การประมาณช่วงความเชื่อมั่นของ p ในการแจกแจงทวินาม
Estimation of the confidence interval for binomial proportion

วันเพ็ญ จันทรังษี[1]

บทคัดย่อ

ปัญหาบางประการในการหาความน่าจะเป็นครอบคลุมของพารามิเตอร์ p สำหรับการแจกแจงทวินาม ซึ่งได้ถูกกล่าวไว้ในงานวิจัยหลาย ๆ ชิ้น อาทิ Brown *et al.* [1, 2] และ Agresti & Coull [3] นอกจากนี้ มีนักวิจัยจำนวนมากได้เสนอช่วงความเชื่อมั่นทางเลือกสำหรับพารามิเตอร์ p และได้หาค่าความน่าจะเป็นครอบคลุมและความกว้างคาดหวังของแต่ละช่วงความเชื่อมั่น จุดประสงค์ของบทความวิจัยนี้ เพื่อเสนอช่วงความเชื่อมั่นของพารามิเตอร์ p แบบใหม่ เพื่อหลีกเลี่ยงปัญหาเรื่องความน่าจะเป็นครอบคลุมและศึกษาลักษณะของช่วงความเชื่อมั่นที่มีอยู่แล้ว การจำลองข้อมูลได้ถูกนำมาใช้เพื่อสร้างตัวอย่างสุ่มสำหรับเปรียบเทียบความน่าจะเป็นครอบคลุมและความกว้างคาดหวังสำหรับแต่ละช่วงความเชื่อมั่น สำหรับการศึกษาในครั้งนี้จะมีการแนะนำช่วงความเชื่อมั่นของ p ว่าช่วงความเชื่อมั่นใดถูกนำมาใช้อย่างเหมาะสมในสถานการณ์ใด

คำสำคัญ: สัดส่วนทวินาม, ช่วงความเชื่อมั่น, ความน่าจะเป็นครอบคลุม, ความกว้างคาดหวัง

Abstract

The problematic behavior of the coverage probability of standard interval for binomial proportion (p) has been pointed out in several research articles such as in Brown *et al.* [1, 2] and Agresti & Coull [3]. A number of researchers have invented alternative confidence intervals for p. Each interval is investigated its coverage probability and expected length. The aims of this paper are to: 1) introduce the new confidence interval for p without the chaotic coverage properties, and 2) study the characteristics of the existing confidence intervals for p. Simulations are used to generate random samples in comparison to coverage probability and expected length for each confidence interval. Based on this study, the recommendations are given on which confidence interval should be used in what circumstances.

Keywords: binomial proportion, confidence interval, coverage probability, expected length

บทนำ

ในการทดลองเบอร์นูลลี (Bernoulli trials) n ครั้งที่เป็นอิสระต่อกัน ถ้าให้ผลลัพธ์ของการทดลองแต่ละครั้งแทนด้วยตัวแปรสุ่ม โดยที่

$$Y_i = \begin{cases} 1 & \text{ด้วยความน่าจะเป็น } p \\ 0 & \text{ด้วยความน่าจะเป็น } 1-p \end{cases}, \quad i = 1, 2, \ldots, n$$

พบว่า $X_n = \sum_{i=1}^{n} Y_i$ มีการแจกแจงทวินาม ซึ่งเขียนเป็นสัญลักษณ์ว่า $X_n \sim Bi\left(x; n, p\right)$ และ $x = 0, 1, 2, \ldots, n$

การแจกแจงเบอร์นูลลีและการแจกแจงทวินามต่างเป็นการแจกแจงไม่ต่อเนื่องที่รู้จักกันมานาน และเป็นที่ทราบกันดีว่าเมื่อตัวอย่างขนาดใหญ่ ($n \to \infty$) พบว่า

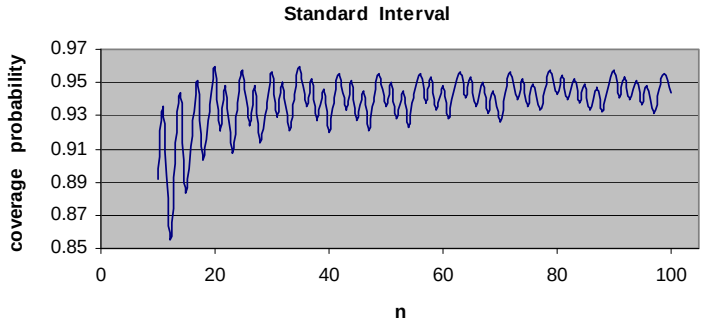
[1]สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม, E-mail: pun_kaset@hotmail.com

(1) $X_n \xrightarrow{D} X$ โดยที่ $X \sim Poisson(x;np)$ เมื่อ $n \to \infty$, $p \to 0$ และ np มีค่าจำกัด (finite) และ

