บทความวิจัย

ช่วงความเชื่อมั่นสำหรับความแตกต่างของสองสัดส่วนโดยขั้นตอนวิธีการเลือกตัวอย่างแบบเลือกซ้ำที่สำคัญ

โชติกา วุฒิสาร รมิดา ศรีเหรา และ พัทธ์ชนก ศรีสุรเดชชัย*
ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0–2564–4444 ต่อ 2100 กด 308 อีเมล: spatchan@tu.ac.th DOI: 10.14416/j.kmutnb.2018.08.003
รับเมื่อ 27 ตุลาคม 2560 ตอบรับเมื่อ 5 มกราคม 2561 เผยแพร่ออนไลน์ 15 สิงหาคม 2561

บทคัดย่อ

ช่วงความเชื่อมั่นสำหรับความแตกต่างของสองสัดส่วน $(p_1 - p_2)$ ได้รับความสนใจโดยนักสถิติตั้งแต่อดีตจนถึงปัจจุบัน มีนักวิจัยเสนอและศึกษาเปรียบเทียบช่วงความเชื่อมั่นประเภทนี้อย่างกว้างขวาง ปัจจุบันคอมพิวเตอร์ได้มีบทบาทอย่างสำคัญในการคำนวณเชิงสถิติ วิธีสร้างช่วงความเชื่อมั่นจึงอาจใช้ขั้นตอนวิธี (Algorithm) แทนการวิเคราะห์ทางตรง ในงานวิจัยนี้ ขั้นตอนวิธีการเลือกตัวอย่างแบบเลือกซ้ำที่สำคัญ (SIR) ซึ่งพบได้บ่อยในสถิติแบบเบส์ได้ถูกนำมาดัดแปลงเพื่อหาช่วงความเชื่อมั่นแล้วนำไปเปรียบเทียบกับวิธีเบส์ วิธีไฮบริด และวิธีวัลด์ ผลการศึกษาพบว่าช่วงจาก SIR ให้ความน่าจะเป็นคุ้มรวมไม่แตกต่างไปจากวิธีเบส์และวิธีไฮบริดในหลายสถานการณ์ และวิธีทั้งสามนี้ให้ผลดีกว่าวิธีวัลด์ในทุกลักษณะประชากรและขนาดตัวอย่างที่ศึกษา เนื่องจากขั้นตอนวิธี SIR ใช้ตัวอย่าง p_1 และ p_2 โดยตรง ช่วงความเชื่อมั่นของ $p_1 - p_2$ ที่ได้จึงอยู่ในช่วง [–1, +1] อย่างแน่นอน ในกรณีที่ $\hat{p}_1 = \hat{p}_2 = 0$ ขั้นตอนวิธี SIR ยังคงสามารถหาช่วงความเชื่อมั่นได้ ในขณะที่ช่วงแบบวัลด์จะไม่สามารถคำนวณได้เนื่องจากความคลาดเคลื่อนมาตรฐานเป็นศูนย์ และลักษณะของประชากรมีผลต่อความยาวช่วงในลักษณะที่ว่า หากทั้ง p_1 และ p_2 ใกล้ .5 (สมมาตร) แล้ว ความยาวของช่วงเชื่อมั่นที่ได้จะมีแนวโน้มมากกว่าความยาวของช่วงที่ประชากรเบ้ เช่น $p_1 = p_2 = .05$ โดยสรุปในหลายสถานการณ์ ขั้นตอนวิธี SIR สามารถสร้างช่วงความเชื่อมั่นได้ดีใกล้เคียงช่วงที่ดีแบบเบส์และแบบไฮบริดซึ่งเป็นหนึ่งในวิธีที่มีประสิทธิภาพสูงสุดจากการทบทวนวรรณกรรม

คำสำคัญ: ความน่าจะเป็นคุ้มรวม, ช่วงความเชื่อมั่นแบบวัลด์, ช่วงความเชื่อมั่นแบบไฮบริด, ช่วงความเชื่อมั่นแบบเบส์

การอ้างอิงบทความ: โชติกา วุฒิสาร รมิดา ศรีเหรา และ พัทธ์ชนก ศรีสุรเดชชัย, “ช่วงความเชื่อมั่นสำหรับความแตกต่างของสองสัดส่วนโดยขั้นตอนวิธีการเลือกตัวอย่างแบบเลือกซ้ำที่สำคัญ,” *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 28, ฉบับที่ 4, หน้า 859–868, ต.ค.–ธ.ค. 2561.

Research Article

Confidence Intervals for the Difference between Two Proportions by Sampling Importance Resampling Algorithm

Chotikar Vuthisarn, Ramidha Srihera and Patchanok Srisuradetchai*
Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Rangsit Campus, Pathum-Thani, Thailand

* Corresponding Author, Tel. 0–2564–4444 Ext. 2100, E–mail: spatchan@tu.ac.th
DOI: 10.14416/j.kmutnb.2018.08.003
Received 27 October 2017; Accepted 7 January 2518; Published online: 15 August 2018

Abstract

Confidence intervals for the difference between two proportions ($p_1 - p_2$) have gained interest by statisticians from past to present. Several researchers have proposed and compared the confidence intervals. Nowadays, computers play an important role in computational statistics. Methods of confidence interval constructions could employ algorithms instead of direct analytical method. In this research, Sampling Importance Resampling (SIR) algorithm frequently found in Bayesian statistics is applied to construct the confidence intervals which are then compared to those obtained from Bayes, hybrid, and Wald confidence intervals. The results show that the confidence intervals obtained from SIR give the coverage probabilities similar to those of Bayes and hybrid methods, and these 3 methods are superior to the Wald method in all population characteristics and sample sizes. Because SIR algorithm directly uses samples of p_1 and p_2, the confidence intervals of $p_1 - p_2$ are definitely in [–1, +1]. In case of $\hat{p}_1 = \hat{p}_2 = 0$, SIR algorithm is able to construct a confidence interval while the Wald interval is not when the standard error equals zero. The characteristic of population also affects the length of intervals. If p_1 and p_2 are close to .5 (symmetric), the interval length tends to be greater than that calculated from skewed population, for example $p_1 = p_2 = .05$. To conclude, in many situations, SIR algorithm is able to produce a good confidence interval similar to Bayes and hybrid method, one of the most efficient methods in literature review.

