

การแจกแจงโลแม็กซ์-เรย์ลีตัดปลายทั้งสองด้านและการประยุกต์

Double Truncated Lomax-Rayleigh Distribution and its Application

กนิษฐา ยิ้มนาค^{1*} และ วิมลมาศ บำรุงเศรษฐพงษ์¹
Kanittha Yimnak^{1*} and Wimonmas Bamrungsetthapong¹

Received: 1 May 2023

Revised: 3 September 2023

Accepted: 9 November 2023

บทคัดย่อ

ในบทความวิจัยนี้นำเสนอการแจกแจงโลแม็กซ์-เรย์ลีตัดปลาย (TLRD) ซึ่งทำการตัดปลายทั้งสองด้าน พร้อมนำเสนอคุณสมบัติทางสถิติบางประการ ได้แก่ ฟังก์ชันการรอดชีพ ฟังก์ชันพิบัติ โมเมนต์ ตลอดจนการประมาณค่าพารามิเตอร์ ในส่วนของระเบียบวิธีเชิงตัวเลขมีการศึกษาฟังก์ชันควอนไทล์ เพื่อใช้ในการสร้างตัวแปรสุ่มที่มีการแจกแจงตามแบบ TLRD ซึ่งผลการศึกษาพบว่าค่าความคลาดเคลื่อนมาตรฐาน (SE) จากการประมาณค่าพารามิเตอร์จะลดลงเมื่อขนาดตัวอย่าง n เพิ่มขึ้น ในส่วนของการศึกษาประสิทธิภาพของการแจกแจงที่พัฒนาขึ้นโดยการประยุกต์การแจกแจงดังกล่าวกับข้อมูลจริงที่เป็นข้อมูลช่วงชีวิตจำนวนห้าชุดซึ่งเป็นข้อมูลอายุการใช้งานของอุปกรณ์อิเล็กทรอนิกส์และข้อมูลทางการแพทย์ โดยเปรียบเทียบการทดสอบภาวะสารูปดีกับการแจกแจงโลแม็กซ์-เรย์ลี (LRD) ซึ่งเป็นการแจกแจงต้นกำเนิด ผลการศึกษาพบว่า การแจกแจงโลแม็กซ์-เรย์ลีตัดปลายเป็นการแจกแจงที่มีความสอดคล้องกับข้อมูลจริงทั้งห้าชุด

คำสำคัญ: การแจกแจงโลแม็กซ์-เรย์ลี การตัดปลายการแจกแจง ข้อมูลช่วงชีวิต

¹ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ปทุมธานี ประเทศไทย 12110

¹Department of Mathematics and Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand 12110

* ผู้ติดต่อประสานงาน (Corresponding author) e-mail: kanittha_y@rmutt.ac.th

Abstract

In this research article, the Lomax-Rayleigh distribution (LRD) is developed by truncating both sides along with the presentation of some statistical properties such as the survival function, hazard function, moment, and parameter estimation. In terms of numerical experiments, the quantile function is studied as being used to create a random variable corresponding to a TLRD distribution. The results show that the standard error (SE) of the parameter estimation decreases as the sample size n increases. To study the efficiency of the distribution, the developed distribution is applied to the five real datasets, which are the lifetime of electronic devices and medical information. The goodness-of-fit test is used for the performance comparison. The results reveal that the truncated Lomax-Rayleigh distribution is the one that is consistent with all five real datasets.

