

---

ข้อสังเกตบางประการเกี่ยวกับการเปรียบเทียบแบบจำลองของการแจกแจงปัวซองและการแจกแจงวิยุตที่เกี่ยวข้อง  
Some Remarks on the Model Comparison of Poisson Distribution and Discrete Related  
Distribution

มานัดธุ์ คำทอง\*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

Manad Khamkong\*

Department of Statistics, Faculty of Science, Chiang Mai University.

---

**บทคัดย่อ**

บทความนี้มุ่งเสนอแนะเกี่ยวกับการเปรียบเทียบแบบจำลองของการแจกแจงปัวซองและการแจกแจงวิยุตที่เกี่ยวข้องที่ควรพิจารณาคุณลักษณะต่างๆ ของการแจกแจงและความสัมพันธ์ระหว่างการแจกแจง ทั้งนี้เพื่อเป็นแนวทางให้นักวิจัยเลือกใช้การแจกแจงวิยุตที่เหมาะสมกับลักษณะของข้อมูลในแต่ละสถานการณ์

**คำสำคัญ :** การแจกแจงปัวซอง การแจกแจงวิยุต การเลือกแบบจำลอง

**Abstract**

The focus of this article is to recommend the model comparison of Poisson distribution and discrete related distribution by considering their characterizations and relationship. As a guide, the researchers chose the appropriate discrete distribution for the data in each case.

**Keywords :** Poisson distribution, Discrete distribution, Model selection

---

\*E-mail: manad.k@cmu.ac.th

## บทนำ

ปัจจุบันมีงานวิจัยหลายสาขาที่ตัวแปรสุ่มที่สนใจศึกษาเป็นตัวแปรสุ่มวิฤต (discrete random variable) ซึ่งข้อมูลมีลักษณะเป็นข้อมูลจำนวนนับ (count data) เช่น จำนวนอุบัติเหตุบนถนนสายหลักเชียงใหม่-เชียงใหม่ในแต่ละเดือน จำนวนการขอรับสินไหมทดแทนจากอุบัติเหตุทางรถยนต์ของบริษัทประกันภัยแห่งหนึ่งในรอบไตรมาส จำนวนผู้ป่วยที่เป็นโรคใช้หวัดคนที่มารับการรักษาในแต่ละเดือน จำนวนสินค้าที่ผลิตไม่ได้มาตรฐานตามที่กำหนดในรอบของการผลิต เป็นต้น เมื่อ ตัวแปรสุ่มที่สนใจศึกษามีการเก็บรวบรวมข้อมูลที่เป็นจำนวนนับในช่วงเวลาหรือขอบเขตที่สนใจศึกษา นักวิจัยส่วนใหญ่มีข้อสมมุติให้ตัวแปรสุ่มวิฤตที่สนใจศึกษามีการแจกแจงปัวซอง (Poisson distribution, Poi) ซึ่งนำเสนอโดยนักคณิตศาสตร์ชาวฝรั่งเศส Siméon Denis Poisson ในระหว่างปี ค.ศ. 1781-1840 โดยสมบัติของการแจกแจงปัวซอง คือ ค่าเฉลี่ยเท่ากับค่าความแปรปรวนซึ่งเท่ากับ  $\lambda$ ,  $E[X]=\text{Var}[X]=\lambda$  เรียกว่า อีควิพิชเพอซัน (equi-dispersion) แต่บ่อยครั้งที่ค่าความแปรปรวนของตัวแปรสุ่มมากกว่าค่าเฉลี่ย เรียกว่า โอเวอร์ดิชเพอซัน (over-dispersion) หรือค่าความแปรปรวนของตัวแปรสุ่มน้อยกว่าค่าเฉลี่ย เรียกว่า อันเดอร์ดิชเพอซัน (under-dispersion) ซึ่งนักวิจัยได้ประยุกต์ใช้การแจกแจงทวินามลบ (negative binomial distribution, NB) และการแจกแจงทวินาม (binomial distribution, Bin) เมื่อเกิดปัญหาโอเวอร์และอันเดอร์ดิชเพอซัน ตามลำดับ (Haight, 1967; Lawless, 1992; Cameron & Trivedi, 1998) และบ่อยครั้งที่ข้อมูลของตัวแปรสุ่มวิฤตที่สนใจศึกษาที่เป็นจำนวนนับในช่วงเวลาหรือขอบเขตที่สนใจศึกษานั้นเกิดขึ้นได้น้อยมากซึ่งเป็นอีกสาเหตุหนึ่งที่ทำให้เกิดปัญหาโอเวอร์ดิชเพอซันได้ ในปี ค.ศ. 1992 Lambert ได้ประยุกต์ใช้การแจกแจงปัวซองกรณีที่มีผลกระทบจากศูนย์ (zero-inflated Poisson distribution, ZIP) ในการหาปัจจัยที่ส่งผลต่อคุณภาพในกระบวนการผลิต ที่ไม่ได้มาตรฐาน และต่อมาได้นักสถิติหลายๆ ท่านพยายามพัฒนาการแจกแจงวิฤตให้มีความเหมาะสมกับข้อมูลเชิงจำนวนนับโดยมุ่งเน้นการขยายขอบเขตของการแจกแจงปัวซองให้มีลักษณะที่เป็นแบบทั่วไปที่ประกอบด้วยทั้งสามลักษณะคือ อีควิพิช-โอเวอร์และอันเดอร์ดิชเพอซัน เรียกว่า มิคซิดดิชเพอซัน (mixed-dispersion) เช่น การแจกแจงปัวซองวางนัยทั่วไป (generalized Poisson distribution, GP) การแจกแจงปัวซองวางนัยทั่วไปที่มีผลกระทบจากศูนย์ (zero-inflated generalized Poisson distribution, ZIGP) ในกระบวนการอนุมาณเชิงสถิติ เมื่อตัวอย่างสุ่ม  $X_1, X_2, \dots, X_n$  ขนาด  $n$  ที่สุ่มมาศึกษา ไม่ทราบการแจกแจงที่แท้จริง แต่จะ

