# ใบรับรองวิทยานิพนธ์

**บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ**

เรื่อง การเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลของฟังก์ชันตัวแปรสุ่ม

โดย นางสาวปิยะฉัตร ลีลาศิลปศาสน์

ได้รับอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์

________________________ คณบดีบัณฑิตวิทยาลัย

(อาจารย์ ดร.มงคล หวังสถิตย์วงษ์)

21 พฤษภาคม 2550

คณะกรรมการสอบวิทยานิพนธ์

วินัย โพธิ์สุวรรณ
________________________ ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.วินัย โพธิ์สุวรรณ์)

________________________ กรรมการ

(ผู้ช่วยศาสตรจารย์วิจิตรา พลเยี่ยม)

________________________ กรรมการ

(รองศาสตราจารย์อุดิศักดิ์ พงศ์พูลผลศักดิ์)

________________________ กรรมการ

(รองศาสตราจารย์สอาด นิวิศพงศ์)

# การเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลของฟังก์ชันตัวแปรสุ่ม

นางสาวปิยะฉัตร ลีลาศิลปศาสน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ ภาควิชาสถิติประยุกต์
บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2549

ชื่อ : นางสาวปิยะฉัตร ลีลาศิลปศาสน์
ชื่อวิทยานิพนธ์ : การเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลของฟังก์ชันตัวแปรสุ่ม
สาขาวิชา : สถิติประยุกต์
สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ที่ปรึกษาวิทยานิพนธ์ : ผู้ช่วยศาสตราจารย์ ดร.วินัย โพธิ์สุวรรณ์
ผู้ช่วยศาสตราจารย์วิจิตรา พลเยี่ยม
ปีการศึกษา : 2549

## บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบวิธีการเลือกแบนวิดจ์ (Bandwidth) สำหรับการประมาณความหนาแน่นแบบเคอร์เนล 4 วิธี คือ วิธี Least Square Cross-Validation วิธี Plug-in วิธีSmoothed Bootstrap และวิธี Kernel Contrast สำหรับฟังก์ชันของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบเอกซ์โพเนนเชียล และการแจกแจงแบบโคชีมาตรฐาน และกำหนดใช้ฟังก์ชันเคอร์เนล 3 ฟังก์ชัน คือ ฟังก์ชันเคอร์เนลแบบควอติค แบบอีพานิชนิคอฟ และแบบเกาส์เซียน ทำการทดลองด้วยเทคนิคมอนติคาร์โล 1,000 รอบในแต่ละสถานการณ์ที่ศึกษา พิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) เป็นเกณฑ์ในการเปรียบเทียบ ผลการวิจัยสามารถสรุปได้ดังนี้

1. ในการประมาณค่าแบนวิดจ์ เมื่อใช้ฟังก์ชันเคอร์เนล 1 ฟังก์ชัน วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE และ MISE ต่ำที่สุด ในทุกขนาดตัวอย่างและทุกกรณีศึกษา และในการประมาณค่าแบนวิดจ์ เมื่อใช้ฟังก์ชันเคอร์เนล 2 ฟังก์ชัน วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่า MSE และ MISE ต่ำที่สุด ยกเว้นกรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ซึ่งวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะให้ค่า MSE ต่ำที่สุด
2. วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะให้ค่าแบนวิดจ์ที่มีขนาดเล็กที่สุด
3. เมื่อขนาดตัวอย่างมีขนาดใหญ่ขึ้น ค่า MSE และ MISE มีแนวโน้มที่จะลดลง

(วิทยานิพนธ์นี้มีจำนวนทั้งสิ้น 131 หน้า)

คำสำคัญ : การประมาณความหนาแน่น, ฟังก์ชันเคอร์เนล, การเลือกแบนวิดจ์

__________________________อาจารย์ที่ปรึกษาวิทยานิพนธ์

Name : Miss Piyachat Leelasilapasart

Thesis Title : Bandwidth Selection for Kernel Density Estimation for Function of Random Variable

Major Field : Applied Statistics

King Mongkut's Institute of Technology North Bangkok

Thesis Advisors: Assistant Professor Dr. Winai Bodhisuwan

Assistant Professor Wichitra Phonyiem

Academic Year : 2006

## Abstract

The objective of this research was to compare the bandwidth selection methods. The bandwidth selections compared were least square cross-validation, plug-in, smoothed bootstrap and kernel contrast methods for kernel density estimation for a function of a random variable. The distributions studied were standard normal, exponential and cauchy distributions. The kernel functions used were Gaussian, epanechnikov and quatic. Using 1,000 monte carlo simulations. We compared by mean squared error (MSE) and mean integrated squared error (MISE). The results of this study were as follow:

1. The smoothed bootstrap method gave the smallest MSE and MISE when one kernel function was used for all of the sample sizes. The kernel contrast method gave the smallest MSE and MISE when two kernel functions were used for all of the sample sizes except when the data had the cauchy Distribution for which least square cross-validation method gave the smallest MSE.
2. The bandwidth value of the least square cross-validation method always gave the smallest value of bandwidth.
3. As the sample size increased, the MSE and MISE decreased.

(Total 131 pages)

Keyword : density estimation, kernel function, bandwidth selection

________________________________________Advisor

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาของ ผู้ช่วยศาสตราจารย์ ดร. วินัย โพธิ์สุวรรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และผู้ช่วยศาสตราจารย์วิจิตรา พลเยี่ยม อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ได้ให้คำแนะนำ และข้อคิดเห็นต่างๆ ที่เป็นประโยชน์ในการวิจัย ตลอดจนช่วยเหลือและแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดี จนกระทั่งวิทยานิพนธ์นี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ผู้วิจัยขอขอบพระคุณ รองศาสตราจารย์อดิศักดิ์ พงศ์พูลผลศักดิ์ และ รองศาสตราจารย์สะอาด นิวิศพงศ์ ในฐานะกรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำและตรวจแก้วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

ผู้วิจัยขอขอบพระคุณ ครู และอาจารย์ภาควิชาสถิติประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ ทุกๆ ท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้

สุดท้ายผู้วิจัยขอกราบขอบพระคุณบิดา มารดา คุณยายอวล สมศรี และญาติพี่น้อง ที่ส่งเสริมและสนับสนุนการศึกษาในทุกๆ ด้าน รวมทั้งให้กำลังใจและห่วงใยผู้วิจัยเสมอมา และขอขอบคุณ เพื่อนๆ ทุกคนที่ให้กำลังใจและให้ความช่วยเหลืออย่างดีมาโดยตลอด

ปิยะฉัตร ลีลาศิลปศาสน์

# สารบัญ

**สารบัญ (ต่อ)**

**สารบัญตาราง**

## สารบัญตาราง (ต่อ)

## สารบัญตาราง (ต่อ)

**สารบัญตาราง (ต่อ)**

## สารบัญภาพ

## สารบัญภาพ (ต่อ)

ภาพที่ หน้า

## สารบัญภาพ (ต่อ)

## สารบัญภาพ (ต่อ)

**สารบัญภาพ (ต่อ)**

# บทที่ 1

# บทนำ

## 1.1 ความเป็นมาและความสำคัญของปัญหา

ฟังก์ชันการแจกแจงความหนาแน่นของความน่าจะเป็น (Probability Density Function) เป็นฟังก์ชันที่ใช้อธิบายถึงลักษณะของตัวแปรสุ่ม (Random Variable) ถ้า $X$ เป็นตัวแปรสุ่มซึ่งมีฟังก์ชันความหนาแน่นของความน่าจะเป็น $f(x)$ จะหาความน่าจะเป็นที่ตัวแปรสุ่ม $X$ มีค่าอยู่ระหว่าง $a$ และ $b$ ได้โดย

$$P(a < X < b) = \int_a^b f(x)dx \text{ เมื่อ } a < b$$

ในกรณีที่ไม่ทราบรูปแบบของ $f(x)$ จะต้องทำการประมาณขึ้น การประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็น จะเรียกว่า การประมาณความหนาแน่น (Density Estimation)

การประมาณความหนาแน่น เป็นวิธีการสร้างฟังก์ชันความหนาแน่น $f(x)$ จากตัวอย่างสุ่ม (Random Sample) ซึ่งสามารถใช้ได้ทั้งวิธีการทางพาราเมตริก และวิธีการทางด้านนอนพาราเมตริก (Silverman, 1986)

วิธีการประมาณความหนาแน่นแบบพาราเมตริก ต้องคำนึงถึงค่าที่คาดว่าจะเป็นของพารามิเตอร์ของการแจกแจง เช่น การแจกแจงแบบปกติ ซึ่งมีพารามิเตอร์ค่าเฉลี่ย $\mu$ และพารามิเตอร์ความแปรปรวน $\sigma^2$ ที่ไม่ทราบค่า ในการประมาณความหนาแน่นจึงต้องเริ่มจากประมาณค่าเฉลี่ย $\mu$ และความแปรปรวน $\sigma^2$ จากชุดข้อมูล จากนั้นนำไปแทนที่พารามิเตอร์ในฟังก์ชันแจกแจงแบบปกติ ทำให้การประมาณความหนาแน่นแบบพาราเมตริกไม่แตกต่างวิธีการประมาณค่าพารามิเตอร์ของการแจกแจงในสถิติอนุมาน (Inferential Statistics)

ส่วนวิธีการประมาณความหนาแน่นแบบนอนพาราเมตริกไม่ต้องคำนึงถึงค่าพารามิเตอร์ใดๆ กล่าวคือ การประมาณความหนาแน่นแบบนอนพาราเมตริกเป็นการประมาณฟังก์ชันความหนาแน่นจากชุดข้อมูลโดยตรง

วิธีการประมาณความหนาแน่นแบบนอนพาราเมตริกที่นิยมใช้ในปัจจุบันมีหลายวิธี เช่น วิธีการแบบฮีสโตแกรม (Histogram) และวิธีการประมาณอย่างง่าย (Naïve Estimator) เป็นต้น แต่

วิธีการประมาณความหนาแน่นที่นิยมนำมาใช้มากที่สุดคือ วิธีการประมาณความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation) เสนอโดย Rosenblatt ในปี ค.ศ. 1956 โดยการประมาณความหนาแน่นแบบเคอร์เนลจะมีการถ่วงน้ำหนักข้อมูลในแต่ละตัวอย่างสุ่ม ซึ่งข้อดีของการประมาณความหนาแน่นวิธีนี้ คือ ได้ฟังก์ชันโครงสร้างข้อมูลทางคณิตศาสตร์ที่สามารถปรับเปลี่ยนรูปแบบได้ง่าย โดยมีตัวปรับแบบเคอร์เนล (Kernel Smoother) เรียกว่า แบนวิดจ์ (Bandwidth) ซึ่งมีรูปแบบทั่วไป ดังนี้ (Wolfgang Hardle, 1990)

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$$

โดยที่ $h$ คือ ค่าแบนวิดจ์ (Bandwidth)

$K(X)$ คือ ฟังก์ชันเคอร์เนล

ถ้า $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันการแจกแจงความหนาแน่นของความน่าจะเป็น $f(x)$ ซึ่งไม่ทราบรูปแบบ ตัวประมาณค่าความหนาแน่นแบบเคอร์เนล (Kernel Estimator) ของ $f(x)$ คือ

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \qquad (1\text{-}1)$$

โดยที่ $h$ คือ ค่าแบนวิดจ์ (Bandwidth)

$x$ คือ ตัวแปรที่ไม่ทราบค่า

$X_i$ คือ ตัวอย่างสุ่ม $X$ โดยที่ $i$ มีค่าตั้งแต่ $i = 1, ..., n$

$K$ คือ ฟังก์ชันเคอร์เนล (Kernel function) ในกรณีนี้จะเรียก $K$ เป็นฟังก์ชันเคอร์เนลลำดับที่ $k$

หลักการของประมาณความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation) คือ หาโครงสร้างของชุดข้อมูล เมื่อไม่ทราบฟังก์ชันความหนาแน่นของความน่าจะเป็น ซึ่งจากการศึกษาทั้งทางด้านทฤษฎีและการจำลอง Ahmad and Mugdadi (2004) พบว่าในกรณีที่ตัวแปรสุ่มเป็นตัวแปรสุ่มที่เป็นอิสระและมีการแจกแจงเดียวกัน ฟังก์ชันเคอร์เนลที่แตกต่างกัน ไม่ใช่สิ่งที่สำคัญใน

การประมาณความหนาแน่น แต่สิ่งที่สำคัญที่สุดในการประมาณแบบความหนาแน่นแบบเคอร์เนล คือ การเลือกค่าพารามิเตอร์ปรับหรือแบนวิดจ์ที่เหมาะสม

สาเหตุที่ทำให้การเลือกแบนวิดจ์มีความสำคัญต่อการแสดงลักษณะของข้อมูลสำหรับตัวประมาณค่าความหนาแน่นแบบเคอร์เนล ก็คือ ถ้าค่าแบนวิดจ์มีขนาดเล็ก อาจจะทำให้การประมาณความหนาแน่นที่ได้มีลักษณะที่แกว่งและแสดงให้เห็นถึงลักษณะที่ผิดพลาด แต่ในทางกลับกันถ้าค่าแบนวิดจ์มีขนาดใหญ่ ก็จะส่งผลให้ค่าความเอนเอียง (Bias) มากขึ้น ส่งผลให้โครงสร้างของข้อมูลผิดพลาด ทำให้ไม่เห็นลักษณะโครงสร้างที่แท้จริงของข้อมูล ซึ่ง Wolfgang Hardle (1990) ได้สรุปลักษณะของแบนวิดจ์แบ่งเป็นกรณีได้ดังนี้

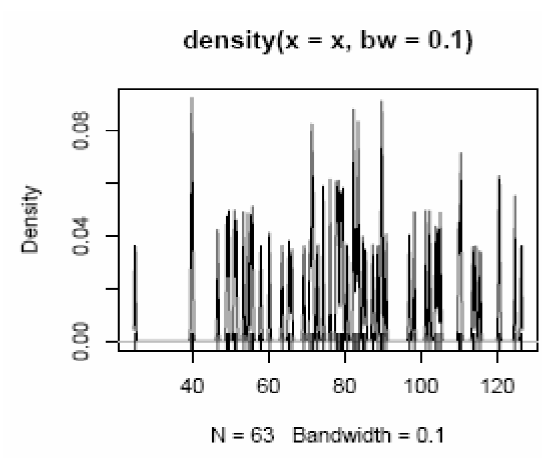
1. ค่าแบนวิดจ์มีค่าเข้าใกล้ 0 $(h \to 0)$ ลักษณะโครงสร้างข้อมูลจะมีลักษณะวุ่นวายแกว่งไปมาไม่แสดงถึงลักษณะของข้อมูลอย่างแท้จริ

2. เมื่อค่าแบนวิดจ์มีค่าเล็ก ลักษณะโครงสร้างข้อมูลจะปรับเรียบมากขึ้น

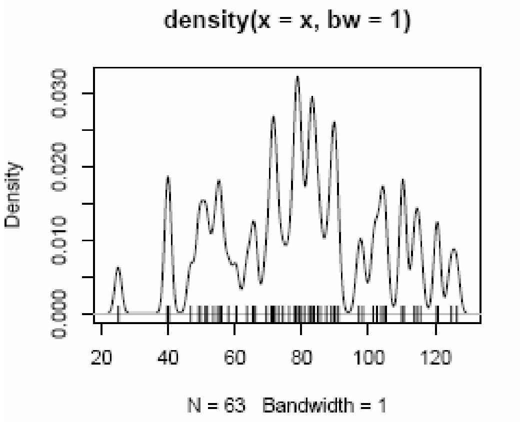
3. เมื่อค่าแบนวิดจ์มีค่ามาก ลักษณะโครงสร้างข้อมูลจะปรับเรียบ และได้ผลการประมาณที่ดี

4. เมื่อค่ามีค่าเข้าใกล้ระยะอนันต์ $(h \to \infty)$ ลักษณะโครงสร้างข้อมูลจะแบนราบ

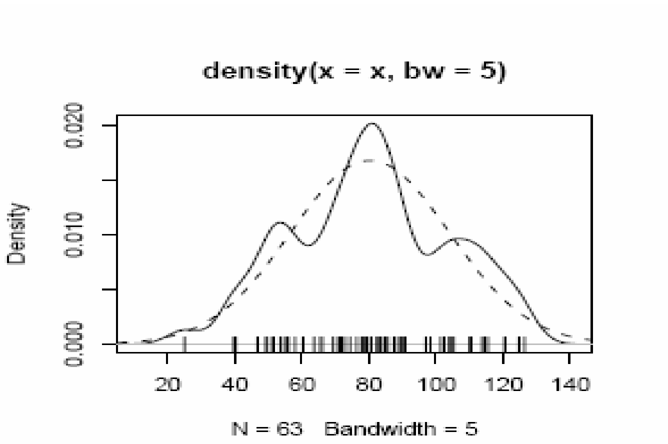
ภาพที่ 1-1 ถึง 1-4 เป็นภาพที่แสดงถึงการประมาณค่าความหนาแน่นที่มีค่าแบนวิดจ์ที่แตกต่างกัน โดยนำข้อมูลปริมาณของหิมะ ในเมืองบัพฟาโล มลรัฐนิวยอร์ก ประเทศสหรัฐอเมริกา ตั้งแต่ปี 1910 – 1972



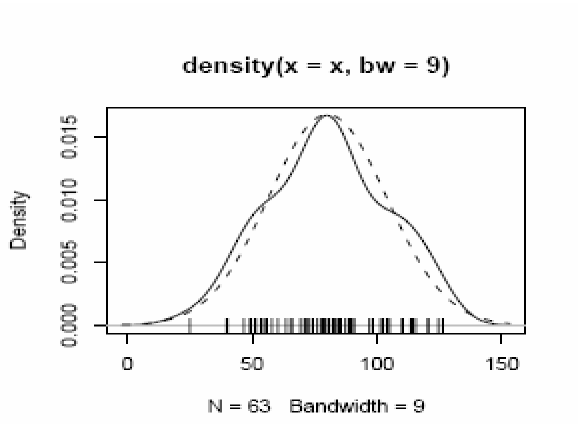

**ภาพที่ 1-1** การประมาณความหนาแน่นที่มีค่าแบนวิดจ์มีค่าเข้าใกล้ 0




**ภาพที่ 1-2** การประมาณความหนาแน่นที่มีค่าแบนวิดจ์มีค่าเล็ก




**ภาพที่ 1-3** การประมาณความหนาแน่นที่มีค่าแบนวิดจ์มีค่ามาก




**ภาพที่ 1-4** การประมาณความหนาแน่นที่มีค่าแบนวิดจ์มีค่าเข้าใกล้ระยะอนันต์

ในการประมาณค่าเพื่ออธิบายถึงลักษณะของตัวประมาณค่าแบบเคอร์เนล จำเป็นที่จะต้องเลือกเครื่องมือวัดความแตกต่างระหว่างฟังก์ชันความหนาแน่นน่าจะเป็นที่แท้จริงกับฟังก์ชันความหนาแน่นน่าจะเป็นที่ประมาณขึ้น ดังนั้นจำเป็นต้องพิจารณาถึงความคลาดเคลื่อนของการประมาณค่าในการเลือกแบนวิดจ์ ซึ่งการหาความคลาดเคลื่อนสามารถทำได้หลายวิธี เช่น ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (Mean Integrated Square Error: MISE) โดย

$$MSE\left(\hat{f}(x)\right) = E\left[\hat{f}_h(x) - f(x)\right]^2 \tag{1-2}$$

$$MISE\left(\hat{f}(x)\right) = E\int\left[\hat{f}_h(x) - f(x)\right]^2 dx \tag{1-3}$$

สำหรับการวิจัยครั้งนี้ ผู้วิจัยมีความสนใจศึกษาวิธีการเลือกแบนวิดจ์สำหรับการประมาณค่าความหนาแน่นแบบเคอร์เนลของฟังก์ชันของตัวแปรสุ่ม โดยจะศึกษาวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation, วิธีการเลือกแบนวิดจ์แบบ Plug-in, วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จากนั้นทำการเปรียบเทียบค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ภายใต้ฟังก์ชันเคอร์เนลต่างๆ แล้วเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์ โดยพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (Mean Integrated Square Error: MISE) ซึ่งวิธีการเลือกแบนวิดจ์วิธีการใดให้ค่า MSE และ MISE ต่ำที่สุด จะเป็นวิธีการเลือกแบนวิดจ์ที่มีประสิทธิภาพมากที่สุด

## 1.2 วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมาย ดังนี้

1.2.1 เพื่อศึกษาวิธีการเลือกแบนวิดจ์ ในการประมาณค่าความหนาแน่นแบบเคอร์เนล คือ วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation, วิธีการเลือกแบนวิดจ์แบบ Plug-in, วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast

1.2.2 เพื่อเปรียบเทียบค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ในการประมาณค่าความหนาแน่นแบบเคอร์เนล ด้วยการเปรียบเทียบค่า MSE และ MISE

## 1.3 สมมติฐานการวิจัย

ภายใต้สถานการณ์เดียวกัน วิธีการเลือกแบนวิดจ์ที่ต่างกัน จะให้ผลประสิทธิภาพในการเลือกแบนวิดจ์ที่แตกต่างกัน

## 1.4 ขอบเขตของการวิจัย

1.4.1 กำหนดให้ตัวอย่างสุ่มเป็นตัวอย่างสุ่มที่เป็นอิสระต่อกัน และมีฟังก์ชันการแจกแจงน่าจะเป็นเดียวกัน (independent and identically distributed; iid)

1.4.2 กำหนดให้ใช้ฟังก์ชันเคอร์เนล 3 ฟังก์ชัน คือ

1.4.1.1 ฟังก์ชันเกาส์เซียน (Gaussian Kernel Function) โดยมีรูปแบบของฟังก์ชัน คือ $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$

1.4.1.2 ฟังก์ชันอีพานิชนิคอฟ (Epanechnikov Kernel Function) โดยมีรูปแบบของฟังก์ชัน คือ $\frac{3}{4}\left(1-u^2\right)I\left(|u|\le 1\right)$

1.4.1.3 ฟังก์ชันควอติค (Quartic Kernel Function) โดยมีรูปแบบของฟังก์ชัน คือ $\frac{15}{16}\left(1-u^2\right)^2 I\left(|u|\le 1\right)$

1.4.3 ตัวแปรสุ่มที่นำมาศึกษามาจาก 3 การแจกแจง คือ การแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบเอกซ์โพเนนเชียล และการแจกแจงแบบโคชี

1.4.4 ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 10, 25, 50, 100 และ 500

1.4.5 ฟังก์ชันของตัวแปรสุ่มที่ศึกษา คือ $g\left(X_1, X_2, ..., X_m\right) = \sum_{i=1}^{m} X_i$ ที่ $m = 2$

1.4.6 กำหนดค่า $k = 2$

## 1.5 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพ

การวิจัยครั้งนี้จะใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (Mean Integrated Square Error: MISE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพสำหรับการเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลของฟังก์ชันของตัวแปรสุ่ม ภายใต้สถานการณ์ที่กำหนดขึ้น วิธีการเลือกแบนวิดจ์วิธีใดที่ให้ค่า MSE และ MISE ต่ำที่สุดจะเป็นวิธีการเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลของฟังก์ชันของตัวแปรสุ่มที่มีประสิทธิภาพมากกว่า

**1.6 ประโยชน์ของการวิจัย**

1.6.1 เพื่อเป็นแนวทางในการตัดสินใจในการเลือกค่าแบนวิดจ์ที่เหมาะสมสำหรับการประมาณความหนาแน่นแบบเคอร์เนลและประสิทธิภาพในการในการประมาณฟังก์ชันความหนาแน่น

**1.7 นิยามศัพท์**

1.7.1 ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Estimation) คือ ฟังก์ชันที่อธิบายถึงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่อง (Continuous Random Variable) สามารถนำเสนอเป็นกราฟโค้งความถี่ต่อเนื่องและค่าความน่าจะเป็นเป็นค่าของพื้นที่ใต้โค้ง

1.7.2 การประมาณความหนาแน่น (Density Estimation) คือ กระบวนการทางด้านนอนพาราเมตริกในการประมาณฟังก์ชันความหนาแน่น (Density Function) ของการแจกแจงความน่าจะเป็น

1.7.3 ตัวอย่างสุ่ม (Random Sample) คือ เซตของตัวแปรเชิงสุ่ม $X_1, X_2, ..., X_n$ ที่เป็นอิสระกันในเชิงสถิติ (Identically Independent) และแต่ละตัวมีความหนาแน่นเช่นเดียวกันกับความหนาแน่นของประชากร

1.7.4 ตัวแปรสุ่ม (Random Variable) คือ ฟังก์ชันที่มีโดเมนเป็นแซมเปิลสเปซ (Sample Space) ที่มีฟังก์ชันความน่าจะเป็นและมีเรนจ์เป็นเซตของจำนวนจริง

1.7.5 ตัวประมาณเคอร์เนล (Kernel Estimator) คือ วิธีทางนอนพาราเมตริกในการประมาณฟังก์ชันความหนาแน่นของการแจกแจงความน่าจะเป็น โดยคำนวณจากตัวอย่างขนาด $n$ โดยแทนที่แต่ละค่าของข้อมูลด้วย “เคอร์เนล” ของพื้นที่ $\frac{1}{n}$ ผลลัพธ์จะแสดงโค้งของฟังก์ชันความหนาแน่นที่ประมาณได้ ซึ่งคล้ายกับการปรับขนาดแท่งความถี่ฮีสโตแกรม

# บทที่ 2
# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดทฤษฎีต่างๆ วิธีการเลือกแบนวิดจ์ และการแจกแจงที่ใช้ในการวิจัยรวมทั้งผลงานการวิจัยที่เกี่ยวข้อง ซึ่งมีรายละเอียดต่างๆ ดังนี้

### 2.1 การประมาณความหนาแน่น (Density Estimation)

ฟังก์ชันการแจกแจงความหนาแน่นของความน่าจะเป็น (Probability Density Function) เป็นฟังก์ชันที่ใช้อธิบายถึงลักษณะของตัวแปรสุ่ม เมื่อพิจารณาตัวอย่างสุ่ม $X$ ซึ่งมีฟังก์ชันความหนาแน่นน่าจะเป็นและความน่าจะเป็น

$$P(a < X < b) \ = \ \int_a^b f(x)dx \ \text{ เมื่อ } a < b \tag{2-1}$$

ถ้า $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด $n$ จากประชากร โดยไม่ทราบฟังก์ชันความหนาแน่นน่าจะเป็น ต้องทำการประมาณฟังก์ชันความหนาแน่น โดยวิธีการประมาณนี้ เรียกว่า การประมาณความหนาแน่น (Density Estimation) เป็นวิธีที่สร้างฟังก์ชันความหนาแน่นจากตัวอย่างสุ่มสามารถแบ่งเป็น 2 วิธี คือ วิธีการทางด้านพาราเมตริก และวิธีการทางด้านนอนพาราเมตริก (Silverman, 1986)

การประมาณความหนาแน่นแบบพาราเมตริก จะต้องคำนึงถึงค่าพารามิเตอร์ของการแจกแจงที่คาดว่าจะเป็น เช่น การแจกแจงแบบปกติ ซึ่งมีพารามิเตอร์ค่าเฉลี่ย $\mu$ และพารามิเตอร์ความแปรปรวน $\sigma^2$ ที่ไม่ทราบค่า ในการประมาณความหนาแน่นจึงต้องเริ่มจาก ประมาณค่าเฉลี่ย $\mu$ และความแปรปรวน $\sigma^2$ จากชุดข้อมูล จากนั้นนำไปแทนที่พารามิเตอร์ในฟังก์ชันแจกแจงแบบปกติ

การประมาณความหนาแน่นแบบนอนพาราเมตริก ไม่ต้องคำนึงถึงค่าพารามิเตอร์ใดๆ กล่าวคือ การประมาณความหนาแน่นแบบนอนพาราเมตริกเป็นการประมาณฟังก์ชันความหนาแน่นจากชุดข้อมูลโดยตรง

วิธีการประมาณความหนาแน่นแบบนอนพาราเมตริกที่ใช้กันในปัจจุบันมีหลายวิธี ซึ่งส่วนใหญ่มีพื้นฐานจากการประมาณความหนาแน่นแบบฮีสโตแกรม

การประมาณค่าความหนาแน่นแบบฮีสโตแกรม (Histogram) เป็นวิธีการประมาณความหนาแน่นแบบนอนพาราเมตริกที่เก่าแก่ที่สุด (Wand and Jones, 1995) ซึ่งถ้ากำหนดให้ $X_1,\ldots,X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันและมีฟังก์ชันการแจกแจกน่าจะเป็นเดียวกัน (Independent and Identically Distributed; iid)

กำหนดให้ $x_0$ เป็นจุดเริ่มต้น และ $h$ เป็นความกว้างของแท่ง (Binwidth) โดยที่ $h$ มีค่ามากกว่า 0 และกำหนดให้ความกว้างของแท่งมีค่าแบบช่วงระหว่าง $[x_0+mh, x_0(m+1)h)$ โดยที่ $m$ เป็นจำนวนเต็มบวกหรือจำนวนเต็มลบ

การคำนวณเริ่มต้นจากกำหนดให้ $a_0 < x_{(1)}$ และ $a_n > x_{(n)}$ โดยที่ $a_0 < a_1 < \cdots < a_n$ โดย $x_{(1)} = \min(X_1,\ldots,X_n)$ และ $x_{(n)} = \max(X_1,\ldots,X_n)$ จากนั้นแบ่งช่วงของ $(a_0, a_n]$ ออกเป็น $k$ ช่วง โดยแต่ละช่วงมีความกว้างเท่าๆ กัน โดยแต่ละช่วงมีความกว้างเท่ากับ $h = \dfrac{a_n - a_0}{k}$ แต่ก็สามารถกำหนดให้ค่าความกว้างแต่ละช่วงมีค่าไม่เท่ากันก็ได้ แต่โดยส่วนมากจะนิยมใช้ในกรณีที่กำหนดให้มีค่าความกว้างของช่วงเท่ากัน

จะกล่าวว่าอัตราส่วนของจำนวนของตัวแปรสุ่มที่มีค่าอยู่ในแต่ละช่วงต่อจำนวนตัวแปรสุ่มทั้งหมด นั่นคือ ความถี่ของช่วงที่ $j$ จะมีค่าเท่ากับ $\dfrac{n_j}{n}$ โดยที่ $n_j$ คือจำนวนของตัวแปรสุ่มที่อยู่ในช่วงที่ $j$ ดังนั้นตัวประมาณความหนาแน่นแบบฮีสโตแกรม คือ

$$\begin{aligned}\hat{f}(x) &= \sum_{j=1}^{k} \frac{n_j}{nh_j} \mathrm{I}_{(a_0,a_n]}(x) \\ &= \frac{1}{n}\sum_{j=1}^{k}\frac{1}{h}\mathrm{I}_{(a_0,a_n]}(x)\sum_{i=1}^{n}\mathrm{I}_{(a_0,a_n]}(X_i)\end{aligned}$$

