การทดสอบภาวะสารูปดีสำหรับการแจกแจงปัวซงกรณีที่ไม่ทราบค่าพารามิเตอร์โดยใช้ตารางค่าวิกฤตที่สร้างใหม่

Goodness of Fit Tests for the Poisson Distribution with Unknown Parameter Using the New Critical Value Tables

วนิดา พงษ์ศักดิ์ชาติ*

Vanida Pongsakchat*

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

Department of Mathematics, Faculty of Science, Burapha University

Received : 24 July 2022

Revised : 25 December 2022

Accepted : 8 January 2023

บทคัดย่อ

การศึกษานี้ได้สร้างตารางค่าวิกฤตของการทดสอบภาวะสารูปดีสำหรับการแจกแจงปัวซงที่ไม่ทราบค่าพารามิเตอร์ โดยการทดสอบที่ศึกษาคือการทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง และศึกษาประสิทธิภาพของการทดสอบทั้งสามวิธีนี้เมื่อใช้ตารางค่าวิกฤตที่สร้างขึ้นใหม่โดยพิจารณาจากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และค่าประมาณกำลังการทดสอบ ผลการศึกษาพบว่าการทดสอบคอลโมโกรอฟ-สมีร์นอฟไม่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ดีเท่าการทดสอบคราเมอร์-ฟอน-มิสส์และการทดสอบแอนเดอร์สัน-ดาร์ลิง โดยที่การทดสอบคราเมอร์-ฟอน-มิสส์เป็นการทดสอบที่ให้ค่าประมาณกำลังการทดสอบสูงที่สุดในเกือบทุกกรณี และค่าประมาณกำลังการทดสอบของการทดสอบทั้ง 3 วิธี จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น

คำสำคัญ : การแจกแจงปัวซง ; การทดสอบภาวะสารูปดี ; ค่าวิกฤต ; กำลังการทดสอบ ; ความผิดพลาดแบบที่ 1

Abstract

This study aims to establish the critical values tables for Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling tests for the Poisson distribution with unknown parameter. Additionally, examined how well these tests performed when using the new critical value tables based on type I error rate and power of the test. The results showed that the Kolmogorov-Smirnov test could not control type I error as well as the Cramer-von Mises and Anderson-Darling tests. In almost every situation, the Cramer-von Mises test provided the most powerful test. Furthermore, as the sample size increase, the power of each test increases.

Keywords : Poisson distribution ; Goodness of fit test ; Critical values ; Power of the test ; Type I error

***Corresponding author. E-mail** : vanida@buu.ac.th

บทนำ

การทดสอบภาวะสารูปดีเป็นวิธีเชิงสถิติที่ใช้สำหรับตรวจสอบการแจกแจงความน่าจะเป็นของข้อมูล ซึ่งวิธีที่นิยมใช้และเป็นที่รู้จักกันดี ได้แก่ การทดสอบคอลโมโกรอฟ-สมีร์นอฟ (Kolmogorov-Smirnov test) การทดสอบคราเมอร์-ฟอน-มิสส์ (Cramer-von Mises test) และการทดสอบแอนเดอร์สัน-ดาร์ลิง (Anderson-Darling test) ซึ่งเป็นการทดสอบที่ถูกนำมาใช้ตรวจสอบข้อมูลที่มีการแจกแจงต่อเนื่อง (Continuous distribution) เช่น การแจกแจงปรกติ (Normal distribution) การแจกแจงแกมมา (Gamma distribution) และการแจกแจงแบบเลขชี้กำลัง (Exponential Distribution) เป็นต้น (Gibbons & Chakraborti, 2003) อย่างไรก็ตามการทดสอบทั้งสามชนิดนี้ได้รับการพัฒนาให้สามารถนำมาใช้ได้กับข้อมูลที่มีการแจกแจงไม่ต่อเนื่อง (Discrete distribution)

จากงานวิจัยที่ผ่านมาได้มีการศึกษาและพัฒนาการทดสอบภาวะสารูปดีสำหรับการแจกแจงไม่ต่อเนื่องมาอย่างต่อเนื่อง โดยการทดสอบคอลโมโกรอฟ-สมีร์นอฟเป็นการทดสอบชนิดแรกที่ได้รับการพัฒนาให้ใช้ได้กับข้อมูลที่มีการแจกแจงไม่ต่อเนื่อง โดย Conover (1972) ได้หาวิธีการคำนวณค่าสถิติทดสอบและความน่าจะเป็นของการปฏิเสธสมมุติฐานว่างที่แท้จริง (Exact critical level) สำหรับการทดสอบคอลโมโกรอฟ-สมีร์นอฟ โดยเป็นการทดสอบสมมุติฐานแบบทางเดียว ต่อมา Horn (1977) ได้นำวิธีของ Conover (1972) มาอธิบายเพิ่มเติม และ Petitt & Stephens (1977) ได้พัฒนาการทดสอบชนิดนี้ให้เป็นการทดสอบสมมุติฐานแบบสองทาง Choulakian, Lockhart & Stephens (1994) ได้พัฒนาการทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิงสำหรับตรวจสอบข้อมูลที่มีการแจกแจงไม่ต่อเนื่อง และ Spinelli & Stephens (1997) ได้นำการทดสอบทั้งสองชนิดนี้มาใช้สำหรับการแจกแจงปัวซง ต่อมา Lockhart, Spinelli & Stephens (2007) ได้นำการทดสอบของ Choulakian, Lockhart & Stephens (1994) มาประยุกต์ใช้ในกรณีที่ไม่ทราบค่าพารามิเตอร์ของการแจกแจงปัวซง