(2) $$\frac{\sqrt{n}\left(\overline{X}_n - p\right)}{\sqrt{p(1-p)}} \xrightarrow{D} Z$$ โดยที่ $Z \sim N(0,1)$ เมื่อ $n \to \infty$, $p \to \frac{1}{2}$

ดังนั้น จึงมีงานวิจัยมากมายนำการแจกแจงทวินามไปใช้ แต่ยังมีบางประเด็นที่ยังเป็นปัญหา (Brown *et al.* [1]) คือ การประมาณช่วงความเชื่อมั่นของ p ที่ใช้กันอย่างกว้างขวางในรูปของ $\hat{p} \pm z_{\alpha/2} n^{-1/2} \left(\hat{p}\left(1-\hat{p}\right)\right)^{1/2}$ เมื่อ $\hat{p}$ คือ อัตราส่วนของจำนวนครั้งที่สำเร็จต่อการทดลองสุ่มทั้งหมด n ครั้ง และ $z_{\alpha/2}$ คือ เปอร์เซ็นต์ไทล์ที่ 100(1-α/2) ของการแจกแจงปกติ ซึ่งเรียกว่า ช่วงความเชื่อมั่น Wald หรือ ช่วงความเชื่อมั่นมาตรฐาน (standard interval) ที่คุ้นเคยและเป็นที่นิยมนั้นยังเป็นที่สงสัยว่ามีความถูกต้องมากน้อยเพียงใด

รูปภาพที่ 1 เป็นการหาความน่าจะเป็นครอบคลุม เมื่อพิจารณาที่ระดับนัยสำคัญ 0.05 โดยที่ $p = 0.5$ และโดยการจำลองข้อมูลจาก $Bi\left(n, 0.5\right)$, $n = 10, ..., 100$ พบว่า เมื่อ $n = 97$ มีความน่าจะเป็นครอบคลุมค่า p เท่ากับ 0.933 ซึ่งไม่เท่ากับ 0.95 ตามที่ต้องการแต่เมื่อ $n = 17$ กลับมีความน่าจะเป็นครอบคลุมค่า p เท่ากับ 0.951 ซึ่งไม่เท่ากับ 0.95 เช่นกัน จึงไม่สามารถที่จะรับประกันได้ว่าช่วงความเชื่อมั่นมาตรฐานที่ใช้อยู่ ทำให้มีค่าความน่าจะเป็นครอบคลุมค่า p เท่ากับ 0.95 เมื่อ n มีค่าเพิ่มขึ้นหรือไม่ Brown *et al.* [1] ได้กล่าวไว้ว่า ความน่าจะเป็นครอบคลุมมีการแกว่ง (oscillation) เกิดมาจากความไม่ต่อเนื่องของการแจกแจงทวินาม (discreteness of the binomial distribution)



รูปภาพที่ 1 ความน่าจะเป็นครอบคลุม เมื่อ p = 0.5, n = 10,...,100

ในการประมาณช่วงความเชื่อมั่นของ p ดังกล่าวช่วงความเชื่อมั่นต้องมีคุณสมบัติ 2 ประการ คือ ประการแรก ช่วงความเชื่อมั่นต้องคลุมพารามิเตอร์ที่ต้องการประมาณค่า นั่นคือ ความน่าจะเป็นที่ช่วงสุ่มนี้จะคลุมค่าที่แท้จริงของ p (coverage probability) สูง ประการที่สอง ช่วงความเชื่อมั่นที่มีความกว้าง (length) ต่ำสุดหรือ มีค่าคาดหมายของความกว้าง (expected length) ต่ำสุด ดังนั้น นักวิจัยจึงได้เสนอแนวทางอื่นในการประมาณช่วงความเชื่อมั่นของ p ไว้หลายแนวทาง อาทิ ช่วงความเชื่อมั่น Agresti-Coull [3] ช่วงความเชื่อมั่น Wilson [4] ช่วงความเชื่อมั่น recentered [5] เป็นต้น

ด้วยเหตุนี้ การวิจัยครั้งนี้จึงสนใจการประมาณช่วงความเชื่อมั่นของ p เพื่อพิจารณาว่า วิธีการประมาณช่วงความเชื่อมั่นของ p ต่าง ๆ วิธีใดมีความน่าจะเป็นครอบคลุมและมีความยาวช่วงเหมาะสมกว่า และเหมาะสมกว่าในสถานการณ์ใด

ช่วงความเชื่อมั่นที่สนใจศึกษา

ข้อมูลภายใต้การแจกแจงทวินาม เมื่อ ตัวแปรสุ่ม X มีการแจกแจงแบบทวินามที่มีพารามิเตอร์ n และ p โดย $x = 0,1,...,n$ ช่วงความเชื่อมั่นของ p ที่ใช้ศึกษาเปรียบเทียบ ได้แก่

1. ช่วงความเชื่อมั่นมาตรฐาน (standard interval หรือ Wald interval)

