การศึกษาเปรียบเทียบการประมาณพารามิเตอร์โดยตัวประมาณแบบกราฟที่มีการถ่วงน้ำหนักจากข้อมูลที่ถูกตัดปลายทางขวา

A Comparative Study on Parameter Estimation by Weighted Graphical Estimator from Right–Censored Data

อชิรญา สดรัมย์[1], อนุภาพ สมบูรณ์สวัสดี[2]

Achiraya Sodram, Anupap Somboonsavatdee

[1]นิสิตปริญญาโท ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

[2]อาจารย์ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

Abstract

The objective of this study is to compare parameter estimations of type-II right censored data by Maximum Likelihood Estimation (MLE), Graphical Estimation (GE) and Weighted Graphical Estimation Methods (WGE). The distribution of data under considerations in this study is Smallest Extreme Value (SEV). The comparison was done under conditions of sample sizes n = 20, 40, 80 and 120 with the censoring proportion p = 10%, 20% and 30% of sample. From the study, Relative Efficiency (RE) of Varience and Mean Square Error (MSE) shomefhat WGE was more efficient than the GE. When the WGE and the GE were compared under distribution of data, it found that the parameter estimations for μ (location parameter) and σ (scale parameter) by the WGE of Increasing Weighting were the most efficient estimation for SEV.

Keywords: *Maximum Likelihood Estimaion, Graphical Estimation, Weighted Graphical Estimation, Right-Censored Data*.

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบการประมาณค่าพารามิเตอร์ของข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2 ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method; MLE Method) วิธีการประมาณแบบกราฟ (Graphical Estimation Method; GE Method) และวิธีการประมาณแบบกราฟที่มีการถ่วงน้ำหนัก (Weighted Graphical Estimation Method; WGE Method) โดยการแจกแจงที่สนใจศึกษาคือ การแจกแจงค่าต่ำสุดขีด (Smallest Extreme Value Distribution) ข้อมูลที่ใช้ในการศึกษาได้จากการจำลองโดยมีขนาดตัวอย่าง (Sample Size; n) เท่ากับ 20, 40, 80 และ 120 ด้วยสัดส่วนของข้อมูลตัดปลายทางขวา (Censoring Proportion; *p*) เป็น 10%, 20% และ 30% ของขนาดตัวอย่าง จากการเปรียบเทียบค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency; RE) ของค่าความแปรปรวน และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย พบว่า การประมาณค่าพารามิเตอร์ด้วยวิธี WGE จะมีประสิทธิภาพมากกว่าการประมาณ

ค่าพารามิเตอร์ด้วยวิธี GE และเมื่อพิจารณาเปรียบเทียบวิธี WGE กับวิธี GE ภายใต้การแจกแจงที่กำหนดจะได้ว่า การประมาณค่าพารามิเตอร์ μ (location parameter) และ σ (scale parameter) ด้วยวิธี WGE แบบ Increasing Weighting จะมีประสิทธิภาพดีที่สุดสำหรับข้อมูลที่มีการแจกแจง SEV

คำสำคัญ: *การประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด, การประมาณแบบกราฟ, การประมาณแบบกราฟที่มีการถ่วงน้ำหนัก, ข้อมูลที่ถูกตัดปลายทางขวา*

บทนำ

การประมาณค่าพารามิเตอร์ เป็นขั้นตอนหนึ่งที่สำคัญในการอนุมานเชิงสถิติ (Statistical Inference) เป็นระเบียบวิธีทางสถิติที่ใช้ในการศึกษาวิจัย ดำเนินงานต่างๆ เพื่อหาข้อสรุปเกี่ยวกับประชากร

ในการศึกษาค่าที่บอกถึงลักษณะประชากร ซึ่งเรียกว่า พารามิเตอร์ นั้นจะทราบได้ก็ต่อเมื่อ ทำการแจงนับจากทุก ๆ หน่วยในประชากร สำหรับการสำรวจหรือการทดลองต่าง ๆ ซึ่งเป็นการศึกษาที่ต้องใช้ระยะเวลาเก็บข้อมูลที่เกิดขึ้นยาวนานมาก รวมทั้งค่าใช้จ่ายก็สูงไปด้วย เนื่องจากผู้วิจัยไม่สามารถคาดเดาได้ว่า ระยะเวลาการทดลองจะสิ้นสุดเมื่อใด เพื่อแก้ปัญหาเหล่านี้ ผู้ที่ทำการวิจัยอาจจะออกแบบการทดลองโดย กำหนดเวลาที่สิ้นสุดการเก็บข้อมูลไว้ล่วงหน้า (Fixed Censoring Time) และกำหนดจำนวนค่าสังเกตที่จะเกิดความล้มเหลวไว้ล่วงหน้า (Fixed Number of Uncensored Failure) โดยทำการทดลองไม่ครบตามขนาดตัวอย่าง ซึ่งมีผลทำให้ได้ข้อมูลที่เก็บรวบรวมมานั้นเป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete Data) โดยจะพบว่า ข้อมูลที่ได้นั้นเป็นข้อมูลที่ได้ถูกตัดปลายทางขวา (Right Censored Data) ข้อมูลลักษณะนี้เหมาะสำหรับงานวิจัยด้านการศึกษาทางการแพทย์ การศึกษาทางด้านประชากรศาสตร์ การศึกษาทางด้านธุรกิจประกันภัย และการศึกษาทางด้านอุตสาหกรรม เป็นต้น

เนื่องจากงานวิจัยส่วนใหญ่ มักพบว่าการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด เป็นวิธีที่มีประสิทธิภาพดีกว่าวิธีการประมาณอื่นๆ แต่ในทางปฏิบัติ การวิเคราะห์ข้อมูลที่ไม่สมบูรณ์ด้วยวิธีภาวะน่าจะเป็นสูงสุดนั้นทำได้ยาก เนื่องด้วยวิธีภาวะน่าจะเป็นสูงสุด ไม่มีรูปแบบสมการที่แน่นอน ทำให้คำนวณได้ยากและซับซ้อน และพบว่าวิธีการประมาณแบบกราฟมีความสะดวกและง่ายต่อการคำนวณมากกว่าวิธีภาวะน่าจะเป็นสูงสุด ดังงานวิจัยของ ขวัญรัตน์ ตั้งพิษฐานสกุล (2554) และประภาศิริ สุนทรศิริเวช (2555) จากการวิจัยในครั้งนี้ได้ทำการศึกษาจากตัวอย่างสุ่มที่มาจากประชากรที่มีการแจกแจงแบบ Location-Scale โดยศึกษาจากข้อมูลที่มีการแจกแจงที่มีลักษณะเบ้ซ้าย (การแจกแจงค่าต่ำสุดขีด) ในการวิจัยครั้งนี้ ผู้วิจัยมีความสนใจการประมาณแบบกราฟที่มีการถ่วงน้ำหนักที่แตกต่างกันออกไป

ดังนั้นผู้วิจัยสนใจที่จะทำการศึกษาเปรียบเทียบการประมาณพารามิเตอร์จากข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2 ซึ่งจะกำหนดจำนวนค่าสังเกตที่จะเกิดความล้มเหลวไว้ล่วงหน้า ด้วยวิธีภาวะน่าจะเป็นสูงสุด วิธีการประมาณแบบกราฟและวิธีการประมาณแบบกราฟที่มีการถ่วงน้ำหนัก

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาการประมาณค่าพารามิเตอร์ของข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2 คือ กำหนดจำนวนค่าสังเกตที่จะเกิดความล้มเหลวไว้ล่วงหน้า (Fixed Number of Uncensored Failure) ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method; MLE Method), วิธีการประมาณแบบกราฟ (Graphical Estimation Method; GE Method) และวิธีการประมาณแบบกราฟที่มีการถ่วงน้ำหนัก (Weighted Graphical Estimation

Method; WGE Method) โดยการแจกแจงที่สนใจศึกษาคือ การแจกแจงค่าต่ำสุดขีด (Smallest Extreme Value Distribution)

2. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณพารามิเตอร์จากวิธีต่างๆ ในข้อที่ 1 กับวิธีการประมาณแบบกราฟ (Graphical Estimation Method; GE Method)

ทฤษฎีที่เกี่ยวข้อง

ข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2

ข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2 คือ การกำหนดจำนวนค่าสังเกตที่จะเกิดความล้มเหลวไว้ล่วงหน้าด้วยค่าคงที่ r ซึ่งจะเรียกว่า Fixed Number of Uncensored Failure

ถ้า n คือจำนวนข้อมูลทั้งหมด และกำหนด r คือจำนวนค่าสังเกตที่ไม่ถูกตัดทิ้งโดยที่ $r \le n$ ให้ $T_{(1)} \le T_{(2)} \le ... \le T_{(r)}$ เป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง และ $T_{(r+1)} \le T_{(r+2)} \le ... \le T_{(n)}$ เป็นค่าสังเกตที่ถูกตัดทิ้ง ซึ่ง $T_{(i)} \ge T_{(r)}$; $i = r+1, r+2, ..., n$ ไม่ทราบค่าที่แท้จริงของค่าสังเกต ดังนั้น X_i เป็นตัวแปรสุ่มของค่าสังเกต ซึ่ง

$$X_i = \begin{cases} T_{(i)} & ; i \le r \\ T_{(r)} & ; i = r+1, r+2, ..., n \end{cases}$$

และมีฟังก์ชันภาวะน่าจะเป็นดังนี้

$$L = \prod_{i=1}^{r} f\left(x_i\right) \cdot S\left(T_{(r)}\right)^{n-r}$$

ตารางที่ 1 ขอบเขตการวิจัย

ขอบเขตการวิจัย	สัญลักษณ์	กำหนดขอบเขต
สัดส่วนของการตัดปลายทางขวา	p	10%, 20%, 30%
ขนาดตัวอย่าง	n	20, 40, 80, 120
ขนาดการถ่วงน้ำหนัก	w	2 เท่า, 4 เท่า
จำนวนการทำซ้ำ	-	5000

การแจกแจงที่ใช้ในการวิจัย

การแจกแจงที่ใช้ในการวิจัยครั้งนี้ได้แก่ การแจกแจงค่าต่ำสุดขีด (Smallest Extreme Value distribution) โดยให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงค่าต่ำสุดขีด ด้วยพารามิเตอร์ μ และ σ เขียนแทนด้วย $X \sim SEV(\mu, \sigma)$ โดยที่ μ เป็นพารามิเตอร์แสดงตำแหน่ง (Location parameter) และ σ เป็นพารามิเตอร์แสดงขนาด (Scale parameter) จะได้ว่าการแจกแจงค่าต่ำสุดขีดเป็นการแจกแจงที่มีฟังก์ชันความหนาแน่นน่าจะเป็น (Probability Density Function; PDF) อยู่ในรูปของ

$$f\left(x; \mu, \sigma\right) = \frac{1}{\sigma} \exp\left[\frac{x-\mu}{\sigma} - \exp\left(\frac{x-\mu}{\sigma}\right)\right] \quad , -\infty < x < \infty$$

เมื่อ $\sigma > 0$ และ $-\infty < \mu < \infty$ และมีฟังก์ชันการแจงแจงสะสม (Cumulative Distribution Function; CDF) อยู่ในรูปของ

$$P(X \le x) = F(x;\mu,\sigma) = \Phi_{SEV}\left(\frac{x-\mu}{\sigma}\right)$$

เมื่อ $\Phi_{SEV}(Z) = 1-\exp\left[-\exp(Z)\right]$ ซึ่งมีค่าเฉลี่ยและค่าความแปรปรวน คือ

$$E(x) = \mu - \sigma\gamma \text{ และ } Var(x) = \frac{\sigma^2\pi^2}{6} \text{ ตามลำดับ}$$

และ $\gamma \approx 0.5772$ คือค่าคงที่ของออยเลอร์ (Euler's constant)

การประมาณค่าพารามิเตอร์

1. วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method; MLE Method)

ให้ $X_1,...,X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีฟังก์ชันความหนาแน่นน่าจะเป็น $f(x;\theta)$

ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ของตัวอย่างสุ่ม ได้แก่ ฟังก์ชันความหนาแน่นน่าจะเป็นร่วมของ $X_1,...,X_n$ โดยถือว่าเป็นฟังก์ชันของพารามิเตอร์ θ ซึ่งเรามักแทนฟังก์ชันภาวะน่าจะเป็นด้วย L หรือ $L(\theta;x_1,...,x_n)$ หรือ $L(\theta)$ นั่นคือ $L = f(x_1;\theta),...,f(x_n;\theta)$ โดยถือว่า L เป็นฟังก์ชันของ θ

การประมาณค่าพารามิเตอร์ด้วยวิธีนี้เป็นการหาค่าของพารามิเตอร์ θ ที่ทำให้ $L(\theta;x_1,...,x_n)$ มีค่ามากที่สุด (ธีระพร 2536)

$$\hat{\theta} = \arg\max_{\theta} L(\theta;x_1,...,x_n)$$

2. วิธีการประมาณแบบกราฟ (Graphical Estimation Method; GE Method)

ให้ X เป็นตัวแปรสุ่มจากฟังก์ชันการแจกแจง F โดยที่

$$P(X \le x;\mu,\sigma) = F(x;\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

โดยที่ $\Phi(x)$ เป็นฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่ม X เมื่อ $\mu = 0$ และ $\sigma = 1$

และ F มาจากการแจกแจงแบบ location-scale family ก็ต่อเมื่อ

$$F(x;\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \text{เมื่อ } \Phi(x) = F(x;0,1)$$

โดยจะเรียก μ ว่า Location parameter เมื่อ $-\infty < \mu < \infty$ และ σ ว่า Scale parameter เมื่อ $\sigma > 0$

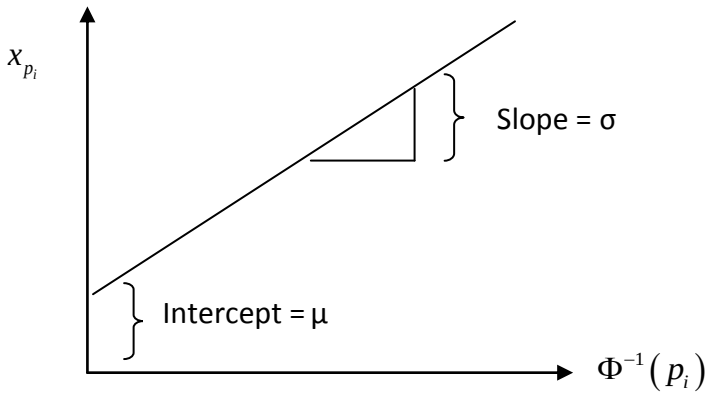
เนื่องจาก $F(x;\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ เมื่อ $\Phi(x) = F(x;0,1)$

จะได้ว่า $x_{p_i} = \mu + \sigma\Phi^{-1}(p_i)$

โดยที่ $\Phi^{-1}(p_i)$ คือ ฟังก์ชันควอนไทล์ (Quantile function) ของ F เมื่อ i เป็นลำดับที่ของข้อมูลเมื่อเรียงจากน้อยไปมาก และ x_{p_i} คือ ควอนไทล์ตัวที่ p_i ของ X

จะเห็นว่า ควอนไทล์ของ x_{p_i} เป็นฟังก์ชันเชิงเส้นของควอนไทล์จาก $\Phi^{-1}(p_i)$ โดยที่ ถ้าเราพล็อตกราฟระหว่าง x_{p_i} กับ $\Phi^{-1}(p_i)$ โดยพล็อต x_{p_i} บนแกน y และพล็อต $\Phi^{-1}(p_i)$ บนแกน x จะได้ μ และ σ เป็นจุดตัดบนแกน y (intercept) และค่าความชัน (slope) ของเส้นตรงตามลำดับ

ภาพที่ 1 แสดงการประมาณด้วยวิธีแบบกราฟ



จากความสัมพันธ์เชิงเส้นตรงของ x_{p_i} และ $\Phi^{-1}(p_i)$ เราสามารถนำแนวคิดนี้มาใช้ในการประมาณโดยใช้กราฟ ด้วยการพล็อตกราฟระหว่าง $x_{(i)}$ กับ $\Phi^{-1}\left(p_{(i)}\right)$ โดยให้ $x_{(i)}$ เป็น sample quantile และ $p_{(i)}$ เป็นลำดับควอนไทล์ของ $x_{(i)}$ ดังนั้นการประมาณค่า จุดตัดบนแกน y (intercept) ประมาณค่า μ และค่าความชัน (slope) ของเส้นตรงประมาณค่า σ ได้

3. วิธีการประมาณแบบกราฟที่มีการถ่วงน้ำหนัก (Weighted Graphical Estimation Method; WGE Method)

วิธีนี้ใช้หลักการเดียวกันกับการประมาณแบบกราฟ กล่าวคือ เป็นการพล็อตกราฟระหว่าง $x_{(i)}$ กับ $\Phi^{-1}\left(p_{(i)}\right)$ ที่มีการถ่วงน้ำหนักในแต่ละอันดับ ดังนั้นการประมาณค่า จุดตัดบนแกน y (intercept) ประมาณค่า μ และค่าความชัน (slope) ของเส้นตรงประมาณค่า σ ได้เช่นเดียวกัน

เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพของการประมาณ

1. การประมาณค่าแบบจุด จะพิจารณาเปรียบเทียบจากค่าคลาดเคลื่อนกำลังสอง (Mean Square Error; MSE) ของตัวประมาณพารามิเตอร์ μ และ σ

กำหนดให้ μ เป็นพารามิเตอร์แสดงตำแหน่ง (Location parameter)

σ เป็นพารามิเตอร์แสดงขนาด (Scale parameter)

$\hat{\mu}_i$ เป็นค่าประมาณของพารามิเตอร์ μ จากตัวอย่างที่ i

$\hat{\sigma}_i$ เป็นค่าประมาณของพารามิเตอร์ σ จากตัวอย่างที่ i

$$MSE(\mu) = \frac{\sum_{i=1}^{N}(\hat{\mu}_i - \mu)^2}{N} \text{ และ } MSE(\sigma) = \frac{\sum_{i=1}^{N}(\hat{\sigma}_i - \sigma)^2}{N}$$

เมื่อ N = จำนวนครั้งของการทดลองในแต่ละสถานการณ์ของการทดลอง

โดยถ้าวิธีการประมาณค่าแบบใด ที่ทำให้ได้ค่า MSE ต่ำกว่า จะถือว่าการประมาณค่าจากวิธีนั้น เป็นวิธีที่มีประสิทธิภาพมากกว่า

2. ประสิทธิภาพสัมพัทธ์ (Relative Efficiency; RE) ของตัวประมาณพารามิเตอร์ μ และ σ จะพิจารณาเปรียบเทียบคือ ค่าประสิทธิภาพสัมพัทธ์ที่คิดจากค่าสัดส่วนของ MSEs ของตัวประมาณพารามิเตอร์ โดยจะเปรียบเทียบกับการประมาณด้วยวิธีการประมาณแบบกราฟ (GE Method)

กำหนดให้ RE เป็นค่าประสิทธิภาพสัมพัทธ์ที่คิดจากค่าสัดส่วนของ MSE

MSE_{GE} เป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณด้วยวิธี GE

$MSE_{METHOD1}$ เป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณด้วยวิธี METHOD1

$$RE(METHOD1) = \frac{MSE_{GE}}{MSE_{METHOD1}}$$

เกณฑ์ที่ใช้วัด คือ

ถ้า $RE(METHOD1)$ มีค่ามากกว่า 1 หมายความว่า ตัวประมาณ METHOD1 มีประสิทธิภาพมากกว่า GE

ถ้า $RE(METHOD1)$ มีค่าน้อยกว่า 1 หมายความว่า ตัวประมาณ METHOD1 มีประสิทธิภาพน้อยกว่า GE

ถ้า $RE(METHOD1)$ มีค่าเท่ากับ 1 หมายความว่า ตัวประมาณ METHOD1 และ GE มีประสิทธิภาพเท่ากัน

วิธีการวิจัย

1. ศึกษาวิธีการประมาณค่าพารามิเตอร์ของการแจกแจงแบบ location-scale และเปรียบเทียบประสิทธิภาพของการประมาณค่า

1.1 ศึกษาค้นคว้าเอกสารต่าง ๆ ทั้งทฤษฎี และงานวิจัยที่เกี่ยวกับการประมาณค่าพารามิเตอร์จากข้อมูลที่ถูกตัดปลายทางขวาประเภทที่ 2

1.2 สร้างข้อมูลที่มีการแจกแจงค่าต่ำสุดขีด ขนาดตัวอย่าง (n) และสัดส่วนการตัดปลาย (p) ตามขอบเขตการวิจัย ดังตารางที่ 1

1.3 สร้างขนาดการถ่วงน้ำหนัก (w) ตามขอบเขตการวิจัย ดังตารางที่ 1 โดยกำหนดให้ค่าน้ำหนักของค่าสังเกตที่น้อยสุด มีค่าเท่ากับ 1

2. คำนวณค่าพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุด (MLE), วิธีการประมาณแบบกราฟ (GE) และวิธีการประมาณแบบกราฟที่มีการถ่วงน้ำหนัก (WGE) ซึ่งประกอบไปด้วย 5 วิธี คือ วิธี Increasing Weighting, วิธี Decreasing Weighting, วิธี Symmetric-Triangular Weighting, วิธี Right-Skewed-Triangular Weighting และวิธี Left-Skewed-Triangular Weighting

3. ทำซ้ำข้อ 1-2 จำนวน 5,000 รอบในแต่ละสถานการณ์

4. คำนวณค่าความเอนเอียง ค่าความแปรปรวน และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

5. คำนวณค่าประสิทธิภาพสัมพันธ์ (RE)

6. สรุปผลการศึกษา

ผลการวิจัย

ส่วนที่ 1: การประมาณค่าพารามิเตอร์ μ

เมื่อ $p = 0$; พบว่า วิธี WGE แบบ Increasing Weighting จะมีประสิทธิภาพมากที่สุด โดยแบบ Increasing Weighting ที่มี $w = 4$ มีประสิทธิภาพมากกว่าแบบ Increasing Weighting ที่มี $w = 2$

เมื่อ $p \neq 0$; พบว่า วิธี WGE แบบ Increasing Weighting จะมีประสิทธิภาพมากที่สุด โดยแบบ Increasing Weighting ที่มี $w = 4$ มีประสิทธิภาพมากกว่าแบบ Increasing Weighting ที่มี $w = 2$

ส่วนที่ 2: การประมาณค่าพารามิเตอร์ σ

เมื่อ $p = 0$; พบว่า วิธี WGE แบบ Increasing Weighting จะมีประสิทธิภาพมากที่สุด โดยแบบ Increasing Weighting ที่มี $w = 4$ มีประสิทธิภาพมากกว่าแบบ Increasing Weighting ที่มี $w = 2$

เมื่อ $p \neq 0$; พบว่า วิธี WGE แบบ Increasing Weighting จะมีประสิทธิภาพมากที่สุด โดยแบบ Increasing Weighting ที่มี $w = 4$ มีประสิทธิภาพมากกว่าแบบ Increasing Weighting ที่มี $w = 2$

การอภิปรายผล

จากผลการศึกษาพบว่า เมื่อข้อมูลมีการแจกแจงค่าต่ำสุดขีดที่มีพารามิเตอร์ $\mu = 0$ และ $\sigma = 1$ การประมาณค่าพารามิเตอร์ μ และ σ ด้วยวิธี WGE แบบ Increasing Weighting มีประสิทธิภาพมากกว่าวิธี GE ในทุกกรณี ทั้งนี้อาจเป็นผลมาจากการแจกแจงค่าต่ำสุดขีด มีลักษณะเบ้ซ้าย ข้อมูลส่วนที่ให้รายละเอียดได้ดีมักจะอยู่ทางขวา วิธี Increasing Weighting นั้นเป็นการปรับค่าการถ่วงน้ำหนักให้ค่าสังเกตแต่ละอันดับเพิ่มขึ้นครั้งละเท่าๆ กัน โดยกำหนดให้ค่าน้ำหนักของค่าสังเกตที่น้อยสุด มีค่าเท่ากับ 1

นอกจากนี้ยังพบว่าวิธีการประมาณแบบ Increasing Weighting ที่มี $w = 4$ มีประสิทธิภาพมากกว่าแบบ Increasing Weighting ที่มี $w = 2$ ในทุกกรณีเช่นกัน

ข้อเสนอแนะ

จากงานวิจัยนี้พบว่า วิธีการประมาณพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดยังคงเป็นวิธีการประมาณที่มีประสิทธิภาพมากที่สุด เมื่อเทียบกับการประมาณแบบกราฟและการประมาณแบบกราฟที่มีการถ่วงน้ำหนัก และถ้าเปลี่ยนการแจกแจงในการศึกษาเป็นแบบอื่น ซึ่งเป็นการแจกแจงแบบ Location-Scale แล้ว โดยรวมการประมาณค่าแบบกราฟที่มีการถ่วงน้ำหนัก ยังจะให้ผลดีกว่าแบบกราฟหรือไม่

เอกสารอ้างอิง

ธีระพร วีระถาวร. การอนุมานเชิงสถิติขั้นกลาง: โครงสร้างและความหมาย. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2539.

ขวัญรัตน์ ตั้งพิษฐานสกุล. การศึกษาเปรียบเทียบการประมาณพารามิเตอร์โดยตัวประมาณ แบบกราฟด้วยข้อมูลบางส่วนจากข้อมูลที่ถูกตัดปลายทางขวา. วิทยานิพนธ์ปริญญามหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย. 2554.

ประภาศิริ สุนทรศิริเวช. การประมาณแบบกราฟด้วยข้อมูลบางส่วนจากข้อมูลที่ถูกตัดปลายกำหนดเวลา. วิทยานิพนธ์ปริญญามหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย. 2555.

Meeker, W.Q., Escobar, L.A. Statistical Methods for Reliability Data. New York: John Wiley, 1998.

ตารางที่ 2 แสดงค่าประสิทธิภาพสัมพันธ์ (RE) ของการประมาณค่าพารามิเตอร์ μ สำหรับการแจกแจงค่าต่ำสุดขีด ที่มีพารามิเตอร์ $\mu = 0$ และ $\sigma = 1$ โดยจำแนกตามขนาดตัวอย่าง และสัดส่วนของการตัดปลายทางขวา

Censoring Proportion p	Sample Size n	GE	MLE	WGE									
				Increasing Weighting		Decreasing Weighting		Symmetric-Triangular Weighting		Right-Skewed-Triangular Weighting		Left-Skewed-Triangular Weighting	
				$w=2$	$w=4$	$w=2$	$w=4$	$w=2$	$w=4$	$w=2$	$w=4$	$w=2$	$w=4$
0	20	1	1.120	1.017	1.020	0.944	0.848	0.937	0.970	0.918	0.959	0.994	0.982
	40	1	1.135	1.022	1.025	0.934	0.826	0.937	0.969	0.914	0.956	0.996	0.984
	80	1	1.140	1.026	1.030	0.924	0.806	0.936	0.968	0.911	0.954	0.997	0.987
	120	1	1.163	1.034	1.041	0.910	0.781	0.935	0.967	0.907	0.951	1.000	0.990
0.1	20	1	1.147	1.058	1.080	0.895	0.767	0.914	0.951	0.890	0.939	0.987	0.975
	40	1	1.173	1.068	1.095	0.880	0.743	0.923	0.955	0.890	0.937	1.001	0.998
	80	1	1.203	1.079	1.111	0.867	0.721	0.927	0.957	0.891	0.937	1.005	1.005
	120	1	1.242	1.092	1.131	0.853	0.699	0.933	0.960	0.894	0.937	1.012	1.016
0.2	20	1	1.274	1.098	1.153	0.860	0.717	0.913	0.945	0.885	0.933	1.000	1.002
	40	1	1.324	1.114	1.177	0.845	0.694	0.931	0.955	0.891	0.935	1.015	1.028
	80	1	1.372	1.127	1.200	0.832	0.677	0.945	0.962	0.900	0.939	1.025	1.045
	120	1	1.428	1.142	1.227	0.819	0.658	0.957	0.969	0.910	0.942	1.034	1.062
0.3	20	1	1.462	1.129	1.216	0.842	0.694	0.927	0.950	0.894	0.941	1.001	1.011
	40	1	1.549	1.151	1.255	0.824	0.670	0.952	0.963	0.907	0.943	1.033	1.067
	80	1	1.628	1.166	1.283	0.813	0.656	0.974	0.975	0.924	0.950	1.047	1.092
	120	1	1.713	1.182	1.314	0.802	0.640	0.994	0.985	0.940	0.958	1.059	1.115

ตารางที่ 3 แสดงค่าประสิทธิภาพสัมพันธ์ (RE) ของการประมาณค่าพารามิเตอร์ σ สำหรับการแจกแจงค่าต่ำสุดขีด ที่มีพารามิเตอร์ $\mu = 0$ และ $\sigma = 1$ โดยจำแนกตามขนาดตัวอย่าง และสัดส่วนของการตัดปลายทางขวา

Censoring Proportion p	Sample Size n	GE	MLE	WGE									
				Increasing Weighting		Decreasing Weighting		Symmetric-Triangular Weighting		Right-Skewed-Triangular Weighting		Left-Skewed-Triangular Weighting	
				$w=2$	$w=4$	$w=2$	$w=4$	$w=2$	$w=4$	$w=2$	$w=4$	$w=2$	$w=4$
0	20	1	1.741	1.162	1.329	0.858	0.743	1.042	1.090	1.050	1.063	1.058	1.133
	40	1	1.807	1.177	1.365	0.850	0.732	1.053	1.115	1.058	1.084	1.072	1.165
	80	1	1.897	1.186	1.387	0.845	0.726	1.057	1.124	1.063	1.097	1.075	1.173
	120	1	1.923	1.193	1.402	0.840	0.719	1.059	1.128	1.064	1.099	1.078	1.179
0.1	20	1	1.740	1.144	1.299	0.875	0.774	1.045	1.100	1.053	1.066	1.059	1.137
	40	1	1.801	1.157	1.330	0.868	0.764	1.056	1.127	1.063	1.098	1.077	1.178
	80	1	1.817	1.160	1.337	0.866	0.762	1.063	1.143	1.071	1.120	1.083	1.192
	120	1	1.844	1.165	1.347	0.862	0.756	1.065	1.147	1.073	1.124	1.086	1.196
0.2	20	1	1.724	1.133	1.279	0.884	0.790	1.044	1.098	1.053	1.074	1.061	1.142
	40	1	1.788	1.146	1.308	0.877	0.780	1.055	1.125	1.062	1.098	1.076	1.175
	80	1	1.790	1.150	1.319	0.874	0.777	1.064	1.146	1.072	1.125	1.084	1.193
	120	1	1.799	1.154	1.325	0.872	0.773	1.067	1.152	1.076	1.133	1.087	1.198
0.3	20	1	1.742	1.123	1.256	0.892	0.803	1.042	1.095	1.051	1.058	1.054	1.126
	40	1	1.794	1.136	1.288	0.884	0.793	1.053	1.121	1.060	1.095	1.073	1.169
	80	1	1.819	1.142	1.303	0.881	0.789	1.063	1.145	1.071	1.126	1.083	1.191
	120	1	1.816	1.146	1.309	0.878	0.784	1.068	1.156	1.078	1.139	1.087	1.200