เนื่องจากการทดสอบคอลโมโกรอฟ-สมีร์นอฟ คราเมอร์-ฟอน-มิสส์ และแอนเดอร์สัน-ดาร์ลิง ที่ถูกพัฒนามาใช้สำหรับตรวจสอบข้อมูลที่มีการแจกแจงไม่ต่อเนื่องจะต้องมีการกำหนดค่าพารามิเตอร์ลงในฟังก์ชันการแจกแจงความน่าจะเป็นสะสมของการแจกแจงที่ต้องการทดสอบ ซึ่ง Lockhart, Spinelli & Stephens (2007) ได้ดัดแปลงการทดสอบนี้ให้สามารถนำมาใช้ในกรณีที่มีการประมาณพารามิเตอร์ได้ และได้เสนอการแจกแจงเชิงเส้นกำกับ (Asymptotic distribution) ของการทดสอบเหล่านี้ อย่างไรก็ตามการหาความน่าจะเป็นของการปฏิเสธสมมุติฐานว่างหรือค่าพี (p-value) ต้องอาศัยความรู้ทางคณิตศาสตร์ค่อนข้างมาก ดังนั้นการนำการทดสอบเหล่านี้ไปใช้จึงเป็นไปได้ยากสำหรับผู้ศึกษาหรือนักวิจัยสาขาอื่น ๆ ที่ไม่ได้มีความรู้ทางคณิตศาสตร์มากพอ

สำหรับการแจกแจงไม่ต่อเนื่องที่พบได้มากในการศึกษาและงานวิจัยในหลายสาขาได้แก่การแจกแจงปัวซง (Poisson distribution) โดยการแจกแจงชนิดนี้พบในข้อมูลที่แสดงจำนวนครั้งของการเกิดสิ่งที่สนใจในช่วงเวลาหรือในขอบเขตที่กำหนด เช่น ปริมาณการใช้งานเครือข่ายมือถือ (Mobile network)ในหนึ่งวัน จำนวนผู้เข้าชมเว็บไซต์ในหนึ่งชั่วโมง จำนวนลูกค้าที่ประสบภาวะล้มละลายในแต่ละเดือน จำนวนของการเรียกร้องค่าสินไหมทดแทนในหนึ่งไตรมาสของบริษัทประกันภัย และจำนวนผู้เสียชีวิตด้วยโรคมะเร็งปอดต่อปี เป็นต้น

ดังนั้นในงานวิจัยนี้จึงต้องการหาวิธีที่จะทำให้ผู้ที่จะนำการทดสอบเหล่านี้ไปใช้สามารถใช้งานได้ง่ายขึ้นโดยไม่จำเป็นต้องมีความรู้ทางคณิตศาสตร์ขั้นสูง นั่นคือการสร้างตารางค่าวิกฤตสำหรับ สถิติทดสอบของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง สำหรับการตรวจสอบ

ข้อมูลว่ามีการแจกแจงปัวซงหรือไม่ ในกรณีที่ไม่ทราบค่าพารามิเตอร์และประมาณค่าพารามิเตอร์จากข้อมูลตัวอย่าง และเปรียบเทียบประสิทธิภาพของการทดสอบทั้งสามชนิดนี้โดยพิจารณาจากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และกำลังการทดสอบ

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการศึกษาด้วยวิธีมอนติคาร์โล (Monte Carlo Method) เพื่อสร้างตารางค่าวิกฤตของสถิติทดสอบของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง สำหรับการแจกแจงปัวซง และศึกษาประสิทธิภาพของการทดสอบเหล่านี้โดยพิจารณาจาก 1) ความสามารถในการควบคุมการเกิดความผิดพลาดแบบที่ 1 และ 2) กำลังการทดสอบ

การทดสอบภาวะสารูปดีที่ศึกษา

การทดสอบภาวะสารูปดีที่นำมาศึกษาได้แก่ การทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง โดยในงานวิจัยนี้จะใช้สถิติทดสอบของการทดสอบ 3 ชนิด จาก Spinelli & Stephens (1997) และ Lockhart, Spinelli & Stephens (2007) มีรายละเอียดดังนี้

ให้ x คือตัวอย่างสุ่มขนาด n จากการแจกแจงปัวซงมีค่าเฉลี่ยคือ λ ที่ไม่ทราบค่า โดย x มีค่าแตกต่างกันจำแนกได้เป็น k กลุ่ม และมี p_i คือความน่าจะเป็นที่ตัวอย่างสุ่มมีค่าอยู่ในกลุ่มที่ i หาได้จาก

$$p_i = \frac{e^{-\lambda}\lambda^x}{x!}, \quad i = 1, 2, \ldots, k; \; x = 0, 1, 2, \ldots \tag{1}$$