โดยที่ $h$ คือ ความกว้างของแท่ง

$n$ คือ จำนวนข้อมูลทั้งหมด

$\mathrm{I}_{(a_0,a_n]}$ คือ จำนวนข้อมูลที่อยู่ในช่วงข้อมูล

## 2.2 การประมาณความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation)

การประมาณความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation) เป็นวิธีการประมาณความหนาแน่นที่นิยมใช้กันมากที่สุด (Silverman,1986) เป็นวิธีการที่สังเกตแต่ละค่าจะมีน้ำหนักถ่วง เสนอเป็นครั้งแรกโดย Rosenblatt ในปี ค.ศ. 1956 ตัวประมาณค่าความหนาแน่นแบบเคอร์เนล ( $\hat{f}_h(x)$ ) คือ

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{2-2}$$

โดยที่ $h$ คือ ค่าพารามิเตอร์ปรับเรียบหรือแบนวิดจ์ (Smoother Parameter or Bandwidth)

$x$ คือ ตัวแปรที่ไม่ทราบค่า $x$

$X_i$ คือ ตัวแปรสุ่ม $X$ โดยที่ $i = 1, 2, \ldots, n$

$K$ คือ ฟังก์ชันเคอร์เนล (Kernel Function)

ทั้งนี้ Wolfgang Hardle (1990) ได้สรุปฟังก์ชันเคอร์เนล ดังตารางที่ 2-1

**ตารางที่ 2-1** ตารางรูปแบบของฟังก์ชันเคอร์เนล

| ฟังก์ชันเคอร์เนล | $K(u)$ |
|---|---|
| 1. ฟังก์ชันเคอร์เนลแบบอีพานีชนิคอฟ (Epanechnikov) | $\frac{3}{4}(1-u^2) I(\lvert u\rvert \le 1)$ |
| 2. ฟังก์ชันเคอร์เนลแบบควอติค (Quartic) | $\frac{15}{16}(1-u^2)^2 I(\lvert u\rvert \le 1)$ |
| 3. ฟังก์ชันเคอร์เนลไตรเวท (Triweight) | $\frac{35}{32}(1-u^2)^3 I(\lvert u\rvert \le 1)$ |
| 4. ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian) | $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$ |
| 5. ฟังก์ชันเคอร์เนลแบบสามเหลี่ยม (Triangular) | $(1-\lvert u\rvert) I(\lvert u\rvert \le 1)$ |
| 6. ฟังก์ชันเคอร์เนลแบบยูนิฟอร์ม (Uniform) | $\frac{1}{2} I(\lvert u\rvert \le 1)$ |
| 7. ฟังก์ชันเคอร์เนลแบบคอซายอัส (Cosinus) | $\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right) I(\lvert u\rvert \le 1)$ |

โดยที่ $u = \frac{x - X_i}{h}$

จากตารางที่ 2-1 ที่แสดงถึงฟังก์ชันเคอร์เนลชนิดต่างๆ สามารถหาสมการฟังก์ชันเคอร์เนลได้จาก

$$K(u,p) = \frac{\left(1-u^2\right)^p}{2^{2p+1}B(p+1,p+1)} \text{ โดยที่ } \{|u \le 1|\} \tag{2-3}$$

โดยที่ $u$ คือ ตัวแปร $u$ โดยที่ $u = \dfrac{x - X_i}{h}$

$p$ คือ ค่าคงที่ $p$ โดยที่ $p$ มีค่าตั้งแต่ 0 เป็นต้นไป

และ $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ โดยที่ $\Gamma(c) = (c-1)!$ เมื่อ $c$ เป็นค่าคงที่ ดังนั้นฟังก์ชันเคอร์เนลมีรูปแบบแตกต่างกันไปตามลักษณะแนวโน้มของการแจกแจงของข้อมูล ยกตัวอย่างเช่น

ในกรณีที่ $p = 0$ จากสมการที่ 2-3 จะได้ว่า

$$K(u,p) = \frac{\left(1-u^2\right)^p}{2^{2p+1}B(p+1,p+1)} \text{ โดยที่ } \{|u \le 1|\}$$

$$K(u,0) = \frac{\left(1-u^2\right)^0}{2^{2(0)+1}B(0+1,0+1)}$$

$$= \frac{1}{2B(1,1)}$$

จาก $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ จะกล่าวได้ว่า $B(1,1) = \dfrac{\Gamma(1)\Gamma(1)}{\Gamma(1+1)}$

$$K(u,0) = \frac{1}{2} \text{ โดยที่ } \{|u \le 1|\}$$

ฟังก์ชันเคอร์เนลที่ได้ก็คือ ฟังก์ชันเคอร์เนลแบบยูนิฟอร์ม

ในกรณีที่ $p = 1$ จากสมการที่ 2-3 จะได้ว่า

$$K(u,p) = \frac{\left(1-u^2\right)^p}{2^{2p+1}B(p+1,p+1)} \text{ โดยที่ } \{|u \le 1|\}$$

$$K(u,1) = \frac{\left(1-u^2\right)^1}{2^{2(1)+1}B(1+1,1+1)}$$

$$K(u,1) = \frac{\left(1-u^2\right)}{2^3 B(2,2)}$$

จาก $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ จะกล่าวได้ว่า $B(1,1)=\frac{\Gamma(2)\Gamma(2)}{\Gamma(2+2)}$

$$K(u,1) = \frac{3}{4}\left(1-u^2\right) \text{ โดยที่ } \{|u \le 1|\}$$

ฟังก์ชันเคอร์เนลที่ได้ก็คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

คุณสมบัติของฟังก์ชันเคอร์เนล

1. ฟังก์ชันเคอร์เนลเป็นฟังก์ชันแบบสมมาตร โดยมีค่าระหว่าง 0 ถึง 1 ซึ่งกล่าวได้ว่า $K(u)=K(-u)$

2. ถ้าฟังก์ชันเคอร์เนลเป็นฟังก์ชันความหนาแน่น ตัวประมาณค่าความหนาแน่นแบบเคอร์เนลจะเป็นฟังก์ชันความหนาแน่นด้วยเช่นกัน กล่าวคือ ถ้า $\int_{-\infty}^{\infty} K(u)du=1$ แล้ว $\int_{-\infty}^{\infty} \hat{f}_h(u)\,du=1$

3. เมื่อกำหนดค่าแบนวิดจ์ $h$ และฟังก์ชันความหนาแน่นเคอร์เนล $K$ ในข้อมูล 1 ชุด จะมีฟังก์ชันเคอร์เนลได้เพียง 1 ฟังก์ชัน

4. ถ้าฟังก์ชันเคอร์เนล $K$ สามารถหาอนุพันธ์ได้อย่างต่อเนื่อง ตัวประมาณค่าความหนาแน่นแบบเคอร์เนลก็จะสามารถหาอนุพันธ์อย่างต่อเนื่องเช่นกัน

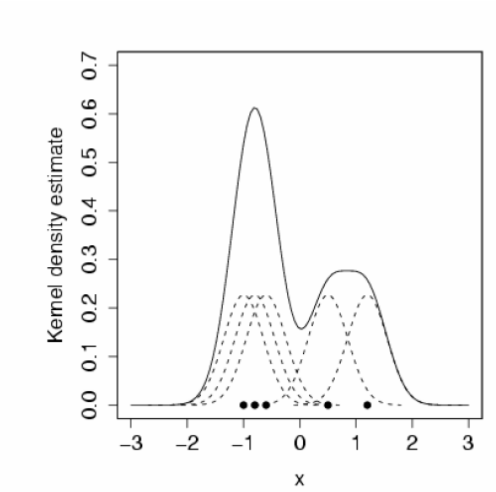
5. $\int_{-\infty}^{\infty} u^j K(u)\,du = 0$ สำหรับ $j=1,...,k-1$

6. $\int_{-\infty}^{\infty} u^k K(u)\,du \neq 0$

โดยทั่วไปแล้วจะเรียกฟังก์ชันเคอร์เนล $K$ ว่าฟังก์ชันเคอร์เนลลำดับที่ $k$ ซึ่งส่วนใหญ่จะศึกษาในกรณีที่ $k=2$ เนื่องจากคุณสมบัติของฟังก์ชันเคอร์เนล นั่นคือ $\int_{-\infty}^{\infty} u^k K(u)\,du \neq 0$ อาจจะกล่าวได้ว่าเมื่อ $k=2$ ค่าของ $K$ ที่ได้จะมีค่าเป็นบวก กล่าวคือ ฟังก์ชันเคอร์เนล $K$ จะกลายเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น และตัวประมาณความหนาแน่นแบบเคอร์เนลจะเข้าไปแทนที่ที่ค่าสังเกต $X_i$ แต่ละค่า แต่สำหรับกรณีที่ $k>2$ ค่าของ $K$ ที่ได้อาจจะมีค่าเป็นค่า

ลบและส่งผลให้ตัวประมาณค่าความหนาแน่นแบบเคอร์เนลไม่มีการเคลื่อนที่ และค่าประมาณความหนาแน่นแบบเคอร์เนลก็จะเป็นค่าลบด้วยเช่นกัน

ในการวิจัยนี้จะศึกษาการประมาณความหนาแน่นแบบเคอร์เนลแบบตัวแปรสุ่มเดียว ดังแสดงในภาพที่ 2-4 และในกรณีที่ $k=1$ และ $k=2$ โดยที่เมื่อ $k=1$ จะได้ว่า $\int_{-\infty}^{\infty} K(x)\,dx \neq 0$ และเมื่อ $k=2$ จะได้ว่า $\int_{-\infty}^{\infty} x^2 K(x)\,dx \neq 0$




**ภาพที่ 2-1** การประมาณความหนาแน่นแบบเคอร์เนลแบบตัวแปรสุ่มเดียว

2.2.1 ฟังก์ชันเคอร์เนลที่ใช้ในการวิจัย

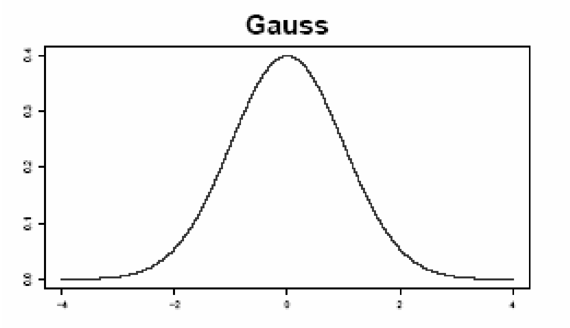
2.2.1.1 ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian Kernel Function) ฟังก์ชันนี้มีรูปแบบเหมือนกับฟังก์ชันการแจกแจงความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐาน โดย

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right) \tag{2-4}$$

โดยที่ $u = \frac{x - X_i}{h}$

$h$ คือ ค่าแบนวิดจ์

$X_i$ คือ ค่าของตัวอย่างสุ่ม




**ภาพที่ 2-2** กราฟแสดงฟังก์ชันเคอร์เนลแบบเกาส์เซียน

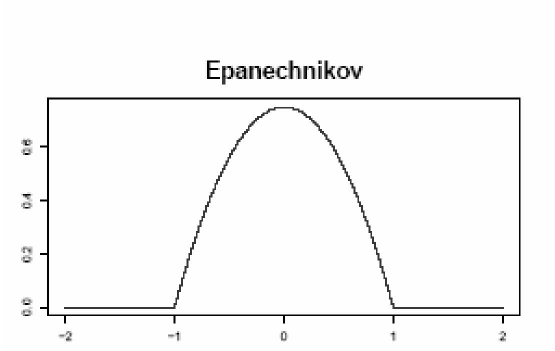
2.2.1.2 ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov Kernel Function) มีรูปแบบฟังก์ชัน คือ

$$K(u) = \frac{3}{4}\left(1-u^2\right)I\left(|u|\le 1\right) \tag{2-5}$$

โดยที่ $u = \dfrac{x - X_i}{h}$

$h$ คือ ค่าแบนวิดจ์

$X_i$ คือ ค่าของตัวอย่างสุ่ม




**ภาพที่ 2-3** กราฟแสดงฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

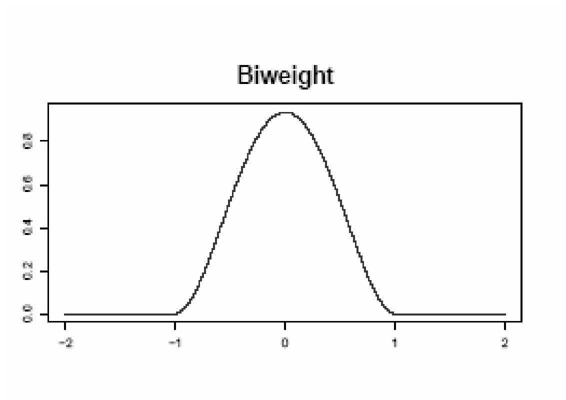
2.2.1.3 ฟังก์ชันเคอร์เนลแบบควอติค (Quartic Kernel Function) หรืออาจเรียกฟังก์ชันเคอร์เนลนี้ว่า ฟังก์ชันเคอร์เนลแบบไบเวท (Biweight) มีรูปแบบฟังก์ชัน คือ

$$K(u) \quad = \quad \frac{15}{16}\left(1-u^2\right)^2 I\left(|u| \le 1\right) \tag{2-6}$$

โดยที่ $u = \dfrac{x - X_i}{h}$

$h$ คือ ค่าแบนวิดจ์

$X_i$ คือ ค่าของตัวอย่างสุ่ม




**ภาพที่ 2-4** กราฟแสดงฟังก์ชันเคอร์เนลแบบควอติค

2.2.2 ประสิทธิภาพของฟังก์ชันเคอร์เนล

Silverman (1986) ได้กล่าวถึงฟังก์ชันเคอร์เนลที่เหมาะสม ว่าควรจะเป็นฟังก์ชันเคอร์เนลที่ทำให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีค่าต่ำ และได้ให้สูตรการคำนวณหาประสิทธิภาพสัมพัทธ์ (Relative Efficiency) ของฟังก์ชันเคอร์เนลดังนี้

$$eff(K) \; = \; \left(\frac{5}{4}\right) C(K) \left\{ \int f''(x)\,dx \right\}^{\frac{1}{5}} n^{-\frac{4}{5}}$$

โดยที่ $C(K) \; = \; k_2^{\frac{2}{5}} \left\{ \int K(t)^2 \, dx \right\}^{\frac{4}{5}}$

สำหรับประสิทธิภาพของฟังก์ชันเคอร์เนลที่ศึกษาแสดงในตารางที่ 2-2

**ตารางที่ 2-2** แสดงค่าประสิทธิภาพสัมพัทธ์ของฟังก์ชันเคอร์เนล

| **ฟังก์ชันเคอร์เนล** | **ค่าประสิทธิภาพสัมพัทธ์** |
|---|---|
| ฟังก์ชันอีพานิชนิคอฟ (Epanechnikov) | 1.0000 |
| ฟังก์ชันควอติค (Quartic) | 0.9939 |
| ฟังก์ชันสามเหลี่ยม (Triangular) | 0.9859 |
| ฟังก์ชันเกาส์เซียน (Gaussian) | 0.9512 |

**2.3 วิธีการเลือกแบนวิดจ์ (Bandwidth Selection)**

2.3.1 วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation

วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เสนอโดย Rudemo ในปี ค.ศ. 1982 เป็นวิธีการเลือกแบนวิดจ์ที่นิยมนำมาใช้และศึกษามากที่สุด

วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะพิจารณาจากผลต่างระหว่างฟังก์ชันความหนาแน่นน่าจะเป็นที่ประมาณได้กับฟังก์ชันความหนาแน่นที่แท้จริง แบนวิดจ์แบบ Least Square Cross-Validation ( $ISE(h)$ ) คือ

$$ISE(h) = \int_{-\infty}^{\infty} \left( \hat{f}_h(x) - f(x) \right)^2 dx$$

$$ISE(h) = \int_{-\infty}^{\infty} \hat{f}_h^2(x)dx - 2\int_{-\infty}^{\infty} \hat{f}_h(x) f(x)dx + f^2(x)dx \tag{2-7}$$

จากสมการที่ 2-4 จะเห็นว่า

$\int_{-\infty}^{\infty} \hat{f}_h^2(x)dx$ คำนวณได้จากข้อมูล

$\int_{-\infty}^{\infty} \hat{f}_h(x) f(x)dx$ ประมาณค่าได้จากข้อมูล

$\int_{-\infty}^{\infty} f^2(x)dx$ เป็นส่วนที่เป็นอิสระกับค่าแบนวิดจ์

ดังนั้น

$$ISE(h) - \int_{-\infty}^{\infty} f^2(x)dx = \int_{-\infty}^{\infty} \hat{f}_h^2(x)dx - 2\int_{-\infty}^{\infty} \hat{f}_h(x) f(x)dx \tag{2-8}$$

แนวคิดพื้นฐานของการเลือกแบนวิดจ์แบบ Least Square Cross-Validation คือ ประมาณค่า $ISE(h)-\int_{-\infty}^{\infty} f^2(x)dx$ จากข้อมูลและทำให้มีค่าประมาณที่ได้นั้นมีค่าน้อยที่สุด จากสมการที่ 2-5 เนื่องจาก $\int_{-\infty}^{\infty} \hat{f}_h(x)\, f(x)\, dx \;=\; E_x\left[\hat{f}_h(x)\right] \;=\; n^{-1}\sum_{i=1}^{n} f_{h,i}(x_i)$ ดังนั้นสมการที่ 2-5 สามารถเขียนรูปแบบสมการของการเลือกแบนวิดจ์แบบ Least Square Cross-Validation ($LSCV(h)$) ดังนี้

$$LSCV(h) \;=\; \int_{-\infty}^{\infty} \hat{f}_h^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} f_{h,i}(x_i) \tag{2-9}$$

$LSCV(h)$ จึงเป็นตัวประมาณค่าที่ไม่เอนเอียงของ $\int_{-\infty}^{\infty} \hat{f}_h^2(x)\,dx - 2\int_{-\infty}^{\infty} \hat{f}_h(x)\, f(x)\,dx$ และเป็นตัวประมาณที่ทำให้ $LSCV(h) \;=\; ISE(h)-\int_{-\infty}^{\infty} f^2(x)dx$ มีค่าน้อยที่สุด

2.3.2 วิธีการเลือกแบนวิดจ์แบบ Plug-in

วิธีการเลือกแบนวิดจ์แบบ Plug-in เสนอขึ้นเป็นครั้งแรกโดย Woodroofe ในปี ค.ศ. 1970 โดยเสนอสมการสำหรับประมาณค่าแบนวิดจ์ที่เหมาะสม (Optimal Bandwidth; $h_{opt}$ ) ดังนี้

$$h_{opt} \;=\; \left(\frac{R(K)}{\mu_2^2(K)R\left(f^{(2)}\right)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{2-10}$$

โดยที่ $R(K) \;=\; \int_{-\infty}^{\infty} K(x)^2 dx$ และ $\mu_2(K) \;=\; \int_{-\infty}^{\infty} x^2 K(x)\,dx$ ซึ่งจากสมการ 2-7 จะเห็นว่ามีเพียง $R\left(f^{(2)}\right)$ เท่านั้นที่ไม่ทราบถึงคุณสมบัติ

ต่อมาในปี ค.ศ. 1991 Sheather และ Jones นำเสนอการประมาณค่า $R\left(f^{(p)}\right)$ โดยใช้ $R\left(\hat{f}_h^{(p)}\right)$ ของการเลือกแบนวิดจ์แบบ Plug-in ซึ่งวิธีการประมาณนี้มีข้อดี คือ ให้ค่าประมาณที่เป็นบวกเสมอ

การเลือกแบนวิดจ์โดยวิธีการ Plug-in จะพิจารณาจาก MISE แต่เนื่องจากสมการสำหรับประมาณค่าแบนวิดจ์ที่เหมาะสม (สมการที่ 2-7) มีลักษณะแบบ asymptotic กล่าวคือ เมื่อ $n \to \infty$ แล้ว $h = h(n) \to 0$ ดังนั้นสมการ Asymptotic Mean Integrated Squared Error (AMISE) คือ

$$AMISE(h) \;=\; (nh)^{-1} R(K) + \frac{1}{4} h^4 \sigma_K^4 R(f'') \tag{2-11}$$

โดย $R(K)=\int_{-\infty}^{\infty}K^2(x)dx$ และ $\sigma_K^2=\int_{-\infty}^{\infty}x^2K(x)dx$

ดังนั้นจึงเขียนสมการ 2-11 ในรูป

$$\psi(h) \quad = \quad (nh)^{-1}R(K)+\frac{1}{4}h^4\sigma_K^4\hat{S}(\alpha)$$

โดยที่ $\hat{S}(\alpha)$ เป็นค่าประมาณแบบเคอร์เนลของ $R(f'')$ เมื่อใช้ค่าแบนวิดจ์ที่เหมาะสม ดังนั้นค่าแบนวิดจ์ที่เหมาะสมของวิธีการ Plug-in เมื่อใช้ค่าประมาณของ $R\left(f^{(p)}\right)$ ที่เสนอโดย Sheather และ Jones คือ

$$h \quad = \quad \left[\frac{R(K)}{\sigma_K^4\hat{S}_D(\hat{\alpha}(h))}\right]^{\frac{1}{5}}n^{-\frac{1}{5}} \tag{2-12}$$

โดยที่ $\hat{S}(\alpha) \quad = \quad \{n(n-1)\}^{-1}\alpha^{-5}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi^{iv}\left\{\alpha^{-1}\left(X_i-X_j\right)\right\}$ และ

$$\hat{\alpha}(h) \quad = \quad 1.357\left\{\frac{\hat{S}_D(a)}{\hat{T}_D(b)}\right\}^{\frac{1}{7}}h^{\frac{5}{7}}$$

$$\hat{T}_D(b) \quad = \quad -\{n(n-1)\}^{-1}b^{-7}\sum_{i=1}^{n}\sum_{i=1}^{n}\phi^{vi}\left\{b^{-1}\left(X_i-X_j\right)\right\}$$

2.3.3 วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap

วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap เสนอเป็นครั้งแรกในปี ค.ศ. 1989 โดย Taylor โดยพิจารณาถึงค่าแบนวิดจ์ที่ทำให้ค่า MISE มีค่าน้อยที่สุด

วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap แตกต่างจากวิธีการ Bootstrap ที่ใช้ทั่วๆ ไป คือวิธีการ Smoothed Bootstrap สามารถคำนวณค่า MISE ได้โดยตรงในขณะที่วิธีการ Bootstrap ทั่วไปจะใช้วิธีการจำลอง (Simulation) จึงทำให้วิธีการแบบ Smoothed Bootstrap มีการคำนวณที่รวดเร็วกว่าวิธีอื่นๆ (Jones, marron and Sheather 1996)

ค่าแบนวิดจ์แบบ Smoothed Bootstrap เป็นค่าแบนวิดจ์ที่ให้ค่า MISE น้อยที่สุด ในขั้นตอนการคำนวณจะต้องทำการสุ่มตัวอย่างซ้ำ (Resample) จากการแจกแจงที่คาดว่าจะเป็น จากนั้นจึงทำการสร้างตัวประมาณแบบ Smoothed Bootstrap ดังขั้นตอนต่อไปนี้

กำหนดให้ $X_1^*,...X_n^*$ เป็นตัวอย่างสุ่มจากการสุ่มตัวอย่างซ้ำ จะได้ตัวประมาณค่าความหนาแน่นแบบ Bootstrap โดย

$$\hat{f}_{nj}^*(x,h) \quad = \quad \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i^*}{h}\right) \tag{2-13}$$

เมื่อ $j=1,...B$ โดยที่ $B$ คือ จำนวนครั้งของการ Bootstrap

ความแปรปรวนของ $f_n(x,h)$ คือ

$$B^{-1}\sum_{j=1}^{B}\int_{-\infty}^{\infty}\left(f_{nj}^*(x,h)-\overline{f}_{nj}^*(x,h)\right)^2 dx$$

โดยที่ $\overline{f}_{nj}^*(x,h) \;=\; B^{-1}\sum_{j=1}^{B} f_{nj}^*(x,h)$

จากนั้นสร้างการประมาณความหนาแน่น $\hat{f}_n(x,h_0)$ และทำการสุ่มตัวอย่างซ้ำจาก $\hat{f}_n(x,h_0)$ จากขั้นตอนนี้แต่ละค่าของตัวอย่างสุ่มจะเพิ่มค่าความคลาดเคลื่อน ($\varepsilon$) เข้าไปในแต่ละค่า โดยที่ $\varepsilon$ มีฟังก์ชันความหนาแน่นน่าจะเป็น $K$ จากนั้นจึงทำการประมาณค่าความเอนเอียง โดย $\hat{f}_n(x,h_0)-\overline{f}_{nj}^*(x,h)$ และทำการประมาณค่า $MISE(h)$ โดยใช้

$$MISE(h) \quad = \quad B^{-1}\sum_{j=1}^{B}\int_{-\infty}^{\infty}\left(f_{nj}^*(x,h)-\hat{f}_n(x,h_0)\right)^2 dx \tag{2-14}$$

ดังนั้นจึงเลือกค่าแบนวิดจ์ที่ทำให้ค่า $MISE(h)$ มีค่าน้อยที่สุด

2.3.4 วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast

Ahmad และ Ran ได้เสนอวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในปี ค.ศ. 2003 โดยวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เป็นวิธีที่ขึ้นอยู่กับฟังก์ชันเคอร์เนล ($k_i$) $q$ ฟังก์ชันที่แตกต่างกันและค่าสัมประสิทธิ์ความแตกต่าง (Coefficient of Contrast) จะมีผลรวมเท่ากับศูนย์ จากตัวประมาณค่าสามารถสร้างสมการ Contrast สำหรับค่าแบนวิดจ์ที่ประมาณได้ ดังนี้

$$con\left(\hat{f}_h(x)\right) \quad = \quad \sum_{i=1}^{q} p_i\hat{f}(x) \tag{2-15}$$

โดยที่ $p_i$ คือ ค่าสัมประสิทธิ์ความแตกต่าง โดยที่ $i=1,...,q$

การเลือกแบนวิดจ์แบบ Kernel Contrast สามารถทำตามขั้นตอนได้ดังนี้

1. สร้างฟังก์ชัน Contrast ของตัวประมาณความหนาแน่นแบบเคอร์เนล โดย

$$\hat{f}_h(x) \quad = \quad \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$$

2. เลือกขอบเขตวิกฤตเพื่อวัดถึงความแตกต่างระหว่าง $\hat{f}_h(x)$ กับ $f(x)$ โดยที่

$$con\left(f_h(x)\right) \quad = \quad f(x)\sum_{i=1}^{q} p_i \quad = \quad 0$$

3. หาค่า $\hat{h}$ ที่ทำให้ขอบเขตวิกฤตน้อยที่สุด

4. ประมาณค่า $f(x)$ ด้วย $\hat{f}_h(x) = \sum_{i=1}^{q} c_i \hat{f}_{h,i}(x)$ โดยที่ $\sum_{i=1}^{q} c_i = 1$ และ $0 \le c_i \le 1$ สำหรับ $i = 1,...,q$

จากนั้นเลือกค่าแบนวิดจ์ที่ทำให้ค่า $MISE(h)$ มีค่าน้อยที่สุด คือ

$$MISE(h) \quad = \quad E\left[\int_{-\infty}^{\infty}\left(\sum_{i=1}^{q} p_i \hat{f}_{h,i}(x) - \sum_{i=1}^{q} p_i f(x)\right)^2\right]dx$$

แต่เนื่องจาก $\sum_{i=1}^{q} p_i f(x) = f(x)\sum_{i=1}^{q} p_i = 0$ ดังนั้น

$$MISE(h) \quad = \quad E\left[\int_{-\infty}^{\infty}\left(\sum_{i=1}^{q} p_i \hat{f}_{h,i}(x)\right)^2\right]dx \qquad (2\text{-}16)$$

ดังนั้น $ISE(h) = \int_{-\infty}^{\infty}\left(\sum_{i=1}^{q} p_i \hat{f}_{h,i}(x)\right)^2 dx$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของ $MISE(h)$ ดังนั้นจึงเลือกค่าแบนวิดจ์ที่ทำให้ค่า MISE มีค่าน้อยที่สุด

**2.4 เกณฑ์การพิจารณา**

การพิจารณาถึงความเหมาะสมของตัวประมาณค่าโดยทั่วไปจะพิจารณาจากความแตกต่างระหว่างฟังก์ชันความหนาแน่นที่แท้จริง $f(x)$ และฟังก์ชันความหนาแน่นที่ประมาณขึ้น $f_h(x)$ ดังนี้

2.4.1 ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Squared Error: MSE) เป็นวิธีการวัดความแตกต่างของตัวประมาณความหนาแน่น $\hat{f}_h(x)$ จากความหนาแน่นที่แท้จริง ซึ่งมีหลายวิธีที่นำมาศึกษา แต่เมื่อพิจารณาถึงการประมาณแบบจุด มักจะนิยมใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง โดย

$$MSE\left(\hat{f}(x)\right) = E\left[\hat{f}_h(x) - f(x)\right]^2$$

$$MSE\left(\hat{f}(x)\right) = \left[E\hat{f}(x) - f(x)\right]^2 + \operatorname{var}\hat{f}(x) \qquad (2\text{-}17)$$

จากสมการที่ 2-17 สามารถเขียนรูปแบบอย่างง่าย คือ

$$MSE\left(\hat{f}(x)\right) = \left[bias\,\hat{f}(x)\right]^2 + \operatorname{var}\hat{f}(x) \qquad (2\text{-}18)$$

เนื่องจาก

$$bias\,\hat{f}(x) = \frac{1}{2}h^2 f''(x)k_2 + o\left(h^2\right)$$

$$\operatorname{var}\left(\hat{f}(x)\right) = n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt + o(nh)^{-1}$$

ดังนั้น

$$MSE\left(\hat{f}(x)\right) = \left[\frac{1}{2}h^2 f''(x)k_2 + o\left(h^2\right)\right]^2 + n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt + o(nh)^{-1} \qquad (2\text{-}19)$$

2.4.2 ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (Mean Integrated Squared Error: MISE)

$$MISE\left(\hat{f}(x)\right) = E\int\left[\hat{f}_h(x) - f(x)\right]^2 dx$$

$$= \int E\left\{\hat{f}_h(x) - E\left[\hat{f}_h(x)\right] + E\left[\hat{f}_h(x)\right] - f(x)\right\}^2 dx$$

$$MISE\left(\hat{f}(x)\right) = \int E\left\{\hat{f}_h(x) - E\left[\hat{f}_h(x)\right]\right\}^2 + \left\{E\left[\hat{f}_h(x)\right] - f(x)\right\}^2 dx \qquad (2\text{-}20)$$