ลิมิตความเชื่อมั่น $100(1-\alpha)\%$ คือ $\hat{p} \pm z_{\alpha/2} n^{-1/2} \left(\hat{p}\left(1-\hat{p}\right)\right)^{1/2}$

2. ช่วงความเชื่อมั่น recentered หรือช่วง BCD (recentered interval)

ช่วงความเชื่อมั่นนี้เสนอโดย Brown *et al.* [1] นั่นคือ $100(1-\alpha)\%$ ของช่วง BCD คือ $\left[p_{BCD}^{L}, p_{BCD}^{U}\right]$ เมื่อ

$$p_{BCD}^{L} = \tilde{p} - z_{\alpha/2} n^{-1/2} \left(\hat{p}\hat{q}\right)^{1/2} \text{ และ } p_{BCD}^{U} = \tilde{p} + z_{\alpha/2} n^{-1/2} \left(\hat{p}\hat{q}\right)^{1/2},\ \tilde{p} = \frac{X + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2}$$

3. ช่วงความเชื่อมั่น Agresti-Coull

ช่วงความเชื่อมั่น Agresti-Coull [3] นี้ปรับปรุงมาจากช่วงความเชื่อมั่นมาตรฐาน จึงมีสูตรที่คล้ายคลึงกัน แต่ต่างกันที่ใช้ $\tilde{p}$ แทน $\hat{p}$ ดังนั้น $100(1-\alpha)\%$ ของช่วง Agresti-Coull คือ $\left[p_{AC}^{L}, p_{AC}^{U}\right]$ เมื่อ

$$p_{AC}^{L} = \tilde{p} - z_{\alpha/2} n^{-1/2} \left(\tilde{p}\tilde{q}\right)^{1/2} \text{ และ } p_{AC}^{U} = \tilde{p} + z_{\alpha/2} n^{-1/2} \left(\tilde{p}\tilde{q}\right)^{1/2},\ \tilde{q} = 1-\tilde{p}$$

กรณี $\alpha = 0.05$ จะใช้ $z_{\alpha/2}^2 = 1.96^2 \approx 4$ และ $\tilde{p} = \frac{X + z_{\alpha/2}^2/2}{n + z_{\alpha/2}^2} \approx \frac{X+2}{n+4}$

4. ช่วงความเชื่อมั่น Wilson (Wilson interval)

ช่วงความเชื่อมั่นนี้ถูกพัฒนาโดย Wilson [4] โดยใช้ทฤษฎีแนวโน้มเข้าสู่ศูนย์กลาง (central limit theorem) ได้ว่า $\frac{n^{1/2}\left(\hat{p}-p\right)}{\sqrt{p\left(1-p\right)}} \xrightarrow{D} N(0,1)$ ดังนั้น

$$P\left(-z_{\alpha/2} \le \frac{n^{1/2}\left(\hat{p}-p\right)}{\sqrt{p\left(1-p\right)}} \le z_{\alpha/2}\right) = P\left(\frac{n\left(\hat{p}-p\right)^2}{p\left(1-p\right)} \le z_{\alpha/2}^2\right) = 1-\alpha$$

การหาช่วงของ p ทำได้โดยแก้สมการ $\frac{n\left(\hat{p}-p\right)^2}{p\left(1-p\right)} = z_{\alpha/2}^2$

$$n\hat{p}^2 - 2n\hat{p}p + np^2 = z_{\alpha/2}^2 p - z_{\alpha/2}^2 p^2$$

$$(n + z_{\alpha/2}^2)p^2 - (2n\hat{p} + z_{\alpha/2}^2)p + n\hat{p}^2 = 0$$

สมการจะอยู่ในรูป $ap^2 + bp + c = 0$ มีผลเฉลยคือ $p = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

นั่นคือ $a = (n + z_{\alpha/2}^2),\ b = -(2n\hat{p} + z_{\alpha/2}^2),\ c = n\hat{p}^2$ ทำให้

$$p = \frac{(2n\hat{p} + z_{\alpha/2}^2) \pm \sqrt{(2n\hat{p} + z_{\alpha/2}^2)^2 - 4(n + z_{\alpha/2}^2)(np^2)}}{2(n + z_{\alpha/2}^2)}$$

$$= \frac{(2n\hat{p} + z_{\alpha/2}^2) \pm \sqrt{4n^2\hat{p}^2 + 4nz_{\alpha/2}^2\hat{p} + z_{\alpha/2}^4 - 4n^2\hat{p}^2 - 4nz_{\alpha/2}^2\hat{p}^2}}{2(n + z_{\alpha/2}^2)}$$

$$= \frac{n\hat{p} + z^2_{\alpha/2}/2}{n + z^2_{\alpha/2}} \pm \frac{z_{\alpha/2}}{n + z^2_{\alpha/2}}\sqrt{n\hat{p}(1-\hat{p}) + \frac{z^2_{\alpha/2}}{4}}$$

$$= \frac{X + z^2_{\alpha/2}/2}{n + z^2_{\alpha/2}} \pm \frac{z_{\alpha/2}}{n + z^2_{\alpha/2}}\sqrt{n\hat{p}(1-\hat{p}) + \frac{z^2_{\alpha/2}}{4}}$$

