บทที่ 2

สถิติที่ใช้ในการวิจัย

ในการประมาณค่าแบบช่วงความเชื่อมั่นสำหรับพารามิเตอร์ λ ของการแจกแจงแบบปัวส์ซองมีด้วยกันหลายวิธี สำหรับวิธีการประมาณที่ทำการศึกษาเปรียบเทียบในวิจัยครั้งนี้คือ วิธีการประมาณแบบปกติ วิธีการประมาณแบบไค-สแควร์ วิธีการประมาณแบบเบส์ และวิธีการประมาณแบบวอร์เคล ในบทนี้จะกล่าวถึงรายละเอียดของทฤษฎีที่เกี่ยวข้องกับการวิจัย และวิธีการประมาณแต่ละวิธี ดังต่อไปนี้

2.1 การประมาณค่าพารามิเตอร์แบบช่วง

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n จากการแจกแจงซึ่งมี θ เป็นพารามิเตอร์ที่ไม่ทราบค่า สมมติสามารถหาตัวสถิติ $L(X_1, X_2, ..., X_n)$ และ $U(X_1, X_2, ..., X_n)$ สำหรับค่าจริง θ ใด ๆ โดยที่

$$P[L(X_1, X_2, ..., X_n < \theta < U(X_1, X_2, ..., X_n)] = 1-\alpha$$

เมื่อความน่าจะเป็น $1-\alpha$ เป็นค่าคงที่ (0 < α < 1)

นั่นคือถ้า $x_1, x_2, ..., x_n$ เป็นค่าสังเกตของ $X_1, X_2, ..., X_n$ ดังนั้น $l(x_1, x_2, ..., x_n)$ แทนด้วย l และ $u(x_1, x_2, ..., x_n)$ แทนด้วย u เป็นค่าสังเกตของ $L(X_1, X_2, ..., X_n)$ และ $U(X_1, X_2, ..., X_n)$ ตามลำดับ

จากนี้ เมื่อทราบค่าของตัวแปรสุ่ม และหาค่าตัวสถิติ L และ U ได้แล้ว ก็สามารถที่จะสร้างช่วง (l,u) และเรียกช่วง (l,u) ที่ได้ว่าช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ θ [$100(1-\alpha)\%$ confidence interval for θ] หรือกล่าวได้ว่า ช่วงค่าประมาณที่ได้ซึ่งสามารถจะคลุมค่าจริง θ จะตกอยู่ในช่วง (l,u) ด้วยความเชื่อมั่น $100(1-\alpha)\%$ และเรียกค่า l ว่าขีดจำกัดความเชื่อมั่นล่าง (lower confidence limit) เรียกค่า u ว่าขีดจำกัดความเชื่อมั่นบน (upper confidence limit) และเรียกค่า 1 - α ว่าสัมประสิทธิ์ความเชื่อมั่น (confidence coefficient)

2.2 ทฤษฎีบทลิมิตเข้าสู่ศูนย์กลาง (The Central Limit Theorem)

เมื่อสุ่มตัวอย่างขนาดใหญ่จากประชากรที่มีการแจกแจงแบบใด ๆ ก็ตามที่มีความแปรปรวนจำกัด ค่าเฉลี่ยตัวอย่างจะมีการแจกแจงลู่เข้าสู่การแจกแจงปกติ นั่นคือ

ถ้า $\overline{X}$ คือค่าเฉลี่ยของตัวอย่างสุ่ม $X_1, X_2, \ldots, X_n$ จากการแจกแจงที่มีค่าเฉลี่ย μ และมีค่าความแปรปรวน $\sigma^2 < \infty$ แล้ว เมื่อ $n \to \infty$ ได้ว่า

$$\overline{X} \sim \text{normal}\left(\mu, \sigma^2_{\overline{X}}\right) \qquad \text{(โดยประมาณ)}$$

หรือ

$$Z = \frac{\overline{X} - \mu}{\sigma_X / \sqrt{n}} \sim \text{normal}\,(0,1) \qquad \text{(โดยประมาณ)}$$

2.3 การแจกแจงแบบปัวส์ซอง (Poisson Probability distribution)

การทดลองสุ่มใด ๆ ก็ตามที่จำนวนผลสำเร็จที่เกิดขึ้นในแต่ละช่วงเวลาเป็นอิสระต่อกัน และความน่าจะเป็นของการเกิดผลสำเร็จหนึ่งครั้งในช่วงเวลาสั้น ๆ แปรผันตรงกับความยาวของช่วงเวลาและมีความน่าจะเป็นน้อยมากที่จะเกิดผลสำเร็จมากกว่าหนึ่งครั้งในช่วงเวลาสั้น ๆ สามารถกล่าวได้ว่า การทดลองนั้นเป็นการทดลองแบบปัวส์ซอง และมีฟังก์ชันความน่าจะเป็น (probability function) ดังนี้

$$f(x;\lambda) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2,\ldots$$

$$= 0 \text{ ,ค่า } x \text{ อื่นๆ}$$

โดยที่ X คือ จำนวนผลสำเร็จที่เกิดขึ้นต่อหนึ่งช่วงเวลา

λ คือ ค่าเฉลี่ยของจำนวนผลสำเร็จที่เกิดขึ้นต่อหนึ่งช่วงเวลา

และ $e =$ 2.7182...