ในกรณีที่ไม่ทราบค่า λ จะใช้ค่าประมาณจากวิธีภาวะน่าจะเป็นสูงสุด (Maximum likelihood method) คือ $\hat{\lambda} = \bar{x} = \sum x \big/ n$ แทน จะได้

$$\hat{p}_i = \frac{e^{-\hat{\lambda}}\hat{\lambda}^x}{x!} \tag{2}$$

เป็นค่าประมาณของ p_i

กำหนดให้ O_i คือจำนวนค่าสังเกตในกลุ่มที่ i และ $E_i = n \times \hat{p}_i$ คือจำนวนค่าสังเกตที่คาดหมายในกลุ่มที่ i และให้ $S_j = \sum_{i=1}^{j} O_i$, $\hat{T}_j = \sum_{i=1}^{j} E_i$ แล้ว $S_j \big/ n$ และ $\hat{H}_j = \hat{T}_j \big/ n$ คือความถี่สะสมสัมพัทธ์ของค่าที่สังเกตได้และค่าที่คาดหมาย ซึ่งตรงกันกับฟังก์ชันการแจกแจงเชิงประจักษ์ (Empirical distribution function) ของข้อมูล ($F_n(x)$) และฟังก์ชันการแจกแจงของการแจกแจงที่ต้องการทดสอบ ($F(x)$) ในกรณีของการแจกแจงต่อเนื่อง โดยที่ $\hat{Z}_j = S_j - \hat{T}_j$, j = 1,2,…,k จะได้ค่าสถิติทดสอบของการทดสอบทั้ง 3 วิธี ดังนี้

1. การทดสอบคอลโมโกรอฟ-สมีร์นอฟ (KS) มีค่าสถิติทดสอบคือ

$$KS = \max_{1 \le j \le k} \left| \sum_{i=1}^{j} \left(O_i - E_i \right) \right|$$

2. การทดสอบคราเมอร์-ฟอน-มิสส์ (CVM) มีค่าสถิติทดสอบคือ

$$CVM = W^2 = n^{-1} \sum_{j=1}^{k} \hat{z}_j^2 \hat{p}_j$$

3. การทดสอบแอนเดอร์สัน-ดาร์ลิง (AD) มีค่าสถิติทดสอบคือ

$$AD = A^2 = n^{-1} \sum_{j=1}^{k} \frac{\hat{z}_j^2 \hat{p}_j}{\hat{H}_j \left(1 - \hat{H}_j \right)}$$

การสร้างตารางค่าวิกฤต

ในการสร้างตารางค่าวิกฤตของสถิติทดสอบได้ใช้วิธีที่นำเสนอโดย Abidin, Adam and Midi (2012) ซึ่งหาค่าวิกฤตโดยการจำลองข้อมูลจากการแจกแจงที่สนใจที่มีพารามิเตอร์แตกต่างกันจำนวนหลาย ๆ ชุด นำข้อมูลที่ได้แต่ละชุดมาคำนวณค่าสถิติทดสอบของการทดสอบที่ศึกษา จากนั้นนำค่าสถิติทดสอบที่ได้จากข้อมูลที่มีการแจกแจงชนิดเดียวกันขนาดตัวอย่างเดียวกันแต่มีค่าพารามิเตอร์ต่างกันมาหาค่าเฉลี่ย โดยค่าเฉลี่ยของค่าสถิติทดสอบที่ได้นี้ถือเป็นค่าวิกฤตของสถิติทดสอบนั้น โดยมีขั้นตอนดังนี้

1. สร้างข้อมูลที่มีการแจกแจงปัวซงที่มีพารามิเตอร์ λ = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 และมีขนาดตัวอย่าง n = 10, 15, 20, 25, 30, 40, 50, 60, 70, 100, 150

2. ประมาณค่าพารามิเตอร์ของการแจกแจงปัวซงด้วยวิธีภาวะน่าจะเป็นสูงสุด จากนั้นนำไปแทนค่าในฟังก์ชันการแจกแจงความน่าจะเป็นสะสม $F_n(x, \hat{\lambda})$

3. คำนวณค่าสถิติทดสอบของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง

4. ทำซ้ำข้อ 1-3 จำนวน 100,000 รอบ

5. นำค่าสถิติทดสอบที่ได้มาหาค่าเปอร์เซ็นต์ไทล์ที่ 80, 85, 90, 95 และ 99 ซึ่งจะได้เป็นค่าวิกฤตที่ระดับนัยสำคัญ 0.20, 0.15, 0.10, 0.05 และ 0.01 ตามลำดับ

6. นำค่าวิกฤตในข้อ 5 มาหาค่าเฉลี่ยในทุกค่าพารามิเตอร์เพื่อเป็นตัวแทนของค่าวิกฤตในแต่ละขนาดตัวอย่าง

การศึกษาประสิทธิภาพของการทดสอบภาวะสารูปดี

สำหรับการศึกษาประสิทธิภาพของการทดสอบจะพิจารณาจากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และกำลังการทดสอบของการทดสอบทั้ง 3 ชนิด