Keywords: Coverage Probability, Wald Confidence Interval, Hybrid Confidence Interval, Bayes Confidence Interval

Please cite this article as: C. Vuthisarn, R. Srihera, and P. Srisuradetchai, “Confidence intervals for the difference between two proportions by sampling importance resampling algorithm,” *The Journal of KMUTNB*, vol. 28, no. 4, pp. 859–868, Oct.–Dec. 2018 (in Thai).

1. บทนำ

การประมาณค่าแบบช่วง (Interval Estimation) สำหรับสัดส่วนประชากร (Proportion) นั้นเป็นที่สนใจและถูกนำไปศึกษาอย่างกว้างขวาง พารามิเตอร์สัดส่วน (p) ของการแจกแจงทวินามนี้แทน ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ และมีฟังก์ชันมวลความน่าจะเป็น $f(y) = {}^{n}C_{y}\, p^{y}(1-p), y = 0, 1,..., n$ โดยที่ตัวแปรสุ่ม Y แทน จำนวนครั้งของการเกิดสิ่งที่สนใจจากการทดลองสุ่มทั้งหมด n ครั้ง หรือ $Y \sim B(n, p)$ สำหรับช่วงความเชื่อมั่นที่มักพบในวิชาสถิติพื้นฐาน คือ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ที่ใช้การแจกแจงเชิงเส้นกำกับแบบปรกติมาตรฐาน $\hat{p} \pm Z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$ โดยที่ $\hat{p} = y/n$ และ $Z_{1-\alpha/2}$ แทน ควอนไทล์ที่ $1-\alpha/2$ ของการแจกแจงปรกติมาตรฐาน [1] และ Vollset [2] ได้ทำการศึกษาเปรียบเทียบ 7 วิธี ที่ใช้หาช่วงความเชื่อมั่นของสัดส่วนโดยใช้เกณฑ์ต่างๆ เช่น ความน่าจะเป็นคุ้มรวม (Coverage Probability) สัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient) ความกว้างของช่วง (Interval Width) ในการเปรียบเทียบ

ในกรณีของ 2 ประชากร โดยใช้แนวคิดของการแจกแจงเชิงเส้นกำกับ ถ้าให้ $Y_1 \sim B(n_1, p_1)$ และ $Y_2 \sim B(n_2, p_2)$ แล้ว ช่วงความเชื่อมั่นของ $p_1 - p_2$ ที่มีสัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)$ แสดงได้ดังสมการที่ (1)

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \qquad (1)$$

โดยที่ $\hat{p}_1 = y_1/n_1$ และ $\hat{p}_2 = y_2/n_2$ ช่วงความเชื่อมั่นนี้เป็นที่นิยมดังจะเห็นได้จากการปรากฏในซอฟต์แวร์ต่างๆ เช่น SPSS, SAS และ Minitab เป็นต้น อย่างไรก็ตาม ช่วงความเชื่อมั่นที่ใช้การแจกแจงเชิงเส้นกำกับนี้มีแนวโน้มที่ให้ความน่าจะเป็นคุ้มรวมต่ำและต้องใช้ตัวอย่างค่อนข้างใหญ่ [3] นอกจากนี้ยังมีการปรับให้ต่อเนื่องด้วยวิธีของ Yates [4] จะทำให้ได้ช่วงความเชื่อมั่นดังสมการที่ (2)

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\alpha/2} \times \left[\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right] \quad (2)$$

Beal [5] ได้เสนอ 2 ช่วงความเชื่อมั่นของ $p_1 - p_2$ ที่ใช้การแจกแจงเชิงเส้นกำกับเป็นพื้นฐาน โดยช่วงที่ได้นี้เป็นช่วงที่ได้จากการดัดแปลงจากช่วงความเชื่อมั่นของ Haldane [6] และ Good [7] และเมื่อเปรียบเทียบกับช่วงที่เสนอโดย Mee [8], Miettinen และ Nurminen [9] ผลการศึกษาพบว่า ช่วงใหม่ที่พัฒนาขึ้นให้ค่าความน่าจะเป็นคุ้มรวมสูงกว่าวิธีอื่นๆ ในบางกรณีที่ศึกษาโดยเฉพาะในกรณีที่ขนาดตัวอย่างเล็ก Beal กล่าวถึงวิธีของ Miettinen และ Nurminen ว่าเป็นวิธีที่ซับซ้อนที่ต้องเขียนโปรแกรมคอมพิวเตอร์ในการคำนวณ นอกจากนี้ Miettinen และ Nurminen [9] ยังได้เสนอช่วงความเชื่อมั่นที่ใช้ฟังก์ชันภาวะน่าจะเป็นโปรไฟล์ (Profile Likelihood) Newcombe [3] ได้ปรับปรุงช่วงความเชื่อมั่นที่ใช้ฟังก์ชันภาวะน่าจะเป็นโปรไฟล์ดังกล่าวนี้ โดยมีการคำนวณค่าความน่าจะเป็นทางปลายของการแจกแจงที่แท้จริง (Exact Distribution) และแบบที่เรียกว่า “mid-p” ซึ่งผู้สนใจสามารถอ่านเพิ่มเติมได้จาก Berry และ Armitage [10]

