บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการวิเคราะห์เชิงสถิตินั้น เมื่อต้องการที่จะทราบลักษณะหรือคุณสมบัติบางประการของประชากร จะต้องอาศัยข้อมูลที่สุ่มจากประชากรเพื่อใช้ในการวิเคราะห์และอ้างอิงในเชิงสถิติ โดยทั่วไปการวิเคราะห์ข้อมูลทางสถิติเพื่อศึกษาความสัมพันธ์เชิงฟังก์ชันระหว่างตัวแปรมักมีข้อสมมติเบื้องต้นที่ว่า ตัวแปรสุ่มมีรูปแบบการแจกแจงความน่าจะเป็นรูปแบบใดรูปแบบหนึ่ง รูปแบบการแจกแจงที่ผู้วิเคราะห์จำนวนมากนิยมใช้คือรูปแบบการแจกแจงแบบปกติ ซึ่งถูกนำมาใช้โดยปราศจากการตรวจสอบอย่างถี่ถ้วนว่าเหมาะสมกับข้อมูลที่มีหรือไม่ และเนื่องจากรูปแบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มมีความไม่แน่นอนและเป็นไปได้หลายรูปแบบ ส่งผลให้ตัวแปรสุ่มที่มีอาจไม่ได้มีการแจกแจงที่เป็นไปตามข้อสมมติดังที่ผู้วิเคราะห์พิจารณาก็เป็นได้ ดังนั้นผู้วิเคราะห์จะแน่ใจได้อย่างไรว่ารูปแบบการแจกแจงความน่าจะเป็นที่ได้เลือกใช้ในการวิเคราะห์เชิงสถิตินั้นมีความเหมาะสมกับข้อมูล การศึกษาวิธีการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นจึงเป็นสิ่งที่น่าสนใจและสำคัญอย่างยิ่ง เพราะหากทราบฟังก์ชันความหนาแน่นของความน่าจะเป็นหรือสามารถประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวแปรสุ่ม จะทำให้ทราบได้ว่าข้อมูลมีลักษณะอย่างไร การกระจายของข้อมูลเป็นอย่างไร มีความเบ้เกิดขึ้นหรือไม่ มีฐานนิยมหลายแห่งหรือไม่ ทำให้เห็นลักษณะของการแจกแจงความน่าจะเป็น สามารถใช้อ้างอิงเพื่อบอกลักษณะต่าง ๆ รวมถึงอธิบายรายละเอียดต่าง ๆ ของตัวแปรได้ สิ่งเหล่านี้จะเป็นประโยชน์ในการวิเคราะห์ข้อมูล การสรุปผล และการประยุกต์ใช้ในหลาย ๆ ด้าน ส่งผลให้ผู้ที่พบเห็นสามารถเข้าใจสิ่งที่เกี่ยวข้องได้ง่ายและรวดเร็ว ตลอดจนทำให้ทราบคุณลักษณะต่าง ๆ ของข้อมูล เช่น ค่าเฉลี่ย ความแปรปรวน เป็นต้น (ประชุม สุวัตถี, 2539 , Hardle, 1990)

การประมาณความหนาแน่นเป็นวิธีการสร้างฟังก์ชันความหนาแน่นของความน่าจะเป็นจากตัวแปรสุ่ม ในกรณีที่ไม่ทราบรูปแบบฟังก์ชันความหนาแน่นของความน่าจะเป็น ต้องทำการประมาณขึ้น เรียกว่า การประมาณความหนาแน่น (Density estimation) โดยแนวทางหลักที่นิยมใช้ คือ การประมาณความหนาแน่นแบบอิงพารามิเตอร์ (Parametric density estimation) และการประมาณความหนาแน่นแบบไม่อิงพารามิเตอร์ (Nonparametric density estimation)