Keywords: Lomax-Rayleigh Distribution; Truncated Distribution; Lifetime Data

บทนำ

การตัดปลาย (truncation) เป็นหนึ่งในการพัฒนาการแจกแจงของข้อมูลช่วงชีวิต (lifetime) ซึ่งเป็นข้อมูลที่ใช้ในการอธิบายอายุการใช้งานของสินค้าหรือผลิตภัณฑ์ อีกทั้งยังนำมาใช้ในการอธิบายงานทางด้านคณิตศาสตร์ประกันภัย เศรษฐศาสตร์ วิศวกรรมศาสตร์ วิทยาศาสตร์สิ่งแวดล้อม การจัดการ ฯลฯ นักวิจัยหลายๆท่านได้คิดค้นและพัฒนาการแจกแจงของข้อมูลช่วงชีวิตด้วยพารามิเตอร์ชนิดต่างๆเพื่อเพิ่มประสิทธิภาพในการอธิบายปรากฏการณ์ของข้อมูลจริง ยกตัวอย่างเช่นการแจกแจงที่เป็นการผสมผสานกันระหว่างการแจกแจงแบบเลขชี้กำลัง (Exponential distribution) และการแจกแจงแกมมา (Gamma distribution) จากการศึกษาที่ผ่านมา เช่น Lindley [1] นำเสนอการแจกแจงลินลีย์พารามิเตอร์เดียว (one-parameter Lindley distribution) และคุณสมบัติทางสถิติบางประการเพื่อให้มีความสอดคล้องและยืดหยุ่นกับข้อมูลจริงมากยิ่งขึ้น ถัดมา Shanker, et al. [2] พัฒนาการแจกแจงลินลีย์พารามิเตอร์เดียวโดยการเพิ่มพารามิเตอร์รูปร่าง (shape parameter) ให้กลายเป็นการแจกแจงที่มีสองพารามิเตอร์ พร้อมกับนำเสนอคุณสมบัติทางสถิติที่สำคัญ เช่น โมเมนต์ การประมาณค่าพารามิเตอร์ตลอดจนการประยุกต์กับข้อมูลจริง Shanker [3] นำเสนอการแจกแจงอakash (Akash distribution) และการนำเสนอการประยุกต์กับข้อมูลจริงทางด้านวิศวกรรมศาสตร์การแจกแจงช่วงชีวิตที่พัฒนามาจากการแจกแจงแบบเกาส์ (Gaussian distribution) และการประยุกต์ เช่น Chankham, et al. [4] นำเสนอการสร้างช่วงความเชื่อมั่นแบบใหม่สำหรับค่าสัมประสิทธิ์ของความแปรผันของการแจกแจงแบบเกาส์แบบผกผัน (Inverse Gaussian distribution) และนำไปใช้ในการวัดการกระจายตัวของ PM 2.5 ในประเทศไทย นอกจากนี้ยังมีการแจกแจงช่วงชีวิตที่พัฒนามาจากการแจกแจงพาเรโตชนิดที่ 2 (Pareto distribution Type II) ซึ่งเป็นการแจกแจงที่มีลักษณะหางหนา (heavy tailed distributions) สำหรับตัวแปรสุ่มต่อเนื่องที่เป็นจำนวนจริงบวก ได้แก่ การแจกแจงโลแม็กซ์ (Lomax distribution) นำเสนอโดย Lomax [5] ซึ่งถูกนำมาประยุกต์ใช้ในงานทางธุรกิจ เศรษฐศาสตร์ ทฤษฎีแถวคอย การสร้างแบบจำลองการรับส่งข้อมูลทางอินเทอร์เน็ต เป็นต้น Rayleigh [6] ได้คิดค้นตัวแบบการแจกแจงเรย์ลีย์ (Rayleigh distribution) ซึ่งเป็นหนึ่งในตัวแบบที่พัฒนามาจากการแจกแจงไวบูล (Weibull distribution) ต่อมา Venegas, et al. [7] เสนอการแจกแจงโลแม็กซ์-เรย์ลีย์ (Lomax-Rayleigh : LRD) ด้วยพารามิเตอร์ α และ θ ซึ่งเป็นการผสมกันระหว่างการแจกแจงโลแม็กซ์และการแจกแจงเรย์ลีย์ และมีการนำเสนอฟังก์ชันการรอดชีพ (survival function) ฟังก์ชันพิบัติ (hazard function) สถิติอันดับ (order statistics) การประมาณค่าพารามิเตอร์และการประยุกต์กับข้อมูลจริง การแจกแจงโลแม็กซ์-เรย์ลีย์มีตัวแบบที่ค่อนข้างง่ายไม่ซับซ้อน อีกทั้งยังพบว่ามีคุณสมบัติที่สอดคล้องกับข้อมูลจริง ฟังก์ชันการแจกแจงสะสม (cumulative distribution function : CDF) และฟังก์ชันความหนาแน่นความน่าจะเป็น (probability density function : PDF) ของการแจกแจงโลแม็กซ์-เรย์ลีย์ แสดงดังสมการที่ (1) และ (2) ตามลำดับ

$$F(x; \theta, \alpha) = 1 - \left(\frac{\theta}{\theta + x^2} \right)^\alpha \quad (1)$$

$$f(x; \theta, \alpha) = \frac{2\alpha\theta^\alpha x}{(\theta + x^2)^{\alpha+1}} \quad (2)$$

เมื่อ $\alpha > 0$ คือ พารามิเตอร์รูปร่าง (shape parameter) และ $\theta > 0$ คือ พารามิเตอร์บ่งขนาด (scale parameter) Falgore, et al. [8] เสนอตัวแบบย่อยของโลแม็กซ์จีแฟมิลี่ผกผัน (inverse Lomax G family) ซึ่งถูกเรียกว่าการแจกแจงโลแม็กซ์-เรย์ลีผกผัน (Inverse Lomax Rayleigh distribution) โดยมีการนำเสนอคุณสมบัติที่สำคัญบางประการ ได้แก่ ฟังก์ชันความหนาแน่นความน่าจะเป็น ฟังก์ชันการแจกแจงสะสม ฟังก์ชันพิบัติและฟังก์ชันการรอดชีพ

การแจกแจงช่วงชีวิตที่ถูกตัดปลาย (truncated lifetime distribution) เป็นหนึ่งในวิธีการพัฒนาการแจกแจงให้มีความสอดคล้องใกล้เคียงกับปรากฏการณ์ของข้อมูลจริงมากขึ้นเนื่องจากการจำกัดโดเมน (domain) ให้อยู่ในช่วงที่ต้องการศึกษา การตัดปลายการแจกแจงทำได้สองวิธีได้แก่ การตัดปลายเพียงด้านเดียว (ด้านซ้ายหรือด้านขวา) และการตัดปลายทั้งสองด้าน (ทั้งด้านซ้ายและด้านขวา) [9] มีนักสถิติหลายท่านได้ศึกษาการตัดปลายการแจกแจงและศึกษาประสิทธิภาพของการแจกแจงที่พัฒนาขึ้น ตัวอย่างเช่น Johnson [10] นำเสนอการแจกแจงปกติตัดปลายและคุณสมบัติทางสถิติบางประการ ซึ่งเป็นที่ยอมรับในการประยุกต์ใช้ในงานหลายๆ ด้าน Iwueze [11] นำเสนอแนวคิดการแจกแจงปกติที่ตัดปลายทางซ้ายที่ค่าศูนย์สำหรับการประยุกต์ใช้ในการสร้างแบบจำลองของข้อมูลอนุกรมเวลา Aryuyuen [12] นำเสนอการแจกแจงลินเลย์สองพารามิเตอร์ตัดปลายพร้อมทั้งศึกษาคุณสมบัติทางสถิติและนำเสนอการประยุกต์กับข้อมูลจริง Shukla และ Shanker [13] ได้พัฒนาการแจกแจงอค์พารามิเตอร์เดียวตัดปลายและคุณสมบัติทางสถิติของการแจกแจงที่พัฒนาขึ้น อีกทั้งยังมีการนำเสนอการประยุกต์ใช้กับข้อมูลจริงที่เป็นข้อมูลช่วงชีวิตจำนวน 3 ชุด ซึ่งการแจกแจงอค์ที่พัฒนาขึ้นมานี้มีความเหมาะสมกับข้อมูลจริงมากกว่าการแจกแจงอค์ที่เป็นการแจกแจงต้นกำเนิด Hammood, et al. [14] นำเสนอการแจกแจงเรย์ลี-โลแม็กซ์ตัดปลาย (truncated Rayleigh Lomax distribution) พร้อมคุณสมบัติทางสถิติและมีการเปรียบเทียบประสิทธิภาพกับการแจกแจงอื่นๆ ซึ่งผลการศึกษาพบว่าการแจกแจงดังกล่าวมีประสิทธิภาพและมีความสอดคล้องยิ่งยวดกับข้อมูลจริงมากกว่าการแจกแจงเรย์ลี-โลแม็กซ์ที่เป็นการแจกแจงต้นกำเนิด (baseline distribution)