Newcombe [3] ได้ศึกษาเปรียบเทียบการประมาณค่าแบบช่วงสำหรับผลต่างของสัดส่วน โดยเปรียบเทียบทั้งหมด 11 วิธี ได้แก่ 1) วิธีช่วงความเชื่อมั่นแบบวัลด์ (Wald) 2) วิธีช่วงความเชื่อมั่นแบบวัลด์ที่มีการปรับค่าแบบต่อเนื่อง (Wald with Continuity Correction) 3) วิธีของ Beal’s Haldane 4) วิธีของ Beal’s Jeffreys-Perks 5) วิธีของ Mee 6) วิธีของ Miettinen และ Nurminen 7) วิธีที่ใช้ภาวะน่าจะเป็นโปรไฟล์ 8) วิธีที่ใช้ภาวะน่าจะเป็นโปรไฟล์ที่ขึ้นอยู่กับพื้นที่ปลายหางที่แท้จริง 9) วิธีที่ใช้ภาวะน่าจะเป็นโปรไฟล์ที่ขึ้นอยู่กับพื้นที่ปลายหางแบบ “mid-p” 10) วิธีไฮบริด (Hybrid Method) ที่ขึ้นอยู่กับฟังก์ชันสกอร์ของ Wilson และ 11) วิธีไฮบริดที่ขึ้นอยู่กับฟังก์ชันสกอร์ของ Wilson แบบปรับปรุง สองวิธีสุดท้ายนี้เป็นวิธี Newcombe พัฒนาขึ้นมา จากผลการศึกษาพบว่า วิธีที่ง่ายต่อการคำนวณมีแนวโน้มที่จะมีประสิทธิภาพต่ำ เมื่อพิจารณาค่าความน่าจะเป็นคุ้มรวมพบว่าวิธีที่ใช้ภาวะน่าจะเป็นโปรไฟล์แบบใช้พื้นที่ปลายหางให้ความน่าจะเป็นคุ้มรวมสูงสุด แต่วิธีดังกล่าวนี้ยากต่อการคำนวณ วิธีที่ใช้ฟังก์ชันสกอร์ (Score Function) หรือวิธีที่ 10) และ 11) เป็นวิธีที่มีประสิทธิภาพโดยรวมดีเมื่อพิจารณาปัจจัยอื่นๆ ประกอบกัน

เช่น ความซับซ้อนในการคำนวณ ค่าความน่าจะเป็นคุ้มรวม เป็นต้น

Agresti และ Coull [11] ได้ศึกษาช่วงความเชื่อมั่นที่แท้จริง (Exact Confidence Interval) ที่ได้จากการแปลงกลับของการทดสอบทวินาม (Binomial Test) ว่ามีแนวโน้มที่จะให้ค่าความน่าจะเป็นคุ้มรวมสูงเกินไปและช่วงแบบวัลด์ซึ่งขึ้นกับการแจกแจงเชิงเส้นกำกับว่ามีแนวโน้มที่จะให้ค่าความน่าจะเป็นคุ้มรวมต่ำเกินไป

Agresti และ Caffo [12] พบว่า วิธีการเพิ่ม 2 ค่าสังเกตเทียมที่สนใจ (Success) และ 2 ค่าสังเกตเทียมที่ไม่สนใจ (Failure) นั่นคือ $p_i = (y_i + 1)/(n_i + 2)$, $i = 1, 2$ จะให้ช่วงความเชื่อมั่นที่มีพฤติกรรมคล้ายคลึงกับช่วงที่ขึ้นอยู่กับฟังก์ชันสกอร์ของ Wilson นอกจากนี้ เมื่อพิจารณาช่วงความเชื่อมั่นแบบเบส์ซึ่งนำเสนอโดย Carlin และ Louis [13] ช่วงความเชื่อมั่นนี้ใช้การแจกแจงก่อนแบบเอกรูป (Uniform Distribution) แล้วตัวประมาณค่าแบบจุดของ $p_1 - p_2$ จะเหมือนกับตัวประมาณจากวิธีการเพิ่มค่าสังเกตเทียม แต่จะแตกต่างกันตรงตัวหารในความคลาดเคลื่อนมาตรฐาน ซึ่ง Agresti และ Caffo ใช้ $n_i + 2$ แต่ช่วงความเชื่อมั่นแบบเบส์ใช้ $n_i + 3$

Reed III [14] ได้ทำการศึกษาเปรียบเทียบช่วงความเชื่อมั่นทั้ง 4 วิธี ได้แก่ 1) วิธีช่วงความเชื่อมั่นแบบวัลด์ 2) วิธีช่วงความเชื่อมั่นแบบวัลด์ที่มีการปรับค่าแบบต่อเนื่อง 3) วิธีไฮบริดของ Newcombe [1] และ 4) วิธีของ Agresti-Caffo จากผลการศึกษาพบว่า ช่วงเชื่อมั่นแบบวัลด์ให้ค่าความน่าจะเป็นคุ้มรวมต่ำกว่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด Reed III แนะนำให้เลือกใช้วิธีไฮบริดของ Newcombe และของ Agresti-Caffo เนื่องจากให้ความน่าจะเป็นคุ้มรวมใกล้เคียงกับสัมประสิทธิ์ความเชื่อมั่นตามที่กำหนด

ขั้นตอนวิธีการเลือกตัวอย่างแบบเลือกซ้ำที่สำคัญ (Sampling Importance Resampling; SIR) ซึ่งนำเสนอโดย Rubin [15], [16] ขั้นตอนวิธีนี้มีจุดประสงค์หลักในการสร้างตัวอย่างซึ่งมีการแจกแจงภายหลัง (Posterior Distribution) ที่ยากในการวิเคราะห์โดยตรง โดยมีขั้นตอนดังนี้ ขั้นที่ 1 ประมาณการแจกแจงภายหลัง $p(\theta|\underset{\sim}{y})$ ด้วย $p^*(\theta|y)$ ซึ่งในบางครั้งจะระบุค่อนข้างยาก จึงมักประมาณด้วยการแจกแจงก่อน $p(\theta)$ ขั้นที่ 2 สุ่มตัวอย่าง $\theta_1,..., \theta_m$ จาก $p(\theta)$ ทั้งหมด m ตัว ขั้นที่ 3 คำนวณน้ำหนัก $w(\theta_i)$ แต่ละตัวจากสมการที่ (3)

$$\begin{aligned} w(\theta_i) &= \frac{p(\theta_i|\underset{\sim}{y})}{p^*(\theta|\underset{\sim}{y})} \approx \frac{1}{p(\theta_i)} \times p(\theta_i|\underset{\sim}{y}) \\ &= \frac{1}{p(\theta_i)} \times \frac{p(\theta_i) f(\underset{\sim}{y}|\theta)}{\int p(\theta_i) f(\underset{\sim}{y}|\theta) d\theta} \\ &= \frac{f(\underset{\sim}{y}|\theta)}{\int p(\theta_i) f(\underset{\sim}{y}|\theta) d\theta} \\ &= k\, L(\theta_i) \end{aligned} \tag{3}$$

