การประมาณค่าความแปรปรวนของการแจกแจงพาเรโต: การศึกษาเชิงจำลอง
The Variance Estimation for Pareto Distribution: A Simulation Study

วราฤทธิ์ พานิชกิจโกศลกุล[1]*

[1]ผู้ช่วยศาสตราจารย์ ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต จังหวัดปทุมธานี 12121

บทคัดย่อ

การวิจัยครั้งนี้เสนอวิธีการประมาณค่าความแปรปรวนของการแจกแจงพาเรโต และเปรียบเทียบวิธีการประมาณค่าความแปรปรวน 3 วิธี คือ 1) วิธีอย่างง่าย 2) วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และ 3) วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ โดยการเปรียบเทียบค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวประมาณค่าความแปรปรวน การวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล ผลการวิจัยสรุปได้ดังนี้ วิธีอย่างง่าย ให้ค่า $|Bias|$ ต่ำที่สุด เกือบทุกกรณีที่ศึกษา รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ตามลำดับ ส่วนวิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ให้ค่า MAE และ MSE ต่ำที่สุด ในทุกกรณีที่ศึกษา รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีอย่างง่าย ตามลำดับ

Abstract

This research proposes the variance estimation method of Pareto data and to compare three variance estimation methods. Those methods are simple method, the adjusted variance method with term $(n - 1)/n$, and the adjusted variance method with term $(n - 2)/n$. The research was considered by the absolute of bias ($|Bias|$), the mean absolute errors (MAE) and the mean square errors (MSE) of variance estimator. This research used Monte Carlo Simulation method. Results of this research show that the $|Bias|$ of simple method is the lowest for almost all conditions. We can arrange the methods as follows: the adjusted variance method with term $(n - 1)/n$ and the adjusted variance method with term $(n - 2)/n$, respectively. For all situations, the MAE and MSE the of adjusted variance method with term $(n - 2)/n$ is the lowest. We can arrange the methods as follows: the adjusted variance method with term $(n - 1)/n$ and simple method, respectively.

คำสำคัญ : การจำลองแบบมอนติคาร์โล การประมาณค่าความแปรปรวน การแจกแจงพาเรโต
Keywords : Monte Carlo Simulation, Variance Estimation, Pareto Distribution

* *ผู้นิพนธ์ประสานงานไปรษณีย์อิเล็กทรอนิกส์ wararit@mathstat.sci.tu.ac.th โทร. 02-564-4440 ต่อ 2101-3*

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

1. บทนำ

การแจกแจงพาเรโต (Pareto Distribution) เริ่มเป็นที่รู้จักตั้งแต่ปี ค.ศ. 1895 (Pareto 1985) ในครั้งแรกเรียกการแจกแจงดังกล่าวว่า การแจกแจงรายได้ (Income Distribution) เนื่องจากมีที่มาจากรูปแบบการกระจายรายได้ นักเศรษฐศาสตร์หรือผู้สนใจทั่วไปสามารถใช้การแจกแจงพาเรโตในการอธิบายลักษณะของการกระจายรายได้ โดยจากข้อมูลรายได้ของกลุ่มตัวอย่างที่เก็บมาอย่างสุ่ม สามารถนำไปใช้อนุมานเพื่อหารายได้ของประชากรทั้งหมด นักเศรษฐศาสตร์หรือผู้สนใจก็จะทราบได้ว่าประชากรส่วนใหญ่มีรายได้เท่าใด มีประชากรกี่คนที่มีรายได้ในช่วงที่สนใจ ซึ่งจากข้อมูลการกระจายรายได้เหล่านี้สามารถหาความเหลื่อมล้ำของรายได้ประชากร ผู้บริหารหรือนักพัฒนาก็จะสามารถนำข้อมูลการกระจายรายได้มาวางแผนเพื่อพัฒนา หรือหากลยุทธ์ที่เหมาะสมในการส่งเสริมคุณภาพชีวิตของประชากรให้ดีขึ้น นอกจากนั้น ค่าของเบี้ยความเสียหายในธุรกิจประกันภัย (Insurance Claims Data) มักมีการแจกแจงพาเรโต (ศึกษาเพิ่มเติมใน Brazauskasand Serfling 2003, Cebrin et al. 2003, Zisheng and Chi 2006) นอกจากนี้ ยังมีการประยุกต์ใช้การแจกแจงพาเรโตในงานวิจัยด้านต่าง ๆ อาทิ วิศวกรรมลม (Wind Engineering) (Holmes and Moriarty 1999) ปริมาณน้ำฝน (Van Montfort and Witter 1986) และช่วงเวลาที่ใช้รับส่งไฟล์บนเว็บไซต์ (File Transmission Duration in Web) (Nossenson and Attiya 2004) เป็นต้น

ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function: pdf) ของการแจกแจงพาเรโต เขียนได้ดังนี้ (Evans et al. 2000)

$$f(x|a,b) = \frac{ba^b}{x^{b+1}}, \quad x \geq a > 0, \; b > 0$$

ค่าเฉลี่ยและความแปรปรวนของการแจกแจงพาเรโต มีดังนี้ ค่าเฉลี่ย $(\mu) = \frac{ab}{b-1}, \; b > 1$ และความแปรปรวน $(\sigma^2) = \frac{ba^2}{(b-1)^2(b-2)}, \; b > 2$