สำหรับการประมาณความหนาแน่นแบบอิงพารามิเตอร์จะมีข้อสมมติว่า เซตของข้อมูลตัวอย่างมีการแจกแจงรูปแบบใดรูปแบบหนึ่ง ขั้นตอนในการประมาณความหนาแน่นแบบอิงพารามิเตอร์ คือ กำหนดรูปแบบการแจกแจงให้กับข้อมูล จากนั้นหาค่าประมาณของพารามิเตอร์ทุกตัวโดยอาศัยข้อมูลจากตัวอย่างสุ่มมาประมาณค่าพารามิเตอร์ เพื่อนำไปแทนค่าในรูปแบบของฟังก์ชัน-การแจกแจงที่กำหนด เช่น กำหนดให้ข้อมูลมีการแจกแจงแบบปกติ ซึ่งมีพารามิเตอร์คือค่าเฉลี่ยและความแปรปรวน คือ μ และ σ^2 ตามลำดับ จะต้องหาค่าประมาณของ μ และ σ^2 โดยอาศัยข้อมูลจากตัวอย่างสุ่ม จากนั้นนำค่าที่ประมาณได้จากตัวอย่างสุ่มไปแทนที่พารามิเตอร์ในฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบปกติ อย่างไรก็ตามการประมาณความหนาแน่นแบบอิงพารามิเตอร์จะต้องเลือกการแจกแจงให้เหมาะสมและใกล้เคียงกับข้อมูลให้มากที่สุด หากเลือกการแจงแจงที่ไม่เหมาะสมกับข้อมูล จะส่งผลให้ได้ผลลัพธ์ที่ไม่แม่นยำ นำไปสู่การสรุปผลหรือการอ้างอิงที่เกิดความผิดพลาด

ส่วนการประมาณความหนาแน่นแบบไม่อิงพารามิเตอร์ ถือเป็นเครื่องมือสำคัญที่จะช่วยในการวิเคราะห์และนำเสนอโครงสร้างของข้อมูลได้มีประสิทธิภาพ มีข้อสมมติน้อยกว่าการ-ประมาณความหนาแน่นแบบอิงพารามิเตอร์ ไม่มีการกำหนดการแจกแจงให้กับเซตของข้อมูล-ตัวอย่าง การประมาณความหนาแน่นแบบไม่อิงพารามิเตอร์สามารถทำได้หลายวิธี โดยวิธีฮิสโตแกรมถือเป็นวิธีที่ง่ายและเก่าแก่ที่สุด นิยมใช้อย่างกว้างขวางเพื่อแสดงรูปร่างของฟังก์ชัน-ความหนาแน่นของความน่าจะเป็น (Wand & Jones, 1995, pp. 5) วิธีฮิสโตแกรมจะแสดงรูปร่างของฟังก์ชันความหนาแน่นของความน่าจะเป็นเป็นกราฟแท่งความถี่หลาย ๆ แท่งเรียงต่อกัน การ-ประมาณความหนาแน่นโดยใช้ฮิสโตแกรมจะต้องคำนึงถึงการเลือกจุดเริ่มต้นและการกำหนดความกว้างของช่วง เนื่องจากส่งผลต่อรูปร่างของฮิสโตแกรมที่จะแตกต่างกันออกไป การใช้ฮิสโตแกรมมักเกิดข้อบกพร่องในการประมาณความหนาแน่น นั่นคือ อาจเกิดความไม่ต่อเนื่องของกราฟฟังก์ชันความหนาแน่นของความน่าจะเป็น มีสาเหตุจากการมีจำนวนช่วงที่มากเกินไป หรือกำหนดค่าความกว้างของช่วงน้อยเกินไป ทำให้บางช่วงไม่มีค่าสังเกตตกอยู่ จากปัญหาดังกล่าวทำให้ในเวลาต่อมาได้มีผู้คิดวิธีการต่าง ๆ มากมายเพื่อประมาณฟังก์ชันความหนาแน่นของความ-น่าจะเป็น เช่น วิธีตัวประมาณอย่างง่าย (Fix & Hodges, 1951) วิธีตัวประมาณแบบเคอร์เนล (Rosenblatt, 1956 , Parzen, 1962) วิธีฟังก์ชันถ่วงน้ำหนักทั่วไป (Whittle, 1958) วิธีย่านใกล้สุด (Loftsgaarden & Quesenberry, 1965) วิธีเคอร์เนลแปรผัน (Breiman, Meisel, & Purcell, 1977) เป็นต้น วิธีการที่กล่าวข้างต้นส่วนใหญ่ล้วนมีพื้นฐานจากวิธีฮิสโตแกรมแทบทั้งสิ้น