โดยที่ $1 = 1, 2, ...m, k = \frac{1}{\int p(\theta_i) f(\underset{\sim}{y}|\theta) d\theta}$ ขั้นที่ 4 สุ่มตัวอย่างซ้ำแบบคืนที่ $\theta_1^*, \ldots, \theta_n^*$ มา n ตัว จากตัวอย่างเดิม $\theta_1,..., \theta_m$ ด้วยน้ำหนัก $w(\theta_1),..., w(\theta_m)$ ตามลำดับ ขั้นที่ 5 จะได้ตัวอย่างจากการแจกแจงภายหลังซึ่งคือ $\{\theta_1^*, \ldots, \theta_n^*\}$ ตัวอย่างนี้เสมือนว่ามีการแจกแจงโดยประมาณแบบการแจกแจงภายหลัง

ผู้วิจัยจึงเล็งเห็นว่า สามารถนำขั้นตอนวิธีนี้ไปขยายกับปัญหาในกรณีที่มีพารามิเตอร์ 2 ตัว ช่วงความเชื่อมั่นสำหรับความแตกต่างของสองสัดส่วนเป็นที่สนใจในสถิติมาช้านานจนกระทั่งถึงปัจจุบัน หากใช้ขั้นตอนวิธี SIR ซึ่งมีข้อดี คือ ใช้ตัวอย่างของพารามิเตอร์โดยตรง จะให้ขอบเขตบนและขอบเขตล่างของช่วงความเชื่อมั่น $p_1 - p_2$ ในช่วง $[-1, 1]$ เสมอ ซึ่งหลายช่วงความเชื่อมั่น ดังเช่นในสมการที่ (1) และ (2) ไม่ได้จำกัดอยู่ในช่วง $[-1, 1]$ นอกจากนี้ ในกรณีที่ค่าของตัวแปรสุ่ม y_1 และ y_2 (จากการแจกแจงทวินาม $B(n_1, p_1)$ และ $B(n_2, p_2)$ ตามลำดับ) มีค่าเป็นศูนย์ ช่วงความเชื่อมั่นแบบวัลด์จะไม่สามารถคำนวณได้และจากการทบทวนวรรณกรรมพบว่า วิธีไฮบริดของ Newcombe [3] เป็นวิธีที่ดีทั้งในแง่ของความน่าจะเป็นคุ้มรวมและความไม่ซับซ้อนในการคำนวณ จึงจะนำมาเปรียบเทียบกับช่วงที่ได้จากขั้นตอนวิธี SIR นอกจากนี้ วิธีแบบเบส์ยังไม่เคยถูกนำมาเปรียบเทียบกับวิธีอื่นๆ จึงนำมาศึกษาในครั้งนี้ด้วย รวมถึงช่วงความเชื่อมั่นแบบวัลด์ซึ่งเป็นที่นิยมและพบได้ในโปรแกรมสำเร็จรูปทางสถิติทั่วไป

2. วิธีการวิจัย

กำหนดให้ $Y_1 \sim B(n_1, p_1)$ และ $Y_2 \sim B(n_2, p_2)$ โดยที่ $\hat{p}_1$ และ $\hat{p}_2$ คือ สัดส่วนตัวอย่างที่มีลักษณะที่สนใจจากประชากรกลุ่มที่ 1 และกลุ่มที่ 2 และ n_1 และ n_2 คือ ขนาดตัวอย่างของกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ ในการศึกษาจะมีขั้นตอนหลักๆ เป็นดังนี้

1) ทบทวนวรรณกรรมที่เกี่ยวข้องเกี่ยวกับช่วงความเชื่อมั่นของผลต่างของสองสัดส่วน

2) ศึกษาแนวคิดของขั้นตอนวิธี SIR พร้อมทั้งนำเสนอขั้นตอนวิธีการหาช่วงความเชื่อมั่นของผลต่างของสองสัดส่วนโดยใช้แนวคิดของ SIR

3) ศึกษาเชิงการจำลองโดยแบ่งเป็น 4 สถานการณ์ ดังนี้

สถานการณ์ที่ 1: ผลต่าง $p_1 - p_2$ มีค่าน้อยมากและประชากรทั้งสองสมมาตร กล่าวคือ ให้ $p_1 = p_2 = 0.5$

สถานการณ์ที่ 2: ผลต่าง $p_1 - p_2$ มีค่าน้อยมากและประชากรทั้งสองเบ้ กล่าวคือ ให้ $p_1 = p_2 = 0.05$

สถานการณ์ที่ 3: ผลต่าง $p_1 - p_2$ มีค่าในระดับกลาง กล่าวคือ ให้ $p_1 = 0.75, p_2 = 0.25$

สถานการณ์ที่ 4: ผลต่าง $p_1 - p_2$ มีค่าในระดับสูง กล่าวคือ ให้ $p_1 = 0.95, p_2 = 0.05$

ในแต่ละสถานการณ์กำหนดขนาดตัวอย่าง n_i ที่ใช้กรณีศึกษาของทั้งสองประชากรมีค่าเป็น 10, 30, 50, 100

4) จำลองค่า y_1 และ y_2 จาก $B(n_1, p_1)$ และ $B(n_2, p_2)$ แล้วคำนวณช่วงความเชื่อมั่น 95% (α = .05) ของผลต่างของสัดส่วนตามสูตร ดังนี้

วิธีวัลด์: $$\left(\hat{p}_1 - \hat{p}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1} + \frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}$$

วิธีเบส์: $$\left(\hat{p}_1 - \hat{p}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{V(\hat{p}_1) + V(\hat{p}_2)} \qquad [13]$$

โดยที่ $\hat{p}_i = \dfrac{y_i + 1}{n_i + 2}$ และ $V\left(\hat{p}_i\right) = \dfrac{\hat{p}_i\left(1-\hat{p}_i\right)}{n_i + 3}, i = 1, 2$

วิธีไฮบริด:

$$LCL = \left(\hat{p}_1 - \hat{p}_2\right) - Z_{\frac{\alpha}{2}} \sqrt{\frac{l_1\left(1-l_1\right)}{n_1} + \frac{u_2\left(1-u_2\right)}{n_2}}$$

$$UCL = \left(\hat{p}_1 - \hat{p}_2\right) + Z_{\frac{\alpha}{2}} \sqrt{\frac{u_1\left(1-u_1\right)}{n_1} + \frac{l_2\left(1-l_2\right)}{n_2}} \qquad [3]$$