1. สมมุติฐานที่ทดสอบ คือ

H_0 : ข้อมูลมีการแจกแจงปัวซง

H_a : ข้อมูลไม่มีการแจกแจงปัวซง

2. ในกรณีของการศึกษาความสามารถในการควบคุมความผิดพลาดแบบที่ 1 สร้างตัวอย่างสุ่มที่มีการแจกแจงปัวซงที่มีค่าเฉลี่ย λ = 1, 2, 3, 5, 10, 20 และขนาดตัวอย่าง n = 10, 30, 50, 100 ในกรณีของศึกษากำลังการทดสอบ สร้างตัวอย่างสุ่มที่มีการแจกแจงทวินาม ($B(m,p)$) ที่มีค่าเฉลี่ย ($m \times p$) เท่ากับ 1, 2, 3, 5, 10 และ 20 ตามค่าเฉลี่ยของการแจกแจงปัวซงที่ศึกษา โดยกำหนดพารามิเตอร์ m (จำนวนครั้งของการทำการทดลอง) และ p (ความน่าจะเป็นของการเกิดความสำเร็จหรือสิ่งที่สนใจ) ดัง Table 1

3. ประมาณค่าพารามิเตอร์ λ ของการแจกแจงปัวซง และนำไปแทนค่าในฟังก์ชัน $F(x, \hat{\lambda})$

4. คำนวณค่าสถิติทดสอบของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง

5. เปรียบเทียบค่าสถิติของการทดสอบแต่ละวิธีกับค่าวิกฤตในตารางสร้างไว้ พิจารณาผลการทดสอบสมมุติฐาน

6. ทำซ้ำข้อ 1 – 5 จำนวน 10,000 รอบ

7. ในกรณีของการควบคุมความผิดพลาดแบบที่ 1 คำนวณค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ($\hat{\alpha}$) จาก

$$\hat{\alpha} = \frac{\text{Number of reject } H_0}{10{,}000}$$

การพิจารณาความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ของการทดสอบทั้ง 3 ชนิด จะนำค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 เปรียบเทียบกับเกณฑ์ของ Bradley (Bradley, 1978) หากค่าประมาณที่ได้อยู่ในช่วง [0.025, 0.075] ที่ระดับนัยสำคัญ 0.05 หรืออยู่ในช่วง [0.005, 0.015] ที่ระดับนัยสำคัญ 0.01 จะถือว่าการทดสอบชนิดนั้นสามารถควบคุมความผิดพลาดแบบที่ 1 ได้

ส่วนในกรณีของกำลังการทดสอบจะได้ค่าประมาณของกำลังการทดสอบ ($1-\beta$) จาก

$$1-\beta = \frac{\text{Number of reject } H_0}{10{,}000}$$

Table 1 Parameters of binomial distribution

Mean	*m*	*p*	*m*	*p*	*m*	*p*
1	5	1/5	3	1/3	2	1/2
2	10	1/5	6	1/3	4	1/2
3	15	1/5	9	1/3	6	1/2
5	25	1/5	15	1/3	10	1/2
10	50	1/5	30	1/3	20	1/2
20	100	1/5	60	1/3	40	1/2

ผลการวิจัย

จากการดำเนินการวิจัยได้ผลการวิจัยดังนี้

ตารางค่าวิกฤต

Table 2 – 4 แสดงค่าวิกฤตของสถิติทดสอบของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง ตามลำดับ ซึ่งจะเห็นได้ว่าค่าวิกฤตของการทดสอบทั้ง 3 ชนิด มีค่าสูงขึ้นเมื่อระดับนัยสำคัญลดลง นอกจากนั้นค่าวิกฤตของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ มีค่าลดลงเมื่อขนาดตัวอย่างใหญ่ขึ้น ในขณะที่ค่าวิกฤตของการทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิงมีค่าเพิ่มขึ้นเมื่อขนาดตัวอย่างขนาดใหญ่ขึ้น

Table 2 Critical values of Kolmogorov-Smirnov test

	Significance level (α)				
n	0.20	0.15	0.10	0.05	0.01
10	0.1861	0.1978	0.2126	0.2344	0.2793
15	0.1527	0.1620	0.1743	0.1932	0.2308
20	0.1297	0.1379	0.1487	0.1655	0.1988
25	0.1185	0.1260	0.1357	0.1504	0.1804
30	0.1082	0.1150	0.1238	0.1374	0.1649
40	0.0937	0.0997	0.1074	0.1194	0.1432
50	0.0822	0.0875	0.0945	0.1053	0.1271
60	0.0765	0.0813	0.0876	0.0975	0.1172
70	0.0709	0.0754	0.0812	0.0903	0.1088
100	0.0593	0.0631	0.0680	0.0757	0.0911
150	0.0485	0.0515	0.0555	0.0618	0.0745