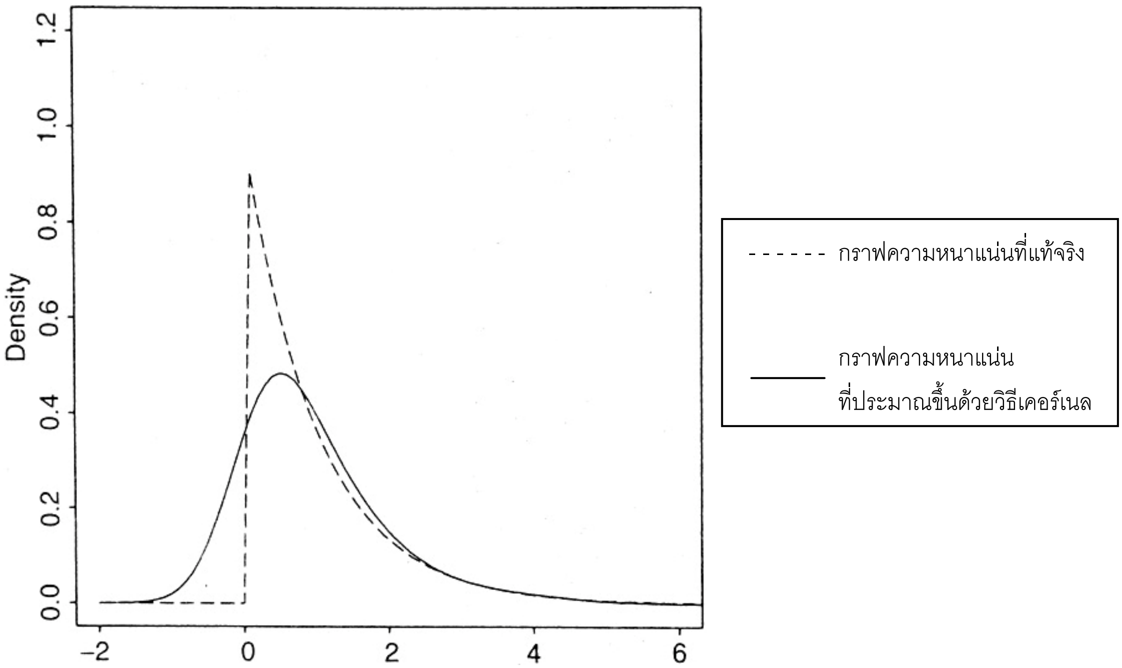
เมื่อกล่าวถึงแนวทางการประมาณความหนาแน่นแบบไม่อิงพารามิเตอร์ การประมาณความหนาแน่นด้วยวิธีเคอร์เนลถือเป็นวิธีหนึ่งที่มีความสำคัญ การประมาณความหนาแน่นแบบเคอร์เนลถูกเสนอขึ้นอย่างเป็นทางการในปีค.ศ. 1956 โดย Murray Rosenblatt เป็นวิธีที่มีผู้ศึกษาไว้มากที่สุด มีบทความทางวิชาการออกมาเป็นจำนวนมากและเป็นที่นิยมอย่างแพร่หลาย การเลือกใช้วิธีนี้ไม่ได้หมายความว่าเป็นวิธีที่ดีที่สุดสำหรับทุก ๆ สถานการณ์ แต่มีหลายเหตุผลที่จะพิจารณาวิธีนี้เป็นอันดับต้น ๆ นั่นคือ สามารถใช้ดำเนินการได้จริงและมีความซับซ้อนไม่มาก รวมทั้งคุณสมบัติทางคณิตศาสตร์ที่เกี่ยวข้องนั้นสามารถทำความเข้าใจได้ไม่ยากและถือเป็นพื้นฐานสำคัญที่นำไปสู่การศึกษาวิธีการประมาณความหนาแน่นแบบไม่อิงพารามิเตอร์วิธีอื่น ๆ การประมาณความหนาแน่นด้วยวิธีเคอร์เนลจะมีการถ่วงน้ำหนักข้อมูลของตัวแปรสุ่มผ่านฟังก์ชันถ่วงน้ำหนัก นั่นคือ ฟังก์ชันเคอร์เนล ในการวิเคราะห์แบบตัวแปรเดียว (univariate) ถือว่าการใช้วิธีการประมาณแบบเคอร์เนลเป็นวิธีที่มีประสิทธิภาพ สิ่งสำคัญที่มีอิทธิพลต่อการประมาณความหนาแน่นด้วยวิธีเคอร์เนลมีสองส่วน คือ การเลือกค่าความกว้างของช่วง (แบนวิดธ์ (h)) และ การกำหนดรูปแบบฟังก์ชันเคอร์เนล ($K(u)$) (Silverman, 1986) สำหรับการกำหนดรูปแบบฟังก์ชันเคอร์เนลไม่ใช่สิ่งที่มีผลต่อการประมาณความหนาแน่นมากนัก แต่สิ่งที่สำคัญที่สุดคือการเลือกแบนวิดธ์ให้มีความเหมาะสมกับข้อมูลที่ต้องการวิเคราะห์ (Mugdadi & Ahmad, 2004) การกำหนดค่าแบนวิดธ์ให้เหมาะสมจะส่งผลให้การวิเคราะห์ข้อมูลเป็นไปอย่างมีประสิทธิภาพ เนื่องจากเมื่อกำหนดค่าแบนวิดธ์แตกต่างกัน จะส่งผลให้รูปร่างของกราฟความหนาแน่นมีความแตกต่างกัน (Jones, Marron, & Sheather, 1996) รูปร่างกราฟความหนาแน่นตามค่าแบนวิดธ์ที่ใช้ในประมาณความหนาแน่นมีลักษณะคือ เมื่อค่าแบนวิดธ์มีค่าเข้าใกล้ศูนย์ $(h \to 0)$ กราฟความหนาแน่นจะมีลักษณะแหลมและมีลักษณะที่ไม่เรียบ เกิดการขึ้นลงบ่อยครั้ง นำเสนอลักษณะของข้อมูลได้ไม่ดี นั่นคือ เกิดความเรียบน้อยกว่าปกติ (undersmooth) เมื่อแบนวิดธ์มีค่าเพิ่มขึ้นจะทำให้ลักษณะกราฟความหนาแน่นมีความเรียบยิ่งขึ้น และเมื่อกำหนดค่าแบนวิดธ์มีค่ามากเป็นอนันต์ $(h \to \infty)$ กราฟความหนาแน่นจะมีลักษณะเรียบเป็นเส้นตรง นั่นคือเกิดความเรียบมากกว่าปกติ (oversmooth) ส่งผลให้ลักษณะกราฟแสดงรายละเอียดของการเข้าถึงข้อมูลได้ไม่ชัดเจน (Hardle, 1990, pp. 48)