เราเรียกการแจกแจงความน่าจะเป็นที่มีฟังก์ชันความน่าจะเป็นข้างต้นว่า การแจกแจงความน่าจะเป็นแบบปัวส์ซอง (Poisson probability distribution) แทนด้วยสัญลักษณ์ $X \sim Poisson(\lambda)$

ในการอนุมานเกี่ยวกับพารามิเตอร์ λ ของการแจกแจงแบบปัวส์ซองที่เป็นการประมาณค่า (Estimation) และในการประมาณค่าแบบจุด (Point estimation) จะได้ว่า

ถ้าสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงแบบปัวส์ซอง ที่มีพารามิเตอร์ λ จะได้ตัวอย่างสุ่ม คือ $X_1, X_2, \ldots, X_n$ ดังนั้นตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum likelihood) ของ λ คือ

$$\hat{\lambda} = (1/n)\sum_{i=1}^{n} X_i = \overline{X}$$

เมื่อ $\sum_{i=1}^{n} X_i$ มีการแจกแจงแบบปัวส์ซองที่มีพารามิเตอร์ คือ $n\lambda$ และ $\hat{\lambda}$ มีคุณสมบัติไม่เอนเอียงและมีความแปรปรวนต่ำสุด (Minimum variance unbiased estimator : MVUE)

และจะได้ว่า

$$E(\hat{\lambda}) = (1/n)\sum_{i=1}^{n} E(X_i)$$

$$= (1/n)\sum_{i=1}^{n} \lambda$$

$$= (1/n)(n\lambda)$$

ฉะนั้น $V(\hat{\lambda}) \;=\; \lambda$

$$V(\hat{\lambda}) = V(\overline{X})$$

$$= V(X)/n^2$$

ฉะนั้น $V(\hat{\lambda}) \;= \lambda / n$

ดังนั้นถ้าตัวอย่าง มีขนาดใหญ่เพียงพอ (n $\rightarrow \infty$) จากทฤษฎีบทลิมิตสู่ศูนย์กลาง $\hat{\lambda}$ จะมีการแจกแจงปกติโดยประมาณ และมีค่าเฉลี่ยของการแจกแจงเป็น λ และความแปรปรวน คือ λ / n

2.4 **การแจกแจงแกมมา** (Gamma Distribution)

ตัวแปรสุ่ม X เรียกว่า มีการแจกแจงแกมมาซึ่งมี พารามิเตอร์ $\alpha > 0, \beta > 0$ ถ้ามีฟังก์ชันความน่าจะเป็น(Probability function) ดังนี้

$$f(x) = P(X = x) = \frac{x^{\alpha-1}e^{-\beta x}}{\beta^{\alpha}\Gamma(\alpha)}, x > 0$$

การแจกแจงแกมมาแทนด้วยสัญลักษณ์ $X \sim gamma(\alpha, \beta)$

โดยมีค่าเฉลี่ย คือ α / β

และค่าความแปรปรวน คือ α / β^2

2.5 **การแจกแจงไค-สแควร์**(Chi-Square distribution)

พิจารณากรณีเฉพาะของการแจกแจงแกมมาให้ตัวแปรสุ่ม X มีการแจกแจงแกมมาซึ่งมี $\alpha = n/2$ และ $\beta = 1/2$ เมื่อ n เป็นจำนวนเต็มบวก จะได้ฟังก์ชันความน่าจะเป็น (Probability function) ดังนี้

$$f(x) = P(X = x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1}e^{-x/2}, x > 0$$

เราเรียก X ว่ามีการแจกแจงไค-สแควร์ ที่มีองศาความเป็นอิสระเท่ากับ n แทนด้วยสัญลักษณ์ $X \sim \chi_n^2$

โดยมีค่าเฉลี่ย คือ n

และ ค่าความแปรปรวน คือ $2n$

2.6 ความสัมพันธ์ระหว่างการแจกแจงแบบปัวส์ซองกับการแจกแจงแกมมา

จากตัวแปรสุ่ม X ที่มีการแจกแจง $gamma(\alpha,\beta)$ เมื่อ α เป็นจำนวนเต็มบวก ดังนั้นสำหรับแต่ละ x จะได้

$$P(X \le x) = P(Y \ge \alpha)$$

เมื่อ $Y \sim Poisson(\beta x)$[1]

ในทางกลับกัน พิสูจน์ได้ว่า ในช่วงเวลาห่างระหว่างเกิดเหตุการณ์ $T_i, i = 1,2,...$ เป็นอิสระกันและต่างมีการแจกแจงเอกซโพเนนเชียล ดังนั้นจำนวนเหตุการณ์ที่เกิดขึ้นในช่วงเวลา $(0,t]$ มีการแจกแจงปัวส์ซอง:

ให้ $P_n(t) =$ ฟังก์ชันการแจกแจงสะสมของจำนวนเหตุการณ์เกิดขึ้นในช่วง $(0,t]$ และให้ $\sum_{i=1}^{n+1} T_i$ คือผลบวกของช่วงเวลาห่างเกิดเหตุการณ์ $n+1$ ช่วง เพราะฉะนั้น X มีการแจกแจงแกมมาด้วยพารามิเตอร์ $(n+1,\lambda)$ และได้ว่า

$$P_n(t) = P(X > t)$$

$$= \int_t^{\infty} \frac{\lambda(\lambda x)^n e^{-\lambda x}}{n!} dx$$

จากนี้ให้ $u = x - t$ ได้

$$P_n(t) = \int_0^{\infty} \frac{\lambda^{n+1}(u+t)^n}{n!} e^{-\lambda t} e^{-\lambda u} du$$

จากทฤษฎีบททวินาม (binomial theorem) ได้

---

[1] Casella G. and Berger.R.L, Statistical Inference,(California:Wadsworth ,1990) , p.101.

$$(u+t)^n = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} u^{n-i} t^i$$

เพราะฉะนั้น

$$P_n(t) \qquad = \sum_{i=0}^{n} \frac{\lambda^{n+1} e^{-\lambda t} t^i}{i!(n-1)!} \int_0^{\infty} e^{-\lambda u} u^{n-i} du$$

$$= \sum_{i=0}^{n} \frac{e^{-\lambda t} (\lambda t)^i \Gamma(n-i-1)}{i!(n-1)!}$$

$$= \sum_{i=0}^{n} \frac{e^{-\lambda t} (\lambda t)^i}{i!}$$

ซึ่งคือ ฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปัวส์ซองนั่นเอง

2.7 วิธีการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ของการแจกแจงแบบปัวส์ซอง

การประมาณค่าแบบช่วงสำหรับพารามิเตอร์ λ ของการแจกแจงแบบปัวส์ซอง ที่ใช้ในการศึกษาวิจัยมีรายละเอียดดังนี้

2.7.1 วิธีการประมาณแบบปกติ

ให้ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มจากการแจกแจงแบบปัวส์ซอง ที่มีพารามิเตอร์คือ λ และ $\hat{\lambda} = \bar{X}$ เป็นค่าเฉลี่ยตัวอย่าง

เมื่อให้ $\hat{\lambda}$ เป็นตัวประมาณแบบจุดของ λ จะได้

$$Z = \frac{\hat{\lambda} - \lambda}{(\lambda / n)^{1/2}} \sim N(0,1) \quad , \; n \to \infty$$

และ

$$P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1 - \alpha$$

$$P(-Z_{1-\alpha/2} < \frac{\hat{\lambda} - \lambda}{(\lambda / n)^{1/2}} < Z_{1-\alpha/2)}) = 1 - \alpha$$

$$P\left[-Z_{1-\alpha/2}\{\lambda/n\}^{1/2} < \hat{\lambda}-\lambda < Z_{1-\alpha/2}\{\lambda/n\}^{1/2}\right]=1-\alpha$$

$$P\left[\hat{\lambda}-Z_{1-\alpha/2}\{\lambda/n\}^{1/2} < \lambda < \hat{\lambda}+Z_{1-\alpha/2}\{\lambda/n\}^{1/2}\right]=1-\alpha$$

แต่เนื่องจากไม่ทราบค่าเฉลี่ยของประชากร (λ) จึงแทนค่า λ ด้วย $\hat{\lambda}=\overline{X}$ จะได้

$$P\left[\hat{\lambda}-Z_{1-\alpha/2}\{\hat{\lambda}/n\}^{1/2} < \lambda < \hat{\lambda}+Z_{1-\alpha/2}\{\hat{\lambda}/n\}^{1/2}\right]=1-\alpha$$

และดังนั้น ช่วงความเชื่อมั่น $100(1-\infty)\%$ สำหรับ λ คือ

$$\left(\hat{\lambda}-Z_{1-\alpha/2}\{\hat{\lambda}/n\}^{1/2}, \hat{\lambda}+Z_{1-\alpha/2}\{\hat{\lambda}/n\}^{1/2}\right)$$

2.7.2. วิธีการประมาณแบบไค-สแควร์

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบปัวส์ซอง ที่มีพารามิเตอร์ คือ λ

ให้ $Y=\sum X_i$ ฉะนั้น $Y \sim Poisson(n\lambda)$

จากนี้หาช่วงความเชื่อมั่นของ λ จากการแก้สมการต่อไปนี้หาค่า λ

$$\sum_{k=y}^{\infty} \frac{e^{-n\lambda}(n\lambda)^k}{k!} = \frac{\alpha}{2} \qquad (2.1)$$

และ

$$\sum_{k=0}^{y} \frac{e^{-n\lambda}(n\lambda)^k}{k!} = \frac{\alpha}{2} \qquad (2.2)$$

