ตัวประมาณแจ็คไนฟ์

(The Jackknife Estimator)

ปรีชา วิจิตรธรรมรส*
Preecha Vichitthamaros

Abstract

The Jackknife estimator is a robust statistics. This technique's purpose is to decrease the bias of an estimator and provide an approximate confidence interval for the parameter of interest. It was introduced by Quenouille in 1949 and named by Tukey in 1958. The name 'Jackknife' was coined due to the fact that, like a Scout's trusty jackknife, it is rough and ready in all situations. The usefulness of the jackknife is truly remarkable: Once an estimate has been calculated from the sample-any estimate whatever-the jackknife will provide a confidence interval around it.

บทคัดย่อ

ตัวประมาณแจ็คไนฟ์เป็นตัวสถิติที่มีความแกร่งตัวหนึ่งที่พัฒนาขึ้นเพื่อใช้เป็นเทคนิคในการที่จะลดความเอนเอียงในการประมาณค่าแบบจุดและการประมาณช่วงความเชื่อมั่นสำหรับพารามิเตอร์ ตัวประมาณแจ็คไนฟ์พัฒนาขึ้นโดย Quenouille ในปี ค.ศ. 1949 ต่อมาในปี ค.ศ. 1958 Tukey เป็นผู้ตั้งชื่อให้แก่ตัวประมาณดังกล่าว เนื่องจากตัวประมาณดังกล่าวเปรียบเสมือนมีดพับของลูกเสือซึ่งมีความสมบุกสมบันและพร้อมเสมอที่จะนำมาใช้ภายใต้สถานการณ์ต่าง ๆ ประโยชน์ของตัวประมาณแจ็คไนฟ์ที่น่าสนใจคือ ในการประมาณค่าพารามิเตอร์โดยให้ข้อมูลจากตัวอย่าง ไม่ว่าจะเป็นการประมาณค่าพารามิเตอร์ตัวใดก็ตาม วิธีแจ็คไนฟ์สามารถให้ช่วงความเชื่อมั่นที่ครอบคลุมค่าพารามิเตอร์ดังกล่าว

* ผู้ช่วยศาสตราจารย์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

1. บทนำ

ตัวประมาณแจ็คไนฟ์ (Jackknife Estimator) เป็นตัวสถิติที่มีความแกร่งตัวหนึ่งที่พัฒนาขึ้นเพื่อใช้เป็นเทคนิคในการที่จะลดความเอนเอียง (bias) ในการประมาณค่าพารามิเตอร์ ตัวประมาณ แจ็คไนฟ์พัฒนาขึ้นโดย Quenouille ในปี ค.ศ. 1949 ต่อมาในปี ค.ศ. 1958 Tukey เป็นผู้ตั้งชื่อให้แก่ตัวประมาณดังกล่าว เนื่องจากตัวประมาณดังกล่าวเปรียบเสมือนมีดพับของลูกเสือซึ่งมีความสมบุกสมบันและพร้อมเสมอที่จะนำมาใช้ภายใต้สถานการณ์ต่าง ๆ

นอกจากนี้ตัวประมาณแจ็คไนฟ์ยังสามารถใช้ในการสร้างช่วงความเชื่อมั่นโดยประมาณสำหรับพารามิเตอร์ที่สนใจใด ๆ ซึ่งในบางครั้งการสร้างช่วงความเชื่อมั่นสำหรับพารามิเตอร์บางตัวทำได้ยุ่งยากหากใช้วิธีการประมาณช่วงแบบที่ใช้อยู่ทั่วไป เช่น ในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนและความเบี่ยงเบนมาตรฐานของประชากร โดยวิธีที่ใช้กันอยู่ทั่ว ๆ ไปต้องอาศัยข้อสมมติเบื้องต้น (assumptions) ที่ว่าประชากรจะต้องมีการแจกแจงแบบปกติหรือใกล้เคียงปกติ หากข้อตกลงดังกล่าวไม่เป็นจริงการประมาณช่วงจะไม่น่าเชื่อถือ ซึ่งในกรณีนี้เราสามารถใช้วิธีการประมาณช่วงแบบแจ็คไนฟ์ในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนและความเบี่ยงเบนมาตรฐานของประชากรได้ ซึ่งจะให้ผลที่น่าเชื่อถือกว่า (more reliable)