ดังนั้น $100(1-\alpha)\%$ ของช่วงความเชื่อมั่น Wilson คือ $\left[p^L_W, p^U_W\right]$ เมื่อ $p^L_W = \tilde{p} - \frac{z_{\alpha/2}}{\tilde{n}}\left(n\hat{p}\hat{q} + \frac{z^2_{\alpha/2}}{4}\right)^{1/2}$ และ

$p^U_W = \tilde{p} + \frac{z_{\alpha/2}}{\tilde{n}}\left(n\hat{p}\hat{q} + \frac{z^2_{\alpha/2}}{4}\right)^{1/2}$ ที่ซึ่ง $\tilde{p} = \frac{\tilde{X}}{\tilde{n}}$, $\tilde{X} = n\hat{p} + z^2_{\alpha/2}/2$, $\tilde{n} = n + z^2_{\alpha/2}$

การสร้างช่วงความเชื่อมั่นของ p แบบใหม่

สำหรับช่วงความเชื่อมั่นที่จะเสนอนี้ จะใช้ชื่อว่า ช่วง WP เป็นช่วงความเชื่อมั่นที่ผู้วิจัยสร้างขึ้นจากการประยุกต์ใช้ช่วงความเชื่อมั่น Wilson [4] เมื่อ

$$\frac{X + z^2_{\alpha/2}/2}{n + z^2_{\alpha/2}} \pm \frac{z_{\alpha/2}}{n + z^2_{\alpha/2}}\sqrt{n\hat{p}(1-\hat{p}) + \frac{z^2_{\alpha/2}}{4}}$$

$$= \frac{n}{n + z^2_{\alpha/2}}\left(\frac{X + z^2_{\alpha/2}/2}{n} \pm \frac{z_{\alpha/2}}{n}\sqrt{n\hat{p}(1-\hat{p}) + \frac{z^2_{\alpha/2}}{4}}\right)$$

จะเห็นว่าเมื่อ $n \to \infty$ และ $np \to a$ โดยที่ a เป็นค่าคงตัว $\frac{n}{n + z^2_{\alpha/2}} \to 1$, $\frac{z^2_{\alpha/2}}{2n} \to 0$ และ $\frac{1}{n} \to 0$ ดังนั้น ช่วงความเชื่อมั่น Wilson ก็จะเข้าใกล้ช่วงความเชื่อมั่นมาตรฐาน แต่จากการศึกษาพบว่า ช่วงความเชื่อมั่นมาตรฐานมีความน่าจะเป็นครอบคลุมต่ำ เมื่อ $p \to 0$ หรือ $p \to 1$ โดยเฉพาะกรณีที่ n มีค่าน้อย ๆ แต่การประมาณช่วงจะดีขึ้นเมื่อ n มีค่าเพิ่มขึ้น

กรณีที่ n มีค่าเพิ่มขึ้น $\frac{n}{n + z^2_{\alpha/2}}$ มีค่าเข้าใกล้ 1 สำหรับช่วง WP จึงตัดพจน์นี้ทิ้งไป แต่อาจมีผลกระทบกับช่วงความเชื่อมั่นเมื่อ n มีค่าน้อย จะได้

$$\left(\frac{X}{n} + \frac{z^2_{\alpha/2}}{2n}\right) \pm \frac{z_{\alpha/2}}{n}\sqrt{n\hat{p}(1-\hat{p}) + \frac{z^2_{\alpha/2}}{4}}$$

ปรับจุดศูนย์กลางช่วงโดยตัดพจน์ $\frac{z^2_{\alpha/2}}{2n}$ ออก เพราะเมื่อ n มีค่าเพิ่มขึ้น ทำให้พจน์นี้มีค่าเข้าใกล้ 0 และประมาณ $z^2_{\alpha/2} = 1.96^2 \approx 4$ จะได้ $\left(\frac{X}{n}\right) \pm \frac{z_{\alpha/2}}{n}\sqrt{n\hat{p}(1-\hat{p}) + 1}$ นั่นคือ $100(1-\alpha)\%$ ช่วง WP คือ $\left[p^L_{WP}, p^U_{WP}\right]$ เมื่อ

$$p^L_{WP} = \hat{p} - \frac{z_{\alpha/2}}{n}\left(n\hat{p}\hat{q} + 1\right)^{1/2} \text{ และ } p^U_{WP} = \hat{p} + \frac{z_{\alpha/2}}{n}\left(n\hat{p}\hat{q} + 1\right)^{1/2}$$

เมื่อ $\hat{q} = 1 - \hat{p}$ ซึ่งช่วง WP จะถูกนำไปใช้เปรียบเทียบกับช่วงความเชื่อมั่นอื่น