จากความสัมพันธ์ระหว่างการแจกแจงปัวส์ซองและการแจกแจงแกมมา เราเขียนได้ว่า

$$\frac{\alpha}{2} = \sum_{k=y}^{\infty} \frac{e^{-n\lambda}(n\lambda)^k}{k!} = P(Y \ge y)$$

$$= P(U < \lambda), U \sim gamma \quad (y, n)$$

$$= P(X < 2n\lambda), X = 2nU \sim \chi^2_{2y}$$

เนื่องจาก $X = 2nU$ มีการแจกแจงไค-สแควร์ด้วยองศาความเป็นอิสระเท่ากับ $2y$ ดังนั้น ได้ค่าของ λ ให้เป็น λ_L ดังนี้

$$2n\lambda_L = \chi^2_{2y,\alpha/2}$$

$$\lambda_L = \frac{1}{2n}\chi^2_{2y,\alpha/2}$$

ในทำนองเดียวกันจากสมการ (2.2) หาค่าของ λ ให้เป็น λ_U ได้

$$\lambda_U = \frac{1}{2n}\chi^2_{2(y+1),1-\alpha/2}$$

และดังนั้น ช่วงความเชื่อมั่น $100(1-\infty)\%$ สำหรับ λ คือ

$$\left(\frac{1}{2n}\chi^2_{2y,\alpha/2}, \frac{1}{2n}\chi^2_{2(y+1),1-\alpha/2}\right)$$

2.7.3 วิธีการประมาณแบบเบส์

วิธีการประมาณค่าแบบช่วงด้วยวิธีการเบส์สำหรับพารามิเตอร λ ของการแจกแจงแบบปัวส์ซอง จะใช้ฟังก์ชันความหนาแน่นภายหลัง(Posterior density function)ของ Θ คือ $g(\lambda|\underset{\sim}{x})$ โดยพิจารณาดังนี้

ให้ $X_1, X_2, ..., X_n$ มีการแจกแจงเหมือนกันและอิสระจากกันด้วยการแจกแจงแบบปัวส์ซอง ที่มีพารามิเตอร์คือ $\Theta = \lambda$ (Poisson(λ)) และฟังก์ชันความหนาแน่นก่อน(Prior density function)ของ Θ มีการแจกแจงแบบ $gamma(\alpha,\beta)$ แทนโดย $\Theta \sim gamma(\alpha,\beta)$ และมีฟังก์ชันความสูญเสีย(Loss function)อยู่ในรูปของกำลังสองของความแตกต่างระหว่างตัวประมาณและตัวพารามิเตอร์ที่ถูกประมาณ ดังนั้นจะได้ฟังก์ชันความหนาแน่นภายหลังของ Θ เมื่อกำหนด $\sum X = \sum x$ โดยมีรายละเอียดดังนี้

พิจารณาฟังก์ชันแบบมีเงื่อนไขของ $\underset{\sim}{x}$ เมื่อกำหนด $\Theta = \lambda$ คือ

$$h(\underset{\sim}{x}|\lambda) = \frac{e^{-n\lambda}\lambda^{X}}{x_1!x_2!...x_n!}, \lambda > 0, x = \sum_{i=1}^{n} x_i$$

$h(\underset{\sim}{x}|\lambda)$ คือ ฟังก์ชันความหนาแน่นร่วมแบบมีเงื่อนไขของ $\underset{\sim}{x}$ เมื่อกำหนด λ และ พารามิเตอร์ λ เป็นค่าของตัวแปรสุ่ม Θ ซึ่งมีฟังก์ชันความหนาแน่นก่อน(prior density function) คือ

$$g(\lambda) = gamma(\alpha,\beta) = \frac{e^{-\lambda/\beta}\lambda^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}, \lambda > 0, \alpha, \beta > 0$$

เราเรียก $g(\lambda)$ ว่า ฟังก์ชันความหนาแน่นก่อนของ Θ และจะได้ $g(\lambda|\underset{\sim}{x})$ เป็นฟังก์ชันความหนาแน่นแบบมีเงื่อนไขของ λ เมื่อกำหนด $X = x$ และเรียก $g(\lambda|\underset{\sim}{x})$ ว่า ฟังก์ชันความหนาแน่นภายหลัง (posterior density function) ของ λ เมื่อกำหนด $X = x$ ดังนี้

$$g(\lambda|\underset{\sim}{x}) = \frac{g(\lambda)h(\underset{\sim}{x}|\lambda)}{\int_0^{\infty} g(\lambda)h(\underset{\sim}{x}|\lambda)d\lambda}$$

พิจารณา

$$\int_0^{\infty} g(\lambda)h(\underset{\sim}{x}|\lambda)d\lambda = \int_0^{\infty} \frac{e^{-\lambda/\beta}\lambda^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \frac{e^{-n\lambda}\lambda^{X}}{(X_1!X_2!...X_n!)} d\lambda$$

$$= \frac{1}{\beta^{\alpha}\Gamma(\alpha)(X_1!X_2!...X_n!)}\int_0^{\infty} e^{-\lambda/((n\beta+1)/\beta)}\lambda^{(X+\alpha)-1}d\lambda$$

$$= \frac{(\beta/\{n\beta+1\})^{X+\alpha}\Gamma(X+\alpha)}{\beta^{\alpha}\Gamma(\alpha)(X_1!X_2!...X_n!)}\int_0^{\infty}\frac{e^{-\lambda/(\beta/\{n\beta+1\})}\lambda^{(X+\alpha)-1}}{(\beta/\{n\beta+1\})^{X+\alpha}\Gamma(X+\alpha)}d\lambda$$

ฉะนั้น $= \dfrac{B^{A}\Gamma(A)}{\beta^{\alpha}\Gamma(\alpha)(X_1!X_2!...X_n!)}$ ซึ่ง $A = X+\alpha$ และ $B = \dfrac{\beta}{n\beta+1}$