กราฟของฟังก์ชันความหนาแน่นของความน่าจะเป็นแสดงดังรูปที่ 1

วิธีการประมาณค่าพารามิเตอร์ a และ b มีการศึกษาอย่างละเอียดโดย Johnson, Kotz และ Balakrishnan (1994) ซึ่งสามารถประมาณค่าได้หลายวิธี เช่น วิธีกำลังสองน้อยที่สุด (Least Squares Method) วิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method: ML) และวิธีการประมาณแบบไม่เอนเอียงซึ่งมีความแปรปรวนต่ำสุดสม่ำเสมอ (Uniformly Minimum Variance Unbiased Estimation Method: UMVUE) เป็นต้น (Kleiber and Kotz 2003) เมื่อได้ค่าประมาณพารามิเตอร์ $\hat{a}$ และ $\hat{b}$ แล้ว ขั้นตอนต่อไปจะนำค่าประมาณพารามิเตอร์ไปคำนวณค่าสถิติต่าง ๆ เช่น ค่าเฉลี่ย หรือค่าความแปรปรวน เป็นต้น วิธีดังกล่าว เรียกว่า วิธีอย่างง่าย (Simple Method)

Like (1969) ได้เสนอตัวประมาณแบบไม่เอนเอียงซึ่งมีความแปรปรวนต่ำสุด (Minimum Variance Unbiased Estimator: MVUE) ตัวประมาณ MVUE ของ a คือ

$$a^* = X_{(1)}\left[1 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\ln X_{(i)} - \ln X_{(1)}\right)\right]$$

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนครฯ

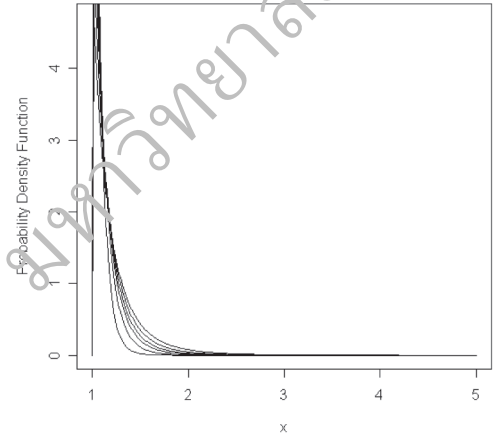
โดยที่ $X_{(i)}$ แทน สถิติลำดับอันดับที่ i (i th order statistics) นั่นคือ $X_{(i)} \leq ... \leq X_{(n)}$, $nb > 1$ และ $n \geq 2$ ส่วนตัวประมาณ MVUE ของ b คือ

$$b^* = \frac{n-2}{\sum_{i=1}^{n}(\ln X_{(i)} - \ln X_{(1)})}, n \geq 3$$

Baxter (1980) ปรับปรุงสูตรของตัวประมาณแบบไม่เอนเอียงซึ่งมีความแปรปรวนต่ำสุดให้ง่ายขึ้น โดยมีการนำเทอมค่าคงที่ซึ่งขึ้นอยู่กับขนาดตัวอย่างมาปรับในตัวประมาณค่าพารามิเตอร์ โดยตัวประมาณของ Baxter (1980) แสดงได้ดังนี้

$$b' = \frac{n-2}{n}\hat{b}$$

โดยที่ $\hat{b}$ คือ ตัวประมาณความควรจะเป็นสูงสุดของ ดังนั้น ผู้วิจัยจึงใช้แนวคิดดังกล่าวโดยนำเทอมค่าคงที่มาปรับในสูตรของตัวประมาณค่าความแปรปรวน ซึ่งเทอมค่าคงที่นี้จะทำให้ความคลาดเคลื่อนของตัวประมาณค่าความแปรปรวนลดลง โดยทดลองกับเทอมค่าคงที่ในรูปแบบต่าง ๆ



รูปที่ 1 ฟังก์ชันความหนาแน่นของการแจกแจงพาเรโต เมื่อ a = 1 และ b = ‘ 5,6,7,8,10 และ 15 (จากขวาไปซ้าย)

ดังนั้น ในการวิจัยครั้งนี้จึงเสนอวิธีการประมาณค่าความแปรปรวนของการแจกแจงพาเรโต และเปรียบเทียบวิธีการประมาณค่าความแปรปรวน 3 วิธี คือ 1) วิธีอย่างง่าย 2) วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และ 3) วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่

$(n - 2)/n$ โดยการเปรียบเทียบค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: *MAE*) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: *MAE*) ของตัวประมาณค่า

สำหรับขอบเขตของการวิจัยครั้งนี้ มีดังต่อไปนี้

1. กำหนดขนาดตัวอย่างที่ใช้ (n) เท่ากับ 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 200, 300, 500 และ 1000

2. กำหนดค่าพารามิเตอร์ a เท่ากับ 1

3. กำหนดค่าพารามิเตอร์ b เท่ากับ 5, 6, 7, 8, 10 และ 15

4. ประมาณค่าพารามิเตอร์ a และ b ด้วยวิธีความควรจะเป็นสูงสุด

โปรแกรมที่ใช้ในการจำลองเขียนด้วยโปรแกรม R เวอร์ชัน 2.10.1 (The R Development Core Team 2010) และกำหนดให้เครื่องคอมพิวเตอร์ทำงานซ้ำๆ กัน 10,000 ครั้งในแต่ละสถานการณ์