การเลือกแบนวิดธ์ด้วยวิธีง่าย ๆ อาจทำได้โดยใช้ดุลยพินิจของผู้วิเคราะห์มาพิจารณาลักษณะของกราฟความหนาแน่นที่ได้จากแบนวิดธ์ค่าต่าง ๆ ว่ารูปกราฟใดมีลักษณะที่แสดงถึงการเข้าถึงข้อมูลได้ดีที่สุด มีการขึ้นลงของกราฟและมีความเรียบพอเหมาะ หรืออาจใช้วิธีการกำหนดค่าแบนวิดธ์โดยเริ่มตั้งแต่ค่าน้อย ๆ ไปถึงค่ามาก ๆ จนกระทั่งเห็นรูปร่างหรือโครงสร้าง

กราฟความหนาแน่นที่แสดงการเข้าถึงโครงสร้างของข้อมูลได้ดี จึงตัดสินใจเลือกค่าแบนวิดธ์นั้น (Silverman, 1986, pp. 44) นอกจากวิธีดังกล่าวแล้ว ยังมีการเลือกแบนวิดธ์วิธีอื่น ๆ ที่ถูกเสนอขึ้นโดยอาศัยหลักการทางคณิตศาสตร์ เช่น วิธี Direct plug-in (Woodroofe, 1970 , Nadaraya, 1974) วิธี Rules of thumb (Deheuvels, 1977) วิธี Least squares cross-validation (Rudemo, 1982 , Bowman, 1984) วิธี Silverman’s rules of thumb (Silverman, 1986) วิธี Biased cross-validation (Scott & Terrell, 1987) วิธี Smoothed bootstrap (Taylor, 1989 , Faraway & Jhun, 1990) วิธี Park and Marron’s plug-in (Park & Marron, 1990) และ วิธี Solve the equation plug-in (Sheather & Jones, 1991) เป็นต้น

ถึงแม้ว่าการประมาณความหนาแน่นด้วยวิธีเคอร์เนลมีประโยชน์และมีข้อดีหลายประการ แต่เป็นที่ทราบกันดีเช่นกันว่ากราฟความหนาแน่นที่เกิดจากการประมาณความหนาแน่นด้วยวิธีเคอร์เนลมักเกิดปัญหาเมื่อประมาณความหนาแน่นบริเวณขอบเขต นั่นคือ ในการประมาณความหนาแน่นบริเวณขอบเขตสิ้นสุดมักประมาณความหนาแน่นได้ไม่ตรงกับรูปแบบความหนาแน่นที่แท้จริง ซึ่งปัญหาที่เกิดขึ้นเรียกว่าปัญหาการประมาณความหนาแน่นบริเวณขอบเขต (ปัญหา Boundary effect) สาเหตุของปัญหาเกิดจากตัวประมาณความหนาแน่นด้วยวิธีเคอร์เนลนั้นไม่ทราบขอบเขตสิ้นสุดที่แน่ชัด ส่งผลให้การประมาณความหนาแน่นบริเวณขอบเขตเกิดความผิดพลาด นั่นคือ ได้กราฟความหนาแน่นที่เลยออกไปในบริเวณที่ไม่มีตัวอย่างสุ่มตกอยู่ และขอบเขตสิ้นสุดของกราฟความหนาแน่นที่ประมาณได้ไม่สอดคล้องกับขอบเขตสิ้นสุดที่แท้จริง การประมาณความหนาแน่นในบริเวณขอบเขตจึงถือเป็นปัญหาสำคัญในการประมาณความหนาแน่นด้วยวิธีเคอร์เนล ดังนั้นการประมาณความหนาแน่นในบริเวณดังกล่าวจึงจำเป็นที่จะต้องมีรูปแบบเฉพาะ เพื่อใช้ในการประมาณความหนาแน่น เพื่อให้ผลในการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นมีความสอดคล้องกับความเป็นจริง (Jones & Foster, 1996 , Karunamuni & Zhang, 2008 , Wand & Jones, 1995) ปัญหาดังกล่าวแสดงได้จากภาพที่ 1.1

ภาพที่ 1.1
ฟังก์ชันความหนาแน่นของความน่าจะเป็นที่เกิดปัญหา
การประมาณความหนาแน่นบริเวณขอบเขต



จากภาพที่ 1.1 แสดงตัวอย่างกราฟฟังก์ชันความหนาแน่นของความน่าจะเป็นที่เกิดปัญหาการประมาณความหนาแน่นบริเวณขอบเขตเมื่อประมาณฟังก์ชันความหนาแน่นของความ-น่าจะเป็นด้วยวิธีเคอร์เนล ขนาดตัวอย่างคือ 1,000 ฟังก์ชันความหนาแน่นของความน่าจะเป็นที่แท้จริงคือการแจกแจงแบบเอกซ์โพเนนเชียลที่มีพารามิเตอร์เท่ากับ 1 (Exp(1)) นั่นคือ $f(x) = e^{-x}; x > 0$ (Wand & Jones, 1995, pp. 46)