นั่นคือ จาก $g(\lambda|\underset{\sim}{x}) = \dfrac{g(\lambda)h(\underset{\sim}{x}|\lambda)}{\int_0^{\infty} g(\lambda)h(\underset{\sim}{x}|\lambda)d\lambda}$

$$= \frac{e^{-\lambda/\beta}\lambda^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\ \frac{e^{-n\lambda}\lambda^{X}}{(X_1!X_2!...X_n!)}\ \frac{\beta^{\alpha}\Gamma(\alpha)(X_1!X_2!...X_n!)}{B^{A}\Gamma(A)}$$

$$= \frac{e^{-\lambda/B}\lambda^{A-1}}{B^{A}\Gamma(A)}$$

ดังนั้นฟังก์ชันความหนาแน่นภายหลังของ Θ เมื่อกำหนด $X = x$ คือ $gamma(A,B)$ เมื่อ $A = X+\alpha$ และ $B = \dfrac{\beta}{n\beta+1}$

เมื่อพิจารณาหาช่วงความเชื่อมั่น $100(1-\alpha)\%$ แบบเบส์จะพิจารณาจากฟังก์ชันความหนาแน่นภายหลังของ Θ เมื่อกำหนด $X = x$ ในกรณีเฉพาะเมื่อ $\alpha = \beta = 1$ จะได้ดังนี้

$$P(\lambda_L \le \Theta \le \lambda_U|\underset{\sim}{x}) = \int_{\lambda_L}^{\lambda_U}\frac{B^{A}\lambda^{A-1}e^{-B\lambda}}{\Gamma(A)}d\lambda$$

ให้ $v = 2B\lambda$ ดังนั้น $\lambda = v/2B$

$$= \int_{\lambda_L}^{\lambda_U}\frac{B^{A}\left(\frac{v}{2B}\right)^{A-1}e^{-v/2}}{\Gamma(A)}d\left(\frac{v}{2B}\right)$$

$$= \frac{1}{2^{A}\Gamma(A)}\int_{2B\lambda_L}^{2B\lambda_U} v^{A-1}e^{-v/2}dv \qquad (2.\ 3)$$

เมื่อ $A = X+1$ และ $B = n+1$

พิจารณาจากการอินติเกรตใน (2.3) พบว่า คือฟังก์ชันการแจกแจงแบบไค-สแควร์ที่มีองศาความเป็นอิสระ คือ $2(X+1)$
ฉะนั้นจะได้ขีดจำกัดความเชื่อมั่นล่างคือ

$$2B\lambda_L = \chi^2_{2(y+1),\alpha/2} \quad \text{เมื่อ} \quad B=(n+1) \qquad \text{และให้} \ y \ \text{แทน} \ X$$

$$\lambda_L = \frac{1}{2(n+1)}\chi^2_{2(y+1),\alpha/2}$$

ทำนองเดียวกันจะได้ขีดจำกัดความเชื่อมั่นบน คือ

$$2B\lambda_U = \chi^2_{2(y+1),\alpha/2} \quad \text{เมื่อ} \quad B=(n+1)$$

$$\lambda_U = \frac{1}{2(n+1)}\chi^2_{2(y+1),1-\alpha/2}$$

และดังนั้นช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ λ คือ

$$\left(\frac{1}{2(n+1)}\chi^2_{2(y+1),\alpha/2}, \frac{1}{2(n+1)}\chi^2_{2(y+1),1-\alpha/2}\right)$$

2.7.4 วิธีการประมาณแบบวอร์เดล

วิธีการประมาณโดยการหาช่วงที่สั้นที่สุด เสนอโดย ดอน จี วอร์เดล (Don.G.Wardell:1995) ได้ พัฒนาการศึกษามาจาก วิธีการทางสถิติ ของ มูด แกรบิลและบอส (Mood,Graybill,Boss;1974) โดยมีการศึกษาดังนี้

ให้ X เป็นตัวแปรสุ่มแบบไม่ต่อเนื่องที่มี $f_X(x;\lambda)$ เป็นฟังก์ชันความน่าจะเป็นและมี λ คือพารามิเตอร์ n คือ ขนาดตัวอย่าง Θ เป็นปริภูมิของช่วงของพารามิเตอร์และให้ $T=t(X_1,...,X_n)$ เป็นตัวสถิติที่พอเพียงที่ได้จากตัวอย่าง

จะได้ T มีฟังก์ชันการแจกแจงสะสมคือ $F_T(t;\lambda)$ ถ้าให้ t_0 เป็นค่าสังเกตใดๆ ดังนั้นระดับความเชื่อมั่น $100(1-p_1-p_2)\%$ เมื่อ $0<p_1,p_2<1$ สำหรับพารามิเตอร์ λ สามารถหาได้จาก

$$p_1 = F_T(t_0;\lambda_H) \qquad (2.4)$$

และ

$$p_2 = 1 - F_T(t_0 - 1;\lambda_L) \qquad (2.5)$$

เมื่อ λ_L คือค่าขีดจำกัดความเชื่อมั่นล่างของช่วงแรกเริ่มที่จะนำไปใช้ในการหาค่า λ_L^* (ค่าขีดจำกัดความเชื่อมั่นล่างที่ต่ำที่สุด)

