

การแจกแจงปัวซองพาราลอจิสติก: สมบัติและการประยุกต์

The Poisson-Paralogistic Distribution: Properties and Applications

บุญฤทธิ์ ชูประดิษฐ์¹ และ สิริทิพ วะสินรัตน์^{2*}

Boonyarit Choopradit¹ and Sirithip Wasinrat^{2*}

¹ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ประเทศไทย

²คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏจันทรเกษม ประเทศไทย

¹Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Thailand

²Faculty of Science, Chandrakasem Rajabhat University, Thailand

บทคัดย่อ

งานวิจัยนี้เสนอการแจกแจงผสมระหว่างการแจกแจงพาราลอจิสติกกับการแจกแจงปัวซอง เรียกว่าการแจกแจงปัวซองพาราลอจิสติก ที่ประกอบด้วย 3 พารามิเตอร์ มีการศึกษาสมบัติเบื้องต้นของการแจกแจงปัวซองพาราลอจิสติกบางประการ ได้แก่ ตัวสถิติอันดับ ฟังก์ชันควอนไทล์ และมูลค่าเสี่ยงภัย การประมาณค่าพารามิเตอร์ของการแจกแจงด้วยวิธีภาวะน่าจะเป็นสูงสุด และประเมินประสิทธิภาพของการประมาณค่าพารามิเตอร์ของการแจกแจงใหม่ ด้วยการจำลองข้อมูลภายใต้สถานการณ์ต่าง ๆ พบว่าค่าประมาณพารามิเตอร์เฉลี่ยมีค่าใกล้เคียงกับค่าพารามิเตอร์จริง ค่าความเอนเอียง และรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าลดลง เมื่อขนาดตัวอย่างเพิ่มขึ้น และการเปรียบเทียบประสิทธิภาพของการแจกแจงใหม่ ในการประยุกต์ใช้กับข้อมูลจริง 3 ชุด พบว่าการแจกแจงปัวซองพาราลอจิสติกให้ค่าล็อกภาวะน่าจะเป็นสูงกว่าการแจกแจงพาราลอจิสติก การแจกแจงพาราลอจิสติกผกผัน และการแจกแจงล็อกพาราลอจิสติก

คำสำคัญ: การแจกแจงผสม, ฟังก์ชันควอนไทล์, วิธีภาวะน่าจะเป็นสูงสุด

ABSTRACT

In this paper, the Poisson-paralogistic distribution (PPD), a new distribution compounding the paralogistic and Poisson distributions, is proposed. Properties of the new distribution, such as order statistics, quantile function, and value-at-risk (VaR), are obtained. The maximum likelihood estimation (MLE) is then studied to obtain a parameter estimation. A simulation study shows that the average of PPD estimated parameters approximates the real parameters, bias values, and root mean square error value decrease when sample size increases. The proposed distribution is applied to three real datasets, potentially providing a better fit than other related paralogistic, inverse paralogistic, and log-paralogistic distributions, with higher log-likelihood values.

KEYWORDS: mixture distribution, quantile function, maximum likelihood method

*Corresponding Author: sirithip.w@chandra.ac.th

Received: 24/08/2022; Revised: 28/02/2023; Accepted: 10/04/2023

1. บทนำ

การศึกษาลักษณะการแจกแจงที่เหมาะสมกับข้อมูลสามารถพบได้ในหลากหลายศาสตร์ เช่น Khajorntaewawong et al. (2020) ได้ศึกษาการแจกแจงความล่าช้าของเที่ยวบินขาออก Choopradit et al. (2021) ศึกษาการแจกแจงค่าใช้จ่ายต่อหัวของนักท่องเที่ยวในกลุ่มจังหวัดภาคใต้ฝั่งอ่าวไทยก่อนสถานการณ์การระบาดของโควิด-19 ทั่วโลก และ Choopradit & Wasinrat (2021) การแจกแจงของจำนวนพายุหมุนเขตร้อนที่เคลื่อนเข้าสู่ประเทศไทย อย่างไรก็ตามในบางกรณีชุดข้อมูลไม่สามารถหาลักษณะการแจกแจงที่เหมาะสมได้นั้น

ในปัจจุบันมีการพัฒนาฟังก์ชันการแจกแจงความน่าจะเป็นใหม่ ๆ เพิ่มขึ้นมากมายอย่างต่อเนื่อง เพื่อให้สอดคล้องกับลักษณะของข้อมูลที่มีความหลากหลายมากขึ้นและสอดคล้องกับสถานการณ์ต่าง ๆ ที่เปลี่ยนแปลงไป ซึ่งวิธีที่จะพัฒนาการแจกแจงขึ้นมาใหม่วิธีหนึ่งคือ การผสมการแจกแจงที่มีอยู่แล้วเข้าด้วยกัน โดยมีผู้ศึกษาทั้งที่เป็นแบบการแจกแจงผสมอันตะ (Finite mixture distribution) (Balakrishnan et al., 2009; Erisoglu et al., 2013; Hall & Zhou, 2003; McLachlan & Peel, 2000) และ แบบการแจกแจงผสมอนันต์ (Infinite mixture distribution) (Bulmer, 1974; Emilio et al., 2008; Withers &

Nadarajah, 2011) จากงานวิจัยของ Rodrigues et al. (2009) ได้เสนอการรวมกันของตัวแบบการรอดชีพระยะยาว (Unification of long-term survival models) ซึ่งเป็นการผสมผสานกันระหว่างตัวแปรสุ่มจำนวนสาเหตุของการเกิดเหตุการณ์ที่ไม่ทราบค่า ที่มาจากการแจกแจงของตัวแปรสุ่มไม่ต่อเนื่องกับตัวแปรสุ่มระยะเวลาของการเกิดเหตุการณ์ที่เป็นตัวแปรสุ่มต่อเนื่อง หรือที่เรียกว่าตัวแบบการรักษาแบบผสม (Mixture cure model) โดย Wasinrat & Choopradit (2023) ได้นำตัวแบบของการรอดชีพระยะยาวนี้ไปพัฒนาเป็นการแจกแจงใหม่ โดยผสมการแจกแจงปัวซองและการแจกแจงพาราโตนกผกผัน เรียกว่าการแจกแจงปัวซองพาราโตนกผกผัน (Poisson inverse pareto distribution)

การแจกแจงพาราลอจิสติก (Paralogistic distribution: PD) เป็นฟังก์ชันความน่าจะเป็นแบบต่อเนื่อง ถูกนำเสนอครั้งแรกโดย McDonald (1984) ซึ่งพัฒนามาจากการแจกแจงเบอร์ (Burr distribution) กรณีที่พารามิเตอร์บ่งรูปร่าง (Shape parameter) ทั้งสองตัวมีค่าเท่ากัน (Kleiber & Kotz, 2003) สำหรับการประยุกต์ใช้การแจกแจงพาราลอจิสติกนั้น สามารถนำไปประยุกต์ใช้กับตัวแบบทางเศรษฐศาสตร์ การเงิน และทางด้านการประกันภัยอีกด้วย (Kleiber & Kotz, 2003; Klugman et al., 2019) ที่ผ่านมามีผู้วิจัยได้ขยายการแจกแจงพาราลอจิสติก เช่น งานวิจัยของ Idemudia & Ekhsosuehi (2019) ได้เสนอการแจกแจงพาราลอจิสติกใหม่ที่มี 3 พารามิเตอร์ โดยการเพิ่มพารามิเตอร์บ่งตำแหน่ง (Location parameter) ซึ่งเพิ่มความยืดหยุ่นในการประยุกต์ใช้กับชุดข้อมูลจริง โดยการเปรียบเทียบกับ การแจกแจงที่ใกล้เคียง ได้แก่ การแจกแจงพาราลอจิสติก การแจกแจงพาราลอจิสติกผกผัน การแจกแจงล็อกลอจิสติก การแจกแจงแกมมา พบว่าจากการทดสอบภาวะสารถูปติการแจกแจงพาราลอจิสติกใหม่ 3 พารามิเตอร์มีความเหมาะสมมากกว่าเมื่อเทียบกับการแจกแจงอื่น และงานวิจัยของ Bhati et al. (2019) ได้ขยายการแจกแจงพาราลอจิสติกที่ชื่อว่า การแจกแจงพาราลอจิสติก-เอ็มพีแอลจี (Paralogistic-MPLG distribution) ซึ่งเป็นการแจกแจงที่มี 4 พารามิเตอร์ จากการนำไปประยุกต์ใช้กับข้อมูลจริงเปรียบเทียบกับ การแจกแจงอื่น เช่น การแจกแจงไวบูล-เอ็มพีแอลจี การแจกแจงล็อกนอร์มัล-เอ็มพีแอลจี การแจกแจงไวบูล-พาราโตนกผกผัน การแจกแจงล็อกนอร์มัล-เอ็มพีแอลจี พบว่าการทดสอบภาวะสารถูปติการแจกแจงพาราลอจิสติก-เอ็มพีแอลจี มีความเหมาะสมเป็นลำดับที่ 2 เมื่อเทียบกับการแจกแจงอื่น เห็นได้ว่าการขยายการแจกแจงพาราลอจิสติกสามารถ

นำไปประยุกต์ใช้กับข้อมูลได้ดีเมื่อเทียบกับการแจกแจงอื่น อย่างไรก็ตามยังไม่มีการพัฒนาหรือการขยายของการแจกแจงพาราลอจิสติกที่กว้างขวางมากนัก

