การประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการใช้การแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ

An Approximation of Probability of Binomial Distribution with Poisson Distribution, Generalized Poisson Distribution and Normal Distribution

ณัฐติยา โนนกอง[1] และ บรรทม สุระพร[1*]

[1]ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

[*]Corresponding Author, E-mail: bunthom.s@ubu.ac.th

บทคัดย่อ

วัตถุประสงค์ของงานวิจัยนี้คือ ศึกษาการประมาณค่าของการแจกแจงทวินาม ซึ่งในกรณีที่ n มีค่ามาก ($n \to \infty$) และ p มีค่าน้อย ($p \to 0$) นั้นทำให้เกิดปัญหาในการคำนวณ ดังนั้นงานวิจัยนี้จึงสนใจเรื่องการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป การแจกแจงปรกติ และจะเปรียบเทียบประสิทธิภาพของการประมาณทั้ง 3 วิธีข้างต้น โดยพิจารณาที่ค่าความผิดพลาดของการประมาณ ซึ่งจากการวิจัย พบว่าการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซงนัยทั่วไปเป็นวิธีที่มีความผิดพลาดในการประมาณน้อยที่สุด

ABSTRACT

The aim of this research is to study the estimation of binomial distribution. In which case n is large ($n \to \infty$) and p is small ($p \to 0$), causes that computation is difficult. So this research suggested in approximate binomial probability distribution with Poisson distribution, generalized Poisson distribution, and normal distribution. Moreover, we compare efficiency of each estimators and error of estimated. The result, we found that an approximation of probability of binomial distribution with generalized Poisson distribution is lowest error of estimated.

คำสำคัญ: การประมาณค่าความน่าจะเป็น การแจกแจงทวินาม การแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป การแจกแจงปรกติ

Keywords: Probability estimation, Binomial distribution, Usual Poisson distribution, Generalized Poisson distribution, Normal distribution

1. บทนำ

1.1 ความเป็นมาและความสำคัญของการวิจัย

การแจกแจงทวินามเป็นการแจกแจงของตัวแปรสุ่มไม่ต่อเนื่องที่เกิดขึ้นในการทดลองซ้ำๆ กัน n ครั้งที่อิสระกัน ซึ่งในการทดลองแต่ละครั้งจะมีผลลัพธ์ที่เป็น ไปได้คือประสบความสำเร็จหรือล้มเหลว โดยที่ p แทนความน่าจะเป็นของการประสบความสำเร็จในแต่ละครั้งของการทดลอง ซึ่ง p มีค่าคงที่เสมอ และ $q = 1 - p$ แทนความน่าจะเป็นของการล้มเหลวในแต่ละครั้งของการทดลอง โดยมีฟังก์ชันความน่าจะเป็นดังสมการ $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$; $x = 0, 1, 2, \ldots, n$ ซึ่งในกรณีที่ n มีค่ามาก ($n \to \infty$) และ p มีค่าน้อย ($p \to 0$) นั้นทำให้เกิดปัญหาในการคำนวณ งานวิจัยเรื่องนี้ จึงได้นำเสนอการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ มาใช้ในการประมาณเพื่อแก้ปัญหาที่เกิดขึ้นดังกล่าว ดำเนินการเปรียบเทียบประสิทธิภาพของการประมาณทั้ง 3 วิธีข้างต้น โดยพิจารณาจากค่าความผิดพลาดของการประมาณเพื่อหาวิธีการที่ทำให้ความผิดพลาดของการประมาณต่ำที่สุด

1.2 วัตถุประสงค์การวิจัย

1. ศึกษาการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซง
2. ศึกษาการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซงนัยทั่วไป
3. ศึกษาการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปรกติ
4. เปรียบเทียบประสิทธิภาพของการประมาณทั้ง 3 วิธี โดยพิจารณาจากความผิดพลาดของการประมาณ

1.3 ขอบเขตของการวิจัย

ในการทำวิจัยเรื่องการประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงปัวซงนัยทั่วไป ผู้วิจัยได้ใช้ความรู้ในเรื่องการแจกแจงทวินาม การแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ รวมถึงการใช้โปรแกรม Microsoft Excel ในการคำนวณค่าความน่าจะเป็นของการแจกแจงทวินาม และการหาค่าของการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซง การแจกแจง ปัวซงนัยทั่วไป และการแจกแจงปรกติ โดยการคำนวณค่า n และ p ในกรณีต่างๆ เพื่อนำค่าที่ได้มาหาค่าความผิดพลาดของการประมาณ ว่าวิธีใดที่ให้ค่าใกล้เคียงกับค่าความน่าจะเป็นของการแจกแจงทวินามมากที่สุด

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถแก้ปัญหาในการคำนวณค่าความน่า จะเป็นของการแจกแจงทวินามในกรณีที่ n มีค่ามาก ($n \to \infty$) และ p มีค่าน้อย ($p \to 0$) ได้
2. ได้เรียนรู้วิธีการใหม่ๆ ในการหาค่าประมาณของความน่าจะเป็นทวินาม ด้วยการแจกแจงแบบต่างๆ

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรูปแบบของการแจกแจงแบบต่างๆ และกล่าวถึงนิยามตัวแปรสุ่ม ฟังก์ชัน ค่าคาดหวังและความแปรปรวนของการแจกแจงแบบต่างๆ รวมทั้งการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ ซึ่งมีรายละเอียดดังนี้

2.1 การแจกแจงทวินาม (BInomial Distribution: BID)

การแจกแจงทวินาม ใช้แทนข้อมูลที่มีลักษณะดังนี้

1. เป็นการแจกแจงของตัวแปรสุ่มแบบไม่ต่อเนื่อง
2. เป็นการทดลองซ้ำๆ กัน n ครั้ง โดยมีผลลัพธ์เป็นไปได้สองอย่างเท่านั้นคือได้สิ่งที่สนใจ หรืออาจเรียก ว่าสำเร็จ (Success) และได้สิ่งที่ไม่สนใจหรือเรียกว่าล้ม เหลว (Failure)
3. ความน่าจะเป็นที่ได้ผลลัพธ์ตามสิ่งที่สนใจ หรือสำเร็จ แทนด้วย p และคงที่ในทุกๆ ครั้ง ของการทดลอง

4. การทดลองแต่ละครั้งเป็นอิสระกัน

2.1.1 ตัวแปรสุ่มและฟังก์ชันการแจกแจง

นิยาม 2.1 สำหรับการทดลองทวินาม กำหนดตัวแปรสุ่ม X คือ จำนวนครั้งของการสำเร็จ โดยมีฟังก์ชันความน่าจะเป็นดังสมการ

$$f(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & ;x = 0,1,2,\ldots,n \\ 0 & ;\text{ที่อื่นๆ} \end{cases}$$

เมื่อ n คือ จำนวนครั้งที่ทำการทดลองสุ่ม

x คือ จำนวนครั้งของการสำเร็จที่เกิดขึ้น

p คือ ความน่าจะเป็นของการสำเร็จ ซึ่งคงที่ของแต่ละครั้งของการทดลองสุ่ม ($0 \le p \le 1$)

$1 - p$ หรือ q คือ ความน่าจะเป็นของการล้มเหลวจากการทดลองสุ่ม ($0 \le q \le 1$)

จะได้ว่า ตัวแปรสุ่ม X มีการแจกแจงทวินาม โดยมีพารามิเตอร์ n, p เขียนได้เป็น X~Bin($x;n,p$)

2.1.2 ค่าคาดหวัง (Expectation: *E*(*X*)) และค่าความแปรปรวน (Variance: *V*(*X*))

ค่าคาดหวัง หรือค่าเฉลี่ย สามารถหาได้จาก $E(X) = \sum_{all\ x} x.f(x)$ ซึ่งมีค่าเท่ากับ np และค่าความแปรปรวน หาได้จาก $V(X) = E(X^2) - [E(X)]^2$ มีค่าเท่ากับ $np(1-p) = npq$

2.2 การแจกแจงปัวซง (Usual Poisson Distribution: UPD)

การแจกแจงปัวซง เป็นการแจกแจงของตัวแปรสุ่มแบบไม่ต่อเนื่องที่เกิดขึ้นในการทดลองใดๆ ที่ให้ความสนใจต่อจำนวนครั้งของการเกิดเหตุการณ์ ในช่วงเวลาหรือบริเวณที่กำหนด เราเรียกว่าเป็นการทดลองปัวซง ซึ่งมีลักษณะดังนี้

1. จำนวนผลลัพธ์ที่เกิดขึ้นในช่วงเวลาหนึ่ง หรือบริเวณที่กำหนด เป็นอิสระกับจำนวนที่เกิดขึ้นในช่วงเวลาอื่นหรือบริเวณอื่น

2. ความน่าจะเป็นที่เกิดผลลัพธ์ในช่วงเวลาที่สั้นมากหรือบริเวณที่เล็กมาก จะเป็นปฏิภาคโดยตรงกับช่วง เวลาหรือขนาดของบริเวณนั้นๆ และไม่ขึ้นอยู่กับจำนวนผลลัพธ์ที่เกิดขึ้นนอกช่วงเวลาหรือบริเวณดังกล่าว

2.2.1 ตัวแปรสุ่มและฟังก์ชันการแจกแจง

นิยาม 2.2 สำหรับการทดลองปัวซง กำหนดตัวแปรสุ่ม X คือ จำนวนครั้งของการเกิดเหตุการณ์ที่สนใจในช่วงเวลา หรือบริเวณที่กำหนด โดยมีฟังก์ชันความน่าจะเป็น ดังสมการ

$$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & ;x = 0,1,2,\ldots \\ 0 & ;\text{ที่อื่นๆ} \end{cases}$$

เมื่อ λ คือ ค่าเฉลี่ยของจำนวนครั้งของการเกิดเหตุการณ์นั้น ($\lambda > 0$)

e คือ ค่าคงที่ ซึ่ง $e \simeq 2.71828...$

จะได้ว่า ตัวแปรสุ่ม X มีการแจกแจงปัวซง โดยมีพารามิเตอร์ λ เขียนได้เป็น X~Poi(x; λ)

2.2.2 ค่าคาดหวัง และค่าความแปรปรวน

ค่าคาดหวังหรือค่าเฉลี่ย ซึ่งมีค่าเท่ากับ λ และค่าความแปรปรวนมีค่าเท่ากับ λ

2.2.3 การประมาณค่าการแจกแจงทวินาม ด้วยการแจกแจงปัวซง

สำหรับการแจกแจงทวินาม เมื่อ n มีค่ามากๆ จะทำให้การคำนวณหาค่าความน่าจะเป็นค่อนข้างยุ่งยากมาก จึงสามารถใช้การแจกแจงปัวซงประมาณค่าความน่าจะเป็นของการแจกแจงทวินามได้ โดยที่ ถ้า n มีค่ามากๆ และค่า p จึงต้องมีค่าน้อยๆ เข้าใกล้ 0

เพื่อจะทำให้ค่าเฉลี่ย (λ) ต้องคงที่ ($n \to \infty$, $p \to 0$) จะใช้การแจกแจงปัวซงในการประมาณ ค่าความน่าจะเป็นของการแจกแจงทวินาม โดยใช้ $\lambda = np$

2.3 การแจกแจงปัวซงนัยทั่วไป (Generalized Poisson Distribution: GPD)

ถ้าการทดลองใดๆ ที่ให้ความสนใจต่อจำนวนครั้งของการเกิดเหตุการณ์ ในช่วงเวลา หรือบริเวณที่กำหนด เราเรียกว่าเป็นการทดลองปัวซงนัยทั่วไป โดยมีลักษณะคล้ายกับการแจกแจงปัวซง แต่จะแตกต่างจากการแจกแจงปัวซง คือจะมีสองพารามิเตอร์ ซึ่งพารามิเตอร์แรกเหมือนกับการแจกแจงปัวซง กล่าวคืออธิบายค่าเฉลี่ยของจำนวนครั้งที่เกิดขึ้น แต่พารามิเตอร์ที่สองจะอธิบายอัตราส่วนระหว่างค่าเฉลี่ยกับความแปรปรวน ดังนั้นถ้าพารามิเตอร์ที่สองมีค่าเป็น 0 การแจกแจงปัวซงนัยทั่วไป จะกลายมาเป็นการแจกแจงปัวซง

2.3.1 ตัวแปรสุ่มและฟังก์ชันการแจกแจง

นิยาม 2.3 สำหรับการทดลองปัวซงนัยทั่วไป กำหนดตัวแปรสุ่ม X คือ จำนวนครั้งของการเกิดเหตุการณ์ที่สนใจในช่วงเวลาหรือบริเวณที่กำหนด โดยมีฟังก์ชันความน่าจะเป็น ดังสมการ

$$f(x;\lambda,\theta) = \begin{cases} \dfrac{\lambda(\lambda+\theta x)^{x-1} e^{-\lambda-\theta x}}{x!} & ; x = 0,1,2,\ldots \\ 0 \quad \text{สำหรับ} \quad x > m, \text{ เมื่อ} \quad \theta < 0 \end{cases}$$

เมื่อ λ คือ ค่าเฉลี่ยของจำนวนครั้งของการเกิดเหตุการณ์นั้น ($\lambda > 0$)

θ คือ ค่าพารามิเตอร์ของ GPD โดยมีค่า $\max(-1, -\lambda/m < \theta \le 1)$ และ m (≥ 4) เป็นจำนวนเต็มบวกมากที่สุดสำหรับ $\lambda + m\theta > 0$ เมื่อ θ มีค่าเป็นลบ

e คือ ค่าคงที่ ซึ่ง $e \simeq 2.71828\ldots$

จะได้ว่า ตัวแปรสุ่ม X มีการแจกแจงปัวซงนัยทั่วไป โดยมีพารามิเตอร์ λ, θ เขียนได้เป็น $X \sim \text{GPoi}(x; \lambda, \theta)$

2.3.2 ค่าคาดหวัง และค่าความแปรปรวน ค่าคาดหวัง หรือค่าเฉลี่ย $E(X) = \dfrac{\lambda}{1-\theta}$ และค่าความแปรปรวน $V(X) = \dfrac{\lambda}{(1-\theta)^3}$

2.3.3 การประมาณค่าการแจกแจงทวินาม ด้วยการแจกแจงปัวซงนัยทั่วไป

สำหรับการแจกแจงทวินาม เมื่อ n มีค่ามากๆ และค่า p มีค่าน้อยๆ เข้าใกล้ 0 ($n \to \infty$, $p \to 0$) สามารถใช้การแจกแจงปัวซงนัยทั่วไปประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม โดยใช้ $np = \dfrac{\lambda}{1-\theta}$ และ $np(1-p) = \dfrac{\lambda}{(1-\theta)^3}$ โดยให้ค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงแบบทวินามเท่ากับค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงปัวซงนัยทั่วไป ดังนั้นจึงได้ว่า

$\dfrac{E(X)}{V(X)} = \dfrac{\lambda}{1-\theta}\dfrac{(1-\theta)^3}{\lambda} = \dfrac{np}{np(1-p)}$ จึงได้ $(1-\theta)^2 = \dfrac{1}{1-p}$ ดังนั้น $\hat{\theta} = 1-\dfrac{1}{\sqrt{1-p}}$ และ $np = \dfrac{\lambda}{1-\theta}$ จึงได้ $np(1-\theta) = \lambda$ ดังนั้น $\hat{\lambda} = np(1-1+\dfrac{1}{\sqrt{1-p}}) = \dfrac{np}{\sqrt{1-p}}$ เนื่องจาก $\lambda > 0$ ดังนั้น $\hat{\lambda} = \dfrac{np}{\sqrt{1-p}} > 0 \to np > 0$ และจาก $\theta \le 1$ ดังนั้น $\hat{\theta} = 1-\dfrac{1}{\sqrt{1-p}} \le 1$, $\dfrac{\sqrt{1-p}-1}{\sqrt{1-p}} \le 1$, $\sqrt{1-p} - 1 \le \sqrt{1-p}$, $1-2\sqrt{1-p} \le 0$ เราจะได้ว่า $p \le 3/4 = 0.75$ ดังนั้นในการประมาณจึงใช้ค่า p ไม่เกิน 0.75

2.4 การแจกแจงปรกติ (Normal Distribution: NOD)

การแจกแจงปรกติ เป็นการแจกแจงความน่าจะเป็นแบบต่อเนื่องที่สำคัญที่สุด ในบรรดาการแจกแจงความน่าจะเป็นแบบต่อเนื่อง เนื่องจากข้อมูลส่วนใหญ่จะมีลักษณะเป็นการแจกแจงปรกติหรือใกล้เคียง ซึ่งการแจกแจงปรกติจะมีลักษณะดังนี้

1. เป็นการแจกแจงของตัวแปรสุ่มแบบต่อเนื่องที่มีค่าเป็นจำนวนจริง

2. เมื่อเรานำข้อมูลแสดงด้วยกราฟแจกแจงความถี่ จะมีลักษณะเป็นรูปโค้งระฆังคว่ำและสมมาตรที่ค่าเฉลี่ย (μ) มีการกระจายของข้อมูลด้วยความแปรปรวน (σ^2)

2.4.1 ตัวแปรสุ่มและฟังก์ชันการแจกแจง

นิยาม 2.4 สำหรับการทดลองปรกติ

กำหนดตัวแปรสุ่ม X คือ ค่าของตัวแปรสุ่มที่สนใจ ซึ่งมีค่าเป็นจำนวนจริง โดยมีฟังก์ชันความน่าจะเป็น ดังสมการ

$$f(x;\mu,\sigma^2) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\ \sigma^2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} & ;-\infty < x < \infty, \sigma^2 > 0 \\ 0 & ;\text{ที่อื่นๆ} \end{cases}$$

เมื่อ μ คือ ค่าเฉลี่ยของข้อมูลตัวแปรสุ่มของเหตุการณ์นั้น ($-\infty < \mu < \infty$)

σ^2 คือ ค่าความแปรปรวนของเหตุการณ์นั้น ($\sigma^2 > 0$)

e คือ ค่าคงที่ ซึ่ง $e \simeq 2.71828...$

จะได้ว่า ตัวแปรสุ่ม X มีการแจกแจงปรกติ โดยมีพารามิเตอร์ μ, σ^2 เขียนได้เป็น X~Nor(x; μ,σ^2)

2.4.2 ค่าคาดหวัง (Expectation: $E(X)$) และค่าความแปรปรวน (Variance: $V(X)$)

ค่าคาดหวังหรือค่าเฉลี่ย $E(X) = \mu$ และค่าความแปรปรวน $V(X) = \sigma^2$

2.4.3 การประมาณค่าการแจกแจงทวินาม ด้วยการแจกแจงปรกติ

สำหรับการแจกแจงข้อมูลใดๆ สามารถแปลงให้เป็นข้อมูลปรกติมาตรฐาน โดยวิธีเอาข้อมูลนั้นลบออกด้วยค่าเฉลี่ยและหารด้วยส่วนเบี่ยงเบนมาตรฐานของการแจกแจงนั้น กล่าวคือ $Z = \dfrac{X - E(X)}{\sqrt{V(X)}}$ เพื่อที่จะหาความน่าจะเป็นโดยประมาณได้ง่ายขึ้น ซึ่งในกรณีการแจกแจงทวินาม เมื่อ n มีค่ามากๆ และค่า p มีค่าเข้าใกล้ 0.5 ($n \longrightarrow \infty$, $p \longrightarrow 0.5$) หรือ $np > 5$ และ $nq > 5$ จะทำให้การคำนวณหาค่าความน่าจะเป็นค่อนข้างยุ่งยากมาก จึงสามารถใช้การแจกแจงปรกติประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม โดยใช้ $\mu = np$ และ $\sigma^2 = np(1-p)$ ดังนั้น จึงได้ว่า $Z = \dfrac{X-\mu}{\sigma}$ เนื่องจากการแจกแจงปรกติเป็นแบบต่อเนื่อง จึงต้องปรับ แก้ความต่อเนื่อง โดยการ ± 0.5 ซึ่งการปรับแก้นี้ เรียกว่า ปัจจัยในการปรับให้ต่อเนื่อง (Continuity correction factor) และแทนค่า จะได้การประมาณคือ $Z = \dfrac{(X \pm 0.5) - np}{\sqrt{np(1-p)}}$

หมายเหตุ: การนำการแจกแจงปรกติมาประมาณค่าการแจกแจงทวินาม โดยค่าความน่าจะเป็นหรือค่า ประมาณใกล้เคียงค่าความน่าจะเป็นทวินาม โดยค่าความน่าจะเป็นจะยิ่งใกล้เคียงมากขึ้น เมื่อ n มีขนาดใหญ่ และค่า p เข้าใกล้ 0.50 อย่างไรก็ดีไม่ได้หมายความว่าเราจะใช้การแจกแจงปกติมาประมาณค่าความน่าจะเป็นทวินามไม่ได้ หาก p ไม่ใกล้เคียง 0.50 เหตุที่เมื่อ p เข้าใกล้ 0.50 แล้วทำให้การประมาณค่ามีความใกล้เคียงการแจกแจงทวินามมากขึ้นนั้นเพราะจะมีความสมมาตรเกิดขึ้น ซึ่งใกล้ เคียงกับคุณสมบัติของการแจกแจงปรกติ

3. วิธีการดำเนินการวิจัย

สำหรับในบทนี้จะแบ่งออกเป็น 2 ส่วน ซึ่งในส่วนแรกจะกล่าวถึงการนำโปรแกรม Microsoft Excel มาคำนวณค่าความน่าจะเป็นของการแจกแจงทวินาม และคำนวณหาการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ โดยการแทนค่าจากฟังก์ชันของการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงทั้ง 3 วิธี ซึ่งจะแบ่งการคำนวณออกเป็นกรณีต่างๆ และในส่วนที่สอง แสดงวิธีการคำนวณค่าความน่าจะเป็นของการแจกแจงทวินาม และการประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงแบบต่างๆ ซึ่งมีรายละเอียดดังนี้

3.1 การคำนวณใช้โปรแกรม Microsoft Excel

ในหัวข้อนี้จะใช้โปรแกรม Microsoft Excel ในการคำนวณหาค่าความน่าจะเป็นของการแจกแจงทวินาม และคำนวณหาค่าประมาณความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงปัวซง การแจกแจง ปัวซงนัยทั่วไป และการแจกแจงปรกติ รวมถึงการหาค่าความผิดพลาดของการประมาณ โดยใช้ค่าจริงจากการหาค่าความน่าจะเป็นของการแจกแจงทวินาม ลบออกด้วยค่าประมาณความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ โดยใช้โปรแกรม Microsoft Excel ซึ่งจะแบ่งกรณีในการคำนวณตามค่าของ n (4, 10, 20, 100, 150) ค่าของ p (0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01) ดังตารางต่อไปนี้

ตารางที่ 1 ค่าความน่าจะเป็นของการแจกแจงทวินาม ค่าประมาณความน่าจะเป็นของการแจกแจงปัวซง (UPD) การแจกแจงปัวซงนัยทั่วไป (GPD) การแจกแจงปรกติ (NOD) และความผิดพลาดของการประมาณแบบต่างๆ กรณีที่ $x = \frac{n}{2}$

p	BID (n, p)	UPD (np)	GPD (λ, θ)	NOD (μ, σ)	UPD Error	GPD Error	NOD Error
n =4							
0.7	0.264600	0.238375	0.277844	0.412577	0.026225	0.013244	0.147977
0.6	0.345600	0.261268	0.359130	0.405054	0.084332	0.013530	0.059454
0.5	0.375000	0.270671	0.382786	0.352065	0.104329	0.007786	0.022935
0.4	0.345600	0.258428	0.347545	0.267028	0.087172	0.001945	0.078572
0.3	0.264600	0.216860	0.263574	0.159181	0.047740	0.001026	0.105419
0.2	0.153600	0.143785	0.152425	0.052151	0.009815	0.001175	0.101449
0.1	0.048600	0.053626	0.048300	0.001454	0.005026	0.000300	0.047146
0.05	0.013538	0.016375	0.013490	0.000001	0.002837	0.000048	0.013537
0.01	0.000588	0.000769	0.000588	1.31E-33	0.000181	4.78E-07	0.000588
		Average			**0.040851**	**0.004340**	**0.064120**
n =10							
0.7	0.102919	0.127716	0.104334	0.161117	0.024797	0.001415	0.058198
0.6	0.200658	0.160623	0.205585	0.244447	0.040035	0.004927	0.043789
0.5	0.246094	0.175468	0.248149	0.240008	0.070626	0.002055	0.006086
0.4	0.200658	0.156293	0.199541	0.161149	0.044365	0.001117	0.039509
0.3	0.102919	0.100818	0.101938	0.062162	0.002101	0.000981	0.040757
0.2	0.026424	0.036089	0.026370	0.006860	0.009665	0.000054	0.019564
0.1	0.001488	0.003066	0.001513	0.000005	0.001578	0.000025	0.001483
0.05	0.000061	0.000158	0.000063	2.16E-12	0.000097	0.000002	0.000061
0.01	2.40E-08	7.54E-08	2.50E-08	1.39E-64	5.14E-08	1.03E-09	2.40E-08
		Average			**0.021474**	**0.001175**	**0.023272**

ตารางที่ 1 ค่าความน่าจะเป็นของการแจกแจงทวินาม ค่าประมาณความน่าจะเป็นของการแจกแจงปัวซง (UPD) การแจกแจงปัวซงนัยทั่วไป (GPD) การแจกแจงปรกติ (NOD) และความผิดพลาดของการประมาณแบบต่างๆ กรณีที่ $x = \frac{n}{2}$ (ต่อ)

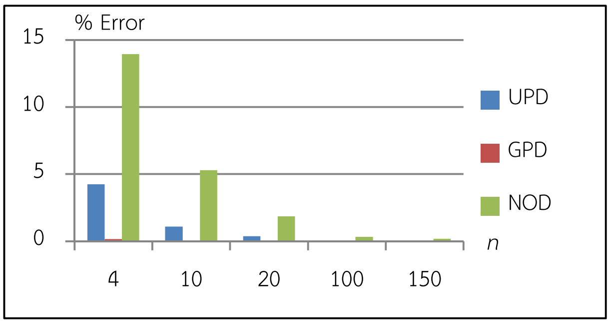
p	BID (*n, p*)	UPD (*np*)	GPD (λ, θ)	NOD (μ, σ)	UPD Error	GPD Error	NOD Error
n =20							
0.7	0.030817	0.066282	0.029944	0.045283	0.035465	0.000873	0.014466
0.6	0.117142	0.104838	0.119142	0.144046	0.012304	0.002000	0.026904
0.5	0.176197	0.125110	0.176932	0.174007	0.051087	0.000735	0.002190
0.4	0.117142	0.099262	0.116208	0.094961	0.017880	0.000934	0.022181
0.3	0.030817	0.041303	0.030730	0.017471	0.010486	0.000087	0.013346
0.2	0.002031	0.005292	0.002080	0.000303	0.003261	0.000049	0.001728
0.1	0.000006	0.000038	0.000007	5.72E-10	0.000032	4.72E-07	0.000006
0.05	1.08E-08	1.01E-07	1.20E-08	9.62E-22	9.06E-08	1.20E-09	1.08E-08
0.01	1.67E-15	2.31E-14	1.90E-15	4.01E-117	2.14E-14	2.31E-16	1.67E-15
		Average			**0.014502**	**0.000520**	**0.008980**
n =100							
0.7	0.000013	0.002351	0.000009	0.000010	0.002338	0.000004	0.000003
0.6	0.010338	0.023272	0.010216	0.012424	0.012934	0.000122	0.002086
0.5	0.079589	0.056325	0.079655	0.079390	0.023264	0.000066	0.000199
0.4	0.010338	0.017708	0.010364	0.008190	0.007370	0.000026	0.002148
0.3	0.000013	0.000221	0.000014	0.000004	0.000208	0.000001	0.000009
0.2	1.62E-11	7.63E-09	2.10E-11	2.36E-14	7.61E-09	4.79E-12	1.62E-11
0.1	5.20E-24	1.49E-19	8.87E-24	3.54E-41	1.49E-19	3.67E-24	5.20E-24
0.05	6.90E-38	1.97E-32	1.40E-37	4.17E-96	1.97E-32	7.10E-38	6.90E-38
0.01	6.10E-72	1.21E-65	1.44E-71	0.0E+00	1.21E-65	8.32R-72	6.10E-72
		Average			**0.005127**	**0.000024**	**0.000494**
n =150							
0.7	1.36E-07	0.000392	8.02E-08	7.12E-08	0.000392	5.60E-08	6.50E-08
0.6	0.003044	0.012220	0.002960	0.003585	0.009176	0.000084	0.000541
0.5	0.065039	0.046015	0.065075	0.064930	0.019024	0.000036	0.000109
0.4	0.003044	0.008111	0.003078	0.002364	0.005067	0.000034	0.000680
0.3	1.36E-07	0.000011	1.56E-07	2.75E-08	0.000011	1.97E-08	1.09E-07
0.2	1.89E-16	2.29E-12	2.84E-16	1.51E-20	2.29E-12	9.49E-17	1.89E-16
0.1	3.43E-35	1.99E-28	7.84E-35	1.45E-60	1.99E-28	4.41E-35	3.43E-35
0.05	5.24E-56	9.50E-48	1.55E-55	1.8E-142	9.50E-48	1.03E-55	5.24E-56
0.01	4.4E-107	1.45E-97	1.6E-106	0.0E+00	1.45E-97	1.2E-106	4.4E-107
		Average			**0.003741**	**0.000017**	**0.000148**

ตารางที่ 2 ค่าเฉลี่ยของความผิดพลาดของการประมาณค่าความน่าจะเป็นทวินามด้วยการแจกแจงแบบต่างๆ เมื่อ p = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7

กรณีที่ $x = 0$

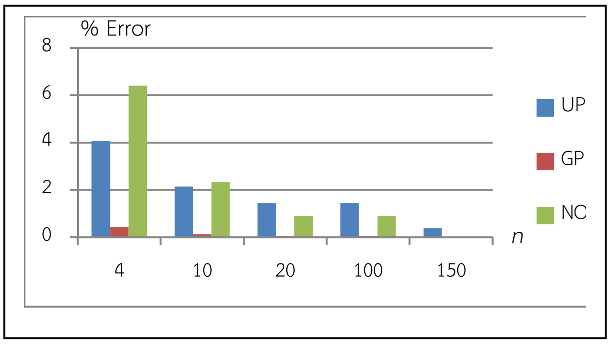
n	UPD Error	GPD Error	NOD Error
4	0.042491	0.001573	0.139583
10	0.010918	0.000191	0.052993
20	0.003644	0.000709	0.018660
100	0.000298	2.45E-07	0.003255
150	0.000197	1.99E-07	0.001858

กราฟแสดงร้อยละของความผิดพลาดของการประมาณ



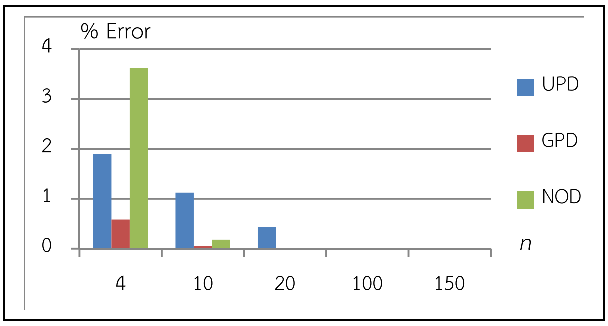
กรณีที่ $x = \frac{n}{2}$

n	UPD Error	GPD Error	NOD Error
4	0.040851	0.004340	0.064120
10	0.021474	0.001175	0.023272
20	0.014502	0.000520	0.008980
100	0.014502	0.000520	0.008980
150	0.003741	0.000017	0.000148



กรณีที่ $x = n$

n	UPD Error	GPD Error	NOD Error
4	0.018940	0.005803	0.036158
10	0.011242	0.000545	0.001784
20	0.004386	0.000129	0.000060
100	0.000015	2.66E-13	2.33E-12
150	7.35E-07	1.00E-18	4.23E-17



จากกราฟจะเห็นได้ว่าเมื่อค่า x ซึ่งคือจำนวนที่สำเร็จในการทดลอง n ครั้ง มีค่าเพิ่มขึ้นค่าความคลาดเคลื่อนของการประมาณทุกรูปแบบการแจกแจง (การแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ) มีค่าลดลงซึ่งกล่าวคือมีการประมาณค่าความน่าจะเป็นของทวินามได้ใกล้เคียงมากขึ้น

3.2 วิธีการคำนวณค่าความน่าจะเป็นของการแจกแจงทวินาม และการประมาณค่าความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจงแบบต่างๆ

ในหัวข้อนี้จะแสดงวิธีการคำนวณค่าความน่าจะเป็นของการแจกแจงทวินาม และการประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงแบบต่างๆ ซึ่งจะขอยกตัวอย่างการแสดงวิธีการคำนวณเพียงบางกรณีดังนี้

กรณีที่ $n = 20, x = \frac{n}{2} = 10$

1) $p = 0.7$; $BID(n,p) = \binom{20}{10}(0.7)^{10}(1-0.7)^{20-10} = 0.030817$

$$; \quad UPD(np) = \frac{e^{-14}(14)^{10}}{10!} = 0.066282$$

$$; \quad GPD(\lambda,\theta) = \frac{(25.56039)(25.56039+(-0.82574(10)))^{9}e^{-25.56039-(-0.82574(10))}}{10!} = 0.029944$$

$$; \quad NOD(\mu,\sigma^2) = \frac{(10\pm 0.5)-14}{\sqrt{14(1-0.7)}} = P(9.5 \le Z \le 10.5) = 0.045283$$

$$2)\ p = 0.5 \ ; \quad BID(n,p) = \binom{20}{10}(0.5)^{10}(1-0.5)^{10} = 0.176197$$

$$; \quad UPD(np) = \frac{e^{-10}(10)^{10}}{10!} = 0.125110$$

$$; \quad GPD(\lambda,\theta) = \frac{(14.14214)(14.14214+(-0.41421(10)))^{9}e^{-14.14214-(-0.41421(10))}}{10!} = 0.176932$$

$$; \quad NOD(\mu,\sigma^2) = \frac{(10\pm 0.5)-10}{\sqrt{10(1-0.5)}} = P(-0.22 \le Z \le 0.22) = 0.174007$$

$$3)\ p = 0.3 \ ; \quad BID(n,p) = \binom{20}{10}(0.3)^{10}(1-0.3)^{10} = 0.030817$$

$$; \quad UPD(np) = \frac{e^{-6}(6)^{10}}{10!} = 0.041303$$

$$; \quad GPD(\lambda,\theta) = \frac{(7.171372)(7.171372+(-0.19523(10)))^{9}e^{-7.171372-(-0.19523(10))}}{10!} = 0.030730$$

$$; \quad NOD(\mu,\sigma^2) = \frac{(10\pm 0.5)-6}{\sqrt{6(1-0.3)}} = P(1.71 \le Z \le 2.2) = 0.017471$$

$$4)\ p = 0.1 \ ; \quad BID(n,p) = \binom{20}{10}(0.1)^{10}(1-0.1)^{10} = 0.000006$$

$$; \quad UPD(np) = \frac{e^{-2}(2)^{10}}{10!} = 0.000038$$

$$; \quad GPD(\lambda,\theta) = \frac{(2.108185)(2.108185+(-0.05409(10)))^{9}e^{-2.108185-(-0.05409(10))}}{10!} = 0.000007$$

$$; \quad NOD(\mu,\sigma^2) = \frac{(10\pm 0.5)-2}{\sqrt{2(1-0.1)}} = P(5.59 \le Z \le 6.34) = 5.718 \text{ x } 10^{-10}$$

$$5)\ p = 0.01 \ ; \quad BID(n,p) = \binom{20}{10}(0.01)^{10}(1-0.01)^{10} = 1.671 \text{ x } 10^{-15}$$

$$; \quad UPD(np) = \frac{e^{-0.2}(0.2)^{10}}{10!} = 2.310 \text{ x } 10^{-14}$$

$$; \quad GPD(\lambda,\theta) = \frac{(0.201008)(0.201008+(-0.00504(10)))^{9}e^{-0.201008-(-0.00504(10))}}{10!} = 1.902 \text{ x } 10^{-15}$$

$$; \quad NOD(\mu,\sigma^2) = \frac{(10\pm 0.5)-0.2}{\sqrt{0.2(1-0.01)}} = P(20.9 \le Z \le 23.15) = 4.012 \text{ x } 10^{-117}$$

4. สรุปผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์ในการศึกษาการประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงปัวซง การแจกแจงปัวซงนัยทั่วไป และการแจกแจงปรกติ และหาวิธีการที่ให้ค่าของการประมาณใกล้เคียงกับค่าความน่าจะเป็นของการแจกแจงทวินามมากที่สุด เพื่อนำไปใช้คำนวณแทนการหาค่าความน่าจะเป็นของการแจกแจงทวินาม ในกรณีที่เกิดปัญหาในการคำนวณ คือ กรณีที่ n มีค่ามาก และ p มีค่าน้อย ($n \rightarrow \infty$, $p \rightarrow 0$)

ผลการวิจัยพบว่า การประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม ด้วยการแจกแจงปัวซงนัยทั่วไป ให้ค่าที่ใกล้เคียงกับค่าความน่าจะเป็นของการแจกแจงทวินามมากที่สุด ดังนั้นการประมาณค่าความน่าจะเป็นของการแจกแจงทวินาม ด้วยการ

แจกแจงปัวส์ซงนัยทั่วไป จึงเป็นวิธีการหนึ่งที่เหมาะสมในการนำมาคำนวณ แทนการหาค่าความน่าจะเป็นของการแจกแจงทวินามในกรณีที่เกิดปัญหาในการคำนวณ

5. ข้อเสนอแนะ

เนื่องจากโปรแกรม Microsoft Excel มีข้อจำกัดในการคำนวณหาค่าแฟกทอเรียลได้ในจำนวนที่ไม่มากนัก ดังนั้นหากผู้ที่สนใจในงานวิจัยนี้ควรหาโปรแกรม หรือวิธีการอื่นๆ ที่สามารถคำนวณค่าแฟกทอเรียลได้จำนวนมากๆ มาคำนวณแทน

มีข้อสนใจในการศึกษาเกี่ยวกับ dispersion คืออัตราส่วนของค่าเฉลี่ยกับความแปรปรวนของการเกิดเหตุ การณ์นั้นจะสอดคล้องหรือสัมพันธ์กับผลการประมาณค่า ความน่าจะเป็นของการแจกแจงทวินามด้วยการแจกแจง ปัวซงนัยทั่วไป ซึ่งควรเป็นการศึกษาวิจัยในครั้งต่อไป

6. เอกสารอ้างอิง

กฤษณะ เนียมมณี. (2542). ทฤษฎีความน่าจะเป็น. กรุงเทพฯ: พิทักษ์การพิมพ์.

สรชัย พิศาลบุตร. (2555). หลักสถิติ. กรุงเทพฯ: วิทยพัฒน์.

Suraporn, B. (2013). Parameters Estimation when Random Variables are Generalized Poisson Distribution. KKU Science Journal 41(2): 84 – 94.

Suraporn, B. (2014). Properties of Estimators for Generalized Poisson Distribution. KKU Science Journal 42(1): 373 – 382.

❑❑❑❑❑