ในปัจจุบันมีวิธีแก้ปัญหาการประมาณความหนาแน่นหลากหลายวิธีที่สามารถแก้ไขปัญหาการประมาณความหนาแน่นบริเวณขอบเขต วิธีทั่วไปที่ถือเป็นวิธีพื้นฐาน เช่น The reflection method (Boneva, Kendall, & Stefanov, 1971) เป็นวิธีที่ถูกออกแบบมาสำหรับใช้ในกรณีที่แทนค่าศูนย์หรือ a โดยที่ $0 < a \leq \infty$ ลงในอนุพันธ์อันดับที่หนึ่งของฟังก์ชันความ-หนาแน่นของความน่าจะเป็นได้เท่ากับศูนย์ จะมีการสะท้อนค่าของตัวแปรสุ่มให้มีเครื่องหมายตรงข้ามกับตัวแปรสุ่มเดิม และกำหนดเครื่องหมายของฟังก์ชันถ่วงน้ำหนักให้เป็นลบ The boundary kernel method (Gasser & Muller, 1979) เป็นวิธีที่ใช้ฟังก์ชันเคอร์เนลรูปแบบเฉพาะสำหรับประมาณความหนาแน่นบริเวณขอบเขตสิ้นสุด สามารถใช้ได้สำหรับทุกรูปแบบฟังก์ชัน-

ความหนาแน่นของความน่าจะเป็น แต่มีข้อบกพร่องคือ ค่าประมาณความหนาแน่นอาจมีค่าติดลบ โดยเฉพาะกรณีที่แทนค่าศูนย์หรือ a โดยที่ $0 < a \le \infty$ ลงในฟังก์ชันความหนาแน่นของความ-น่าจะเป็น แล้วได้ค่าใกล้เคียงศูนย์ The transformation method (Copas & Fryer, 1980) เป็นวิธีที่ใช้ในกรณีที่ตัวแปรสุ่มมีค่าอยู่ในช่วงจำกัด $[a, b]$ จะอาศัยการแปลงข้อมูลของตัวแปรสุ่มผ่านฟังก์ชันการแปลง เป็นต้น นอกจากวิธีที่ได้กล่าวมาแล้ว มีหลากหลายวิธีถูกเสนอขึ้นโดยการปรับปรุงหรือประยุกต์จากวิธีพื้นฐาน เช่น The Jones and Foster nonnegative adaptation estimator (Jones & Foster, 1996) The pseudodata method (Cowling & Hall, 1996) The method of Zhang et al. (Zhang, Karunamuni, & Jones, 1999) The method of Hall and Park (Hall & Park, 2002) The method of Karunamuni and Alberts (Karunamuni & Alberts, 2005) เป็นต้น

จากการศึกษาที่ผ่านมาได้มีผู้เสนอวิธีการประมาณความหนาแน่นเพื่อแก้ไขปัญหาการประมาณความหนาแน่นบริเวณขอบเขตออกมาอย่างต่อเนื่อง และทำการเปรียบเทียบประสิทธิภาพการแก้ไขปัญหาดังกล่าวด้วยวิธีที่เสนอขึ้นเทียบกับวิธีอื่น ๆ โดยใช้ค่าแบนวิดธ์ที่เหมาะสม (Optimal bandwidth) เท่านั้น โดยไม่ได้นำวิธีการเลือกแบนวิดธ์เข้ามาพิจารณาเพื่อกำหนดค่าแบนวิดธ์แต่อย่างใด ด้วยเหตุนี้ผู้วิจัยจึงเห็นว่า หากนำวิธีการเลือกแบนวิดธ์ซึ่งเป็นส่วนสำคัญที่สุดในการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็น มาร่วมพิจารณาเพื่อประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นที่มีปัญหาการประมาณความหนาแน่นบริเวณขอบเขต โดยใช้ตัวประมาณความหนาแน่นที่สามารถแก้ไขปัญหาดังกล่าวได้ น่าจะทำให้การประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นของตัวแปรสุ่มมีความถูกต้องมากยิ่งขึ้น

1.2 วัตถุประสงค์ของการวิจัย

เปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดธ์ 6 วิธี คือ วิธี Direct plug-in วิธี Rules of thumb วิธี Least squares cross-validation วิธี Silverman's rules of thumb วิธี Biased cross-validation และ วิธี Solve the equation plug-in ในการประมาณความหนาแน่นด้วยวิธีของ Zhang และคณะ (1999) เมื่อฟังก์ชันความหนาแน่นของความน่าจะเป็นเกิดปัญหาการประมาณความหนาแน่นบริเวณขอบเขต

1.3 สมมติฐานของการวิจัย

1. การเลือกแบนวิดธ์ด้วยวิธี Rules of thumb และวิธี Biased cross-validation มีประสิทธิภาพในการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นบริเวณขอบเขตได้ดีที่สุด

2. การเลือกแบนวิดธ์ด้วยวิธี Least squares cross-validation มีประสิทธิภาพในการประมาณฟังก์ชันความหนาแน่นของความน่าจะเป็นบริเวณขอบเขตที่ด้อยกว่าวิธีอื่น ๆ

1.4 ขอบเขตของการวิจัย

1. กำหนดตัวแปรสุ่ม $X_1, X_2, ..., X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน และมีฟังก์ชัน-การแจกแจงความน่าจะเป็นเดียวกัน (Independent and identical distributed : i.i.d.)

2. ฟังก์ชันความหนาแน่นของความน่าจะเป็นที่ใช้สำหรับสร้างตัวแปรสุ่มจำนวน 2 รูปแบบ คือ $f(x) = 5e^{-5x}$ เมื่อ $x > 0$ และ $f(x) = \frac{5}{4}(1+15x)e^{-5x}$ เมื่อ $x > 0$

3. ตัวประมาณฟังก์ชันความหนาแน่นที่ใช้ในการวิจัย คือ ตัวประมาณความหนาแน่นของ Zhang และคณะ (1999)

4. ฟังก์ชันเคอร์เนลที่ใช้ในการวิจัย คือ ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian kernel) และ ฟังก์ชันเคอร์เนลของจุดสิ้นสุด (End-point kernel)

5. วิธีการเลือกแบนวิดธ์ที่ใช้ในการวิจัยจำนวน 6 วิธี คือ การเลือกแบนวิดธ์ด้วยวิธี Direct plug-in วิธี Rules of thumb วิธี Least squares cross-validation วิธี Silverman’s rules of thumb วิธี Biased cross-validation และ วิธี Solve the equation plug-in

6. ขนาดตัวอย่างที่ใช้ในการวิจัย คือ 30 , 50 , 100 , 150 และ 200

7. จำนวนการทำซ้ำ 1,000 รอบ

1.5 เกณฑ์ที่ใช้ในการเปรียบเทียบ

ในการเปรียบเทียบประสิทธิภาพของวิธีการเลือกแบนวิดธ์ สำหรับรูปแบบฟังก์ชัน-ความหนาแน่นของความน่าจะเป็นที่เกิดปัญหาการประมาณความหนาแน่นบริเวณขอบเขตทั้งสองรูปแบบ จะใช้ค่าความคลาดเคลื่อนกำลังสองรวมเฉลี่ย (Mean Integrated Squared Error : $MISE$) เป็นเกณฑ์ในการชี้วัด

1.6 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

เป็นแนวทางในการเลือกแบนวิดธ์ให้มีความเหมาะสมสำหรับรูปแบบฟังก์ชันความหนาแน่นของความน่าจะเป็นที่เกิดปัญหาการประมาณความหนาแน่นบริเวณขอบเขต

1.7 คำจำกัดความของศัพท์ที่ใช้ในการวิจัยและสัญลักษณ์ที่ใช้ในการวิจัย

1.7.1 คำจำกัดความของศัพท์ที่ใช้ในการวิจัย

1. ตัวแปรสุ่ม (Random variable) คือ ฟังก์ชันค่าจริงที่นิยามได้บนปริภูมิตัวอย่าง S ที่มีโดเมน (Domain) เป็นเซตของปริภูมิตัวอย่าง S และมีเรนจ์ (Range) เป็นเซตย่อยของจำนวนจริง

2. ตัวอย่างสุ่ม (Random sample) คือ เซตของตัวแปรสุ่ม $X_1, X_2,..., X_n$ ที่เป็นอิสระต่อกัน และมีฟังก์ชันการแจกแจงความน่าจะเป็นเดียวกัน (Independent and identical distributed : i.i.d.)

3. ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability density function) เขียนแทนด้วย $f(x)$ คือ ฟังก์ชันที่ใช้อธิบายถึงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่อง (Continuous random variable) ที่นิยามบนจำนวนจริง $\mathbb{R} = (-\infty, \infty)$ สามารถนำเสนอเป็นกราฟโค้งความถี่ต่อเนื่อง ค่าของความน่าจะเป็นคือค่าของพื้นที่ใต้โค้ง ซึ่ง $f(x)$ มีคุณสมบัติคือ $f(x) \geq 0$ เมื่อ $-\infty < x < \infty$ และ $\int_{-\infty}^{\infty} f(x)dx = 1$

4. ค่าประมาณความหนาแน่นแบบเคอร์เนล (Kernel density estimate) คือ ค่าความหนาแน่นที่ประมาณขึ้นโดยวิธีการประมาณความหนาแน่นแบบเคอร์เนล

5. ฟังก์ชันที่มีฐานนิยมเดียว (Unimodal function) คือ ฟังก์ชัน f มีฐานนิยมค่าเดียวที่ M ก็ต่อเมื่อ f เป็นฟังก์ชันไม่ลด (Nondecreasing function) ที่ x เมื่อ $x < M$ และ f เป็นฟังก์ชันไม่เพิ่ม (Nonincreasing function) ที่ x เมื่อ $x > M$

1.7.2 สัญลักษณ์ที่ใช้ในการวิจัย

สัญลักษณ์ต่าง ๆ ที่ใช้ในการวิจัย มีรายละเอียดดังนี้

DPI	แทน	การเลือกแบนวิดธ์ด้วยวิธี Direct plug-in
ROT	แทน	การเลือกแบนวิดธ์ด้วยวิธี Rules of thumb
LSCV	แทน	การเลือกแบนวิดธ์ด้วยวิธี Least squares cross-validation
SROT	แทน	การเลือกแบนวิดธ์ด้วยวิธี Silverman’s rules of thumb
BCV	แทน	การเลือกแบนวิดธ์ด้วยวิธี Biased cross-validation
STE	แทน	การเลือกแบนวิดธ์ด้วยวิธี Solve the equation plug-in
h	แทน	ค่าแบนวิดธ์ มีค่าเป็นจำนวนจริงบวก
h_{opt}	แทน	ค่าแบนวิดธ์ที่เหมาะสม (Optimal bandwidth)
h_{DPI}	แทน	ค่าแบนวิดธ์ที่ได้จากวิธี Direct plug-in
h_{ROT}	แทน	ค่าแบนวิดธ์ที่ได้จากวิธี Rules of thumb
h_{LSCV}	แทน	ค่าแบนวิดธ์ที่ได้จากวิธี Least squares cross-validation
h_{SROT}	แทน	ค่าแบนวิดธ์ที่ได้จากวิธี Silverman’s rules of thumb
h_{BCV}	แทน	ค่าแบนวิดธ์ที่ได้จากวิธี Biased cross-validation
h_{STE}	แทน	ค่าแบนวิดธ์ที่ได้จากวิธี Solve the equation plug-in
MISE	แทน	ความคลาดเคลื่อนกำลังสองรวมเฉลี่ย
n	แทน	ขนาดตัวอย่าง
$\hat{f}(x)$	แทน	ตัวประมาณฟังก์ชันความหนาแน่นด้วยวิธีเคอร์เนล
$\tilde{f}(x)$	แทน	ตัวประมาณฟังก์ชันความหนาแน่นของ Zhang และคณะ (1999)
$K(u)$	แทน	ฟังก์ชันเคอร์เนล
$K_{(0)}(u)$	แทน	ฟังก์ชันเคอร์เนลของจุดสิ้นสุด (End-point kernel)
$\hat{\sigma}$	แทน	ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม
IQR	แทน	พิสัยระหว่างควอร์ไทล์ (Interquartile range)
A	แทน	ค่าที่น้อยที่สุดระหว่างค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม กับ $\frac{IQR}{1.34}$