โดยที่ค่า l_i และ u_i คือ p_i ค่าที่ได้จากการแก้สมการ

$$Z_{\frac{\alpha}{2}} = \frac{\left|\hat{p}_i - p_i\right|}{\sqrt{\dfrac{p_i\left(1-p_i\right)}{n_i}}}$$

และ $u_i > l_i, i = 1, 2$

5) เปรียบเทียบความยาวช่วง ค่าประมาณความน่าจะเป็นคุ้มรวม (Coverage Probability; CP) และเปอร์เซ็นต์ของขอบเขตบนและล่างของช่วงที่เกิน +1 และต่ำกว่า –1 ที่ได้จากวิธีวัลด์ วิธีเบส์ วิธีไฮบริดและขั้นตอนวิธี SIR ที่นำเสนอขึ้นในงานวิจัยครั้งนี้ โดยใช้การจำลองมอนติคาร์โลซึ่งมีจำนวนทำซ้ำ 5,000 ครั้ง

6) ทดสอบว่าช่วงที่ได้จะ CP เท่ากับ .95 หรือไม่นั้น โดยพิจารณาจากทดสอบสมมติฐาน $H_0: \mathrm{CP} \geq 0.95$ ซึ่งเป็นไปตามนิยามที่ให้ไว้ใน Rohde [17] ในที่นี้กำหนดระดับนัยสำคัญเท่ากับ .01 หากค่า CP จากการจำลองมอนติคาร์โลมากกว่าหรือเท่ากับ $0.95 - z_{.99}\sqrt{\dfrac{0.95(0.05)}{5000}} \approx 0.94$ จะไม่สามารถปฏิเสธสมมติฐานว่างได้ และสรุปว่าช่วงความเชื่อมั่นนี้มี CP เท่ากับ .01 ที่ระดับนัยสำคัญ

3. ผลการวิจัย

ผลการวิจัยจะแบ่งออกเป็น 2 ส่วน คือ ช่วงความเชื่อมั่นที่ใช้แนวคิดขั้นตอนวิธี SIR ที่นำเสนอและผลการศึกษาเชิงจำลอง

3.1 ขั้นตอนวิธีที่ใช้หาช่วงความเชื่อมั่น

นำเข้า:

1) ค่าสังเกต y_1 และ y_2 จากการทดลองทวินาม $B(n_1, p_1)$ และ $B(n_2, p_2)$ โดยที่ n_1 และ n_2 เป็นค่าที่ทราบ และ p_1 และ p_2 เป็นพารามิเตอร์ที่สนใจที่ไม่ทราบ

2) การแจกแจงก่อนของ p_1 และ p_2 ซึ่งกำหนดเป็นการแจกแจงเอกรูปต่อเนื่องบนช่วง 0 ถึง 1

ขั้นตอนวิธี:

1) สุ่มตัวอย่าง $\{(p_1, p_2)_1,..., (p_1, p_2)_m\}$ จากการแจกแจงก่อน โดยที่ p_i และ p_j ใดๆ เป็นอิสระต่อกัน

2) คำนวณน้ำหนักให้แต่ละ $(p_1, p_2)_i$ จาก

$$w\left(\left(p_1, p_2\right)_i\right) = \frac{L\left(\left(p_1, p_2\right)_i\right)}{\sum_{i=1}^{m} L\left(\left(p_1, p_2\right)_i\right)} \quad (4)$$

โดยที่ $L(p_1, p_2) = p_1^{y_1}(1-p_1)^{n_1-y_1}\, p_2^{y_2}(1-p_2)^{n_2-y_2}$

3) สุ่มตัวอย่าง $\{(p_1, p_2)_1^*, \ldots, (p_1, p_2)_n^*\}$ (ขนาด n) แบบคืนที่จาก $\{(p_1, p_2)_1, \ldots, (p_1, p_2)_m\}$ ด้วยความน่าจะเป็นตามสมการที่ (4)

4) คำนวณ $\{(p_1 - p_2)_1^*, \ldots, (p_1 - p_2)_n^*\}$ แล้วหาควอนไทล์ที่ $\alpha/2$ และ $1 - \alpha/2$ ซึ่งคือ ขอบล่างและบนของช่วงความเชื่อมั่น $(1 - \alpha)100\%$

นำออก: ขอบล่างและบนของช่วงความเชื่อมั่น $(1 - \alpha)100\%$

ในขั้นตอนวิธีข้างต้นนี้ ผู้วิจัยได้ทดลองหาค่า (m, n) ที่เหมาะสมโดยลองผิดลองถูก (Trial and Error) พบว่า $m = 5{,}000$ และ $n = 3{,}000$ จะให้ช่วงความเชื่อมั่นที่มีความน่าเป็นจะคุ้มรวมสูงแต่ใช้เวลาไม่มาก (น้อยกว่า 10 วินาที)

3.2 ผลการศึกษาเชิงจำลอง

ในการเปรียบเทียบค่าประมาณความน่าจะเป็นคุ้มรวม ความยาวของช่วงและเปอร์เซ็นต์ที่ขอบเขตล่างและบนออกนอกช่วง [−1, +1] โดยได้ทำการทดลองมอนติคาร์โลในทั้ง 4 สถานการณ์ สรุปได้ดังนี้

ในกรณีผลต่าง $p_1 - p_2 = 0$ (มีค่าน้อยมาก) และประชากรทั้งสองสมมาตรหรือทั้ง p_1 และ p_2 เท่ากับ 0.5 พบว่าช่วงแบบวัลด์เป็นวิธีเดียวที่ให้ความน่าจะเป็นคุ้มรวมต่ำกว่าที่กำหนด ในกรณีขนาดตัวอย่างของอย่างน้อย 1 กลุ่ม มีขนาดเล็ก ($n_i = 10$) ดังที่ขีดเส้นใต้ไว้ในตารางที่ 1 กล่าวคือ CP ยังมีไม่เท่ากับ .95 เมื่อเทียบกับ .94 ที่ระดับนัยสำคัญ .01 ตามที่ระบุในขั้นตอนการศึกษา หรือกล่าวได้ว่ายังไม่ผ่านเกณฑ์ที่กำหนด โดยทั่วไปหาก $\min(n_1, n_2) \geq 30$ ทั้ง 4 วิธีจะให้ช่วงที่มีความน่าจะเป็นคุ้มรวมใกล้เคียง .95 โดยวิธีวัลด์มีแนวโน้มที่จะให้ช่วงที่มีความน่าจะเป็นคุ้มรวมต่ำที่สุดและมีความยาวของช่วงมากที่สุด ส่วนวิธี SIR ให้ช่วงที่มีความน่าจะเป็นคุ้มรวมสูงกว่าวิธีอื่นๆ และความยาวช่วงน้อยเป็นอันดับที่ 2 รองจากวิธีไฮบริดที่ให้ความยาวช่วงต่ำสุด นอกจากนี้ทั้ง 4 วิธีให้ช่วงที่อยู่ใน [−1, +1] ในทุกกรณี