เกณฑ์ที่ใช้ในการเปรียบเทียบช่วงความเชื่อมั่น

จะเปรียบเทียบช่วงความเชื่อมั่นทั้ง 5 ช่วง ที่ได้แสดงไปแล้วก่อนหน้านี้ โดยการใช้เกณฑ์ต่าง ๆ คือ ความน่าจะเป็นครอบคลุม ความน่าจะเป็นครอบคลุมเฉลี่ย ความกว้างคาดหวังเฉลี่ย และความคลาดเคลื่อนรากกำลังสองเฉลี่ย

1. ความน่าจะเป็นครอบคลุม (coverage probability)
ความน่าจะเป็นครอบคลุม คือ ความน่าจะเป็นที่ช่วงความเชื่อมั่นจะคลุมค่าพารามิเตอร์ p นิยามโดย

$$C(p,n) = P_p(p \in CI) = \sum_{x=0}^{n} I(x,p)\binom{n}{x} p^x (1-p)^{1-x}, \quad 0 < p < 1$$

ซึ่ง $I(x,p) = 1$ เมื่อช่วงความเชื่อมั่นคลุมค่า p และเท่ากับ 0 เมื่อช่วงความเชื่อมั่นไม่คลุมค่า p จุดประสงค์ คือ ต้องการสร้างช่วงความเชื่อมั่นที่คลุมค่า p ซึ่งมีความน่าจะเป็นครอบคลุมเท่ากับ $1-\alpha$

2. ความน่าจะเป็นครอบคลุมเฉลี่ย (average coverage probability หรือ ACP)
ความน่าจะเป็นครอบคลุมเฉลี่ย คือ ความน่าจะเป็นที่ช่วงความเชื่อมั่นจะคลุมค่าพารามิเตอร์ p โดยเฉลี่ย นิยามโดย

$$\overline{C}(n) = \int_0^1 C(p,n)\,dp = \int_0^1 \sum_{x=0}^{n} I(x,p)\binom{n}{x} p^x (1-p)^{n-x}\,dp$$

3. ความกว้างคาดหวังเฉลี่ยช่วงความเชื่อมั่น (average expected length หรือ AEL)
ความกว้างคาดหวังเฉลี่ยช่วงความเชื่อมั่น นิยามโดย

$$\int_0^1 E_{p,n}(length(CI))\,dp = \sum_{x=0}^{n} (U(x,n) - L(x,n))\binom{n}{x} \int_0^1 p^x (1-p)^{n-x}\,dp$$

เมื่อ $E_{p,n}(length(CI)) = \sum_{x=0}^{n} (U(x,n) - L(x,n))\binom{n}{x} p^x (1-p)^{n-x}$ ที่ซึ่ง U และ L เป็น เป็นลิมิตบนและลิมิตล่างของช่วงความเชื่อมั่น

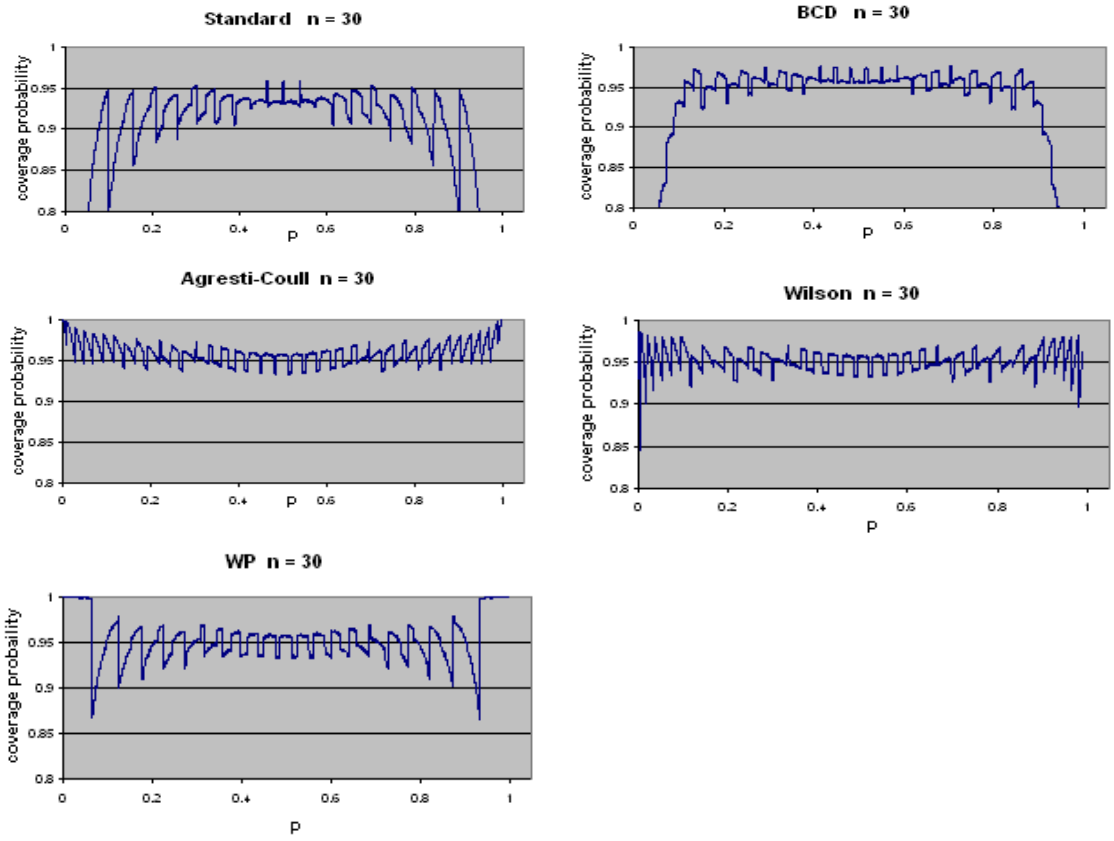
4. ความคลาดเคลื่อนรากกำลังสองเฉลี่ย (root mean square error หรือ RMSE)