ในบทความนี้จะกล่าวถึงการประมาณค่าพารามิเตอร์แบบจุดและแบบช่วงโดยใช้ตัวประมาณแจ็คไนฟ์ในแง่ทฤษฎี เพื่อให้ทราบและเข้าใจในหลักการประมาณค่าด้วยแจ็คไนฟ์

2. การประมาณค่าแบบจุด (Point Estimation)

ขั้นตอนการประมาณค่าแบบจุดโดยใช้วิธีแจ็คไนฟ์มีดังนี้ ถ้าให้ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ซึ่งสุ่มมาจากประชากรที่มีพารามิเตอร์คือ θ และให้ $\hat{\theta}$ คือตัวประมาณของ θ ถ้าแบ่งตัวอย่างสุ่มออกเป็น N กลุ่ม โดยแต่ละกลุ่มมีขนาดเท่า ๆ กันเท่ากับ $m = \frac{n}{N}$ (N เป็นตัวประกอบหนึ่งของ n) จากนั้นให้ตัดกลุ่มตัวอย่างที่แบ่งออกมาทิ้งทีละกลุ่มและทำการประมาณค่า θ โดยใช้ข้อมูลที่เหลืออยู่ (N-1)m ตัว วิธีการประมาณให้ใช้วิธีเดียวกันกับการประมาณพารามิเตอร์เมื่อใช้ขนาดตัวอย่างเท่ากับ n ให้ $\hat{\theta}_i$ คือค่าประมาณของ θ ซึ่งได้จากการประมาณโดยตัดตัวอย่างกลุ่มที่ i ทิ้ง ดังนั้นค่า $\hat{\theta}_i$ จะมีทั้งหมด N ค่า (i = 1, 2,..., N) ค่า $\hat{\theta}_i$ นี้จะเรียกว่า **ตัวสถิติแจ็คไนฟ์** (Jackknife Statistic) เมื่อได้ค่า $\hat{\theta}_i$ แล้ว จากนั้นให้คำนวณค่าแฝง (Pseudovalues) ดังสูตร

$$J_i = N\hat{\theta}-(N-1)\hat{\theta}_i \; ; \; i = 1, 2, \ldots, N$$

นิยาม 1 ตัวประมาณแจ็คไนฟ์ของพารามิเตอร์ θ (Jackknife Estimator of θ) คือ

$$J(\hat{\theta}) = \frac{1}{N}\sum_{i=1}^{N} J_i = \frac{1}{N}\sum_{i=1}^{N}(N\hat{\theta}-(N-1)\hat{\theta}_i)$$

$$= N\hat{\theta}-(N-1)\bar{\hat{\theta}} \text{ โดย } \bar{\hat{\theta}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i$$

ข้อสังเกต

1) ตัวประมาณแจ็คไนฟ์สามารถเขียนใหม่ได้ในรูป

$$J(\hat{\theta}) = \hat{\theta}+(N-1)(\hat{\theta}-\bar{\hat{\theta}})$$

ซึ่งแสดงให้เห็นว่า ตัวประมาณ $J(\hat{\theta})$ คือการปรับค่า $\hat{\theta}$ ด้วยเทอมปรับแก้ซึ่งเป็นฟังก์ชันของผลต่างของ $\hat{\theta}$ และ $\bar{\hat{\theta}}$

2) กรณีพิเศษคือ เมื่อ m = 1 ซึ่งเป็นกรณีที่ใช้มากในการนำตัวประมาณแจ็คไนฟ์ไปใช้ ดังนั้นสูตรจะเป็นดังนี้

$$J(\hat{\theta}) = n\hat{\theta} - (n-1)\bar{\hat{\theta}}; N=n$$

การใช้ m = 1 จะทำให้ได้ค่าแฝง (J_i) จำนวน n ค่า ซึ่งมีจำนวนเท่ากับขนาดตัวอย่างเมื่อเริ่มต้นพอดี ทำให้จำนวนองศาอิสระ (degree of freedom) ไม่เปลี่ยนแปลง ถ้าหากใช้ m > 1 จะทำให้องศาความเป็นอิสระลดลง และหากขนาดตัวอย่าง มีค่าน้อยอยู่แล้ว การใช้ m > 1 จะทำให้ค่าแฝง (J_i) ถูกสร้างขึ้นจากขนาดตัวอย่างที่เล็ก ซึ่งอาจทำให้การประมาณไม่น่าเชื่อถือ (unreliable estimate) ดังนั้นสิ่งที่เราต้องการคือ จำนวนค่าแฝง (J_i) จำนวนมากและแต่ละค่าสร้างจากค่าสังเกต (X_i) จำนวนมากเช่นกัน

