

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีและสถิติที่ใช้ในการวิจัย

การแจกแจงเบอร์นูลลี (Bernoulli distribution)

เมื่อทำการทดลองสุ่มหนึ่งๆ ที่มีผลลัพธ์ที่เป็นไปได้ 2 ลักษณะ คือ ลักษณะที่สนใจหรือเกิดผลสำเร็จ (Success) และลักษณะที่ไม่สนใจหรือไม่เกิดผลสำเร็จ (Failure) โดยผลลัพธ์ทั้งสองลักษณะแยกจากกันโดยเด็ดขาด ความน่าจะเป็นที่เกิดผลสำเร็จในการทดลองเท่ากับ p และความน่าจะเป็นที่ไม่เกิดผลสำเร็จในการทดลองเท่ากับ $1 - p$ การทดลองนี้เรียกว่าการทดลองเบอร์นูลลี

ให้ X แทน จำนวนครั้งที่เกิดผลสำเร็จในการทดลอง

ตัวแปรสุ่ม X มีการแจกแจงเบอร์นูลลีที่มีฟังก์ชันความน่าจะเป็น

$$f(x) = p^x (1 - p)^{1-x} \quad , \quad x = 0, 1 \quad , \quad 0 \leq p \leq 1$$

แล้ว $E(X) = p$ และ $Var(X) = p(1 - p)$

การแจกแจงปัวซอง (Poisson distribution)

การแจกแจงปัวซองเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม ที่พิจารณาถึงจำนวนลักษณะที่สนใจที่เกิดขึ้น ในช่วงเวลาใดเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่ง

ให้ X แทน จำนวนครั้งที่เกิดลักษณะที่สนใจ ในช่วงเวลาใดเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่ง

ตัวแปรสุ่ม X มีการแจกแจงปัวซองที่มีฟังก์ชันความน่าจะเป็น

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad , \quad x = 0, 1, 2, \dots \quad , \quad \lambda > 0$$

แล้ว $E(X) = \lambda$ และ $Var(X) = \lambda$

เมื่อ λ แทน ค่าเฉลี่ยของจำนวนเหตุการณ์ที่เกิดลักษณะที่สนใจ ในช่วงเวลาใดเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่ง

การแจกแจงเรขาคณิต (Geometric distribution)

การแจกแจงเรขาคณิตเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม ที่มาจากการทดลองเบอร์นูลลีซ้ำๆ กัน ไปเรื่อยๆ จนกระทั่งเกิดลักษณะที่สนใจหรือผลสำเร็จเป็นครั้งแรกจึงหยุดการทดลอง โดยการทดลองแต่ละครั้งเป็นอิสระต่อกัน ความน่าจะเป็นที่เกิดผลสำเร็จในการทดลองแต่ละครั้งเท่ากับ p และความน่าจะเป็นที่ไม่เกิดผลสำเร็จในการทดลองแต่ละครั้งเท่ากับ $1 - p$ การทดลองนี้เรียกว่าการทดลองเรขาคณิต

ให้ X แทน จำนวนครั้งที่ทำการทดลองจนกระทั่งเกิดผลสำเร็จเป็นครั้งแรก

ตัวแปรสุ่ม X มีการแจกแจงเรขาคณิตที่มีฟังก์ชันความน่าจะเป็น

$$f(x) = p(1 - p)^{x-1} \quad , \quad x = 1, 2, 3, \dots \quad , \quad 0 \leq p \leq 1$$

แล้ว

$$E(X) = \frac{1}{p} \quad \text{และ} \quad \text{Var}(X) = \frac{(1-p)}{p^2}$$

การแจกแจงแบบเลขชี้กำลัง (Exponential distribution)

เมื่อกล่าวถึงการทดลองปัวซอง ที่พิจารณาจำนวนเหตุการณ์ที่เกิดลักษณะที่สนใจในช่วงเวลาใดเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่ง จำนวนเหตุการณ์ที่เกิดเป็นตัวแปรสุ่มแบบไม่ต่อเนื่องที่มีการแจกแจงปัวซอง ไม่เพียงแต่จำนวนเหตุการณ์ที่เกิดลักษณะที่สนใจเป็นตัวแปรสุ่มเท่านั้น ช่วงเวลาของการรอคอยที่เกิดลักษณะที่สนใจ ก็เป็นตัวแปรสุ่มเช่นกัน

ให้ X แทน ช่วงเวลาของการรอคอย จนกระทั่งเกิดลักษณะที่สนใจหรือเกิดผลสำเร็จเป็นครั้งแรก

ตัวแปรสุ่ม X มีการแจกแจงแบบเลขชี้กำลังที่มีฟังก์ชันความน่าจะเป็น

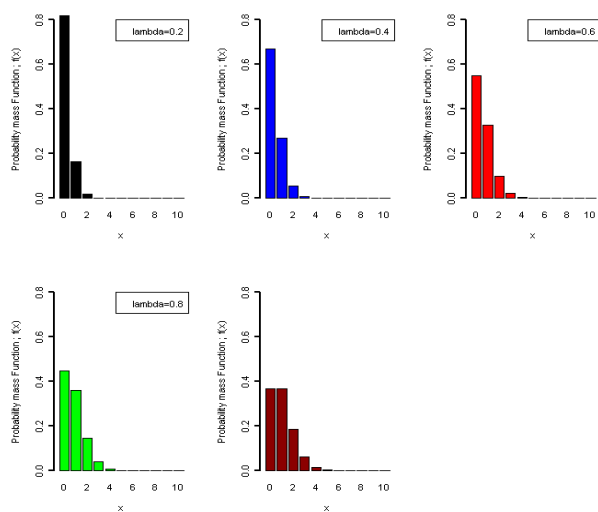
$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad , \quad x > 0 \quad , \quad \beta > 0$$

แล้ว

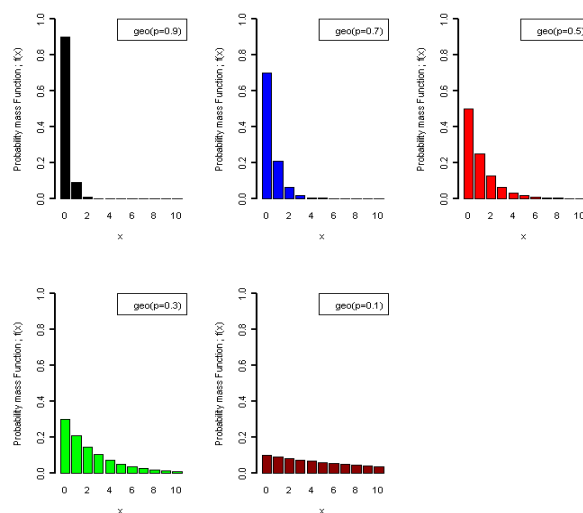
$$E(X) = \beta \quad \text{และ} \quad \text{Var}(X) = \beta^2$$

เมื่อ $\beta = \frac{1}{\lambda}$ แทน ค่าเฉลี่ยของช่วงเวลาของการรอคอย จนกระทั่งเกิดลักษณะที่สนใจหรือเกิดผลสำเร็จเป็นครั้งแรก

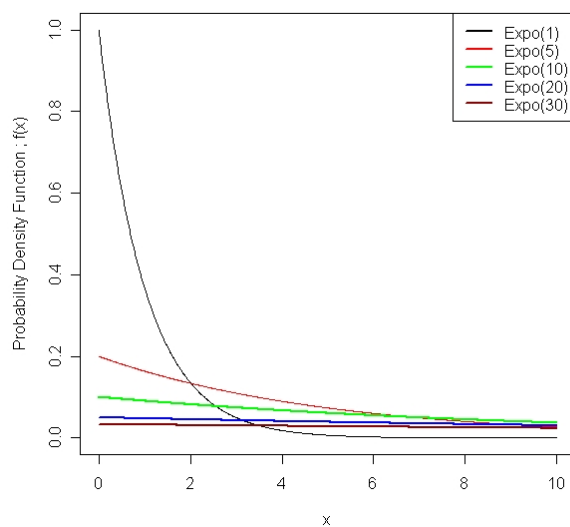
λ แทน ค่าเฉลี่ยของจำนวนเหตุการณ์ที่เกิดลักษณะที่สนใจ ในช่วงเวลาใดเวลาหนึ่งหรือขอบเขตใดขอบเขตหนึ่ง



ภาพที่1 แสดงฟังก์ชันมวลความน่าจะเป็นของการแจกแจงปัวซอง



ภาพที่2 แสดงฟังก์ชันมวลความน่าจะเป็นของการแจกแจงเรขาคณิต



ภาพที่3 แสดงฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงแบบเลขชี้กำลัง

ทฤษฎีการลู่เข้าสู่ส่วนกลาง (Central limit theorem)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันและมีการแจกแจงเดียวกัน ที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 ถ้าตัวอย่างมีขนาดใหญ่ ($n \rightarrow \infty$)

$$\text{จะได้ว่า} \quad Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

นั่นคือ Z_n ลู่เข้าสู่การแจกแจงปกติมาตรฐาน

ความสัมพันธ์ระหว่างการแจกแจงแกมมากับการแจกแจงไคกำลังสอง

$$\text{ถ้าตัวแปรสุ่ม } X \sim \text{Gamma}(\alpha, \beta) \text{ แล้วตัวแปรสุ่ม } Y = \frac{2X}{\beta} \sim \chi^2_{2\alpha}$$

(ดูพิสูจน์ในภาคผนวก)

ความสัมพันธ์ระหว่างการแจกแจงเอฟกับการแจกแจงบีต้า

$$\text{ถ้าตัวแปรสุ่ม } X \sim F(m, n) \text{ แล้วตัวแปรสุ่ม } Y = \frac{n}{n + mX} \sim \text{Beta}\left(\frac{n}{2}, \frac{m}{2}\right)$$

(ดูพิสูจน์ในภาคผนวก)