λ_H คือค่าขีดจำกัดความเชื่อมั่นบนของช่วงแรกเริ่มที่จะนำไปใช้ในการหาค่า λ_H^* (ค่าขีดจำกัดความเชื่อมั่นบนที่ต่ำที่สุด)

p_1 คือ ความน่าจะเป็นที่จะเกิดจำนวนผลสำเร็จเท่ากับ t_0 หรือต่ำกว่าเมื่อขนาดตัวอย่างเท่ากับ n ถ้าค่าเฉลี่ยตัวอย่างมีค่าเท่ากับ λ_H

p_2 คือ ความน่าจะเป็นที่จะเกิดจำนวนผลสำเร็จเท่ากับ t_0 หรือมากกว่าเมื่อขนาดตัวอย่างเท่ากับ n ถ้าค่าเฉลี่ยตัวอย่างมีค่าเท่ากับ λ_L

ค่าของ λ_H และ λ_L ได้มาจากการอ้างอิงผลของค่าประมาณแบบช่วงของวิธีการประมาณแบบปกติมาใช้เป็นค่าเริ่มต้นในการหาค่าประมาณที่ดีที่สุด

Mood,Graybill,Boss;(1974) กล่าวว่า การพิจารณาการแก้สมการ (2.4),(2.5) จะใช้วิธีการเชิงตัวเลขในการหาค่า และวอร์เดลได้นำวิธีของ Mood,Graybill,Boss มาใช้เฉพาะเพื่อหาคำตอบที่จะเป็นไปได้ คือค่า p_1, p_2 เพื่อให้เป็นไปตามเงื่อนไข 2 ข้อที่จะกล่าวต่อไป การหาคำตอบที่ดีที่สุด วอร์เดลได้นำเสนอวิธีการดังกล่าว มีรายละเอียดดังนี้

วิธีการศึกษา

การหาช่วงความเชื่อมั่น ที่สั้นที่สุดจะกระทำดังนี้

ให้

CL แทน ระดับความเชื่อมั่นที่ต้องการพิจารณา ในการวิจัยครั้งนี้ ใช้ 0.90,0.95,0.99 ตามลำดับ

ดังนั้นการพิจารณาหาค่าที่เหมาะสมแบบไม่เชิงเส้น(nonlinear optimization) พิจารณาโดย

$$\text{minimize}_{\lambda_H,\lambda_L}\ \lambda_H - \lambda_L \qquad (3a)$$

ภายใต้เงื่อนไข

$$CL \le 1 - p_1 - p_2 \le 1 \qquad (3b)$$

$$\lambda_H - \lambda_L \ge 0 \qquad (3c)$$

$$\lambda_H, \lambda_L \in \Theta \qquad (3d)$$

เมื่อ p_1 และ p_2 ได้มาจาก (2.4)และ(2.5)

นอกจากนี้ การแก้ปัญหา (3) ต้องอาศัยเงื่อนไขที่เหมาะสมเพื่อพิจารณาหาค่าที่ดีที่สุดโดยจะหาคำตอบภายใต้เงื่อนไขที่เหมาะสมตาม (4)ซึ่งการหาเงื่อนไข(4)จะประยุกต์จากวิธีการแบบไม่เชิงเส้นของ Karush-Kuhn-Tucker (Hillier and Lieberman 1990)

สมมติให้ $\lambda_H \neq \lambda_L$ ดังนั้นจะหาเงื่อนไขที่เหมาะสม จาก

$$-1-\lambda^* \frac{\partial P_1}{\partial \lambda_H}\Big|\lambda_H = \lambda_H^* = 0 \qquad (4a)$$

$$-1-\lambda^* \frac{\partial P_2}{\partial \lambda_L}\Big|\lambda_L = \lambda_L^* = 0 \qquad (4b)$$

$$p_1 + p_2 - \alpha = 0 \qquad (4c)$$

โดยที่ λ_H^*, λ_L^* คือ ขีดจำกัดความเชื่อมั่นบนและขีดจำกัดความเชื่อมั่นล่างที่ต่ำที่สุด

λ^* คือ ตัวดำเนินการของลากรังจ์

α คือ 1-CL

เพื่อขจัด λ^* จึงรวม (4a) และ(4b) เข้าด้วยกัน จึงได้เงื่อนไขในการหาคำตอบที่เหมาะสมดังนี้

$$-\frac{\partial p_1}{\partial \lambda_H}\Big|_{\lambda_H=\lambda_H^*} = \frac{\partial p_2}{\partial \lambda_L}\Big|_{\lambda_L=\lambda_L^*} \qquad (4d)$$

เมื่อพิจารณาในกรณีการแจกแจงแบบปัวส์ซอง จะมีรายละเอียด ดังนี้

$f_x(x;\lambda) = (e^{-\lambda}\lambda^x / x!)$ เมื่อ $\lambda \geq 0$ และ x =0,1,2,...