Table 3 Critical values of Cramer-von Mises test

n	Significance level (α)				
	0.20	0.15	0.10	0.05	0.01
10	0.1019	0.1145	0.1329	0.1630	0.2308
15	0.1031	0.1163	0.1349	0.1668	0.2395
20	0.1039	0.1173	0.1360	0.1679	0.2421
25	0.1041	0.1176	0.1367	0.1692	0.2449
30	0.1042	0.1178	0.1370	0.1697	0.2460
40	0.1043	0.1179	0.1371	0.1699	0.2466
50	0.1044	0.1181	0.1372	0.1702	0.2473
60	0.1045	0.1182	0.1373	0.1703	0.2474
70	0.1046	0.1182	0.1374	0.1705	0.2477
100	0.1046	0.1183	0.1376	0.1707	0.2487
150	0.1045	0.1182	0.1376	0.1708	0.2491

Table 4 Critical values of Anderson-Darling test

n	Significance level (α)				
	0.20	0.15	0.10	0.05	0.01
10	0.6176	0.6956	0.8079	1.0035	1.5225
15	0.6446	0.7233	0.8392	1.0384	1.5384
20	0.6596	0.7406	0.8546	1.0557	1.5517
25	0.6677	0.7497	0.8649	1.0664	1.5624
30	0.6743	0.7571	0.8748	1.0779	1.5742
40	0.6828	0.7662	0.8836	1.0866	1.5802
50	0.6879	0.7709	0.8887	1.0925	1.5871
60	0.6929	0.7761	0.8926	1.0958	1.5886
70	0.6935	0.7769	0.8957	1.0995	1.5954
100	0.6996	0.7835	0.9035	1.1091	1.6005
150	0.7040	0.7878	0.9076	1.1141	1.6046

เปรียบเทียบประสิทธิภาพของการทดสอบทั้งสามวิธีเมื่อใช้ตารางค่าวิกฤตที่สร้างขึ้นใหม่

1. ความสามารถในการควบคุมความผิดพลาดแบบที่ 1

จาก Table 5 การทดสอบแอนเดอร์สัน-ดาร์ลิงสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ทุกกรณีทั้งที่ระดับนัยสำคัญ 0.05 และ 0.01 ในขณะที่การทดสอบคราเมอร์-ฟอน-มิสส์สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ทุกกรณีเฉพาะที่ระดับนัยสำคัญ 0.05 แต่ที่ระดับนัยสำคัญ 0.01 จะสามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อ $\lambda = 2$ ขึ้นไป ส่วนการทดสอบคอลโมโกรอฟ-สมีร์นอฟ ไม่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ เมื่อ $\lambda = 1, 20$ ที่ทุกขนาดตัวอย่าง และเมื่อ $\lambda = 2$ ขนาดตัวอย่างเท่ากับ 10 ทั้งที่ระดับนัยสำคัญ 0.05 และ 0.01

2. กำลังการทดสอบ

จาก Table 6 – 8 จะเห็นได้ว่าค่าประมาณกำลังการทดสอบมีค่าเพิ่มขึ้นเมื่อขนาดตัวอย่างใหญ่ขึ้น และ p มีค่าเพิ่มขึ้น ในกรณีที่ค่าเฉลี่ยของการแจกแจงทวินามเท่ากับ 1 – 5 ที่ทุกค่า p การทดสอบคราเมอร์-ฟอน-มิสส์เป็นการทดสอบที่มีค่าประมาณกำลังการทดสอบมากที่สุด ยกเว้นเมื่อค่าเฉลี่ยเท่ากับ 5 และขนาดตัวอย่างเท่ากับ 100 ที่การทดสอบแอนเดอร์สัน-ดาร์ลิงมีค่าประมาณกำลังการทดสอบมากที่สุด

จาก Table 6 เมื่อค่าเฉลี่ยเท่ากับ 10 และ 20 ขนาดตัวอย่างเท่ากับ 10 – 50 การทดสอบคอลโมโกรอฟ-สมีร์นอฟมีค่าประมาณกำลังการทดสอบมากที่สุด ในขณะที่เมื่อขนาดตัวอย่างเท่ากับ 100 การทดสอบแอนเดอร์สัน-ดาร์ลิงมีค่าประมาณกำลังการทดสอบมากที่สุด

จาก Table 7 เมื่อค่าเฉลี่ยเท่ากับ 10 และขนาดตัวอย่างเท่ากับ 10 การทดสอบคอลโมโกรอฟ-สมีร์นอฟมีค่าประมาณกำลังการทดสอบมากที่สุด เมื่อขนาดตัวอย่างเท่ากับ 30 และ 50 การทดสอบคราเมอร์-ฟอน-มิสส์เป็นการทดสอบที่มีค่าประมาณกำลังการทดสอบมากที่สุด ในขณะที่เมื่อขนาดตัวอย่างเท่ากับ 100 การทดสอบแอนเดอร์สัน-ดาร์ลิง มีค่าประมาณกำลังการทดสอบมากที่สุด