ตัวอย่างที่ 1 กรณีตัวแปรสุ่มไม่ต่อเนื่อง

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n ซึ่งสุ่มมาจากประชากรที่มีการแจกแจง แบบ Bernoulli คือ $P(X_i = 0) = 1-P$ และ $P(X_i = 1) = P$ ให้ $X = X_1+X_2+...+X_n$ จะได้ว่า $\hat{P} = \frac{X}{n}$ เป็นตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator : MLE) สำหรับ P

เนื่องจาก $\theta = f(P) = PQ = P(1-P)$ เป็นฟังก์ชัน 1-1 ดังนั้นตัวประมาณภาวะน่าจะเป็นสูงสุดสำหรับ $\theta = PQ$ คือ

$$
\begin{aligned}
\hat{\theta} &= \hat{P}\hat{Q} = \frac{X}{n}\left[1-\frac{X}{n}\right] \\
E(\hat{\theta}) &= E\left[\frac{X}{n}\left(1-\frac{X}{n}\right)\right] \\
&= E\left[\frac{X}{n}-\frac{X^2}{n^2}\right] \\
&= \frac{E(X)}{n}-\frac{E(X^2)}{n^2} \\
&= \frac{nP}{n}-\frac{nPQ+n^2P^2}{n^2} \\
&= P-P^2-\frac{PQ}{n} \\
&= P(1-P)-\frac{PQ}{n} \\
&= PQ-\frac{PQ}{n} \\
&= \frac{n-1}{n}PQ
\end{aligned}
$$

ดังนั้น $\hat{\theta} = \hat{P}\hat{Q}$ เป็นตัวประมาณ *ที่เอนเอียง* สำหรับ $\theta = PQ$

โดยวิธีการแจ็คไนฟ์เมื่อใช้ m = 1 คือตัดค่าสังเกตคราวละ 1 ค่า แล้วจึงคำนวณค่าประมาณ $\hat{P}$ ดังนี้

$$
\hat{P}_i = \begin{cases} \frac{X-1}{n-1} & ; X_i = 1 \text{ นั่นคือ ถ้าค่าสังเกตตัวที่ i คือผลสำเร็จ} \\ \frac{X}{n} & ; X_i = 0 \text{ นั่นคือ ถ้าค่าสังเกตตัวที่ i คือผลล้มเหลว} \end{cases}
$$

ดังนั้น ตัวสถิติแจ็คไนฟ์ $\hat{\theta}_J = \hat{P}\hat{Q}$ จะคำนวณได้ดังนี้

$$\hat{\theta}_i = \begin{cases} \left(\frac{X-1}{n-1}\right)\left(1-\frac{X-1}{n-1}\right) = \left(\frac{X-1}{n-1}\right)\left(\frac{n-X}{n-1}\right) & ; X_i = 1 \\ \left(\frac{X}{n-1}\right)\left(1-\frac{X}{n-1}\right) = \left(\frac{X}{n-1}\right)\left(\frac{n-X-1}{n-1}\right) & ; X_i = 0 \end{cases}$$

เนื่องจากในตัวอย่างสุ่ม n มีจำนวนผลสำเร็จ X ครั้ง และจำนวนผลล้มเหลวมี n - X ครั้ง ดังนั้นจะมี $\hat{\theta}_i$ ที่ใช้สูตรเมื่อ $X_i = 1$ จำนวน X ครั้ง จะมี $\hat{\theta}_i$ ที่ใช้สูตรเมื่อ $X_i = 0$ จำนวน n - X ครั้ง

$$\begin{aligned}
\bar{\hat{\theta}} &= \frac{1}{n}\sum_{i=1}^{N}\hat{\theta}_i \\
&= \frac{1}{n}\left[X\left(\frac{X-1}{n-1}\right)\left(\frac{n-X}{n-1}\right) + (n-X)\left(\frac{X}{n-1}\right)\left(\frac{n-X-1}{n-1}\right)\right] \\
&= \hat{P}\frac{(X-1)(n-X)}{(n-1)^2} + (1-\hat{P})\frac{X(n-X-1)}{(n-1)^2} \\
&= \hat{P}\hat{Q}\frac{n(X-1)}{(n-1)^2} + \hat{P}\hat{Q}\,n\frac{(n-X-1)}{(n-1)^2} \\
&= \frac{\hat{P}\hat{Q}}{(n-1)^2}(nX-n+n^2-nX-n) \\
&= \frac{\hat{P}\hat{Q}}{(n-1)^2}(n^2-2n) \\
&= \frac{n(n-2)}{(n-1)^2}\hat{P}\hat{Q}
\end{aligned}$$