$$RMSE = \sqrt{\int_0^1 \left[C(p,n) - (1-\alpha)\right]^2 dp}$$

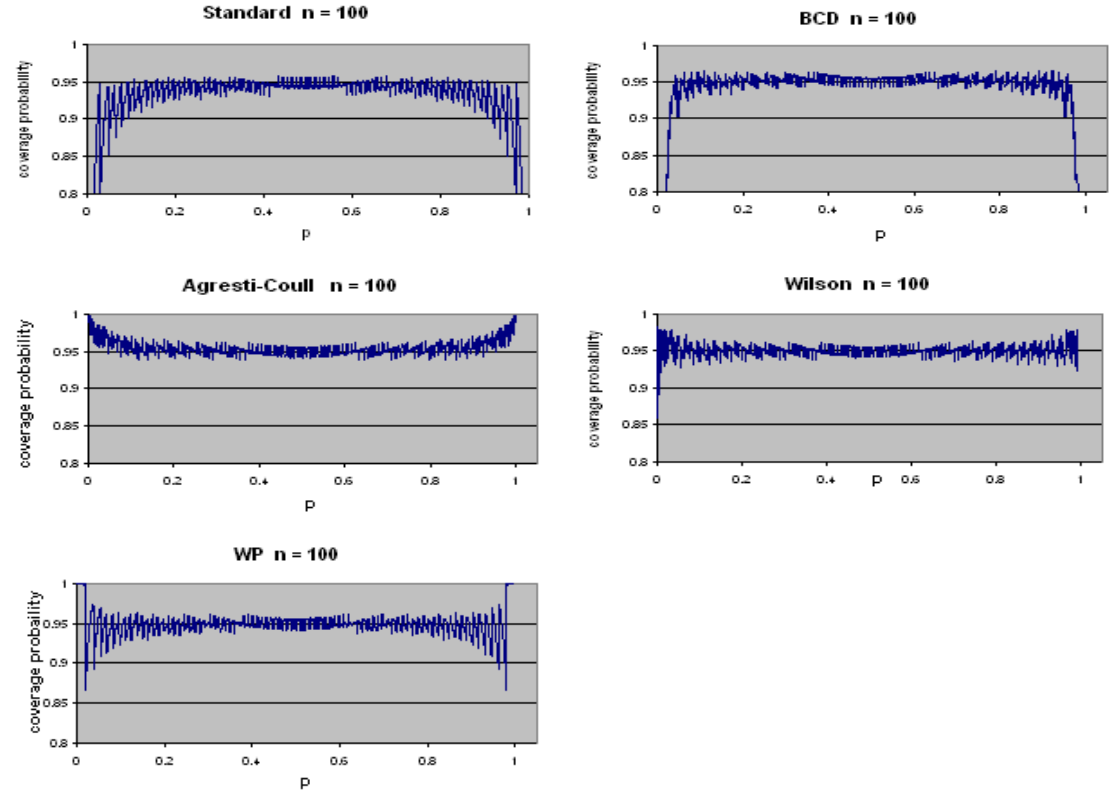
เมื่อ $C(p,n)$ คือ ความน่าจะเป็นครอบคลุม เป็นการวัดความน่าจะเป็นครอบคลุมห่างจาก $1-\alpha$ โดยเฉลี่ยแล้วเป็นระยะทางมากน้อยเพียงใด

การเปรียบเทียบช่วงความเชื่อมั่น

โดยการใช้เทคนิคการจำลอง (simulation technique) และโปรแกรม Matlab ในการคำนวณ และการสร้างตัวอย่างสุ่มได้จำลองข้อมูลจาก $Bi(30,p)$ และ $Bi(100,p)$ เมื่อ $p = 0.0005(0.0005)1$ เพื่อเปรียบเทียบช่วงความเชื่อมั่นของค่า p จากช่วงความเชื่อมั่นทั้ง 5 ช่วง แสดงในรูปภาพที่ 2 และ 3 ตามลำดับ