สำหรับกรณีที่ผลต่าง $p_1 - p_2 = 0$ (มีค่าน้อยมาก) แต่ประชากรในทั้งสองเบ้รุนแรง ($p_1 = p_2 = .05$) ผลการจำลองสรุปดังในตารางที่ 2 ในกรณีนี้จะพบว่า ช่วงแบบวัลด์มีความน่าจะเป็นคุ้มรวมต่ำกว่า .95 ในหลายกรณี ส่วนอีก 3 วิธีที่เหลือส่วนใหญ่ให้ค่าของ CP เท่ากับ .95 ถึงแม้ตัวอย่างของทั้ง 2 กลุ่มมีขนาดเล็ก โดยวิธีเบส์มีแนวโน้มให้ค่า CP สูงที่สุดและมีความยาวช่วงต่ำที่สุดในหลายกรณีที่ศึกษา จะเห็นว่าหาก $\min(n_1, n_2) \geq 30$ ทุกวิธี ยกเว้นวิธีวัลด์จะให้ช่วงที่มี CP เท่ากับ .95 ตามที่ต้องการ ซึ่งในทุกกรณีทั้ง 4 วิธีจะให้ช่วงที่อยู่ใน [−1, +1]

ในกรณีที่ผลต่าง $p_1 - p_2 = .05$ และประชากรทั้งสองเบ้ ($p_1 = 0.75, p_2 = 0.25$) ผลการจำลองสรุปในตารางที่ 3 วิธีวัลด์ยังต้องใช้ตัวอย่างขนาดใหญ่ในทั้ง 2 กลุ่มหรือ $\min(n_1, n_2) \geq 50$ จึงจะให้ CP เท่ากับ .95 สำหรับ 3 วิธีที่เหลือให้ CP เท่ากับ .95 ถึงแม้ตัวอย่างในทั้ง 2 กลุ่มจะมีขนาดเล็กก็ตาม CP ที่ได้นี้มีค่าใกล้เคียงกันเช่นเดียวกับความยาวของช่วงความเชื่อมั่นและทั้ง 4 วิธีให้ช่วงที่อยู่ใน [−1, +1] ในทุกกรณีที่ศึกษา

เนื่องจากช่วงความเชื่อมั่นสำหรับความแตกต่างของสองสัดส่วนที่ได้จากทั้ง 4 วิธี ในสถานการณ์ที่ 1–3 ให้ช่วงที่อยู่ใน [−1, +1] ทุกกรณี ดังนั้น ในตารางที่ 1 ถึงตารางที่ 3 จึงนำเสนอเพียงค่าความน่าจะเป็นคุ้มรวมและความยาวของช่วงเท่านั้น

ในสถานการณ์สุดท้าย ผลการจำลองได้สรุปดังในตารางที่ 4 ในกรณีพารามิเตอร์ที่แท้จริง $p_1 - p_2 = .9$ เข้าใกล้ 1 จะทำให้ขอบเขตบนที่ได้ในหลายวิธีจะมากกว่าหนึ่ง เช่น วิธีเบส์ กรณี ($n_1 = 10, n_2 = 30$) ถึงแม้ค่าของ CP จะมากกว่า .95 อยู่เพียงวิธีเดียวแต่ขอบบนมากกว่าหนึ่งมากถึง 55% วิธีวัลด์ยังคงเป็นวิธีที่แย่กว่าวิธีอื่นๆ อยู่มาก กล่าวคือ ค่าของ CP

ตารางที่ 1 ค่าประมาณความน่าจะเป็นคุ้มรวม (CP) ความยาว (Length) ของช่วงความเชื่อมั่น 95% สำหรับสถานการณ์ที่ 1

$p_1 = 0.5, p_2 = 0.5$				
n_1	n_2	วิธี	CP	Length
10	10	Wald	0.9070	0.8295
		Hybrid	0.9586	**0.7228**
		Bayes	0.9586	0.7407
		SIR	0.9587	0.7345
	30	Wald	0.9209	0.6843
		Hybrid	0.9476	**0.5921**
		Bayes	0.9459	0.6226
		SIR	0.9553	0.6158
	50	Wald	0.9141	0.6477
		Hybrid	0.9430	**0.5602**
		Bayes	0.9420	0.5880
		SIR	0.9541	0.5798
	100	Wald	0.9077	0.6176
		Hybrid	0.9514	**0.5343**
		Bayes	0.9345	0.5578
		SIR	0.9581	0.5477
30	30	Wald	0.9537	0.4974
		Hybrid	0.9540	**0.4698**
		Bayes	0.9540	0.4753
		SIR	0.9542	0.4731
	50	Wald	0.9452	0.4462
		Hybrid	0.9503	**0.4211**
		Bayes	0.9503	0.4291
		SIR	0.9538	0.4270
	100	Wald	0.9364	0.4020
		Hybrid	0.9452	**0.3794**
		Bayes	0.9429	0.3870
		SIR	0.9474	0.3846
50	50	Wald	0.9445	0.3880
		Hybrid	0.9445	**0.3743**
		Bayes	0.9445	0.3772
		SIR	0.9501	0.3755
	100	Wald	0.9467	0.3367
		Hybrid	0.9469	**0.3247**
		Bayes	0.9469	0.3288
		SIR	0.9538	0.3273
100	100	Wald	0.9423	0.2758
		Hybrid	0.9423	**0.2707**
		Bayes	0.9423	0.2718
		SIR	0.9472	**0.2707**

หมายเหตุ: ขีดเส้นใต้ หมายถึง CP < .94 ที่ระดับนัยสำคัญ .01
ตัวหนา หมายถึง ความยาวของช่วงต่ำสุด