2. วิธีการศึกษา

การวิจัยครั้งนี้ใช้วิธีการจำลองแบบมอนติคาร์โล ซึ่งมีวิธีการดำเนินงาน ดังนี้

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

2.1 การจำลองข้อมูลที่ใช้ในการวิจัย

จำลองข้อมูลที่ใช้โดยการสร้างตัวแปรสุ่ม X ซึ่งมีการแจกแจงพาเรโตด้วยพารามิเตอร์ a และ b ตามที่กำหนดไว้ในขอบเขตของการวิจัย

2.2 การประมาณค่าพารามิเตอร์ a และ b ด้วยวิธีความควรจะเป็นสูงสุด

เมื่อสร้างตัวแปรสุ่ม X ที่มีการแจกแจงพาเรโตแล้ว ประมาณค่าพารามิเตอร์ a และ b ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) (Malik 1970, Krishnamoorthy 2006) ซึ่งตัวประมาณค่าความควรจะเป็นสูงสุดเป็นตัวประมาณที่คงเส้นคงวา (Consistent Estimators) (Quandt 1966) ตัวประมาณความควรจะเป็นสูงสุดของ a คือ

$$\hat{a} = X_{(1)} = \min\{X_1,...,X_n\}$$

ตัวประมาณความควรจะเป็นสูงสุดของ b คือ

$$\hat{b} = \frac{1}{\ln\left(\frac{GM}{\hat{a}}\right)}$$

โดยที่ $\ln(\cdot)$ แทน ลอการิทึมธรรมชาติ และ GM แทน ค่าเฉลี่ยเรขาคณิต (Geometric Mean) ซึ่งมีค่าเท่ากับ $\left(\prod_{i=1}^{n} X_i\right)^{1/n}$

2.3 การประมาณค่าความแปรปรวนของข้อมูลด้วยวิธีการที่ศึกษา

เมื่อประมาณค่าพารามิเตอร์ a และ b นำค่าประมาณ $\hat{a}$ และ $\hat{b}$ ที่ได้ ประมาณค่าความแปรปรวน ด้วยวิธีการที่ศึกษา

2.3.1 วิธีอย่างง่าย ตัวประมาณค่าความแปรปรวน คือ

$$\hat{\sigma}_S^2 = \frac{\hat{b}\hat{a}^2}{(\hat{b}-1)^2(\hat{b}-2)}$$

2.3.2 วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ ตัวประมาณค่าความแปรปรวน คือ

$$\hat{\sigma}_{C1}^2 = \left(\frac{n-1}{n}\right)\frac{\hat{b}\hat{a}^2}{(\hat{b}-1)^2(\hat{b}-2)}$$

2.3.3 วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ตัวประมาณค่าความแปรปรวน คือ

$$\hat{\sigma}_{C2}^2 = \left(\frac{n-2}{n}\right)\frac{\hat{b}\hat{a}^2}{(\hat{b}-1)^2(\hat{b}-2)}$$

2.4 การคำนวณค่าความเอนเอียงสัมบูรณ์ ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณจากวิธีการประมาณค่าแต่ละวิธี แล้วทำการเปรียบเทียบ

ค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ของตัวประมาณ คำนวณจาก

$$|Bias| = \left|\frac{\sum_{i=1}^{10,000}\hat{\sigma}_{ki}^2}{10,000} - \sigma^2\right|$$

ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* ของตัวประมาณ คำนวณจาก

$$MAE = \frac{\sum_{i=1}^{10,000}\left|\sigma^2 - \hat{\sigma}_{ki}^2\right|}{10,000}$$

ค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ของตัวประมาณ คำนวณจาก

$$MSE = \frac{\sum_{i=1}^{10,000}(\sigma^2 - \hat{\sigma}_{ki}^2)^2}{10,000}$$

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

เมื่อ σ^2 แทน ค่าความแปรปรวนของประชากร

$\hat{\sigma}^2_{ki}$ แทน ค่าประมาณของความแปรปรวน ซึ่งคำนวณจากวิธีที่ k ในการทำซ้ำครั้งที่ i

k แทน วิธีการประมาณค่าพารามิเตอร์

3. ผลการศึกษาและอภิปรายผล

ในการนำเสนอผลการวิจัยเพื่อความสะดวก จะใช้สัญลักษณ์แทนความหมายต่าง ๆ ดังนี้

S หมายถึง วิธีอย่างง่าย

C_1 หมายถึง วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$

C_2 หมายถึง วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$

$|Bias|$ หมายถึง ค่าความเอนเอียงสัมบูรณ์

MAE หมายถึง ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย

MSE หมายถึง ค่าคลาดเคลื่อนกำลังสองเฉลี่ย

ผลการเปรียบเทียบค่าความเอนเอียงสัมบูรณ์ ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ นำเสนอดังตารางที่ 1 สรุปรายละเอียด ดังนี้

3.1 การเปรียบเทียบค่าความเอนเอียงสัมบูรณ์

เมื่อพิจารณาในภาพรวม เมื่อค่าพารามิเตอร์ b เท่ากับ 5 วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ให้ค่าต่ำที่สุด รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีอย่างง่าย ตามลำดับ