จากสมการที่ 2-17 สามารถเขียนรูปแบบอย่างง่าย คือ

$$MISE\left(\hat{f}(x)\right) \quad = \int \mathrm{var}\, \hat{f}_h(x)\,dx + \int \mathrm{bias}^2\, \hat{f}_h(x)\,dx \tag{2-21}$$

เนื่องจาก

$$bias\ \hat{f}(x) = \frac{1}{2}h^2 f''(x)k_2 + o\left(h^2\right)$$

$$\mathrm{var}\left(\hat{f}(x)\right) = n^{-1}h^{-1}f(x)\int_{-\infty}^{\infty} K(t)^2\,dt + o\left(nh\right)^{-1}$$

ดังนั้น

$$MISE\left(\hat{f}(x)\right) = n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt\, f(x)\,dx + \frac{h^4}{4}\,k_2^2\int_{-\infty}^{\infty} f''(x)^2\,dx + o\left(nh\right)^{-1} + o\left(h^4\right) \tag{2-22}$$

แสดงวิธีการคำนวณอย่างละเอียดในภาคผนวก ข

## 2.5 การแจกแจงต่างๆ ที่ใช้ในงานวิจัย

2.5.1 การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงแบบปกติเป็นการแจกแจงแบบต่อเนื่องที่พบมากที่สุดและมีความสำคัญมากทั้งทางด้านสถิติประยุกต์และสามารถนำไปประโยชน์ในการประมาณค่าของประชากรและทดสอบสมมติฐานทางสถิติ

อับราฮัม เดอร์มัวร์ (Abraham De Moire, 1667-1754) นักคณิตศาสตร์ชาวฝรั่งเศส เป็นผู้ค้นพบการแจกแจงแบบปกติ เมื่อปี ค.ศ. 1733 ต่อมา ปิแอร์ ลาปลาส (Pierre Laplace, 1749-1827)

นักคณิตศาสตร์ชาวฝรั่งเศส กับคาร์ล เกาส์ (Carl Gauss, 1777-1855) นักคณิตศาสตร์และดาราศาสตร์ชาวเยอรมัน ได้ค้นพบการแจกแจงแบบปกติโดยไม่ทราบผลงานของอับราฮัม เดอร์มัวร์ มาก่อนเลย ซึ่งพบว่าการแจกแจงของความคลาดเคลื่อนในการวัดทางวิทยาศาสตร์กายภาพสามารถประมาณได้อย่างใกล้เคียงโดยใช้โค้งปกติ ซึ่งเขาเรียกว่า “โค้งปกติของความคลาดเคลื่อน” (The Normal curve of Error) และถือได้ว่าเป็นกฎของความน่าจะเป็น (The Laws of chance) ผลงานของลาปลาส และ เกาส์ เป็นที่รู้จักกันแพร่หลายและนำไปใช้ประโยชน์อย่างกว้างขวาง (วินัย, 2537) ในบางครั้งมักจะเรียกการแจกแจงแบบปกติว่า “การแจกแจงลาปลาส” (Laplacian Distribution) หรือ “การแจกแจงเกาส์เซียน” (Gaussian Distribution) การแจกแจงแบบปกติมีฟังก์ชันความหนาแน่นน่าจะเป็น (pdf) คือ

ให้ $X$ เป็นตัวแปรสุ่ม จะกล่าวว่า $X$ มีการแจกแจงปกติ ถ้า $X$ มีฟังก์ชัน กำหนดโดย

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty$$

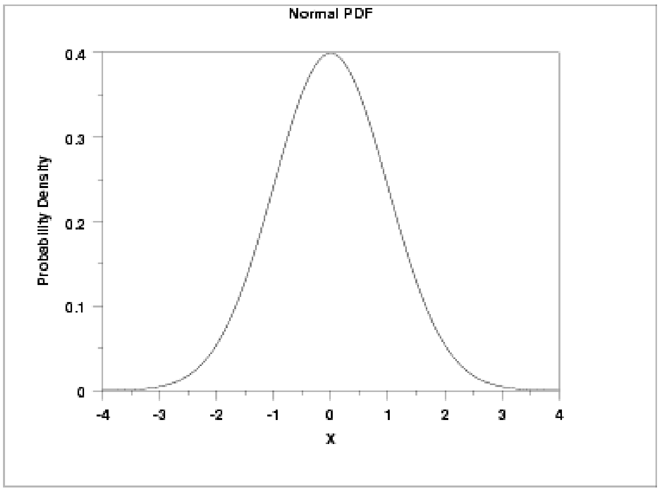
เมื่อ

$\sigma^2$ คือ ความแปรปรวนของประชากร

$\sigma$ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร

$\mu$ คือ ค่าเฉลี่ยของประชากร

$\pi \simeq$ 3.14159




**ภาพที่ 2-5** แสดงโค้งการแจกแจงแบบปกติ

คุณสมบัติของการแจกแจงแบบปกติ

1. ลักษณะโค้งเป็นรูประฆังคว่ำ (bell shaped) สมมาตรกับแกนตั้งที่ลากผ่านเส้นค่าเฉลี่ย $\mu$ ฟังก์ชันความหนาแน่นน่าจะเป็น มีลักษณะเป็นรูประฆังคว่ำที่สมมาตร ดังนั้นค่าเฉลี่ยของตัวแปรสุ่ม $X$ คือ

$$\mu_x = E(X) = \mu$$

ความแปรปรวนของตัวแปรสุ่ม $X$ คือ

$$\sigma_x^2 = E[(x-\mu)^2] = \sigma^2$$

2. กราฟลดลงอย่างต่อเนื่องทั้งสองข้าง มีจุดเปลี่ยนเว้าที่ $x = \mu \pm \sigma'$ ปลายโค้งจะเข้าใกล้แกน $X$ เมื่อ $x$ มีค่าห่างจาก $\mu$ ออกไป แต่จะไม่ตัดแกน $X$

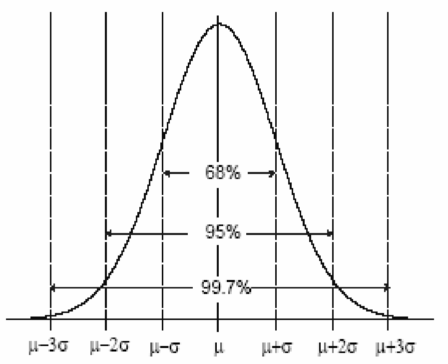
3. เป็นโค้งที่มีจุดสูงสุดเพียงจุดเดียว (Unimodal) อยู่ที่ $x = \mu$ ซึ่งเป็นค่าฐานนิยม

4. ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมมีค่าเท่ากัน อยู่ที่จุด $x = \mu$

5. มีค่าความโด่ง (Kurtosis) เท่ากับ 3 และค่าความเบ้ (Skewness) เท่ากับ 0

6. $\mu$ และ $\sigma^2$ เป็นค่าพารามิเตอร์ของการแจกแจงแบบปกติ โดย $\mu$ และ $\sigma^2$ จะเป็นตัวกำหนดตำแหน่งที่ตั้งของเส้นโค้งและลักษณะของเส้นโค้งว่าจะแบนหรือโด่ง

7. พื้นที่ใต้โค้งปกติที่อยู่ระหว่าง $\mu \pm 1\sigma$ , $\mu \pm 2\sigma$ และ $\mu \pm 3\sigma$ จะมีค่าเป็น 68.27%, 95.45% และ 99.73% ตามลำดับ




**ภาพที่ 2-6** กราฟแสดงพื้นที่ใต้โค้งปกติ

ในกรณีที่การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 จะเรียกการแจกแจงแบบปกตินี้ว่า การแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution) สามารถเขียนได้ว่า $Z \sim N(0,1)$ เมื่อ $Z$ คือตัวแปรที่มีการแจกแจงแบบปกติมาตรฐาน

2.5.2 การแจกแจงเอกซ์โพเนนเชียล (Exponential Distribution)

การแจกแจงเอกซ์โพเนนเชียล เป็นการแจกแจงของตัวแปรสุ่มที่แสดงระยะเวลาระหว่างเหตุการณ์ที่สนใจศึกษา (วินัย, 2534) เช่น ระยะเวลาที่รอคอยในการใช้บริการฝากเงินในธนาคาร หรือแสดงถึงอายุในการใช้งาน (Life Time) ของวัตถุสิ่งของ เช่น อายุการใช้งานของเครื่องซักผ้า ตัวแปรสุ่มต่อเนื่อง $X$ ที่มีการแจกแจงเอกซ์โพเนนเชียลพารามิเตอร์ $\theta > 0$ จะมีฟังก์ชันความหนาแน่นในรูป

$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \quad \text{เมื่อ} \quad x > 0$$
$$= 0 \quad \text{เมื่อ} \quad x \text{ มีค่าอื่นๆ}$$

คุณสมบัติของการแจกแจงเอกซ์โพเนนเชียล

1. ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มที่มีการแจกแจงแบบเอกซ์โพเนนเชียลจะขึ้นอยู่กับพารามิเตอร์ $\theta$ โดยมีค่าเฉลี่ยเท่ากับ $\theta$ และความแปรปรวนเท่ากับ $\theta^2$

2. ตัวแปรสุ่มแบบเอกซ์โพเนนเชียล มีคุณสมบัติไร้ความจำ (Memory less) คือ

$$P(X > t + x | X > t) = P(X > t)$$

กล่าวคือ

$$P(X > t + x | X > t) = \frac{P(X > t + x)}{P(X > t)}$$

$$= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}}$$

$$= e^{-\lambda x} \quad = P(X > x)$$

จากคุณสมบัติดังกล่าวสามารถยกตัวอย่างได้ดังนี้

สมมติว่าอายุการใช้งานของหลอดไฟฟ้ายี่ห้อหนึ่ง มีการแจกแจงเอกซ์โพเนนเชียล ที่มีอายุการใช้งานเฉลี่ยเท่ากับ 1,000 ชั่วโมง ให้ $X$ แทนอายุการใช้งานของหลอดไฟฟ้า

$$P(X \geq x) = \int_x^\infty 1e^{-\frac{t}{1000}}dt$$
$$= e^{-\frac{x}{1000}}$$

ถ้าใช้หลอดไฟฟ้าหลอดหนึ่งมาแล้ว 500 ชั่วโมง ความน่าจะเป็นที่หลอดไฟฟ้าหลอดนี้จะมีอายุการใช้งานต่อไปอีกอย่างน้อย 800 ชั่วโมง นั่นคือ

$$P(X > 1300 | X > 500) = \frac{P(X > 1300)}{P(X > 500)}$$

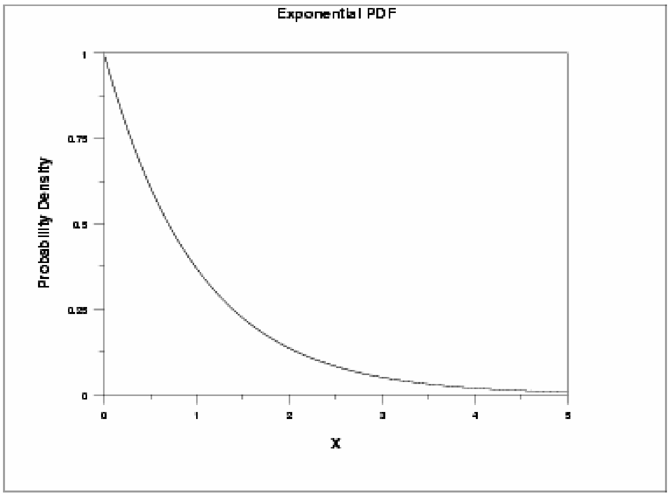
$$= \frac{e^{-\frac{1300}{1000}}}{e^{-\frac{500}{1000}}}$$

$$= e^{-\frac{8}{10}}$$

ซึ่งเท่ากับ $P(X > 800) \;=\; e^{-\frac{8}{10}}$

ดังนั้น $P(X > 1300 | X > 500) \;=\; P(X > 800)$

แสดงให้เห็นว่าความน่าจะเป็นที่หลอดไฟฟ้านั้นจะมีอายุการใช้งานอย่างน้อย 1,300 ชั่วโมง เมื่อถูกใช้มาแล้วอย่างน้อย 500 ชั่วโมง จะมีค่าเท่ากับความน่าจะเป็นที่หลอดไฟฟ้านี้จะมีอายุการใช้งานอย่างน้อย 800 ชั่วโมง นับตั้งแต่เริ่มต้นใช้งาน




**ภาพที่ 2-7** กราฟแสดงการแจกแจงแบบเอกซ์โพเนนเชียล

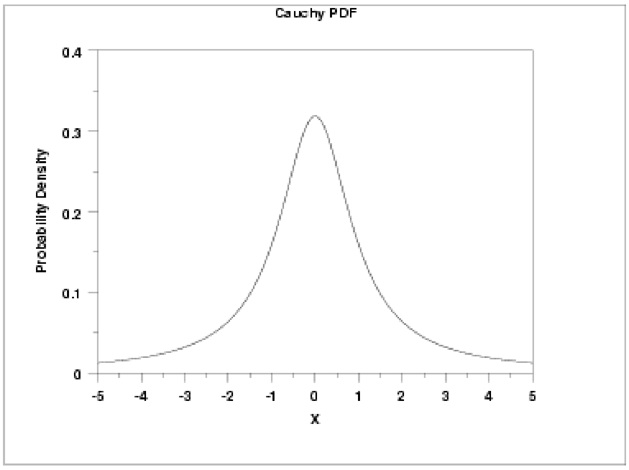
2.5.3 การแจกแจงแบบโคชี (Cauchy Distribution)

โอกุส แดง ลุยส์ โคชี (Augustin Louis Cauchy) นักคณิตศาสตร์ชาวฝรั่งเศสเป็นผู้ค้นพบการแจกแจงแบบโคชี (Cauchy Distribution) จะพบการแจกแจงแบบโคชีในงานทางด้านวิทยาศาสตร์และวิศวกรรมศาสตร์ เช่น ในกลศาสตร์และทฤษฎีไฟฟ้า (มานพ, 2545)

ตัวแปรสุ่มต่อเนื่อง $X$ มีการแจกแจงแบบโคชีด้วยพารามิเตอร์ $\theta$ โดยที่ $-\infty < \theta < \infty$ และ $\beta > 0$ ฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงแบบโคชี คือ

$$f(x) = \frac{1}{\pi\beta\left[1+\left(\frac{x-\theta}{\beta}\right)^2\right]}, \qquad -\infty < \theta < \infty$$

เขียนแทนด้วย $X \sim C(\theta,\beta)$ หาก $\theta = 0$ และ $\beta = 1$ จะเขียนแทนด้วย $X \sim C(0,1)$ และจะเรียกว่า $X$ มีการแจกแจงแบบโคชีมาตรฐาน (มานพ, 2547)




**ภาพที่ 2-8** รูปแสดงฟังก์ชันความหนาแน่นแบบโคชีมาตรฐาน

**2.6 งานวิจัยที่เกี่ยวข้อง**

จากผลงานวิจัยที่ผ่านมา มีนักสถิติหลายท่านได้ทำการศึกษาถึงการเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลสำหรับฟังก์ชันของตัวแปรสุ่ม ซึ่งมีผลสรุปได้ดังนี้

Taylor (1989) ทำการศึกษาวิธีการเลือกแบนวิดจ์ และค่า ISE โดยทำการเปรียบเทียบจากการจำลองข้อมูลโดยแบ่งกรณีศึกษาเป็น 2 กรณี คือ

กรณีที่1 เปรียบเทียบค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ Cross-Validation วิธีการเลือกแบนวิดจ์แบบ Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Asymptotic Normal Formula กำหนดขนาดตัวอย่าง 3 ขนาด คือ 100, 200 และ 300 โดยจำลองข้อมูลจากการแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution) การแจกแจงแบบปกติผสม (Mixture Normal Distribution) การแจกแจงแบบโคชีมาตรฐาน (Standard Cauchy

Distribution) การแจกแจงแบบไคสแควร์ (Chi-Square Distribution) ที่องศาอิสระเท่ากับ 6 โดยเปรียบเทียบค่าแบนวิดจ์จากอัตราส่วนของ $\frac{\text{ISE }\left(h_i\right)}{\text{ISE }\left(h_1\right)}$

กรณีที่2 ศึกษาค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Bootstrap โดยใช้ค่า ISE เป็นหลัก โดยเปรียบเทียบค่าต่างๆ คือ 1) $B^*\left(h\right)-ISE\left(h\right)$ 2) $\frac{B^*\left(h\right)-ISE\left(h\right)}{ISE\left(h\right)}$ และ 3) $\frac{\left[B^*\left(h\right)-ISE\left(h\right)\right]^2}{ISE\left(h\right)}$ จากการเปรียบเทียบนี้จะเห็นว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นอัตราการลู่เข้าของ $\text{var}\left\{B^*\left(h\right)\right\}$ จะเร็วกว่า $ISE\left(h\right)$

ผลการวิจัยพบว่า ค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Bootstrap จะให้ผลลัพธ์ที่ดีกว่าค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Cross-Validation ยกเว้นในกรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ส่วนในกรณีข้อมูลมีการแจกแจงแบบไคสแควร์ ที่องศาอิสระเท่ากับ 6 เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Bootstrap จะให้ค่าผลลัพธ์ที่ดีกว่าค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Asymptotic Normal Formula อีกทั้งสรุปได้ว่าถึงแม้ว่าค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Bootstrap จะมีขนาดใหญ่ แต่ก็ยังให้ผลลัพธ์ที่ดีกว่าค่าแบนวิดจ์ขนาดเล็กจากวิธีการเลือกแบนวิดจ์แบบ Cross-Validation

Faraway และ Jhun (1990) ทำการศึกษาถึงวิธีการเลือกแบนวิดจ์แบบ Bootstrap โดยจำลองข้อมูลจากการแจกแจง 6 การแจกแจง คือ การแจกแจงปกติมาตรฐาน (Standard Normal Distribution), การแจกแจงแบบ bimodal, การแจกแจงแบบ Contaminated Normal, การแจกแจงแบบลอกนอร์มอล (Lognormal Distribution), การแจกแจงแบบโคชี (Cauchy Distribution) และการแจกแจงแบบเบต้า (Beta Distribution) โดยกำหนดขนาดตัวอย่างเท่ากับ 50 และ 400 ทำการ Bootstrap จำนวน 100 ครั้ง โดยแสดงผลลัพธ์เป็นค่าแบนวิดจ์เป็นค่าแบนวิดจ์ที่ทำให้ค่า MISE น้อยที่สุด ค่าแบนวิดจ์ที่ทำให้ ISE น้อยที่สุด ค่าอัตราส่วนของ MISE ที่ประมาณได้กับค่า MISE ของ ISE และค่าความคลาดเคลื่อนประมาณ 2% ในแต่ละ Bootstrap

สามารถสรุปได้ว่า ค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Bootstrap จะให้ผลลัพธ์ที่ดีกว่าค่าแบนวิดจ์ที่ได้จากการเลือกแบนวิดจ์แบบ Cross-Validation โดยค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Bootstrap จะมีค่าแบนวิดจ์ขนาดใหญ่กว่าค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Cross-Validation แต่ให้ความแปรปรวนต่ำกว่า กล่าวคือวิธีการเลือกแบนวิดจ์แบบ Bootstrap จะให้ผลลัพธ์ที่ดีกว่า

Hall, Sheather, Marron และ Jones (1991) ทำการศึกษาและเปรียบเทียบค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ Additive Procedure วิธีการเลือกแบนวิดจ์

แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Gaussian Reference Distribution เมื่อขนาดตัวอย่างมีขนาดตัวอย่างเท่ากับ 50 และ 100 และใช้การแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution) การแจกแจงแบบปกติผสม (Mixture Normal Distribution) ด้วยค่าเฉลี่ยที่แตกต่างกัน และการแจกแจงแบบปกติผสม (Mixture Normal Distribution) ที่มีค่าความแปรปรวนต่างกัน

Altman และ Leger (1995) ศึกษาและเปรียบเทียบค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 3 วิธี เมื่อขนาดตัวอย่างมีขนาดเล็ก คือ 1) วิธีการเลือกแบนวิดจ์แบบ Cross-Validation 2) วิธีการเลือกแบนวิดจ์แบบ Leave One Out 3) วิธีการเลือกแบนวิดจ์แบบ Plug-in ศึกษาเฉพาะฟังก์ชันเคอร์เนลแบบควอดราติค ทำการจำลองข้อมูลจากการแจกแจง 5 การแจกแจง คือ 1) การแจกแจงแบบปกติมาตรฐาน 2) การแจกแจงแบบเอกซ์โพเนนเชียล 3)การแจกแจงแบบ Skewed Unimodal 4) การแจกแจงแบบ Asymmetric Bimodal และ 5) การแจกแจงแบบคลอร์ (Claw Distribution) ใช้ขนาดตัวอย่างเท่ากับ 10, 100 และ 1,000 ได้ผลการศึกษาดังนี้

1. เมื่อขนาดตัวอย่างเท่ากับ 10 ค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Cross-Validation และวิธีการเลือกแบนวิดจ์แบบ Leave One Out จะมีค่าน้อยกว่าค่าน้อยกว่าค่าแบนวิดจ์ที่ทำให้ค่า ASE มีค่าน้อยที่สุด แต่วิธีการเลือกแบนวิดจ์ทั้ง 2 วิธีการนี้ยังให้ผลลัพธ์ที่ดี

2. เมื่อขนาดตัวอย่างเท่ากับ 100 และ 1,000 วิธีการเลือกแบนวิดจ์แบบ Cross-Validation และวิธีการเลือกแบนวิดจ์แบบ Leave One Out ยังคงเลือกค่าแบนวิดจ์ขนาดเล็ก ซึ่งเป็นผลลัพธ์ที่ไม่ดี แต่วิธีการเลือกแบนวิดจ์แบบ Plug-in ยังเป็นวิธีการที่ให้ผลลัพธ์ที่น่าพอใจ

Altman และ Leger สรุปว่า วิธีการเลือกแบนวิดจ์แบบ Cross-Validation และวิธีการเลือกแบนวิดจ์แบบ Leave One Out จะได้ผลลัพธ์ที่น่าพอใจเมื่อมีขอบเขตของขนาดตัวอย่าง ในขณะที่วิธีการเลือกแบนวิดจ์แบบ Plug-in จะให้ผลลัพธ์ที่น่าพอใจทุกขนาดของตัวอย่าง

Jones, Marron และ Sheather (1996) ทำการศึกษาและเปรียบเทียบวิธีการเลือกแบนวิดจ์ โดยแบ่งวิธีการเลือกแบนวิดจ์เป็น 2 ช่วงเวลา คือ 1) ช่วงเวลาที่ 1วิธีการเลือกแบนวิดจ์ในช่วงเวลานี้ เช่น วิธีการเลือกแบนวิดจ์ Rule of Thumb วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation และวิธีการเลือกแบนวิดจ์แบบ Biased Cross-Validation เป็นต้น 2) ช่วงเวลาที่ 2 วิธีการเลือกแบนวิดจ์ในช่วงเวลานี้ เช่น วิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap เป็นต้น โดยกำหนดสมมติฐานในการวิจัย 3 ข้อ คือ

1. วิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 จะให้ผลลัพธ์ที่ดีกว่าวิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 1

2. วิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 สามารถนำมาใช้กับโปรแกรมคอมพิวเตอร์ได้ง่ายกว่าวิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 1

3. วิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in เป็นวิธีการเลือกแบนวิดจ์ที่ดีที่สุดในวิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 จากวิธีการเลือกแบนวิดจ์ทั้งหมด

Jones, Marron และ Sheather ทำการเปรียบเทียบโดยใช้ข้อมูลจริงจาก The Australian Institute of Sport Data สามารถสรุปได้ว่า

1. วิธีการเลือกแบนวิดจ์แบบ Rule of Thumb จะให้ผลลัพธ์ที่ขาดลักษณะที่สำคัญของข้อมูลไปบางส่วน

2. วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะให้ผลลัพธ์ในทิศทางที่ด้อยกว่าและเชื่อถือไม่ค่อยได้

3. วิธีการเลือกแบนวิดจ์แบบ Biased Cross-Validation จะให้ผลลัพธ์ไม่แน่นอน

สรุปได้ว่าวิธีการเลือกแบนวิดจ์จากช่วงเวลาที่ 1 นั้น ไม่เหมาะสมที่จะนำไปใช้อย่างแพร่หลาย แต่วิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 เช่น วิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in ให้ผลลัพธ์ที่สม่ำเสมอและคงที่ แสดงให้เห็นว่าเหมาะสมที่จะนำไปใช้ในโปรแกรมทางคอมพิวเตอร์ นอกจากนี้ Jones, Marron และ Sheather ได้ศึกษาโดยการจำลองข้อมูลจากการแจกแจงแบบปกติผสม (Mixture Normal Distribution) 15 ฟังก์ชัน โดยมีขนาดตัวอย่างเท่ากับ 100 และ 1,000 ได้ผลการศึกษา ดังนี้

1. การแจกแจงของค่าแบนวิดจ์จากการแบนวิดจ์แบบ Rule of Thumb มีค่าเฉลี่ยค่อนข้างมาก แต่ค่าความแปรปรวนจะน้อยกว่าวิธีเลือกแบนวิดจ์อื่นๆ เนื่องจากมาจากการประมาณแบบสเกล จึงทำให้ลักษณะไม่เป็นไปโดยสุ่ม

2. การแจกแจงของค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าเฉลี่ยใกล้เคียงกับค่าแบนวิดจ์ของ MISE แต่มีค่าความแปรปรวนมากกว่าวิธีอื่น โดยลักษณะจะเป็นไปในทางต่ำกว่าค่าประมาณ

3. การแจกแจงของค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ Biased Cross-Validation หารูปแบบได้ยาก เนื่องจากโค้งที่ได้มีลักษณะที่ไม่แน่นอน เมื่อขนาดตัวอย่างเท่ากับ 100 จะมีค่าเฉลี่ยมาก ในขณะที่ขนาดตัวอย่างเท่ากับ 1,000 ค่าเฉลี่ยจะมีค่าไม่แน่นอน บางครั้งมีค่ามากและบางครั้งมีค่าเข้าใกล้ค่าแบนวิดจ์ที่ประมาณได้

4. ลักษณะของค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in จะมีลักษณะคล้ายกับค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Rule of Thumb และวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation การกระจายจะมีจุดศูนย์กลางใกล้กับค่าแบนวิดจ์ที่เหมาะสม แต่การกระจายจะน้อยกว่าค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าความเอนเอียงมากกว่า เนื่องจากมีตัวแปรมากกว่า

5. ลักษณะของค่าแบนวิดจ์ที่ได้จากการเลือกแบนวิดจ์แบบ Smoothed Bootstrap โดยทั่วไปจะมีลักษณะคล้ายกับค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in แต่ค่าที่ได้จะมีค่ามากกว่า

Chiu (1996) ศึกษาถึงคุณสมบัติทางทฤษฎี วิธีการดำเนินงานและการแสดงผลลัพธ์ภายใต้สถานการณ์ที่แตกต่างกัน พร้อมทั้งศึกษาจากการจำลองข้อมูลภายใต้สถานการณ์ต่างๆ คือ 1) การแจกแจงแบบปกติมาตรฐาน 2) การแจกแจงแบบโคชี 3) การแจกแจงแบบไคสแควร์ กรณีศึกษา $\frac{\left(\chi_4^2-4\right)}{\sqrt{8}}$ 4) การแจกแจงแบบลอกนอร์มอล และพิจารณาถึงการแจกแจงแบบปกติผสม 7 กรณี และกำหนดขนาดตัวอย่างเท่ากับ 100, 400 และ 1,600

Mugdadi และ Ahman (2003) ศึกษาวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast และทำการจำลองข้อมูลโดยใช้การแจกแจงแบบปกติมาตรฐาน โดยศึกษาฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบไบเวท จากนั้นทำการเปรียบเทียบกราฟของฟังก์ชันของตัวแปรสุ่ม $g\left(X_1, X_2, X_3\right) = X_1 + X_2 + X_3$ โดยใช้ขนาดตัวอย่างเท่ากับ 25 พร้อมทั้งนำข้อมูลจริงมาคำนวณ

Hansen (2004) ได้ศึกษาวิธีการเลือกแบนวิดจ์แบบ Plug-in โดยใช้ค่า AMISE เป็นเกณฑ์ในการพิจารณา ทำการจำลองข้อมูลจากการแจกแจงปกติผสม 9 การแจกแจง กำหนดขนาดตัวอย่างเท่ากับ 10, 30, 60 และ 120 และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ทำการวนซ้ำจำนวน 1,000 รอบในแต่ละตัวแบบและขนาดตัวอย่าง ได้ผลการวิจัย ดังนี้

1. ค่า MISE และค่า AMISE มีรูปร่างคล้ายกันในหลายๆ กรณี โค้งการแจกแจงของข้อมูลจะใกล้เคียงกัน โดยเฉพาะเมื่อค่าแบนวิดจ์มีขนาดเล็ก สำหรับในกรณีที่ค่าแบนวิดจ์มีขนาดใหญ่ ค่า MISE จะแสดงลักษณะที่สำคัญผิดพลาด

2. ค่าแบนวิดจ์ที่ทำให้ค่า MISE และ AMISE น้อยที่สุด จะมีค่าใกล้เคียงกัน

3. โค้งข้อมูลที่ได้จากค่าแบนวิดจ์ที่ใช้ค่า MISE เป็นเกณฑ์ จะผันแปรตามการแจกแจง

Mugdadi และ Ahman (2004) อธิบายขั้นตอนวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast และทำการเปรียบเทียบทั้ง 2 วิธี โดยใช้วิธีการจำลองแบบมอนติคาร์โล และขนาดตัวอย่างมี 2 ขนาด คือ 25 และ 50 และกำหนดให้ตัวแปรสุ่มมีรูปแบบ $g\left(X_1, X_2, ..., X_m\right) = \sum_{i=1}^{m} X_i$ โดยที่ $m = 2$ โดยศึกษาฟังก์ชันเคอร์เนล 3 ฟังก์ชัน คือ 1) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน 2) ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov Kernel Function) และ 3) ฟังก์ชันเคอร์เนลแบบไบเวท (Biweight Kernel Function) ทำการจำลองข้อมูลจากการแจกแจง 3 การแจกแจง คือ 1) การแจกแจงแบบปกติ 2) การแจกแจง

แบบเอกซ์โพเนนเชียล และ 3) การแจกแจงแบบโคชี ทำการเปรียบเทียบโดยนำค่าความแปรปรวนและค่าความเอนเอียงของค่าแบนวิดจ์ที่ประมาณได้มาพิจารณา สามารถสรุปผลการวิจัยได้ดังนี้