การจำลองข้อมูลด้วยวิธีบูตสเตรป

ในการอนุมานเชิงสถิติแบบเดิมนั้นจะเปรียบเทียบข้อมูลที่ได้จากการสังเกตกับทฤษฎีการแจกแจงของตัวสถิติ การสุ่มตัวอย่างซ้ำ (Resampling) เป็นการปฏิวัติวิธีการทางสถิติใหม่ โดยที่วิธีการสุ่มตัวอย่างซ้ำเป็นวิธีการที่ปราศจากทฤษฎีการแจกแจงของตัวสถิติ การอนุมานค่าสถิติของวิธีการสุ่มตัวอย่างซ้ำอาศัยการสุ่มตัวอย่างซ้ำจากตัวอย่างที่เราสังเกตได้หรือสร้างข้อมูลที่มีลักษณะที่เราสนใจแล้วสุ่มข้อมูลจากข้อมูลที่สร้างขึ้น คล้ายวิธีการจำลองแบบมอนติคาร์โล ที่สร้างข้อมูลและสรุปผลจากรูปแบบของข้อมูลที่เกิดขึ้นได้หลายลักษณะ แต่ในวิธีการสุ่มตัวอย่างซ้ำ จะสร้างรูปแบบของข้อมูลที่หลากหลายจากข้อมูลเพียงชุดเดียว โดยวิธีบูตสเตรป (Bootstrap method) เป็นรูปแบบหนึ่งของวิธีการสุ่มตัวอย่างซ้ำ โดยที่วิธีบูตสเตรปจะนำข้อมูลที่สังเกตได้นำมาสุ่มใหม่ โดยที่จะสุ่มข้อมูลแบบสุ่มแล้วใส่คืนตามจำนวนชุดที่ต้องการ แล้วนำข้อมูลที่ได้แต่ละชุดมาหาค่าความน่าจะเป็น หรือค่าสถิติที่ต้องการ แล้วนำค่าความน่าจะเป็นหรือค่าสถิติที่ได้ทั้งหมด มาหาค่าความน่าจะเป็นหรือค่าสถิติอีกที (จำลอง วงษ์ประเสริฐ, ย่อหน้าที่ 1-3) ดังนั้นหลักเกณฑ์ที่สำคัญของวิธีบูตสเตรป คือ พิจารณาตัวอย่างที่เกิดขึ้นรวบรวมมาจากประชากรเปรียบเสมือนว่าตัวอย่างเป็นประชากรใหม่ แล้วทำการสุ่มตัวอย่างซ้ำ โดยสุ่มแบบใส่คืนจากตัวแบบนี้ โดยทำซ้ำตามจำนวนเท่าที่ต้องการ (สุชาติ กิระนันท์, 2538 หน้า 356)

แนวคิดและขั้นตอนในการหาวิธีการประมาณค่าแบบช่วง

1. วิธีวาลด์ (Wald method)

การหาช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับพารามิเตอร์ θ โดยวิธีวาลด์ คือ การพิจารณาตัวประมาณค่าแบบจุดของ θ โดยวิธีความควรจะเป็นสูงสุด (Maximum likelihood method)

นั่นคือ ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความน่าจะเป็น $f(x; \theta)$ ฟังก์ชันควรจะเป็น (Likelihood function) ของ $X_1, X_2, \dots, X_n$ คือ

$$\begin{aligned} L(\theta) &= f(x_1, x_2, \dots, x_n; \theta) \\ &= \prod_{i=1}^n f(x_i; \theta) \end{aligned}$$

การหาตัวประมาณค่าแบบจุดของ θ ที่ทำให้ฟังก์ชันควรจะเป็นมีค่าสูงสุดนั้น พิจารณาจากการหาอนุพันธ์อันดับที่ 1 ของ $\ln L(\theta)$ เทียบกับ θ และกำหนดให้มีค่าเท่ากับ 0 $\left(\frac{\partial}{\partial \theta} \ln L(\theta) = 0\right)$ เพื่อแก้สมการหาค่า θ ซึ่งจะได้ $\hat{\theta}$ เป็นตัวประมาณความควรจะเป็นสูงสุดของ θ

หลังจากได้ตัวประมาณค่าแบบจุด พิจารณาหาช่วงความเชื่อมั่นโดยใช้ทฤษฎีการลู่เข้าสู่ส่วนกลาง นั่นคือ

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{\text{Var}(\hat{\theta})}} \sim N(0,1)$$

$$\text{จะได้ว่า } P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

จากนั้นทำการแก้สมการเพื่อหาช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับพารามิเตอร์ θ

2. วิธีสกอร์ (Score method)

อเกรสตี และ คูลล์ (Agresti and Coull, 1998) ได้ศึกษาวิธีประมาณแบบสกอร์ โดยได้รับการพัฒนามาจาก วิลสัน (Wilson, 1927) ซึ่งช่วงความเชื่อมั่นที่สร้างขึ้น ได้จากการพิจารณาตัวสถิติที่ใช้ในการทดสอบสมมติฐานแบบสองทาง จะได้ว่าสมมติฐานที่ใช้ในการทดสอบคือ $H_0: \theta = \theta_0$ เทียบกับ $H_1: \theta \neq \theta_0$ ที่ระดับนัยสำคัญ α โดยมีพื้นฐานจากทฤษฎีการลู่เข้าสู่ส่วนกลาง นั่นคือ จะยอมรับสมมติฐานหลัก H_0 เมื่อ

$$\frac{|\hat{\theta} - \theta_0|}{\sqrt{\text{Var}(\hat{\theta})}} \leq Z_{\frac{\alpha}{2}} \quad \text{โดย } \hat{\theta} \text{ คือ ตัวประมาณของ } \theta \text{ โดยวิธี MLE}$$

จากนั้นทำการแก้สมการเพื่อหาช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับพารามิเตอร์ θ

3. วิธีปรับค่าของวาลด์ (Adjusted Wald method)

อเกรสตี และ คูลล์ (Agresti and Coull, 1998) ได้พิจารณาหาช่วงความเชื่อมั่นแบบปรับค่าของวาลด์ โดยมีแนวคิดจากการนำค่าจุดศูนย์กลางของช่วงความเชื่อมั่นของ θ ที่ได้จากวิธีสคอร์ ซึ่งค่าจุดศูนย์กลางที่ได้นั้น กำหนดให้เป็น $\tilde{\theta}$ แล้วนำค่า $\tilde{\theta}$ แทนลงในการประมาณค่าแบบช่วงของวิธีวาลด์

4. วิธีเบส์ (Bayes' method)

โดยทั่วไป ค่าพารามิเตอร์ θ จะถือว่าเป็นค่าคงที่ที่ไม่ทราบค่า แต่แนวคิดของเบส์จะถือว่าค่าพารามิเตอร์ θ เป็นค่าของตัวแปรสุ่ม Θ ที่มีการแจกแจงความน่าจะเป็นในรูปแบบใดรูปแบบหนึ่ง ซึ่งเรียกว่าการแจกแจงก่อน (Prior distribution) เมื่อมีการรวบรวมข้อมูล จะใช้ความรู้เดิมเกี่ยวกับ θ เพื่อนำไปใช้ประโยชน์ในการประมาณค่า θ ให้แม่นยำยิ่งขึ้น ซึ่งการแจกแจงที่ได้จากการพิจารณาความรู้เดิมกับข้อมูลที่ได้จากการเก็บรวบรวมนั้นจะเรียกว่าการแจกแจงภายหลัง (Posterior distribution)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความน่าจะเป็น $f(x|\theta)$ โดยที่ θ เป็นค่าของตัวแปรสุ่ม Θ ที่มีการแจกแจงรูปแบบหนึ่ง ดังนั้นสามารถหาฟังก์ชันความหนาแน่นของความน่าจะเป็นภายหลังของ Θ เมื่อกำหนด $X_1 = x_1, \dots, X_n = x_n$ โดยพิจารณา ดังนี้

ให้ $\pi(\theta)$ เป็น ฟังก์ชันความหนาแน่นของความน่าจะเป็นก่อนของ θ

ฟังก์ชันความน่าจะเป็นร่วมมีเงื่อนไขของ $X_1, X_2, \dots, X_n$ เมื่อกำหนด $\Theta = \theta$ คือ

$$f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta)$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมของ $X_1, X_2, \dots, X_n$ และ Θ คือ

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1, x_2, \dots, x_n | \theta) \pi(\theta)$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมตามขอบของ $X_1, X_2, \dots, X_n$ คือ

$$g(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n; \theta) d\theta$$

ดังนั้น ฟังก์ชันความหนาแน่นของความน่าจะเป็นภายหลังของ Θ

เมื่อกำหนด $X_1 = x_1, \dots, X_n = x_n$ คือ

$$h(\theta | x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n; \theta)}{g(x_1, x_2, \dots, x_n)}$$

เพราะฉะนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับพารามิเตอร์ θ โดยวิธีเบส์ สามารถหาได้จากการพิจารณาความสัมพันธ์ของการแจกแจง หรือพิจารณาจากนิยามที่ว่า

$$\begin{aligned} P(\theta_L < \theta < \theta_U) &= 1 - \alpha \\ &= \int_{\theta_L}^{\theta_U} h(\theta | x_1, x_2, \dots, x_n) d\theta \end{aligned}$$

โดยที่ θ_U และ θ_L เป็นค่าฟังก์ชันของค่าสังเกต $x_1, x_2, \dots, x_n$

5. วิธีบูตสเตรปสคอร์ (Bootstrap Score method)

มีขั้นตอนดังนี้

1. สุ่มตัวอย่างขนาด n แบบใส่คืน $x = (x_1, x_2, \dots, x_n)$
2. คำนวณสูตรการประมาณค่าแบบช่วงสำหรับ θ ด้วยวิธีสคอร์ โดยใช้ข้อมูลตัวอย่างที่ได้จากการสุ่มข้อ 1 ซึ่งจะได้ค่าประมาณขีดจำกัดล่าง (θ_{LBS}^*) และขีดจำกัดบน (θ_{UBS}^*) ของช่วงความเชื่อมั่นในการประมาณค่า θ
3. ทำซ้ำใน ข้อ 1 และ 2 จำนวน B รอบ แล้วหาค่าเฉลี่ยของขีดจำกัดล่างและขีดจำกัดบนที่ได้จากการสุ่มตัวอย่างโดยวิธีบูตสเตรปทั้ง B รอบ นั่นคือ

$$\theta_{LBS} = \frac{\sum_{i=1}^B \theta_{LBS}^*}{B}$$

$$\theta_{UBS} = \frac{\sum_{i=1}^B \theta_{UBS}^*}{B}$$

โดยที่ θ_{LBS}^* แทน ขีดจำกัดล่างของช่วงความเชื่อมั่นที่ได้จากวิธีสคอร์
จากการสุ่มซ้ำด้วยวิธีบูตสเตรปในแต่ละรอบ
 θ_{UBS}^* แทน ขีดจำกัดบนของช่วงความเชื่อมั่นที่ได้จากวิธีสคอร์
จากการสุ่มซ้ำด้วยวิธีบูตสเตรปในแต่ละรอบ

ดังนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ θ โดยวิธีบูตสเตรปสคอร์ คือ $(\theta_{LBS}, \theta_{UBS})$

6. วิธีบูตสแตรปเบส (Bootstrap Bayes' method)

มีขั้นตอนดังนี้

1. สุ่มตัวอย่างขนาด n แบบใส่คืน $x = (x_1, x_2, \dots, x_n)$
2. คำนวณสูตรการประมาณค่าแบบช่วงสำหรับ θ ด้วยวิธีเบส โดยใช้ข้อมูลตัวอย่างที่ได้จากการสุ่มข้อ 1 ซึ่งจะได้ค่าประมาณขีดจำกัดล่าง (θ_{LBB}^*) และขีดจำกัดบน (θ_{UBB}^*) ของช่วงความเชื่อมั่นในการประมาณค่า θ
3. ทำซ้ำใน ข้อ 1 และ 2 จำนวน B รอบ แล้วหาค่าเฉลี่ยของขีดจำกัดล่างและขีดจำกัดบนที่ได้จากการสุ่มตัวอย่างโดยวิธีบูตสแตรปทั้ง B รอบ นั่นคือ

$$\theta_{LBB} = \frac{\sum_{i=1}^B \theta_{LBB}^*}{B}$$

$$\theta_{UBB} = \frac{\sum_{i=1}^B \theta_{UBB}^*}{B}$$

โดยที่ θ_{LBB}^* แทน ขีดจำกัดล่างของช่วงความเชื่อมั่นที่ได้จากวิธีเบส
จากการสุ่มซ้ำด้วยวิธีบูตสแตรปในแต่ละรอบ
 θ_{UBB}^* แทน ขีดจำกัดบนของช่วงความเชื่อมั่นที่ได้จากวิธีเบส
จากการสุ่มซ้ำด้วยวิธีบูตสแตรปในแต่ละรอบ

ดังนั้น ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ θ โดยวิธีบูตสแตรปเบส คือ $(\theta_{LBB}, \theta_{UBB})$

วิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากร เมื่อตัวแปรสุ่มมีการแจกแจงปัวซอง

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปัวซอง โดยมีพารามิเตอร์ λ

พิจารณาการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากรจากการแจกแจงปัวซอง

1. วิธีวาลด์ (Wald method)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปัวซอง โดยมีพารามิเตอร์ λ มีฟังก์ชันมวลความน่าจะเป็นดังนี้

$$f(x_i; \lambda) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \quad , \quad x_i = 0, 1, 2, \dots \quad , \quad \lambda > 0 \quad , \quad i = 1, 2, \dots, n$$

หาตัวประมาณ λ โดยวิธีความควรจะเป็นสูงสุด (MLE)

จะได้ว่าตัวประมาณ λ โดยวิธี MLE คือ $\hat{\lambda} = \bar{X}$

$$\begin{aligned} \text{จะได้ว่า} \quad E(\hat{\lambda}) &= E(\bar{X}) \\ &= \lambda \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{\lambda}) &= \text{Var}(\bar{X}) \\ &= \frac{\lambda}{n} \end{aligned}$$

จากทฤษฎีการลู่เข้าสู่ส่วนกลาง เมื่อตัวอย่างมีขนาดใหญ่ ค่าเฉลี่ยตัวอย่าง $\hat{\lambda}$ จะมีการแจกแจงปกติที่มีค่าเฉลี่ย λ ความแปรปรวน $\frac{\lambda}{n}$

$$\text{นั่นคือ} \quad \hat{\lambda} \sim N\left(\lambda, \frac{\lambda}{n}\right)$$

$$\text{จะได้ว่า} \quad Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \sim N(0, 1)$$

เนื่องจากไม่ทราบค่า λ ดังนั้นจึงประมาณค่าความแปรปรวนของ $\hat{\lambda}$ ด้วย $\frac{\hat{\lambda}}{n}$

$$\text{นั่นคือ} \quad Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}} \sim N(0, 1)$$

$$\text{ดังนั้น} \quad P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\hat{\lambda} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}} < \lambda < \hat{\lambda} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}}\right) = 1 - \alpha$$

$\therefore$ ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ λ โดยวิธีวาลด์ คือ $\hat{\lambda} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}}$

2. วิธีสคออร์ (Score method)

พิจารณาการทดสอบสมมติฐาน

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda \neq \lambda_0$$

กำหนดระดับนัยสำคัญ α

ดังนั้น จะยอมรับสมมติฐานหลัก H_0 เมื่อ

$$\frac{|\hat{\lambda} - \lambda_0|}{\sqrt{\frac{\lambda_0}{n}}} \leq Z_{\frac{\alpha}{2}}$$

$$(\hat{\lambda} - \lambda_0)^2 \leq Z_{\frac{\alpha}{2}}^2 \frac{\lambda_0}{n}$$

ให้ $Z = Z_{\frac{\alpha}{2}}$ และ $\lambda = \lambda_0$

$$\hat{\lambda}^2 - 2\hat{\lambda}\lambda + \lambda^2 - Z^2 \frac{\lambda}{n} \leq 0$$

$$\lambda^2 - \left(2\hat{\lambda} + \frac{Z^2}{n}\right)\lambda + \hat{\lambda}^2 \leq 0$$

กำหนดให้สมการข้างต้นมีค่าเท่ากับ 0 นั่นคือ $\lambda^2 - \left(2\hat{\lambda} + \frac{Z^2}{n}\right)\lambda + \hat{\lambda}^2 = 0$

จะเห็นว่าสมการที่ได้อยู่ในรูปของสมการกำลังสอง คือ $ax^2 + bx + c = 0$

ดังนั้นสามารถแก้สมการหาค่า x ได้จากสูตร $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

โดยที่ $a = 1$, $b = -\left(2\hat{\lambda} + \frac{Z^2}{n}\right)$, $c = \hat{\lambda}^2$ และ $x = \lambda$

$$\begin{aligned}
\text{จะได้ว่า } \lambda &= \frac{\left(2\hat{\lambda} + \frac{Z^2}{n}\right) \pm \sqrt{\left(-\left(2\hat{\lambda} + \frac{Z^2}{n}\right)\right)^2 - 4\hat{\lambda}^2}}{2} \\
&= \frac{\left(2\hat{\lambda} + \frac{Z^2}{n}\right) \pm \sqrt{4\hat{\lambda}^2 + 4\hat{\lambda}\frac{Z^2}{n} + \frac{Z^4}{n^2} - 4\hat{\lambda}^2}}{2} \\
&= \frac{\left(2\hat{\lambda} + \frac{Z^2}{n}\right) \pm \sqrt{4\hat{\lambda}\frac{Z^2}{n} + \frac{Z^4}{n^2}}}{2} \\
&= \hat{\lambda} + \frac{Z^2}{2n} \pm Z\sqrt{\frac{\hat{\lambda}}{n} + \frac{Z^2}{4n^2}}
\end{aligned}$$

เนื่องจากสมการที่ต้องการหาค่า คือ $\lambda^2 - \left(2\hat{\lambda} + \frac{Z^2}{n}\right)\lambda + \hat{\lambda}^2 \leq 0$

ดังนั้น เราจะได้ว่า λ อยู่ในช่วง $\left(\hat{\lambda} + \frac{Z^2}{2n} - Z\sqrt{\frac{\hat{\lambda}}{n} + \frac{Z^2}{4n^2}}, \hat{\lambda} + \frac{Z^2}{2n} + Z\sqrt{\frac{\hat{\lambda}}{n} + \frac{Z^2}{4n^2}}\right)$

$\therefore$ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ λ โดยวิธีสคออร์ คือ $\hat{\lambda} + \frac{Z_{\alpha/2}^2}{2n} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}}{n} + \frac{Z_{\alpha/2}^2}{4n^2}}$

3. วิธีปรับค่าของวาลด์ (Adjusted Wald method)

หาช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ λ โดยวิธีปรับค่าของวาลด์ ได้จากพิจารณา
ค่าจุดศูนย์กลางของช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ λ โดยวิธีสคออร์

$$\begin{aligned}
\text{นั่นคือ } \hat{\lambda} + \frac{Z^2}{2n} &= \frac{\sum_{i=1}^n X_i}{n} + \frac{Z^2}{2n} \\
&= \frac{\sum_{i=1}^n X_i + \frac{Z^2}{2}}{n} \\
\text{กำหนดให้ } \tilde{\lambda} &= \frac{\sum_{i=1}^n X_i + \frac{Z_{\alpha/2}^2}{2}}{n}
\end{aligned}$$

นำค่า $\tilde{\lambda}$ จากข้างต้นแทนที่ $\hat{\lambda}$ ในช่วงความเชื่อมั่นของ λ โดยวิธีวาลด์

เนื่องจากช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ λ โดยวิธีวาลด์ คือ $\hat{\lambda} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{\lambda}}{n}}$

$\therefore$ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ λ โดยวิธีปรับค่าของวาลด์ คือ $\tilde{\lambda} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\tilde{\lambda}}{n}}$

4. วิธีเบย์ (Bayes' method)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปัวซอง (λ) โดยที่ λ เป็นค่าของตัวแปรสุ่ม Θ ที่มีการแจกแจงแกมมา (α_0, β_0) ทราบค่า α_0 และ β_0 ฟังก์ชันความน่าจะเป็นร่วมมีเงื่อนไขของ $X_1, X_2, \dots, X_n$ เมื่อกำหนด $\Theta = \lambda$ คือ

$$\begin{aligned} f(x_1, x_2, \dots, x_n | \lambda) &= \prod_{i=1}^n f(x_i | \lambda) \\ &= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i)!} \quad , \quad x_i = 0, 1, 2, \dots \quad , \quad i = 1, 2, \dots, n \\ &\quad , \quad \lambda > 0 \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นก่อนของ Θ คือ