เมื่อค่าพารามิเตอร์ b เท่ากับ 6 วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ ให้ค่า $|Bias|$ ต่ำที่สุด รองลงมา คือ วิธีอย่างง่าย และวิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ตามลำดับ

เมื่อค่าพารามิเตอร์ b เท่ากับ 7, 8, 10 และ 15 วิธีอย่างง่าย ให้ค่า $|Bias|$ ต่ำที่สุด รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ และวิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ตามลำดับ

3.2 การเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย

ในทุกระดับของขนาดตัวอย่างและทุกระดับของค่าพารามิเตอร์ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ให้ค่า MAE ต่ำที่สุด รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีอย่างง่าย ตามลำดับ

3.3 การเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ย

ในทุกระดับของขนาดตัวอย่างและทุกระดับของค่าพารามิเตอร์ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ให้ค่า MSE ต่ำที่สุด รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีอย่างง่าย ตามลำดับ

จากผลการจำลอง พบว่า เมื่อ b มีค่าเพิ่มขึ้น จะทำให้ค่า $|Bias|$, MAE และ MSE ลดลง นั่นคือ ค่า $|Bias|$, MAE และ MSE แปรผกผันกับค่า b นอกจากนี้ ค่า $|Bias|$, MAE และ MSE แปรผกผันกับขนาดตัวอย่าง (n)

4. สรุป

4.1 สรุปผลการศึกษาและอภิปราย

ในการเปรียบเทียบค่าความเอนเอียงสัมบูรณ์

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับการประมาณค่าความแปรปรวนของการแจกแจงพาเรโต สรุปผลได้ ดังนี้

วิธีอย่างง่าย ให้ค่าความเอนเอียงสัมบูรณ์ต่ำที่สุด เกือบทุกกรณีที่ศึกษา รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ตามลำดับ

วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ ให้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* และค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ต่ำที่สุด ในทุกระดับของขนาดตัวอย่าง และทุกระดับของค่าพารามิเตอร์ รองลงมา คือ วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 1)/n$ และวิธีอย่างง่าย ตามลำดับ

จากผลการจำลอง พบว่า เมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นเข้าใกล้อนันต์ $(n \to \infty)$ ค่าความเอนเอียง (Bias) ของตัวประมาณทุกวิธีมีค่าเข้าใกล้ศูนย์ นั่นคือ ตัวประมาณค่าทุกวิธีที่นำเสนอเป็นตัวประมาณค่าที่ไม่เอนเอียงเมื่อขนาดตัวอย่างเข้าใกล้อนันต์ (Asymptotically Unbiased Estimators) ดังนั้นเมื่อ $n \to \infty$ ค่าประมาณของความแปรปรวนลู่เข้าสู่ค่าความแปรปรวนของประชากร $\sigma^2 = ba^2/[(b-1)^2(b-2)]$

ดังนั้น ในการประมาณค่าความแปรปรวนของการแจกแจงพาเรโต ควรใช้วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n - 2)/n$ เนื่องจากเป็นวิธีที่ให้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* และค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ต่ำที่สุด

ตารางที่ 1 ค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* และค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ของตัวประมาณค่าความแปรปรวนของการแจกแจงพาเรโต เมื่อค่าพารามิเตอร์ a เท่ากับ 0.5

n	b	$\|Bias\|$			*MAE*			*MSE*		
		วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2
5	5	0.0273777	0.0166938	0.0060099	0.6744705	0.4314601	0.2425757	0.0702653	0.0587800	0.0479355
	6	0.0017141	0.0016287	0.0049716	0.1084494	0.0694084	0.0390655	0.0226606	0.0195874	0.0169263
	7	0.0013456	0.0008679	0.0030815	0.0010246	0.0006554	0.0003777	0.0103777	0.0092734	0.0084518
	8	0.0008067	0.0007152	0.0022371	0.0017347	0.0011103	0.0006293	0.0070304	0.0063141	0.0058088
	10	0.0001328	0.0008779	0.0016229	0.0000384	0.0000253	0.0000164	0.0033432	0.0030668	0.0029144
	15	0.0001990	0.0004536	0.0007081	0.0000029	0.0000020	0.0000015	0.0011136	0.0010478	0.0010346
6	5	0.0122309	0.0058521	0.0005266	0.7850373	0.5450951	0.3488393	0.0645266	0.0556867	0.0473550
	6	0.0029421	0.0000483	0.0030386	0.0038952	0.0026990	0.0017366	0.0166368	0.0149719	0.0136126
	7	0.0008098	0.0009455	0.0027009	0.0003926	0.0002730	0.0001815	0.0090655	0.0083074	0.0077541
	8	0.0002458	0.0009290	0.0021037	0.0001492	0.0001044	0.0000707	0.0059804	0.0055156	0.0052102
	10	0.0000697	0.0007011	0.0013325	0.0000282	0.0000201	0.0000143	0.0030701	0.0028564	0.0027389
	15	0.0001654	0.0003832	0.0006009	0.0000023	0.0000017	0.0000014	0.0010310	0.0009818	0.0009697