ดังนั้น ตัวประมาณแจ็คไนฟ์สำหรับ $\theta = PQ$ คือ

$$\begin{aligned}
J(\hat{\theta}) &= n\hat{\theta}-(n-1)\bar{\hat{\theta}} \\
&= n\hat{P}\hat{Q}-\frac{n(n-2)}{n-1}\hat{P}\hat{Q} \\
&= \frac{n}{n-1}\hat{P}\hat{Q}
\end{aligned}$$

เพราะว่า $E[J(\hat{\theta})] = \frac{n}{n-1}E(\hat{P}\hat{Q}) = PQ$ ดังนั้น ตัวประมาณแจ็คไนฟ์สำหรับ θ เป็นตัวประมาณที่ไม่เอนเอียง นอกจากนี้ยังสามารถพิสูจน์ได้ว่าตัวประมาณแจ็คไนฟ์กรณีนี้เป็นตัวประมาณไม่เอนเอียงที่มีความแปรปรวนต่ำสุดสำหรับ θ ด้วย

ตัวอย่างที่ 2 กรณีตัวแปรสุ่มต่อเนื่อง

ให้ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n ซึ่งสุ่มประชากรที่มีการแจกแจงแบบปกติ N(θ, 1) ตัวประมาณของ θ คือ $\hat{\theta} = \bar{X}$ เมื่อใช้วิธีการประมาณแบบแจ็คไนฟ์โดยใช้ m = 1 จะได้

$$\hat{\theta}_i = \frac{1}{n-1}\sum_{j\neq i} X_j$$

$$\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{N}\hat{\theta}_i$$

$$= \frac{1}{n(n-1)} \sum_{j \neq i}^{N} X_j$$

$$= \frac{1}{n(n-1)} \left[(n-1) \sum_{i=1}^{N} X_i \right]$$

$$= \bar{X}$$

ดังนั้น $J(\hat{\theta}) = n\bar{X}-(n-1)\bar{X} = \bar{X}$

ตัวประมาณแจ็คไนฟ์สำหรับ θ คือ $\bar{X}$ ซึ่งจะเห็นว่าเป็นตัวประมาณตัวเดียวกับตัวประมาณไม่เอนเอียงที่มีความแปรปรวนต่ำสุดของ θ

ตัวอย่างที่ 3 ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงแบบปกติ $N(\mu, \sigma^2)$ โดยที่ μ และ σ^2 ไม่ทราบค่า พิจารณาการประมาณ σ^2 ตัวประมาณภาวะน่าจะเป็นสูงสุดสำหรับ σ^2 คือ

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i-\bar{X})^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2-\bar{X}^2$$

เป็นตัวประมาณที่เอนเอียง โดยวิธีการแจ็คไนฟ์ ให้ m = 1 จะได้ว่า

$$\hat{\sigma}_i^2 = \frac{1}{n-1} \sum_{j \neq i} X_j^2 - \left(\frac{1}{n-1} \sum_{j \neq i} X_j\right)^2$$

$$= \frac{1}{n-1} \sum_{j \neq i} X_j^2 - \left(\frac{n\bar{X}-X_i}{n-1}\right)^2$$

$$= \frac{1}{n-1} \sum_{j \neq i} X_j^2 - \frac{n^2\bar{X}^2-2nX_iX+X_i^2}{(n-1)^2}$$