รูปภาพที่ 2 ความน่าจะเป็นครอบคลุมค่า p ที่ระดับนัยสำคัญ 0.05 เมื่อ n =30



รูปภาพที่ 3 ความน่าจะเป็นครอบคลุมค่า p ที่ระดับนัยสำคัญ 0.05 เมื่อ n =100

ค่าความน่าจะเป็นครอบคลุมสำหรับช่วงความเชื่อมั่น 95% เมื่อ $n = 30$ ได้ข้อสังเกตดังนี้

1. ช่วงความเชื่อมั่นมาตรฐาน (standard interval): ความน่าจะเป็นครอบคลุมพารามิเตอร์ p มีค่าถึง 0.95 มากกว่าเมื่อ $n = 10$ แต่โดยส่วนใหญ่ ค่าความน่าจะเป็นครอบคลุมยังไม่ถึง 0.95 โดยเฉพาะเมื่อ p เข้าใกล้ 0 หรือ 1 ความน่าจะเป็นครอบคลุม p จะมีค่าต่ำมาก เช่น $C(0.005, 30) = 0.13831$ และ $C(0.995, 30) = 0.13975$

2. ช่วง BCD: ความน่าจะเป็นครอบคลุมพารามิเตอร์ p ยังคงมีค่าต่ำมากเมื่อ p เข้าใกล้ 0 หรือ 1 และ p บางค่าที่อยู่ใกล้ค่าขอบมีความน่าจะเป็นครอบคลุมเป็นศูนย์ เช่น $C(0.0005, 10) = 0$ แต่เมื่อ $0.0075 < p < 0.9925$ มีความน่าจะเป็นครอบคลุมมากกว่า 0.7

3. ช่วง Agresti-Coull: ความน่าจะเป็นครอบคลุมพารามิเตอร์ p โดยเฉลี่ยมีค่าสูงกว่า 0.95 รวมทั้งเมื่อ p เข้าใกล้ 0 หรือ 1 ถึงแม้ว่ามีบางค่าของ p จะมีความน่าจะเป็นครอบคลุมไม่ถึง 0.95 แต่ก็ไม่ต่ำกว่า 0.93

4. ช่วง Wilson: ความน่าจะเป็นครอบคลุมพารามิเตอร์ p โดยเฉลี่ยมีค่าสูงกว่า 0.95 รวมทั้งเมื่อ p เข้าใกล้ 0 หรือ 1

5. ช่วง WP: ความน่าจะเป็นครอบคลุมพารามิเตอร์ p มีค่าสูงกว่า 0.95 เมื่อ p เข้าใกล้ 0 หรือ 1 ซึ่งสูงกว่าช่วงทั้ง 4 แบบ ข้างต้น โดยความน่าจะเป็นครอบคลุมพารามิเตอร์ p ไม่ต่ำกว่า 0.85

จากรูปภาพที่ 3 ช่วงความเชื่อมั่นมาตรฐานและ BCD เมื่อ p มีค่าเข้าใกล้ 0 หรือ 1 ค่าความน่าจะเป็นครอบคลุมยังคงมีค่าต่ำเมื่อเทียบกับช่วงอื่น ๆ

ตารางที่ 1 ความน่าจะเป็นครอบคลุมเฉลี่ย (ACP) ความกว้างคาดหวังเฉลี่ย (AEL) ค่าคลาดเคลื่อนรากกำลังสองเฉลี่ย (RMSE) ที่ระดับความเชื่อมั่น 95%

ช่วงความเชื่อมั่น	เกณฑ์ที่ใช้	$n = 10$	$n = 30$	$n = 50$	$n = 100$	$n = 200$
Standard	ACP	0.76924	0.87494	0.90045	0.92223	0.9345
	AEL	0.40537	0.25507	0.20112	0.14415	0.10265
	RMSE	0.28637	0.16819	0.13125	0.09303	0.0662
BCD	ACP	0.7934	0.88399	0.90629	0.92565	0.93667
	AEL	0.41607	0.25739	0.20222	0.14462	0.10283
	RMSE	0.30453	0.22299	0.18295	0.13645	0.09949
Agresti-Coull	ACP	0.96375	0.95988	0.95777	0.95546	0.95349
	AEL	0.44633	0.26868	0.20859	0.14729	0.10388
	RMSE	0.02191	0.01419	0.01193	0.00895	0.00659
Wilson	ACP	0.95405	0.95253	0.95163	0.951	0.95077
	AEL	0.41467	0.25749	0.20238	0.1447	0.10287
	RMSE	0.02191	0.01419	0.01193	0.00895	0.00659
WP	ACP	0.9666	0.95355	0.95089	0.94943	0.9489
	AEL	0.59139	0.29432	0.21983	0.15096	0.10508
	RMSE	0.03507	0.02454	0.02019	0.01512	0.01153