ขีดเส้นใต้ หมายถึง วิธีการประมาณค่าที่ให้ค่า $|Bias|$ ต่ำที่สุด หรือค่า *MAE* ต่ำที่สุด หรือค่า *MSE* ต่ำที่สุด

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ตารางที่ 1 ค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* และค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ของตัวประมาณค่าความแปรปรวนของการแจกแจงพาเรโต เมื่อค่าพารามิเตอร์ a เท่ากับ 0.5 (ต่อ)

n	b	$\|Bias\|$			*MAE*			*MSE*		
		วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2
7	5	0.0129865	0.0074111	0.0018356	0.1429927	0.1049869	0.0728728	0.0420832	0.0376442	0.0335790
	6	0.0032515	0.0006442	0.0019632	0.0028941	0.0021189	0.0014751	0.0152229	0.0139328	0.0128838
	7	0.0009081	0.0006105	0.0021291	0.0003037	0.0002229	0.0001591	0.0085017	0.0078575	0.0073832
	8	0.0000769	0.0009059	0.0018887	0.0000996	0.0000740	0.0000544	0.0054134	0.0050651	0.0048308
	10	0.0001246	0.0006580	0.0011913	0.0000197	0.0000149	0.0000114	0.0028162	0.0026534	0.0025629
	15	0.0001247	0.0003172	0.0005096	0.0000020	0.0000015	0.0000013	0.0009717	0.0009276	0.0009129
8	5	0.0078485	0.0048163	0.0017842	1.6760710	1.2832620	0.9428495	0.0448877	0.0404822	0.0364037
	6	0.0027353	0.0005184	0.0016985	0.0045803	0.0035013	0.0025751	0.0140302	0.0130718	0.0122468
	7	0.0008197	0.0004980	0.0018158	0.0002745	0.0002099	0.0001573	0.0079934	0.0074739	0.0070892
	8	0.0003209	0.0005696	0.0014600	0.0001272	0.0000976	0.0000736	0.0051927	0.0048749	0.0046580
	10	0.0001035	0.0005728	0.0010421	0.0000163	0.0000128	0.0000102	0.0026162	0.0024827	0.0024086
	15	0.0000880	0.0002609	0.0004339	0.0000018	0.0000015	0.0000012	0.0009296	0.0008887	0.0008718
9	5	0.0131934	0.0088340	0.0044745	0.1609880	0.1271409	0.0973025	0.0321933	0.0296540	0.0273572
	6	0.0022124	0.0002999	0.0016125	0.0019780	0.0015591	0.0011962	0.0128403	0.0120435	0.0114128
	7	0.0007992	0.0003699	0.0015389	0.0002103	0.0001658	0.0001292	0.0075200	0.0070832	0.0067551
	8	0.0000911	0.0006749	0.0014408	0.0000685	0.0000542	0.0000432	0.0047750	0.0045363	0.0043792
	10	0.0000661	0.0004874	0.0009087	0.0000140	0.0000113	0.0000093	0.0025101	0.0023952	0.0023290
	15	0.0000916	0.0002450	0.0003983	0.0000014	0.0000011	0.0000010	0.0008487	0.0008170	0.0008053
10	5	0.0133202	0.0093840	0.0054478	0.4502253	0.3646268	0.2880603	0.0315317	0.0292885	0.0272577
	6	0.0015817	0.0000764	0.0017346	0.0004912	0.0003958	0.0003158	0.0117277	0.0110808	0.0105794
	7	0.0004391	0.0005770	0.0015932	0.0001492	0.0001210	0.0000979	0.0069460	0.0066042	0.0063540
	8	0.0001894	0.0004093	0.0012091	0.0000590	0.0000480	0.0000392	0.0046576	0.0044400	0.0042888
	10	0.0000639	0.0004433	0.0008227	0.0000123	0.0000102	0.0000085	0.0023794	0.0022798	0.0022204
	15	0.0000613	0.0002024	0.0003434	0.0000013	0.0000011	0.0000010	0.0008278	0.0007970	0.0007850
15	5	0.0037786	0.0017859	0.0002018	0.0014018	0.0012119	0.0010423	0.0189252	0.0181359	0.0174663
	6	0.0009186	0.0001426	0.0012039	0.0002197	0.0001907	0.0001659	0.0093426	0.0090082	0.0087451
	7	0.0004580	0.0002207	0.0008994	0.0000732	0.0000636	0.0000556	0.0056738	0.0054780	0.0053334
	8	0.0000344	0.0004214	0.0008772	0.0000298	0.0000261	0.0000231	0.0037404	0.0036256	0.0035450
	10	0.0001031	0.0003534	0.0006037	0.0000069	0.0000061	0.0000056	0.0019164	0.0018656	0.0018384
	15	0.0000431	0.0001384	0.0002336	0.0000009	0.0000008	0.0000007	0.0006855	0.0006696	0.0006631
20	5	0.0021635	0.0007532	0.0006570	0.0006080	0.0005451	0.0004891	0.0155497	0.0150973	0.0147214
	6	0.0003851	0.0003841	0.0011534	0.0001333	0.0001203	0.0001092	0.0078341	0.0076400	0.0074946
	7	0.0002169	0.0002800	0.0007770	0.0000473	0.0000427	0.0000389	0.0047838	0.0046631	0.0045750
	8	0.0000323	0.0003708	0.0007093	0.0000189	0.0000172	0.0000158	0.0031714	0.0031043	0.0030615
	10	0.0000655	0.0002551	0.0004447	0.0000049	0.0000045	0.0000042	0.0016647	0.0016320	0.0016136
	15	0.0000462	0.0001175	0.0001888	0.0000006	0.0000006	0.0000005	0.0005987	0.0005888	0.0005848