$T = \sum_{i=1}^{n} X_i$ เป็นตัวสถิติที่พอเพียงสำหรับ λ และT จะมีการแจกแจงแบบปัวส์ซองที่มีพารามิเตอร์ $n\lambda$

t_0 ในกรณีการแจกแจงแบบปัวส์ซองนี้ คือ ค่า X (จำนวนหน่วยที่เกิดผลสำเร็จในช่วงเวลาที่กำหนด)

ดังนั้นกรณีการแจกแจงแบบปัวส์ซอง พิจารณา (3)จะมีเงื่อนไขเป็นดังนี้

minimize $_{\lambda_H,\lambda_L}\ \lambda_H - \lambda_L$ (5a)

ภายใต้เงื่อนไข

$$CL \leq 1 - p_1 - p_2 \leq 1 \qquad (5b)$$

$$\lambda_H - \lambda_L \geq 0 \qquad (5c)$$

$$\lambda_H \geq 0 \qquad (5d)$$

$$\lambda_L \geq 0 \qquad (5e)$$

และการจะได้คำตอบที่เหมาะสมสำหรับหาช่วงที่สั้นที่สุด พิจารณาจาก 2 สมการคือ

$$p_1 + p_2 - \alpha = 0 \qquad (6a)$$

และ $$t_0 e^{n(\lambda_H^* - \lambda_L^*)} - n\lambda_H^* (\frac{\lambda_H^*}{\lambda_L^*})^{t_0 - 1} = 0 \qquad (6b)$$

เมื่อ λ_L^* คือ ขีดจำกัดความเชื่อมั่นล่างของการแจกแจงที่มีค่าต่ำที่สุด

λ_H^* คือ ขีดจำกัดความเชื่อมั่นบนของการแจกแจงที่มีค่าต่ำที่สุด

ในการหาคำตอบ λ_L^*, λ_H^* จะได้จากการแก้สมการ (6a) และ (6b)เมื่อ (6b)ได้มาจากการศึกษาตาม (4b) มีรายละเอียดดังนี้

ให้ $p_1 = F_{(T)}(i_0;\lambda_H) = \sum_{t=0}^{t_0} \frac{e^{n_{\lambda_H}} (n\lambda_H)^t}{t!}$;

$$p_2 = 1 - F_{T}(t_0 - 1;\lambda_L)$$
$$= \sum_{t=0}^{t_0-1} \frac{e^{n\lambda_L} (n\lambda_L)^t}{t!}$$

และจาก (4d)

$$-\frac{\partial p_1}{\partial \lambda_H}\bigg|_{\lambda_H = \lambda_H^*} = \frac{\partial p_2}{\partial \lambda_L}\bigg|_{\lambda_L = \lambda_L^*}$$

$$\frac{\partial p_1}{\partial \lambda_H} = \sum_{t=0}^{t_0} \left[\frac{e^{-n\lambda_H}(nt)(n\lambda_H)^{t-1}}{t!} + \frac{-ne^{-n\lambda_H}(n\lambda_H)^t}{t!} \right]$$

$$= n\sum_{t=1}^{t_0} \frac{e^{-n\lambda_H}(n\lambda_H)^{t-1}}{(t-1)!} - n\sum_{t=0}^{t_0} \frac{e^{-n\lambda_H}(n\lambda_H)^t}{t!}$$

$$= ne^{-n\lambda_H}\left[1 + (n\lambda_H) + \frac{(n\lambda_H)^2}{2} + \ldots + \frac{(n\lambda_H)^{t_0-1}}{(t_0-1)!} - 1 - (n\lambda_H) - \frac{(n\lambda_H)^2}{2} - \ldots - \frac{(n\lambda_H)^{t_0-1}}{(t_0-1)!} - \frac{(n\lambda_H)^{t_0}}{t_0!}\right]$$

$$= \frac{-ne^{-n\lambda_H}(n\lambda_H)^{t_0}}{t_0!}$$

ในทำนองเดียวกันจะได้

$$\frac{\partial p_2}{\partial \lambda_L} = \frac{ne^{-n\lambda_L}(n\lambda_L)^{t_0-1}}{(t_0-1)!}$$

ดังนั้น $$-\frac{\partial p_1}{\partial \lambda_H}\bigg|_{\lambda_H = \lambda_H^*} = \frac{\partial p_2}{\partial \lambda_L}\bigg|_{\lambda_L = \lambda_L^*}$$

$$\frac{ne^{-n\lambda^*_H}(n\lambda^*_H)^{t_0}}{t_0!} = \frac{ne^{-n\lambda^*_L}(n\lambda^*_L)^{t_0}}{(t_0-1)!}$$

$$\frac{n\lambda_H(n\lambda^*_H)^{t_0-1}}{(n\lambda^*_L)^{t_0-1}} = \frac{t_0e^{-n\lambda^*_L}}{e^{-n\lambda^*_H}} = t_0e^{n(\lambda^*_H-\lambda^*_L)}$$

$$t_0e^{n(\lambda^*_H-\lambda^*_L)} - n\lambda^*_H(\frac{\lambda^*_H}{\lambda^*_L})^{t_0-1} = 0 \qquad (6b)$$

ในการคำนวณหา λ^*_L , λ^*_H จะหาได้จากการแก้สมการ (6a) และ(6b) ซึ่งในการวิจัยครั้งนี้ใช้การหาคำตอบที่ดีที่สุดโดยใช้โปรแกรมสำเร็จรูป EXCEL โดยใช้วิธีการ Non-linear programming โดยแทนเงื่อนไขต่างๆเพื่อหาคำตอบที่ดีที่สุดตามเงื่อนไขที่ได้ตาม (5)