ตารางที่ 2 ค่าประมาณความน่าจะเป็นคุ้มรวม (CP) ความยาว (Length) ของช่วงความเชื่อมั่น 95% สำหรับสถานการณ์ที่ 2

$p_1 = 0.05, p_2 = 0.05$				
n_1	n_2	วิธี	CP	Length
10	10	Wald	0.6310	0.2850
		Hybrid	0.9986	0.5910
		Bayes	0.9986	**0.4968**
		SIR	0.9986	0.5176
	30	Wald	0.8383	0.2516
		Hybrid	0.9940	0.4700
		Bayes	0.9940	0.3935
		SIR	0.9770	**0.3991**
	50	Wald	0.8092	0.2337
		Hybrid	0.9837	0.4340
		Bayes	0.9934	0.3738
		SIR	0.9587	**0.3719**
	100	Wald	0.5558	0.2162
		Hybrid	0.9688	0.3997
		Bayes	0.9915	0.3614
		SIR	0.9428	**0.3525**
30	30	Wald	0.9245	0.2052
		Hybrid	0.9930	0.2875
		Bayes	0.9930	**0.2533**
		SIR	0.9848	0.2595
	50	Wald	0.9223	0.1867
		Hybrid	0.9889	0.2566
		Bayes	0.9889	**0.2230**
		SIR	0.9732	0.2267
	100	Wald	0.8335	0.1662
		Hybrid	0.9860	0.2260
		Bayes	0.9883	0.1993
		SIR	0.9603	**0.1991**
50	50	Wald	0.9428	0.1651
		Hybrid	0.9799	0.2063
		Bayes	0.9799	**0.1875**
		SIR	0.9693	0.1903
	100	Wald	0.9160	0.1435
		Hybrid	0.9757	0.1776
		Bayes	0.9774	**0.1599**
		SIR	0.9566	0.1603
100	100	Wald	0.9470	**0.1186**
		Hybrid	0.9641	0.1345
		Bayes	0.9659	0.1267
		SIR	0.9507	0.1267

หมายเหตุ: ขีดเส้นใต้ หมายถึง CP < .94 ที่ระดับนัยสำคัญ .01
ตัวหนา หมายถึง ความยาวของช่วงต่ำสุด

ตารางที่ 3 ค่าประมาณความน่าจะเป็นคุ้มรวม (CP) ความยาว (Length) ของช่วงความเชื่อมั่น 95% สำหรับสถานการณ์ที่ 3

$p_1 = 0.75, p_2 = 0.25$				
n_1	n_2	วิธี	CP	Length
10	10	Wald	0.8937	0.7112
		Hybrid	0.9570	**0.6575**
		Bayes	0.9556	0.6738
		SIR	0.9554	0.6650
	30	Wald	0.8923	0.5848
		Hybrid	0.9341	0.5369
		Bayes	0.9505	0.5605
		SIR	0.9474	**0.5534**
	50	Wald	0.8903	0.5502
		Hybrid	0.9428	**0.5070**
		Bayes	0.9506	0.5288
		SIR	0.9546	0.5210
	100	Wald	0.9135	0.5271
		Hybrid	0.9512	**0.4851**
		Bayes	0.9508	0.5050
		SIR	0.9582	0.4953
30	30	Wald	0.9388	0.4297
		Hybrid	0.9450	**0.4143**
		Bayes	0.9454	0.4194
		SIR	0.9468	0.4165
	50	Wald	0.9294	0.3849
		Hybrid	0.9474	**0.3708**
		Bayes	0.9448	0.3770
		SIR	0.9472	0.3745
	100	Wald	0.9259	0.3467
		Hybrid	0.9463	**0.3340**
		Bayes	0.9443	0.3399
		SIR	0.9496	0.3371
50	50	Wald	0.9470	0.3354
		Hybrid	0.9498	**0.3276**
		Bayes	0.9514	0.3303
		SIR	0.9471	0.3283
	100	Wald	0.9418	0.2911
		Hybrid	0.9486	**0.2843**
		Bayes	0.9476	0.2874
		SIR	0.9483	0.2856
100	100	Wald	0.9421	0.2386
		Hybrid	0.9505	**0.2357**
		Bayes	0.9498	0.2367
		SIR	0.9487	**0.2352**

หมายเหตุ: ขีดเส้นใต้ หมายถึง CP < .94 ที่ระดับนัยสำคัญ .01
ตัวหนา หมายถึง ความยาวของช่วงต่ำสุด

ตารางที่ 4 ค่าประมาณความน่าจะเป็นคุ้มรวม (CP) ความยาว (Length) และเปอร์เซ็นต์ของขอบเขตบนที่มากกว่าหนึ่ง (% > 1) ของช่วงความเชื่อมั่น 95% สำหรับสถานการณ์ที่ 4

$p_1 = 0.95, p_2 = 0.05$					
n_1	n_2	วิธี	CP	Length	% > 1
10	10	Wald	0.6400	0.2842	61.51
		Hybrid	0.9275	0.4804	35.72
		Bayes	0.9275	0.4966	73.62
		SIR	0.8108	0.4804	0.00
	30	Wald	0.6632	0.2505	76.86
		Hybrid	0.8899	0.3911	12.96
		Bayes	0.9692	0.3933	55.28
		SIR	0.8648	0.3824	0.00
	50	Wald	0.6730	0.2339	75.72
		Hybrid	0.9183	0.3693	4.47
		Bayes	0.9787	0.3738	54.34
		SIR	0.8805	0.3629	0.00
	100	Wald	0.6211	0.2106	49.49
		Hybrid	0.9307	0.3507	0.55
		Bayes	0.9856	0.3592	54.81
		SIR	0.9051	0.3468	0.00
30	30	Wald	0.8000	0.2051	59.66
		Hybrid	0.9216	0.2506	4.82
		Bayes	0.9694	0.2532	19.67
		SIR	0.9038	0.2466	0.00
	50	Wald	0.8531	0.1875	40.50
		Hybrid	0.9167	0.2251	1.67
		Bayes	0.9693	0.2235	8.52
		SIR	0.9014	0.2179	0.00
	100	Wald	0.8609	0.1659	23.23
		Hybrid	0.9320	0.2012	0.16
		Bayes	0.9726	0.1992	4.14
		SIR	0.9171	0.1940	0.00
50	50	Wald	0.8745	0.1655	24.91
		Hybrid	0.9338	0.1864	0.47
		Bayes	0.9662	0.1879	3.49
		SIR	0.9149	0.1833	0.00
	100	Wald	0.9188	0.1435	10.39
		Hybrid	0.9354	0.1612	0.01
		Bayes	0.9651	0.1600	1.50
		SIR	0.9232	0.1559	0.00
100	100	Wald	0.9231	0.1190	0.79
		Hybrid	0.9452	0.1266	0.00
		Bayes	0.9664	0.1272	0.03
		SIR	0.9208	0.1241	0.00