ขีดเส้นใต้ หมายถึง วิธีการประมาณค่าที่ให้ค่า $|Bias|$ ต่ำที่สุด หรือค่า *MAE* ต่ำที่สุด หรือค่า *MSE* ต่ำที่สุด

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ตารางที่ 1 ค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* และค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ของตัวประมาณค่าความแปรปรวนของการแจกแจงพาเรโต เมื่อค่าพารามิเตอร์ a เท่ากับ 0.5 (ต่อ)

n	b	$\|Bias\|$			*MAE*			*MSE*		
		วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2
25	5	0.0025044	0.0013625	0.0002207	0.0005587	0.0005110	0.0004676	0.0143368	0.0139854	0.0136879
	6	0.0007402	0.0001106	0.0005190	0.0001137	0.0001043	0.0000960	0.0072166	0.0070514	0.0069193
	7	0.0001343	0.0002599	0.0006542	0.0000352	0.0000325	0.0000302	0.0043028	0.0042205	0.0041630
	8	0.0000409	0.0002329	0.0005066	0.0000155	0.0000143	0.0000134	0.0028758	0.0028223	0.0027861
	10	0.0000344	0.0001873	0.0003402	0.0000041	0.0000038	0.0000035	0.0015062	0.0014813	0.0014674
	15	0.0000357	0.0000931	0.0001506	0.0000005	0.0000005	0.0000004	0.0005371	0.0005297	0.0005267
30	5	0.0017333	0.0008075	0.0001183	0.0003280	0.0003044	0.0002832	0.0124488	0.0121903	0.0119754
	6	0.0004004	0.0001129	0.0006263	0.0000861	0.0000804	0.0000753	0.0065485	0.0064350	0.0063470
	7	0.0001897	0.0001407	0.0004711	0.0000292	0.0000273	0.0000256	0.0039454	0.0038794	0.0038311
	8	0.0000605	0.0001683	0.0003971	0.0000122	0.0000114	0.0000108	0.0025845	0.0025433	0.0025155
	10	0.0000385	0.0001658	0.0002931	0.0000033	0.0000031	0.0000030	0.0013780	0.0013586	0.0013473
	15	0.0000270	0.0000752	0.0001234	0.0000004	0.0000004	0.0000004	0.0004890	0.0004835	0.0004810
40	5	0.0006893	0.0000211	0.0006472	0.0002149	0.0002038	0.0001939	0.0104805	0.0103466	0.0102390
	6	0.0003375	0.0000459	0.0004294	0.0000565	0.0000536	0.0000511	0.0055322	0.0054558	0.0053964
	7	0.0001005	0.0001450	0.0003906	0.0000203	0.0000193	0.0000184	0.0033558	0.0033160	0.0032880
	8	0.0000580	0.0001135	0.0002850	0.0000089	0.0000084	0.0000080	0.0022537	0.0022266	0.0022081
	10	0.0000297	0.0001254	0.0002211	0.0000025	0.0000023	0.0000023	0.0012070	0.0011956	0.0011894
	15	0.0000076	0.0000442	0.0000808	0.0000003	0.0000003	0.0000003	0.0004234	0.0004194	0.0004176
50	5	0.0008495	0.0003116	0.0002462	0.0001698	0.0001624	0.0001558	0.0094427	0.0093377	0.0092522
	6	0.0001650	0.0001383	0.0004416	0.0000430	0.0000413	0.0000398	0.0049372	0.0048870	0.0048494
	7	0.0000227	0.0001721	0.0003671	0.0000157	0.0000151	0.0000146	0.0029999	0.0029754	0.0029589
	8	0.0000514	0.0000857	0.0002228	0.0000069	0.0000066	0.0000064	0.0020190	0.0019988	0.0019841
	10	0.0000035	0.0000737	0.0001509	0.0000019	0.0000019	0.0000018	0.0010842	0.0010744	0.0010679
	15	0.0000071	0.0000560	0.0000849	0.0000002	0.0000002	0.0000002	0.0003763	0.0003744	0.0003738
75	5	0.0004887	0.0001350	0.0002188	0.0001061	0.0001030	0.0001003	0.0077652	0.0077108	0.0076666
	6	0.0001416	0.0000603	0.0002622	0.0000277	0.0000269	0.0000263	0.0040501	0.0040205	0.0039968
	7	0.0000494	0.0000809	0.0002112	0.0000099	0.0000096	0.0000094	0.0024373	0.0024215	0.0024104
	8	0.0000345	0.0000566	0.0001478	0.0000045	0.0000044	0.0000043	0.0016354	0.0016248	0.0016174
	10	0.0000060	0.0000574	0.0001087	0.0000013	0.0000012	0.0000012	0.0008771	0.0008721	0.0008691
	15	0.0000140	0.0000334	0.0000529	0.0000002	0.0000002	0.0000001	0.0003100	0.0003087	0.0003082
100	5	0.0005022	0.0002367	0.0000287	0.0000782	0.0000765	0.0000749	0.0066811	0.0066380	0.0066026
	6	0.0000927	0.0000582	0.0002092	0.0000202	0.0000197	0.0000194	0.0034564	0.0034399	0.0034278
	7	0.0000139	0.0000835	0.0001808	0.0000074	0.0000072	0.0000071	0.0021093	0.0020992	0.0020923
	8	0.0000057	0.0000736	0.0001416	0.0000033	0.0000033	0.0000032	0.0014251	0.0014189	0.0014149
	10	0.0000141	0.0000526	0.0000910	0.0000009	0.0000009	0.0000009	0.0007664	0.0007633	0.0007615
	15	0.0000034	0.0000181	0.0000328	0.0000001	0.0000001	0.0000001	0.0002678	0.0002668	0.0002662