1. ค่าความแปรปรวนของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีค่าน้อยกว่าค่าความแปรปรวนของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เมื่อจำลองโดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบไบเวท

2. ในกรณีอื่นๆ ค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะมีค่าน้อยกว่าค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation กรณีที่ข้อมูลมีการแจกแจงแบบปกติ โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบไบเวท และกรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล ในฟังก์ชันเคอร์เนลแบบเกาส์เซียน

3. ค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าน้อยกว่าค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast กรณีที่ข้อมูลมีการแจกแจงแบบโคชีและการแจกแจงแบบเอกซ์โพเนนเชียล เมื่อใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบไบเวท

4. ค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีค่าน้อยกว่าค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Square Cross-Validation เมื่อข้อมูลมีการแจกแจงแบบโคชี เมื่อใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบไบเวท

5. ค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation มีค่าน้อยกว่าค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนทั้ง 2 ฟังก์ชัน

6. ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ KernelContrast จะมีค่าน้อยกว่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เมื่อทำการจำลองข้อมูลจากการแจกแจงแบบปกติและการแจกแจงแบบเอกซ์โพเนนเชียล แต่ในทางกลับกันเมื่อทำการจำลองข้อมูลจากการแจกแจงแบบโคชี ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าน้อยกว่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast

7. ในทุกๆ กรณีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) จะลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น

นอกจากนี้ Mugdadi และ Ahman ยังนำข้อมูลจริงโดยใช้ข้อมูลที่เกี่ยวข้องกับการประกันภัยจากบริษัทออโต้ การประกันภัย ประเทศสหรัฐอเมริกา เดือนกันยายน ในปี 2001 มาวิเคราะห์ โดยใช้ฟังก์ชันของตัวแปรสุ่ม คือ $g\left(X_1, X_2, \ldots, X_m\right) = \sum_{i=1}^{m} X_i$ โดยที่ $m = 2$ ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ได้ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เท่ากับ 0.29 และได้ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เท่ากับ 0.42 จากนั้น Mugdadi และ Ahman ยังนำข้อมูลจริงข้างต้นมาศึกษาในวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast โดยใช้ฟังก์ชันเซียนโดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนที่แตกต่างกัน 4 ค่า คือ 1, 4, 9 และ 16 โดยมีสัมประสิทธิ์ความแตกต่าง (Contrast Coefficient) แทนด้วย $p_1$, $p_2$, $p_3$ และ $p_4$ ตามลำดับ โดยสรุปผลการทดลองได้ดังนี้

1. วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation ไม่เหมาะสมสำหรับการประมาณค่าความหนาแน่น ในกรณีที่ $m = 1$

2. เมื่อสัมประสิทธิ์ความแตกต่างมีค่าเท่ากับ $\frac{1}{2}$ และ $\frac{1}{2}$ จะได้การประมาณความหนาแน่นเช่นเดียวกันกับกรณีที่สัมประสิทธิ์ความแตกต่างมีค่าเท่ากับ $\frac{1}{3}$ และ $\frac{2}{3}$ แต่จะคล้ายกับกรณีที่สัมประสิทธิ์ความแตกต่างมีค่าเท่ากับ $\frac{1}{10}$ และ $\frac{9}{10}$ แต่ไม่เหมือนกันทุกประการ

3. การประมาณความหนาแน่นที่ใช้ 2 ฟังก์ชันเคอร์เนลหรือ 4 ฟังก์ชันเคอร์เนล เมื่อฟังก์ชันมีค่าถ่วงน้ำหนักเดียวกันจะได้โค้งที่เหมือนกัน

4. เมื่อใช้ค่าสัมประสิทธิ์ความแตกต่าง ค่าแบนวิดจ์ที่ได้จากการประมาณจะมีค่าเล็กกว่าวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation แต่ถ้าใช้ฟังก์ชันเคอร์เนล 4 ฟังก์ชันในการประมาณความหนาแน่น วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่าแบนวิดจ์และประมาณความหนาแน่นได้ปรับเรียบมากกว่า

Mugdadi และ Ahman สรุปว่าตัวเลือกของแบนวิดจ์โดยใช้ข้อมูลจริงจะสม่ำเสมอ และจากการจำลองข้อมูล จะเห็นว่าค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่าแบนวิดจ์และการประมาณความหนาแน่นที่ปรับเรียบมากกว่า

# บทที่ 3
# วิธีการดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงทดลอง (Experimental Research) โดยมีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบวิธีการเลือกแบนวิดจ์ 4 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ Least Square Cross – Validation วิธีการเลือกแบนวิดจ์แบบ Plug-in วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast สำหรับการประมาณความหนาแน่นแบบเคอร์เนลสำหรับฟังก์ชันของตัวแปรสุ่ม โดยสร้างเลขคล้ายสุ่มด้วยเทคนิคการจำลองมอนติคาร์โล

ในแต่ละวิธีการเลือกแบนวิดจ์จะกระทำภายใต้สถานการณ์ต่างๆ โดยในแต่ละสถานการณ์จะมีการทำซ้ำในสถานการณ์เดียวกัน จำนวน 1,000 รอบและทำการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลสำหรับฟังก์ชันของตัวแปรสุ่มทั้ง 4 วิธี โดยเปรียบเทียบจากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าเฉลี่ยของคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) และกำหนดให้มีขนาดตัวอย่าง 5 ขนาด คือ 10, 25, 50, 100 และ 500

### 3.1 ขอบเขตงานวิจัย

3.1.1 ฟังก์ชันเคอร์เนลที่ใช้ในงานวิจัย มี 3 ฟังก์ชัน คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบควอติค โดยแบ่งกรณีศึกษาได้ดังนี้

3.1.1.1 กรณีศึกษาที่ 1 กำหนดให้ $K$ เป็นฟังก์ชันเคอร์เนลแบบเกาส์เซียน

3.1.1.2 กรณีศึกษาที่ 2 กำหนดให้ $K$ เป็นฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

3.1.1.3 กรณีศึกษาที่ 3 กำหนดให้ $K$ เป็นฟังก์ชันเคอร์เนลแบบควอติค

3.1.1.4 กรณีศึกษาที่ 4 กำหนดให้ $K_1$ เป็นฟังก์ชันเคอร์เนลแบบเกาส์เซียน และ $K_2$ เป็นฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

3.1.1.5 กรณีศึกษาที่ 5 กำหนดให้ $K_1$ เป็นฟังก์ชันเคอร์เนลแบบเกาส์เซียน และ $K_2$ เป็นฟังก์ชันเคอร์เนลแบบควอติค

3.1.1.6 กรณีศึกษาที่ 6 กำหนดให้ $K_1$ เป็นฟังก์ชันเคอร์เนลแบบควอติค และ $K_2$ เป็นฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

3.1.2 ฟังก์ชันของตัวแปรสุ่มที่สนใจศึกษา คือ $g\left(X_1, X_2, ..., X_m\right) = X_1 + X_2$

3.1.3 กำหนดให้มีขนาดตัวอย่าง 5 ขนาด คือ 10, 25, 50, 100 และ 500

3.1.4 ตัวแปรสุ่มที่นำมาศึกษาครั้งนี้ใช้มาจากการแจกแจง 3 การแจกแจง คือ การแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบเอกซ์โพเนนเชียล และการแจกแจงแบบโคชี

3.1.5 ค่าแบนวิดจ์ที่ได้จะต้องเป็นจำนวนจริงบวก

## 3.2 แผนการทดลอง

การวิจัยครั้งนี้เป็นการเลือกค่าแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลโดยกำหนดให้มีฟังก์ชันเคอร์เนล 3 ฟังก์ชัน แบ่งเป็น 6 กรณีการศึกษา สำหรับฟังก์ชันของตัวแปรสุ่มที่มาจากการแจกแจง 3 การแจกแจง เริ่มต้นจากการสร้างข้อมูลให้มีการแจกแจงตามที่ต้องการ จากนั้นคำนวณหาค่าแบนวิดจ์ เพื่อนำค่าแบนวิดจ์ที่ได้ไปคำนวณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ขนาดตัวอย่าง 5 ขนาด พิจารณาค่า MSE และ MISE เพื่อเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์

## 3.3 ขั้นตอนการวิจัย

ผู้วิจัยทำการทดลองโดยการสร้างข้อมูลและทำการวิเคราะห์จากโปรแกรมคอมพิวเตอร์ด้วยการเขียนฟังก์ชันในโปรแกรม R โดยขั้นตอนของการวิจัยจะแบ่งเป็น 4 ขั้นตอนดังนี้

1. สร้างข้อมูลของตัวแปรสุ่ม $X_1$ และ $X_2$ ให้มีการแจกแจงที่กำหนด
2. กำหนดฟังก์ชันแบบเคอร์เนลตามสถานการณ์ที่กำหนด
3. ประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 4 วิธี
4. คำนวณค่า MSE และค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์

ซึ่งแต่ละขั้นตอนมีรายละเอียดดังนี้

1. สร้างข้อมูลของตัวแปรสุ่ม $X_1$ และ ตัวแปรสุ่ม $X_2$ ให้มีการแจกแจงที่กำหนด ดังนี้

1.1 สร้างตัวแปรสุ่ม $X_1$ และ $X_2$ ให้มีการแจกแจงตามที่กำหนด

ผู้วิจัยจะสร้างข้อมูลของตัวแปรสุ่ม $X_1$ และตัวแปรสุ่ม $X_2$ ให้มีการแจกแจงและขนาดที่กำหนด คือ

1.1.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

1.1.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

1.1.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

1.2 สร้างฟังก์ชันของตัวแปรสุ่ม เนื่องจากผู้วิจัยกำหนดให้ฟังก์ชันของตัวแปร

สุ่มมีลักษณะ $g\left(X_1, X_2, ..., X_m\right) = \sum_{i=1}^{m} X_i$ ดังนั้นผู้วิจัยจึงสร้างข้อมูลให้มีลักษณะดังกล่าว คือ $X_1 + X_2$

1.3 กำหนดให้ฟังก์ชันเคอร์เนลเป็นฟังก์ชันลำดับที่ 2 กล่าวคือ $k = 2$

2. กำหนดฟังก์ชันแบบเคอร์เนลเป็นไปตามสถานการณ์ ซึ่งสามารถแบ่งเป็นกรณีได้ดังนี้

2.1 กรณีศึกษาที่ 1 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ย เท่ากับ 0 และความแปรปรวนเท่ากับ 1

2.2 กรณีศึกษาที่ 2 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์อีพานิชนิคอฟ

2.3 กรณีศึกษาที่ 3 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติค

2.4 กรณีศึกษาที่ 4 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

2.5 กรณีศึกษาที่ 5 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์เนลแบบควอติค

2.6 กรณีศึกษาที่ 6 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และ $K_2$ คือ ฟังก์ชันเคอร์เนลแบบควอติค

โดยในแต่ละสถานการณ์ที่กำหนดจะสร้างตัวแปรสุ่มตามการแจกแจงที่กล่าวไปแล้วข้างต้น

3. ประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 4 วิธี คือ

3.1 วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation

$$LSCV\left(h\right) \quad = \quad \int_{-\infty}^{\infty} \hat{f}_h^2\left(x\right)dx - \frac{2}{n}\sum_{i=1}^{n} f_{h,i}\left(x_i\right)$$

3.2 วิธีการเลือกแบนวิดจ์แบบ Plug-in

$$h \quad = \quad \left[\frac{R\left(K\right)}{\sigma_K^4 \hat{S}_D\left(\hat{\alpha}\left(h\right)\right)}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}$$

3.3 วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap

$$MISE\left(h\right) \quad = \quad B^{-1}\sum_{j=1}^{B}\int_{-\infty}^{\infty}\left(f_{nj}^*\left(x,h\right) - \hat{f}_n\left(x,h_0\right)\right)^2 dx$$

3.4 วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast

$$ISE(h) = \int_{-\infty}^{\infty}\left(\sum_{i=1}^{q} p_i \hat{f}_{h,i}(x)\right)^2 dx$$

4. คำนวณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความของคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 4 วิธี

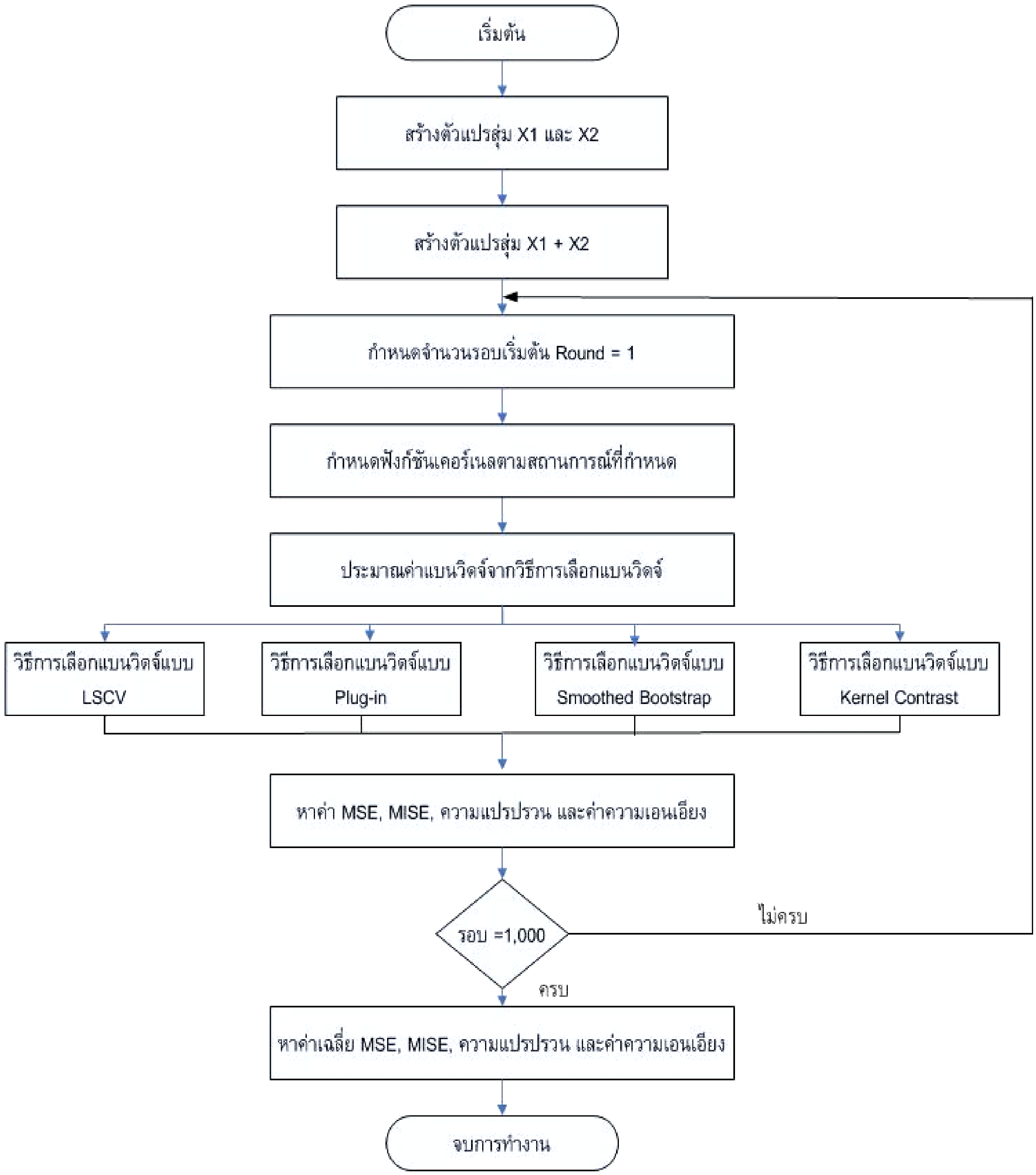
4.1 ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง

$$MSE\left(\hat{f}(x)\right) = \frac{\sum_{i=1}^{1000} E\left[\hat{f}_h(x) - f(x)\right]^2}{1,000}$$

4.2 ค่าเฉลี่ยของอนุพันธ์ของความคลาดเคลื่อนกำลังสอง

$$MISE\left(\hat{f}(x)\right) = \frac{\sum_{i=1}^{1000} E\int\left[\hat{f}_h(x) - f(x)\right]^2 dx}{1,000}$$

ทุกขั้นตอนของการวิจัยเป็นไปตามภาพที่ 3-1




**ภาพที่ 3-1** แผนผังวิธีการดำเนินการวิจัย

# บทที่ 4
# ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาข้อสรุปที่เหมาะสมในการเลือกค่าแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลสำหรับฟังก์ชันของตัวแปรสุ่ม โดยผู้วิจัยได้ทำการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ของวิธีการเลือกแบนวิดจ์ 4 วิธี คือวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation วิธีการเลือกแบนวิดจ์แบบ Plug-in วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ภายใต้สถานการณ์ที่กำหนด โดยในการวิเคราะห์ผลผู้วิจัยใช้เกณฑ์การเปรียบเทียบคือ หากวิธีใดให้ค่า MSE และ MISE ต่ำกว่าจะเป็นวิธีการที่มีประสิทธิภาพมากกว่า

ผลการวิจัยในครั้งนี้ จำแนกได้ 2 ลักษณะ คือ ค่าเฉลี่ยของความเคลื่อนกำลังสอง (MSE) และค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ซึ่งได้นำเสนอในรูปแบบตารางและรูปภาพกราฟ ซึ่งใช้สัญลักษณ์ในการนำเสนอ ดังนี้

| | | |
|---|---|---|
| LSCV | หมายถึง | วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation |
| PI | หมายถึง | วิธีการเลือกแบนวิดจ์แบบ Plug-in ของ Shealter และ Jones |
| SB | หมายถึง | วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap |
| KC | หมายถึง | วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast |
| $h$ | หมายถึง | ค่าแบนวิดจ์ |
| MSE | หมายถึง | ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง |
| MISE | หมายถึง | ค่าเฉลี่ยของความของคลาดเคลื่อนกำลังสองเฉลี่ยรวม |

ผู้วิจัยได้วางแผนทำการทดลองโดยกำหนดให้ฟังก์ชันเคอร์เนลมีลำดับที่ $k = 2$ ภายใต้สถานการณ์ที่กำหนด ดังนี้คือ

1. กรณีศึกษาที่ 1 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

1.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

1.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

1.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

2. กรณีศึกษาที่ 2 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

2.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

2.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

2.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

3. กรณีศึกษาที่ 3 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติคโดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

3.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

3.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

3.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

4. กรณีศึกษาที่ 4 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

4.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

5. กรณีศึกษาที่ 5 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

5.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

5.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

5.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

6. กรณีศึกษาที่ 6 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และ $K_2$ คือ ฟังก์ชันเคอร์ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

6.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

6.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

6.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ต่อไปนี้จะแสดงตาราง กราฟ และสรุปรายละเอียดตามกรณีศึกษา ตามที่กล่าวมาข้างต้น

**4.1 กรณีศึกษาที่ 1**

กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.1.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

4.1.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.1.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ผลการวิจัยนำเสนอในตารางที่ 4-1 ถึง 4-6 และภาพที่ 4-1 ถึง 4-6 โดยแสดงค่าแบนวิดจ์, ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการเลือกแบนวิดจ์ทั้ง 3 วิธี สามารถสรุปรายละเอียด ได้ดังนี้

**ตารางที่ 4-1** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนโดยจำแนกตามขนาดของตัวอย่าง

| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 0.4294 |
| | PI | 1.8231 |
| | SB | 2.438 |
| 25 | LSCV | 0.4102 |
| | PI | 1.7649 |
| | SB | 2.2783 |
| 50 | LSCV | 0.3958 |
| | PI | 1.4311 |
| | SB | 2.0748 |
| 100 | LSCV | 0.3347 |
| | PI | 1.2847 |
| | SB | 2.0035 |

**ตารางที่ 4-1 (ต่อ)**

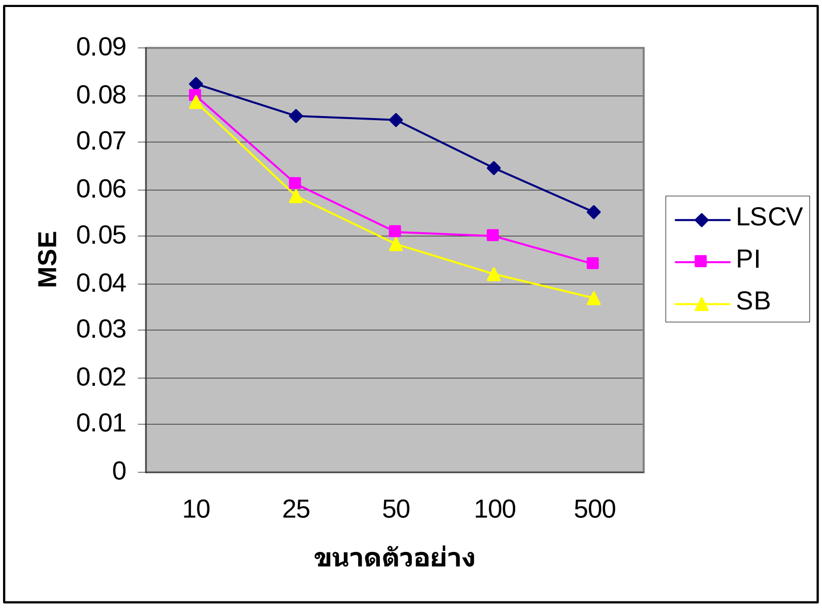
| $n$ | วิธีการเลือกแบนวิดจ์ | $h$ |
|---|---|---|
| 500 | LSCV | 0.3417 |
| | PI | 1.2258 |
| | SB | 1.9024 |

จากตารางที่ 4-1 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 3 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยที่ค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็กที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่าแบนวิดจ์ขนาดใหญ่ที่สุด

**ตารางที่ 4-2** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน โดยจำแนกตามขนาดของตัวอย่าง

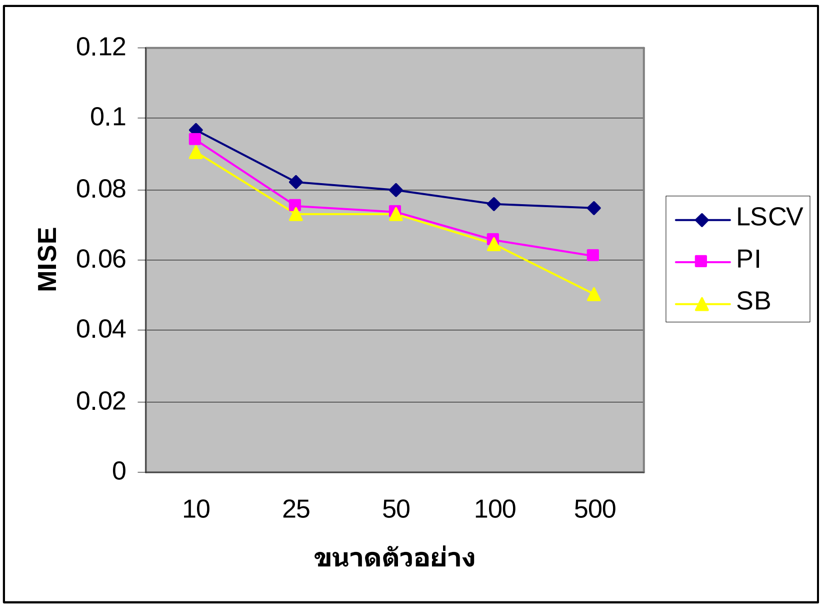
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.0825 | 0.0969 |
| | PI | 0.0800 | 0.0938 |
| | SB | 0.0784 | 0.0908 |
| 25 | LSCV | 0.0755 | 0.0820 |
| | PI | 0.0613 | 0.0755 |
| | SB | 0.0585 | 0.0730 |
| 50 | LSCV | 0.0749 | 0.0797 |
| | PI | 0.0510 | 0.0735 |
| | SB | 0.0484 | 0.0730 |
| 100 | LSCV | 0.0646 | 0.0756 |
| | PI | 0.0499 | 0.0655 |
| | SB | 0.0421 | 0.0646 |
| 500 | LSCV | 0.0550 | 0.0749 |
| | PI | 0.0443 | 0.0612 |
| | SB | 0.0369 | 0.0501 |




**ภาพที่ 4-1** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน

จากตารางที่ 4-2 และภาพที่ 4-1 แสดงให้เห็นถึงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ LSCV วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกค่าขนาดตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE สูงที่สุด




**ภาพที่ 4-2** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน

จากตารางที่ 4-2 และภาพที่ 4-2 แสดงให้เห็นถึงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ LSCV วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MISE มากที่สุด

**ตารางที่ 4-3** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนโดยจำแนกตามขนาดของตัวอย่าง

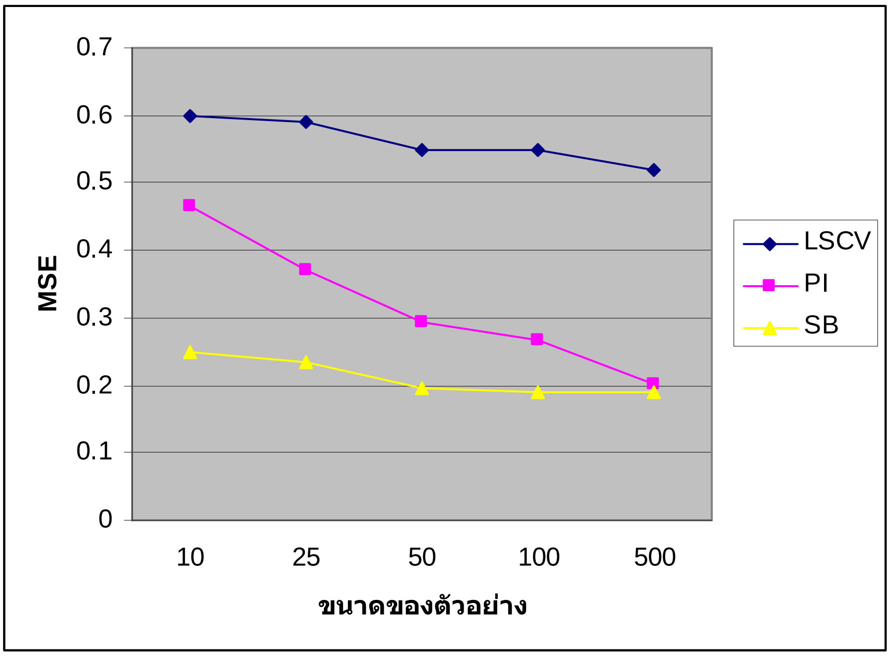
| $n$ | วิธีการเลือกแบนวิดจ์ | $h$ |
|---|---|---|
| 10 | LSCV | 0.4311 |
| | PI | 1.9147 |
| | SB | 2.1690 |
| 25 | LSCV | 0.5345 |
| | PI | 2.3945 |
| | SB | 2.6485 |
| 50 | LSCV | 0.5825 |
| | PI | 2.3582 |
| | SB | 2.8912 |
| 100 | LSCV | 0.5060 |
| | PI | 1.7899 |
| | SB | 2.1560 |
| 500 | LSCV | 0.1956 |
| | PI | 1.0752 |
| | SB | 1.4592 |

จากตารางที่ 4-3 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 3 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะมีค่าขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap

**ตารางที่ 4-4** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนโดยจำแนกตามขนาดของตัวอย่าง

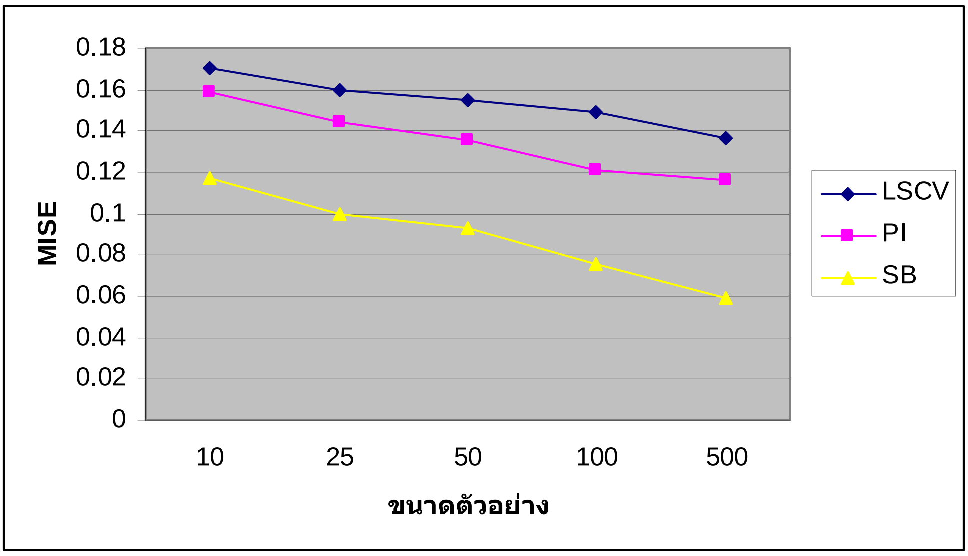
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.5997 | 0.1702 |
| | PI | 0.4668 | 0.1585 |
| | SB | 0.2503 | 0.1172 |
| 25 | LSCV | 0.5911 | 0.1601 |
| | PI | 0.3709 | 0.1441 |
| | SB | 0.2344 | 0.0999 |
| 50 | LSCV | 0.5479 | 0.1544 |
| | PI | 0.2929 | 0.1356 |
| | SB | 0.1966 | 0.0928 |
| 100 | LSCV | 0.5478 | 0.1492 |
| | PI | 0.2664 | 0.1208 |
| | SB | 0.1905 | 0.0759 |
| 500 | LSCV | 0.5184 | 0.1360 |
| | PI | 0.2023 | 0.1166 |
| | SB | 0.1904 | 0.0586 |




**ภาพที่ 4-3** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน

จากตารางที่ 4-4 และภาพที่ 4-3 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี คือวิธีการเลือกแบนวิดจ์แบบ LSCV วิธีการเลือกแบนวิดจ์แบบ plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกค่าขนาดตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE สูงที่สุด




**ภาพที่ 4-4** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน

จากตารางที่ 4-4 และภาพที่ 4-4 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ LSCV ตามลำดับ

**ตารางที่ 4-5** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SBเมื่อข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน โดยจำแนกตามขนาดของตัวอย่าง