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha_0) \beta_0^{\alpha_0}} \lambda^{\alpha_0-1} e^{-\lambda/\beta_0} \quad , \quad \alpha_0, \beta_0 > 0 \quad , \quad \lambda > 0$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมของ $X_1, X_2, \dots, X_n$ และ Θ คือ

$$\begin{aligned} f(x_1, x_2, \dots, x_n; \lambda) &= f(x_1, x_2, \dots, x_n | \lambda) \pi(\lambda) \\ &= \frac{e^{-\left(\frac{n\beta_0+1}{\beta_0}\right)\lambda} \lambda^{\sum_{i=1}^n x_i + \alpha_0 - 1}}{\prod_{i=1}^n (x_i)! \Gamma(\alpha_0) \beta_0^{\alpha_0}} \quad , \quad \alpha_0, \beta_0 > 0 \quad , \quad \lambda > 0 \\ &\quad , \quad x_i = 0, 1, 2, \dots \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมตามขอบของ $X_1, X_2, \dots, X_n$ คือ

$$\begin{aligned} g(x_1, x_2, \dots, x_n) &= \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n; \lambda) d\lambda \\ &= \frac{\Gamma\left(\sum_{i=1}^n x_i + \alpha_0\right)}{\left(\prod_{i=1}^n (x_i)!\right) \Gamma(\alpha_0) \beta_0^{\alpha_0}} \left[\frac{\beta_0}{n\beta_0 + 1} \right]^{\sum_{i=1}^n x_i + \alpha_0} \quad , \quad \alpha_0, \beta_0 > 0 \quad , \quad x_i = 0, 1, 2, \dots \\ &\quad , \quad i = 1, 2, \dots, n \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นภายหลังของ Θ เมื่อกำหนด $X_1 = x_1, \dots, X_n = x_n$

$$\begin{aligned}
 h(\lambda \mid x_1, x_2, \dots, x_n) &= \frac{f(x_1, x_2, \dots, x_n; \lambda)}{g(x_1, x_2, \dots, x_n)} \\
 &= \frac{1}{\Gamma\left(\sum_{i=1}^n x_i + \alpha_0\right) \left[\frac{\beta_0}{n\beta_0 + 1}\right]^{\sum_{i=1}^n x_i + \alpha_0}} \lambda^{\sum_{i=1}^n x_i + \alpha_0 - 1} e^{-\left(\frac{n\beta_0 + 1}{\beta_0}\right)\lambda}, \quad \lambda > 0 \\
 &\quad , \quad \alpha_0, \beta_0 > 0, x_i = 0, 1, 2, \dots
 \end{aligned}$$

$$\therefore \Theta = \lambda \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \sim \text{Gamma}\left(\sum_{i=1}^n x_i + \alpha_0, \frac{\beta_0}{n\beta_0 + 1}\right)$$

จากความสัมพันธ์ระหว่างการแจกแจงแกมมากับการแจกแจงไคกำลังสอง

$$\text{ถ้าตัวแปรสุ่ม } X \sim \text{Gamma}(\alpha, \beta) \text{ แล้วตัวแปรสุ่ม } Y = \frac{2X}{\beta} \sim \chi_{2\alpha}^2$$

$$\text{เนื่องจาก } \Theta = \lambda \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \sim \text{Gamma}\left(\sum_{i=1}^n x_i + \alpha_0, \frac{\beta_0}{n\beta_0 + 1}\right)$$

$$\text{จะได้ว่า} \quad \frac{\frac{2\lambda}{\beta_0}}{n\beta_0 + 1} \sim \chi_{2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2$$

นั่นคือ

$$P\left(\chi_{\frac{\alpha}{2}, 2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2 < \frac{2\lambda}{\frac{\beta_0}{n\beta_0 + 1}} < \chi_{1-\frac{\alpha}{2}, 2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2\right) = 1 - \alpha$$

$$P\left(\frac{\beta_0}{2(n\beta_0 + 1)} \chi_{\frac{\alpha}{2}, 2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2 < \lambda < \frac{\beta_0}{2(n\beta_0 + 1)} \chi_{1-\frac{\alpha}{2}, 2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2\right) = 1 - \alpha$$

ดังนั้น ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ λ โดยวิธีเบส์ คือ

$$\left(\frac{\beta_0}{2(n\beta_0 + 1)} \chi_{\frac{\alpha}{2}, 2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2, \frac{\beta_0}{2(n\beta_0 + 1)} \chi_{1-\frac{\alpha}{2}, 2\left(\sum_{i=1}^n x_i + \alpha_0\right)}^2\right)$$

วิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากร เมื่อตัวแปรสุ่มมีการแจกแจงเรขาคณิต

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงเรขาคณิต โดยมีพารามิเตอร์ p

พิจารณาการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากรจากการแจกแจงเรขาคณิต

1. วิธีวาลด์ (Wald method)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงเรขาคณิต โดยมีพารามิเตอร์ p มีฟังก์ชันมวลความน่าจะเป็นดังนี้

$$f(x_i; p) = p(1-p)^{x_i-1}, \quad x_i = 1, 2, 3, \dots, \quad 0 < p < 1, \quad i = 1, 2, \dots, n$$

หาตัวประมาณ $\frac{1}{p}$ โดยวิธีความควรจะเป็นสูงสุด (MLE)

พิจารณาจากตัวประมาณ p

$$\text{จะได้ว่า ตัวประมาณ } p \text{ โดยวิธี MLE คือ } \hat{p} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}}$$

$$\text{นั่นคือ ตัวประมาณ } \frac{1}{p} \text{ โดยวิธี MLE คือ } \frac{1}{\hat{p}} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

$$\begin{aligned} \text{จะได้ว่า} \quad E\left(\frac{1}{\hat{p}}\right) &= E(\bar{X}) \\ &= \frac{1}{p} \\ \text{Var}\left(\frac{1}{\hat{p}}\right) &= \text{Var}(\bar{X}) \\ &= \frac{(1-p)}{np^2} \end{aligned}$$

จากทฤษฎีการลู่เข้าสู่ส่วนกลาง เมื่อตัวอย่างมีขนาดใหญ่ ค่าเฉลี่ยตัวอย่าง $\frac{1}{\hat{p}}$ จะมี

การแจกแจงปกติ ที่มีค่าเฉลี่ย $\frac{1}{p}$ ความแปรปรวน $\frac{(1-p)}{np^2}$

นั่นคือ
$$\frac{1}{\hat{p}} \sim N\left(\frac{1}{p}, \frac{(1-p)}{np^2}\right)$$

จะได้ว่า
$$Z = \frac{\frac{1}{\hat{p}} - \frac{1}{p}}{\sqrt{\frac{(1-p)}{np^2}}} \sim N(0,1)$$

เนื่องจากไม่ทราบค่า $\frac{1}{p}$ ดังนั้นจึงประมาณค่าความแปรปรวนของ $\frac{1}{\hat{p}}$ ด้วย $\frac{(1-\hat{p})}{n\hat{p}^2}$

นั่นคือ
$$Z = \frac{\frac{1}{\hat{p}} - \frac{1}{p}}{\sqrt{\frac{(1-\hat{p})}{n\hat{p}^2}}} \sim N(0,1)$$

ดังนั้น

$$P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

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

$$P\left(\frac{1}{\hat{p}} - Z_{\frac{\alpha}{2}} \sqrt{\frac{(1-\hat{p})}{n\hat{p}^2}} < \frac{1}{p} < \frac{1}{\hat{p}} + Z_{\frac{\alpha}{2}} \sqrt{\frac{(1-\hat{p})}{n\hat{p}^2}}\right) = 1 - \alpha$$

$\therefore$ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีวาลด์ คือ $\frac{1}{\hat{p}} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{(1-\hat{p})}{n\hat{p}^2}}$

2. วิธีสคออร์ (Score method)

พิจารณาการทดสอบสมมติฐาน

$$H_0 : \frac{1}{p} = \frac{1}{p_0}$$

$$H_1 : \frac{1}{p} \neq \frac{1}{p_0}$$

กำหนดระดับนัยสำคัญ α

ดังนั้น จะยอมรับสมมติฐานหลัก H_0 เมื่อ

$$\frac{\left| \frac{1}{\hat{p}} - \frac{1}{p_0} \right|}{\sqrt{\frac{(1-p_0)}{np_0^2}}} \leq Z_{\frac{\alpha}{2}}$$

$$\left(\frac{1}{\hat{p}} - \frac{1}{p_0} \right)^2 \leq Z_{\frac{\alpha}{2}}^2 \left(\frac{1-p_0}{np_0^2} \right)$$

ให้ $Z = Z_{\frac{\alpha}{2}}$ และ $p = p_0$

$$\frac{1}{\hat{p}^2} - \frac{2}{\hat{p}p} + \frac{1}{p^2} - \frac{Z^2}{np^2} + \frac{Z^2}{np} \leq 0$$

$$\left(1 - \frac{Z^2}{n} \right) \frac{1}{p^2} + \left(\frac{Z^2}{n} - \frac{2}{\hat{p}} \right) \frac{1}{p} + \frac{1}{\hat{p}^2} \leq 0$$

กำหนดให้สมการข้างต้นมีค่าเท่ากับ 0 นั่นคือ $\left(1 - \frac{Z^2}{n} \right) \frac{1}{p^2} + \left(\frac{Z^2}{n} - \frac{2}{\hat{p}} \right) \frac{1}{p} + \frac{1}{\hat{p}^2} = 0$

จะเห็นว่าสมการที่ได้อยู่ในรูปของสมการกำลังสอง คือ $ax^2 + bx + c = 0$

ดังนั้นสามารถแก้สมการหาค่า x ได้จากสูตร $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