จาก Table 8 ที่ระดับนัยสำคัญ 0.05 เมื่อค่าเฉลี่ยเท่ากับ 10 และ 20 ขนาดตัวอย่างเท่ากับ 10 การทดสอบคอลโมโกรอฟ-สมีร์นอฟมีค่าประมาณกำลังการทดสอบมากที่สุด เมื่อขนาดตัวอย่างเท่ากับ 30 การทดสอบคราเมอร์-ฟอน-มิสส์เป็นการทดสอบที่มีค่าประมาณกำลังการทดสอบมากที่สุด ในขณะที่เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 การทดสอบแอนเดอร์สัน-ดาร์ลิงมีค่าประมาณกำลังการทดสอบมากที่สุด สำหรับที่ระดับนัยสำคัญ 0.01 เมื่อค่าเฉลี่ยเท่ากับ 10 และ 20 ขนาดตัวอย่างเท่ากับ 10 การทดสอบคอลโมโกรอฟ-สมีร์นอฟมีค่าประมาณกำลังการทดสอบมากที่สุด เมื่อขนาดตัวอย่างเท่ากับ 30 และ 50 การทดสอบคราเมอร์-ฟอน-มิสส์เป็นการทดสอบที่มีค่าประมาณกำลังการทดสอบมากที่สุด ในขณะที่เมื่อขนาดตัวอย่างเท่ากับ 100 การทดสอบแอนเดอร์สัน-ดาร์ลิงมีค่าประมาณกำลังการทดสอบมากที่สุด

Table 5 Simulated probability of type I error

λ	n	Significance level (α)					
		0.05			0.01		
		KS	CVM	AD	KS	CVM	AD
1	10	0.0132*	0.0615	0.0316	0.0026*	0.0224*	0.0056
	30	0.0173*	0.0673	0.0352	0.0041*	0.0217*	0.0092
	50	0.0169*	0.0597	0.0288	0.0032*	0.0191*	0.0070
	100	0.0174*	0.0620	0.0323	0.0029*	0.0205*	0.0081
2	10	0.0196*	0.0635	0.0435	0.0049*	0.0148	0.0080
	30	0.0260	0.0584	0.0469	0.0052	0.0135	0.0102
	50	0.0306	0.0585	0.0510	0.0056	0.0140	0.0110
	100	0.0271	0.0550	0.0484	0.0053	0.0137	0.0121
3	10	0.0332	0.0568	0.0506	0.0066	0.0148	0.0099
	30	0.0378	0.0613	0.0561	0.0084	0.0124	0.0113
	50	0.0355	0.0497	0.0492	0.0074	0.0109	0.0110
	100	0.0343	0.0541	0.0532	0.0074	0.0125	0.0114
5	10	0.0453	0.0520	0.0543	0.0112	0.0115	0.0106
	30	0.0442	0.0493	0.0502	0.0077	0.0095	0.0096
	50	0.0491	0.0533	0.0555	0.0110	0.0096	0.0100
	100	0.0442	0.0493	0.0537	0.0086	0.0093	0.0109
10	10	0.0646	0.0462	0.0502	0.0132	0.0081	0.0086
	30	0.0608	0.0488	0.0497	0.0128	0.0084	0.0094
	50	0.0658	0.0485	0.0481	0.0135	0.0094	0.0093
	100	0.0602	0.0453	0.0463	0.0122	0.0084	0.0086
20	10	0.0794*	0.0410	0.0464	0.0184*	0.0084	0.0090
	30	0.0827*	0.0477	0.0481	0.0195*	0.0092	0.0091
	50	0.0876*	0.0469	0.0486	0.0196*	0.0092	0.0085
	100	0.0791*	0.0460	0.0445	0.0165*	0.0071	0.0078

Note: * The test that cannot control type I error.

Table 6 Simulated power for binomial distribution with p=1/5

m	*n*	Significance level					
		0.05			0.01		
		KS	CVM	AD	KS	CVM	AD
5	10	0.0183	0.0906*	0.0241	0.0043	0.0406*	0.0041
	30	0.0419	0.1256*	0.0592	0.0140	0.0543*	0.0181
	50	0.0716	0.1613*	0.0871	0.0214	0.0775*	0.0289
	100	0.1265	0.2684*	0.1757	0.0417	0.1321*	0.0675
10	10	0.0243	0.0807*	0.0336	0.0071	0.0227*	0.0058
	30	0.0433	0.1043*	0.0702	0.0096	0.0341*	0.0141
	50	0.0621	0.1368*	0.1110	0.0131	0.0486*	0.0281
	100	0.1051	0.2230*	0.2140	0.0297	0.0892*	0.0788
15	10	0.0348	0.0660*	0.0308	0.0080	0.0187*	0.0037
	30	0.0609	0.1040*	0.0750	0.0140	0.0303*	0.0163
	50	0.0835	0.1327*	0.1118	0.0228	0.0415*	0.0279
	100	0.1252	0.2113	0.2140*	0.0375	0.0795*	0.0765
25	10	0.0478	0.0567*	0.0334	0.0099	0.0112*	0.0034
	30	0.0597	0.0856*	0.0639	0.0139	0.0215*	0.0112
	50	0.0944	0.1173*	0.1039	0.0239	0.0375*	0.0269
	100	0.1356	0.1957	0.2056*	0.0375	0.0795*	0.0765
50	10	0.0668*	0.0523	0.0280	0.0127*	0.0095	0.0029
	30	0.0841*	0.0815	0.0617	0.0206*	0.0198	0.0111
	50	0.1121*	0.1079	0.0957	0.0300*	0.0283	0.0212
	100	0.1675	0.1927	0.2040*	0.0484	0.0672	0.0683*
100	10	0.0857*	0.0456	0.0235	0.0154*	0.0070	0.0016
	30	0.1095*	0.0770	0.0560	0.0224*	0.0178	0.0094
	50	0.1408*	0.1077	0.0960	0.0342*	0.0260	0.0193
	100	0.1910*	0.1772	0.1910*	0.0552	0.0591	0.0633*

Note: * The test with the highest simulated power.