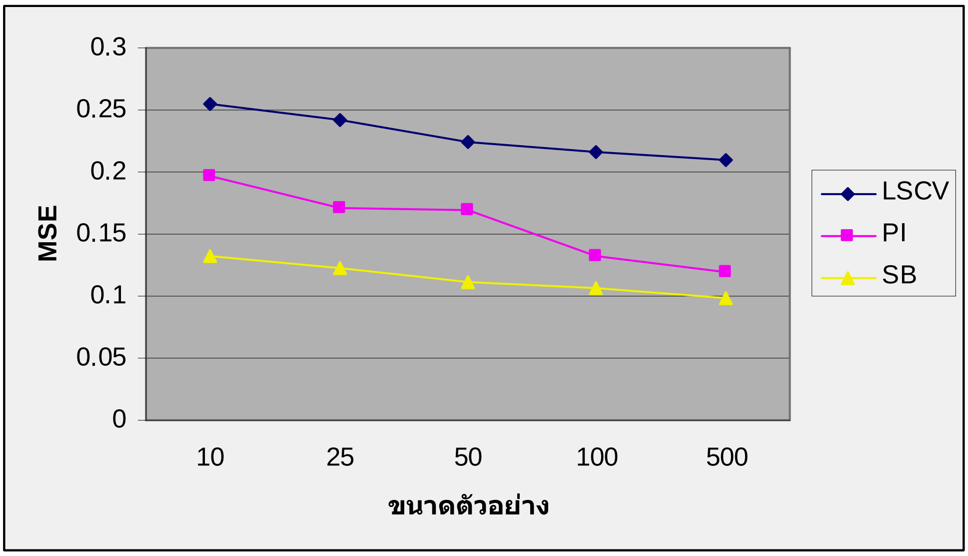
| $n$ | วิธีการเลือกแบนวิดจ์ | $h$ |
|---|---|---|
| 10 | LSCV | 0.6463 |
| | PI | 3.0744 |
| | SB | 3.2640 |
| 25 | LSCV | 1.2197 |
| | PI | 5.4873 |
| | SB | 4.8592 |
| 50 | LSCV | 0.8008 |
| | PI | 4.3982 |
| | SB | 4.7852 |
| 100 | LSCV | 0.8090 |
| | PI | 3.4777 |
| | SB | 3.8945 |
| 500 | LSCV | 0.4818 |
| | PI | 2.6140 |
| | SB | 3.1245 |

จากตารางที่ 4-5 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 3 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 เมื่อข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็ก เมื่อเปรียบเทียบค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap โดยค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่าแบนวิดจ์ที่มีขนาดใหญ่ที่สุด ยกเว้นในกรณีที่ $n = 25$

**ตารางที่ 4-6** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน โดยจำแนกตามขนาดของตัวอย่าง

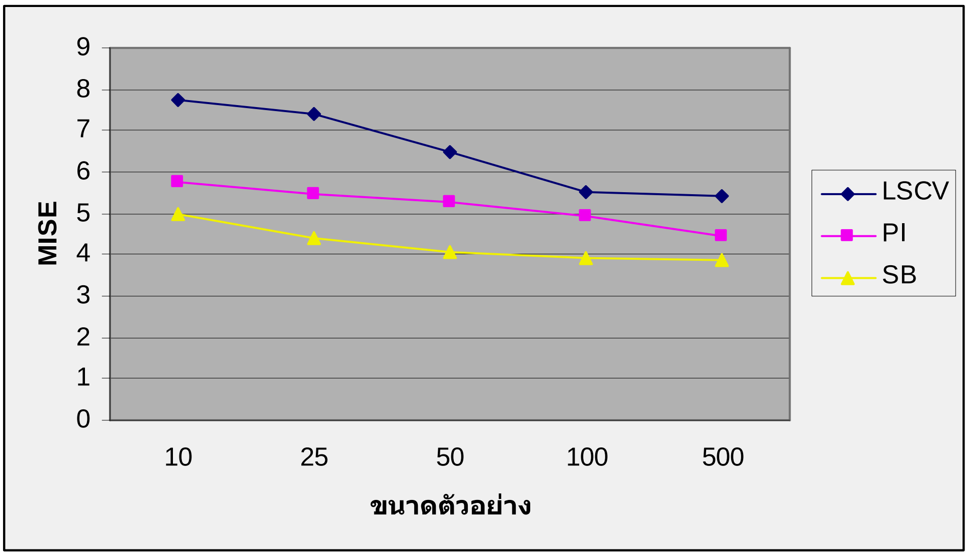
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.2548 | 7.7626 |
| | PI | 0.1965 | 5.7614 |
| | SB | 0.1327 | 4.9955 |
| 25 | LSCV | 0.2417 | 7.4162 |
| | PI | 0.1713 | 5.4128 |
| | SB | 0.1226 | 4.3842 |
| 50 | LSCV | 0.2298 | 6.4788 |
| | PI | 0.1688 | 5.2506 |
| | SB | 0.1108 | 4.0498 |
| 100 | LSCV | 0.2156 | 5.5081 |
| | PI | 0.1315 | 4.9448 |
| | SB | 0.1058 | 3.9333 |
| 500 | LSCV | 0.2104 | 5.4403 |
| | PI | 0.1190 | 4.4495 |
| | SB | 0.0979 | 3.8779 |




**ภาพที่ 4-5** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB เมื่อข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน

จากตารางที่ 4-6 และภาพที่ 4-5 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ LSCV ตามลำดับ




**ภาพที่ 4-6** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน

จากตารางที่ 4-6 และภาพที่ 4-6 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ LSCV ตามลำดับ

**4.2 กรณีศึกษาที่ 2**

กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.2.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

4.2.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.2.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ผลการวิจัยนำเสนอในตารางที่ 4-7 ถึง 4-12 และภาพที่ 4-7 ถึง ภาพที่ 4-12 โดยแสดงค่าแบนวิดจ์, ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการเลือกแบนวิดจ์ทั้ง 3 วิธี สามารถสรุปรายละเอียด ได้ดังนี้

**ตารางที่ 4-7** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยจำแนกตามขนาดของตัวอย่าง

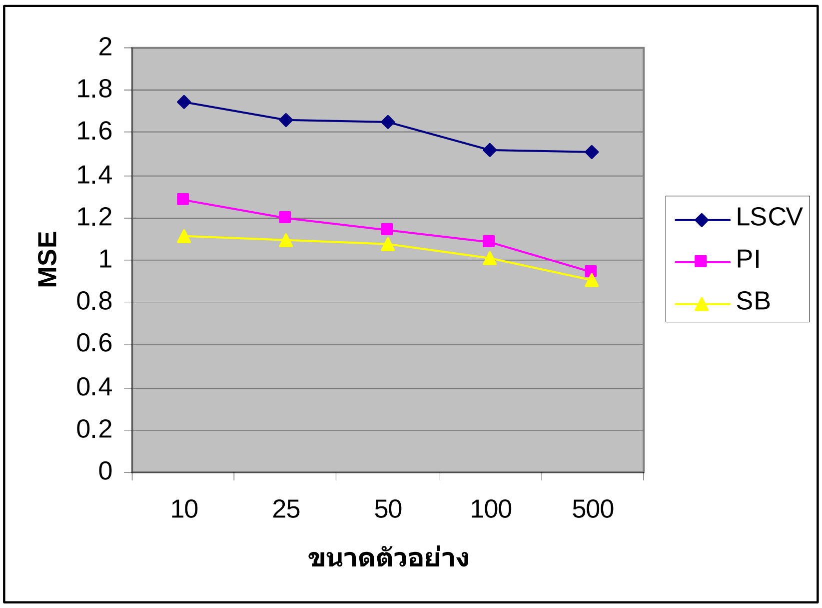
| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 0.4324 |
| | PI | 1.2142 |
| | SB | 1.8975 |
| 25 | LSCV | 0.4821 |
| | PI | 2.4612 |
| | SB | 2.695 |
| 50 | LSCV | 0.8233 |
| | PI | 2.6398 |
| | SB | 3.1236 |
| 100 | LSCV | 0.7492 |
| | PI | 2.4804 |
| | SB | 2.974 |
| 500 | LSCV | 0.5463 |
| | PI | 1.6589 |
| | SB | 2.254 |

จากตารางที่ 4-7 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 4 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะมีค่าขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap โดยค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่าแบนวิดจ์ที่มีขนาดใหญ่ที่สุด

**ตารางที่ 4-8** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยจำแนกตามขนาดของตัวอย่าง

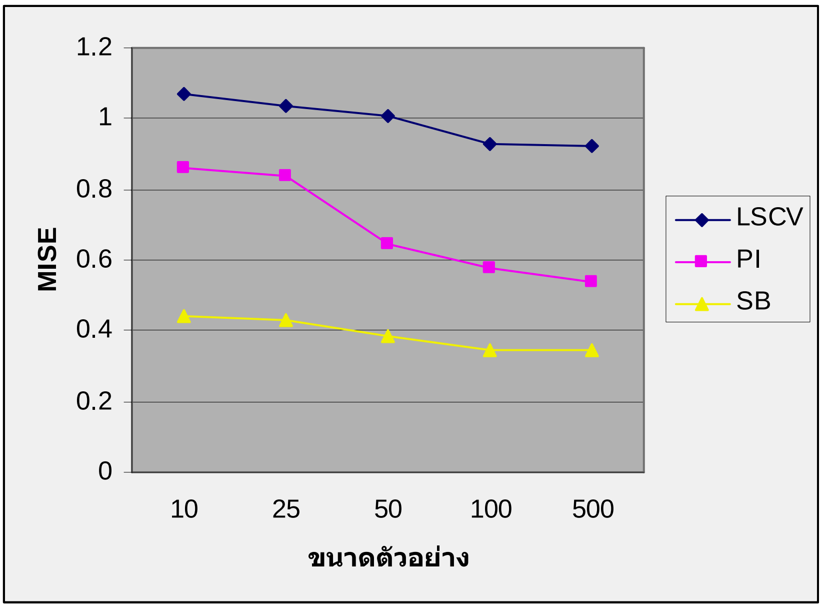
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 1.7472 | 1.0697 |
| | PI | 1.2799 | 0.8618 |
| | SB | 1.1086 | 0.4398 |
| 25 | LSCV | 1.665 | 1.0379 |
| | PI | 1.1937 | 0.8367 |
| | SB | 1.0981 | 0.4280 |
| 50 | LSCV | 1.6500 | 1.0053 |
| | PI | 1.1374 | 0.6425 |
| | SB | 1.0765 | 0.3861 |
| 100 | LSCV | 1.5197 | 0.9310 |
| | PI | 1.0880 | 0.5790 |
| | SB | 1.0112 | 0.3444 |
| 500 | LSCV | 1.5094 | 0.9243 |
| | PI | 0.9413 | 0.5403 |
| | SB | 0.9069 | 0.3443 |




**ภาพที่ 4-7** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-8 และภาพที่ 4-7 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MSE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in โดยที่วิธีการเลือกแบนวิดจ์แบบ LSCV ให้ค่า MSE มากที่สุด




**ภาพที่ 4-8** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-8 และภาพที่ 4-8 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยที่วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ LSCV ตามลำดับ

**ตารางที่ 4-9** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟโดยจำแนกตามขนาดของตัวอย่าง

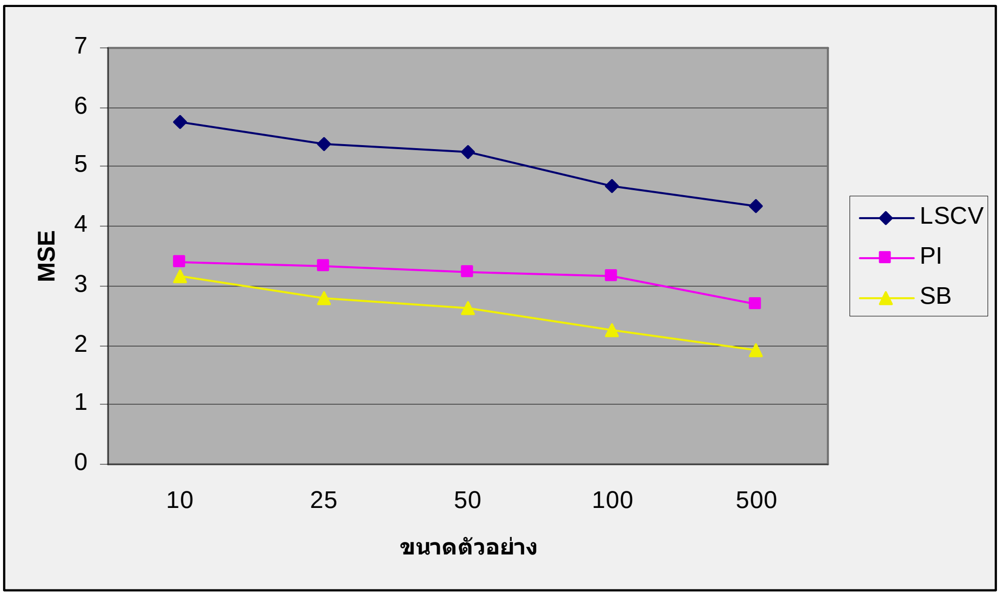
| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 1.0456 |
| | PI | 4.4758 |
| | SB | 3.562 |
| 25 | LSCV | 0.4847 |
| | PI | 2.7412 |
| | SB | 3.124 |
| 50 | LSCV | 0.4672 |
| | PI | 1.9697 |
| | SB | 2.189 |
| 100 | LSCV | 0.3773 |
| | PI | 1.4452 |
| | SB | 1.568 |
| 500 | LSCV | 0.2362 |
| | PI | 1.1355 |
| | SB | 1.3264 |

จากตารางที่ 4-9 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 4 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่าแบนวิดจ์ที่ประมาณได้ที่มีขนาดใหญ่กว่าทั้ง 2 วิธีที่เหลือ มีบางกรณีเท่านั้นที่ให้ค่าแบนวิดจ์ที่มีขนาดเล็กกว่า

**ตารางที่ 4-10** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบLSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยจำแนกตามขนาดของตัวอย่าง

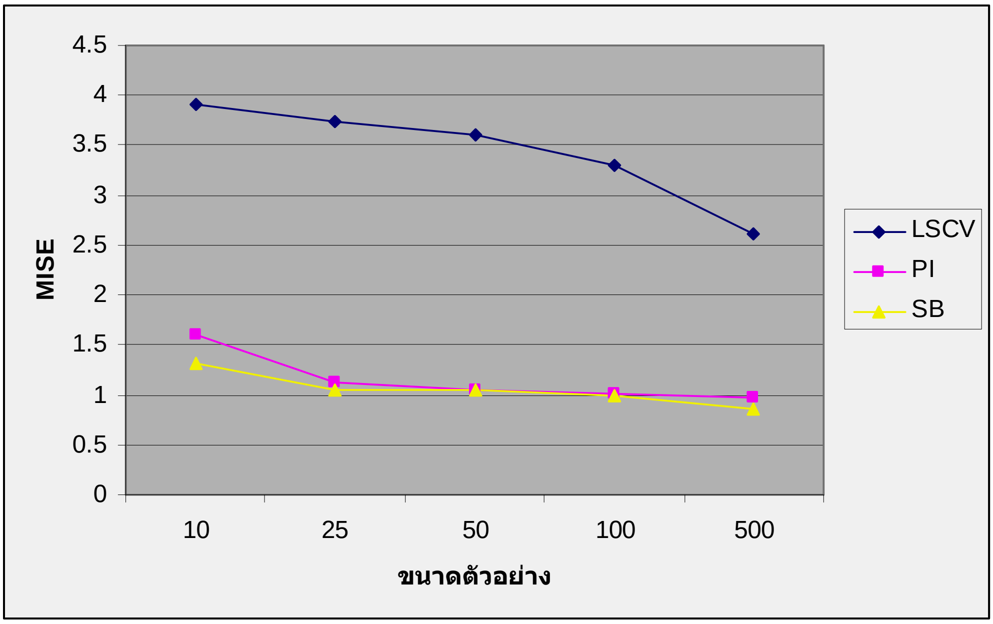
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 5.7529 | 3.9160 |
| | PI | 3.3923 | 1.5933 |
| | SB | 3.1472 | 1.3126 |
| 25 | LSCV | 2.3841 | 3.7340 |
| | PI | 3.3155 | 1.1179 |
| | SB | 2.7799 | 1.0549 |
| 50 | LSCV | 5.2354 | 3.5957 |
| | PI | 3.2382 | 1.0477 |
| | SB | 2.6348 | 1.0451 |
| 100 | LSCV | 4.6688 | 3.3073 |
| | PI | 3.1612 | 1.0115 |
| | SB | 2.2511 | 0.9834 |
| 500 | LSCV | 4.3482 | 2.6178 |
| | PI | 2.7088 | 0.9785 |
| | SB | 1.9258 | 0.8551 |




**ภาพที่ 4-9** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-10 และภาพที่ 4-9 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ กรณี ทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MISE มากที่สุด




**ภาพที่ 4-10** แสดงการเปรียบเทียบค่า MISE จากวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-10 และภาพที่ 4-10 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ที่ได้จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี มีแนวโน้มต่ำลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MISE มากที่สุด

และจากภาพที่ 4-10 จะได้ว่าเมื่อขนาดตัวอย่างเพิ่มขึ้น ค่า MISE ที่ได้วิธีการเลือกแบนวิดจ์แบบ Plug-in และ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะมีค่าใกล้เคียงกัน

**ตารางที่ 4-11** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยจำแนกตามขนาดของตัวอย่าง

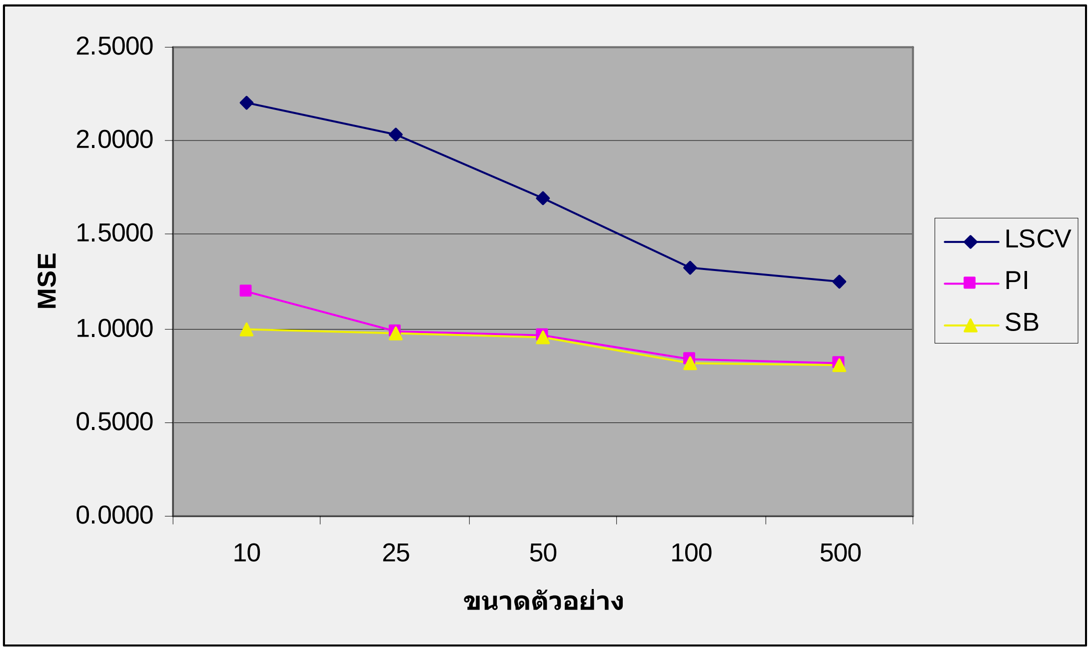
| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 1.9010 |
| | PI | 9.4682 |
| | SB | 6.7820 |
| 25 | LSCV | 0.6972 |
| | PI | 4.3120 |
| | SB | 4.2562 |
| 50 | LSCV | 0.1389 |
| | PI | 3.8113 |
| | SB | 3.6871 |
| 100 | LSCV | 0.1071 |
| | PI | 2.0918 |
| | SB | 2.516 |
| 500 | LSCV | 0.0709 |
| | PI | 1.4296 |
| | SB | 1.5234 |

จากตารางที่ 4-11 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 4 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ เมื่อข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และเมื่อขนาดตัวอย่างเพิ่มขึ้นค่าของแบนวิดจ์ที่ประมาณได้ของแต่ละวิธีมีขนาดน้อยลงตามลำดับ

**ตารางที่ 4-12** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยจำแนกตามขนาดของตัวอย่าง

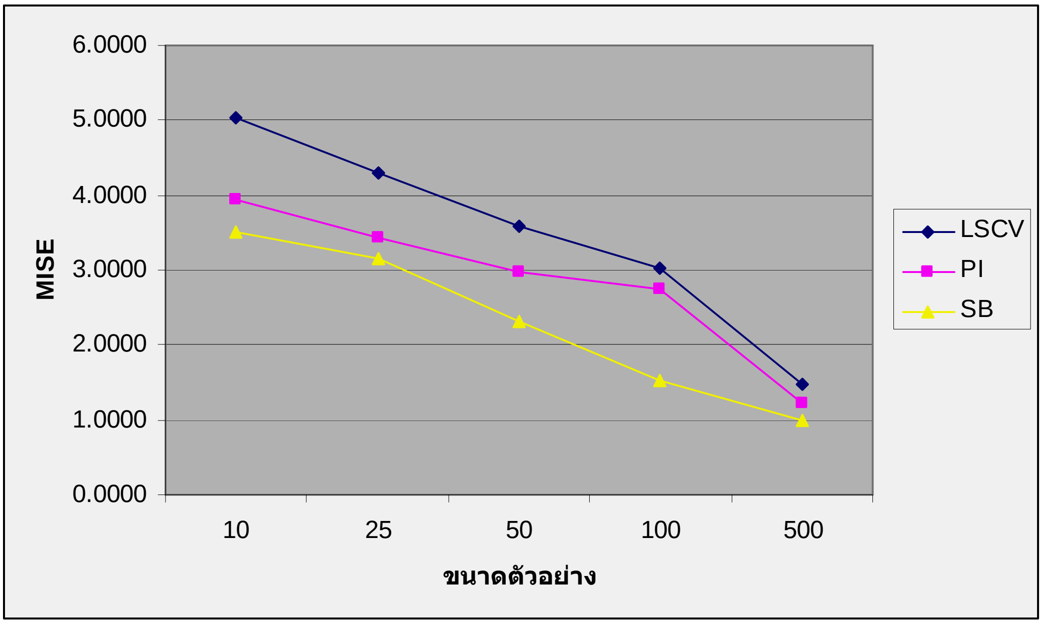
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 2.2040 | 5.0366 |
| | PI | 1.1963 | 3.9386 |
| | SB | 0.9921 | 3.5147 |
| 25 | LSCV | 2.0287 | 4.3067 |
| | PI | 0.9873 | 3.4306 |
| | SB | 0.9694 | 3.1535 |
| 50 | LSCV | 1.6957 | 3.5797 |
| | PI | 0.9654 | 2.9633 |
| | SB | 0.9575 | 2.3197 |
| 100 | LSCV | 1.3288 | 3.0132 |
| | PI | 0.8372 | 2.7557 |
| | SB | 0.8209 | 1.5370 |
| 500 | LSCV | 1.2467 | 1.4699 |
| | PI | 0.8190 | 1.2089 |
| | SB | 0.8001 | 0.9979 |




**ภาพที่ 4-11** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-12 และภาพที่ 4-11 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ที่มีค่าต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in สำหรับวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE สูงที่สุดในทุกขนาดของตัวอย่าง จากภาพที่ 4-11 จะเห็นว่าเมื่อขนาดตัวอย่างเท่ากับ 25, 50, 100 และ 500 ค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่าใกล้เคียงกัน




**ภาพที่ 4-12** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-18 และภาพที่ 4-12 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ของวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ที่มีค่าต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in สำหรับวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MISE สูงที่สุดในทุกขนาดของตัวอย่าง

**4.3 กรณีศึกษาที่ 3**

กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.3.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

4.3.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.3.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ผลการวิจัยนำเสนอในตารางที่ 4-13 ถึง 4-18 และภาพที่ 4-13 ถึง 4-18 แสดงค่าแบนวิดจ์, ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการเลือกแบนวิดจ์ทั้ง 3 วิธี สามารถสรุปรายละเอียด ได้ดังนี้

**ตารางที่ 4-13** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบควอติค โดยจำแนกตามขนาดของตัวอย่าง

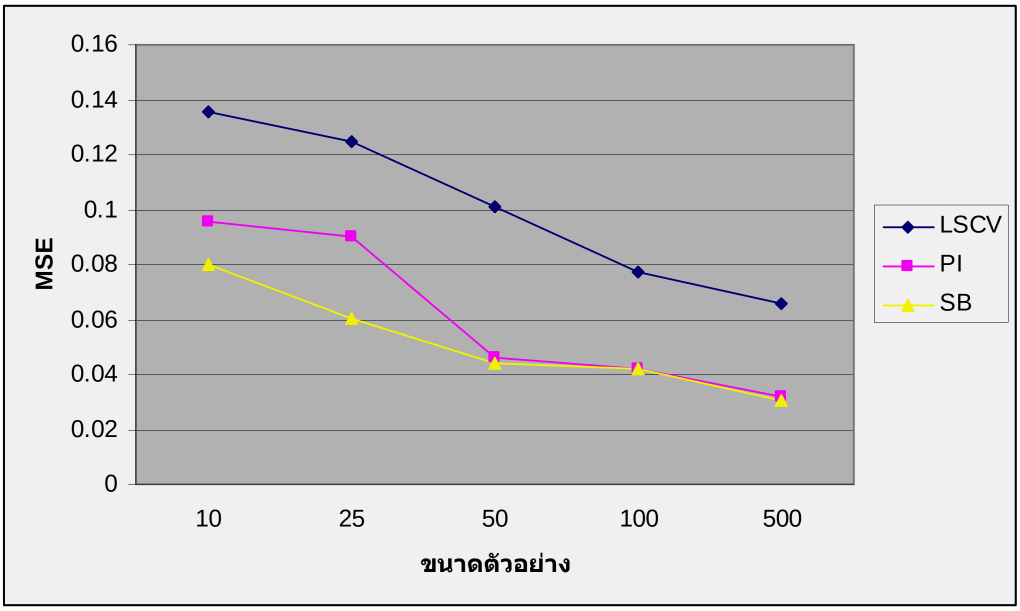
| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 1.3199 |
| | PI | 3.5807 |
| | SB | 3.7580 |
| 25 | LSCV | 1.1146 |
| | PI | 3.3025 |
| | SB | 3.4562 |
| 50 | LSCV | 0.6770 |
| | PI | 2.1819 |
| | SB | 2.9564 |
| 100 | LSCV | 0.5896 |
| | PI | 1.9237 |
| | SB | 2.023 |
| 500 | LSCV | 0.5612 |
| | PI | 1.8372 |
| | SB | 1.965 |

จากตารางที่ 4-13 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 4 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติค เมื่อข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และเมื่อขนาดตัวอย่างเพิ่มขึ้นค่าของแบนวิดจ์ที่ประมาณได้ของแต่ละวิธีมีขนาดน้อยลงตามลำดับ

**ตารางที่ 4-14** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบควอติค โดยจำแนกตามขนาดของตัวอย่าง

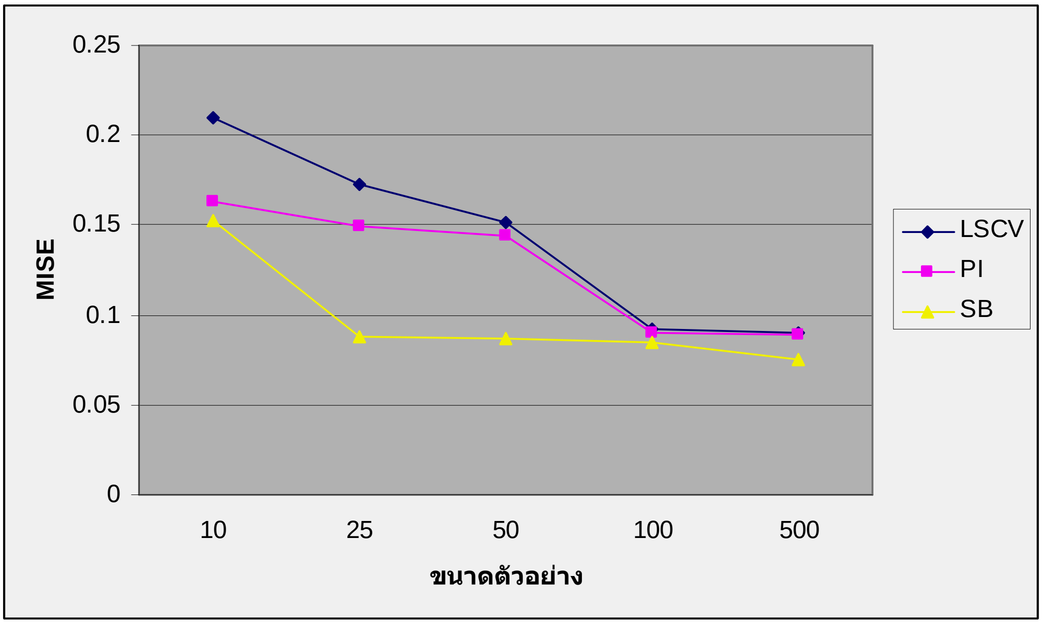
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.1358 | 0.2095 |
| | PI | 0.0958 | 0.1627 |
| | SB | 0.0798 | 0.1521 |
| 25 | LSCV | 0.1247 | 0.1730 |
| | PI | 0.0902 | 0.1498 |
| | SB | 0.0606 | 0.0883 |
| 50 | LSCV | 0.1008 | 0.1514 |
| | PI | 0.0463 | 0.1444 |
| | SB | 0.0442 | 0.0872 |
| 100 | LSCV | 0.0775 | 0.0917 |
| | PI | 0.0422 | 0.0901 |
| | SB | 0.0420 | 0.0852 |
| 500 | LSCV | 0.0661 | 0.0905 |
| | PI | 0.0316 | 0.0889 |
| | SB | 0.0304 | 0.0757 |




**ภาพที่ 4-13** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-14 และภาพที่ 4-13 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบควอติค และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยที่วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุดในทุกขนาดของตัวอย่าง โดยจากภาพที่ 4-13 จะเห็นว่าเมื่อขนาดตัวอย่างมีขนาดใหญ่ขึ้น ค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ smoothed Bootstrap กับค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ Plug-in จะมีค่าใกล้เคียงกัน




**ภาพที่ 4-14** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-14 และภาพที่ 4-14 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบควอติค และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยที่วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุดในทุกขนาดของตัวอย่าง

**ตารางที่ 4-15** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบควอติคโดยจำแนกตามขนาดของตัวอย่าง

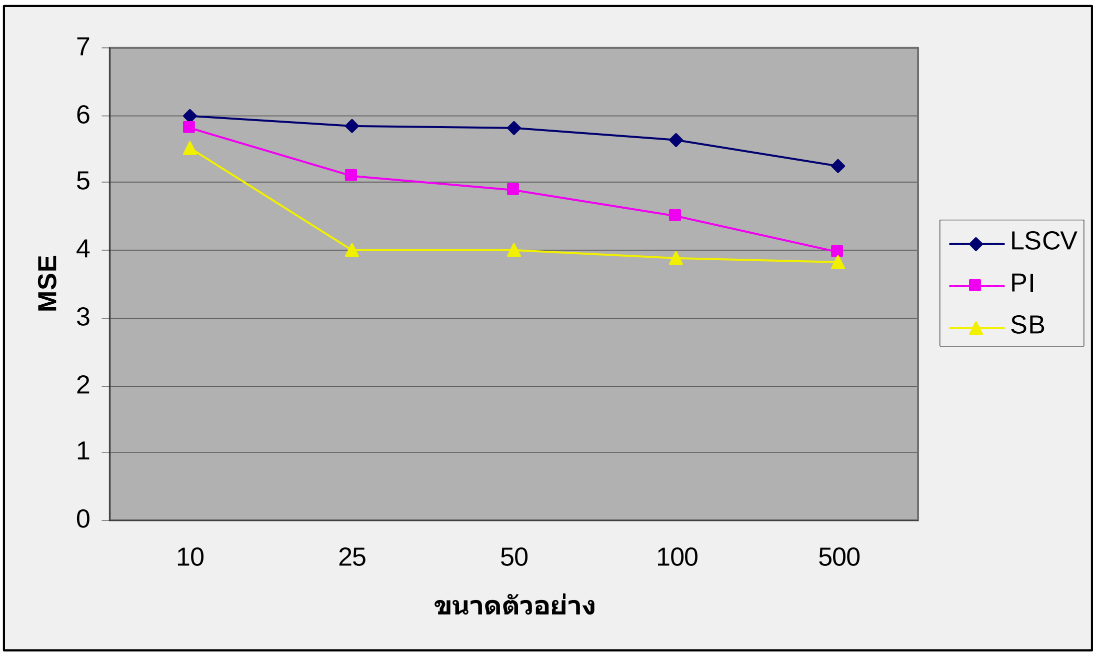
| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 0.644 |
| | PI | 2.397 |
| | SB | 3.128 |
| 25 | LSCV | 0.561 |
| | PI | 2.357 |
| | SB | 2.856 |
| 50 | LSCV | 0.390 |
| | PI | 1.607 |
| | SB | 2.145 |
| 100 | LSCV | 0.270 |
| | PI | 1.418 |
| | SB | 1.984 |
| 500 | LSCV | 0.071 |
| | PI | 1.041 |
| | SB | 1.563 |

จากตารางที่ 4-22 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 4 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติค เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และเมื่อขนาดตัวอย่างเพิ่มขึ้นค่าของแบนวิดจ์ที่ประมาณได้ของแต่ละวิธีมีขนาดน้อยลงตามลำดับ

**ตารางที่ 4-16** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบควอติค โดยจำแนกตามขนาดของตัวอย่าง

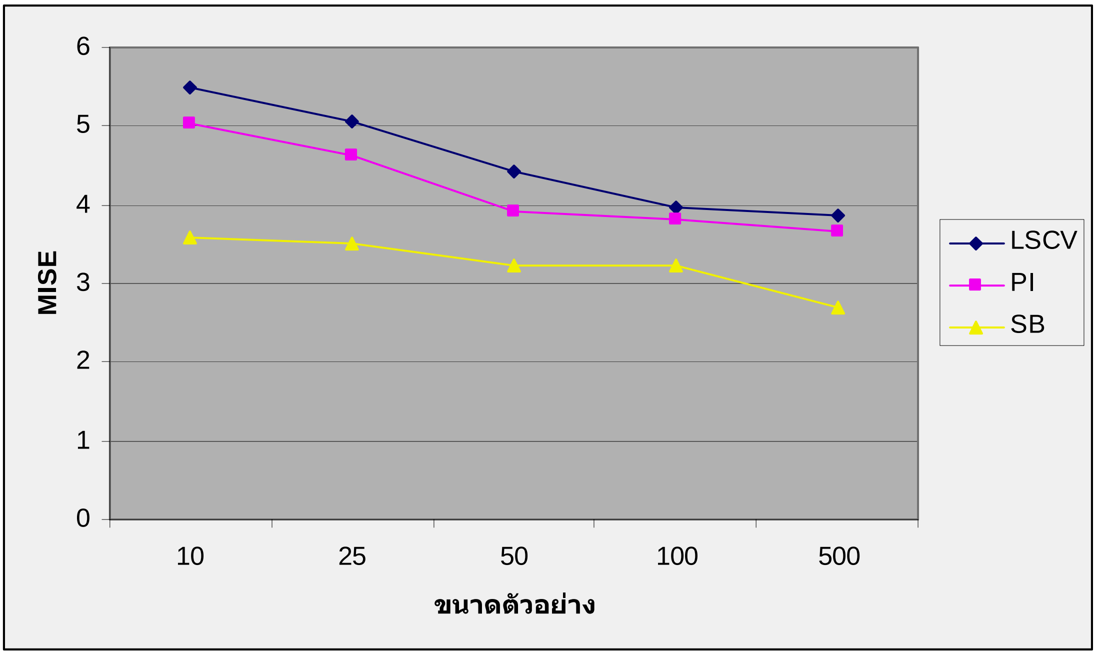
| *n* | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 5.9869 | 5.4873 |
| | PI | 5.8090 | 5.0358 |
| | SB | 4.5211 | 3.5814 |
| 25 | LSCV | 5.8298 | 5.0592 |
| | PI | 5.0897 | 4.6336 |
| | SB | 4.0144 | 3.5106 |
| 50 | LSCV | 5.8047 | 4.4220 |
| | PI | 4.9011 | 3.9081 |
| | SB | 3.9947 | 3.2293 |
| 100 | LSCV | 5.6269 | 3.9545 |
| | PI | 4.5092 | 3.8240 |
| | SB | 3.8979 | 3.2177 |
| 500 | LSCV | 5.2356 | 3.8671 |
| | PI | 3.9800 | 3.6569 |
| | SB | 3.8224 | 2.705 |




**ภาพที่ 4-15** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-23 และภาพที่ 4-15 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบควอติค และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยที่วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุด




**ภาพที่ 4-16** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-24 และภาพที่ 4-16 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบควอติค และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น โดยที่ค่า MISE จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MISE มากที่สุด

**ตารางที่ 4-17** แสดงค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ฟังก์ชันเคอร์เนลแบบควอติค โดยจำแนกตามขนาดของตัวอย่าง

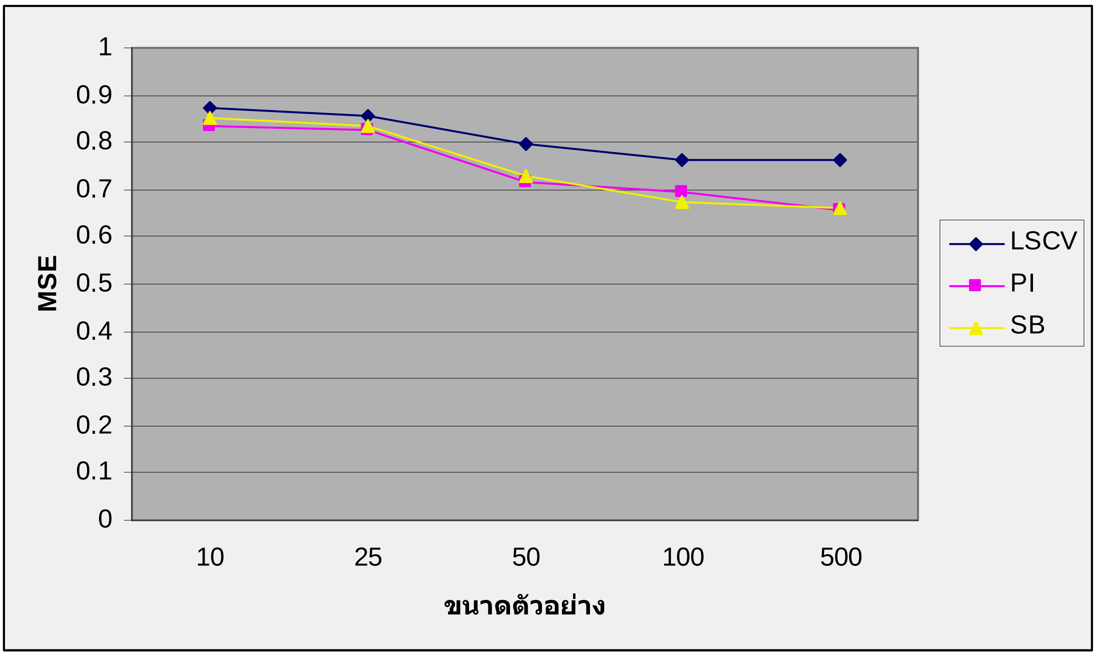
| $n$ | **วิธีการเลือกแบนวิดจ์** | $h$ |
|---|---|---|
| 10 | LSCV | 1.5709 |
| | PI | 7.3530 |
| | SB | 6.5482 |
| 25 | LSCV | 1.3836 |
| | PI | 6.6077 |
| | SB | 5.8463 |
| 50 | LSCV | 1.3077 |
| | PI | 5.2898 |
| | SB | 4.9521 |
| 100 | LSCV | 0.1042 |
| | PI | 3.2249 |
| | SB | 2.4632 |
| 500 | LSCV | 0.0882 |
| | PI | 0.8316 |
| | SB | 1.4392 |

จากตารางที่ 4-17 แสดงค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์ ทั้ง 4 วิธี เมื่อกำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติค เมื่อข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สำหรับขนาดตัวอย่าง 10, 25, 50, 100 และ 500 ซึ่งสามารถสรุปผลได้ดังนี้

ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่าแบนวิดจ์ที่มีขนาดเล็ก เมื่อเทียบกับค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Plug-in และค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และเมื่อขนาดตัวอย่างเพิ่มขึ้นค่าของแบนวิดจ์ที่ประมาณได้ของแต่ละวิธีมีขนาดน้อยลงตามลำดับ

**ตารางที่ 4-18** แสดงการเปรียบเทียบประสิทธิภาพของการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ฟังก์ชันเคอร์เนลแบบควอติค โดยจำแนกตามขนาดของตัวอย่าง

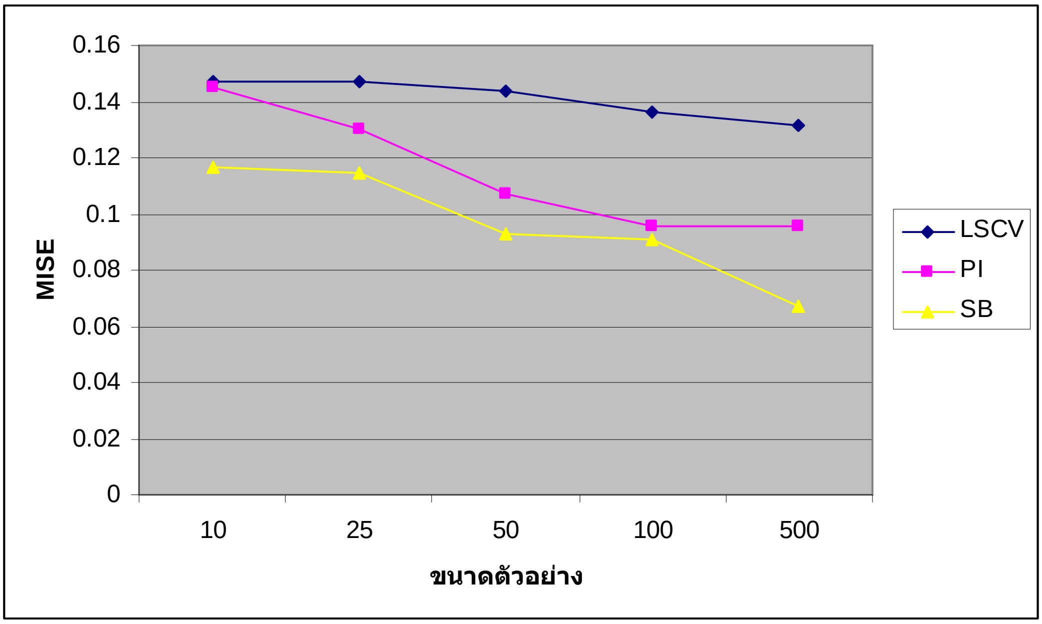
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.8714 | 0.1473 |
| | PI | 0.8337 | 0.1452 |
| | SB | 0.8512 | 0.1167 |
| 25 | LSCV | 0.8593 | 0.1472 |
| | PI | 0.8262 | 0.1300 |
| | SB | 0.8355 | 0.1145 |
| 50 | LSCV | 0.7952 | 0.1437 |
| | PI | 0.7150 | 0.1070 |
| | SB | 0.7302 | 0.0930 |
| 100 | LSCV | 0.7648 | 0.1365 |
| | PI | 0.6962 | 0.0956 |
| | SB | 0.6748 | 0.0909 |
| 500 | LSCV | 0.7606 | 0.1318 |
| | PI | 0.6586 | 0.0955 |
| | SB | 0.6616 | 0.0668 |




**ภาพที่ 4-17** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-18 และภาพที่ 4-17 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบควอติค และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุด โดยจากภาพที่ 4-18 เมื่อขนาดตัวอย่างเพิ่มขึ้น จะเห็นว่าค่า MSE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Plug-in จะมีค่าใกล้เคียงกัน




**ภาพที่ 4-18** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ SB กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-18 และภาพที่ 4-18 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบควอติค และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยที่วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MISE มากที่สุดในทุกขนาดของตัวอย่าง

## 4.4 กรณีศึกษาที่ 4

กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟโดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.4.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

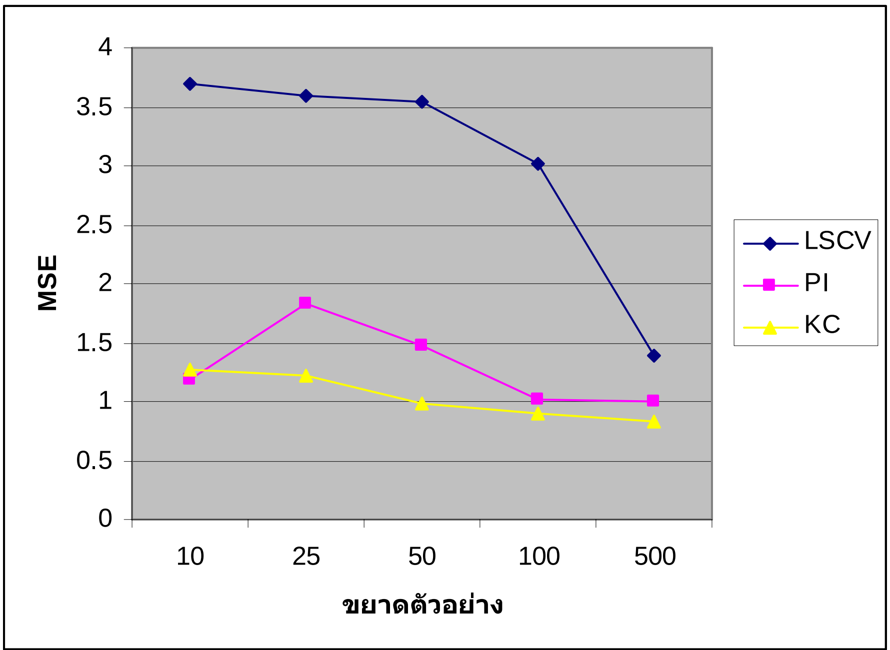
4.4.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.4.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ผลการวิจัยนำเสนอในตารางที่ 4-19 ถึง 4-21 และภาพที่ 4-19 ถึง ภาพที่ 4-24 โดยแสดงค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการเลือกแบนวิดจ์ทั้ง 3 วิธี สามารถสรุปรายละเอียด ได้ดังนี้

**ตารางที่ 4-19** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและอีพานิชนิคอฟ จำแนกตามขนาดของตัวอย่าง

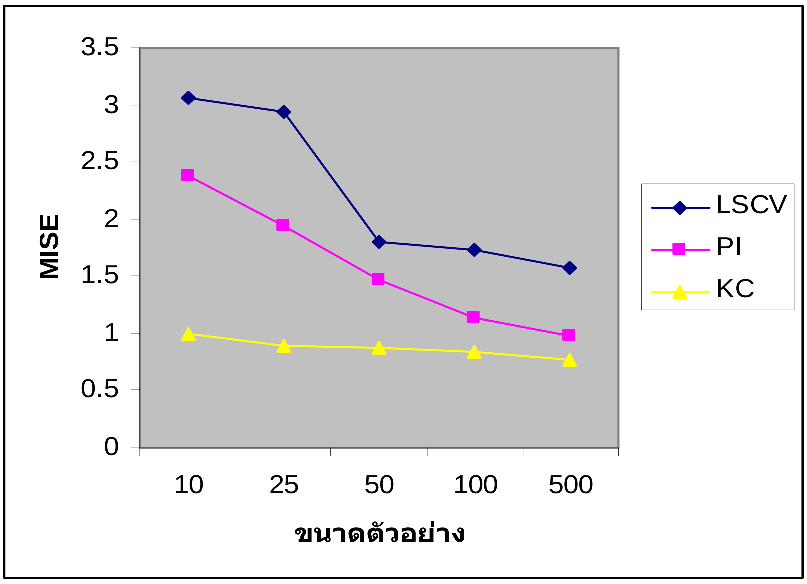
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 3.6962 | 3.0565 |
| | PI | 1.1939 | 2.3828 |
| | KC | 1.2668 | 0.9969 |
| 25 | LSCV | 3.5874 | 2.9389 |
| | PI | 1.8359 | 1.9456 |
| | KC | 1.2280 | 0.8901 |
| 50 | LSCV | 3.5358 | 1.8092 |
| | PI | 1.4829 | 1.4787 |
| | KC | 0.9751 | 0.8783 |
| 100 | LSCV | 3.0139 | 1.7263 |
| | PI | 1.0101 | 1.1361 |
| | KC | 0.9017 | 0.8484 |
| 500 | LSCV | 1.3841 | 1.5735 |
| | PI | 1.0010 | 0.9777 |
| | KC | 0.8359 | 0.7657 |




**ภาพที่ 4-19** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและอีพานิชนิคอฟ

จากตารางที่ 4-19 และภาพที่ 4-19 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug- in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุดในทุกขนาดของตัวอย่าง



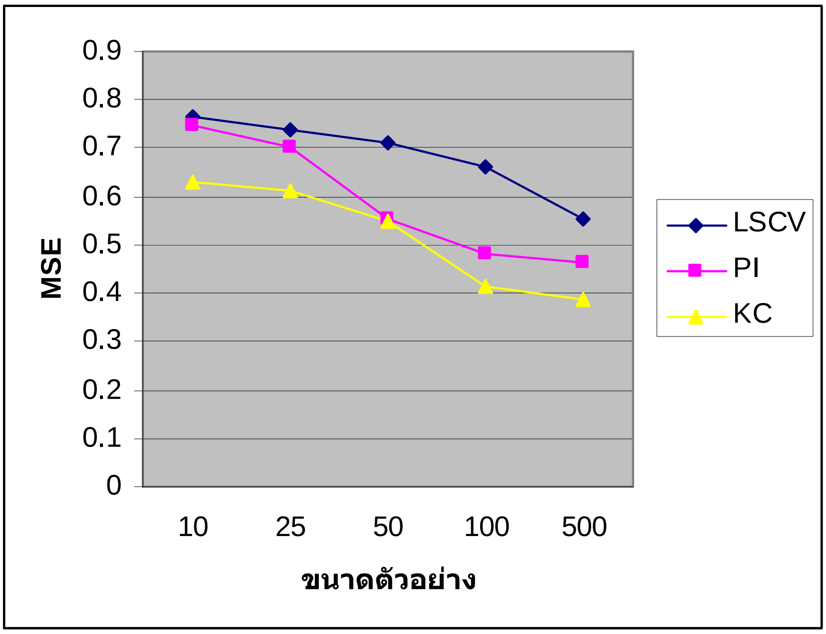

**ภาพที่ 4-20** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-19 และภาพที่ 4-20 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug- in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MISE มากที่สุดในทุกขนาดของตัวอย่าง

**ตารางที่ 4-20** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียลโดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ จำแนกตามขนาดของตัวอย่าง

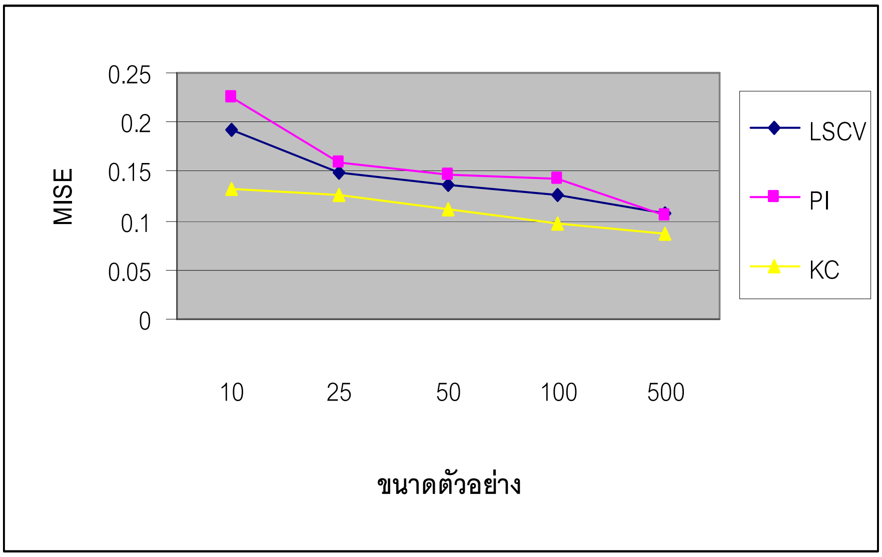
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.7651 | 0.1918 |
| | PI | 0.7489 | 0.2243 |
| | KC | 0.6305 | 0.1332 |
| 25 | LSCV | 0.7386 | 0.1489 |
| | PI | 0.7019 | 0.1589 |
| | KC | 0.6134 | 0.1257 |
| 50 | LSCV | 0.7127 | 0.1355 |
| | PI | 0.5529 | 0.1474 |
| | KC | 0.5489 | 0.1122 |
| 100 | LSCV | 0.6600 | 0.1260 |
| | PI | 0.4823 | 0.1424 |
| | KC | 0.4130 | 0.097 |
| 500 | LSCV | 0.5521 | 0.1082 |
| | PI | 0.4657 | 0.1048 |
| | KC | 0.3865 | 0.087 |




**ภาพที่ 4-21** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-20 และภาพที่ 4-21 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น ทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ให้ค่า MSE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MSE มากที่สุดในทุกขนาดของตัวอย่าง



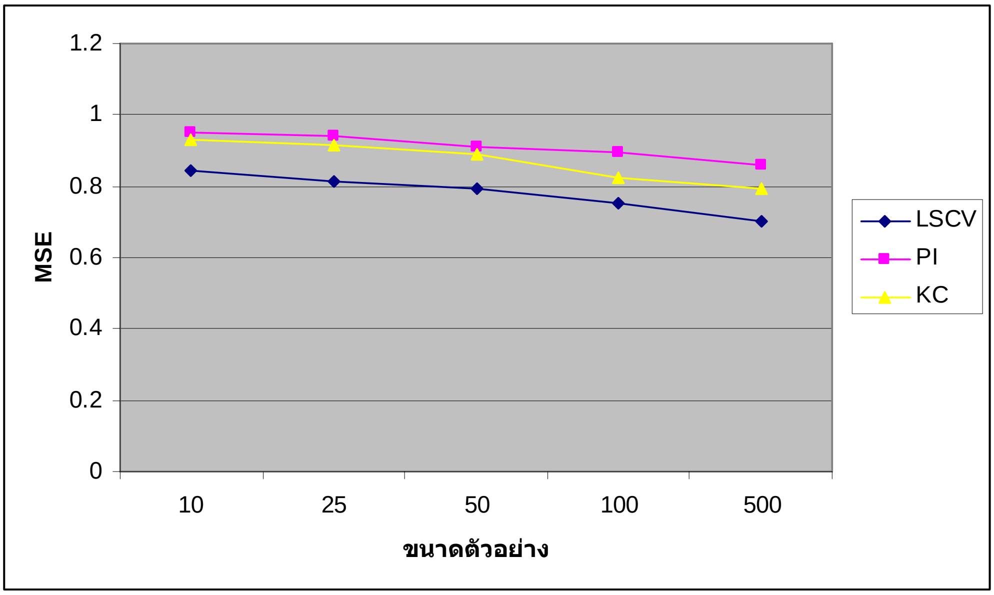

**ภาพที่ 4-22** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-20 และภาพที่ 4-22 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast กรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่า MISE ต่ำที่สุดในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MISE มากที่สุด ในทุกขนาดของตัวอย่าง

**ตารางที่ 4-21** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ จำแนกตามขนาดของตัวอย่าง

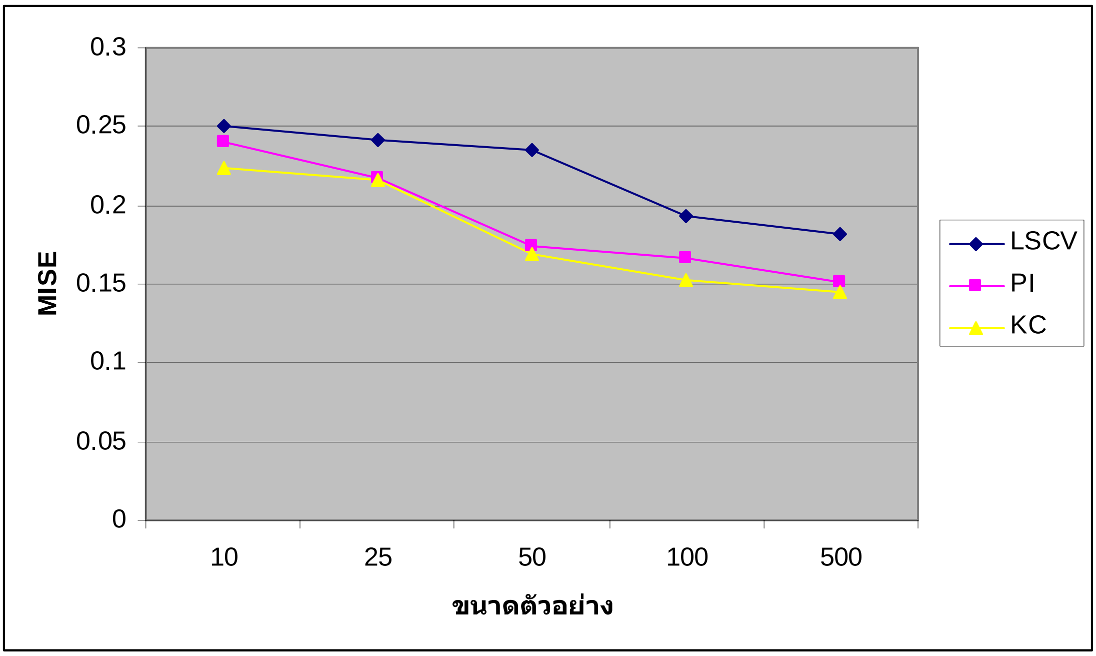
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 0.8432 | 0.2498 |
| | PI | 0.9528 | 0.2405 |
| | KC | 0.9321 | 0.2238 |
| 25 | LSCV | 0.8143 | 0.2417 |
| | PI | 0.9411 | 0.2180 |
| | KC | 0.9128 | 0.2165 |
| 50 | LSCV | 0.7929 | 0.2357 |
| | PI | 0.9772 | 0.1736 |
| | KC | 0.8921 | 0.1692 |
| 100 | LSCV | 0.7541 | 0.1926 |
| | PI | 0.8945 | 0.1662 |
| | KC | 0.8213 | 0.1520 |
| 500 | LSCV | 0.7008 | 0.1818 |
| | PI | 0.8593 | 0.1511 |
| | KC | 0.7914 | 0.1447 |




**ภาพที่ 4-23** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐานและใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-21 และภาพที่ 4-23 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, การเลือกแบนวิดจ์แบบ Plug-in และการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยค่า MSE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV มีค่าต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ส่วนวิธีการเลือกแบนวิดจ์แบบ Plug-in เป็นวิธีการเลือกแบนวิดจ์ที่มีค่ามากที่สุด ในทุกกรณีของขนาดตัวอย่าง




**ภาพที่ 4-24** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-21 และภาพที่ 4-24 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยค่า MISE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีค่าต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่มีค่ามากที่สุด ในทุกกรณีของขนาดตัวอย่าง

### 4.5 กรณีศึกษาที่ 5

กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.5.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

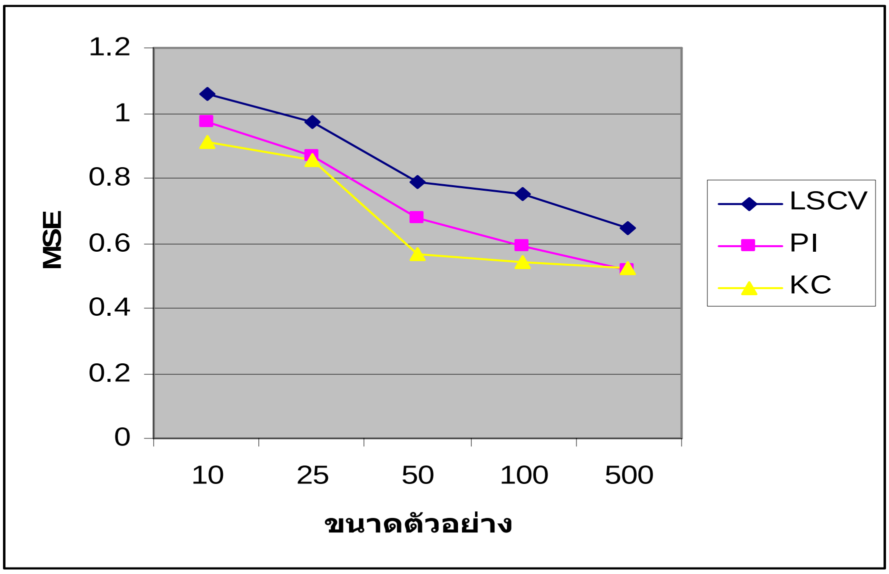
4.5.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.5.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ผลการวิจัยนำเสนอในตารางที่ 4-22 ถึง 4-24 และภาพที่ 4-25 ถึง ภาพที่ 4-30 โดยแสดงค่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการเลือกแบนวิดจ์ทั้ง 3 วิธี สามารถสรุปรายละเอียด ได้ดังนี้