ดังนั้น $\bar{\hat{\sigma}}^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_i^2$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} X_j^2 - \frac{n^2\bar{X}^2-2n\bar{X}^2+\sum_{i=1}^{n} X_i^2/n}{(n-1)^2}$$

$$= \frac{(n-1)}{n(n-1)} \sum_{i=1}^{n} X_i^2 - \frac{(n^2-2n)\bar{X}^2}{(n-1)^2} - \frac{1}{n(n-1)^2} \sum_{i=1}^{n} X_i^2$$

$$= \frac{(n-1)^2-1}{n(n-1)^2} \sum_{i=1}^{n} X_i^2 - \frac{(n^2-2n)\bar{X}^2}{(n-1)^2}$$

$$= \frac{n^2-2n}{n(n-1)^2} \sum_{i=1}^{n} X_i^2 - \frac{(n^2-2n)\bar{X}^2}{(n-1)^2}$$

$$= \frac{n(n-2)}{n(n-1)^2} \left[\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 \right]$$

$$= \frac{n(n-2)}{(n-1)^2} \hat{\sigma}^2$$

ตัวประมาณความแปรปรวนโดยวิธีแจ็คไนฟ์ คือ

$$J(\hat{\sigma}^2) = n\hat{\sigma}^2-(n-1)\bar{\hat{\sigma}}^2$$

$$= n\hat{\sigma}^2 - \frac{n(n-2)}{n-1}\hat{\sigma}^2$$

$$= \frac{n(n-1)-n(n-2)}{n-1}\hat{\sigma}^2$$

$$= \frac{n(n-1-n+2)}{n-1}\hat{\sigma}^2$$

$$= \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = S^2$$

ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงสำหรับ σ^2 จะเห็นว่า วิธีการของแจ็คไนฟ์ไม่เพียงแต่ลดความเอนเอียงเท่านั้น แต่กลับพบว่าวิธีการแจ็คไนฟ์ทำให้ได้ตัวประมาณใหม่ซึ่งไม่เอนเอียงของ σ^2

ตัวอย่างที่ 4 กรณีที่ค่าที่เป็นไปได้ของ X_i ขึ้นอยู่กับพารามิเตอร์ θ

ให้ $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงแบบ Displaced Exponential ที่มี p.d.f.

$$f(x) = e^{-(x-\theta)}I_{(\theta,\infty)}(x)$$

ให้ $Y_1 \le Y_2 \le \ldots \le Y_n$ เป็นตัวสถิติอันดับ (Order Statistics) ของตัวอย่างสุ่ม จะได้ว่า $E(Y_1) = \theta+\frac{1}{n}$, $E(Y_2) = \theta+\frac{2n-1}{n(n-1)}$ ตัวประมาณภาวะน่าจะเป็นสูงสุดสำหรับ θ คือ $Y_1 = \min\{X_1, \ldots, X_n\}$ ซึ่งเป็นตัวประมาณที่เอนเอียง โดยใช้วิธีการแจ็คไนฟ์ให้ $m = 1$ จะได้

$$\hat{\theta}_i = \begin{cases} Y_1 \; ; X_i \ne Y_1 \\ Y_2 \; ; X_i = Y_1 \end{cases}$$

และ $\bar{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{N}\hat{\theta}_i = \frac{1}{n}\left[(n-1)Y_1+Y_2\right]$

ตัวประมาณแจ็คไนฟ์สำหรับ θ คือ

$$J(\hat{\theta}) = n\hat{\theta} - (n-1)\bar{\hat{\theta}}_i$$

$$= nY_1-(n-1)\left[\frac{n-1}{n}Y_1+\frac{1}{n}Y_2\right]$$

$$= nY_1-\frac{(n-1)^2}{n}Y_1-\frac{(n-1)}{n}Y_2$$

$$= Y_1-\frac{n-1}{n}(Y_2-Y_1)$$

$$E\left[J(\hat{\theta})\right] = E(Y_1)+\frac{n-1}{n}\left[E(Y_1)-E(Y_2)\right]$$

$$= \theta+\frac{1}{n}+\frac{n-1}{n}\left[\theta+\frac{1}{n}-\theta-\frac{2n-1}{n(n-1)}\right]$$

$$= \theta$$

จะเห็นว่า $J(\hat{\theta})$ เป็นตัวประมาณที่ไม่เอนเอียงสำหรับ θ

นิยาม 2 ให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณ 2 ตัวสำหรับ θ และให้ R เป็นค่าคงที่ค่าหนึ่ง โดย $R \neq 1$ ตัวประมาณแจ็คไนฟ์แบบทั่วไป (Generalized Jackknife Estimator; $G(\hat{\theta}_1, \hat{\theta}_2)$) คือ

$$G(\hat{\theta}_1, \hat{\theta}_2) = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1-R}$$