Table 7 Simulated power for binomial distribution with p=1/3

m	*n*	Significance level					
		0.05			0.01		
		KS	CVM	AD	KS	CVM	AD
3	10	0.0345	0.1376*	0.0331	0.0073	0.0646*	0.0071
	30	0.1160	0.2785*	0.1431	0.0472	0.1405*	0.0446
	50	0.2230	0.3974*	0.2460	0.0865	0.2315*	0.1038
	100	0.4675	0.6861*	0.5701	0.2630	0.4862*	0.3484
6	10	0.0355	0.1089*	0.0368	0.0119	0.0349*	0.0029
	30	0.0984	0.2123*	0.1489	0.0246	0.0836*	0.0457
	50	0.1655	0.3325*	0.2847	0.0534	0.1515*	0.1068
	100	0.3697	0.6182*	0.6112	0.1524	0.3860*	0.3625
9	10	0.0508	0.0977*	0.0356	0.0124	0.0281*	0.0036
	30	0.1241	0.2037*	0.1551	0.0288	0.0769*	0.0403
	50	0.1956	0.3119*	0.2850	0.0688	0.1352*	0.1067
	100	0.4006	0.5912	0.6161*	0.1720	0.3413	0.3596*
15	10	0.0658	0.0888*	0.0353	0.0127	0.0195*	0.0023
	30	0.1233	0.1844*	0.1452	0.0356	0.0619*	0.0389
	50	0.2205	0.2933*	0.2829	0.0723	0.1224*	0.1024
	100	0.4081	0.5549	0.6026*	0.1741	0.3054	0.3410*
30	10	0.0820*	0.0715	0.0325	0.0207*	0.0156	0.0023
	30	0.1588	0.1796*	0.1467	0.0448	0.0564*	0.0322
	50	0.2569	0.2836*	0.2800	0.0845	0.1098*	0.0971
	100	0.4557	0.5353	0.5969*	0.1993	0.2863	0.3249*
60	10	0.1016*	0.0609	0.0259	0.0199*	0.0119	0.0015
	30	0.1964*	0.1697	0.1388	0.0556*	0.0532	0.0312
	50	0.2940*	0.2709	0.2687	0.1007	0.1022*	0.0879
	100	0.5037	0.5364	0.5947*	0.2321	0.2839	0.3262*

Note: * The test with the highest simulated power.

Table 8 Simulated power for binomial distribution with p=1/2

m	n	Significance level					
		0.05			0.01		
		KS	CVM	AD	KS	CVM	AD
2	10	0.0866	0.2594*	0.0711	0.0222	0.1568*	0.0170
	30	0.3720	0.6905*	0.3375	0.1964	0.4774*	0.1399
	50	0.6786	0.9170*	0.5628	0.4371	0.7963*	0.3269
	100	0.9650	0.9993*	0.9137	0.8573	0.9944*	0.7739
4	10	0.0708	0.2007*	0.0726	0.0268	0.0737*	0.0077
	30	0.2909	0.5559*	0.4235	0.1076	0.2971*	0.1749
	50	0.5380	0.8010*	0.7341	0.2662	0.5701*	0.4666
	100	0.9066	0.9900*	0.9776	0.6897	0.9468*	0.9100
6	10	0.0898	0.1790*	0.0632	0.0247	0.0557*	0.0066
	30	0.3299	0.5108*	0.4217	0.1149	0.2711*	0.1669
	50	0.5677	0.7639*	0.7336	0.2997	0.5209*	0.4519
	100	0.9107	0.9812	0.9849*	0.7129	0.9175	0.9243*
10	10	0.1156	0.1524*	0.0595	0.0313	0.0415*	0.0050
	30	0.3453	0.4828*	0.4192	0.1330	0.2356*	0.1589
	50	0.6019	0.7457*	0.7425	0.3071	0.4919*	0.4523
	100	0.9180	0.9755	0.9844*	0.7191	0.8989	0.9252*
20	10	0.1417*	0.1366	0.0559	0.0368*	0.0341	0.0032
	30	0.3879	0.4573*	0.4098	0.1579	0.2169*	0.1502
	50	0.6458	0.7256	0.7366*	0.3408	0.4602*	0.4374
	100	0.9320	0.9696	0.9840*	0.7503	0.8836	0.9186*
40	10	0.1803*	0.1188	0.0512	0.0447*	0.0294	0.0027
	30	0.4384	0.4471*	0.4055	0.1779	0.2034*	0.1448
	50	0.6952	0.7205	0.7387*	0.3859	0.4529*	0.4403
	100	0.9469	0.9707	0.9855*	0.7816	0.8773	0.9230*

Note: * The test with the highest simulated power.