โดยที่ $a = 1 - \frac{Z^2}{n}$, $b = \frac{Z^2}{n} - \frac{2}{\hat{p}}$, $c = \frac{1}{\hat{p}^2}$ และ $x = \frac{1}{p}$

$$\text{จะได้ว่า } \frac{1}{p} = \frac{-\left(\frac{Z^2}{n} - \frac{2}{\hat{p}} \right) \pm \sqrt{\left(\frac{Z^2}{n} - \frac{2}{\hat{p}} \right)^2 - 4\left(1 - \frac{Z^2}{n} \right) \frac{1}{\hat{p}^2}}}{2\left(1 - \frac{Z^2}{n} \right)}$$

$$\frac{1}{p} = \frac{\frac{2}{\hat{p}} - \frac{Z^2}{n} \pm \sqrt{\frac{Z^4}{n^2} - \frac{4Z^2}{n\hat{p}} + \frac{4}{\hat{p}^2} - \frac{4}{\hat{p}^2} + \frac{4Z^2}{n\hat{p}^2}}}{2\left(\frac{n - Z^2}{n} \right)}$$

$$= \frac{\frac{n}{\hat{p}} - \frac{Z^2}{2} \pm Z \sqrt{\frac{Z^2}{4} - \frac{n}{\hat{p}} \left(1 - \frac{1}{\hat{p}} \right)}}{n - Z^2}$$

เนื่องจากสมการที่ต้องการหาค่า คือ $\left(1 - \frac{Z^2}{n}\right) \frac{1}{p^2} + \left(\frac{Z^2}{n} - \frac{2}{\hat{p}}\right) \frac{1}{p} + \frac{1}{\hat{p}^2} \leq 0$

ดังนั้น เราจะได้ว่า $\frac{1}{p}$ อยู่ในช่วง $\left(\frac{\frac{n}{\hat{p}} - \frac{Z^2}{2} - Z \sqrt{\frac{Z^2}{4} - \frac{n}{\hat{p}} \left(1 - \frac{1}{\hat{p}}\right)}}{n - Z^2}, \frac{\frac{n}{\hat{p}} - \frac{Z^2}{2} + Z \sqrt{\frac{Z^2}{4} - \frac{n}{\hat{p}} \left(1 - \frac{1}{\hat{p}}\right)}}{n - Z^2} \right)$

$\therefore$ ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีสคอว์ คือ $\frac{\frac{n}{\hat{p}} - \frac{Z_{\alpha/2}^2}{2} \pm Z_{\alpha/2} \sqrt{\frac{Z_{\alpha/2}^2}{4} - \frac{n}{\hat{p}} \left(1 - \frac{1}{\hat{p}}\right)}}{n - Z_{\alpha/2}^2}$

3. วิธีปรับค่าของวาลด์ (Adjusted Wald method)

หาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีปรับค่าของวาลด์ ได้จาก

พิจารณาค่าจุดศูนย์กลางของช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีสคอว์

$$\text{นั่นคือ} \quad \frac{\frac{n}{\hat{p}} - \frac{Z^2}{2}}{n - Z^2} = \frac{n\bar{X} - \frac{Z^2}{2}}{n - Z^2}$$

$$= \frac{\sum_{i=1}^n X_i - \frac{Z^2}{2}}{n - Z^2}$$

$$\text{กำหนดให้} \quad \frac{1}{\tilde{p}} = \frac{\sum_{i=1}^n X_i - \frac{Z_{\alpha/2}^2}{2}}{n - Z_{\alpha/2}^2}$$

$$= \frac{\sum \tilde{X}_i}{\tilde{n}}$$

$$\text{นั่นคือ} \quad \frac{\sum \tilde{X}_i}{\tilde{n}} = \frac{\sum_{i=1}^n X_i - \frac{Z_{\alpha/2}^2}{2}}{n - Z_{\alpha/2}^2}$$

$$\tilde{n} = n - Z_{\alpha/2}^2$$

นำค่า $\frac{1}{\tilde{p}}$ จากข้างต้นแทนที่ $\frac{1}{\hat{p}}$ ในช่วงความเชื่อมั่นของ $\frac{1}{p}$ โดยวิธีวาลด์

เนื่องจากช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีวาลด์ คือ $\frac{1}{\hat{p}} \pm Z_{\alpha/2} \sqrt{\frac{(1 - \hat{p})}{n\hat{p}^2}}$

ดังนั้น ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีปรับค่าของวาลด์ คือ $\frac{1}{\tilde{p}} \pm Z_{\alpha/2} \sqrt{\frac{(1 - \tilde{p})}{\tilde{n}\tilde{p}^2}}$

4. วิธีเบส์ (Bayes' method)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงเรขาคณิต (p) โดยที่ p เป็นค่าของตัวแปรสุ่ม Θ ที่มีการแจกแจงบีต้า (α_0, β_0) ทราบค่า α_0 และ β_0 ฟังก์ชันความน่าจะเป็นร่วมมีเงื่อนไขของ $X_1, X_2, \dots, X_n$ เมื่อกำหนด $\Theta = p$ คือ

$$\begin{aligned} f(x_1, x_2, \dots, x_n | p) &= \prod_{i=1}^n f(x_i | p) \\ &= p^n (1-p)^{\sum_{i=1}^n x_i - n} \quad , \quad x_i = 1, 2, 3, \dots \quad , \quad i = 1, 2, \dots, n \\ &\quad , \quad 0 < p < 1 \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นก่อนของ Θ คือ

$$\pi(p) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0-1} (1-p)^{\beta_0-1} \quad , \quad \alpha_0, \beta_0 > 0, 0 < p < 1$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมของ $X_1, X_2, \dots, X_n$ และ Θ คือ

$$\begin{aligned} f(x_1, x_2, \dots, x_n; p) &= f(x_1, x_2, \dots, x_n | p) \pi(p) \\ &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} p^{\alpha_0+n-1} (1-p)^{\beta_0+\sum_{i=1}^n x_i - n-1} \quad , \quad 0 < p < 1 \\ &\quad , \quad \alpha_0, \beta_0 > 0, x_i = 1, 2, 3, \dots \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมตามขอบของ $X_1, X_2, \dots, X_n$ คือ

$$\begin{aligned} g(x_1, x_2, \dots, x_n) &= \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n; p) dp \\ &= \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \frac{\Gamma(\alpha_0 + n) \Gamma\left(\beta_0 + \sum_{i=1}^n x_i - n\right)}{\Gamma\left(\alpha_0 + \beta_0 + \sum_{i=1}^n x_i\right)} \quad , \quad \alpha_0, \beta_0 > 0 \\ &\quad , \quad x_i = 1, 2, 3, \dots \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นภายหลังของ Θ เมื่อกำหนด $X_1 = x_1, \dots, X_n = x_n$

$$\begin{aligned} h(p | x_1, x_2, \dots, x_n) &= \frac{f(x_1, x_2, \dots, x_n; p)}{g(x_1, x_2, \dots, x_n)} \\ &= \frac{\Gamma\left(\alpha_0 + \beta_0 + \sum_{i=1}^n x_i\right)}{\Gamma(\alpha_0 + n) \Gamma\left(\beta_0 + \sum_{i=1}^n x_i - n\right)} p^{\alpha_0 + n - 1} (1 - p)^{\beta_0 + \sum_{i=1}^n x_i - n - 1}, \quad 0 < p < 1 \\ &\quad, \alpha_0, \beta_0 > 0, \quad x_i = 1, 2, 3, \dots \end{aligned}$$

$$\therefore \Theta = p | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \sim \text{Beta}\left(\alpha_0 + n, \beta_0 + \sum_{i=1}^n x_i - n\right)$$

ความสัมพันธ์ระหว่างการแจกแจงเอพกับการแจกแจงเบต้า

$$\text{ถ้าตัวแปรสุ่ม } X \sim F(a, b) \text{ แล้วตัวแปรสุ่ม } Y = \frac{b}{b + aX} \sim \text{Beta}\left(\frac{b}{2}, \frac{a}{2}\right)$$

$$\text{เนื่องจาก } \Theta = p | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \sim \text{Beta}\left(\alpha_0 + n, \beta_0 + \sum_{i=1}^n x_i - n\right)$$

$$\text{จะได้ว่า } \frac{2(\alpha_0 + n)}{2(\alpha_0 + n) + 2(\beta_0 + \sum_{i=1}^n x_i - n)F_{a,b}} \sim \text{Beta}\left(\alpha_0 + n, \beta_0 + \sum_{i=1}^n x_i - n\right)$$

$$\text{โดยที่ } b = 2(\alpha_0 + n), \quad a = 2\left(\beta_0 + \sum_{i=1}^n x_i - n\right)$$

$\therefore$ ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ p โดยวิธีเบส์ คือ

$$\left(\frac{2(\alpha_0 + n)}{2(\alpha_0 + n) + 2(\beta_0 + \sum_{i=1}^n x_i - n)F_{1-\frac{\alpha}{2}, a, b}}, \frac{2(\alpha_0 + n)}{2(\alpha_0 + n) + 2(\beta_0 + \sum_{i=1}^n x_i - n)F_{\frac{\alpha}{2}, a, b}} \right)$$

ดังนั้น ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ $\frac{1}{p}$ โดยวิธีเบส์ คือ

$$\left(\frac{2(\alpha_0 + n) + 2(\beta_0 + \sum_{i=1}^n x_i - n)F_{\frac{\alpha}{2}, a, b}}{2(\alpha_0 + n)}, \frac{2(\alpha_0 + n) + 2(\beta_0 + \sum_{i=1}^n x_i - n)F_{1-\frac{\alpha}{2}, a, b}}{2(\alpha_0 + n)} \right)$$

วิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากร เมื่อตัวแปรสุ่มมีการแจกแจงแบบเลขชี้กำลัง

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบเลขชี้กำลัง โดยมีพารามิเตอร์ β พิจารณาการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยประชากรจากการแจกแจงแบบเลขชี้กำลัง

1. วิธีวาลด์ (Wald method)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบเลขชี้กำลัง โดยมีพารามิเตอร์ β มีฟังก์ชันความหนาแน่นความน่าจะเป็นดังนี้

$$f(x_i; \beta) = \frac{1}{\beta} e^{-\frac{x_i}{\beta}} \quad , \quad x_i > 0 \quad , \quad \beta > 0 \quad , \quad i = 1, 2, \dots, n$$

หาตัวประมาณ β โดยวิธีความควรจะเป็นสูงสุด (MLE)