**ตารางที่ 4-22** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบควอติค จำแนกตามขนาดของตัวอย่าง

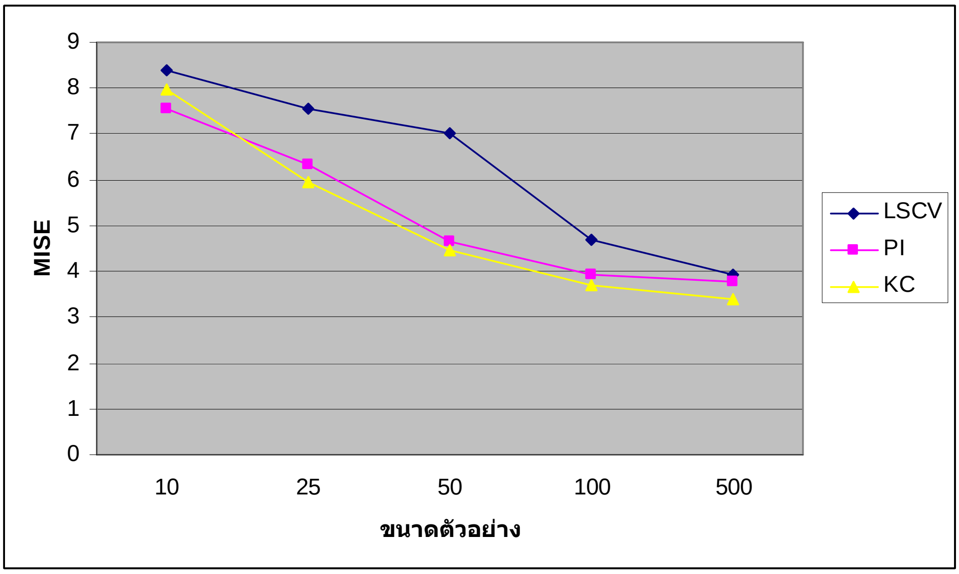
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 1.06065 | 8.38029 |
| | PI | 0.97485 | 7.5542 |
| | KC | 0.90916 | 7.35594 |
| 25 | LSCV | 0.97193 | 7.5533 |
| | PI | 0.86667 | 6.31164 |
| | KC | 0.85329 | 5.95004 |
| 50 | LSCV | 0.78821 | 7.00266 |
| | PI | 0.67907 | 4.63651 |
| | KC | 0.56647 | 4.46144 |
| 100 | LSCV | 0.75289 | 4.69229 |
| | PI | 0.59091 | 3.93716 |
| | KC | 0.53937 | 3.69452 |
| 500 | LSCV | 0.64407 | 3.94271 |
| | PI | 0.51851 | 3.78313 |
| | KC | 0.52105 | 3.40556 |




**ภาพที่ 4-25** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-22 และภาพที่ 4-25 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลง เมื่อขนาดของตัวอย่างเพิ่มขึ้น ในทุกขนาดของตัวอย่าง โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MSE ต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุด และค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ Plug-in จะใกล้เคียงกับค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ที่ขนาดตัวอย่างเท่ากับ 500



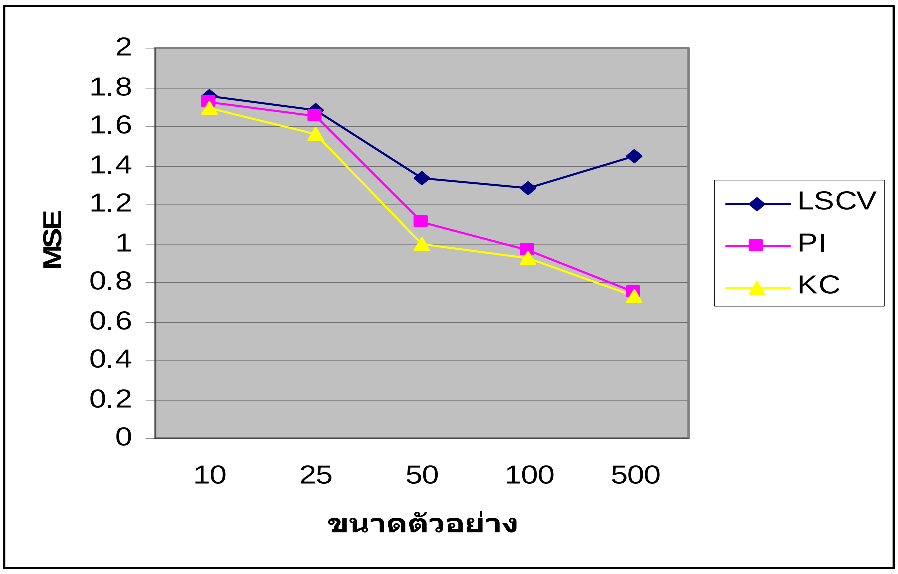

**ภาพที่ 4-26** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-22 และภาพที่ 4-26 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลง เมื่อขนาดของตัวอย่างเพิ่มขึ้น ในทุกขนาดของตัวอย่าง โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MISE ต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE มากที่สุด และค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV จะใกล้เคียงกับค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ Plug-in ที่ขนาดตัวอย่างเท่ากับ 500

**ตารางที่ 4-23** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC เมื่อข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล โดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบควอติค จำแนกตามขนาดของตัวอย่าง

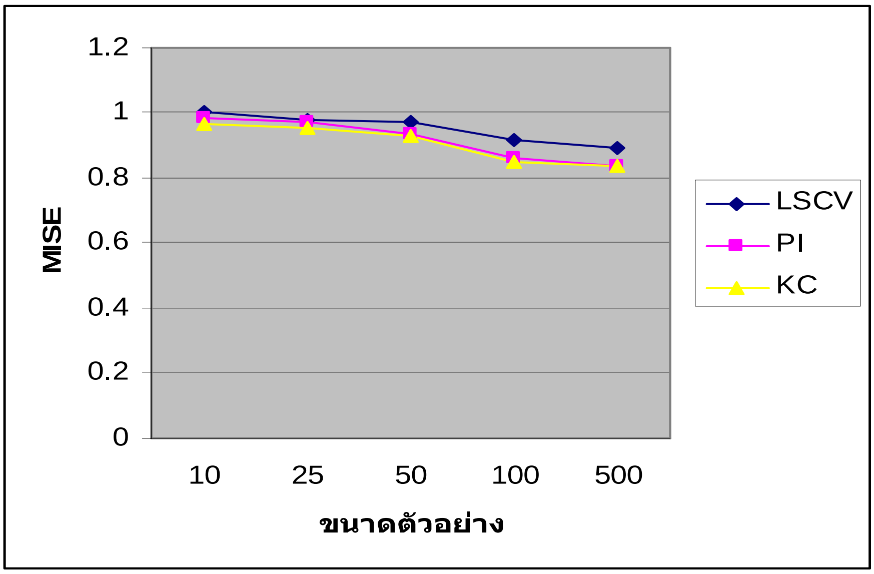
| *n* | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 1.74888 | 1.00431 |
| | PI | 1.72184 | 0.98691 |
| | KC | 1.6928 | 0.96805 |
| 25 | LSCV | 1.67924 | 0.97956 |
| | PI | 1.64878 | 0.97522 |
| | KC | 1.5582 | 0.95241 |
| 50 | LSCV | 1.33763 | 0.97115 |
| | PI | 1.11112 | 0.93771 |
| | KC | 0.9983 | 0.92808 |
| 100 | LSCV | 1.2786 | 0.91385 |
| | PI | 0.96916 | 0.86216 |
| | KC | 0.92362 | 0.84993 |
| 500 | LSCV | 1.4511 | 0.89019 |
| | PI | 0.7472 | 0.83736 |
| | KC | 0.7243 | 0.83499 |




**ภาพที่ 4-27** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-23 และภาพที่ 4-27 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MSE ต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MSE มากที่สุด ในทุกขนาดของตัวอย่าง



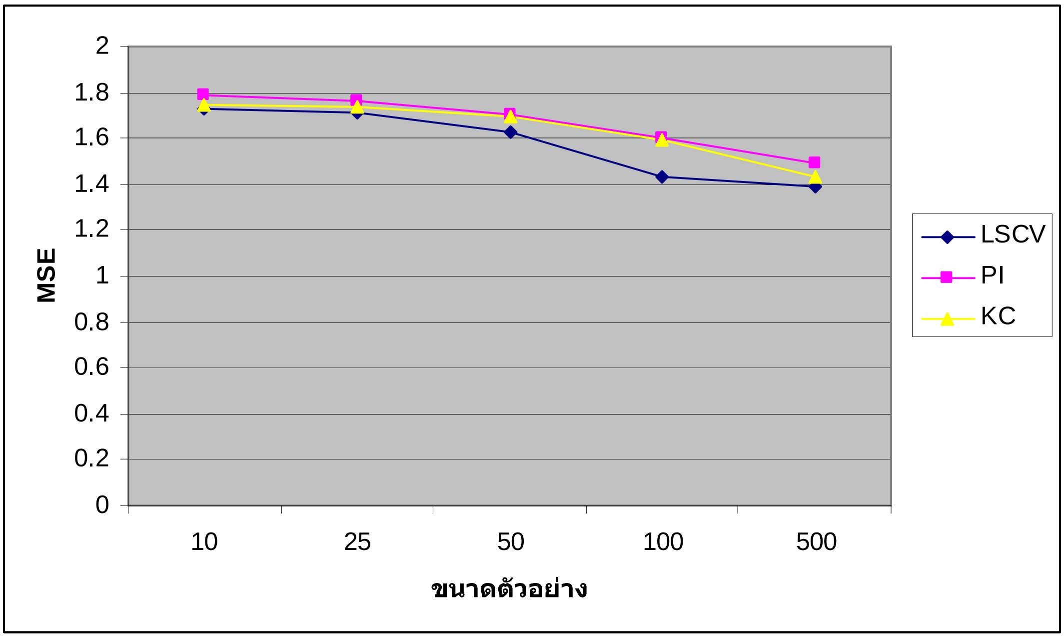

**ภาพที่ 4-28** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-23 และภาพที่ 4-28 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MISE ต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MISE มากที่สุด ในทุกขนาดของตัวอย่าง และจากภาพที่ 4-28 จะเห็นว่าค่า MISE ที่ได้จากการประมาณค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ Plug-in และ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีความใกล้เคียงกันในทุกขนาดของตัวอย่าง

**ตารางที่ 4-24** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบควอติค จำแนกตามขนาดของตัวอย่าง

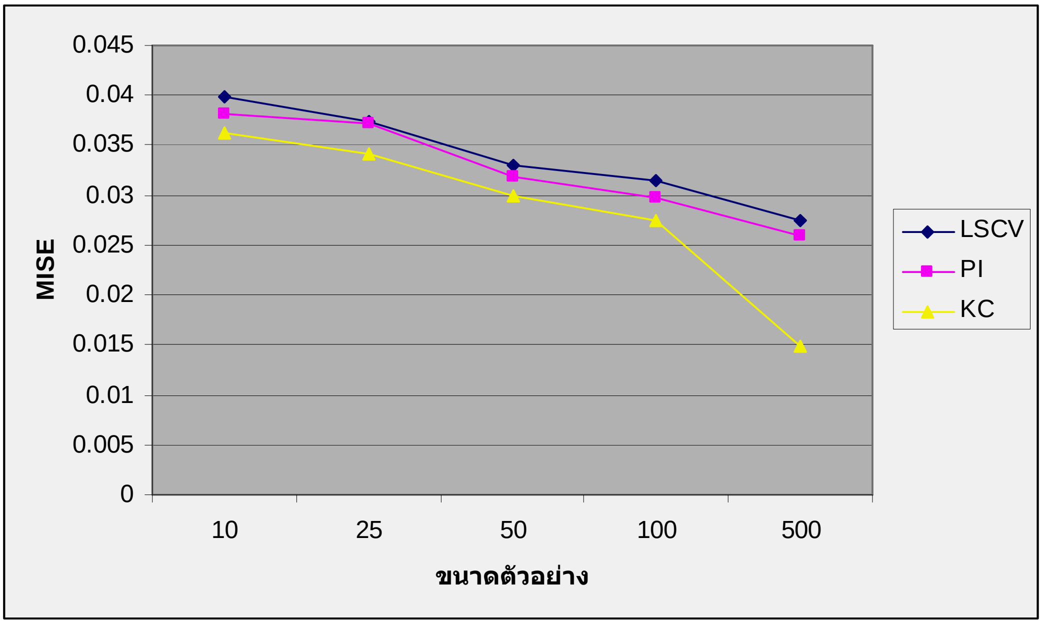
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 1.7327 | 0.0398 |
| | PI | 1.7897 | 0.0381 |
| | KC | 1.7443 | 0.0362 |
| 25 | LSCV | 1.7121 | 0.0374 |
| | PI | 1.7645 | 0.0371 |
| | KC | 1.7397 | 0.0342 |
| 50 | LSCV | 1.6238 | 0.0330 |
| | PI | 1.7029 | 0.0319 |
| | KC | 1.6948 | 0.0300 |
| 100 | LSCV | 1.4318 | 0.0314 |
| | PI | 1.6004 | 0.0298 |
| | KC | 1.5958 | 0.0275 |
| 500 | LSCV | 1.3924 | 0.0274 |
| | PI | 1.4932 | 0.0259 |
| | KC | 1.4324 | 0.0148 |




**ภาพที่ 4-29** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-24 และภาพที่ 4-29 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น โดยค่า MSE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV มีค่าต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ส่วนวิธีการเลือกแบนวิดจ์แบบ Plug-in เป็นวิธีการเลือกแบนวิดจ์ที่มีค่ามากที่สุด ในทุกกรณีของขนาดตัวอย่าง




**ภาพที่ 4-30** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน และฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

จากตารางที่ 4-24 และภาพที่ 4-30 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MISE ต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MISE มากที่สุด ในทุกขนาดของตัวอย่าง

**4.6 กรณีศึกษาที่ 6**

กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และ $K_2$ คือ ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.6.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

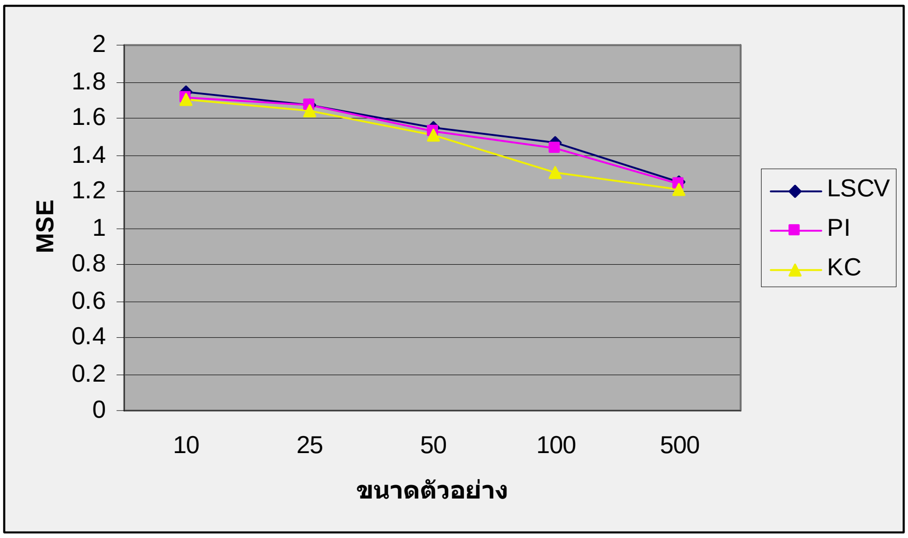
4.6.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.6.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

ผลการวิจัยนำเสนอในตารางที่ 4-25 ถึง 4-27 โดยแสดงค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการเลือกแบนวิดจ์ทั้ง 3 วิธี สามารถสรุปรายละเอียด ได้ดังนี้

**ตารางที่ 4-25** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบควอติค จำแนกตามขนาดของตัวอย่าง

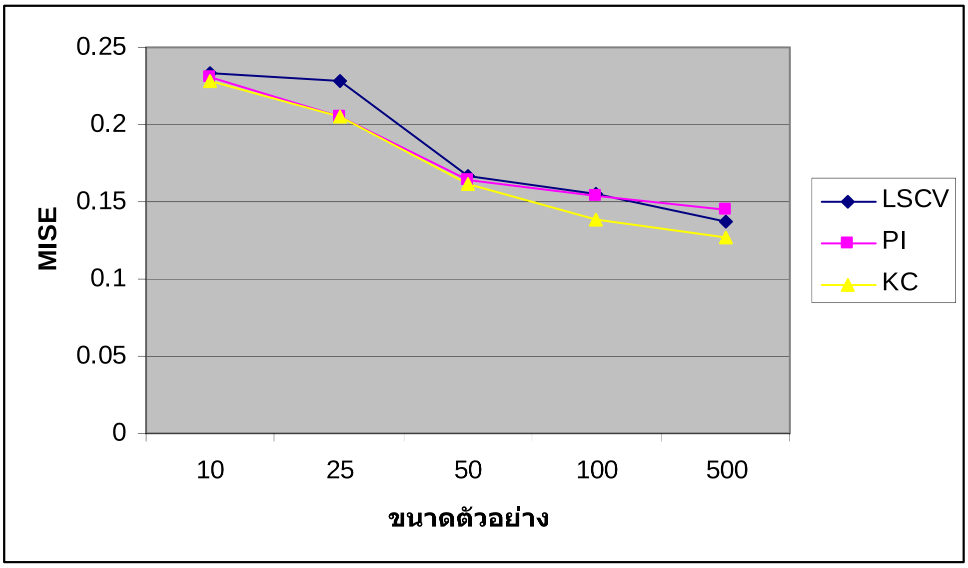
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 1.74778 | 0.23322 |
| | PI | 1.7175 | 0.23103 |
| | KC | 1.70732 | 0.22871 |
| 25 | LSCV | 1.67464 | 0.2278 |
| | PI | 1.66868 | 0.20567 |
| | KC | 1.64576 | 0.20449 |
| 50 | LSCV | 1.55093 | 0.16663 |
| | PI | 1.52813 | 0.16384 |
| | KC | 1.51098 | 0.16163 |
| 100 | LSCV | 1.46722 | 0.15565 |
| | PI | 1.43304 | 0.15361 |
| | KC | 1.30496 | 0.13853 |
| 500 | LSCV | 1.25298 | 0.13751 |
| | PI | 1.23744 | 0.14425 |
| | KC | 1.2096 | 0.1273 |




**ภาพที่ 4-31** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-25 และภาพที่ 4-31 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MSE ต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MSE มากที่สุด ในทุกขนาดของตัวอย่างโดยจากภาพที่ 4-31 จะเห็นว่า ค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีค่าใกล้เคียงกัน เมื่อขนาดตัวอย่างเท่ากับ 10, 25, 50 และ 500 ส่วนในกรณีที่ขนาดตัวอย่างเท่ากับ 100 จะเห็นว่า ค่า MSE ที่ได้ของวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in จะใกล้เคียงกัน



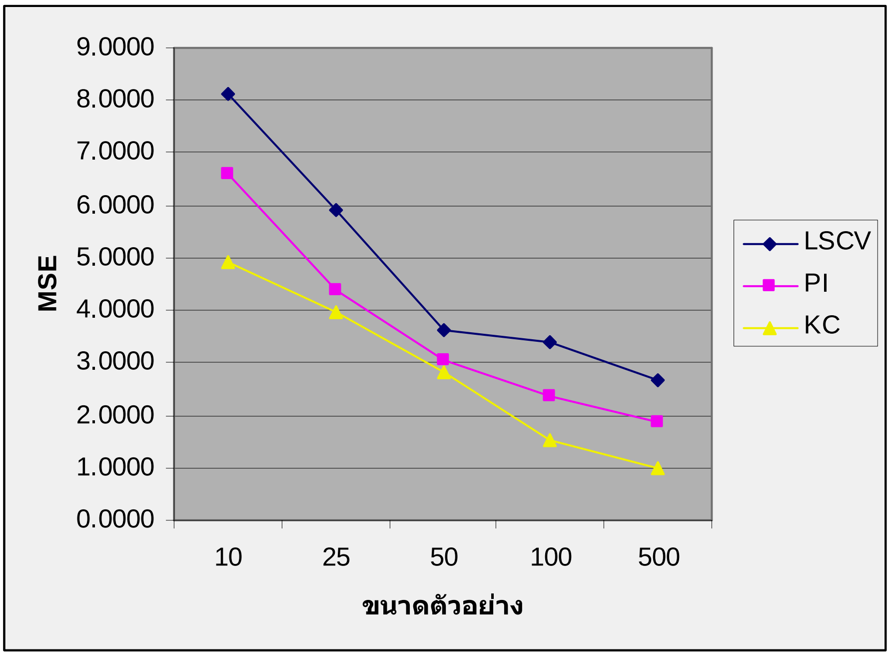

**ภาพที่ 4-32** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-25 และภาพที่ 4-32 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบปกติมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธี โดยวิธีการเลือกแบนวิดจ์แบบ Kernel contrat จะให้ค่า MSE ต่ำที่สุด ในทุกขนาดของตัวอย่าง รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีที่ให้ค่า MISE มากที่สุด ในทุกขนาดของตัวอย่าง ยกเว้นกรณีที่ $n = 500$ จากภาพที่ 4-32 จะเห็นว่า ค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in มีค่าใกล้เคียงกันในทุกขนาดของตัวอย่าง

**ตารางที่ 4-26** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบควอติค จำแนกตามขนาดของตัวอย่าง

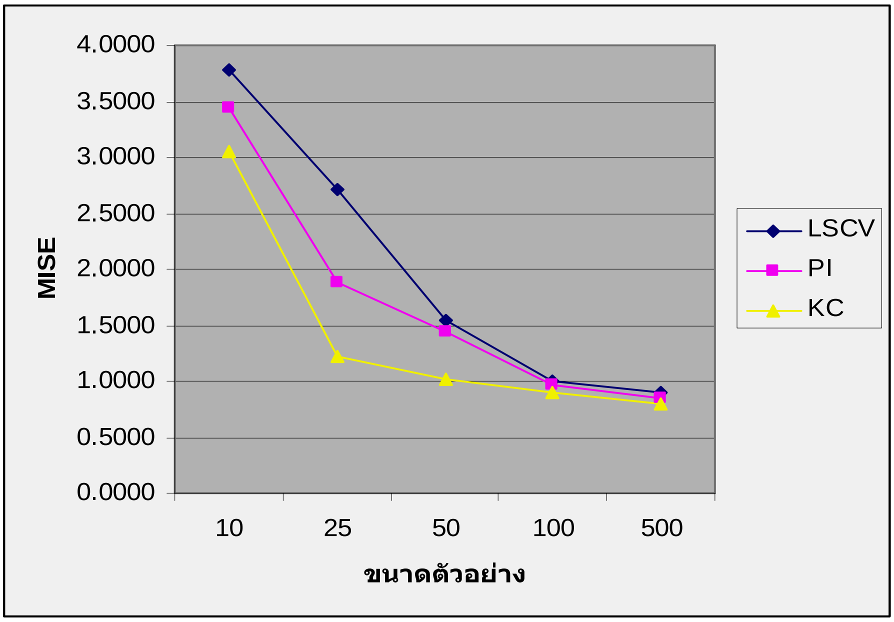
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 8.1176 | 3.7762 |
| | PI | 6.6112 | 3.4362 |
| | KC | 4.9374 | 3.0562 |
| 25 | LSCV | 5.9147 | 2.7076 |
| | PI | 4.3815 | 1.8762 |
| | KC | 3.9490 | 1.2269 |
| 50 | LSCV | 3.6332 | 1.5464 |
| | PI | 3.0583 | 1.4433 |
| | KC | 2.8302 | 1.0207 |
| 100 | LSCV | 3.3975 | 0.9984 |
| | PI | 2.3716 | 0.9728 |
| | KC | 1.5369 | 0.8957 |
| 500 | LSCV | 2.6734 | 0.8924 |
| | PI | 1.8646 | 0.8516 |
| | KC | 0.9891 | 0.7985 |




**ภาพที่ 4-33** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-40 และภาพที่ 4-33 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้นในทุกขนาดของตัวอย่าง ซึ่งวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE ต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE สูงที่สุด



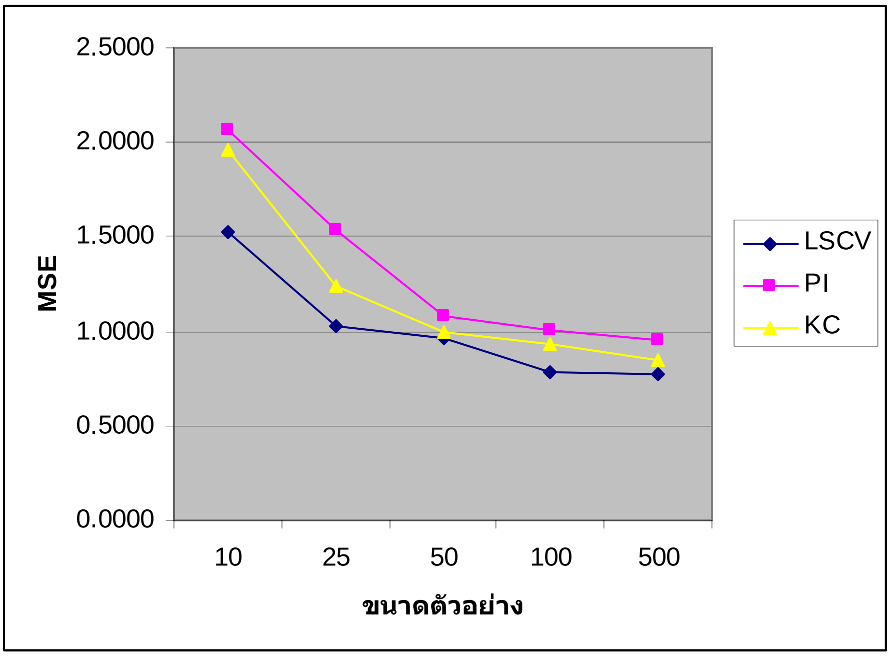

**ภาพที่ 4-34** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-41 และภาพที่ 4-34 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล สามารถสรุปผลได้ดังนี้

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้นในทุกขนาดของตัวอย่าง ซึ่งวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MISE ต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MISE สูงที่สุด จากภาพที่ 4-34 เห็นว่าเมื่อขนาดของตัวอย่างมีขนาดใหญ่ขึ้น ค่า MISE ของทั้ง 3 วิธีการเลือกแบนวิดจ์จะมีค่าใกล้เคียงกัน

**ตารางที่ 4-27** แสดงการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC เมื่อข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟและฟังก์ชันเคอร์เนลแบบควอติค จำแนกตามขนาดของตัวอย่าง

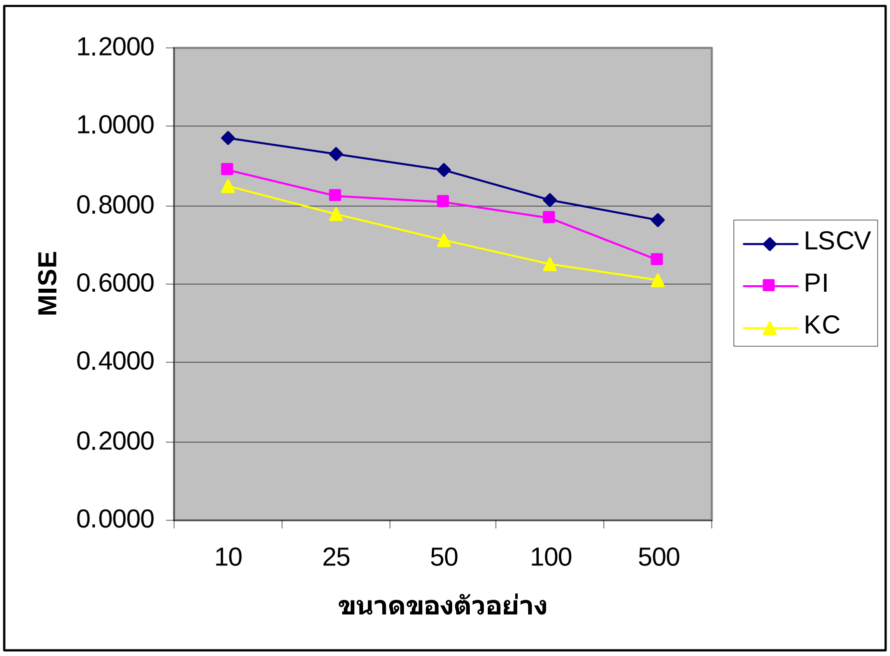
| $n$ | **วิธีการเลือกแบนวิดจ์** | **MSE** | **MISE** |
|---|---|---|---|
| 10 | LSCV | 1.5296 | 0.9701 |
| | PI | 2.0633 | 0.8876 |
| | KC | 1.9646 | 0.8516 |
| 25 | LSCV | 1.0229 | 0.9322 |
| | PI | 1.5397 | 0.8213 |
| | KC | 1.2417 | 0.7783 |
| 50 | LSCV | 0.9639 | 0.8918 |
| | PI | 1.0753 | 0.8107 |
| | KC | 0.9992 | 0.7117 |
| 100 | LSCV | 0.7827 | 0.8127 |
| | PI | 1.0021 | 0.7662 |
| | KC | 0.9296 | 0.6489 |
| 500 | LSCV | 0.7720 | 0.8648 |
| | PI | 0.9539 | 0.6594 |
| | KC | 0.8501 | 0.6108 |




**ภาพที่ 4-35** แสดงการเปรียบเทียบค่า MSE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน และใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-42 และภาพที่ 4-35 แสดงค่า MSE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้นในทุกขนาดของตัวอย่าง ซึ่งวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE ต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE สูงที่สุด




**ภาพที่ 4-36** แสดงการเปรียบเทียบค่า MISE ของวิธีการเลือกแบนวิดจ์แบบ LSCV, PI และ KC กรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบควอติค

จากตารางที่ 4-43 และภาพที่ 4-36 แสดงค่า MISE ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ทั้ง 3 วิธี ในกรณีที่ใช้ฟังก์ชันเคอร์เนลลแบบเกาส์เซียนและฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน สามารถสรุปผลได้ดังนี้

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) มีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้นในทุกขนาดของตัวอย่าง ซึ่งวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MISE ต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in ส่วนวิธีการเลือกแบนวิดจ์แบบ LSCV เป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MISE สูงที่สุด

# บทที่ 5
# สรุปผลการวิจัยและข้อเสนอแนะ

ในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์ โดยศึกษาวิธีการเลือกแบนวิดจ์ 4 วิธี คือ วิธีการเลือกแบนวิดจ์แบบ แบบ Least Square Cross-Validation วิธีเลือกแบนวิดจ์แบบ Plug-in วิธีเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีเลือกแบนวิดจ์แบบ Kernel Contrast ศึกษาในกรณีที่ข้อมูลมีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบเอกซ์โพเนนเชียล และการแจกแจงแบบโคชี โดยศึกษาฟังก์ชันเคอร์เนล 3 ฟังก์ชัน คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และ ฟังก์ชันเคอร์เนลแบบควอติค โดยทำการศึกษาในสถานการณ์ ต่างๆ ดังต่อไปนี้

1. กรณีศึกษาที่ 1 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

   1.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

   1.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

   1.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

2. กรณีศึกษาที่ 2 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

   2.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

   2.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

   2.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

3. กรณีศึกษาที่ 3 กำหนดให้ $K$ คือ ฟังก์ชันเคอร์เนลแบบควอติคโดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

   3.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

   3.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

   3.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

4. กรณีศึกษาที่ 4 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

4.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

4.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

4.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

5. กรณีศึกษาที่ 5 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และ $K_2$ คือ ฟังก์ชันเคอร์ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

5.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

5.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

5.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

6. กรณีศึกษาที่ 6 กำหนดให้ $K_1$ คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และ $K_2$ คือ ฟังก์ชันเคอร์ฟังก์ชันเคอร์เนลแบบควอติค โดยมีขนาดตัวอย่าง ($n$) เท่ากับ 10, 25, 50, 100 และ 500 โดยสร้างข้อมูลให้มีการแจกแจงดังนี้

6.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

6.2 กำหนดให้ $X_1 \sim Exp(1)$ และ $X_2 \sim Exp(1)$

6.3 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดจ์ในแต่ละสถานการณ์ คือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และ ค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ยรวม (MISE) การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง ดำเนินการโดยการสร้างข้อมูล (Generated Data) ที่มีเงื่อนไขตามที่ต้องการและวิเคราะห์ผลด้วยโปรแกรม R โดยในแต่ละสถานการณ์ผู้วิจัยจะทำการทดลองซ้ำๆ กันเป็นจำนวน 1,000 รอบ เพื่อทำการประมาณค่าแบนวิดจ์ตามที่ต้องการ และทำการเปรียบเทียบค่า MSE และ MISE

**5.1 สรุปผลการวิจัย**

ผลสรุปจากการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ของการประมาณค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์ 4 วิธีการภายใต้สถานการณ์ต่างๆ ซึ่งจำแนกตามกรณีศึกษาที่กำหนดไว้ ดังนี้

5.1.1 กรณีศึกษาที่ 1

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MSE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MISE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

5.1.2 กรณีศึกษาที่ 2

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MSE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MISE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

5.1.3 กรณีศึกษาที่ 3

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MSE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MISE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

5.1.4 กรณีศึกษาที่ 4

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MSE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

ในกรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ค่า MSE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE ต่ำที่สุด และมีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ส่วนวิธีการเลือกแบนวิดจ์แบบ Plug-in จะเป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE มากที่สุด

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MISE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MISE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

5.1.5 กรณีศึกษาที่ 5

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MSE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

ในกรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ค่า MSE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE ต่ำที่สุด และมีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ส่วนวิธีการเลือกแบนวิดจ์แบบ Plug-in จะเป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE มากที่สุด

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการ

เลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MISE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

5.1.6 กรณีศึกษาที่ 6

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (MSE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MSE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

ในกรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน ค่า MSE ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ LSCV จะให้ค่า MSE ต่ำที่สุด และมีแนวโน้มลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น รองลงมาคือวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ส่วนวิธีการเลือกแบนวิดจ์แบบ Plug-in จะเป็นวิธีการเลือกแบนวิดจ์ที่ให้ค่า MSE มากที่สุด

เมื่อเปรียบเทียบประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ในการประมาณค่าแบนวิดจ์ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap มีค่า MSE ต่ำที่สุด และต่ำกว่าวิธีการเลือกแบนวิดจ์แบบ LSCV และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ในทุกๆ ขนาดตัวอย่าง และทั้ง 3 วิธีการเลือกแบนวิดจ์จะให้ค่า MISE ที่มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

## 5.2 การอภิปรายผล

5.2.1 สรุปผลโดยรวมในทุกระดับขนาดตัวอย่างที่กำหนดในแผนการทดลองได้ว่า วิธีการเลือกแบนวิดจ์ที่เหมาะสมที่สุดในการประมาณค่าแบนวิดจ์ คือ วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Plug-in และวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation ตามลำดับ

5.2.2 ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation ในกรณีศึกษาที่4 ถึงกรณีศึกษาที่ 6 ในกรณีที่ข้อมูลมีการแจกแจงแบบโคชีมาตรฐาน จะมีค่าต่ำที่สุด รองลงมาคือ วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast และวิธีการเลือกแบนวิดจ์แบบ Plug-in ตามลำดับ ซึ่งสอดคล้องกับงานวิจัยของ Mugdadi และ Ahman (2004) โดยเมื่อขนาดตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) จะลดลงในทุกๆ กรณี

5.2.3 ค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะเลือกค่าแบนวิดจ์ที่มีขนาดเล็กกว่าวิธีการเลือกแบนวิดจ์อื่นๆ

**5.3 ข้อเสนอแนะ**

5.3.1 การนำไปใช้ประโยชน์

การวิจัยครั้งนี้สามารถนำไปประยุกต์ใช้กับข้อมูลที่มีรูปแบบของฟังก์ชันของตัวแปรสุ่มที่คล้ายกับงานวิจัยในครั้งนี้ ในส่วนของการประมาณค่าแบนวิดจ์นั้น ผู้วิจัยขอเสนอวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ทั้งนี้เพราะว่าทั้ง 2 วิธีการนั้นทำให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนเฉลี่ยรวม (MISE) ต่ำที่สุด ซึ่งค่า MSE และ MISE นั้นเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าแบนวิดจ์ในครั้งนี้

5.3.2 การศึกษาวิจัย

1. ในการวิจัยครั้งนี้ได้ทำการศึกษาวิธีการเลือกแบนวิดจ์ด้วยวิธีการประมาณ 3 วิธีการ คือ

วิธีการเลือกแบนวิดจ์แบบ LSCV วิธีการเลือกแบนวิดจ์แบบ Plug-in วิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap และวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ซึ่งในการศึกษาครั้งต่อไปควรศึกษาวิธีอื่นเพิ่มเติ่ม

2. ในการวิจัยครั้งนี้ได้ทำการศึกษาข้อมูลที่มีการแจกแจงแบบปกติมาตรฐาน การแจกแจงแบบโคชีมาตรฐาน และการแจกแจงแบบเอกซโพเนนเชียล ในการศึกษาครั้งต่อไปควรศึกษาการแจกแจงของข้อมูลแบบอื่นเพิ่มเติม

3. ในการวิจัยครั้งนี้ได้ทำการศึกษาฟังก์ชันเคอร์เนล 3 ฟังก์ชัน คือ ฟังก์ชันเกาส์เซียน ฟังก์ชันอีพานิชนิคอฟ และฟังก์ชันควอติค ในการศึกษาครั้งต่อไปควรศึกษาฟังก์ชันเคอร์เนลแบบอื่นเพิ่มเติม

# บรรณานุกรม

## ภาษาไทย


ทองคำ ไม้กลัด. **สถิติเชิงอนุมานเบื้องต้น.** พิมพ์ครั้ง 1: ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2542.

ธีระพร วีระถาวร. **ความน่าจะเป็นเบื้องต้น**: ทฤษฎีและการประยุกต์ใช้. พิมพ์ครั้งที่ 3. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2542.

มานพ วราภักดิ์. **การจำลองเบื้องต้น.** กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย, 2547.

______. **ทฤษฎีความน่าจะเป็น.** กรุงเทพมหานคร : ศูนย์ผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2545.

วินัย โพธิ์สุวรรณ. **การเปรียบเทียบตัวสถิติสำหรับทดสอบความเท่ากันของความแปรปรวน**. วิทยานิพนธ์สถิติศาสตรมหาบัณฑิต. สาขาวิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2534.

วิจิตรา พลเยี่ยม. **ทฤษฎีความน่าจะเป็น.** กรุงเทพมหานคร: งานผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2541.


## ภาษาอังกฤษ


Ahmad, I.A and Mugdadi, A.R. "Analysis of Kernel Density Estimation of Function of Random Variable." **Journal of Nonparametric Statistics**. 15(2003b): 579-605.

Altman, N. and Leger, C. "Bandwidth Selection for kernel distribution function estimation." **Journal of Statistical Planning and Inference**. 46(1995): 195-214.

Bowman, A. W. "An alternative method of cross-validation for smoothing of density estimates." **Biometrika**. 71(1984): 353-360.

Faraway, J.J and Jhun, M "Bootstrap Choice of Bandwidth for Density Estimation." **Journal of the American Statistical Association**. 85(1990): 1119-1122.

Hall, P., Simon J., Sheather, S.J., Jones, M.C and Marron, J.S. "On optimal data-based bandwidth selection kernel density estimation." **Biometrika**. 78(1991): 263-269.

Hansen, B.E. “Bandwidth Selection for Nonparametric Distribution Estimation.” **University of Wisconsin**, Madison. May (2004).

Jones, M.C, Marron, J.S and Sheather, S.J. “A Brief Survey of Bandwidth Selection for Density Estimation.” **Journal of the American Statistical Association**. 91(1996): 401-407.

Mugdadi, A.R. and Ahmad, I.A.”A bandwidth selection for kernel density Estimation of functions of random variables.” **Computation Statistics and Data Analysis**. 47(2004): 49-62.

Shean-Tsong. “A Comparative Review of Bandwidth Selection for Kernel Density Estimation.” **Statistica Sinica**. 6(1996): 129-145.

Sheather, S.J. and Jones, M.C. “A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation” **Journal of Royal Statistical Society B**. 53(1991): 683-690.

Silverman, B.W. **Density Estimation for Statistics and Data Analysis**. New York: Chapman and Hall, 1986.

Taylor, C.C. “Bootstrap choice of the smoothing parameter in kernel density estimation.” **Biometrika**. 76(1989): 705-712.

Turlach, B.A. “Bandwidth Selection in Kernel Density Estimation: A Review.” **www.citeceer.com**

Wolfgang Hardle. **Smoothing Techniques with Implementation in S.** New York: Spinger-Verleg, 1990.

ภาคผนวก ก

# โปรแกรมที่ใช้ในการวิจัย

1. ตัวอย่างโปรแกรม R

โปรแกรมสำหรับจำลองข้อมูลให้มีลักษณะตามสถานการณ์ต่างๆ ที่กำหนด การคำนวณค่าแบนวิดจ์ การคำนวณค่า MSE และการคำนวณค่า MISE

```
***************************************************************
Data Generation from Standard Normal Distribution and using
Gaussian Kernel function.
***************************************************************

library (stats)
library (MASS)
library (KernSmooth)

data<-function (n)
{
      X           <- rep(0,n)
      lower       <- rep(0,n)
      upper       <- rep(0,n)
      y           <- rep(0,n)
      ans         <- rep(0,n)
      gauss.kernel <- rep(0,n)

      x1 <- rnorm(n)
      x2 <- rnorm(n)
      x  <- x1 + x2
      hmax <- 1.144*sqrt(var(x))*n^(-1/5)

for(i in 1:n)
{

      X[i]  <- x[i]
      lower[i]   <- X[i] - hmax
      upper[i] <- X[i] + hmax

      for(j in 1:n)
      {
            if((x[j] > lower[i])||(x[j]<= upper[i]))

            y[j] <- (X[i]-x[j])/hmax

            gauss.kernel[j] <- exp(-(y[j]^2)/2))/(sqrt(2)*pi)

      }

            ans[i] <- sum(gauss.kernel)/(n*hmax)
}

      return(ans)

}
```

```
***************************************************************
Least Square Cross-Validation Bandwidth Selection Method
***************************************************************

LSCV <- bw.ucv(ans, nb = 1000, lower=0.1*hmax, upper = hmax)


***************************************************************
Shealter and Jones Plug-in Bandwidth selection method
***************************************************************

Plugin <-  width.SJ(ans ,nb = 1000,lower=0.1*hmax, upper=hmax,
           method = c("ste", "dpi"))


***************************************************************
Smoothed Bootstrap Bandwidth selection method
***************************************************************
h.window <- function(x, sample.size, window=1)
{
      if(window==0){
            wind.weight<-0}

      if(window==1){
            stdev<-var(x)^0.5
            wind.weight <-(1.06)*stdev*n^ (-1/5)}

      if(window==2){
            quar<-diff(quantile(x,c(0.25,0.75)))
            wind.weight <-(0.79)*quar*n^ (-1/5)}

      if(window==3){
            stdev<-var(x)^0.5
            skew <-sum((x-mean(x))^3)/n
wind.weight <-(((abs(skew)*(1.64^2+2))/(3*stdev*1.64))^0.5)^
               (-1/4)}

      if(window==4){
            wind.weight<-n^(-1/2)}

      if(window==5){
            stdev<-var(x)^0.5
            wind.weight <-(n)^(-3/4)*(stdev)*(2^(1/2))}

      if(window==6){
            stdev<-var(x)^0.5
            wind.weight <-(1.41)*stdev*n^(-3/4)}

      if(window==7){
            stdev<-var(x)^0.5
            quar<-diff(quantile(x,c(0.25,0.75)))
            wind.weight <-min(stdev,(0.90)*(quar/1.34)
```

```
                         *n^(-1/5))}
return(wind.weight)
}

density2 <- function(x, n, num.boot.samp,window)
{
     x.length<- n*num.boot.samp
     x.unif <-  matrix(sample(x, size = x.length,
                replace=T),nrow = num.boot.samp)
     x.h       <- apply(x.unif,1,h.window,n,window)
     x.h       <- matrix(x.h,ncol= n,nrow =num.boot.samp,byrow
                  = F)
     e    <-    matrix(rnorm(x.length),nrow=num.boot.samp)
     y    <-    x.unif +x.h *e
     return(y)
}

**************************************************************
Kernel Contrast Bandwidth selection method
**************************************************************

kc<- function(x,nb =1000, lower=0.1*hmax,upper = hmax)
{

     if ((n <- length(x)) < 2)
     stop("need at least 2 data points")
     if (!is.numeric(x))
          stop("invalid 'x'")

     n    <- length(x)
     p1    <- -1
     p2    <-  1
     hmax <- 1.144 * sqrt(var(x)) * n^(-1/5)
     gaussian.kernel <- (1/sqrt(2*pi))*exp(-1/2)*(x^2)
     quartic.kernel  <-  0.9375*((1-x^2)^2)
     M  <- gaussian.kernel
     O  <- quartic.kernel
     a1 <- (x-M)/hmax
     a2 <- (x-O)/hmax
     b1 <- sum(a1)* (1/(n*hmax))
     b2 <- sum(a2)* (1/(n*hmax))
     contrast <- (p1*b1)+(p2*b2)
     h <- sum(contrast)/n
     if (h < 1.1 * lower | h > upper - 0.1 * lower)
warning("minimum occurred at one end of the range")
     h
}
```

```
************************************************************
Run Program
************************************************************
case1.cauchy<-function (nrun,n)
{
library(stats)
library(MASS)
library(KernSmooth)

      x          <- numeric(n)
      x1         <- numeric(n)
      x2         <- numeric(n)
      XI           <- numeric(n)
      LSCV      <- numeric(n)
      plugin    <- numeric(n)
      contrast  <- numeric(n)
      ftrue     <- numeric(n)
      mixed1.kernel    <- numeric(n)
      kernel.LSCV      <- numeric(n)
      sum.LSCV.kernel <- numeric(n)
      fhat.LSCV        <- numeric(n)
      mse.LSCV   <- numeric(n)
      bias.mse.LSCV    <- numeric(n)
      var.mse.LSCV     <- numeric(n)
      bias.mise.LSCV   <- numeric(n)
      var.mise.LSCV    <- numeric(n)
      mise.LSCV        <- numeric(n)
      kernel.plugin   <- numeric(n)
      sum.plugin.kernel <- numeric(n)
      fhat.plugin      <- numeric(n)
      mse.plugin       <- numeric(n)
      bias.mse.plugin <- numeric(n)
      var.mse.plugin   <- numeric(n)
      bias.mise.plugin <- numeric(n)
      var.mise.plugin       <- numeric(n)
      mise.plugin           <- numeric(n)
      kernel.contrast       <- numeric(n)
      sum.contrast.kernel      <- numeric(n)
      fhat.contrast         <- numeric(n)
      mse.contrast          <- numeric(n)
      bias.mse.contrast     <- numeric(n)
      var.mse.contrast <- numeric(n)
      bias.mise.contrast    <- numeric(n)
      var.mise.contrast     <- numeric(n)
      mise.contrast         <- numeric(n)

      for (i in 1:nrun)
      {
            x1 <-rexp(n,1)
            x2 <-rexp(n,1)
            x  <- x1+x2

            sd<-sqrt(var(x))
```

```
            q1<-quantile(x, 0.25)
            q2<-quantile(x, 0.75)
            r<-(q2-q1)/1.34
            A<-min(sd,r)
            min<-0.1*0.9*A*n^(-0.2)
            max<-0.9*A*n^(-0.2)

            epanechnikov.kernel <- 0.75*(1-x^2)*(abs(x)<=1)
            XI <- epanechnikov.kernel
            LSCV   <- bw.ucv(x, nb =1000, lower = min, upper =
                       max)
            plugin <- width.SJ(x, nb = 1000, lower = min, upper
                       = max, method = "dpi")

            ftrue <- dgamma(x,2,rate=1)

            K.LSCV      <- (x-XI)/LSCV
            K.sum       <- sum(K.LSCV)
            fhat.LSCV  <- (1/(n*LSCV))*K.sum
            mse.LSCV    <- sum((ftrue-fhat.LSCV)^2)/n
            bias.LSCV   <- sum(ftrue-fhat.LSCV)/n

            K.plugin      <- (x-XI)/plugin
            K.sum.plugin  <- sum(K.plugin)
            fhat.plugin   <- (1/(n*plugin))*K.sum.plugin
            mse.plugin    <- sum((ftrue-fhat.plugin)^2)/n
            bias.plugin   <- sum(ftrue-fhat.plugin)/n


}

mse.bar.LSCV     <- sum(mse.LSCV)/nrun
mse.bar.plugin   <- sum(mse.plugin)/nrun
bias.bar.LSCV   <- sum(bias.LSCV)/nrun
bias.bar.plugin <- sum(bias.plugin)/nrun

out <-
cbind(LSCV,plugin,mse.bar.LSCV,mse.bar.plugin,bias.bar.LSCV,bi
as.bar.plugin)
out
}
```

2. ฟังก์ชัน Built In ในโปรแกรม R

ฟังก์ชันที่ใช้การคำนวณค่าแบนวิดจ์จากวิธี LSCV และ PI ที่ปรากฏอยู่ในโปรแกรมตัวอย่างที่ใช้ในการวิจัย

วิธีการเลือกแบนวิดจ์แบบ Least square Cross-Validation

```
bw.ucv
function(x,nb = 1000,lower =0.1*hmax, upper = hmax)
{
    if ((n <- length(x)) < 2)
```

```
        stop("need at least 2 data points")
    if (!is.numeric(x))
        stop("invalid 'x'")
    fucv <- function(h,x,n,d).C(R_band_ucv_bin,as.integer(n),
            as.integer(length(x)),as.double(d), x,as.double(h),
            u = double(1))$u
    hmax <- 1.144 * sqrt(var(x)) * n^(-1/5)
           storage.mode(x) <- "double"

     Z <- .C(R_band_den_bin, as.integer(n),as.integer(nb),
           d = double(1),x, cnt = integer(nb))
     d <- Z$d
     cnt <- as.integer(Z$cnt)
     h <- optimize(fucv, c(lower, upper), tol = 0.1 * lower, x
           = cnt,n = n, d = d)$minimum
    if (h < 1.1 * lower | h > upper - 0.1 * lower)
        warning("minimum occurred at one end of the range")
    h
}
```

วิธีการเลือกแบนวิดจ์แบบ Plugin

```
width.SJ
function (x,nb=1000,lower=0.1*hmax, upper = hmax, method
           = c("ste", "dpi"))
{

fSD <- function(h, x,alph2,c1,n,d) (c1/SDh(x, alph2 * h^(5/7),
       n, d))^(1/5) – h

SDh <- function(x, h, n, d) .C(VR_phi4_bin,as.integer(n),
       as.integer(length(x)),as.double(d), x, as.double(h),u =
       double(1))$u
TDh <- function(x, h, n, d) .C(VR_phi6_bin, as.integer(n),
       as.integer(length(x)),as.double(d), x, as.double(h), u
       = double(1))$u
method <- match.arg(method)
n <- length(x)
if (!n)
           stop("'x' has length zero")
           storage.mode(x) <- "double"

Z <- .C(VR_den_bin, as.integer(n),as.integer(nb),d
     =double(1),x, cnt = integer(nb))
d <- Z$d
cnt <- as.integer(Z$cnt)
hmax <- 1.144 * sqrt(var(x)) * n^(-1/5)
scale <- min(sqrt(var(x)), IQR(x)/1.349)
a <- 1.24 * scale * n^(-1/7)
b <- 1.23 * scale * n^(-1/9)
c1 <- 1/(2 * sqrt(pi) * n)
TD <- -TDh(cnt, b, n, d)
alph2 <- 1.357 * (SDh(cnt, a, n, d)/TD)^(1/7)
```

```
    if (method == "dpi")
res <- (c1/SDh(cnt, (2.394/(n * TD))^(1/7), n, d))^(1/5)
    else {
      if (fSD(lower, cnt, alph2, c1, n, d) * fSD(upper,
           cnt,alph2, c1, n, d) > 0)

stop("no solution in the specified range of bandwidths")

res <- uniroot(fSD, c(lower, upper), tol = 0.1 * lower,
x = cnt, alph2 = alph2, c1 = c1, n = n, d = d)$root
      }
      4 * res
}
```

# ภาคผนวก ข

# การคำนวณค่า MSE และ MISE

เนื่องจากการประมาณความหนาแน่นแบบเคอร์เนล เป็นการประมาณฟังก์ชันแบบถ่วงน้ำหนัก ดังนั้น

กำหนดให้ $\hat{f}(t)$ เป็นการประมาณฟังก์ชันแบบถ่วงน้ำหนัก โดย

$$\hat{f}(t) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} w(X_i,t)$$

ดังนั้นค่าคาดหวัง คือ

$$E\,\hat{f}(t) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} E\ w(X_i,t)$$

$$= \quad \int_{-\infty}^{\infty} w(x,t)\, f(x)\, dx$$

เมื่อ $X_i$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน จะได้ว่า

$$\operatorname{var}\hat{f}(t) \quad = \quad \frac{1}{n}\operatorname{var} w(X_i,t)$$

$$= \frac{1}{n}\left[\int w(x,t)^2\, f(x)\,dx - \left\{\int w(x,t)\, f(x)\,dx\right\}^2\right]$$

กำหนดให้ $\hat{f}(x)$ เป็นการประมาณค่าแบบเคอร์เนล ดังนั้นจะได้ว่า

$$E\,\hat{f}(x) \quad = \quad \frac{1}{h}K\left(\frac{x-y}{h}\right) f(y)\,dy$$

$$\operatorname{var}\hat{f}(t) \quad = \quad \frac{1}{n}\int \frac{1}{h^2}K\left(\frac{x-y}{h}\right)^2 f(y)\,dy - \left\{\frac{1}{h}\int K\left(\frac{x-y}{h}\right) f(y)\,dy\right\}^2$$

กำหนดให้ $y = x - ht$ และใช้คุณสมบัติของฟังก์ชันเคอร์เนล $\int K(x)\,dx \ = \ 1$ จะได้ว่า

$$bias_h(x) = \int K(t)\, f(x-ht) dt - f(x)$$
$$= \int K(t)\{ f(x-ht) - f(x)\}$$

จากอนุกรมเทเลอร์ จะได้ว่า

$$f(x-ht) = f(x) - ht\, f'(x) + \frac{1}{2}h^2t^2 f''(x) + \ldots$$

จะได้ค่าความเอนเอียง ดังนี้

$$bias_h(x) = -ht\, f'(x)\int tK(t)\,dt + \frac{1}{2}h^2t^2 f''(x)\int t^2K(t) + \ldots$$
$$= \frac{1}{2}h^2 f''(x)k_2 + o\left(h^2\right)$$

สำหรับค่า MISE ต้องการค่า

$$\int bias_h(x)^2 \approx \frac{1}{4}h^4k_2\int f''(x)^2\, dx$$

ค่าความแปรปรวน ได้จาก

$$\operatorname{var}\hat{f}(x) = n^{-1}\int h^{-2}K\left(\frac{x-y}{h}\right)^2 f(y)\,dy - n^{-1}\{ f(x) + bias_h(x)\}^2$$
$$= n^{-1}h^{-1}\int f(x-ht)K(t)^2\, dt - n^{-1}\left\{ f(x) + O\left(h^2\right)\right\}^2$$

แทน $y = x - ht$ ลงในสมการ จะได้

$$\operatorname{var}\hat{f}(x) = n^{-1}h^{-1}\int\{ f(x) - ht\, f'(x) + \ldots\} K(t)^2\, dt + O\left(n^{-1}\right)$$
$$= n^{-1}h^{-1} f(x)\int K(t)^2\, dt + O\left(n^{-1}\right)$$
$$\approx n^{-1}h^{-1} f(x)\int K(t)^2\, dt$$

สำหรับค่า MISE ต้องการค่า

$$\int \operatorname{var} \hat{f}(x)\,dx \approx n^{-1}h^{-1}\int K(t)^2\,dt$$

ดังนั้น

$$MSE\left(\hat{f}(x)\right) = \left[E\hat{f}(x) - f(x)\right]^2 + \operatorname{var} \hat{f}(x)$$

$$MSE\left(\hat{f}(x)\right) = \left[\frac{1}{2}h^2 f''(x)k_2 + o\left(h^2\right)\right]^2 + n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt + o(nh)^{-1}$$

และ

$$MISE\left(\hat{f}(x)\right) = \int \operatorname{var} \hat{f}_h(x)\,dx + \int \operatorname{bias}^2 \hat{f}_h(x)\,dx$$

$$MISE\left(\hat{f}(x)\right) = n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt\, f(x)\,dx + \frac{h^4}{4}k_2^2\int_{-\infty}^{\infty} f''(x)^2\,dx + o(nh)^{-1} + o\left(h^4\right)$$

ภาคผนวก ค

ทฤษฎีบทที่เกี่ยวข้อง

**ทฤษฎีบทที่ 1** เมื่อ $x_1$ และ $x_2$ เป็นตัวแปรสุ่มที่มาจากการแจกแจงแบบปกติมาตรฐาน และเป็นอิสระกัน จะได้ว่า $x_1 + x_2$ มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 2

**พิสูจน์**

$$
\begin{aligned}
M_x(t) &= E\left[e^{tx}\right] \\
&= \int_{-\infty}^{\infty} e^{tx} f(x)\,dx \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx}.e^{-\frac{x^2}{2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2-2tx}{2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{\frac{(x-t)^2}{2} - \frac{t^2}{2}\right\} dx \\
&= e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{(x-t)^2}{2}}\,dx \\
M_x(t) &= e^{-\frac{t^2}{2}}
\end{aligned}
$$

ถ้า $x_1 \sim N(0,1)$ และ $x_2 \sim N(0,1)$ จะได้

$$
\begin{aligned}
M_{x_1+x_2}(t) &= E\left[e^{t(x_1+x_2)}\right] \\
&= E\left[e^{tx_1}\right] E\left[e^{tx_2}\right] \\
&= M_{x_1}(t) M_{x_2}(t)
\end{aligned}
$$

ดังนั้นจะได้ว่า เมื่อ $x_1$ และ $x_2$ เป็นอิสระกัน จะได้ว่า $x_1 + x_2$ มีการแจกแจงแบบปกติ ที่มี $\mu = \mu_1 + \mu_2$ และ $\sigma^2 = \sigma_1^2 + \sigma_2^2$

ดังนั้น ถ้า $x_1 \sim N(0,1)$ และ $x_2 \sim N(0,1)$ จะได้ว่า $x_1 + x_2 \sim N(0,2)$

**ทฤษฎีบทที่ 2** เมื่อ $X$ และ $Y$ เป็นตัวแปรสุ่มที่มาจากการแจกแจงแบบเอกซโพเนนเชียล ที่มีค่าเฉลี่ยเท่ากับ 1และเป็นอิสระกัน จะได้ว่า $X+Y$ มีการแจกแจงแบบแกมมา พารามิเตอร์เท่ากับ 2

**พิสูจน์** กำหนดให้ $X$ และ $Y$ มีการแจกแจงแบบเอกซโพเนนเชียล พารามิเตอร์ $\lambda$ และกำหนดให้ $Z = X+Y$ ดังนั้น

$$
\begin{aligned}
h(Z) &= \int_{-\infty}^{\infty} f(x)g(Z-X)dx \\
&= \int_{-\infty}^{\infty} \lambda e^{-\lambda x}.\lambda e^{-\lambda(Z-X)}dx \\
&= \lambda^2 e^{-\lambda Z}\Big]_0^Z \\
&= Z\lambda^2 e^{-\lambda Z} \ , \ Z > 0
\end{aligned}
$$

จะได้ว่า $Z \sim Gam(2)$

ดังนั้นจะได้ว่า เมื่อ $X$ และ $Y$ มีการแจกแจงแบบเอกซโพเนนเชียล พารามิเตอร์ $\lambda$ จะได้ว่า $Z = X+Y$ มีการแจกแจงแบบแกมมา ด้วยพารามิเตอร์ 2

**ทฤษฎีบทที่ 3** เมื่อ $X$ และ $Y$ เป็นตัวแปรสุ่มที่มาจากการแจกแจงแบบโคชีมาตรฐาน และเป็นอิสระกัน จะได้ว่า $X+Y$ มีการแจกแจงแบบโคชี (0,2)

# ประวัติผู้วิจัย

ชื่อ : นางสาวปิยะฉัตร ลีลาศิลปศาสน์

ชื่อวิทยานิพนธ์ : การเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นแบบเคอร์เนลของฟังก์ชันตัวแปรสุ่ม

สาขาวิชา : สถิติประยุกต์

ประวัติ

ประวัติส่วนตัว เกิดวันที่ 27 ตุลาคม พ.ศ. 2524 ภูมิลำเนาเดิมกรุงเทพมหานคร จบการศึกษาระดับปริญญาตรี จากสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ คณะวิทยาศาสตร์ประยุกต์ สาขาสถิติประยุกต์ ในปี พ.ศ. 2546 จากนั้นจึงเข้าศึกษาต่อในสาขาวิชาสถิติประยุกต์ ภาควิชาสถิติประยุกต์ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ เมื่อปี พ.ศ. 2546