วิจารณ์ผลการวิจัย

ในการศึกษาการทดสอบภาวะสารูปดี 3 ชนิด ได้แก่การทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง เมื่อนำมาใช้ในการทดสอบข้อมูลว่ามีการแจกแจงปัวซงหรือไม่โดยใช้ตารางค่าวิกฤตของการทดสอบแต่ละชนิดที่สร้างขึ้น พบว่าโดยภาพรวมแล้วการทดสอบคราเมอร์-ฟอน-มิสส์เป็นการ

ทดสอบที่มีประสิทธิภาพดีที่สุดเนื่องจากสามารถควบคุมความผิดพลาดแบบที่ 1 และให้ค่าประมาณกำลังการทดสอบมากที่สุดในเกือบทุกกรณี ซึ่งผลการศึกษานี้สอดคล้องกับผลการศึกษาของ Spinelli & Stephens (1997) และ Lockhart, Spinelli & Stephens (2007) ดังนั้นตารางค่าวิกฤตที่สร้างใหม่นี้สามารถนำมาใช้ในการทดสอบว่าข้อมูลมีการแจกแจงปัวซงได้อย่างมีประสิทธิภาพ

นอกจากนั้นจะเห็นได้ว่าเมื่อข้อมูลตัวอย่างถูกสุ่มจากการแจกแจงทวินามที่มีค่า p มีค่าน้อยคือ 1/5 ค่าประมาณกำลังการทดสอบจะมีค่าต่ำเนื่องจากการแจกแจงทวินามที่ p มีค่าน้อยจะมีลักษณะใกล้เคียงกับการแจกแจงปัวซงที่มีค่าเฉลี่ยเท่ากับ $\lambda = m \times p$ และค่าประมาณกำลังการทดสอบจะสูงขึ้นเมื่อ p มีค่าเข้าใกล้ 0.50 มากขึ้น นอกจากนั้นกำลังการทดสอบของการทดสอบทั้ง 3 วิธี ยังมีค่าเพิ่มขึ้นเมื่อ m และขนาดตัวอย่าง n เพิ่มมากขึ้น

สรุปผลการวิจัย

ในการศึกษาประสิทธิภาพของการทดสอบภาวะสารูปดี 3 วิธี คือ การทดสอบคอลโมโกรอฟ-สมีร์นอฟ การทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิง เมื่อนำมาใช้ในการทดสอบข้อมูลว่ามีการแจกแจงปัวซงหรือไม่ โดยใช้ตารางค่าวิกฤตของการทดสอบแต่ละชนิดที่สร้างขึ้นใหม่ พบว่าการทดสอบคราเมอร์-ฟอน-มิสส์ และการทดสอบแอนเดอร์สัน-ดาร์ลิงสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ดีกว่าการทดสอบคอลโมโกรอฟ-สมีร์นอฟ เนื่องจากการทดสอบคอลโมโกรอฟ-สมีร์นอฟไม่สามารถควบคุมความผิดพลาดแบบที่ 1 ได้เมื่อ λ เท่ากับ 1 และ 20 เมื่อพิจารณาจากค่าประมาณกำลังการทดสอบพบว่าการทดสอบคราเมอร์-ฟอน-มิสส์เป็นการทดสอบที่ให้ค่าประมาณกำลังการทดสอบสูงที่สุดในเกือบทุกกรณี และค่าประมาณกำลังการทดสอบของการทดสอบทั้ง 3 วิธี จะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น

เอกสารอ้างอิง

Abidin, N.Z., Adam, M.B., & Midi, H.B. (2012). The Goodness-of-fit Test for Gumbel Distribution: A Comparative Study. *Mathematika, 28*, 35-48.

Bradley, J.V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Choulakian, V., Lockhart, R.A., & Stephens, M.A. (1994). Cramer-von Mises Statistics for Discrete distributions. *The Canadian Journal of Statistics*, *22*(1), 125-137.

Conover, W.J. (1972). A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions. *Journal of the American Statistical Association*, *67*(339), 591-596.

Wood, C. L., & Altavela, M. M. (1978). Large-Sample Results for Kolmogorov-Smirnov Statistics for Discrete Distributions. *Biometrika*, *65*(1), 235–239. https://doi.org/10.2307/2335304

Gibbons, J.D., & Chakraborti, S. (2003). *Nonparametric Statistical Inference: Revised and Expanded.* (4th ed.). CRC Press. https://doi.org/10.4324/9780203911563

Horn, S.D. (1977). Goodness-of-Fit Tests for Discrete Data: A Review and an Application to a Health Impairment Scale. *Biometrics*, *33*(1), 237-247.

Lockhart, R.A., Spinelli, J.J., & Stephens, M.A. (2007). Cramer-von Mises Statistics for Discrete Distributions with Unknow Parameters. *The Canadian Journal of Statistics, 35*(1), 125-133.

Pettitt, A.N., & Stephens, M.A. (1977). The Kolmogorov-Smirnov Goodness-of-Fit Statistic with Discrete and Grouped Data. *Technometrics*, *19*(2), 205-210.

Spinelli, J.J., Stephens, M.A. (1997). Cramer-von Mises Tests of Fit for the Poisson distribution. *The Canadian Journal of Statistics*, *25*(2), 257-268.