จะได้ว่าตัวประมาณ β โดยวิธี MLE คือ $\hat{\beta} = \bar{X}$

จะได้ว่า $E(\hat{\beta}) = E(\bar{X})$

$$= \beta$$

$$Var(\hat{\beta}) = Var(\bar{X})$$

$$= \frac{\beta^2}{n}$$

จากทฤษฎีการลู่เข้าสู่ส่วนกลาง เมื่อตัวอย่างมีขนาดใหญ่ ค่าเฉลี่ยตัวอย่าง $\hat{\beta}$ จะมีการแจกแจงปกติ ที่มีค่าเฉลี่ย β ความแปรปรวน $\frac{\beta^2}{n}$

นั่นคือ
$$\hat{\beta} \sim N\left(\beta, \frac{\beta^2}{n}\right)$$

จะได้ว่า
$$Z = \frac{\hat{\beta} - \beta}{\beta / \sqrt{n}} \sim N(0, 1)$$

เนื่องจากไม่ทราบค่า β ดังนั้นจึงประมาณค่าความแปรปรวนของ $\hat{\beta}$ ด้วย $\frac{\hat{\beta}^2}{n}$

นั่นคือ
$$Z = \frac{\hat{\beta} - \beta}{\hat{\beta} / \sqrt{n}} \sim N(0, 1)$$

ดังนั้น
$$P\left(-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\hat{\beta} - \beta}{\frac{\hat{\beta}}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\hat{\beta} - Z_{\frac{\alpha}{2}} \frac{\hat{\beta}}{\sqrt{n}} < \beta < \hat{\beta} + Z_{\frac{\alpha}{2}} \frac{\hat{\beta}}{\sqrt{n}}\right) = 1 - \alpha$$

∴ ช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ β โดยวิธีไวลด์ คือ $\hat{\beta} \pm Z_{\frac{\alpha}{2}} \frac{\hat{\beta}}{\sqrt{n}}$

2. วิธีสกอร์ (Score method)

พิจารณาการทดสอบสมมติฐาน

$$H_0 : \beta = \beta_0$$

$$H_1 : \beta \neq \beta_0$$

กำหนดระดับนัยสำคัญ α

ดังนั้น จะยอมรับสมมติฐานหลัก H_0 เมื่อ

$$\begin{aligned} \frac{|\hat{\beta} - \beta_0|}{\sqrt{\frac{\beta_0^2}{n}}} &\leq Z_{\frac{\alpha}{2}} \\ (\hat{\beta} - \beta_0)^2 &\leq Z_{\frac{\alpha}{2}}^2 \frac{\beta_0^2}{n} \end{aligned}$$

ให้ $Z = Z_{\frac{\alpha}{2}}$ และ $\beta = \beta_0$

$$\hat{\beta}^2 - 2\hat{\beta}\beta + \beta^2 - Z^2 \frac{\beta^2}{n} \leq 0$$

$$\left(1 - \frac{Z^2}{n}\right)\beta^2 - 2\hat{\beta}\beta + \hat{\beta}^2 \leq 0$$

กำหนดให้สมการข้างต้นมีค่าเท่ากับ 0 นั่นคือ $\left(1 - \frac{Z^2}{n}\right)\beta^2 - 2\hat{\beta}\beta + \hat{\beta}^2 = 0$

จะเห็นว่าสมการที่ได้อยู่ในรูปของสมการกำลังสอง คือ $ax^2 + bx + c = 0$

ดังนั้นสามารถแก้สมการหาค่า x ได้จากสูตร $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

โดยที่ $a = \left(1 - \frac{Z^2}{n}\right)$, $b = -2\hat{\beta}$, $c = \hat{\beta}^2$ และ $x = \beta$

$$\begin{aligned}
\text{จะได้ว่า } \beta &= \frac{2\hat{\beta} \pm \sqrt{(-2\hat{\beta})^2 - 4(1 - Z^2/n)\hat{\beta}^2}}{2(1 - Z^2/n)} \\
&= \frac{2\hat{\beta} \pm \sqrt{4\hat{\beta}^2 - 4\hat{\beta}^2 + 4Z^2\hat{\beta}^2/n}}{2(1 - Z^2/n)} \\
&= \frac{2\hat{\beta} \pm \sqrt{4Z^2\hat{\beta}^2/n}}{2(1 - Z^2/n)} \\
&= \frac{n\hat{\beta} \pm Z\sqrt{n\hat{\beta}^2}}{n - Z^2}
\end{aligned}$$

$$\text{เนื่องจากสมการที่ต้องการทราบค่า คือ } \left(1 - \frac{Z^2}{n}\right)\beta^2 - 2\hat{\beta}\beta + \hat{\beta}^2 \leq 0$$

$$\text{ดังนั้น เราจะได้ว่า } \beta \text{ อยู่ในช่วง } \left(\frac{n\hat{\beta} - Z\sqrt{n\hat{\beta}^2}}{n - Z^2}, \frac{n\hat{\beta} + Z\sqrt{n\hat{\beta}^2}}{n - Z^2} \right)$$

$$\therefore \text{ ช่วงความเชื่อมั่น } (1 - \alpha)100\% \text{ ของ } \beta \text{ โดยวิธีสคอว์ คือ } \frac{n\hat{\beta} \pm Z_{\alpha/2}\sqrt{n\hat{\beta}^2}}{n - Z_{\alpha/2}^2}$$

3. วิธีปรับค่าของวาลด์ (Adjusted Wald method)

หาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ β โดยวิธีปรับค่าของวาลด์ ได้จาก

พิจารณาค่าจุดศูนย์กลางของช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของ β โดยวิธีสคอว์

$$\begin{aligned}
\text{นั่นคือ } \frac{n\hat{\beta}}{n - Z^2} &= \frac{n\bar{X}}{n - Z^2} \\
&= \frac{\sum_{i=1}^n X_i}{n - Z_{\alpha/2}^2} \\
\text{กำหนดให้ } \tilde{\beta} &= \frac{\sum_{i=1}^n X_i}{\tilde{n}} \\
\text{โดยที่ } \tilde{n} &= n - Z_{\alpha/2}^2
\end{aligned}$$

นำค่า $\tilde{\beta}$ จากข้างต้นแทนที่ $\hat{\beta}$ ในช่วงความเชื่อมั่นของ β โดยวิธีวาลด์

$$\text{เนื่องจากช่วงความเชื่อมั่น } (1 - \alpha)100\% \text{ ของ } \beta \text{ โดยวิธีวาลด์ คือ } \hat{\beta} \pm Z_{\frac{\alpha}{2}} \frac{\hat{\beta}}{\sqrt{n}}$$

$$\therefore \text{ ช่วงความเชื่อมั่น } (1 - \alpha)100\% \text{ ของ } \beta \text{ โดยวิธีปรับค่าของวาลด์ คือ } \tilde{\beta} \pm Z_{\frac{\alpha}{2}} \frac{\tilde{\beta}}{\sqrt{\tilde{n}}}$$

4. วิธีเบย์ (Bayes' method)

ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบเลขชี้กำลัง โดยที่ λ เป็นค่าของตัวแปรสุ่ม Θ ที่มีการแจกแจงแกมมา (α_0, β_0) ทราบค่า α_0 และ β_0

ฟังก์ชันความน่าจะเป็นร่วมมีเงื่อนไขของ $X_1, X_2, \dots, X_n$ เมื่อกำหนด $\Theta = \lambda$ คือ

$$\begin{aligned} f(x_1, x_2, \dots, x_n | \lambda) &= \prod_{i=1}^n f(x_i | \lambda) \\ &= \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \quad , \quad x_i > 0 \quad , \quad i = 1, 2, \dots, n \\ &\quad , \quad \lambda > 0 \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นก่อนของ Θ คือ

$$\begin{aligned} \pi(\lambda) &= \frac{1}{\Gamma(\alpha_0) \beta_0^{\alpha_0}} \lambda^{\alpha_0-1} e^{-\lambda/\beta_0} \quad , \quad \alpha_0, \beta_0 > 0 \\ &\quad , \quad \lambda > 0 \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมของ $X_1, X_2, \dots, X_n$ และ Θ คือ

$$\begin{aligned} f(x_1, x_2, \dots, x_n; \lambda) &= f(x_1, x_2, \dots, x_n | \lambda) \pi(\lambda) \\ &= \frac{1}{\Gamma(\alpha_0) \beta_0^{\alpha_0}} \lambda^{n+\alpha_0-1} e^{-\lambda \left(\frac{\beta_0 \sum_{i=1}^n x_i + 1}{\beta_0} \right)} \quad , \quad \alpha_0, \beta_0 > 0 \\ &\quad , \quad \lambda > 0 \quad , \quad x_i > 0 \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นร่วมตามขอบของ $X_1, X_2, \dots, X_n$ คือ

$$\begin{aligned} g(x_1, x_2, \dots, x_n) &= \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n; \lambda) d\lambda \\ &= \frac{\Gamma(n + \alpha_0)}{\Gamma(\alpha_0) \beta_0^{\alpha_0}} \left(\frac{\beta_0}{\beta_0 \sum_{i=1}^n x_i + 1} \right)^{n+\alpha_0} \quad , \quad \alpha_0, \beta_0 > 0 \\ &\quad , \quad x_i > 0. \end{aligned}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็นภายหลังของ Θ เมื่อกำหนด $X_1 = x_1, \dots, X_n = x_n$

$$h(\lambda | x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n; \lambda)}{g(x_1, x_2, \dots, x_n)}$$

$$= \frac{1}{\Gamma(n + \alpha_0)} \left(\frac{\beta_0 \sum_{i=1}^n x_i + 1}{\beta_0} \right)^{n + \alpha_0} \lambda^{n + \alpha_0 - 1} e^{-\lambda \left(\frac{\beta_0 \sum_{i=1}^n x_i + 1}{\beta_0} \right)}, \quad \lambda > 0$$

$$, \quad \alpha_0, \beta_0 > 0, \quad x_i > 0.$$

$$\therefore \Theta = \lambda | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \sim \text{Gamma} \left(n + \alpha_0, \frac{\beta_0}{\beta_0 \sum_{i=1}^n x_i + 1} \right)$$

จากความสัมพันธ์ระหว่างการแจกแจงแกมมากับการแจกแจงไคกำลังสอง

ถ้าตัวแปรสุ่ม $X \sim \text{Gamma}(\alpha, \beta)$ แล้วตัวแปรสุ่ม $Y = \frac{2X}{\beta} \sim \chi_{2\alpha}^2$

$$\text{เนื่องจาก } \Theta = \lambda | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \sim \text{Gamma} \left(n + \alpha_0, \frac{\beta_0}{\beta_0 \sum_{i=1}^n x_i + 1} \right)$$

$$\text{จะได้ว่า } \frac{2\lambda}{\beta_0 \sum_{i=1}^n x_i + 1} \sim \chi_{2(n + \alpha_0)}^2$$

$$\text{นั่นคือ } P \left(\chi_{\frac{\alpha}{2}, 2(n + \alpha_0)}^2 < \frac{2\lambda}{\beta_0 \sum_{i=1}^n x_i + 1} < \chi_{1 - \frac{\alpha}{2}, 2(n + \alpha_0)}^2 \right) = 1 - \alpha$$

$$P \left(\frac{\beta_0}{2 \left(\beta_0 \sum_{i=1}^n x_i + 1 \right)} \chi_{\frac{\alpha}{2}, 2(n + \alpha_0)}^2 < \lambda < \frac{\beta_0}{2 \left(\beta_0 \sum_{i=1}^n x_i + 1 \right)} \chi_{1 - \frac{\alpha}{2}, 2(n + \alpha_0)}^2 \right) = 1 - \alpha$$

ดังนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ λ โดยวิธีเบส์ คือ

$$\left(\frac{\beta_0}{2\left(\beta_0 \sum_{i=1}^n x_i + 1\right)} \chi_{\frac{\alpha}{2}, 2(n+\alpha_0)}^2, \frac{\beta_0}{2\left(\beta_0 \sum_{i=1}^n x_i + 1\right)} \chi_{1-\frac{\alpha}{2}, 2(n+\alpha_0)}^2 \right)$$

ดังนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ β โดยวิธีเบส์ คือ

$$\left(\frac{2\left(\beta_0 \sum_{i=1}^n x_i + 1\right)}{\beta_0 \chi_{1-\frac{\alpha}{2}, 2(n+\alpha_0)}^2}, \frac{2\left(\beta_0 \sum_{i=1}^n x_i + 1\right)}{\beta_0 \chi_{\frac{\alpha}{2}, 2(n+\alpha_0)}^2} \right)$$

5. วิธีบูตสเตรปสคอว์ (Bootstrap score method)

1. สุ่มตัวอย่างขนาด n แบบใส่คืน จากตัวอย่างที่มีการแจกแจงแบบเลขชี้กำลัง $x = (x_1, x_2, \dots, x_n)$ เพื่อกำหนดหา $\hat{\beta} = \bar{X}$
2. หาค่าประมาณขีดจำกัดล่าง (β_{LBS}^*) และขีดจำกัดบน (β_{UBS}^*) ของช่วงความเชื่อมั่นในการประมาณค่า β จะได้ว่า

$$\beta_{LBS}^* = \frac{n\hat{\beta} - Z_{\alpha/2} \sqrt{n\hat{\beta}^2}}{n - Z_{\alpha/2}^2}$$

$$\beta_{UBS}^* = \frac{n\hat{\beta} + Z_{\alpha/2} \sqrt{n\hat{\beta}^2}}{n - Z_{\alpha/2}^2}$$

3. ทำซ้ำใน ข้อ 1 และ 2 จำนวน $B = 200$ รอบ แล้วหาค่าเฉลี่ยของขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่น จะได้ว่า

$$\beta_{LBS} = \frac{\sum_{i=1}^B \beta_{LBS}^*}{B}$$

$$\beta_{UBS} = \frac{\sum_{i=1}^B \beta_{UBS}^*}{B}$$

ดังนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ β โดยวิธีบูตสเตรปสคอว์ คือ $(\beta_{LBS}, \beta_{UBS})$

6. วิธีบูตสเตรปเบส์ (Bootstrap Bayes' method)

1. สุ่มตัวอย่างขนาด n แบบใส่คืน จากตัวอย่างที่มีการแจกแจงแบบเลขชี้กำลัง $x = (x_1, x_2, \dots, x_n)$ เพื่อคำนวณหาค่าองศาอิสระเสรี (Degree of freedom)
2. หาค่าประมาณขีดจำกัดล่าง (β_{LBB}^*) และขีดจำกัดบน (β_{UBB}^*) ของช่วงความเชื่อมั่นในการประมาณค่า β จะได้ว่า

$$\beta_{LBB}^* = \frac{2 \left(\beta_0 \sum_{i=1}^n x_i + 1 \right)}{\beta_0 \chi_{1-\frac{\alpha}{2}, 2(n+\alpha_0)}^2}$$

$$\beta_{UBB}^* = \frac{2 \left(\beta_0 \sum_{i=1}^n x_i + 1 \right)}{\beta_0 \chi_{\frac{\alpha}{2}, 2(n+\alpha_0)}^2}$$

3. ทำซ้ำใน ข้อ 1 และ 2 จำนวน $B = 200$ รอบ แล้วหาค่าเฉลี่ยของขีดจำกัดล่างและขีดจำกัดบนของช่วงความเชื่อมั่น จะได้ว่า

$$\beta_{LBB} = \frac{\sum_{i=1}^B \beta_{LBB}^*}{B}$$

$$\beta_{UBB} = \frac{\sum_{i=1}^B \beta_{UBB}^*}{B}$$

ดังนั้น ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ β โดยวิธีบูตสเตรปเบส์ คือ $(\beta_{LBB}, \beta_{UBB})$

การเปรียบเทียบค่าความน่าจะเป็นครอบคลุมและค่าเฉลี่ยความกว้างของช่วงความเชื่อมั่น

เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยประชากร ศึกษาจากค่าความน่าจะเป็นครอบคลุมว่าวิธีใดบ้างที่ให้ค่าความน่าจะเป็นครอบคลุมไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด พิจารณาจากการทดสอบสมมติฐาน โดยใช้ตัวสถิติทดสอบ Z ดังนี้

H_0 : ค่าความน่าจะเป็นครอบคลุมมีค่าไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ($c \geq c_0$)

H_1 : ค่าความน่าจะเป็นครอบคลุมมีค่าต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ($c < c_0$)

กำหนดระดับนัยสำคัญ α

ตัวสถิติที่ใช้ในการทดสอบ คือ
$$Z = \frac{\hat{c} - c_0}{\sqrt{\frac{c_0(1-c_0)}{M}}}$$

ขอบเขตวิกฤติ คือ $Z < -Z_\alpha$

ดังนั้น จะปฏิเสธสมมติฐานหลัก H_0 เมื่อ $Z < -Z_\alpha$

นั่นคือ จะยอมรับสมมติฐานหลัก H_0 เมื่อ $Z \geq -Z_\alpha$

$$\begin{aligned} \text{จะได้ว่า } \frac{\hat{c} - c_0}{\sqrt{\frac{c_0(1-c_0)}{M}}} &\geq -Z_\alpha \\ \hat{c} &\geq c_0 - Z_\alpha \sqrt{\frac{c_0(1-c_0)}{M}} \end{aligned}$$

เพราะฉะนั้น จะยอมรับสมมติฐานหลัก H_0 เมื่อ $\hat{c} \geq c_0 - Z_\alpha \sqrt{\frac{c_0(1-c_0)}{M}}$

เมื่อ c คือ ค่าความน่าจะเป็นครอบคลุม

$\hat{c}$ คือ ค่าความน่าจะเป็นครอบคลุมที่ได้จากการทดลอง

c_0 คือ ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

α คือ ระดับนัยสำคัญ หรือความน่าจะเป็นที่จะเกิดความผิดพลาดแบบที่ 1

M คือ จำนวนครั้งที่ทำการทดลองซ้ำ

พิจารณาการสร้างเกณฑ์ที่ใช้ในการเปรียบเทียบค่าความน่าจะเป็นครอบคลุม
ที่ระดับความเชื่อมั่น 90%

สมมติฐานการทดสอบ คือ

$$H_0: c \geq 0.90$$

$$H_1: c < 0.90$$

กำหนดระดับนัยสำคัญ $\alpha = 0.10$

นั่นคือ วิธีการประมาณค่าให้ค่าความน่าจะเป็นครอบคลุมมีค่าไม่ต่ำกว่าสัมประสิทธิ์

$$\begin{aligned} \text{ความเชื่อมั่นที่กำหนด} \quad \text{ถ้า} \quad \hat{c} &\geq 0.90 - 1.282 \sqrt{\frac{0.90(1-0.90)}{1000}} \\ &= 0.8878 \end{aligned}$$

พิจารณาการสร้างเกณฑ์ที่ใช้ในการเปรียบเทียบค่าความน่าจะเป็นครอบคลุม
ที่ระดับความเชื่อมั่น 95%

สมมติฐานการทดสอบ คือ

$$H_0: c \geq 0.95$$

$$H_1: c < 0.95$$

กำหนดระดับนัยสำคัญ $\alpha = 0.05$

นั่นคือ วิธีการประมาณค่าให้ค่าความน่าจะเป็นครอบคลุมมีค่าไม่ต่ำกว่าสัมประสิทธิ์

$$\begin{aligned} \text{ความเชื่อมั่นที่กำหนด} \quad \text{ถ้า} \quad \hat{c} &\geq 0.95 - 1.645 \sqrt{\frac{0.95(1-0.95)}{1000}} \\ &= 0.9387 \end{aligned}$$

พิจารณาการสร้างเกณฑ์ที่ใช้ในการเปรียบเทียบค่าความน่าจะเป็นครอบคลุม
ที่ระดับความเชื่อมั่น 99%

สมมติฐานการทดสอบ คือ

$$H_0: c \geq 0.99$$

$$H_1: c < 0.99$$

กำหนดระดับนัยสำคัญ $\alpha = 0.01$

นั่นคือ วิธีการประมาณค่าให้ค่าความน่าจะเป็นครอบคลุมมีค่าไม่ต่ำกว่าสัมประสิทธิ์

$$\begin{aligned} \text{ความเชื่อมั่นที่กำหนด ถ้า } \hat{c} &\geq 0.99 - 2.326 \sqrt{\frac{0.99(1-0.99)}{1000}} \\ &= 0.9827 \end{aligned}$$