ต่ำกว่าวิธีอื่นๆ และเปอร์เซ็นต์ของขอบเขตบนที่มากกว่าหนึ่งสูง ในขณะที่ขอบเขตบนของช่วงจากวิธี SIR มีค่าไม่เกิน 1 แต่ CP ยังมีไม่เท่ากับ .95 (เทียบกับ .94 ใช้ระดับนัยสำคัญ .01 ตามที่ระบุในขั้นตอนการศึกษา) ถึงแม้ขนาดตัวอย่างจะเท่ากับ $n_1 = 100$, $n_2 = 100$ โดยสรุป วิธีเบส์มีแนวโน้มให้ CP สูงกว่าวิธีไฮบริด แต่เปอร์เซ็นต์ที่ขอบเขตบนที่มากกว่าหนึ่งก็มากกว่าเช่นกัน แต่เมื่อขนาดตัวอย่างทั้ง 2 มีใหญ่ขึ้นเท่ากับ 100 ทั้งสองวิธีนี้ก็ให้ CP ใกล้เคียง .95

4. อภิปรายและสรุป

การปรับขั้นตอนวิธี SIR เพื่อนำไปใช้หาช่วงความเชื่อมั่นของผลต่างของสองสัดส่วนนั้น โดยภาพรวม (สถานการณ์ที่ 1, 2 และ 3) ช่วงเชื่อมั่นที่ได้มีค่าประมาณความน่าจะเป็นคุ้มรวมไม่แตกต่างไปจากวิธีเบส์และวิธีไฮบริดซึ่งเป็นวิธีที่พบว่ามีประสิทธิภาพสูงจากการศึกษาเปรียบเทียบ 11 วิธีที่ใช้สร้างช่วงโดย Newcombe [3] ความยาวช่วงของทั้งสามวิธีนี้ก็ไม่แตกต่างกันมากนัก สำหรับสถานการณ์ที่ 4 หากขนาดตัวอย่างไม่ใหญ่มาก ยังไม่มีวิธีใดที่มีคุณสมบัติที่ดีในทุกๆ ด้าน และเมื่อตัวอย่างมีขนาดใหญ่มากวิธีเบส์จะให้ผลที่ดีที่สุด เมื่อพิจารณาลักษณะของประชากรที่มีต่อความยาวช่วงพบว่า หากทั้ง p_1 และ p_2 ใกล้ 0.5 (สมมาตร) ความยาวของช่วงความเชื่อมั่นมีแนวโน้มที่จะสูงกว่าในประชากรที่เบ้ เช่น $p_1 = .05$ และ $p_2 = .05$ โดยสรุป ขั้นตอนวิธี SIR สามารถสร้างช่วงความเชื่อมั่นได้ดีใกล้เคียงช่วงที่ดีแบบไฮบริดและเบส์ได้ในหลายกรณีที่ศึกษา

เอกสารอ้างอิง

[1] R. G. Newcombe, "Two-sided confidence intervals for the single proportion: Comparison of seven methods," *Statistics in Medicine*, vol. 17, pp. 857–872, 1998.

[2] S. E. Vollset, "Confidence intervals for a binomial proportion," *Statistics in Medicine*, vol. 12, pp. 809–824, 1993.

[3] R. G. Newcombe, "Interval estimation for the difference between independent proportions: Comparison of eleven methods," *Statistics in Medicine*, vol. 17, pp. 873–890, 1998.

[4] M. G. Haviland, "Yates's correction for continuity and the analysis of 2 × 2 contingency tables," *Statistics in Medicine*, vol. 9, pp. 363–367, 1990.

[5] S. L. Beal, "Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples," *Biometrics*, vol. 43, pp. 941–950, 1987.

[6] J. B. S. Haldane, "On a Method of Estimating Frequencies," *Biometrika*, vol. 33, pp. 222–225, 1945.

[7] I. J. Good, *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Massachusetts: The MIT Press, 1965.

[8] R. Mee, "Confidence bounds for the difference between two probabilities," *Biometrics*, vol. 40, pp. 1175–1176, 1984.

[9] O. Miettinen and M. Nurminen, "Comparative analysis of two rates," *Statistics in Medicine*, vol. 4, pp. 213–226, 1984.

[10] G. Berry and P. Armitage, "Mid-P confidence intervals: A brief review," *The Statistician*, vol. 44, no. 4, pp. 417–423, 1995.

[11] A. Agresti and B. A. Coull, "Approximate is better than 'exact' for interval estimation of binomial proportions," *The American Statistician*, vol. 52, pp. 119–126, 1998.

[12] A. Agresti and B. Caffo, "Simple and effective confidence intervals for proportions result from adding two successes and two failures," *The American Statistician*, vol. 54, pp. 280–288, 2000.

[13] B. P. Carlin and T. A. Louis, *Bayesian Methods*

for Data Analysis. Florida: Chapman & Hall/CRC, 1996.

[14] J. F. Reed III, “Improved confidence intervals for the difference between two proportions,” *Journal of Modern Applied Statistical Methods*, vol. 8, no. 1, pp. 208–214, 2009.

[15] D. B. Rubin, “A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. Discussion of tanner and wong,” *Journal of the American Statistical Association*, vol. 85, pp. 543–546, 1987.

[16] D. B. Rubin, *Using the SIR algorithm to simulate posterior distributions*. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. Smith. Bayesian Statistics, Clarendon, Oxford, 1988.

[17] C. A. Rohde, *Introductory Statistical Inference with the Likelihood Function*, 1st ed. London: Springer, 2014.