สำหรับการศึกษานี้ ผู้วิจัยได้พัฒนาการแจกแจงที่เป็นการผสมระหว่างการแจกแจงปัวซองกับการแจกแจงพาราลอจิสติกและนำเสนอเป็นการแจกแจงใหม่ พร้อมทั้งพิสูจน์คุณสมบัติเชิงความน่าจะเป็นของการแจกแจงดังกล่าวควบคู่กับการจำลองด้วยโปรแกรมคอมพิวเตอร์ ศึกษาวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีการทางค่าพารามิเตอร์ของการแจกแจงใหม่ด้วยการจำลองข้อมูลภายใต้สถานการณ์ต่าง ๆ นอกจากนี้ มีการประยุกต์การแจกแจงใหม่กับชุดข้อมูลจริง และเปรียบเทียบประสิทธิภาพของการแจกแจงใหม่กับการแจกแจงอื่น ๆ ที่เกี่ยวข้อง

2. วิธีการ

2.1 วิธีของการแจกแจงผสม

(Method of mixture distribution)

งานวิจัยนี้ได้ประยุกต์ใช้วิธีการผสมผสานของตัวแบบการรอดชีพระยะยาว (Unification of long-term survival models) ที่เสนอโดย Rodrigues et al. (2009) ซึ่งเป็นการผสมผสานกันระหว่างตัวแปรสุ่มจำนวนสาเหตุของการเกิดเหตุการณ์ที่ไม่ทราบค่ากับตัวแปรสุ่มเวลาของการเกิดเหตุการณ์ ดังนี้

ให้ N เป็นตัวแปรสุ่มที่ไม่ทราบค่าแทนจำนวนสาเหตุของการเกิดเหตุการณ์ที่สนใจด้วยการแจกแจงความน่าจะเป็น $p_n = P(N = n)$ โดยที่ $n = 0, 1, 2, 3, \dots$ และให้ $Z_i, i = 1, 2, \dots, n$ เป็นตัวแปรสุ่มแทนเวลาของการเกิดเหตุการณ์จากสาเหตุลำดับที่ i และ Z_i เป็นอิสระกับ N ตัวแบบการรอดชีพระยะยาว (Long-term survival function) ของตัวแปรสุ่ม X เขียนแทนด้วย $S_{LT}(x)$ กำหนดโดย

$$\begin{aligned} S_{LT}(x) &= P(N = 0) + P(Z_1 > x, Z_2 > x, \dots, Z_N > x, N \geq 1) \\ &= P(N = 0) + \sum_{n=1}^{\infty} P(N = n) P(Z_1 > x, Z_2 > x, \dots, Z_N > x) \\ &= p_0 + \sum_{n=1}^{\infty} p_n [S(x)]^n \\ &= \sum_{n=0}^{\infty} p_n [S(x)]^n \\ &= A[S(x)] \end{aligned} \quad (1)$$

โดย $A[-]$ คือ ฟังก์ชันก่อกำเนิดของลำดับ p_n และ $S(x)$ คือ ฟังก์ชันการรอดชีพที่ $0 \leq S(x) \leq 1$ ทั้งนี้ เนื่องจาก

$\lim_{x \rightarrow \infty} S_{LT}(x) = p_0$ ใน Rodrigues et al. (2009) ได้แสดง $S_{LT}(x)$ ในรูปของ

$$S_{LT}(x) = p_0 + (1 - p_0)S_M(x) \quad (2)$$

โดยที่ $S_M(x)$ คือ ฟังก์ชันการรอดชีพที่

$$S_M(x) = \frac{\sum_{n=1}^{\infty} p_n [S(x)]^n - p_0}{1 - p_0} \quad \text{และ} \quad \lim_{x \rightarrow \infty} S_M(x) = 0$$

จากสมการที่ (1) และ (2) Wasinrat & Choopradit (2023) ได้แสดงให้เห็นว่า

$$S_M(x) = \frac{A[S(x)] - p_0}{1 - p_0} \quad (3)$$

โดยในงานวิจัยนี้จะกล่าวถึงสมการที่ (3) ในฐานะที่เป็นฟังก์ชันการรอดชีพผสม (Mixed survival function) และด้วยความสัมพันธ์ของฟังก์ชันการรอดชีพและฟังก์ชันการแจกแจง จะนำไปสู่การสร้างเป็นฟังก์ชันการแจกแจงความน่าจะเป็นแบบผสม ซึ่งสอดคล้องกับงานวิจัยของ Wasinrat & Choopradit (2023) ที่ได้นำเสนอการแจกแจงปัวซองพาราลอจิสติก

2.2 การแจกแจงพาราลอจิสติก

ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงพาราลอจิสติก โดยมีฟังก์ชันการแจกแจงสะสม (Cumulative distribution function: cdf) ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function: pdf) และ ฟังก์ชันการรอดชีพ (Survival function) ตามลำดับดังนี้

$$\begin{aligned} F_{PD}(x) &= 1 - \left[\frac{1}{(x/\theta)^\alpha + 1} \right]^\alpha; \quad x > 0, \theta > 0, \alpha > 0 \\ f_{PD}(x) &= \frac{\alpha^2 (x/\theta)^\alpha}{x \left[(x/\theta)^\alpha + 1 \right]^{\alpha+1}}; \quad x > 0, \theta > 0, \alpha > 0 \\ S_{PD}(x) &= \left[\frac{1}{(x/\theta)^\alpha + 1} \right]^\alpha; \quad x > 0, \theta > 0, \alpha > 0 \end{aligned} \quad (4)$$

โดยที่ $\theta > 0$ คือพารามิเตอร์บ่งขนาด (Scale parameter) และ $\alpha > 0$ คือพารามิเตอร์บ่งรูปร่าง (Shape parameter) การแจกแจงพาราลอจิสติกมาจากการแจกแจงเบอร์ (Burr distribution) กรณีที่พารามิเตอร์บ่งรูปร่างทั้งสองตัวมีค่าเท่ากัน ในทำนองเดียวกันการแจกแจงพาราลอจิสติกผกผัน (Inverse paralogistic distribution: IPD) มาจากการแจกแจงเบอร์ผกผัน (Inverse Burr distribution) กรณีที่พารามิเตอร์บ่งรูปร่างทั้งสองตัวมีค่าเท่ากันเช่นกัน และการแจกแจงเบอร์กรณีพารามิเตอร์บ่งรูปร่างมีค่าเท่ากับ 1

จะเป็นการแจกแจงล็อกลอจิสติก (Log-logistic distribution: LLD) (Kleiber & Kotz, 2003) โดยมีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

การแจกแจงพาราลอจิสติกผกผัน

ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงพาราลอจิสติกผกผันโดยมีฟังก์ชันความหนาแน่นความน่าจะเป็น คือ

$$f_{IPD}(x) = \frac{\nu^2 (x/\mu)^{\nu^2}}{x \left[1 + (x/\mu)^\nu \right]^{\nu+1}}, \quad x > 0, \mu > 0, \nu > 0 \quad (5)$$

การแจกแจงล็อกลอจิสติก

ให้ X เป็นตัวแปรสุ่มที่มีการแจกแจงล็อกลอจิสติกโดยมีฟังก์ชันความหนาแน่นความน่าจะเป็น คือ

$$f_{LLD}(x) = \frac{\gamma (x/\kappa)^\gamma}{x \left[1 + (x/\kappa)^\gamma \right]^2}, \quad x > 0, \kappa > 0, \gamma > 0 \quad (6)$$

3. ผลการวิจัย

3.1 การแจกแจงปัวซองพาราลอจิสติก

(Poisson-paralogistic distribution: PPD)

ในขั้นนี้เสนอการแจกแจงผสมใหม่ของการแจกแจงปัวซองและการแจกแจงพาราลอจิสติก โดยใช้ฟังก์ชันการรอดชีพผสมจากสมการที่ (3)

บทนิยาม 1. ให้ $A_p(s)$ แทนฟังก์ชันก่อกำเนิดความน่าจะเป็น (Probability generating function: pgf) ของการแจกแจงปัวซองที่มีพารามิเตอร์ λ โดยที่

$$A_p(s) = \exp[-\lambda(1-s)] \quad (7)$$

โดยที่ $\lambda > 0$ และ $|s| \leq 1$

บทนิยาม 2. ให้ X เป็นตัวแปรสุ่มของการแจกแจงปัวซองพาราลอจิสติกที่มีพารามิเตอร์ λ, θ และ α เขียนแทนด้วย $X \sim PPD(\lambda, \theta, \alpha)$ โดยที่ $\lambda > 0, \theta > 0$ และ $\alpha > 0$

ทฤษฎีบท 1. ฟังก์ชันการรอดชีพของการแจกแจงปัวซองพาราลอจิสติก คือ

$$S_{PPD}(x) = \frac{\exp \left[-\lambda \left(1 - \left[\frac{1}{(x/\theta)^\alpha + 1} \right]^\alpha \right) \right] - \exp(-\lambda)}{1 - \exp(-\lambda)}, \quad x > 0, \lambda > 0, \theta > 0, \alpha > 0 \quad (8)$$

พิสูจน์ ให้ $A[S(x)]$ เป็น pgf ของการแจกแจงปัวซองตามสมการที่ (7) ฟังก์ชันการรอดชีพผสมเขียนได้ดังนี้

$$\begin{aligned} S_M(x) &= \frac{\exp[-\lambda(1-S(x))] - p_0}{1 - p_0} \\ &= \frac{\exp[-\lambda F(x)] - \exp(-\lambda)}{1 - \exp(-\lambda)} \end{aligned}$$

