของความคลาดเคลื่อนกำลังสองรวมเฉลี่ยในการจำลองค่ารอบที่ r โดยใช้แบนวิดธ์ที่ได้จากวิธี Solve the equation plug-in

ขั้นตอนของการวิจัยแสดงเป็นแผนผังได้ดังภาพที่ 3.1

3.2 แผนภาพแสดงขั้นตอนการดำเนินการวิจัย

ภาพที่ 3.1
ขั้นตอนการดำเนินการวิจัย