หมายเหตุ

1) ถ้าเลือก $R = \frac{n-1}{n}$ และ $\hat{\theta}_1 = \hat{\theta}$, $\hat{\theta}_2 = \hat{\theta}_i$ จะได้

$$\begin{aligned} G(\hat{\theta}_1, \hat{\theta}_2) &= \frac{\hat{\theta} - (\frac{n-1}{n})\hat{\theta}_i}{1 - \frac{(n-1)}{n}} \\ &= n\hat{\theta} - (n-1)\hat{\theta}_i \\ &= J(\hat{\theta}) \end{aligned}$$

2) ตัวประมาณแจ็คไนฟ์สามารถนำไปใช้ลดความเอนเอียงในตัวสถิติที่ไม่ใช้พารามิเตอร์ (Nonparametric Statistic) ได้เช่นเดียวกัน

3. การประมาณค่าแบบช่วง (Interval Estimation)

การใช้ตัวประมาณแจ็คไนฟ์มาสร้างช่วงความเชื่อมั่นโดยประมาณสำหรับ θ Tukey แนะนำไว้ว่า ค่าแฝง J_i ที่ได้จากการคำนวณโดยวิธีแจ็คไนฟ์ ก็คือ ตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระกัน (independent and identically distributed random variables) โดยที่เมื่อใช้ m = 1 จะได้ว่า

$$\begin{aligned} T &= \frac{\sqrt{n}\,(J(\hat{\theta}) - \theta)}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(J_i - J(\hat{\theta}))^2\right]^{\frac{1}{2}}} \\ &= \frac{\sqrt{n}\,(J(\hat{\theta}) - \theta)}{\sqrt{S_J^2}} \end{aligned}$$

มีการแจกแจงประมาณได้ด้วยการแจกแจงแบบ t ที่มีองศาความเป็นอิสระเท่ากับ n - 1 และเมื่อขนาดตัวอย่างใหญ่ T จะมีการแจกแจงลู่เข้าสู่การแจกแจงแบบปกติมาตรฐาน N(0, 1) ดังนั้น ขีดจำกัดความเชื่อมั่น $(1-\alpha)$ 100 % สำหรับ θ โดยประมาณ คือ

$$J(\hat{\theta}) \pm Z_{1-\alpha/2} \cdot \frac{S_J}{\sqrt{n}} \quad \text{เมื่อ } n \text{ มีขนาดใหญ่}$$

$$\text{หรือ } J(\hat{\theta}) \pm t_{1-\alpha/2} \cdot \frac{S_J}{\sqrt{n}} \quad \text{เมื่อ } n \text{ มีขนาดเล็ก}$$

หมายเหตุ ค่าของ $\hat{\theta}_i$ ซึ่งคำนวณได้จากการตัดค่าสังเกตที่ i ทิ้ง (กรณีใช้ m – 1) ทั้ง n ตัว จะไม่เป็นอิสระกัน (dependent) เนื่องจากค่าของ $\hat{\theta}_i$ สำหรับ i = 1, 2, ..., n คำนวณมาจากค่าสังเกต (X_i) เกือบเหมือนกันทุกค่า เช่น $\hat{\theta}_1$ คำนวณจาก $X_2, X_3, ..., X_n$ และ $\hat{\theta}_2$ คำนวณจาก $X_1, X_3, ..., X_n$ จะเห็นว่า $\hat{\theta}_1$ และ $\hat{\theta}_2$ ใช้ค่าสังเกต $X_3, ..., X_n$ ในการคำนวณทั้ง 2 ค่า ดังนั้น $\hat{\theta}_1$ และ $\hat{\theta}_2$ จะไม่เป็นอิสระกัน โดยวิธีการแจ็คไนฟ์จะนำค่า $\hat{\theta}_i$ เหล่านี้มาคำนวณหาค่าแฝง (J_i) คือ

$$
\begin{aligned}
J_i &= n\hat{\theta}-(n-1)\hat{\theta}_i \\
&= \hat{\theta}+(n-1)(\hat{\theta}-\hat{\theta}_i) \quad \text{เมื่อใช้ } m = 1
\end{aligned}
$$