โดยที่ $F(x) = 1 - S(x)$ และ $p_0 = P(N=0) = \exp(-\lambda)$
 ดังนั้น แทน $F(x)$ ด้วย $F_{PPD}(x)$ ได้ฟังก์ชันการรอดชีพ
 ของการแจกแจงปัวซองพาราลอจิสติกคือ

$$S_{PPD}(x) = \frac{\exp[-\lambda F_{PPD}(x)] - \exp(-\lambda)}{1 - \exp(-\lambda)}$$

$$= \frac{\exp\left[-\lambda \left(1 - \left[\frac{1}{(x/\theta)^\alpha + 1}\right]^\alpha\right)\right] - \exp(-\lambda)}{1 - \exp(-\lambda)}$$

รูปที่ 1 แสดงฟังก์ชันการรอดชีพของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์ λ, θ, α โดยจะเห็นว่าฟังก์ชันการรอดชีพของการแจกแจงปัวซองพาราลอจิสติกมีรูปร่างลดลง (Decreasing shape) ในกรณีที่ λ มีค่าเพิ่มขึ้น รูปร่างของฟังก์ชันการรอดชีพจะลดลงเร็วขึ้น เมื่อค่าพารามิเตอร์อื่น ๆ คงที่

ฟังก์ชันการแจกแจงสะสม (The cumulative distribution function) เขียนได้ดังนี้

$$F_{PPD}(x) = 1 - \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1 - e^{-\lambda}}, \quad x > 0, \lambda > 0, \theta > 0, \alpha > 0 \quad (9)$$

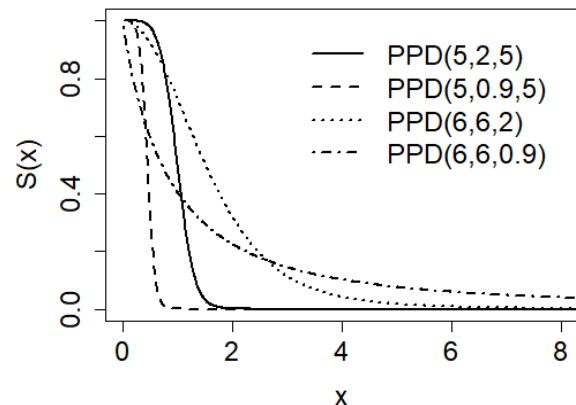
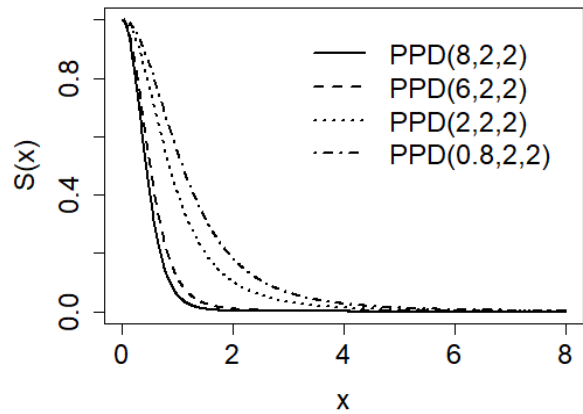
รูปที่ 2 แสดงฟังก์ชันการแจกแจงสะสมของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์ λ, θ, α ในกรณีที่ λ มีค่าเพิ่มขึ้นฟังก์ชันการแจกแจงสะสมจะมีความชันของกราฟที่เพิ่มขึ้น เมื่อค่าพารามิเตอร์อื่น ๆ คงที่

ทฤษฎีบท 2. ฟังก์ชันความหนาแน่นความน่าจะเป็น คือ

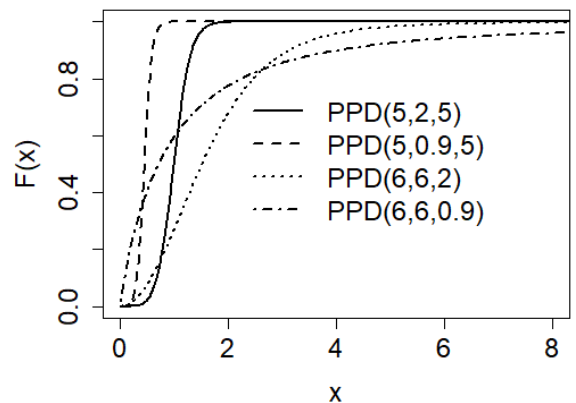
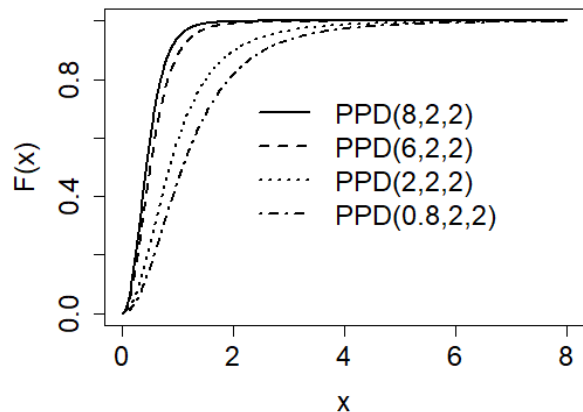
$$f_{PPD}(x) = \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} \lambda \alpha^2 (x/\theta)^\alpha}{(1 - e^{-\lambda}) x \left[(x/\theta)^\alpha + 1\right]^{\alpha+1}}, \quad x > 0, \lambda > 0, \theta > 0, \alpha > 0 \quad (10)$$

พิสูจน์ เมื่อ $f_{PPD}(x) = -\frac{dS_{PPD}(x)}{dx}$ เป็นการเสร็จสิ้นการพิสูจน์ทฤษฎีบท

รูปที่ 3 แสดงฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์ λ, θ, α เส้นโค้งแสดงให้เห็นว่าการแจกแจงปัวซองพาราลอจิสติกมีการแจกแจงเบ้ขวา และยังพบอีกว่าการแจกแจงปัวซองพาราลอจิสติกเป็นการแจกแจงฐานนิยมเดี่ยว (Unimodal distribution) ซึ่งขึ้นอยู่กับค่าของพารามิเตอร์บ่งรูปร่างคือ α อย่างไรก็ดีตาม จากการเพิ่มค่าพารามิเตอร์ λ เข้ามาขยายการแจกแจงพาราลอจิสติกเดิม



รูปที่ 1 ฟังก์ชันการรอดชีพของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์



รูปที่ 2 ฟังก์ชันการแจกแจงสะสมของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์

จะเห็นได้ว่าเมื่อ λ มีค่าลดลง การกระจายของการแจกแจงปัวซองพาราลอจิสติกจะมากขึ้น เมื่อพารามิเตอร์อื่น ๆ คงที่

ทฤษฎีบท 3. ฟังก์ชันพิบัติ (The hazard function) คือ

$$h_{PPD}(x) = \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} \lambda \alpha^2 (x/\theta)^\alpha}{x \left[(x/\theta)^\alpha + 1\right]^{\alpha+1} \left\{ e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda} \right\}}, \quad x > 0, \lambda > 0, \theta > 0, \alpha > 0 \quad (11)$$

พิสูจน์ เมื่อ $h_{PPD}(x) = \frac{f_{PPD}(x)}{S_{PPD}(x)}$ แทน $f_{PPD}(x)$ จากสมการที่ (10) และ $S_{PPD}(x)$ ในสมการที่ (8) จึงได้ผลลัพธ์ เป็นการเสร็จสิ้นการพิสูจน์ทฤษฎีบท

รูปที่ 4 แสดงฟังก์ชันพิบัติของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์ λ, θ, α โดยเส้นโค้งแสดงให้เห็นว่าฟังก์ชันพิบัติของการแจกแจงปัวซองพาราลอจิสติกสามารถสร้างได้หลากหลายรูปร่างดังสมการที่ (11)

3.2 คุณสมบัติเชิงความน่าจะเป็น

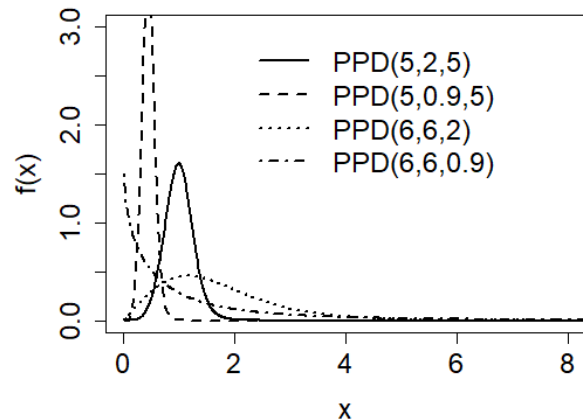
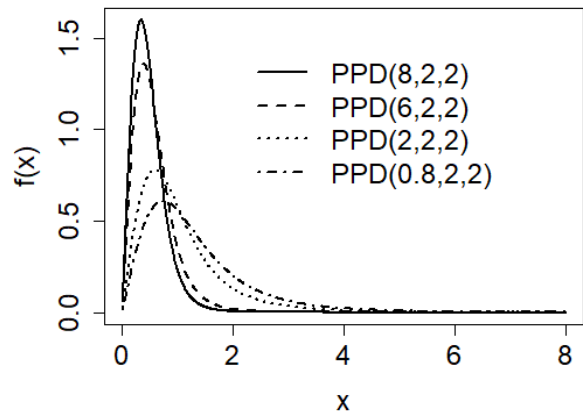
ในส่วนนี้ได้นำเสนอบางสมบัติเบื้องต้นของการแจกแจงปัวซองพาราลอจิสติก เช่น ฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวสถิติอันดับที่ k , ฟังก์ชันควอนไทล์ รวมทั้งมูลค่าเสี่ยงภัยที่นำไปใช้ในการประยุกต์ด้านการประกันภัย

3.2.1 ตัวสถิติอันดับ (Order statistics)

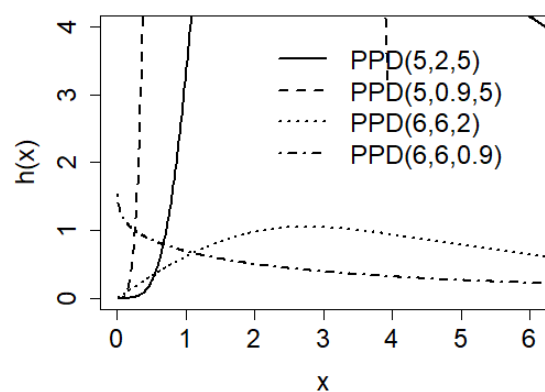
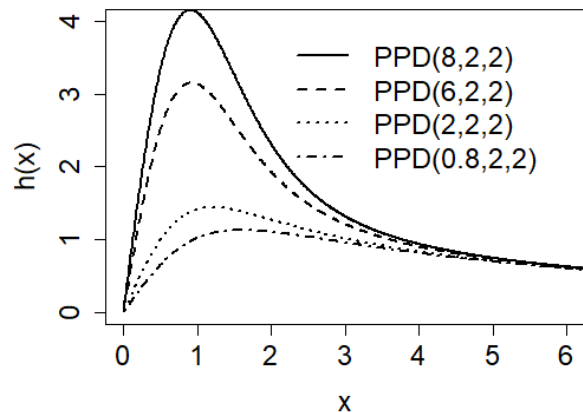
ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n จากการแจกแจงปัวซองพาราลอจิสติกที่มีพารามิเตอร์ λ, θ และ $\alpha, X \sim PPD(\lambda, \theta, \alpha)$, โดยที่ $\lambda > 0, \theta > 0$ และ $\alpha > 0$ ที่มีฟังก์ชันความหนาแน่นความน่าจะเป็นและฟังก์ชันการแจกแจงสะสม ดังสมการที่ (10) และ (9) ตามลำดับ ให้ $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ เป็นตัวสถิติอันดับ แล้วฟังก์ชันความหนาแน่นความน่าจะเป็นของ $X_{k:n}$ คือ $f_{k:n}(x), k = 1, 2, \dots, n$ หาได้จาก

$$f_{k:n}(x) = \frac{n!}{(k-1)!(n-k)!} f(x) [F(x)]^{k-1} [1-F(x)]^{n-k} \quad (12)$$

เมื่อแทนฟังก์ชันความหนาแน่นความน่าจะเป็นและฟังก์ชันการแจกแจงสะสม ดังสมการที่ (10) และ (9) ตามลำดับ ลงในสมการที่ (12) จะได้ว่า



รูปที่ 3 ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์



รูปที่ 4 ฟังก์ชันพิบัติของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์

$$f_{k:n}(x) = \frac{n!}{(k-1)!(n-k)!} \cdot \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} \lambda \alpha^2 (x/\theta)^\alpha}{(1-e^{-\lambda}) x \left[(x/\theta)^\alpha + 1\right]^{\alpha+1}} \left[1 - \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}}\right]^{k-1} \left[\frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}}\right]^{n-k}$$

เมื่อ $k=1$ ฟังก์ชันความหนาแน่นความน่าจะเป็นของ $X_{1:n}$ ได้แก่

$$f_{1:n}(x) = \frac{n!}{(n-1)!} \cdot \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} \lambda \alpha^2 (x/\theta)^\alpha}{(1-e^{-\lambda}) x \left[(x/\theta)^\alpha + 1\right]^{\alpha+1}} \left[\frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}}\right]^{n-1}$$

เมื่อ $k=n$ ฟังก์ชันความหนาแน่นความน่าจะเป็นของ $X_{n:n}$ ได้แก่

$$f_{n:n}(x) = \frac{n!}{(n-1)!} \cdot \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} \lambda \alpha^2 (x/\theta)^\alpha}{(1-e^{-\lambda}) x \left[(x/\theta)^\alpha + 1\right]^{\alpha+1}} \left[\frac{e^{-\lambda \left[1 - \left(\frac{1}{(x/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}}\right]^{n-1}$$

3.2.2 ฟังก์ชันควอนไทล์ (Quantile function: QF)

ให้ $X \square PPD(\lambda, \theta, \alpha)$ โดยที่ $\lambda > 0$, $\theta > 0$ และ $\alpha > 0$ ที่มีฟังก์ชันการแจกแจงสะสม แสดงดังสมการที่ (9) ฟังก์ชันควอนไทล์ (QF; Gilchrist, 2000) เขียนแทนด้วย $Q(p)$ และ $Q(p) = F^{-1}(p)$ โดยที่ $p \in (0,1)$ **ทฤษฎีบท 4.** ถ้า $X \square PPD(\lambda, \theta, \alpha)$ แล้ว QF ของ X คือ

$$Q(p) = \theta \left[1 - \left\{ \frac{\ln[(1-p)(1-e^{-\lambda}) + e^{-\lambda}]}{-\lambda} \right\}^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\alpha}}$$

โดยเฉพาะอย่างยิ่ง มัชฐานของ $X \square PPD(\lambda, \theta, \alpha)$ คือ

$$Q(0.5) = \theta \left[1 - \left\{ \frac{\ln[(1-0.5)(1-e^{-\lambda}) + e^{-\lambda}]}{-\lambda} \right\}^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\alpha}}$$

พิสูจน์ เมื่อ $Q(p) = F^{-1}(p)$ โดยที่ $p \in (0,1)$ นั่นคือ $F(Q(p)) = p$ เมื่อ $X \square PPD(\lambda, \theta, \alpha)$ โดยที่ $\lambda > 0$, $\theta > 0$ และ $\alpha > 0$ ที่มีฟังก์ชันการแจกแจงสะสม แสดงดังสมการที่ (9) เราได้ว่า

$$F(Q(p)) = 1 - \frac{e^{-\lambda \left[1 - \left(\frac{1}{(Q(p)/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}} = p$$

$$\frac{e^{-\lambda \left[1 - \left(\frac{1}{(Q(p)/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}} = 1-p$$

ดังนั้น เมื่อแก้สมการหา $Q(p)$ จะได้ว่า

$$Q(p) = \theta \left[1 - \left\{ \frac{\ln[(1-p)(1-e^{-\lambda}) + e^{-\lambda}]}{-\lambda} \right\}^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\alpha}}$$

ตารางที่ 1 แสดงค่าควอนไทล์ของการแจกแจงปัวซอง พาราลอจิสติกสำหรับบางค่าพารามิเตอร์ λ, θ, α

3.2.3 มูลค่าเสี่ยงภัย (Value-at-Risk: VaR)

ให้ X แทนค่าความเสียหายหรือสูญเสียเป็นผลลัพธ์ที่ไม่พึงประสงค์ มูลค่าเสี่ยงภัย (VaR; Klugman et al., 2019) ของ X คือ ค่าเปอร์เซ็นต์ไทล์ที่ $100p$ ของการแจกแจงของ X เขียนแทนด้วยสัญลักษณ์ $VaR_p(X)$ หรือ π_p พิจารณาการแจกแจงปัวซองพาราลอจิสติก ที่มีพารามิเตอร์ $\lambda > 0$, $\theta > 0$ และ $\alpha > 0$ โดยฟังก์ชันความหนาแน่นความน่าจะเป็นแสดงดัง ทฤษฎีบท 2 จะได้ว่า VaR สำหรับการแจกแจงปัวซองพาราลอจิสติกเขียนได้ดังนี้

$$VaR_p(X) = \pi_p = \theta \left[\frac{1 - \left\{ 1 + \frac{1}{\lambda} \ln[(1-p)(1-e^{-\lambda}) + e^{-\lambda}] \right\}^{-\frac{1}{\alpha}}}{\left\{ 1 + \frac{1}{\lambda} \ln[(1-p)(1-e^{-\lambda}) + e^{-\lambda}] \right\}^{-\frac{1}{\alpha}}} - 1 \right]^{\frac{1}{\alpha}}$$

พิสูจน์ เมื่อ $X \square PPD(\lambda, \theta, \alpha)$ โดยที่ $\lambda > 0$, $\theta > 0$ และ $\alpha > 0$ มีฟังก์ชันความหนาแน่นความน่าจะเป็น แสดงดังสมการที่ (10) แล้วค่าของ π_p หาได้จาก

$$P(X > \pi_p) = \int_{\pi_p}^{\infty} f_{PPD}(x) dx = 1-p \text{ ซึ่งจะได้ว่า}$$

$$\frac{e^{-\lambda \left[1 - \left(\frac{1}{(\pi_p/\theta)^\alpha + 1}\right)^\alpha\right]} - e^{-\lambda}}{1-e^{-\lambda}} = 1-p$$

คูณ $1-e^{-\lambda}$ ทั้งสองข้างของสมการด้านบนและใส่ลอการิทึมธรรมชาติทั้งสองข้าง จะได้

$$-\lambda \left[1 - \left(\frac{1}{(\pi_p/\theta)^\alpha + 1} \right)^\alpha \right] = \ln[(1-p)(1-e^{-\lambda}) + e^{-\lambda}]$$

ดังนั้น เมื่อแก้สมการหา π_p จะได้ว่า

$$\frac{1}{(\pi_p/\theta)^\alpha + 1} = \left\{ 1 + \frac{1}{\lambda} \ln[(1-p)(1-e^{-\lambda}) + e^{-\lambda}] \right\}^{-\frac{1}{\alpha}}$$

ตารางที่ 1 ค่าควอนไทล์ของการแจกแจงปัวซองพาราลอจิสติกสำหรับบางค่าพารามิเตอร์

λ	θ	$\alpha = 6$			$\alpha = 2$			$\alpha = 0.8$		
		Q_1	Q_2	Q_3	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
6	6	2.6943	3.1412	3.5715	0.9448	1.5066	2.2407	0.1899	0.6318	1.8254
	2	0.8981	1.0471	1.1905	0.3149	0.5022	0.7469	0.0633	0.2106	0.6085
	0.8	0.3592	0.4188	0.4762	0.1260	0.2009	0.2988	0.0253	0.0842	0.2434
2	6	3.1731	3.7226	4.2771	1.5540	2.5531	4.0140	0.6851	2.6237	10.1356
	2	1.0577	1.2409	1.4257	0.5180	0.8510	1.3380	0.2284	0.8746	3.3785
	0.8	0.4231	0.4963	0.5703	0.2072	0.3404	0.5352	0.0914	0.3498	1.3514
0.8	6	3.4269	4.0139	4.5902	1.9699	3.2520	5.1268	1.2860	5.2877	23.0711
	2	1.1423	1.3380	1.5301	0.6566	1.0840	1.7089	0.4287	1.7626	7.6903
	0.8	0.4569	0.5352	0.6120	0.2627	0.4336	0.6836	0.1715	0.7050	3.0761

$$\pi_p = \theta \left(\frac{1 - \left\{ 1 + \frac{1}{\lambda} \ln \left[(1-p)(1-e^{-\lambda}) + e^{-\lambda} \right] \right\}^{\frac{1}{\alpha}}}{\left\{ 1 + \frac{1}{\lambda} \ln \left[(1-p)(1-e^{-\lambda}) + e^{-\lambda} \right] \right\}^{\frac{1}{\alpha}}} \right)^{\frac{1}{\alpha}}$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{2n}{\alpha} + \sum_{i=1}^n \frac{\lambda \left\{ -\frac{(x_i/\theta)^\alpha \ln(x_i/\theta) \alpha}{(x_i/\theta)^\alpha + 1} - \ln \left[(x_i/\theta)^\alpha + 1 \right] \right\}}{\left[(x_i/\theta)^\alpha + 1 \right]^\alpha} + \sum_{i=1}^n \ln(x_i/\theta) + \sum_{i=1}^n \left\{ \frac{(x_i/\theta)^\alpha \ln(x_i/\theta)(-\alpha-1)}{(x_i/\theta)^\alpha + 1} - \ln \left[(x_i/\theta)^\alpha + 1 \right] \right\} = 0$$

3.3 การประมาณค่าพารามิเตอร์

ในการประมาณค่าพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุด กำหนดให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n ที่มาจากการแจกแจงปัวซองพาราลอจิสติก ฟังก์ชันภาวะน่าจะเป็นของตัวอย่างสุ่มขนาด n คือ

$$L(\lambda, \theta, \alpha; x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{e^{-\lambda \left[1 - \left(\frac{1}{(x_i/\theta)^\alpha + 1} \right)^\alpha \right]} \lambda \alpha^2 (x_i/\theta)^\alpha}{(1-e^{-\lambda}) x_i \left[(x_i/\theta)^\alpha + 1 \right]^{\alpha+1}}$$

$$= \frac{(\lambda \alpha^2)^n}{(1-e^{-\lambda})^n} \prod_{i=1}^n \frac{(x_i/\theta)^\alpha}{x_i \left[(x_i/\theta)^\alpha + 1 \right]^{\alpha+1}} \cdot e^{-\lambda \sum_{i=1}^n \left[1 - \left(\frac{1}{(x_i/\theta)^\alpha + 1} \right)^\alpha \right]}$$

ฟังก์ชันล็อกภาวะน่าจะเป็นแทนด้วย $\ell(\lambda, \theta, \alpha)$ เขียนได้ดังนี้

$$\ell(\lambda, \theta, \alpha) = n \ln(\lambda) + 2n \ln(\alpha) - n \ln(1-e^{-\lambda}) - \lambda \sum_{i=1}^n \left[1 - \left(\frac{1}{(x_i/\theta)^\alpha + 1} \right)^\alpha \right] + \alpha \sum_{i=1}^n \ln(x_i/\theta) - \sum_{i=1}^n \ln(x_i) - (\alpha+1) \sum_{i=1}^n \ln \left[(x_i/\theta)^\alpha + 1 \right]$$

โดยที่ค่าประมาณพารามิเตอร์ λ, θ และ α สามารถหาได้จากผลลัพธ์ของสมการอนุพันธ์ลำดับที่ 1 ของฟังก์ชันล็อกภาวะน่าจะเป็นเทียบกับพารามิเตอร์แต่ละค่าตามลำดับดังนี้

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \frac{ne^{-\lambda}}{1-e^{-\lambda}} - \sum_{i=1}^n \left[1 - \left(\frac{1}{(x_i/\theta)^\alpha + 1} \right)^\alpha \right] = 0$$

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^n \frac{\alpha^2 \lambda \left[(x_i/\theta)^\alpha + 1 \right]^{-\alpha-1} (x_i/\theta)^\alpha}{\theta} - (\alpha/\theta)^n + \sum_{i=1}^n \frac{(\alpha^2 + \alpha)(x_i/\theta)^\alpha}{\theta \left[(x_i/\theta)^\alpha + 1 \right]} = 0$$

ในกรณีที่ผลลัพธ์ของค่าประมาณพารามิเตอร์ λ, θ และ α โดยวิธีภาวะน่าจะเป็นสูงสุดไม่สามารถจัดรูปแบบได้ สามารถหาค่าประมาณพารามิเตอร์ λ, θ และ α ได้จากการใช้วิธีเชิงตัวเลข สำหรับการศึกษาค้นคว้าหาค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ λ, θ และ α ด้วยวิธีเชิงตัวเลข โดยใช้ฟังก์ชัน mledist ของ fitdistrplus package ในโปรแกรม R (Delignette-Muller & Dutang, 2015)

3.4 การจำลอง

ในส่วนนี้จะนำเสนอผลการศึกษการจำลองเพื่อประเมินประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ λ, θ และ α โดยใช้ฟังก์ชัน mledist ของ fitdistrplus package ในโปรแกรม R

การสร้างตัวแปรสุ่มที่มาจากจากการแจกแจงปัวซองพาราลอจิสติก ใช้วิธีฟังก์ชันการแจกแจงผกผัน ดังนี้ $F(X) = U \Rightarrow X = F^{-1}(U)$ ที่ซึ่ง U คือตัวแปรสุ่มที่มีการแจกแจงเอกรูปมาตรฐาน หรือ $U \sim U(0,1)$ ตัวแปรสุ่ม $X \sim PPD(\lambda, \theta, \alpha)$ สามารถสร้างได้จาก

$$X = \theta \left[\left(1 - \left\{ \frac{\ln \left[(1-U)(1-e^{-\lambda}) + e^{-\lambda} \right]}{-\lambda} \right\} \right)^{-\frac{1}{\alpha}} - 1 \right]^{\frac{1}{\alpha}}$$

การศึกษานี้ ทำการจำลองตัวแปรสุ่มที่มาจากจากการแจกแจงปัวซองพาราลอจิสติก ด้วยการกำหนดค่าพารามิเตอร์เป็นดังนี้ PPD (6, 6, 6) PPD (2, 0.8, 4) และ PPD (4, 6, 2) ที่ขนาดตัวอย่าง 25 50 100 200

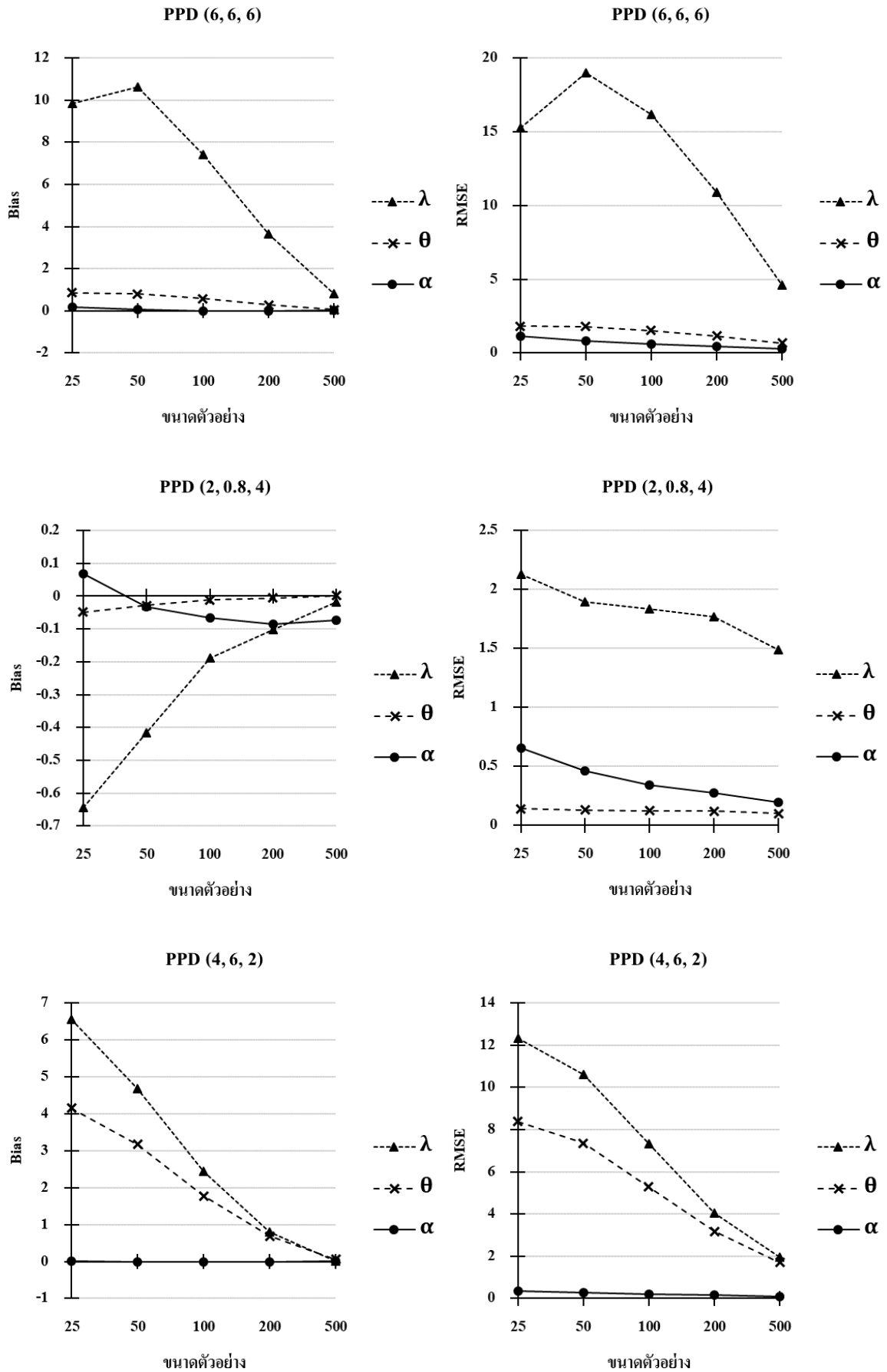
ตารางที่ 2 ค่าประมาณพารามิเตอร์เฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความเอนเอียง และรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

	ขนาดตัวอย่าง	พารามิเตอร์	ค่าประมาณพารามิเตอร์เฉลี่ย	SD	Bias	RMSE
PPD (6, 6, 6)	25	λ	15.8553	11.6821	9.8553	15.2795
		θ	6.8587	1.5541	0.8587	1.8149
		α	6.1672	1.1090	0.1672	1.1210
	50	λ	16.6049	15.7403	10.6049	18.9762
		θ	6.7976	1.6162	0.7976	1.8020
		α	6.0573	0.8005	0.0573	0.8024
	100	λ	13.4196	14.3581	7.4196	16.1587
		θ	6.5749	1.4173	0.5749	1.5292
		α	5.9920	0.5904	-0.0080	0.5903
	200	λ	9.6531	10.2358	3.6531	10.8658
		θ	6.2901	1.1121	0.2901	1.1491
		α	6.0021	0.4547	0.0021	0.4546
	500	λ	6.8125	4.5480	0.8125	4.6189
		θ	6.0503	0.7462	0.0503	0.6793
		α	6.0179	0.3098	0.0179	0.3102
PPD (2, 0.8, 4)	25	λ	1.3572	2.0293	-0.6428	2.1281
		θ	0.7516	0.1334	-0.0484	0.1419
		α	4.0677	0.6511	0.0677	0.6544
	50	λ	1.5844	1.8517	-0.4156	1.8973
		θ	0.7717	0.1283	-0.0283	0.1314
		α	3.9674	0.4602	-0.0326	0.4613
	100	λ	1.8120	1.8258	-0.1880	1.8350
		θ	0.7888	0.1250	-0.0112	0.1256
		α	3.9351	0.3391	-0.0649	0.3451
	200	λ	1.8990	1.7685	-0.1010	1.7709
		θ	0.7951	0.1209	-0.0049	0.1210
		α	3.9152	0.2634	-0.0848	0.2766
	500	λ	1.9829	1.4876	-0.0171	1.4873
		θ	0.8012	0.1020	0.0012	0.1021
		α	3.9259	0.1826	-0.0741	0.1970
PPD (4, 6, 2)	25	λ	10.5523	10.4253	6.5523	12.3112
		θ	10.1508	7.2857	4.1508	8.3836
		α	2.0203	0.3609	0.0203	0.3614
	50	λ	8.6739	9.5226	4.6739	10.6057
		θ	9.1730	6.1415	3.1730	7.3590
		α	1.9862	0.2676	-0.0138	0.2679
	100	λ	6.4359	6.9010	2.4359	7.3167
		θ	7.7729	4.9760	1.7729	5.2813
		α	1.9827	0.2042	-0.0173	0.2048
	200	λ	4.7929	3.9641	0.7929	4.0417
		θ	6.6828	3.0962	0.6828	3.1699
		α	1.9908	0.1549	-0.0092	0.1551
	500	λ	4.0104	1.9464	0.0104	1.9459
		θ	6.0616	1.7058	0.0616	1.7065
		α	2.0009	0.1065	0.0009	0.1064

และ 500 จำนวน 2,000 รอบ ประมาณค่าพารามิเตอร์ ด้วยวิธีภาวน่าจะเป็นสูงสุดในแต่ละค่าพารามิเตอร์และ ขนาดตัวอย่างที่กำหนดในแต่ละรอบ หาค่าเฉลี่ยและส่วน เบี่ยงเบนมาตรฐาน (Standard deviation: SD) ของ ค่าประมาณพารามิเตอร์ดังกล่าว ประเมินประสิทธิภาพของ การประมาณค่าพารามิเตอร์ด้วยวิธีภาวน่าจะเป็นสูงสุด ด้วย

ค่าความเอนเอียง (Bias) และรากของค่าคลาดเคลื่อนกำลัง สองเฉลี่ย (Root mean square error: RMSE) แสดงดัง ตารางที่ 2

จากผลในตารางที่ 2 พบว่า ค่าประมาณพารามิเตอร์เฉลี่ย λ , θ และ α มีค่าใกล้เคียงกับค่าพารามิเตอร์จริงเมื่อ ขนาดตัวอย่างเพิ่มขึ้น และค่าความเอนเอียง (Bias) และ



รูปที่ 5 ค่า Bias และ RMSE ของ PPD พารามิเตอร์ต่าง ๆ

ตารางที่ 3 สถิติเชิงพรรณนาสำหรับข้อมูลชุดที่ 1

จำนวน	ค่าเฉลี่ย	มัธยฐาน	ส่วนเบี่ยงเบนมาตรฐาน	ความเบ้	ความโด่ง
50	2.95	2.94	0.64	0.41	1.02

ตารางที่ 4 ค่าประมาณพารามิเตอร์และเกณฑ์การเปรียบเทียบการแจกแจงของข้อมูลชุดที่ 1

การแจกแจง	ค่าประมาณพารามิเตอร์	LL	AIC	BIC	KS	CM	AD
PPD	$\hat{\lambda} = 3.9945$ $\hat{\theta} = 5.1148$ $\hat{\alpha} = 6.3217$	-46.9383	99.8766	105.6127	0.0790	0.0406	0.3153
PD	$\hat{\alpha} = 5.6043$ $\hat{\theta} = 4.2226$	-48.0954	100.1909	104.0150	0.0978	0.0727	0.5193
IPD	$\hat{\nu} = 4.8719$ $\hat{\mu} = 1.9499$	-53.3155	110.6311	114.4551	0.1328	0.2125	1.4695
LLD	$\hat{\gamma} = 8.4432$ $\hat{K} = 2.9121$	-47.0421	98.0843	101.9084	0.0770	0.0369	0.3180

ตารางที่ 5 สถิติเชิงพรรณนาสำหรับข้อมูลชุดที่ 2

จำนวน	ค่าเฉลี่ย	มัธยฐาน	ส่วนเบี่ยงเบนมาตรฐาน	ความเบ้	ความโด่ง
101	1.03	0.8	1.12	3.05	14.51

ตารางที่ 6 ค่าประมาณพารามิเตอร์และเกณฑ์การเปรียบเทียบการแจกแจงของข้อมูลชุดที่ 2

การแจกแจง	ค่าประมาณพารามิเตอร์	LL	AIC	BIC	KS	CM	AD
PPD	$\hat{\lambda} = 82.3361$ $\hat{\theta} = 98.4441$ $\hat{\alpha} = 0.9420$	-103.0045	212.0089	219.8543	0.0913	0.2032	1.1648
PD	$\hat{\alpha} = 1.2808$ $\hat{\theta} = 0.8290$	-110.4026	224.8051	230.0353	0.1047	0.3497	2.6336
IPD	$\hat{\nu} = 1.1307$ $\hat{\mu} = 0.5230$	-114.2977	232.5954	237.8256	0.1206	0.4431	2.9309
LLD	$\hat{\gamma} = 1.2704$ $\hat{K} = 0.6238$	-112.6862	229.3724	234.6026	0.1113	0.3915	2.7831

ตารางที่ 7 สถิติเชิงพรรณนาสำหรับข้อมูลชุดที่ 3

จำนวน	ค่าเฉลี่ย	มัธยฐาน	ส่วนเบี่ยงเบนมาตรฐาน	ความเบ้	ความโด่ง
63	2.09	2.06	1.25	0.43	2.73

ตารางที่ 8 ค่าประมาณพารามิเตอร์และเกณฑ์การเปรียบเทียบการแจกแจงของข้อมูลชุดที่ 3

การแจกแจง	ค่าประมาณพารามิเตอร์	LL	AIC	BIC	KS	CM	AD
PPD	$\hat{\lambda} = 109.4739$ $\hat{\theta} = 54.3890$ $\hat{\alpha} = 1.6402$	-100.4288	206.8577	213.2871	0.1091	0.0942	0.6561
PD	$\hat{\alpha} = 2.0019$ $\hat{\theta} = 2.8434$	-104.5693	213.1386	217.4249	0.1141	0.1190	1.14079
IPD	$\hat{\nu} = 1.5592$ $\hat{\mu} = 1.2005$	-113.7165	231.4331	235.7194	0.1424	0.3517	2.5551
LLD	$\hat{\gamma} = 2.1352$ $\hat{K} = 1.8016$	-108.7423	221.4846	225.7709	0.1198	0.1817	1.6807

รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และจากรูปที่ 5 สังเกตได้ว่าค่าประมาณของพารามิเตอร์ λ เมื่อตัวอย่างขนาดเล็กให้ค่าความเอนเอียงและรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยมากกว่าค่าประมาณของพารามิเตอร์ตัวอื่นในทุกกรณี นั่นคือมีค่าประมาณที่ห่างจากค่าพารามิเตอร์จริงค่อนข้างมาก แต่อย่างไรก็ตามค่าประมาณของพารามิเตอร์ λ จะมีค่าใกล้เคียงกับค่าพารามิเตอร์จริงเมื่อขนาดตัวอย่างใหญ่

3.5 การประยุกต์

งานวิจัยครั้งนี้ได้ประยุกต์การแจกแจงการแจกแจงปัวซองพาราลอจิสติก (PPD) กับชุดข้อมูลจริง 3 ชุด โดยเปรียบเทียบการแจกแจงใหม่ที่เสนอกับการแจกแจงพาราลอจิสติก (PD) การแจกแจงพาราลอจิสติกผกผัน (IPD) และการแจกแจงล็อกลอจิสติก (LLD)

เกณฑ์การเปรียบเทียบที่ใช้มีดังนี้ ค่าล็อกภาวะน่าจะเป็น (Log-likelihood; LL) ค่าเกณฑ์สารสนเทศของอะกะอิเกะ (Akaike's Information Criterion; AIC) ค่าเกณฑ์สารสนเทศของเบส์ (Bayesian Information Criterion; BIC) ค่าการทดสอบคอลโมโกรอฟ-สมิรโนฟ (Kolmogorov-Smirnov test; KS) ค่าการทดสอบคราเมอร์-ฟอนมีเซส (Cramer-von Mises test; CM) และค่าการทดสอบแอนเดอสันดาร์ลิง (Anderson-Darling test; AD) ซึ่งการแจกแจงที่เหมาะสมที่สุดสำหรับข้อมูลคือ ค่าล็อกภาวะน่าจะเป็น (LL) มีค่าสูงกว่า และค่าเกณฑ์สารสนเทศของอะกะอิเกะ (AIC) ค่าเกณฑ์สารสนเทศของเบส์ (BIC) ค่าการทดสอบคอลโมโกรอฟ-สมิรโนฟ (KS) ค่าการทดสอบคราเมอร์-ฟอนมีเซส (CM) และค่าการทดสอบแอนเดอสันดาร์ลิง (AD) มีค่าน้อยกว่าในส่วนของการประมาณค่าพารามิเตอร์ของการแจกแจงทั้งหมดนั้น คำนวณโดยใช้ฟังก์ชัน fitdistrplus ในโปรแกรม R

3.5.1 ข้อมูลชุดที่ 1

ค่าความต้านทานแรงดึงของเส้นใยคาร์บอนที่สุ่มตัวอย่างมาจากกระบวนการผลิต มีหน่วยเป็นจิกะปาสกาล (Gigapascal: GPa) (Aldahlan et al., 2020)

ตารางที่ 3 แสดงค่าสถิติเชิงพรรณนาสำหรับข้อมูลชุดที่ 1 และตารางที่ 4 แสดงค่าประมาณพารามิเตอร์และเกณฑ์การเปรียบเทียบ LL AIC BIC KS CM และ AD สำหรับการแจกแจงปัวซองพาราลอจิสติก การแจกแจงพาราลอจิสติก การแจกแจงพาราลอจิสติกผกผัน และการแจกแจงล็อกลอจิสติก ของข้อมูลชุดที่ 1 พบว่าเมื่อพิจารณาจากเกณฑ์การเปรียบเทียบ LL การแจกแจงปัวซอง

พาราลอจิสติก มีความเหมาะสมกับข้อมูลชุดที่ 1 มากกว่าการแจกแจงแบบอื่น ๆ เมื่อพิจารณาเกณฑ์การเปรียบเทียบแบบอื่น ๆ พบว่าการแจกแจงล็อกลอจิสติกมีความเหมาะสมกับข้อมูลชุดที่ 1 มากกว่าการแจกแจงปัวซองพาราลอจิสติกเพียงเล็กน้อย และเมื่อเปรียบเทียบค่า LL กับผลศึกษาของ Aldahlan et al. (2020) ที่ประยุกต์ใช้ข้อมูลชุดที่ 1 กับการแจกแจง Exponentiated power generalized Weibull power series (EPGWPS) ที่ประกอบด้วย 5 พารามิเตอร์ ให้ค่า LL เท่ากับ -46.4481 โดยพบว่ามีความใกล้เคียงกับค่า LL ของการแจกแจงปัวซองพาราลอจิสติก ทั้งที่มีจำนวนพารามิเตอร์มากกว่า

3.5.2 ข้อมูลชุดที่ 2

ระยะเวลาที่ใช้ทดสอบเส้นใยเคฟลาร์ 49/อีพ็อกซี่ ภายใต้แรงกดดันอย่างต่อเนื่องที่ระดับ 90% จนกระทั่งเส้นใยขาด มีหน่วยเป็นชั่วโมง (Nasir et al., 2019)

ตารางที่ 5 แสดงค่าสถิติเชิงพรรณนาสำหรับข้อมูลชุดที่ 2 และตารางที่ 6 แสดงค่าประมาณพารามิเตอร์และเกณฑ์การเปรียบเทียบ LL AIC BIC KS CM และ AD สำหรับการแจกแจงปัวซองพาราลอจิสติก การแจกแจงพาราลอจิสติก การแจกแจงพาราลอจิสติกผกผัน และการแจกแจงล็อกลอจิสติก ของข้อมูลชุดที่ 2 ซึ่งพบว่าการแจกแจงปัวซองพาราลอจิสติกมีความเหมาะสมกับข้อมูลชุดที่ 2 มากที่สุด ในทุก ๆ เกณฑ์การเปรียบเทียบ และเมื่อเปรียบเทียบกับผลศึกษาของ Nasir et al. (2019) ที่ประยุกต์ใช้ข้อมูลชุดที่ 2 กับการแจกแจง exponentiated Burr XII (EBXII) power series โดยมีการแจกแจงเฉพาะ ได้แก่ การแจกแจง EBXII geometric (EBXII-G), การแจกแจง EBXII logarithmic (EBXII-L) และการแจกแจง EBXII Poisson (EBXII-P) ซึ่งประกอบด้วย 4 พารามิเตอร์ พบว่าค่า LL ของการแจกแจง EBXII-G, การแจกแจง EBXII-L และการแจกแจง EBXII-P มีค่าเท่ากับ -102.2356, -101.0149 และ -103.4967 ตามลำดับ ซึ่งการแจกแจงปัวซองพาราลอจิสติกมีค่า LL มากกว่าการแจกแจง EBXII-P แต่มีค่า LL น้อยกว่าการแจกแจง EBXII-G และการแจกแจง EBXII-L ในส่วนค่า AIC ของการแจกแจง EBXII-G, การแจกแจง EBXII-L และการแจกแจง EBXII-P มีค่าเท่ากับ 212.4712, 210.0298 และ 214.9934 ตามลำดับ ซึ่งการแจกแจงปัวซองพาราลอจิสติกมีค่า AIC น้อยกว่า การแจกแจง EBXII-G และการแจกแจง EBXII-P แต่มีค่า AIC มากกว่าการแจกแจง EBXII-L แต่ทั้งนี้พารามิเตอร์ของการแจกแจงปัวซองพาราลอจิสติกมีจำนวนที่น้อยกว่า

3.5.3 ข้อมูลชุดที่ 3

ระยะเวลาการใช้งานของกระจกหน้าเครื่องบิน มีหน่วยเป็น 1,000 ชม (Nasir et al., 2019)

ตารางที่ 7 แสดงค่าสถิติเชิงพรรณนาสำหรับข้อมูลชุดที่ 3 และตารางที่ 8 แสดงค่าประมาณพารามิเตอร์และเกณฑ์การเปรียบเทียบ LL AIC BIC KS CM และ AD สำหรับการแจกแจงปัวซองพาราลอจิสติก การแจกแจงพาราลอจิสติก การแจกแจงพาราลอจิสติกผกผัน และการแจกแจงล็อกลอจิสติก ของข้อมูลชุดที่ 3 โดยพบว่าการแจกแจงปัวซองพาราลอจิสติกมีความเหมาะสมกับข้อมูลชุดที่ 3 มากที่สุด ในทุก ๆ เกณฑ์การเปรียบเทียบ และเมื่อเปรียบเทียบกับผลศึกษาของ Nasir et al. (2019) ที่ประยุกต์ใช้ข้อมูลชุดที่ 3 กับการแจกแจง EBXII power series โดยมีการแจกแจงเฉพาะ ได้แก่ การแจกแจง EBXII-G, การแจกแจง EBXII-L และการแจกแจง EBXII-P ซึ่งประกอบด้วย 4 พารามิเตอร์ พบว่าค่า LL ของการแจกแจง EBXII-G, การแจกแจง EBXII-L และการแจกแจง EBXII-P มีค่าเท่ากับ -102.0361, -99.5671, และ -108.0832 ตามลำดับ ซึ่งการแจกแจงปัวซองพาราลอจิสติกมีค่า LL มากกว่าการแจกแจง EBXII-G และการแจกแจง EBXII-P แต่มีค่า LL น้อยกว่าการแจกแจง EBXII-L ในส่วนค่า AIC ของการแจกแจง EBXII-G, การแจกแจง EBXII-L และการแจกแจง EBXII-P มีค่าเท่ากับ 212.0721, 207.1343 และ 224.1664 ตามลำดับ ซึ่งการแจกแจงปัวซองพาราลอจิสติกมีค่า AIC น้อยกว่าทุกการแจกแจงดังกล่าว โดยที่พารามิเตอร์ของการแจกแจงปัวซองพาราลอจิสติกมีจำนวนที่น้อยกว่า

4. สรุปผลการวิจัย

งานวิจัยนี้เสนอการแจกแจงใหม่คือ การแจกแจงปัวซองพาราลอจิสติก ซึ่งเป็นการแจกแจงผสมของการแจกแจงพาราลอจิสติกกับการแจกแจงปัวซอง มีการเสนอสมบัติเบื้องต้นของการแจกแจงปัวซองพาราลอจิสติกบางสมบัติ ประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด และประเมินประสิทธิภาพของการประมาณค่าพารามิเตอร์ของการแจกแจงใหม่ด้วยการจำลองข้อมูลภายใต้สถานการณ์ต่าง ๆ พบว่าค่าประมาณพารามิเตอร์เฉลี่ยของ θ และ α มีค่าใกล้เคียงกับค่าพารามิเตอร์จริงเมื่อขนาดตัวอย่างเพิ่มขึ้น ในขณะที่กรณีตัวอย่างขนาดเล็กค่าประมาณของพารามิเตอร์เฉลี่ยของ λ มีค่าห่างจากค่าพารามิเตอร์จริงค่อนข้างมากเมื่อเทียบกับค่าประมาณพารามิเตอร์ของตัว

อื่น แต่อย่างไรก็ตามค่าประมาณของพารามิเตอร์ λ จะมีค่าใกล้เคียงกับค่าพารามิเตอร์จริง เมื่อตัวอย่างมีขนาดใหญ่

การเปรียบเทียบประสิทธิภาพของการแจกแจงปัวซองพาราลอจิสติกด้วยเกณฑ์การเปรียบเทียบ LL AIC BIC KS CM และ AD กับการแจกแจงอื่น ๆ ที่เกี่ยวข้อง ได้แก่ การแจกแจงพาราลอจิสติก การแจกแจงพาราลอจิสติกผกผัน และการแจกแจงล็อกลอจิสติก โดยการประยุกต์กับชุดข้อมูลจริง 3 ชุด พบว่าการแจกแจงปัวซองพาราลอจิสติกมีความเหมาะสมกับชุดข้อมูลต่าง ๆ มากกว่าการแจกแจงอื่น นอกจากนี้เมื่อเปรียบเทียบค่า LL และ AIC ของการแจกแจงปัวซองพาราลอจิสติกกับการแจกแจง EPGWPS (Aldahlan et al., 2020) ในข้อมูลชุดที่ 1 และการแจกแจง EBXII power series (Nasir et al., 2019) ในข้อมูลชุดที่ 2 และชุดที่ 3 พบว่าการแจกแจงปัวซองพาราลอจิสติกที่มีจำนวนพารามิเตอร์น้อยกว่า ให้ค่า LL และ AIC ใกล้เคียงกับการแจกแจงดังกล่าว

กิตติกรรมประกาศ

ผู้แต่งขอขอบพระคุณผู้ทรงคุณวุฒิทุกท่านที่ได้สละเวลาอันมีค่าอย่างยิ่งในการพิจารณาบทความ และให้ข้อเสนอแนะที่เป็นประโยชน์อย่างยิ่งต่อการปรับปรุงเนื้อหาของบทความฉบับนี้ให้มีความสมบูรณ์มากยิ่งขึ้น

เอกสารอ้างอิง

- Aldahlan, M. A., Jamal, F., Chesneau, C., Elbatal, I., & Elgarhy, M. (2020). Exponentiated power generalized Weibull power series family of distributions: Properties, estimation and applications. *PLoS ONE*, 15(3), e0230004. <https://doi.org/10.1371/journal.pone.0230004>
- Balakrishnan, N., Leiva, V., Sanhueza, A., & Cabrera, E. (2009). Mixture inverse Gaussian distribution and its transformations, moments and applications. *Statistics*, 43(1), 91–104. <https://doi.org/10.1080/02331880701829948>
- Bhati, D., Calderin-Ojeda, E., & Meenakshi, M. (2019). A New Heavy Tailed Class of Distributions Which Includes the Pareto. *Risks*, 7(4), 99. <https://doi.org/10.3390/risks7040099>
- Bulmer, M. G. (1974). On fitting the Poisson lognormal distribution to species abundance data. *Biometrics*, 30(1), 101–110. <https://doi.org/10.2307/2529621>
- Choopradit, B., Thammachote, P., Pianroj, N., & Duangsawat, S. (2021). Identifying Best-fit Distribution for Tourist Expenditure per Person in the Southern Gulf of Thailand Provincial Cluster before Global Covid-19 Pandemic. *Journal of Chandrakasemsarn*, 27(2), 212–227. (in Thai)
- Choopradit, B., & Wasinrat, S. (2021). Distribution of Tropical Cyclone Counts Moving into Thailand. *NKRAFA Journal of Science and Technology*, 17(2), 45–53. (in Thai)

- Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4), 1–34. <https://doi.org/10.18637/jss.v064.i04>
- Emilio, G. D., Sarabia, J. M., & Enrique, C. O. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, 42(1), 39–49. <https://doi.org/10.1016/j.insmatheco.2006.12.001>
- Erisoglu, M., Servi, T., Erisoglu, U., & Calis, N. (2013). Mixture gamma distribution for estimation of wind power potential. *International Journal of Applied Mathematics and Statistics*, 40(10), 223–231.
- Hall, P., & Zhou, X. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals Statistics*, 31(1), 201–224. <https://doi.org/10.1214/aos/1046294462>
- Idemudia, R., & Ekhoesuehi, N. (2019). A New Three-Parameter Paralogistic Distribution: Its Properties and Application. *Journal of Data Science*, 17(2), 239–258. [https://doi.org/10.6339/JDS.201904_17\(2\).0001](https://doi.org/10.6339/JDS.201904_17(2).0001)
- Gilchrist, W. G. (2000). *Statistical modelling with quantile functions*. Chapman & Hall/ CRC.
- Khajorntaewawong, N., Wiriya-preeda, N., Pathumrathaiand, C., & Budsaba, K. (2020). Analysis of Flight Departure Delay Distributions. *Thai Science and Technology Journal*, 28(6), 957–968. (in Thai)
- Kleiber, C., & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley and Sons.
- Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2019). *Loss models: from data to decisions* (5th ed.). John Wiley and Sons.
- McDonald, J. B. (1984). *Some generalized functions for the Size distributions in Economics and Actuarial Sciences*. John Wiley and Sons.
- McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. John Wiley and Sons.
- Nasir, A., Yousof, H. M., Jamal, F., & Korkmaz, M. Ç. (2019). The Exponentiated Burr XII Power Series Distribution: Properties and Applications. *Stats*, 2(1), 15–31. <https://doi.org/10.3390/stats2010002>
- Rodrigues, J., Cancho, V. G., de Castro, M., & Louzada-Neto, F. (2009). On the unification of long-term survival models. *Statistics and Probability Letters*, 79(6), 753–759. <https://doi.org/10.1016/j.spl.2008.10.029>
- Wasinrat, S., & Choopradit, B. (2023). The Poisson Inverse Pareto Distribution and Its Application. *Thailand Statistician*, 21(1), 110–124. <https://ph02.tci-thaijo.org/index.php/thaistat/article/view/248027>
- Withers, C. S., & Nadarajah, S. (2011). On the compound Poisson-gamma distribution. *Kybernetika*, 47(1), 15–37.