ข้อจำกัดของวิธีการนี้ในการทำการวิจัยจากโปรแกรม EXCEL ที่ใช้ในการศึกษาวิธีไม่เชิงเส้น(Non-linear Method) โดยเมื่อขนาดตัวอย่างและค่าพารามิเตอร์เพิ่มขึ้นจะไม่สามารถหาค่าระดับความเชื่อมั่นและค่าความยาวเฉลี่ยได้ เพราะวิธีการประมาณนี้ผู้เสนอคือ วอร์เดล ได้เสนอโปรแกรมมาเพื่อศึกษาวิธีการนี้ในกรณีขนาดตัวอย่างเล็ก เท่านั้น

2.8 เกณฑ์การตรวจสอบ

การตรวจสอบวิธีการประมาณแบบช่วงสำหรับพารามิเตอร์ของการแจกแจงแบบปัวส์ซองทั้ง 4 วิธีนั้น จะทำการตรวจสอบค่าระดับความเชื่อมั่นและเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่กำหนดได้จากแต่ละสถานการณ์ทดลองในการทดลองซ้ำ 2,000 ครั้ง โดยมีรายละเอียดการตรวจสอบดังนี้

การตรวจสอบค่าระดับความเชื่อมั่นเพื่อตรวจสอบว่าวิธีการประมาณใดให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนด เกณฑ์ที่ใช้ในการตรวจสอบจะอาศัยการทดสอบสมมติฐานโดยใช้ตัวสถิติ Z มีรูปแบบดังนี้

$$H_0 : p \geq p_0$$
$$H_1 : p < p_0$$

พิจารณา ความน่าจะเป็นยอมรับ H_0 เมื่อ H_0 เป็นจริงที่ระดับนัยสำคัญ α_0:

$$P\left(-Z_{1-\alpha_0} < \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = 1-\alpha_0$$

เพราะฉะนั้น

$$-Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} - p_0$$

$$p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p}$$

เมื่อพิจารณา ค่าระดับความเชื่อมั่น จะมีค่า $0 < \hat{p} < 1$ ซึ่งจะได้ว่า

$$p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < 1$$

ฉะนั้น จะได้ช่วงของการยอมรับสมมติฐานหลัก H_0 คือ

$$\left(p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{n}}\ ,\ 1\right)$$

ดังนั้นจะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่า p_0 ที่กำหนดถ้า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(p_0 - Z_{1-\alpha_0}\sqrt{\frac{p_0(1-p_0)}{n}}\ ,\ 1\right)$$

โดยที่ α_0 คือระดับนัยสำคัญ ที่กำหนดในการทดสอบ
สำหรับการวิจัยครั้งนี้กำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.05

p คือระดับความเชื่อมั่น

$\hat{p}$ คือระดับความเชื่อมั่นที่ได้จากการทดลอง

p_0 คือระดับความเชื่อมั่นที่กำหนด (0.90, 0.95 และ 0.99)

1. ที่ระดับความเชื่อมั่น 90%

$$H_0 : p \geq 0.90$$
$$H_1 : p < 0.90$$

จะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่า 90 % ถ้า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(0.90 - 1.645\sqrt{\frac{0.90(0.10)}{2,000}}\ ,\ 1\right)$$

หรือมีค่าอยู่ในช่วง $(0.8890\,,1)$

2. ที่ระดับความเชื่อมั่น 95%

$$H_0 : p \geq 0.95$$
$$H_1 : p < 0.95$$

จะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่า 95 % ถ้า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(0.95 - 1.645\sqrt{\frac{0.95(0.05)}{2,000}}\ ,\ 1\right)$$

หรือมีค่าอยู่ในช่วง $(0.9420\,,1)$

3. ที่ระดับความเชื่อมั่น 99%

$$H_0 : p \geq 0.99$$
$$H_1 : p < 0.99$$

จะได้ว่า วิธีการประมาณให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่า 99 % ถ้า $\hat{p}$ มีค่าอยู่ในช่วง

$$\left(0.99-1.645\sqrt{\frac{0.99(0.01)}{2,000}}\ ,\ 1\right)$$

หรือมีค่าอยู่ในช่วง $(0.9864\,,1)$

นั่นคือ ค่าระดับความเชื่อมั่นที่ได้จากการทดลองจะต้องมีค่าไม่ต่ำกว่า 0.8890 , 0.9420 , 0.9864 ที่ระดับความเชื่อมั่น 90% , 95 % และ 99% ตามลำดับ จึงจะถือว่าวิธีการประมาณนั้นให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนดและจะนำเฉพาะวิธีการประมาณที่ให้ค่าระดับความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนดไปทำการเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่อไป โดยจะทำการพิจารณาเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่ได้ว่าวิธีการประมาณใดสามารถให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นสั้นที่สุด จะสรุปว่าวิธีการประมาณนั้นได้ช่วงประมาณที่เหมาะสมที่สุดสำหรับสถานการณ์นั้น ๆ