ดังนั้น ถ้าวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยประชากรวิธีใด ให้ค่าความน่าจะเป็นครอบคลุมที่ได้จากการทดลองมีค่าไม่ต่ำกว่า 0.8878, 0.9387 และ 0.9827 ที่ระดับความเชื่อมั่น 90%, 95% และ 99% ตามลำดับ จะถือว่าวิธีนั้น ให้ค่าความน่าจะเป็นครอบคลุมไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด จากนั้นนำวิธีที่ให้ค่าความน่าจะเป็นครอบคลุมไม่ต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ไปทำการเปรียบเทียบค่าเฉลี่ยความกว้างของช่วงความเชื่อมั่นวิธีใดที่ให้ค่าเฉลี่ยความกว้างของช่วงความเชื่อมั่นแคบที่สุด จะถือว่าวิธีนั้นเป็นวิธีประมาณแบบช่วงสำหรับค่าเฉลี่ยประชากรที่ดีที่สุดสำหรับสถานการณ์นั้น

งานวิจัยที่เกี่ยวข้อง

อเกรสตี และ คูลล์ (Agresti and Coull, 1998) ศึกษาวิธีการประมาณค่าแบบช่วงสำหรับค่าสัดส่วนประชากรในการแจกแจงทวินาม ระหว่างวิธีวาลด์ วิธีสคอร์และวิธีปรับค่าของวาลด์ ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่าง 5 และ 10 พบว่าวิธีสคอร์และวิธีปรับค่าของวาลด์ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด และมีค่าสูงกว่าค่าสัมประสิทธิ์ความเชื่อมั่นโดยวิธีวาลด์ ในทุกขนาดของตัวอย่างที่ทำการศึกษา

เอลเฟสซี และ ไรเนเก (Elfessi and Reineke, 2001) ศึกษาการประมาณค่าพารามิเตอร์ ในการแจกแจงแบบเลขชี้กำลัง ระหว่างวิธีตัวประมาณแบบเดิม (Classical estimator) กับตัวประมาณแบบเบย์ (Bayes' estimator) โดยตัวประมาณแบบเดิมที่ใช้ในการพิจารณามี 3 วิธี คือ ตัวประมาณความควรจะเป็นสูงสุด (MLE) ตัวประมาณไม่เอนเอียงและมีความแปรปรวนต่ำสุดสม่ำเสมอ (UMVUE) และตัวประมาณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำสุด ผู้ศึกษาพบว่าตัวประมาณแบบเบย์มีความสัมพันธ์กับตัวประมาณแบบเดิม เมื่อ $\beta = 0$ นั่นคือ เมื่อ $\alpha = -2, \beta = 0$ ตัวประมาณค่าพารามิเตอร์แบบเบย์ มีค่าเท่ากับตัวประมาณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำสุด เมื่อ $\alpha = -1, \beta = 0$ ตัวประมาณค่าพารามิเตอร์แบบเบย์ มีค่าเท่ากับตัวประมาณไม่เอนเอียงและมีความแปรปรวนต่ำสุดสม่ำเสมอ เมื่อ $\alpha = 0, \beta = 0$ ตัวประมาณค่าพารามิเตอร์แบบเบย์ มีค่าเท่ากับตัวประมาณความควรจะเป็นสูงสุด

บราวน์ ไค และ ดาสกุปตา (Brown, Cai and DasGupta, 2001) ศึกษาการประมาณค่าสัดส่วนประชากรในการแจกแจงทวินาม โดยวิธีวาลด์เป็นวิธีที่ใช้กันอย่างสากล โดยจะแสดงให้เห็นว่าวิธีการประมาณนี้ไม่สามารถให้ค่าความน่าจะเป็นครอบคลุมที่ยอมรับได้ สำหรับตัวอย่างขนาดเล็ก ($n \leq 40$) และ p มีค่าใกล้ 0 หรือ 1 จึงได้พิจารณาการประมาณโดยวิธีอื่น คือ วิธีวิลสันและวิธีเจฟเฟรี สามารถนำไปใช้ได้ดีและทั้งสองวิธีนี้มีความคล้ายคลึงกัน ทั้งค่าคลาดเคลื่อนสมบูรณ์และความกว้างของช่วงความเชื่อมั่นที่ประมาณได้ เมื่อตัวอย่างมีขนาดใหญ่ วิธีวิลสัน วิธีเจฟเฟรีและวิธีปรับค่าของวาลด์ มีประสิทธิภาพเท่าเทียมกัน ทั้งค่าความน่าจะเป็นครอบคลุมและค่าเฉลี่ยความกว้างของช่วงความเชื่อมั่น และวิธีปรับค่าของวาลด์เป็นวิธีที่ง่ายต่อการนำเสนออธิบาย และคำนวณ แม้กระทั่งในกรณีตัวอย่างที่ศึกษามีขนาดเล็ก วิธีปรับค่าของวาลด์สามารถเป็นวิธีมาตรฐานได้วิธีหนึ่ง

นฤดี สมุทรปรีชา (2547) ศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบช่วงของการแจกแจงแบบเลขชี้กำลัง วิธีที่ใช้ในการวิจัย คือ วิธีรากของสมการกำลังสอง วิธีประมาณโดยใช้ตัวอย่างขนาดใหญ่ วิธีการประมาณโดยใช้ตัวประมาณแบบจุด และวิธีการประมาณด้วยช่วงแบบเบส ผู้ศึกษาพบว่า เมื่อตัวอย่างมีขนาดเล็ก ($n < 30$) วิธีการประมาณโดยใช้ตัวประมาณแบบจุด การประมาณแบบเบสโดยวิธีเจฟเฟอรี่จะมีประสิทธิภาพดีกว่าการประมาณแบบเบสโดยวิธีลาปลาซ เมื่อตัวอย่างมีขนาดใหญ่ ($n \geq 30$) การประมาณโดยวิธีรากของสมการกำลังสอง วิธีประมาณโดยใช้ตัวอย่างขนาดใหญ่ การประมาณแบบเบสโดยวิธีเจฟเฟอรี่ มีประสิทธิภาพดีกว่าการประมาณโดยใช้ตัวประมาณแบบจุดกรณีตัวอย่างมีขนาดใหญ่

เคอส์ฮิด อกีล และ โลดี (Khurshid, Ageel and Lodhi, 2005) ศึกษาเปรียบเทียบการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ p ในการแจกแจงทวินามลบ โดยแยกการเปรียบเทียบเป็น 2 ลักษณะ คือ วิธีการประมาณช่วงแบบเดิม (Classical interval estimation methods) 4 วิธี ได้แก่ Large sample normal approximation without transformation, Normal approximation using variance stabilizing transformation, Use of chi-square distribution และ Improved large sample approximation และวิธีบูตสเตรป (Bootstrap methods) 7 วิธี ได้แก่ The standard bootstrap confidence limit, The standard bootstrap confidence limit on adjusted estimate, The first percentile method, The second percentile method, Bias corrected percentile method, Bootstrap-t confidence limit และ Wild bootstrap confidence limit ผู้ศึกษาพบว่า ที่ระดับความเชื่อมั่น 95% วิธี Wild bootstrap จะให้ช่วงความเชื่อมั่นที่แคบที่สุด เมื่อเปรียบเทียบระหว่างวิธีการประมาณช่วงแบบเดิมและวิธีบูตสเตรป แต่ถ้ามพิจารณาเฉพาะวิธีการประมาณช่วงแบบเดิม วิธี Large sample normal approximation without transformation จะให้ช่วงความเชื่อมั่นที่แคบที่สุด

พองนวล วงศ์ตะวัน (2549) ศึกษาเปรียบเทียบวิธีการประมาณค่าสัดส่วนของการแจกแจงทวินาม และค่าเฉลี่ยของการแจกแจงปัวซอง โดยพิจารณาจากค่าสัมประสิทธิ์ความเชื่อมั่นและเปรียบเทียบค่าเฉลี่ยความกว้างของช่วงความเชื่อมั่น เมื่อกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 50, 100 และ 300 วิธีที่ใช้ในการศึกษา 5 วิธี คือ วิธีสคออร์ วิธีประมาณค่าต่ำที่สุด วิธีเบสโดยใช้การแจกแจงเบื้องต้นแบบเจฟเฟอรี่และคอนจูเกต และวิธีบูตสเตรปสคออร์ ผู้ศึกษาพบว่าวิธีสคออร์วิธีประมาณค่าต่ำที่สุด วิธีเบสโดยใช้การแจกแจงเบื้องต้นแบบเจฟเฟอรี่และคอนจูเกต และวิธีบูตสเตรปสคออร์ ให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนดเป็นส่วนใหญ่ ที่ระดับความเชื่อมั่น 90%, 95% และ 99%

วราฤทธิ์ พานิชกิจโกศลกุล (2549) ศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์แบบช่วงของการแจกแจงปัวซอง เมื่อตัวอย่างที่ทำการศึกษามีขนาดเล็กและพารามิเตอร์มีค่าน้อย กำหนดขนาดตัวอย่างเท่ากับ 5, 10, 15, 20, 25 และ 30 ค่าพารามิเตอร์เท่ากับ 0.5, 0.8, 1.0, 1.2 และ 1.5 โดยมีวิธีการประมาณค่า 3 วิธี คือ วิธีของวาลด์แบบปรับค่าความต่อเนื่อง วิธีสคอร์ และวิธีของฟรีแมนและตุกิร์ โดยเปรียบเทียบค่าประมาณสัมประสิทธิ์ความเชื่อมั่นและค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น ผู้ศึกษาพบว่าวิธีของวาลด์แบบปรับค่าความต่อเนื่องและวิธีสคอร์ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด ในทุกระดับขนาดตัวอย่างและค่าพารามิเตอร์ ส่วนวิธีของฟรีแมนและตุกิร์ ให้ค่าประมาณสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด เมื่อขนาดตัวอย่างมีค่ามากกว่า 6 ในทุกระดับค่าพารามิเตอร์ เมื่อพิจารณาค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่น พบว่าวิธีของฟรีแมนและตุกิร์ ให้ค่าความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด ในทุกสถานการณ์ที่ทำการศึกษา