ขีดเส้นใต้ หมายถึง วิธีการประมาณค่าที่ให้ค่า $|Bias|$ ต่ำที่สุด หรือค่า *MAE* ต่ำที่สุด หรือค่า *MSE* ต่ำที่สุด

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ตารางที่ 1 ค่าความเอนเอียงสัมบูรณ์ ($|Bias|$) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย *(MAE)* และค่าคลาดเคลื่อนกำลังสองเฉลี่ย *(MSE)* ของตัวประมาณค่าความแปรปรวนของการแจกแจงพาเรโต เมื่อค่าพารามิเตอร์ a เท่ากับ 0.5 (ต่อ)

n	b	$\|Bias\|$			*MAE*			*MSE*		
		วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2	วิธี S	วิธี C_1	วิธี C_2
200	5	0.0002117	0.0000804	0.0000508	0.0000358	0.0000354	0.0000350	0.0046649	0.0046494	0.0046365
	6	0.0000059	0.0000691	0.0001442	0.0000096	0.0000095	0.0000094	0.0024393	0.0024342	0.0024304
	7	0.0000125	0.0000362	0.0000849	0.0000036	0.0000036	0.0000036	0.0014963	0.0014926	0.0014901
	8	0.0000173	0.0000168	0.0000509	0.0000016	0.0000016	0.0000016	0.0010019	0.0009992	0.0009973
	10	0.0000010	0.0000203	0.0000396	0.0000005	0.0000005	0.0000004	0.0005345	0.0005332	0.0005325
	15	0.0000053	0.0000126	0.0000199	0.0000001	0.0000001	0.0000001	0.0001909	0.0001906	0.0001905
300	5	0.0001324	0.0000452	0.0000421	0.0000238	0.0000237	0.0000235	0.0038399	0.0038270	0.0038214
	6	0.0000572	0.0000070	0.0000431	0.0000065	0.0000064	0.0000064	0.0019991	0.0019954	0.0019924
	7	0.0000190	0.0000134	0.0000459	0.0000024	0.0000024	0.0000024	0.0012290	0.0012266	0.0012248
	8	0.0000030	0.0000257	0.0000483	0.0000011	0.0000011	0.0000011	0.0008241	0.0008230	0.0008224
	10	0.0000019	0.0000110	0.0000238	0.0000003	0.0000003	0.0000003	0.0004395	0.0004388	0.0004384
	15	0.0000005	0.0000054	0.0000103	0.0000000	0.0000000	0.0000000	0.0001549	0.0001546	0.0001545
500	5	0.0000290	0.0000232	0.0000753	0.0000137	0.0000136	0.0000136	0.0029460	0.0029427	0.0029398
	6	0.0000054	0.0000246	0.0000546	0.0000038	0.0000038	0.0000038	0.0015474	0.0015458	0.0015446
	7	0.0000107	0.0000301	0.0000495	0.0000015	0.0000015	0.0000014	0.0009594	0.0009588	0.0009584
	8	0.0000050	0.0000086	0.0000222	0.0000006	0.0000006	0.0000006	0.0006426	0.0006418	0.0006411
	10	0.0000058	0.0000135	0.0000212	0.0000002	0.0000002	0.0000002	0.0003455	0.0003453	0.0003452
	15	0.0000037	0.0000066	0.0000095	0.0000000	0.0000000	0.0000000	0.0001201	0.0001201	0.0001200
1000	5	0.0000569	0.0000308	0.0000047	0.0000069	0.0000069	0.0000069	0.0020797	0.0020786	0.0020777
	6	0.0000076	0.0000074	0.0000225	0.0000019	0.0000019	0.0000019	0.0010869	0.0010864	0.0010861
	7	0.0000026	0.0000123	0.0000220	0.0000007	0.0000007	0.0000007	0.0006635	0.0006632	0.0006630
	8	0.0000026	0.0000042	0.0000110	0.0000003	0.0000003	0.0000003	0.0004499	0.0004497	0.0004496
	10	0.0000018	0.0000020	0.0000057	0.0000001	0.0000001	0.0000001	0.0002424	0.0002423	0.0002422
	15	0.0000026	0.0000040	0.0000055	0.0000000	0.0000000	0.0000000	0.0000847	0.0000847	0.0000847

ขีดเส้นใต้ หมายถึง วิธีการประมาณค่าที่ให้ค่า $|Bias|$ ต่ำที่สุด หรือค่า *MAE* ต่ำที่สุด หรือค่า *MSE* ต่ำที่สุด