จากตารางที่ 1 สำหรับช่วงความเชื่อมั่นมาตรฐานและ BCD มีความน่าจะเป็นครอบคลุมเฉลี่ยต่ำเมื่อ n มีค่าน้อย แต่เมื่อ n ยิ่งมากขึ้น กราฟความน่าจะเป็นครอบคลุมเฉลี่ยจะเริ่มลู่เข้าสู่ 0.95 แต่ก็ยังต่ำกว่าช่วงซึ่งได้จากวิธีของ Agresti-Coull Wilson และ WP อยู่ดี และในบรรดา 3 ช่วงหลังมีความน่าจะเป็นครอบคลุมเฉลี่ยใกล้ 0.95 ทุกค่า n โดยช่วง Agresti-Coull มีความน่าจะเป็นครอบคลุมเฉลี่ยมากที่สุด รองลงมาคือ ช่วง Wilson และ WP ตามลำดับ เมื่อ n มีค่ามากความกว้างคาดหวัง

เฉลี่ยของช่วงความเชื่อมั่นไม่มีความแตกต่างกัน โดยเฉพาะเมื่อ n มากกว่า 100 แต่เมื่อ n มีค่าน้อยกว่า 100 จะได้ว่าความกว้างคาดหวังเฉลี่ยของช่วงความเชื่อมั่นแตกต่างกัน เรียงจากมากไปน้อย ได้ดังนี้ ช่วง WP, ช่วง Agresti-Coull, ช่วง BCD, ช่วง Wilson และช่วงความเชื่อมั่นมาตรฐาน

ค่าคลาดเคลื่อนรากเฉลี่ยกำลังสองของช่วงความเชื่อมั่น 95% จากมากไปน้อยได้ ดังนี้ BCD, standard, WP, Wilson และ Agresti โดยช่วง Wilson กับ Agresti มีค่า RMSE เท่า ๆ กัน ซึ่งแสดงว่า Wilson กับ Agresti มีความน่าจะเป็นครอบคลุมห่างจาก 0.95 โดยเฉลี่ยแล้วเป็นระยะทางน้อยที่สุด

สรุปผลการวิจัยและข้อเสนอแนะ

ช่วงความเชื่อมั่น WP ที่เสนอขึ้นมานั้นมีความกว้างคาดหวังสูงกว่าทั้ง 4 ช่วงที่เหลือ คือ ช่วงความเชื่อมั่นมาตรฐาน ช่วง Wilson ช่วง BCD และ ช่วง Agresti-Coull ซึ่งยังเป็นช่วงความเชื่อมั่นที่ไม่ดีนัก เพราะกรณี n มีค่าน้อย ถึงแม้ว่าช่วง WP จะมีความน่าจะเป็นครอบคลุมที่สูง แต่ก็มีค่าความกว้างคาดหวังสูงกว่าทั้ง 4 ช่วง ถือว่ายังไม่เหมาะสมในการที่จะนำไปใช้ในกรณีนี้

กรณี n มีค่ามาก โดยเฉพาะเมื่อ n มากกว่า 100 ช่วงความเชื่อมั่น WP ให้ค่าความน่าจะเป็นครอบคลุมพารามิเตอร์ p สูง ซึ่งจากผลการวิจัยพบว่าให้ความกว้างคาดหวังของช่วงพอ ๆ กับทั้ง 4 ช่วง ในกรณีนี้ ช่วงความเชื่อมั่น WP เป็นอีกทางเลือกหนึ่งที่สามารถนำไปใช้ได้เพราะสูตรมีความง่ายกว่าสูตรของ Wilson และ Agresti-Coull

ในทางปฏิบัติ นิยมที่จะใช้ช่วงความเชื่อมั่นที่มีความกว้างคาดหวังแคบกว่า แม้ว่าจะมีความน่าจะเป็นครอบคลุมพารามิเตอร์ p ไม่ถึง 0.95 ก็ตาม แต่ถ้ามีความน่าจะเป็นครอบคลุมพารามิเตอร์ p เท่ากัน แน่นอนว่าเราต้องเลือกช่วงความเชื่อมั่นที่แคบกว่า

เอกสารอ้างอิง

[1] Brown, L.D., Cai, T., & Dasgupta, A. (2001). Interval estimation for a binomial proportion (with discussion). **Statistcal Science**, 16 (2), 101-133.

[2] Brown, L.D., Cai, T., & Dasgupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. **The Annals of Statistics**, 30 (1), 160-201.

[3] Agresti, A. & Coull, B.A. (1998). Approximate is better than exact for interval estimation of binomial proportions. **The American Statistician**, 52 (2), 119-126.

[4] Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference. **Journal of the American Statistical Association**, 22 (158), 209-212.

[5] Brown, L. & Li, X. (2005). Confidence intervals for two sample binomial distribution. **Journal of Statistical Planning and Inference,** 130, 359-375.