จะทำให้ได้ค่าแฝง J_i มา n ค่าโดยค่าของ J_i จะมีการกระจายมากขึ้น กล่าวคือค่า J_i แต่ละค่าจะอยู่ห่างจาก $\hat{\theta}$ เท่ากับ n-1 เท่าของค่าที่ $\hat{\theta}_i$ ต่างจาก $\hat{\theta}$ ดังนั้นค่าของ J_i ; i = 1, ..., n จะเปรียบเสมือนเป็นตัวอย่างสุ่มของค่าแฝง (Pseudosample) ซึ่ง J_i แต่ละตัวเป็นอิสระกันและยังมีการแจกแจงเหมือนกันอีกด้วย

4. บทสรุป

จากที่กล่าวมาแล้วในหัวข้อ 2 และ 3 สามารถสรุปวิธีการประมาณค่าโดยวิธีแจ็คไนฟ์ได้ดังขั้นตอนต่อไปนี้

ขั้นตอนที่ 1 จากตัวอย่างสุ่มขนาด n คำนวณค่าประมาณแบบจุดสำหรับพารามิเตอร์ $\hat{\theta}_i$; i = 1, 2, ..., n โดยในการคำนวณค่า $\hat{\theta}_i$ จะตัดข้อมูลตัวที่ i (X_i) ทิ้ง ดังนั้นจะได้ค่า $\hat{\theta}_i$ จำนวน n ค่า โดยแต่ละค่าก็เป็นค่าประมาณของพารามิเตอร์ θ

ขั้นตอนที่ 2 คำนวณค่าแฝง (Pseudovalues) ด้วยสูตร

$$J_i = n\hat{\theta}- (n-1)\hat{\theta}_i \; ; \; i = 1, 2, ..., n$$

โดยที่ $\hat{\theta}$ คือตัวประมาณแบบจุดสำหรับ θ ที่คำนวณโดยใช้ข้อมูลทุกค่า เนื่องจาก $\hat{\theta}$ และ $\hat{\theta}_i$ ต่างก็เป็นตัวประมาณสำหรับ θ ดังนั้นค่าของ J_i แต่ละค่าจะเป็นค่าประมาณสำหรับ θ

ขั้นตอนที่ 3 ค่า $J_1, J_2, ..., J_n$ ที่คำนวณได้จากขั้นตอนที่ 2 จะเปรียบเสมือนกับเป็นตัวอย่างสุ่มชุดใหม่ที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ ให้คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ J_i ดังนี้

$$\bar{J} = \frac{1}{n}\sum_{i=1}^{n} J_i$$

$$S_J = \sqrt{\frac{\sum_{i=1}^{n}(J_i-\bar{J})^2}{n-1}}$$

ขั้นตอนที่ 4 คำนวณค่าความคลาดเคลื่อนมาตรฐาน (standard error) ของ $\bar{J}$

$$S_{\bar{J}} = \frac{S_J}{\sqrt{n}}$$

ขั้นตอนที่ 5 ตัวสถิติ $\bar{J}$ ที่ได้ในขั้นตอนที่ 3 จะเป็นตัวประมาณแบบจุดสำหรับ θ เพราะว่า $\bar{J}$ แต่ละตัวเป็นตัวประมาณสำหรับ θ ส่วนตัวประมาณแบบช่วงสำหรับ θ คำนวณได้ดังนี้

ขีดจำกัดความเชื่อมั่น $(1-\alpha)100\%$ สำหรับ θ คือ

$$\bar{J} \pm tS_{\bar{J}}$$

โดยที่ $t = t_{1-\alpha/2,(n-1)}$

$\bar{J}$ ได้จากขั้นตอนที่ 3 และ $S_{\bar{J}}$ ได้จากขั้นตอนที่ 4

บรรณานุกรม

Dudewicz, E.J. and Mishra, S.N. 1988. **Modern Mathematical Statistics.** New York: John Wiley & Sons.

Mosteller, F. and Tukey, J.W. 1977. **Data Analysis and Regression.** Massachusetts: Addison-Wesley Publishing Company.

Neter, J., Wasserman, W. and Whitmore, G.A. 1988. **Applied Statistics.** 3rd edition. Boston: Allyn and Bacon.

Wonnacott, T.H. and Wonnacott, R.J. 1984. **Introductory Statistics for Business and Economics.** 3rd edition. New York: John Wiley & Sons.