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

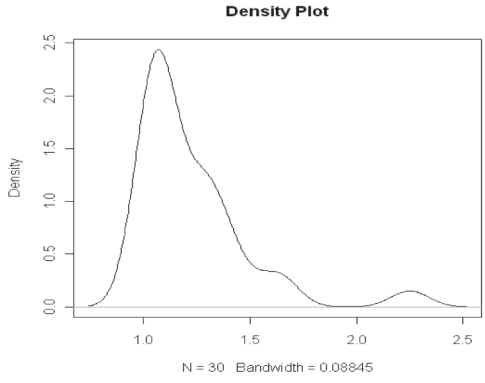
4.2 ตัวอย่างการประยุกต์ใช้

การประมาณค่าความแปรปรวนของการแจกแจงพาเรโตสามารถนำไปประยุกต์ใช้ได้ ดังนี้

สมมติว่าสุ่มตัวอย่างจากประชากรที่มีการแจกแจงพาเรโตจำนวน 30 ค่า ข้อมูลมีดังนี้

1.03, 1.06, 1.01, 1.12, 1.04, 1.01, 1.29, 1.11, 1.16, 1.24, 1.46, 1.28, 1.03, 1.02, 1.12, 1.32, 1.32, 1.05, 1.06, 1.16, 1.41, 2.25, 1.06, 1.62, 1.08, 1.65, 1.35, 1.09, 1.04, 1.29

กราฟแสดงการแจกแจงของข้อมูลตัวอย่างดังรูปที่ 2



รูปที่ 2 การแจกแจงของข้อมูลตัวอย่าง

จากข้อมูลข้างต้นประมาณค่าพารามิเตอร์ของการแจกแจงพาเรโต ดังนี้

$$\hat{a} = \min\{X_1,...,X_n\} = 1.01$$

$$\hat{b} = \frac{1}{\ln\left(\frac{GM}{\hat{a}}\right)} = \frac{1}{\ln\left(\frac{1.202602}{1.01}\right)}$$

$$= 5.729437$$

ขั้นต่อไปทำการประมาณค่าความแปรปรวนของการแจกแจงพาเรโต โดยใช้วิธีปรับค่าความแปรปรวนด้วยเทอมค่าคงที่ $(n-2)/n$ ดังนี้

$$\hat{\sigma}^2_{C2} = \left(\frac{n-2}{n}\right)\frac{\hat{b}\hat{a}^2}{(\hat{b}-1)^2(\hat{b}-2)}$$

$$\hat{\sigma}^2_{C2} = \left(\frac{28}{30}\right)\frac{5.729437(1.01)^2}{(4.729437)^2(3.729437)}$$

$$= 0.2090262$$

5. เอกสารอ้างอิง

Baxter, M.A. 1980. **Minimum variance unbiased estimation of the parameters of the Pareto distribution.** Metrika 27: 133-138.

Brazauskas, V., and Serfling, R. 2003. **Favorable estimates for fitting pareto models: A study using goodness-of-fit measures with actual data.** Astin Bulletin 33: 365-381.

Cebrin, A., Denuit, M., and Lambert, P. 2003. **Generalized pareto fit to the society of actuaries' large claims database.** North American Actuarial Journal 7: 18-36.

Evans, M., Hastings, N., and Peacock, B. 2000. **Statistical Distributions.** New York: John Wiley & Sons.

Holmes, J.D., and Moriarty, W.W. 1999. **Application of the generalized pareto distribution to extreme value analysis in wind engineering.** Journal of Wind Engineering and Industrial Aerodynamics 83: 1-10.

Johnson, N.L., Kotz, S., and Balakrishnan, N. 1994. **Continuous Univariate Distributions Vol.1.** New York: John Wiley & Sons.

Kleiber, C., and Kotz, S. 2003. **Statistical Size Distribution in Economics and Actuarial Sciences.** New York: John Wiley & Sons.

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

Krishnamoorthy, K. 2006. **Handbook of Statistical Distributions with Applications.** Florida: Chapman & Hall.

Like, J. 1969. **Minimum variance unbiased estimation of the parameters of power-function and Pareto's distribution.** Statistische Hefte 10: 104-110.

Malik, H.J. 1970. **Estimation of the parameters of Pareto distribution.** Metrika 15: 126-133.

Nossenson, R., and Attiya, H. 2004. **The distribution of file transmission duration in the web.** International Journal of Communication System 17: 407-419.

Pareto, V. 1985. **La legge della domanda. Giornale degli Economisti** 10: 59-68, 1895. English translation in Rivista di Politica. 1997. Economica 87: 691-700.

Quandt, R.E. 1966. **Old and new methods of estimation and the Pareto distribution.** Metrika 10: 55-82.

The R Development Core Team. 2010. **An introduction to R.** Vienna: R Foundation for Statistical Computing.

Van Montfort, M.A.J., and Witter, J.V. 1986. **The generalized pareto distribution applied to rainfall depths.** Hydrological Sciences Journal 31: 151-162.

Zisheng, O., and Chi, X. 2006. **Generalized pareto distribution fit to medical insurance claims data.** Applied Mathematics-A Journal of Chinese Universities 21: 21-29.

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร