

การเปรียบเทียบตัวประมาณในปัญหาของไนล์ A Comparison of Estimators in the Problem of the Nile

ธรรมรัตน์ กลีบเมฆ¹ วุฒิชัย ศรีโสดาพล²
Tammarat Kleebmek¹, Wuttichai Srisodaphol²

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบตัวประมาณ 2 ตัวในการประมาณค่าพารามิเตอร์ปัญหาของไนล์ นั่นคือ ตัวประมาณความควรจะเป็นสูงสุด กับตัวประมาณที่ผู้วิจัยนำเสนอ โดยใช้เกณฑ์ค่าคลาดเคลื่อนกำลังสองเฉลี่ย เปรียบเทียบตัวประมาณทั้งสอง ผลการวิจัยพบว่า ตัวประมาณที่ผู้วิจัยนำเสนอ มีค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าตัวประมาณความควรจะเป็นสูงสุดเมื่อขนาดตัวอย่างเล็ก และตัวประมาณทั้งสองตัวจะมีค่าคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกันเมื่อขนาดตัวอย่างเพิ่มขึ้น

คำสำคัญ : ตัวประมาณความควรจะเป็นสูงสุด ปัญหาของไนล์ แจ็คไนฟ์

Abstracts

The objective of this research is to compare two estimators for estimating parameter in the problem of the Nile, that is, maximum likelihood estimator (MLE), and the proposed estimator. The mean squared error (MSE) is the criterion to compare these two estimators. The results show that the proposed estimator has higher MSE than maximum likelihood estimator when small sample size. Moreover, these two estimators have a similar MSE when sample size is increasing

Keyword : Maximum likelihood estimator (MLE), Problem of the Nile, Jackknife

¹ นักศึกษาปริญญาโท สาขาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 40002

² อ.ดร., ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 40002

* Corresponding author: E-mail : terry_arist0007@hotmail.com

บทนำ

สมมติให้ X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระกัน โดยที่ X มีการแจกแจงแบบเอ็กซ์โพเนนเชียลซึ่งมีพารามิเตอร์ θ และ Y มีการแจกแจงแบบเอ็กซ์โพเนนเชียลซึ่งมีพารามิเตอร์เป็น $1/\theta$ โดย X และ Y มีฟังก์ชันหนาแน่นความน่าจะเป็นร่วมดังนี้

$$f(x, y; \theta) = \exp\{-(\theta x + \theta^{-1}y)\}; x > 0, y > 0, \theta > 0$$

$$\text{ให้ } \bar{x} = \sum_{i=1}^n x_i / n, \bar{y} = \sum_{i=1}^n y_i / n \text{ และ } T = \sqrt{\bar{y} / \bar{x}}, U = \sqrt{\bar{x} \bar{y}}$$

เมื่อ T เป็นตัวประมาณแบบวิธีความควรจะเป็นสูงสุด (Maximun Likelihood Estimator) ของพารามิเตอร์ θ และ U เป็นตัวสถิติช่วย (Ancillary statistic) ซึ่ง T, U เป็นตัวสถิติร่วมที่มีความเพียงพอแต่ไม่มีความสมบูรณ์ (Jointly not complete sufficient statistic) ซึ่งฟังก์ชันหนาแน่นความน่าจะเป็นในรูปแบบนี้เรียกว่า ปัญหาของไนล์ (Problem of the Nile) ถูกสร้างขึ้นโดย Fisher [1] จากการศึกษาพบว่าเนื่องจาก เป็นตัวประมาณที่เอนเอียง มีผู้วิจัยหลายท่านได้ศึกษาวิธีลดความเอนเอียงของตัวประมาณดังนี้

Quenouille [2] ได้กล่าวถึงวิธีการลดความเอนเอียงของตัวประมาณของพารามิเตอร์ ด้วยวิธีของแจ็กไนล์ ซึ่งจะสามารถหาตัวประมาณตัวใหม่ โดยกำหนดให้ $X_1, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีฟังก์ชันหนาแน่นความน่าจะเป็น $f(x; \theta)$ ซึ่งมี θ เป็นตัวประมาณของพารามิเตอร์ θ และให้ θ เป็นตัวประมาณที่ถูกตัดค่าของตัวแปรสุ่ม X ที่ i ออก โดยที่ตัวประมาณแจ็กไนล์ (θ_J) คือ

$$\theta_J = \sum_{i=1}^n \theta_{-i} / n$$

Schucany, Gray และ Owen [3] ได้เสนอรูปแบบทั่วไปในการประมาณค่าพารามิเตอร์ด้วยวิธีของแจ็กไนล์ โดยสมมติว่ามีตัวประมาณ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณที่มีความเอนเอียงของพารามิเตอร์ θ โดยสามารถหาตัวประมาณ $\tilde{\theta}$ ซึ่งเป็นตัวประมาณที่ไม่มีความเอนเอียงของพารามิเตอร์ θ ได้ดังนี้

$$\theta = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}}$$

ซึ่ง $f_1(n)$ และ $f_2(n)$ หาได้จาก

$$E(\hat{\theta}_1) = \theta + f_1(n)b(\theta), E(\hat{\theta}_2) = \theta + f_2(n)b(\theta)$$

เพื่อให้ได้ตัวประมาณที่ดีขึ้นกว่าเดิม ผู้วิจัยจึงสนใจที่จะนำเสนอตัวประมาณจากวิธีความควรจะเป็นสูงสุด และวิธีแจ็กไนล์ มาหาตัวประมาณใหม่ด้วยวิธีของ Schucany, Gray และ Owen แล้วทำการเปรียบเทียบตัวประมาณจากวิธีความควรจะเป็นสูงสุด โดยใช้เกณฑ์ค่าคลาดเคลื่อนกำลังสองเฉลี่ย ในการวัดความแม่นยำ

จากงานประชุมวิชาการระดับชาติครั้งที่ 22 ปี 2555

อุปกรณ์และวิธีการ

ในการวิจัยครั้งนี้ได้ทำเปรียบเทียบตัวประมาณของพารามิเตอร์ในปัญหาของไนล์โดยหาตัวประมาณจากการประมาณ 2 วิธีคือวิธีความควรจะเป็นสูงสุดและวิธีที่ผู้วิจัยทำเสนอซึ่งเป็นการหาจากตัวประมาณจากวิธีความควรจะเป็นสูงสุด และวิธีแจ๊คไนฟ์ นำมาหาตัวประมาณใหม่โดยวิธีของ Schucany, Gray

ขั้นที่ 1 นำตัวประมาณจากวิธีความควรจะเป็นสูงสุดมาลดความเอนเอียงด้วยวิธีการหาตัวประมาณของแจ๊คไนฟ์โดยที่ $T = \sqrt{y/\bar{x}}$ เป็นตัวประมาณจากวิธีความควรจะเป็นของพารามิเตอร์ θ ซึ่งตัวประมาณแจ๊คไนฟ์ (θ_J) คือ

$$\theta_J = \sum_{i=1}^n T_{-i} / n$$

ขั้นที่ 2 นำตัวประมาณทั้งวิธีความควรจะเป็นสูงสุดและวิธีของแจ๊คไนฟ์ มาหาตัวประมาณตัวใหม่ โดยใช้วิธีการของ Schucany, Gray และ Owen มีวิธีการดังนี้ และ Owen สำหรับตัวประมาณที่นำเสนอมีขั้นตอน ดังนี้

ให้ $\hat{\theta}_1 = T, \hat{\theta}_2 = \hat{\theta}_J$ และ $E(T) = \theta + f_1(n)b(\theta), E(\hat{\theta}_J) = \theta + f_2(n)b(\theta)$ โดยที่

$$\frac{1}{n}, f_2(n) = \frac{1}{n-1} \text{ จากนั้นแทนค่าลงใน } \hat{\theta}_{pr} = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \\ 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}}$$

ตัวประมาณตัวใหม่ที่ไม่เอนเอียงคือ

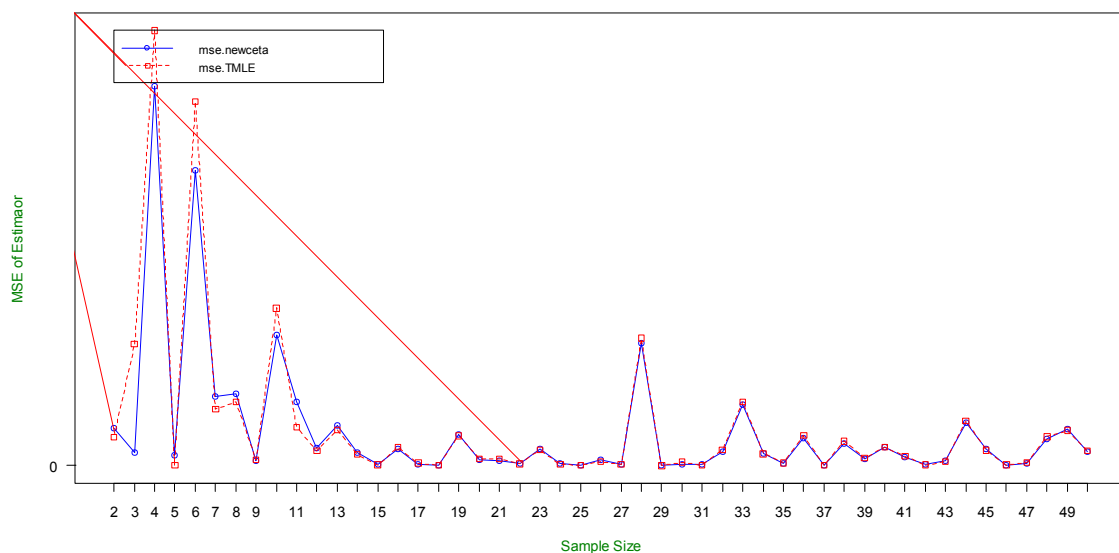
$$\hat{\theta}_{pr} = nT - \frac{(n-1)}{n} \hat{\theta}_J$$

จากนั้นทำการจำลองข้อมูล 2 ชุดที่เป็นอิสระต่อกันด้วยโปรแกรม R นั่นคือ $X=X_1, \dots, X_n$ และ $Y=Y_1, \dots, Y_n$ ซึ่งมาจากฟังก์ชันการแจกแจงแบบ $EXP(\theta)$ และ $EXP(1/\theta)$ ตามลำดับ โดยกำหนดขนาดตัวอย่าง n คือ $n=2,3,\dots,50$ และกำหนดค่าพารามิเตอร์คือ 0.1 0.3 0.5 0.7 และ 0.9 ทำการประมาณค่าพารามิเตอร์จากข้อมูลที่ได้ θ ซึ่งทำซ้ำ 1000 ครั้งเพื่อคำนวณ MSE ของตัวประมาณ T แต่ละวิธี พร้อมทั้งเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ T และ $\hat{\theta}_{pr}$

ผลการวิจัยและอภิปรายผล

การจำลองข้อมูลเมื่อ $n=2,3,\dots,50$ และ $\theta=0.1,0.3,0.5,0.7$ และ 0.9 พบว่าเมื่อตัวอย่างมีขนาดเล็ก ตัวประมาณโดยวิธีที่ผู้วิจัยนำเสนอ นั้นจะมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าวิธีความควรจะเป็นสูงสุด และเมื่อขนาดตัวอย่างเพิ่มขึ้น พบว่าตัวประมาณทั้งสองมีค่าคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกันดังรูปที่ 1-5

$$\theta = 0.1$$



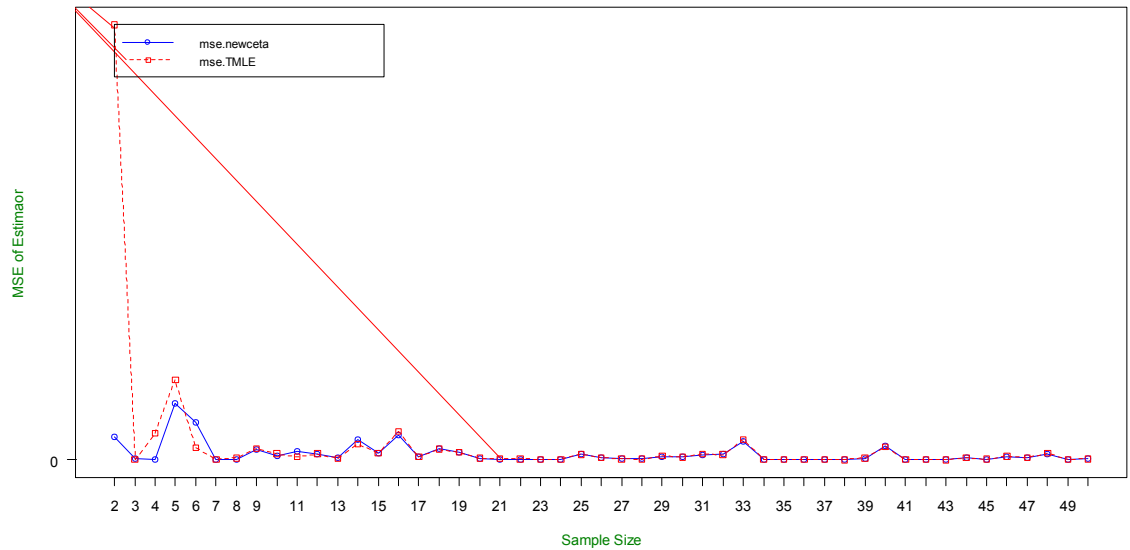
กำหนดให้ $mse.newceta$ คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}_{pr}$

$mse.TMLE$ คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ T

รูปที่ 1 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการประมาณค่าพารามิเตอร์เมื่อ $n=2,3,\dots,50$ และ $\theta=0.1$

จากงานประชุมวิชาการระดับชาติครั้งที่ 22 ปี 2555

$$\theta = 0.3$$

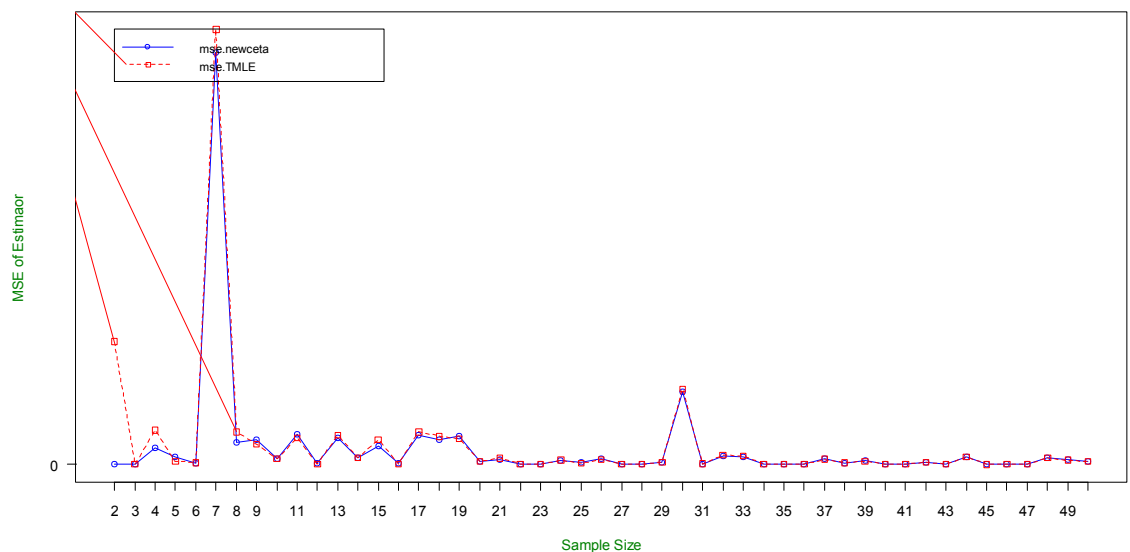


กำหนดให้ mse.newceta คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}_{pr}$

mse.TMLE คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ T

รูปที่ 2 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการประมาณค่าพารามิเตอร์เมื่อ $n = 2, 3, \dots, 50$ และ $\theta = 0.3$

$$\theta = 0.5$$

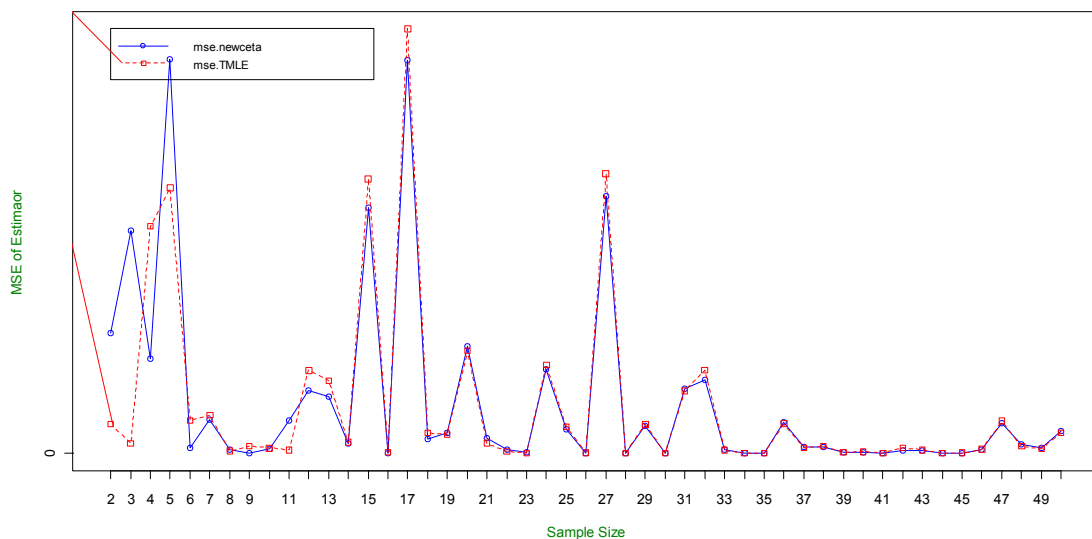


กำหนดให้ mse.newceta คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}_{pr}$

mse.TMLE คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ T

รูปที่ 3 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการประมาณค่าพารามิเตอร์เมื่อ $n = 2, 3, \dots, 50$ และ $\theta = 0.5$

$\theta = 0.7$

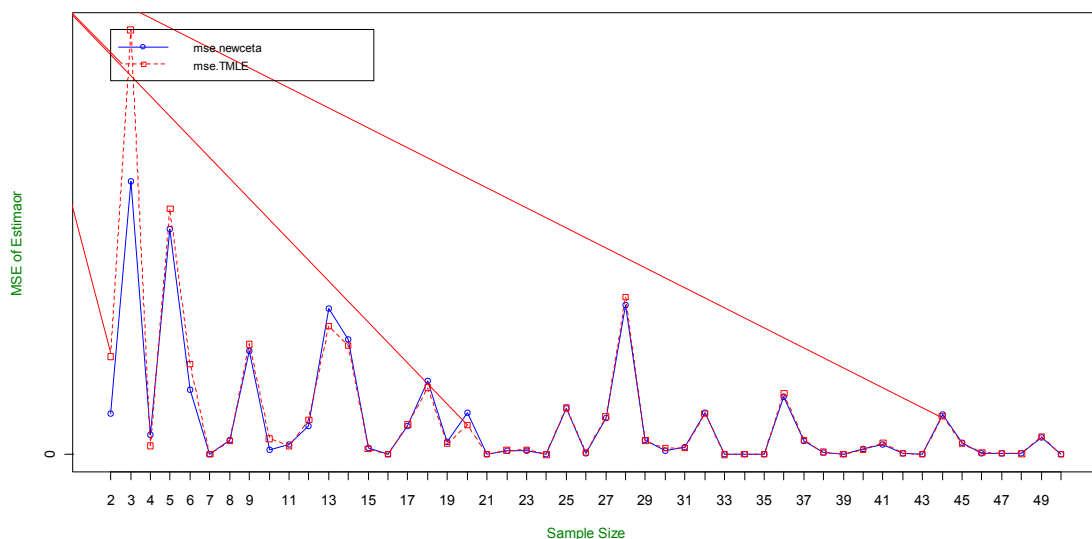


กำหนดให้ mse.newceta คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}_{pr}$

mse.TMLE คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ T

รูปที่ 4 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการประมาณค่าพารามิเตอร์เมื่อ $n = 2, 3, \dots, 50$ และ $\theta = 0.7$

$\theta = 0.9$



กำหนดให้ mse.newceta คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}_{pr}$

mse.TMLE คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ T

รูปที่ 5 ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการประมาณค่าพารามิเตอร์เมื่อ $n = 2, 3, \dots, 50$ และ $\theta = 0.9$

จากงานประชุมวิชาการระดับชาติครั้งที่ 22 ปี 2555

สรุปผลการวิจัย

ตัวประมาณที่ผู้วิจัยนำเสนอมีค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าตัวประมาณจากวิธีความควรจะเป็นสูงสุด เมื่อตัวอย่างมีขนาดเล็ก แต่เมื่อตัวอย่างมีขนาดเพิ่มขึ้นตัวประมาณทั้งสองตัวจะมีค่าคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกัน

เอกสารอ้างอิง

- [1] Fisher, R. A. (1959). **Statistical Methods and Scientific Inference**. London : Oliver and Boyd. Joshi, S.M. and Nabar, S.P. (1987). Estimation of the parameter in the problem of the Nile. Communications in Statistics-Theory and Method, 16, 3149-3156.
- [2] Quenouille, M.H. (1956). **Notes on bias in estimation**. Biometrika 43, 353-60.
- [3] Schucany, W. R., Gray, H.L. & Owen, D.B. (1971). **On bias reduction in estimation**. J. Am. Statist. Assoc. 64, 999-1013.
- [4] ประชุม สุวัฑฒ์. (2553). **ทฤษฎีการอนุมานเชิงสถิติ**. (พิมพ์ครั้งที่3). กรุงเทพฯ : โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันพัฒนาบริหารศาสตร์.
- [5] Lehmann, E. L. (1983). **Theory of Point Estimation**. New York: John Wiley and Sons.
- [6] Lehmann, E. L. and Scheffe, H. (1950). **Completeness, similar regions, and unbiased estimation**. I. Sankhyā, 10, 305-340.
- [7] Miller, R. G. (1968). **Jackknifing variances**. Ann. Math. Statist. 39, 567-582.