จากแนวคิดในการพัฒนาการแจกแจงช่วงชีวิตของ Venegas, et al. [7] และ Hammood, et al. [14] ที่ได้กล่าวมาในข้างต้น ในบทความวิจัยนี้จึงนำเสนอการพัฒนาการแจกแจงโลแม็กซ์-เรย์ลีด้วยการตัดปลาย (truncated Lomax-Rayleigh distribution: TLRD) เนื่องจากตัวแบบการแจกแจงดังกล่าวมีรูปแบบที่ค่อนข้างง่ายไม่ซับซ้อน ซึ่งเป็นอีกแนวทางหนึ่งในการพัฒนาการแจกแจงดังกล่าวให้มีความสอดคล้องใกล้เคียงกับข้อมูลจริงมากขึ้นพร้อมศึกษาคุณสมบัติทางสถิติ ได้แก่ ฟังก์ชันการรอดชีพ ฟังก์ชันพิบัติ โมเมนต์ ฟังก์ชันควอนไทล์ การประมาณค่าพารามิเตอร์ ตลอดจนการประยุกต์กับข้อมูลจริง การแจกแจงที่มีความสอดคล้องใกล้เคียงกับข้อมูลช่วงชีวิตที่เป็นข้อมูลจริงเมื่อนำไปประยุกต์สำหรับงานด้านแผนการควบคุมคุณภาพสินค้าในกระบวนการผลิตหรือการตรวจสอบคุณภาพของสินค้ามีประสิทธิภาพยิ่งขึ้นในอนาคต

การแจกแจงโลแม็กซ์-เรย์ลีตัดปลาย (Truncated Lomax-Rayleigh Distribution หรือ TLRD)

1. การตัดปลายการแจกแจง (Truncating Distribution)

ในการศึกษาครั้งนี้จะนำเสนอการตัดปลายทั้ง 2 ด้าน (both sides truncated distribution) เมื่อกำหนดให้ $X \in (-\infty, \infty)$ เป็นตัวแปรสุ่มแบบต่อเนื่องและการแจกแจงในรูปแบบใดๆ ด้วยค่าพารามิเตอร์ Θ PDF และ CDF ของ X มีค่าจำกัดอยู่ภายในช่วง $[a, b]$ เมื่อ $-\infty < a \leq x \leq b < \infty$ แสดงดังสมการที่ (3) [15]

$$t(x; \Theta, a, b) = \frac{f(x; \Theta)}{F(b; \Theta) - F(a; \Theta)} ; -\infty < a \leq x \leq b < \infty ,$$

(3)

เมื่อ $f(x) = \begin{cases} t(x) ; a \leq x \leq b \\ 0 ; \text{otherwise} \end{cases}$, $t(x|a \leq x \leq b)$, $t(x) \geq 0$, สำหรับทุก ๆ ค่าของ x

2. การแจกแจงโลแม็กซ์-เรย์ลีย์ตัดปลายทั้งสองด้าน (both sides truncated Lomax-Rayleigh distribution)

ให้ X เป็นตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจง TLR ด้วยค่าพารามิเตอร์ α θ a และ b PDF และ CDF ของ X มีค่าจำกัดอยู่ภายในช่วง $[a, b]$, $a \leq x \leq b$ แสดงดังสมการที่ (4) และ (5) ตามลำดับ

$$t(x/a \leq x \leq b) = \frac{2\alpha x}{\left[(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha} \right] (\theta + x^2)^{\alpha+1}} \quad (4)$$

$$T(x/a \leq x \leq b) = \frac{\left[(\theta + x^2)^{-\alpha} - (\theta + a^2)^{-\alpha} \right]}{\left[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha} \right]} \quad (5)$$

เมื่อ $x > 0$ $a \leq x \leq b$ $\alpha > 0$ คือ พารามิเตอร์รูปร่าง (shape parameter) และ $\theta > 0$ คือ พารามิเตอร์บ่งขนาด (scale parameter) โดยภาพที่ 1 กราฟแสดงค่า PDF ของ LRD และ TLRD

$$\begin{aligned} \text{พิสูจน์} \quad \int_a^b \frac{2\alpha x}{\left[(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha} \right] (\theta + x^2)^{\alpha+1}} dx &= \frac{1}{\left[(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha} \right]} \int_a^b \frac{1}{(\theta + x^2)^{\alpha+1}} d(\theta + x^2) \\ &= \frac{\alpha}{\left[(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha} \right]} \left[\frac{(\theta + x^2)^{-\alpha}}{-\alpha} \right]_{x=a}^{x=b} \\ &= \frac{\alpha}{\left[(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha} \right]} \left[\frac{(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}}{-\alpha} \right] = 1 \end{aligned}$$

3. คุณสมบัติทางสถิติของ TLRD

3.1 ฟังก์ชันการรอดชีพและฟังก์ชันพิบัติ (survival function and hazard function)

ฟังก์ชันการรอดชีพ หรือ $s(x)$ คือ ความน่าจะเป็นที่สินค้าหรือผลิตภัณฑ์จะมีอายุการใช้งานเกินเวลา x แสดงดังสมการที่ (6) ส่วนฟังก์ชันพิบัติ หรือ $h(x)$ คือ อัตราส่วนระหว่าง PDF และ $s(x)$ แสดงดังสมการที่ (7) กราฟของ $s(x)$ และ $h(x)$ แสดงดังภาพที่ 2 ซึ่งจะเห็นได้ว่า $s(x)$ มีลักษณะกราฟลดลงอย่างรวดเร็วแบบเอกซ์โพเนนเชียล ส่วน $h(x)$ มีลักษณะกราฟคงที่ใกล้ค่าศูนย์และเพิ่มขึ้นอย่างรวดเร็ว

$$s(x/a \leq x \leq b) = 1 - T(x/a \leq x \leq b) = \frac{\left[(\theta + b^2)^{-\alpha} - (\theta + x^2)^{-\alpha} \right]}{\left[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha} \right]} \quad (6)$$

$$h(x/a \leq x \leq b) = \frac{t(x/a \leq x \leq b)}{s(x/a \leq x \leq b)} = \frac{2\alpha x(\theta + x^2)^{-(\alpha+1)}}{(\theta + x^2)^{-\alpha} - (\theta + b^2)^{-\alpha}} \quad (7)$$

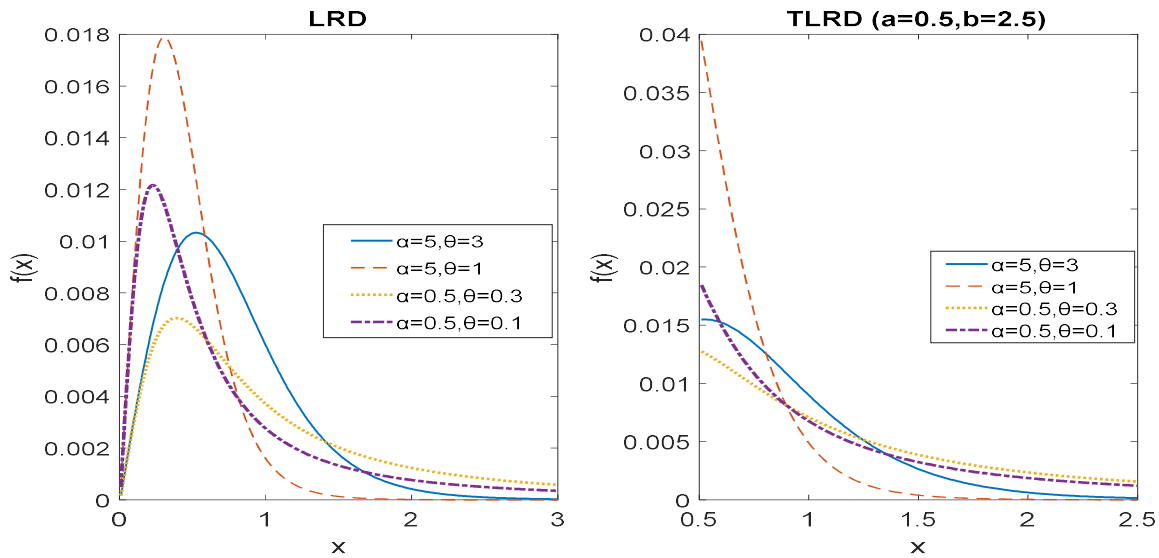


Figure 1 PDF plots of LRD and TLRD with specified parameters α and θ

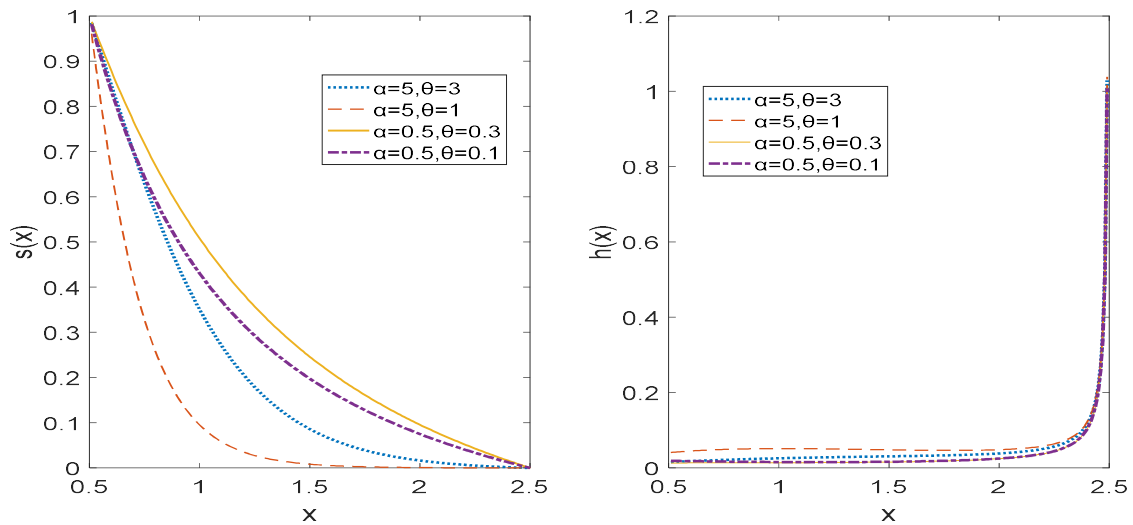


Figure 2 Plots of $s(x)$ and $h(x)$ of TLRD with specified parameters α and θ at $a=0.5$ and $b=2.5$

3.2 โมเมนต์ (Moment)

โมเมนต์ของ TLRD จะให้ค่าคุณลักษณะบางประการของการแจกแจงดังกล่าว เช่น ค่าเฉลี่ย ค่าความแปรปรวน ค่าความเบ้ และค่าความโด่ง เป็นต้น โมเมนต์ของ TLRD แสดงรายละเอียดดังนี้

กำหนดให้ $X \sim \text{TLRD}(\alpha, \theta, a, b)$ เมื่อ $r = 1, 2, \dots$ และ $\alpha > \frac{r}{2}$ โมเมนต์ของ TLRD เขียนได้ดังสมการที่ (8)

$$\mu_r = E(X^r) = \left[\frac{\alpha \theta^{\frac{r}{2}-\alpha}}{(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}} \right] \left[\frac{\Gamma\left(\alpha - \frac{r}{2}\right) \Gamma\left(\frac{r}{2} + 1\right)}{\Gamma(\alpha + 1)} \right] \quad (8)$$

เมื่อกำหนดให้ $\alpha = \alpha - \frac{r}{2}$ $\beta = \frac{r}{2} + 1$ และ $B(\alpha, \beta) = \int_0^1 (u^{\alpha-1} (1-u)^{\beta-1}) du = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}$

3.3 ฟังก์ชันควอนไทล์ (Quantile Function)

ฟังก์ชันควอนไทล์หรือ $Q(u)$ คือค่าของตัวแปรสุ่มซึ่งมีความน่าจะเป็นไม่เกินค่าความน่าจะเป็นของ u หรือเรียกอีกอย่างหนึ่งว่าฟังก์ชันจุดเปอร์เซ็นต์ (percent-point function) หรือฟังก์ชันการแจกแจงสะสมผกผัน (inverse cumulative distribution function) เขียนแทนด้วย $x = Q(u) = F^{-1}(u)$ [16] ซึ่งฟังก์ชันควอนไทล์มักใช้ในการสร้างตัวแปรสุ่มที่มีการแจกแจงตามการแจกแจงที่ต้องการศึกษาสำหรับ $Q(u)$ ของ TLRD แสดงดังสมการที่ (9)

$$Q(u) = \left[\left(u \left[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha} \right] + (\theta + a^2)^{-\alpha} \right)^{-\frac{1}{\alpha}} - \theta \right]^{\frac{1}{2}} \quad (9)$$

พิสูจน์ จาก CDF ในสมการที่ (5) ฟังก์ชันควอนไทล์ของ X จากเทคนิคการแปลงผกผัน (inverse transformation technique) เพื่อใช้ในการสร้างตัวแปรสุ่มสำหรับการสร้าง TLRD เมื่อกำหนด $x = Q(u) = F^{-1}(u)$ และ u เป็นตัวแปรสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม (uniform distribution) ในช่วง (0,1) และ $u = Q(x)$ ดังนั้น

$$\begin{aligned} u &= \frac{[(\theta + x^2)^{-\alpha} - (\theta + a^2)^{-\alpha}]}{[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}]} \\ u[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}] &= [(\theta + x^2)^{-\alpha} - (\theta + a^2)^{-\alpha}] \\ (\theta + x^2)^{-\alpha} &= u[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}] + (\theta + a^2)^{-\alpha} \\ x^2 &= \left(u[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}] + (\theta + a^2)^{-\alpha} \right)^{-\frac{1}{\alpha}} - \theta \\ x &= \left[\left(u[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha}] + (\theta + a^2)^{-\alpha} \right)^{-\frac{1}{\alpha}} - \theta \right]^{\frac{1}{2}} = Q(u) \end{aligned}$$

3.4 การสร้างตัวแปรสุ่มที่มีการแจกแจงโลแม็กซ์-เรย์สัติดปลาย

การจำลองตัวแปรสุ่มของ TLRD ด้วยเทคนิคการแปลงผกผันใช้เพื่อสร้างตัวแปรสุ่มสำหรับ TLRD โดยการกำหนดค่า $x_i = F^{-1}(u_i)$ เมื่อ u_i เป็นตัวแปรสุ่มของการแจกแจงยูนิฟอร์ม บนช่วง (0,1) $X_i, i = 1, 2, \dots, n$ โดยมีขั้นตอนการคำนวณดังนี้

ขั้นที่ 1 สร้าง $u_i, i = 1, 2, \dots, n$ ให้มีการแจกแจงแบบยูนิฟอร์มบนช่วง (0,1)

ขั้นที่ 2 กำหนด

$$x_i = \left[\left(u_i \left[(\theta + b^2)^{-\alpha} - (\theta + a^2)^{-\alpha} \right] + (\theta + a^2)^{-\alpha} \right)^{-\frac{1}{\alpha}} - \theta \right]^{\frac{1}{2}}$$

ตารางที่ 1 และตารางที่ 2 นำเสนอค่าประมาณภาวะน่าจะเป็นสูงสุดสำหรับพารามิเตอร์ α และ θ เมื่อกำหนดค่า $a=0.5$ และ $b=3.5$ สำหรับ TLRD ใน 4 กรณี ได้แก่ กรณีที่ 1 $\alpha, \theta < 1$ ($\alpha=0.8, \theta=0.5, a=0.5, b=3.5$) กรณีที่ 2 $\alpha < 1, \theta > 1$ ($\alpha=0.7, \theta=2.5, a=0.5, b=3.5$) กรณีที่ 3 $\alpha > 1, \theta < 1$ ($\alpha=1.5, \theta=0.5, a=0.5, b=3.5$) และกรณีที่ 4 $\alpha, \theta > 1$ ($\alpha=1.5, \theta=2.5, a=0.5, b=3.5$)

เมื่อกำหนดค่า $n = 50 \ 100 \ 500$ และ 1000 ค่าประมาณพารามิเตอร์ θ คำนวณได้จาก $\hat{\theta} = \frac{\sum_{i=1}^{100} \hat{\theta}_i}{100}$ โดยมีการทำซ้ำ 100 รอบ

และค่าความคลาดเคลื่อนมาตรฐาน (standard error หรือ SE) คำนวณได้จาก $SE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{100} (\hat{\theta}_i - \hat{\theta})^2}{99}} / \sqrt{n}$ ซึ่งการคำนวณ

ดังกล่าวแสดงในตารางที่ 1 และตารางที่ 2 ซึ่งจะเห็นได้ว่าค่าความคลาดเคลื่อนมาตรฐานจากการประมาณค่าพารามิเตอร์จะลดลงเมื่อขนาดตัวอย่าง n เพิ่มขึ้น

3.5 การประมาณค่าพารามิเตอร์ (Parameter Estimation)

การประมาณค่าพารามิเตอร์ของ TLRD จะใช้วิธีตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood estimation) หรือ MLE เมื่อกำหนดให้ $\tilde{x} = (x_1, x_2, \dots, x_n)$ เป็นตัวแปรสุ่มขนาด n และ $X \sim \text{TLRD}(\alpha, \theta, a, b)$ เมื่อกำหนดให้ a และ b คือค่าต่ำที่สุดและค่าสูงที่สุดที่ปรากฏในชุดของข้อมูลซึ่งมีคุณสมบัติเป็นอิสระและมีการแจกแจงเดียวกัน (independent and identically distributed) ของตัวแปรสุ่มตั้งแต่สองค่าขึ้นไป ฟังก์ชันภาวะน่าจะเป็นเขียนได้ดังสมการที่ (10)

$$L(x_i; \alpha, \theta, a, b) = \prod_{i=1}^n \left(\frac{2\alpha x_i}{\left[(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha} \right] (\theta + x_i^2)^{\alpha+1}} \right) \quad (10)$$

ฟังก์ชันล็อกภาวะน่าจะเป็นสูงสุดแสดงดังสมการที่ (11)

$$LL(x_i; \alpha, \theta, a, b) = n \log \left(\frac{2\alpha}{(\theta + a^2)^{-\alpha} - (\theta + b^2)^{-\alpha}} \right) + \sum_{i=1}^n \log(x_i) - (\alpha + 1) \sum_{i=1}^n \log(\theta + x_i^2) \quad (11)$$

การหาคำตอบของค่าประมาณพารามิเตอร์ α และ θ คำนวณโดย

$$\frac{\partial(LL(x_i; \alpha, \theta, a, b))}{\partial \alpha} = 0 \text{ และ } \frac{\partial(LL(x_i; \alpha, \theta, a, b))}{\partial \theta} = 0$$

โดยใช้ระเบียบวิธีนิวตัน-ราฟสัน (Newton-Raphson method)

Table 1 MLE and SE values under TLRD based on simulated data for $\alpha < 1$

$\alpha = 0.8, \theta = 0.5, a = 0.5, b = 3.5$					$\alpha = 0.7, \theta = 2.5, a = 0.5, b = 3.5$				
n	α	SE	θ	SE	n	α	SE	θ	SE
50	0.8746	0.2448	0.5489	0.2117	50	0.8781	0.2650	2.6480	0.9123
100	0.8168	0.0955	0.6142	0.1332	100	0.8566	0.2463	2.6093	0.6517
500	0.8391	0.0515	0.5360	0.0479	500	0.8257	0.1187	2.9205	0.4418
1000	0.8084	0.0358	0.5207	0.0413	1000	0.7227	0.0749	2.5569	0.2417

Table 2 MLE and SE values under TLRD based on simulated data for $\alpha > 1$

$\alpha = 1.5, \theta = 0.5, a = 0.5, b = 3.5$					$\alpha = 1.5, \theta = 2.5, a = 0.5, b = 3.5$				
n	α	SE	θ	SE	n	α	SE	θ	SE
50	1.7860	0.3378	0.7440	0.3110	50	1.4880	0.3517	3.0192	0.8871
100	1.5142	0.1889	0.5193	0.1126	100	1.6899	0.2677	2.9288	0.4935
500	1.5174	0.0470	0.5491	0.0326	500	1.6868	0.0905	2.9142	0.1876
1000	1.5146	0.0171	0.5056	0.0114	1000	1.5908	0.0637	2.7245	0.1326

การประยุกต์กับข้อมูลจริง

TLRD ถูกนำมาประยุกต์กับข้อมูลช่วงชีวิต (lifetime) ที่เป็นข้อมูลจริง 5 ชุดรายละเอียดของข้อมูลแสดงดังนี้

ข้อมูลชุดที่ 1: ระยะเวลาที่ระบบเครื่องปรับอากาศของเครื่องบินจะเสื่อมสภาพ [17] (หน่วย : ไม่ระบุ): 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95

ข้อมูลชุดที่ 2: ระยะเวลาการอยู่รอดของกลุ่มผู้ป่วยที่เป็นมะเร็งศีรษะและคอ ซึ่งรักษาโดยใช้รังสีรักษา [18] (หน่วย : สัปดาห์): 6.53, 7, 10.42, 14.48, 16.10, 22.70, 34, 41.55, 42, 45.28, 49.40, 53.62, 63, 64, 83, 84, 91, 108, 112, 129, 133, 133, 139, 140, 140, 146, 149, 154, 157, 160, 160, 165, 146, 149, 154, 157, 160, 160, 165, 173, 176, 218, 225, 241, 248, 273, 277, 297, 405, 417, 420, 440, 523, 583, 594, 1101, 1146, 1417

ข้อมูลชุดที่ 3: ข้อมูลการชำรุดของแผ่นวาล์วในกระบอกสูบจำนวน 13 รายการ [19] (หน่วย : ไม่ระบุ): 333, 119, 280, 245, 125, 242, 114, 233, 494, 131, 246, 73, 254

ข้อมูลชุดที่ 4-5 เป็นข้อมูลความล้มเหลวของฮาร์ดดิสก์ (Hard Drive Failure Data) [20] (หน่วย : รายวัน) ที่เก็บรวบรวมมาระหว่างเดือนมกราคม 2016 ถึงธันวาคม 2017 จำนวน 2 ชนิด รายการข้อมูลดังนี้

ข้อมูลชุดที่ 4: 14, 3, 14, 21, 104, 16, 23, 3, 21, 48, 15, 27, 37, 52, 1, 17, 46

ข้อมูลชุดที่ 5: 403, 62, 62, 54, 96, 62, 62, 62, 18

ประสิทธิภาพของการแจกแจงที่พัฒนาขึ้นจะทำการเปรียบเทียบโดย ค่าเกณฑ์ข้อสนเทศของอาไคแคะ (Akaike Information Criterion) หรือ AIC ค่าเกณฑ์ข้อสนเทศของเบย์ (Bayesian Information Criterion) หรือ BIC และการทดสอบคอลโมโกรอฟ-สมิร์นอฟ (Kolmogorov–Smirnov test) หรือ KS- test ดังสมการที่ (12), (13) และ (14) ตามลำดับ [21]

$$AIC = -2(LL) + 2p \quad (12)$$

$$BIC = -2(LL) + p \log(n) \quad (13)$$

$$KS = \sup_x |F_n(x) - F_0(x)| \quad (14)$$

เมื่อ $F_0(x)$ คือ ฟังก์ชันการแจกแจงสะสมของตัวอย่าง $F_n(x)$ คือ ฟังก์ชันการแจกแจงสะสมของประชากรที่คาดไว้

n คือ ขนาดตัวอย่าง p คือ จำนวนพารามิเตอร์

ซึ่งค่าเหล่านี้หากมีค่าต่ำแสดงว่าข้อมูลจริงมีความสอดคล้องกับการแจกแจงที่ต้องการทดสอบ การประมาณค่าพารามิเตอร์และการทดสอบภาวะสารูปดี (goodness-of-fit test) แสดงดังตารางที่ 3

Table 3 MLE and statistic values under models based on real datasets

Dataset	Distribution	Maximum likelihood estimates				-LL	AIC	BIC	KS -test (p-value)
		α	θ	a	b				
1	LRD	0.4476	168.0763	-	-	153.3350	310.6699	313.4723	0.1314 (0.6304)
	TLRD	0.0845	38.8942	1	261	148.8096	305.6192	311.2241	0.1037 (0.8704)
2	LRD	0.8408	15,073.04	-	-	372.6788	749.3600	753.4785	0.1374 (0.2039)
	TLRD	0.5774	9,036.50	6.53	1,417	371.0816	750.1600	758.4049	0.1274 (0.2677)
3	LRD	0.02549	3×10^{-13}	-	-	120.4163	244.8327	245.9627	1.5185 (<0.0001)
	TLRD	1.0075	44,005.0796	73	494	76.8658	161.7315	163.9913	0.2169 (0.5055)
4	LRD	1.1693	497.6081	-	-	74.4459	152.8918	154.5582	0.1557 (0.7482)
	TLRD	0.4059	137.5215	1	104	73.5334	155.0668	158.3997	0.1493 (0.8286)
5	LRD	1.1406	5,046.3505	-	-	48.1282	100.2564	100.6508	0.3018 (0.3171)
	TLRD	0.7934	2,761.3980	18	403	47.1753	102.3505	103.1395	0.3042 (0.3085)

จากตารางที่ 3 เมื่อพิจารณาจากค่า KS-test ของข้อมูลจริงทั้ง 5 ชุดจาก LRD และ TLRD เมื่อกำหนดให้ค่า a และ b คือค่าต่ำที่สุดและค่าสูงที่สุดที่ปรากฏในข้อมูลแต่ละชุดตามลำดับ ผลการศึกษาพบว่าข้อมูลที่ 1 2 4 และ 5 มีความสอดคล้องกับทั้ง LRD และ TLRD เนื่องจากมีค่า KS-test ที่ให้ค่า p-value > 0.05 โดยข้อมูลชุดที่ 1 2 และ 4 วิธี TLRD มีค่า KS-test ที่ให้ค่า p-value สูงกว่า LRD เล็กน้อย ส่วนข้อมูลชุดที่ 3 ซึ่งเป็นข้อมูลการชำรุดของแผ่นวาล์วในกระบอกสูบ พบว่าข้อมูลดังกล่าวมีความสอดคล้องกับ TLRD แต่ไม่มีความสอดคล้องกับ LRD เมื่อพิจารณาจากค่า AIC และ BIC ของข้อมูลชุดที่ 2 4 และ 5 พบว่ามีค่าใกล้เคียงกัน โดย TLRD มีค่า AIC และ BIC สูงกว่า LRD เล็กน้อย ยกเว้นข้อมูลชุดที่ 1 และ 3 ซึ่ง TLRD มีค่า AIC และ BIC น้อยกว่า LRD

สรุปผลการวิจัย

ในงานวิจัยนี้นำเสนอการแจกแจงโลแม็กซ์-เรย์ลีที่มีการตัดปลายทั้งสองด้านหรือ TLRD และคุณสมบัติทางสถิติบางประการ ได้แก่ ฟังก์ชันการรอดชีพ ฟังก์ชันพิบัติ โมเมนต์ ฟังก์ชันควอนไทล์ ตลอดจนการประมาณค่าพารามิเตอร์ การแจกแจงดังกล่าวได้ถูกนำมาประยุกต์กับข้อมูลจริง 5 ชุด ที่เป็นข้อมูลทางการแพทย์และข้อมูลความล้มเหลวหรืออายุการใช้งานของอุปกรณ์ และเปรียบเทียบการทดสอบภาวะสารถีกับการแจกแจงโลแม็กซ์-เรย์ลี หรือ LRD ที่เป็นการแจกแจงต้นกำเนิด TLRD มีประสิทธิภาพและมีความสอดคล้องในการอธิบายข้อมูลช่วงชีวิต โดยเฉพาะกรณีข้อมูลชุดที่ 3 ที่เป็นข้อมูลการชำรุดของแผ่นวาล์วในกระบอกสูบ TLRD มีความสอดคล้องใกล้เคียงกับข้อมูลดังกล่าวในขณะที่ LRD ไม่สามารถเป็นตัวแทนในการอธิบายพฤติกรรมของข้อมูลดังกล่าวได้ การแจกแจงช่วงชีวิตที่พัฒนาจนมีความสอดคล้องใกล้เคียงกับข้อมูลจริงจะเป็นประโยชน์ในอนาคตในด้านการประยุกต์ในเรื่องการสร้างตัวแบบเพื่อพยากรณ์อายุการใช้งานของสินค้าหรือผลิตภัณฑ์และงานด้านการพัฒนาแผนการสุ่มตัวอย่างเพื่อการยอมรับ (acceptance sampling plan) ในโรงงานอุตสาหกรรม

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณทุนสนับสนุนการพัฒนายุทธศาสตร์ วิจัยและนวัตกรรม มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ภายใต้กองทุนส่งเสริมวิทยาศาสตร์ วิจัย และนวัตกรรม ประจำปีงบประมาณ พ.ศ.2566 (รหัสทุน ววน. FRB660012/0168) รหัสโครงการวิจัย FRB66E0645O.3

References

- [1] Lindley, D. V. (1958). Fiducial Distributions and Bayes' Theorem. *Journal of the Royal Statistical Society, Series B*, 20(1), 102-107.
- [2] Shanker, R., et al. (2013). A Two-Parameter Lindley Distribution for Modeling Waiting and Survival Times Data. *Applied Mathematics*, 4, 363-368.
- [3] Shanker, R. (2015). Akash Distribution and its Application. *International Journal of Probability and Statistics*, 4(3), 65-75.
- [4] Chankham, W., et al. (2022). Measurement of Dispersion of PM 2.5 in Thailand using Confidence Intervals for the Coefficient of Variation of an Inverse Gaussian Distribution. *PeerJ*, 10, e12988.
- [5] Lomax, K. S. (1954). Business Failures: Another Example of the Analysis of Failure Data. *Journal of the American Statistical Association*, 49, 847-852.
- [6] Rayleigh, L. (1896). *The Theory of Sound, Volume 2*. New York: Dover Publications.
- [7] Venegas, O., et al. (2019). Lomax-Rayleigh Distribution with an Application. *Applied Mathematics & Information Sciences*, 13(5), 741-748.
- [8] Falgore, J. Y., et al. (2021). Inverse Lomax-Rayleigh Distribution with Application. *Heliyon*, 7(11), E08383.
- [9] El-Hadidya, M. A. A. & Alfreedia, A. A. (2019). Internal Truncated Distributions: Applications to Wiener

-
- Process Range Distribution when Deleting a Minimum Stochastic Volatility Interval from its Domain. *Journal of Taibah University for Science*, 13 (1), 201–215.
- [10] Johnson, A. C. (2001). *On The Truncated Normal Distribution: Characteristics of Singly-And Doubly-Truncated Populations of Application in Management Science*, (Ph.D. Dissertation, Illinois Institute of Technology).
- [11] Iwueze, S. (2007). Some Implications of Truncating the $N(1, \sigma^2)$ Distribution to the Left at Zero. *Journal of Applied Sciences*, 7(2), 189-195.
- [12] Aryuyuen, S. (2018). Truncated Two-Parameter Lindley Distribution and its Application. *The Journal of Applied Science*, 17(1), 19-32.
- [13] Shukla, K. K. & Shanker, R. (2020). Truncated Akash Distribution: Properties and Applications. *Biometrics & Biostatistics International Journal*, 9(5), 179-184.
- [14] Hammood, H. K., et al. (2023). Truncated Rayleigh Lomax Distribution. *International Journal of Engineering Research & Technology*, 12(1), 196-210.
- [15] Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. Oxford: The Oxford University Press.
- [16] Aryuyuen, S. & Bodhisuwan, W. (2019). The Truncated Power Lomax Distribution: Properties and Applications. *Walailak Journal of Science and Technology*, 16(9), 655-668.
- [17] Linhart, H. & Zucchini, W. (1986). *Model Selection*. New York: John Wiley.
- [18] Efron, B. (1988). Logistic Regression, Survival Analysis and the Kaplan-Meier Curve. *Journal of the American Statistical Association*, 83(402), 414-425.
- [19] Loukopoulos, P., et al. (2017). Reciprocating Compressor Prognostics of an Instantaneous Failure Mode Utilising Temperature Only Measurements. *Applied Acoustics*, 147, 77-86.
- [20] Zeydan, E. (2019). *Hard Drive Failure Data*. Available from <https://www.kaggle.com/datasets/ezeydan/hard-drive-failure-data?resource=download>. Accessed date: 18 August 2023.
- [21] Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19, 716–723.