



บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน

วิกานดา ผาพันธ์*

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ธิปไตย พงษ์ศาสตร์

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 09 3141 4615 อีเมล: wikanda.p@sci.kmutnb.ac.th DOI: 10.14416/j.kmutnb.2018.12.08

รับเมื่อ 5 เมษายน 2561 แก้ไขเมื่อ 23 กรกฎาคม 2561 ตอปรับเมื่อ 1 พฤศจิกายน 2561 เผยแพร่ออนไลน์ 11 ธันวาคม 2561

© 2018 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

บทคัดย่อ

การแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน (Length-biased Inverse Gaussian Distribution; LBIG) เป็นการแจกแจงที่มีประโยชน์อย่างมากในทางสถิติและมีการประยุกต์ใช้งานในด้านต่างๆ เช่น ฟิสิกส์ วิศวกรรมศาสตร์ ชีววิทยา ซึ่งเหมาะสำหรับใช้ในการวิเคราะห์ข้อมูลที่มีลักษณะเบ้ขวา ในงานวิจัยนี้สนใจศึกษาสมการภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) และหาข้อมูลเมทริกซ์สารสนเทศของฟิชเชอร์ เพื่อสร้างบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับของพารามิเตอร์สำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน โดยพิจารณาค่าความน่าจะเป็นค้ำรวม (Coverage Probability; CP) เปรียบเทียบกับสัมประสิทธิ์ของความเชื่อมั่นที่ 0.98 ของบริเวณวงรีความเชื่อมั่น ที่ขนาดของตัวอย่าง n มีค่าเท่ากับ 10, 20, 30, 50, 60, 100, 500 และ 1,000 พารามิเตอร์ λ มีค่าเท่ากับ 1, 3, 5, 10, 15, 20 และ μ มีค่าเท่ากับ 1 โดยจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo) ทำซ้ำทั้งหมด 10,000 รอบ ในแต่ละโมเดลด้วยโปรแกรม R เวอร์ชัน 3.4.3

คำสำคัญ: การแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน, ข้อมูลเมทริกซ์สารสนเทศของฟิชเชอร์, บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับ



Asymptotic Confidence Ellipses for Length-biased Inverse Gaussian Distribution

Wikanda Phaphan*

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Tippatai Pongsart

Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand

* Corresponding Author, Tel. 09 3141 4615, E-mail: wikanda.p@sci.kmutnb.ac.th DOI: 10.14416/j.kmutnb.2018.12.08

Received 5 April 2018; Revised 23 July 2018; Accepted 1 November 2018; Published online: 11 December 2018

© 2018 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

Abstract

The length-biased inverse Gaussian distribution has been useful in statistics. This distribution has been used in extensive applications, for example, physics, engineering, and biology. It is suited for the right-skewed data analysis. In this research, we are interested in studying the maximum likelihood equations and finding the Fisher information matrix to construct asymptotic confidence ellipse for the length-biased inverse Gaussian distribution by comparing a coverage probability with a confidence coefficient of 0.98 of confidence ellipses for cases of sample sizes $n = 10, 20, 30, 50, 60, 100, 500,$ and $1,000$ parameter $\lambda = 1, 3, 5, 10, 15, 20$ and parameter $\mu = 1$. Monte Carlo simulations are considered with 10,000 iterations by using program R (3.4.3).

Keywords: Length-biased Inverse Gaussian Distribution, Fisher Information Matrix, Asymptotic Confidence Ellipses

1. บทนำ

การแจกแจงอินเวอร์สเกาส์เซียน (Inverse Gaussian Distribution; IG) เป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่องในวงศ์ชั่งกำลัง (Exponential Family) ซึ่งเป็นการแจกแจงทางเลือกที่มีประโยชน์อย่างมากในทางสถิติ เหมาะสำหรับการวิเคราะห์ข้อมูลที่มีลักษณะเบ้ขวา ดังนั้นการแจกแจงนี้จึงเป็นการแจกแจงทางสถิติที่นิยมนำไปประยุกต์ใช้ในศาสตร์ต่างๆ เช่น ฟิสิกส์ วิศวกรรมศาสตร์ ชีววิทยา และการเงิน เป็นต้น ดังที่ได้อธิบายไว้ในงานวิจัยของ Folks และ Chhikara [1] ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงอินเวอร์สเกาส์เซียนประกอบด้วย 2 พารามิเตอร์ คือ

$$f_{IG}(x, \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\} \quad ; x > 0$$

การแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน (Length-biased Inverse Gaussian Distribution; LBIG) เป็นการแจกแจงหนึ่งที่มีลักษณะคล้ายกับการแจกแจงอินเวอร์สเกาส์เซียน ซึ่งงานวิจัยที่เกี่ยวกับความยาวเอนเอียงของการแจกแจงอินเวอร์สเกาส์เซียนได้ถูกนำเสนอโดย Ahsanullah และ Kirmani [2] และ Khattree [3] บางครั้งมีการนำเสนอการแจกแจงความยาวเอนเอียงในรูปแบบที่เรียกว่าการแจกแจงถ่วงน้ำหนักดังเช่นงานวิจัยของ Patil และ Rao [4] เนื่องจากเป็นการแจกแจงที่มีลักษณะเบ้ขวา เช่นเดียวกับการแจกแจงอินเวอร์สเกาส์เซียน จึงมีประยุกต์ใช้งานกับข้อมูลชั่วชีวิต (Lifetime Data) ซึ่งเป็นข้อมูลที่มีค่าไม่เป็นลบ เช่นเดียวกันกับการแจกแจงอินเวอร์สเกาส์เซียน ดังนั้นทั้งสองการแจกแจงจึงเหมาะสำหรับแผนการสุ่มตัวอย่างด้านชีวมิติ และการวิเคราะห์ความอยู่รอด งานวิจัยของ Patil และ Rao [5], Blumenthal [6], Scheaffer [7] และ Simon [8] แสดงให้เห็นถึงการใช้งานการแจกแจงความยาวเอนเอียงในด้านต่างๆ ในงานวิจัยของ Cnaan [9] ได้อธิบายความหมายการแจกแจงความยาวเอนเอียงในการสุ่มตัวอย่างสำหรับการศึกษาโรคหลอดเลือดหัวใจอุดตัน Gupta และ Tripathi [10] เปรียบเทียบรูปแบบธรรมดาและรูปแบบการแจกแจงความยาวเอนเอียง โดยสุ่มตัวอย่างจากทั้งสองแบบของการแจกแจง Sen [11] ได้ศึกษาคุณสมบัติทางคณิตศาสตร์ ค่าเฉลี่ย

ฮาร์โมนิก สัมประสิทธิ์การแปรผันและอธิบายลักษณะของการแจกแจงความยาวเอนเอียง Gupta และ Kirmani [12] ตรวจสอบความสัมพันธ์ระหว่างความยาวเอนเอียงและตัวแปรสุ่มแบบเดิมในบทความเกี่ยวกับการทดสอบอายุการใช้งานด้วยการวัดความเชื่อถือได้ Akman และ Gupta [13] เปรียบเทียบตัวประมาณที่ไม่เอนเอียงที่มีความแปรปรวนต่ำสุด (Uniformly Minimum Variance Unbiased Estimator; UMVUE) และตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator; MLE) ของพารามิเตอร์ μ เมื่อข้อมูลมีการแจกแจงอินเวอร์สเกาส์เซียน $IG(\mu, c\mu^2)$ และความยาวเอนเอียงอินเวอร์สเกาส์เซียน $LBIG(\mu, c\mu^2)$ Gupta และ Akman [14] นำผลลัพธ์บางส่วนจาก Sen [11] มาใช้เพื่อพัฒนาช่วงความเชื่อมั่นและการทดสอบเกี่ยวกับค่าเฉลี่ยและสัมประสิทธิ์การแปรผันของการแจกแจงอินเวอร์สเกาส์เซียนบนพื้นฐานข้อมูลของความยาวเอนเอียง นักวิจัยอื่นๆ ได้ทำการศึกษาเกี่ยวกับการแจกแจงความยาวเอนเอียง ได้แก่ Ahmed และ Abouammoh [15] Vardi และ Cox [16] เป็นต้น

จากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนไม่พบผลงานใดที่เกี่ยวข้องกับบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนที่มีพารามิเตอร์ μ และ λ ดังนั้นในงานวิจัยนี้ผู้ทำการวิจัยจึงได้สนใจศึกษาบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับที่เหมาะสมสำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนที่มีพารามิเตอร์ μ และ λ โดยมีวัตถุประสงค์เพื่อ 1) เพื่อสร้างตัวประมาณที่มีภาวะความน่าจะเป็นสูงสุดของพารามิเตอร์ μ และ λ โดยใช้วิธีภาวะความน่าจะเป็นสูงสุด 2) เพื่อสร้างบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน และ 3) เพื่อหาค่าความน่าจะเป็นค้ำมรวม (Coverage Probability) ของช่วงความเชื่อมั่นที่นำเสนอนี้เปรียบเทียบกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและผู้วิจัยได้กำหนดขอบเขตการวิจัยคือ 1) สร้างเลขสุ่มที่มีการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนที่มีขนาดตัวอย่าง n มีค่าเท่ากับ 10, 20, 30, 50, 60, 100, 500,

1,000 มีพารามิเตอร์ $\mu = 1$ และ $\lambda = 1, 3, 5, 10, 15, 20$ และ 2) ทำการจำลองซ้ำๆ กัน 10,000 รอบ ในแต่ละกรณี ด้วยวิธีมอนติคาร์โล (Monte Carlo) โดยใช้โปรแกรม R เวอร์ชัน 3.4.3

2. วิธีการดำเนินงาน

2.1 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน และการสร้างบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับ

2.2 ประเมินค่าพารามิเตอร์ μ และ λ ด้วยวิธีภาวะความน่าจะเป็นสูงสุดและหาสูตรอย่างง่ายสำหรับบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับ

2.3 สร้างเลขสุ่มที่มีการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนด้วยวิธีปฏิเสธ-ยอมรับ

2.4 ทำการศึกษาโดยการจำลองเพื่อเปรียบเทียบค่าประมาณความน่าจะเป็นคุ่มรวมของช่วงความเชื่อมั่นที่นำเสนอกับสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

2.5 วิเคราะห์ผลการวิจัย

2.6 สรุปผล

3. ผลการวิจัย

3.1 ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน

ถ้าให้ตัวแปรสุ่ม Y มีการแจกแจงอินเวอร์สเกาส์เซียน จะได้ว่า $\frac{\mu^2}{Y}$ มีการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน ดังนั้นอาจกล่าวได้ว่าการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนเป็นส่วนกลับของการแจกแจงอินเวอร์สเกาส์เซียน (Complementary Reciprocal) นั่นเอง โดยมีพารามิเตอร์พื้นฐานคือ μ และ λ ซึ่งการแจกแจงนี้อยู่ในกลุ่มแจกแจงความน่าจะเป็นต่อเนื่องที่ตัวแปรสุ่มมีค่าตั้งแต่ 0 ถึง ∞ ฟังก์ชันความหนาแน่นความน่าจะเป็นสำหรับตัวแปรสุ่ม X ที่มีการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน คือ

$$f_{LBIG}(x, \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x}} \mu^{-1} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}; x > 0$$

โดยที่ $\mu > 0$ เป็นพารามิเตอร์แสดงตำแหน่ง (Location Parameter) หรือพารามิเตอร์ค่าเฉลี่ย (Mean Parameter) และ $\lambda > 0$ เป็นพารามิเตอร์แสดงมาตราส่วน (Shape Parameter) นอกจากนี้ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนมีความสัมพันธ์กับฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงอินเวอร์สเกาส์เซียน ดังสมการ

$$f_{LBIG}(x, \mu, \lambda) = \frac{x \cdot f_{IG}(x, \mu, \lambda)}{E(X)}; x > 0$$

โดยที่ $E(X)$ เป็นโมเมนต์ที่ 1 ของการแจกแจงอินเวอร์สเกาส์เซียน คือ $E(X) = \mu$

3.2 สมการภาวะน่าจะเป็นสูงสุด

เมื่อ X_1, X_2, \dots, X_n เป็นตัวแปรสุ่มจากประชากรที่มีการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน $LBIG(\mu, \lambda)$ จะได้ฟังก์ชันภาวะน่าจะเป็นของพารามิเตอร์ μ และ λ คือ

$$L(\mu, \lambda; x) = \left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} \left(\prod_{i=1}^n (x_i)^{-\frac{1}{2}}\right) \mu^{-n} \exp\left[-\lambda \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\mu^2 x_i}\right]$$

, $\mu > 0, \lambda > 0$.

และจะได้ลอการิทึมของฟังก์ชันภาวะน่าจะเป็นดังนี้

$$\begin{aligned} \ln L(\mu, \lambda; X) &= \frac{n}{2} \ln\left(\frac{\lambda}{2\pi}\right) - \frac{1}{2} \sum_{i=1}^n \ln(x_i) - n \ln(\mu) \\ &\quad - \lambda \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\mu^2 x_i} \\ &= \frac{n}{2} \ln\left(\frac{\lambda}{2\pi}\right) - \frac{1}{2} \sum_{i=1}^n \ln(x_i) - n \ln(\mu) \\ &\quad - \frac{\lambda}{2\mu^2} \sum_{i=1}^n x_i + \frac{n\lambda}{\mu} - \frac{\lambda}{2} \sum_{i=1}^n \left(\frac{1}{x}\right) \end{aligned}$$

จาก

$$\begin{aligned} \frac{\partial \ln L(\mu, \lambda; x)}{\partial \mu} &= -\frac{n}{\mu} + \frac{\lambda}{\mu^3} \sum_{i=1}^n x_i - \frac{n\lambda}{\mu^2} = 0 \\ &= -\mu^2 - \mu\lambda + \lambda \bar{X} = 0 \end{aligned}$$

เมื่อ $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ จะได้



$$\hat{\mu} = - \left[\frac{\lambda \pm \sqrt{(-\lambda)^2 + 4(\lambda \bar{X})}}{2} \right]$$

ซึ่งพารามิเตอร์ μ ต้องมีค่ามากกว่า 0 ดังนั้นตัวประมาณที่มีภาวที่น่าจะเป็นสูงสุดของพารามิเตอร์ μ คือ

$$\hat{\mu}^{(MLE)} = \frac{-\hat{\lambda}^{(MLE)} + \sqrt{\hat{\lambda}^{(MLE)2} + 4\hat{\lambda}^{(MLE)}\bar{X}}}{2}$$

และเนื่องจากไม่สามารถหาตัวประมาณที่มีภาวที่น่าจะเป็นสูงสุดของพารามิเตอร์ λ ในรูปแบบสำเร็จได้จึงใช้ระเบียบวิธีการทำซ้ำเข้ามาช่วย ซึ่งในงานวิจัยชิ้นนี้ใช้การแก้ปัญหาหนึ่งมิติ ด้วยฟังก์ชัน Optimise ในโปรแกรม R โดยพิจารณาฟังก์ชัน $g(\lambda)$

$$\begin{aligned} g(\lambda) &= \left(\frac{n}{2}\right) \ln\left(\frac{\lambda}{2\pi}\right) - n \ln(\mu) - \lambda \sum_1^n \frac{(x_i - \mu)^2}{2\mu^2 x_i} \\ &= \left(\frac{n}{2}\right) \ln\left(\frac{\lambda}{2\pi}\right) - n \ln\left(\frac{-\lambda + \sqrt{\lambda^2 + 4\lambda\bar{X}}}{2}\right) \\ &\quad - \lambda \sum_1^n \frac{\left(x_i - \left(\frac{-\lambda + \sqrt{\lambda^2 + 4\lambda\bar{X}}}{2}\right)\right)^2}{2\left(\frac{-\lambda + \sqrt{\lambda^2 + 4\lambda\bar{X}}}{2}\right)^2 x_i} \end{aligned}$$

(สามารถศึกษาการใช้คำสั่ง Optimize ในการหาตัวประมาณที่มีภาวที่น่าจะเป็นสูงสุดเพิ่มเติมได้จาก Robert และ Casella [17])

3.3 เมทริกซ์สารสนเทศของฟิชเชอร์สำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน

ให้ตัวแปรสุ่ม X มีการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนที่มีพารามิเตอร์ (μ, λ) เมทริกซ์สารสนเทศของฟิชเชอร์สำหรับ θ เมื่อ $\theta = (\mu, \lambda)$ เป็นเวกเตอร์ 2 มิติของพารามิเตอร์เขียนแทนด้วยสัญลักษณ์ $I(\theta)$ เป็นดังนี้

$$I(\theta) = I(\mu, \lambda) = -E \begin{bmatrix} \frac{\partial^2}{\partial \mu^2} \ln f(x, \mu, \lambda) & \frac{\partial^2}{\partial \mu \partial \lambda} \ln f(x, \mu, \lambda) \\ \frac{\partial^2}{\partial \lambda \partial \mu} \ln f(x, \mu, \lambda) & \frac{\partial^2}{\partial \lambda^2} \ln f(x, \mu, \lambda) \end{bmatrix}$$

จาก Jorgensen *et al.* [18] ได้นำเสนอการแจกแจงแบบผสมสำหรับการแจกแจงอินเวอร์สเกาส์เซียน พร้อมทั้งหาคุณสมบัติต่างๆ ของการแจกแจงที่นำเสนอ ซึ่งการแจกแจงนี้เป็นการผสมระหว่างการแจกแจงอินเวอร์สเกาส์เซียนและการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน ดังนั้นเมื่อกำหนดพารามิเตอร์ของการแจกแจงนี้เป็น $\sigma^2 = \frac{1}{\lambda}$, $p = 1$ จะได้การแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนอาศัยความสัมพันธ์นี้จะได้ $E(X)$ ของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน คือ

$$E(X_{LBIG}) = \mu + \frac{\mu^2}{\lambda}$$

จากลอการิทึมของฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน

$$\begin{aligned} \ln f(\mu, \lambda; X) &= \ln\left((2\pi x)^{\frac{1}{2}}\right) - \ln(\mu) - \ln(\lambda^{\frac{1}{2}}) \\ &\quad - \frac{\lambda x}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2x} \end{aligned}$$

จะได้

$$\begin{aligned} E \left[\frac{\partial^2}{\partial \mu^2} \ln f(x; \mu, \lambda) \right] &= E \left[- \frac{\lambda \left(3x - 2\mu - \mu^2 \left(\frac{1}{\lambda} \right) \right)}{\mu^4} \right] \\ &= - \left[\frac{\lambda \left(3E[x] - 2\mu - \mu^2 \left(\frac{1}{\lambda} \right) \right)}{\mu^4} \right] \\ &= - \left[\frac{2\mu^2 + \lambda\mu}{\mu^4} \right] \\ &= - \left[\frac{2\mu + \lambda}{\mu^3} \right] \end{aligned}$$

$$\begin{aligned} E \left[\frac{\partial^2}{\partial \lambda^2} \ln f(x; \mu, \lambda) \right] &= E \left[- \frac{1}{2\lambda^2} \right] \\ &= - \frac{1}{2\lambda^2} \end{aligned}$$

$$\begin{aligned}
 E\left[\frac{\partial^2}{\partial\mu\partial\lambda}\ln f(x;\mu,\lambda)\right] &= E\left[\frac{x-\mu}{\mu^3}\right] \\
 &= \left[\frac{E[x]-\mu}{\mu^3}\right] \\
 &= \left[\frac{\mu+\frac{\mu^2}{\lambda}-\mu}{\mu^3}\right] \\
 &= \frac{1}{\mu\lambda} \\
 &= \frac{1}{2n^2\lambda+2n^2\mu-n^2\mu} \begin{bmatrix} \frac{n}{\lambda^2} & \frac{n}{\mu\lambda} \\ \frac{n}{\mu\lambda} & \frac{2n(\lambda+\mu)}{\mu^3} \end{bmatrix} \\
 &= \frac{\mu^3\lambda^2}{n^2(2\lambda+\mu)} \begin{bmatrix} \frac{n}{\lambda^2} & \frac{n}{\mu\lambda} \\ \frac{n}{\mu\lambda} & \frac{2n(\lambda+\mu)}{\mu^3} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\mu^3}{n(2\lambda+\mu)} & \frac{\mu^2\lambda}{n(2\lambda+\mu)} \\ \frac{\mu^2\lambda}{n(2\lambda+\mu)} & \frac{2\lambda^2(\lambda+\mu)}{n(2\lambda+\mu)} \end{bmatrix}
 \end{aligned}$$

ดังนั้นเมทริกซ์สารสนเทศของพิชเชอร์สำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนคือ

$$I(\theta) = I(\mu, \lambda) = \begin{bmatrix} \frac{2\mu+\lambda}{\mu^3} & -\frac{1}{\mu\lambda} \\ -\frac{1}{\mu\lambda} & \frac{1}{2\lambda^2} \end{bmatrix}$$

เมื่อ θ เป็นเวกเตอร์ 2 มิติของพารามิเตอร์และ $\theta = (\mu, \lambda)$ จะได้เมทริกซ์สารสนเทศของพิชเชอร์สำหรับตัวอย่างขนาด n คือ

$$I_n(\theta) = I_n(\mu, \lambda) = \begin{bmatrix} \frac{(2\mu+\lambda)n}{\mu^3} & -\frac{n}{\mu\lambda} \\ -\frac{n}{\mu\lambda} & \frac{n}{2\lambda^2} \end{bmatrix}$$

3.4 เมทริกซ์ความแปรปรวนร่วม

เมทริกซ์ความแปรปรวนร่วม ซึ่งมีค่าเท่ากับอินเวอร์สของเมทริกซ์สารสนเทศของพิชเชอร์ แทนด้วยสัญลักษณ์ $\Lambda = I_n^{-1}(\theta)$ ดังนั้นเมื่อนั้น n เข้าสู่อินฟินิตี้จะได้

$$\begin{aligned}
 I_n^{-1}(\mu, \lambda) &= \frac{1}{\left(\frac{2n(\lambda+\mu)}{\mu^3}\right)\left(\frac{n}{\lambda^2}\right) - \left(-\frac{n}{\mu\lambda}\right)\left(-\frac{n}{\mu\lambda}\right)} \begin{bmatrix} \frac{n}{\lambda^2} & \frac{n}{\mu\lambda} \\ \frac{n}{\mu\lambda} & \frac{2n(\lambda+\mu)}{\mu^3} \end{bmatrix} \\
 &= \frac{1}{\left(\frac{2n^2(\lambda+\mu)}{\mu^3\lambda^2}\right) - \left(-\frac{n}{\mu\lambda}\right)^2} \begin{bmatrix} \frac{n}{\lambda^2} & \frac{n}{\mu\lambda} \\ \frac{n}{\mu\lambda} & \frac{2n(\lambda+\mu)}{\mu^3} \end{bmatrix}
 \end{aligned}$$

3.5 การแจกแจงเชิงเส้นกำกับ

พิจารณาลำดับของตัวแปรสุ่มจากภาวะความน่าจะเป็นสูงสุดของ $\hat{\theta}_n^{(MLE)}$ โดยใช้ทฤษฎี Delta จะได้ว่า

$$\sqrt{n}(\hat{\theta}_n^{(MLE)} - \theta) \xrightarrow{d} X \sim N_2(0, I^{-1}(\theta))$$

$$\frac{\sqrt{n}(\hat{\theta}_n^{(MLE)} - \theta)}{\sqrt{I^{-1}(\theta)}} \xrightarrow{d} Z \sim N_2(0, I_2)$$

$$\frac{(\hat{\theta}_n^{(MLE)} - \theta)}{\sqrt{I_n^{-1}(\theta)}} \xrightarrow{d} Z \sim N_2(0, I_2)$$

จากอินเวอร์สของเมทริกซ์สารสนเทศของพิชเชอร์มีค่าเท่ากับเมทริกซ์ความแปรปรวนร่วมจะได้

$$\frac{(\hat{\theta}_n^{(MLE)} - \theta)}{\sqrt{\Lambda}} \xrightarrow{d} N_2(0, I_2)$$

ดังนั้น

$$\left(\Lambda^{\frac{1}{2}}\right)^{-1} (\hat{\theta}_n^{(MLE)} - \theta) \xrightarrow{d} Z \sim N_2(0, I_2),$$



โดยที่ $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ เป็นเมทริกซ์เอกลักษณะ และถ้า Z_1, Z_2

เป็นตัวแปรสุ่มที่มีการแจกแจงปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 และเป็นอิสระกัน จะได้ว่า $Z' = (Z_1, Z_2)$ มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal Distribution) $N(0, I_2)$ ดังนั้น

$$\begin{aligned} & (\hat{\theta}_n^{(MLE)} - \theta)' (\Lambda)^{-\frac{1}{2}} (\Lambda)^{-\frac{1}{2}} (\hat{\theta}_n^{(MLE)} - \theta) \\ &= (\hat{\theta}_n^{(MLE)} - \theta)' \Lambda^{-1} (\hat{\theta}_n^{(MLE)} - \theta) \\ &= Z'Z = Z_1^2 + Z_2^2 = \sum_{i=1}^2 Z_i^2 \end{aligned}$$

ซึ่งมีการแจกแจงโคสแควร์ที่มีองศาความเป็นอิสระเท่ากับ $2 (\chi_2^2)$

3.6 ขอบเขตความเชื่อมั่นสำหรับพารามิเตอร์ของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน

จาก Cramér [19], Chew [20], Duangchana และ Budsaba [21] จะได้ว่าบริเวณความเชื่อมั่น $100(1 - \alpha)\%$ สำหรับพารามิเตอร์ θ ของการแจกแจงปกติใน 2 มิติ (Two-Dimensional Normal Distribution) คือวงรี ดังสมการ

$$(\theta_n^{(MLE)} - \theta)' \Lambda^{-1} (\hat{\theta}_n^{(MLE)} - \theta) \leq \chi_2^2(\alpha)$$

เนื่องจากการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนมีอินเวอร์สของเมทริกซ์ความแปรปรวนร่วมคือ

$$\Lambda^{-1} = I_n(\theta) = I_n(\mu, \lambda) = \begin{bmatrix} \frac{(2\mu + \lambda)n}{\mu^3} & -\frac{n}{\mu\lambda} \\ -\frac{n}{\mu\lambda} & \frac{n}{2\lambda^2} \end{bmatrix}$$

จะได้ขอบเขตความเชื่อมั่น $100(1 - \alpha)\%$ สำหรับ พารามิเตอร์ $\theta' = (\mu, \lambda)$ ของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียน คือ

$$\begin{bmatrix} \hat{\mu}_n^{(MLE)} - \mu & \hat{\lambda}_n^{(MLE)} - \lambda \end{bmatrix} \begin{bmatrix} \frac{n(2\mu + \lambda)}{\mu^3} & -\frac{n}{\mu\lambda} \\ -\frac{n}{\mu\lambda} & \frac{n}{2\lambda^2} \end{bmatrix} \begin{bmatrix} \hat{\mu}_n^{(MLE)} - \mu \\ \hat{\lambda}_n^{(MLE)} - \lambda \end{bmatrix} \leq \chi_{(2)}^2(\alpha)$$

ซึ่งจัดรูปใหม่ได้

$$\begin{aligned} & \frac{n(2\mu + \lambda)}{\mu^3} (\hat{\mu}_n^{(MLE)} - \mu)^2 - \frac{2n}{\mu\lambda} (\hat{\mu}_n^{(MLE)} - \mu)(\hat{\lambda}_n^{(MLE)} - \lambda) \\ &+ \frac{n}{2\lambda^2} (\hat{\lambda}_n^{(MLE)} - \lambda)^2 \leq \chi_{(2)}^2(\alpha) \end{aligned}$$

3.7 ผลการศึกษาเชิงจำลอง

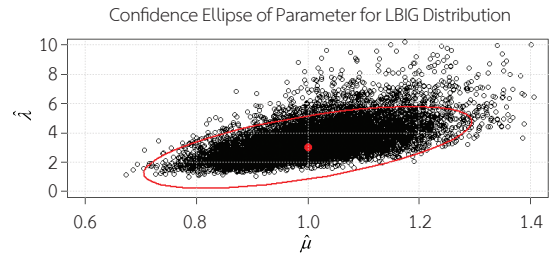
ในหัวข้อนี้จะศึกษาความน่าจะเป็นคัมรวมของบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนเมื่อกำหนดพารามิเตอร์ $\mu = 1, \lambda = 1, 3, 5, 10, 15, 20$ โดยจะทำการจำลองเลขสุ่มของการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนด้วยวิธีปฏิเสธ-ยอมรับ ที่ขนาดตัวอย่าง $n = 10, 20, 30, 50, 60, 100, 500, 1,000$ และนำตัวเลขสุ่มที่ได้ไปประมาณค่าพารามิเตอร์ μ และ λ ด้วยวิธีภาวะความน่าจะเป็นสูงสุดทำการจำลองซ้ำๆ กัน 10,000 รอบ ในแต่ละกรณี จากนั้นทำการสร้างบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับ และคำนวณหาค่าความน่าจะเป็นคัมรวมของบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับกับสัมประสิทธิ์ความเชื่อมั่นที่ 0.98 โดยใช้โปรแกรม R เวอร์ชัน 3.4.3 ได้ผลลัพธ์ดังตารางที่ 1 และรูปที่ 1-4 (ในที่นี้จะแสดงเฉพาะบางกรณี)

4. อภิปรายและสรุป

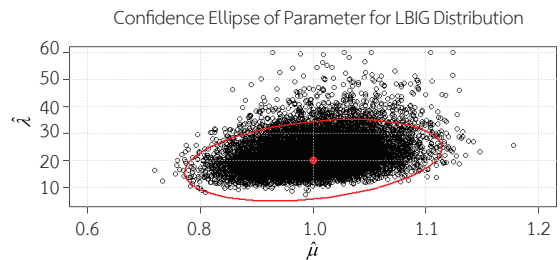
จากผลการศึกษาเชิงจำลองในตารางที่ 1 พบว่าค่าประมาณของตัวประมาณพารามิเตอร์ μ และ λ ที่นำเสนอมีค่ามากกว่าค่าจริง (Overestimate) และจากรูปที่ 1: $n = 30, \mu = 1, \lambda = 3$ จะเห็นว่ามีจำนวนค่าประมาณพารามิเตอร์ที่หลุดออกนอกบริเวณวงรีจำนวนมาก แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 500 ดังรูปที่ 3 จำนวนค่าประมาณพารามิเตอร์ที่หลุดออกนอกบริเวณวงรีจะน้อยลง ในทำนองเดียวกันจากรูปที่ 2: $n = 30, \mu = 1, \lambda = 20$ พบว่ามีจำนวนค่าประมาณพารามิเตอร์ที่หลุดออกนอกบริเวณวงรีจำนวนมาก แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 1,000 ดังรูปที่ 4 จำนวนค่าประมาณพารามิเตอร์ที่หลุดออกนอกบริเวณวงรีลดลงอย่างเห็นได้ชัด นั่นแสดงให้เห็นว่าความน่าจะเป็นคัมรวมของบริเวณวงรี

ตารางที่ 1 ขนาดตัวอย่าง n , ค่าประมาณพารามิเตอร์ μ , ค่าประมาณพารามิเตอร์ λ และค่าความน่าจะเป็นคัมรวม

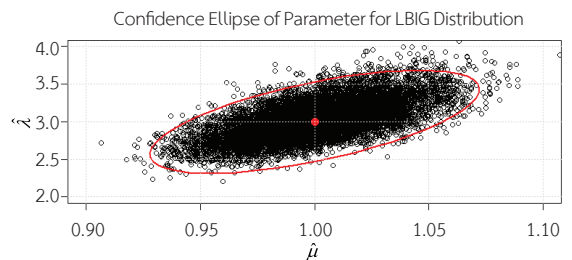
| n | μ | λ | ค่าประมาณพารามิเตอร์ | | CP |
|-------|-------|-----------|----------------------|-----------------|-----------------|
| | | | $\hat{\mu}$ | $\hat{\lambda}$ | |
| 10 | 1 | 1 | 1.087159 | 1.966299 | 0.711300 |
| | | 3 | 1.034578 | 4.703122 | 0.796900 |
| | | 5 | 1.017337 | 7.409270 | 0.822400 |
| | | 10 | 1.003813 | 11.274768 | 0.817300 |
| | | 15 | 1.008734 | 14.603413 | 0.841800 |
| | | 20 | 1.006203 | 28.972971 | 0.843900 |
| 20 | 1 | 1 | 1.033106 | 1.430457 | 0.767600 |
| | | 3 | 1.016462 | 3.720606 | 0.862100 |
| | | 5 | 1.009235 | 6.050067 | 0.883100 |
| | | 10 | 1.004174 | 11.935709 | 0.897500 |
| | | 15 | 1.00373 | 17.93545 | 0.90390 |
| | | 20 | 1.002776 | 23.646658 | 0.908700 |
| 30 | 1 | 1 | 1.019858 | 1.285810 | 0.7862 |
| | | 3 | 1.010831 | 3.470861 | 0.8849 |
| | | 5 | 1.006117 | 5.674344 | 0.9123 |
| | | 10 | 1.002791 | 11.242765 | 0.9189 |
| | | 15 | 1.002793 | 16.788570 | 0.9301 |
| | | 20 | 1.001823 | 22.447316 | 0.9299 |
| 50 | 1 | 1 | 1.005362 | 1.183575 | 0.802900 |
| | | 3 | 1.006037 | 3.255717 | 0.909700 |
| | | 5 | 1.002652 | 5.044425 | 0.926800 |
| | | 10 | 1.002755 | 10.727601 | 0.937400 |
| | | 15 | 1.001092 | 16.011761 | 0.942900 |
| | | 20 | 1.000489 | 21.319404 | 0.944600 |
| 60 | 1 | 1 | 1.002595 | 1.162293 | 0.805400 |
| | | 3 | 1.004304 | 3.205111 | 0.917100 |
| | | 5 | 1.004688 | 5.320601 | 0.928600 |
| | | 10 | 1.001318 | 10.587047 | 0.943700 |
| | | 15 | 1.001164 | 15.052080 | 0.944900 |
| | | 20 | 1.001387 | 21.155376 | 0.952000 |
| 100 | 1 | 1 | 0.996167 | 1.114048 | 0.8086 |
| | | 3 | 1.003798 | 3.141008 | 0.9237 |
| | | 5 | 1.001292 | 5.177069 | 0.9403 |
| | | 10 | 1.001075 | 10.328870 | 0.9562 |
| | | 15 | 1.000450 | 15.460640 | 0.9604 |
| | | 20 | 1.000879 | 20.595330 | 0.9629 |
| 500 | 1 | 1 | 0.989934 | 1.068969 | 0.7608 |
| | | 3 | 1.000519 | 3.030926 | 0.9325 |
| | | 5 | 1.000859 | 5.044425 | 0.9539 |
| | | 10 | 1.000139 | 10.070449 | 0.9670 |
| | | 15 | 1.000296 | 15.112562 | 0.9698 |
| | | 20 | 1.000110 | 20.142080 | 0.9713 |
| 1,000 | 1 | 1 | 0.988601 | 1.061008 | 0.6785 |
| | | 3 | 1.000226 | 3.019076 | 0.9384 |
| | | 5 | 0.999963 | 5.019547 | 0.9571 |
| | | 10 | 0.999996 | 10.023282 | 0.9684 |
| | | 15 | 1.000062 | 15.052080 | 0.9710 |
| | | 20 | 1.000144 | 20.080555 | 0.9739 |



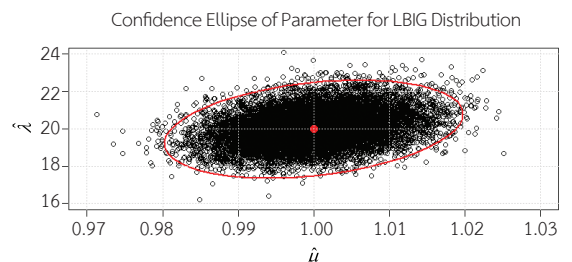
รูปที่ 1 บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับกรณีที่ $n = 30, \mu = 1, \lambda = 3$



รูปที่ 2 บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับกรณีที่ $n = 30, \mu = 1, \lambda = 20$



รูปที่ 3 บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับกรณีที่ $n = 500, \mu = 1, \lambda = 3$



รูปที่ 4 บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับกรณีที่ $n = 1,000, \mu = 1, \lambda = 20$

ความเชื่อมั่นเชิงเส้นกำกับจะมีค่าเพิ่มขึ้นเมื่อขนาดตัวอย่าง n เพิ่มขึ้น โดยที่ค่าความน่าจะเป็นคัมรวมจะมีค่าเข้าใกล้สัมประสิทธิ์ของช่วงความเชื่อมั่น 0.98 ที่ขนาดตัวอย่าง $n = 1,000$ ซึ่งให้ค่าความน่าจะเป็นคัมรวมสูงสุด ทำให้ทราบว่าค่าความน่าจะเป็นคัมรวมของบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับสำหรับการแจกแจงความยาวเอนเอียงอินเวอร์สเกาส์เซียนจะมีค่าสูงขึ้นเมื่อขนาดตัวอย่างมีขนาดใหญ่

นอกจากนี้ค่าพารามิเตอร์ λ ที่แตกต่างกันส่งผลทำให้ค่าความน่าจะเป็นคัมรวมของบริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับแตกต่างกัน และเมื่อเพิ่มค่าพารามิเตอร์ λ ก็จะส่งผลให้ค่าความน่าจะเป็นคัมรวมมีค่าเพื่อขึ้นด้วย ดังจะเห็นได้ว่าที่ค่าพารามิเตอร์ $\lambda = 20$ ให้ค่าความน่าจะเป็นคัมรวมสูงสุดที่ขนาดตัวอย่าง $n = 1,000$ โดยมีค่าความน่าจะเป็นคัมรวมเท่ากับ 0.9739 (แสดงดังรูปที่ 4)

ข้อเสนอแนะสำหรับงานวิจัยครั้งต่อไปควรศึกษาการประยุกต์ใช้บริเวณวงรีความเชื่อมั่นเชิงเส้นกำกับกับข้อมูลจริง

5. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนการวิจัยจากโครงการทุนวิจัยรุ่นใหม่ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ประจำปี พ.ศ. 2561 สัญญารับทุนเลขที่ 6145111 และขอขอบคุณ คุณภากร พานแก้ว คุณดวงใจ พงศ์จันทิก และคุณธันยาพร ชาญบุญยวิทย์ ผู้ช่วยวิจัยทั้ง 3 คน

เอกสารอ้างอิง

- [1] J.L.Folks and R.S.Chhikara, *The Inverse Gaussian Distribution*. New York: Marcel Dekker, 1989.
- [2] M. Ahsanllah and S. Kirmani, "A characterization of the wald distribution," *Naral Research Logistics Quarterly*, vol. 31, pp. 155–158, 1984.
- [3] R. Khattree, "Characterization of inverse-gaussian and gamma distributions through their length-biased distributions," *IEEE Transactions on Reliability*, vol. 38, pp. 610–611, 1989.
- [4] G. Patil and C. Rao, "Weighted distributions: A survey of their applications," in *Applications of statistics*, Amsterdam, Netherlands: North Holland Publishing Company, 1977, pp. 383–405.
- [5] G. Patil and C. Rao, "Weighted distributions and size-biased sampling with applications to wildlife populations and human families," *BIOMETRICS*, vol. 34, pp. 179–189, 1978.
- [6] S. Blumenthal, "Proportional sampling in life length studies," *Technometrics*, vol. 9, no. 2, pp. 205–218, 1967.
- [7] R. L. Scheaffer, "Size-biased sampling," *Technometrics*, vol. 14, no. 3, pp. 635–644, 1972.
- [8] R. Simon, "Length biased sampling in etiologic studies," *American Journal of Epidemiology*, vol. 111, no. 4, pp. 444–452, 1980.
- [9] A. Cnaan, "Survival models with two phases and length biased sampling," *Communications in Statistics-Theory and Methods*, vol. 14, no. 4, pp. 861–886, 1985.
- [10] R. C. Gupta and S. Tripathi, "A comparison between the ordinary and length biased modified power series distributions with applications," *Communication s in Statistics*, vol. 16, no. 4, pp. 1195–1206, 1987.
- [11] P. K. Sen, "What do the arithmetic, geometric and harmonic means tell us in length-biased sampling?," *Statistics and Probability Letters*, vol. 5, pp. 95–98, 1987.
- [12] R. C. Gupta and S. Kirmani, "On order relations between reliability measures of two distributions," *Stochastic Models*, vol. 1, no. 3, pp. 149–156, 1987.
- [13] O. Akman and R. C. Gupta, "A comparison of various estimators of the mean of an inverse Gaussian distribution," *Journal of Statistical*



- Computation and Simulation*, vol. 40, no. 1–2, pp. 71–81, 1991.
- [14] R. C. Gupta and O. Akman, “Statistical inference based on the length biased data for the inverse Gaussian distribution,” *A Journal of Theoretical and Applied Statistics*, vol. 31, no. 4, pp. 325–337, 1998.
- [15] A. Ahmed and A. Abouammoh, “Characterization of gamma, inverse Gaussian, and negative binomial distributions via their length-biased distributions,” *Statistical Paper*, vol. 34, no. 1, pp. 167–173, 1993.
- [16] Y. Vardi, “Nonparametric estimation in the presence of length bias,” *Annals of Statistics*, vol. 10, no. 1, pp. 616–620, 1982.
- [17] C. R. Robert and G. Casella, “Monte Carlo Optimization,” in *Introducing Monte Carlo Methods with R*, New York : Springer, 2009, pp. 125–165.
- [18] B. Jorgensen, V. Seshadri, and G. A. Whitmore, “On the mixture of the inverse Gaussian distribution with its complementary reciprocal,” *Scandinavian Journal of Statistics*, vol. 18, no. 1, pp. 77–89, 1991.
- [19] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ, US: Princeton University Press, 1946.
- [20] V. Chew, “Confidence, prediction, and tolerance regions for the multivariate normal distribution,” *Journal of the American Statistical Association*, vol. 61, no. 315, pp. 605–617, 1966.
- [21] N. Duangchana and K. Budsaba, “Asymptotic confidence ellipses of parameters for the inverse Gaussian distribution,” *Thammasat International Journal of Science and Technology*, vol. 19, no. 2, pp. 22–29, 2014.