สมมุติให้ตัวอย่างสุ่มนั้นถูกสุ่มมาศึกษาอย่างเป็นอิสระต่อกัน และมีรูปแบบการแจกแจงอย่างเดียวกัน (independent and identically distributed, iid) และพยายามสมมุติให้มีการแจกแจงที่เหมาะสมกับลักษณะการเกิดขึ้นของข้อมูลตัวอย่างสุ่มที่สุ่มมาศึกษาชุดนั้นๆ ซึ่งไม่ว่าจะเป็นการประมาณค่าและการทดสอบสมมุติฐานของแบบจำลองที่สร้างขึ้นจะมีความยากในการเลือก การแจกแจงวิฤตที่เหมาะสมกับลักษณะของข้อมูลจำนวนนับเนื่องจากแต่ละการแจกแจงที่กล่าวมา มีความสัมพันธ์และคล้ายคลึงกันในบางสถานการณ์

ดังนั้นบทความวิชาการนี้ผู้ศึกษาจึงมีความสนใจที่จะเสนอแนวทางในการเลือก การแจกแจงวิฤตที่เหมาะสมกับการวิเคราะห์ข้อมูลของแต่ละสถานการณ์ เพื่อให้ผู้สนใจเลือกใช้การแจกแจงวิฤตที่เหมาะสมกับงานวิจัยที่มีตัวแปรสุ่มที่สนใจศึกษาเป็นตัวแปรสุ่มวิฤตที่เกิดขึ้น ในช่วงเวลาหรือขอบเขตที่สนใจศึกษา

### การแจกแจงปัวซองและการแจกแจงวิฤตที่เกี่ยวข้อง

สำหรับตัวแปรสุ่มวิฤตที่สนใจศึกษามีการนับที่ข้อมูลในช่วงเวลาหรือขอบเขตที่สนใจศึกษา นักวิจัยส่วนใหญ่มีข้อสมมุติให้ตัวแปรสุ่มที่สนใจศึกษามีการแจกแจงปัวซอง แทนด้วย  $X \sim \text{Poi}(\lambda)$  มีฟังก์ชันมวลความน่าจะเป็น (probability mass function, pmf) ดังนี้

$$P(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (1)$$

สำหรับ  $x = 0, 1, 2, \dots$ ;  $\lambda > 0$  และที่อื่นๆ  $P(x; \lambda) = 0$  โดยค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ  $\mu = \lambda$  ซึ่งมีโมเมนต์ศูนย์กลาง (central moments) ที่สองเท่ากับโมเมนต์ศูนย์กลางที่สาม คือ  $\mu_2 = \mu_3 = \lambda$  ดังนั้นคุณลักษณะของค่าดัชนีวัดการกระจาย (index of dispersion, ID) คือ  $\mu_2 \mu^{-1} = 1$  และค่าสัมประสิทธิ์ความเบ้ (coefficient of skewness) คือ  $\mu_3 \mu^{-3/2} = \lambda^{-1/2}$  นั่นคือฟังก์ชันมวลความน่าจะเป็นของการแจกแจงปัวซองมีลักษณะเบ้ขวาและจะมีลักษณะสมมาตรเมื่อค่าเฉลี่ยมีค่ามากๆ ( $\lambda \rightarrow \infty$ )

สำหรับการแจกแจงที่มีคุณสมบัติอันเดอร์ดิชเพอซันของตัวแปรสุ่มวิฤตที่ใกล้เคียงกับการแจกแจงปัวซอง คือการแจกแจงทวินาม แทนด้วย  $X \sim \text{Bin}(r, p)$  มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

$$P(x; r, p) = \binom{r}{x} p^x (1-p)^{r-x} \quad (2)$$

สำหรับ  $x = 0, 1, 2, \dots, r$ ;  $0 < p < 1$  และที่อื่นๆ  $P(x; r, p) = 0$  โดยค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ  $rp$  ค่า  $\mu_2 = rp(1-p)$  และ  $\mu_3 = rp(1-p)(1-2p)$  ซึ่งคุณลักษณะของค่าดัชนีวัดการกระจาย คือ  $1-p$  และค่าสัมประสิทธิ์ความเบ้คือ  $1-2p(rp(1-p))^{-1/2}$  นั่นคือฟังก์ชัน

มวลความน่าจะเป็นของการแจกแจงทวินามจะมีลักษณะเบ้ขวา เมื่อ  $p < 0.50$  ลักษณะเบ้ซ้ายเมื่อ  $p > 0.50$  ลักษณะสมมาตรเมื่อ  $p = 0.50$  และเมื่อ  $r$  มีค่าเพิ่มขึ้น ( $r \rightarrow \infty$ ) ฟังก์ชันมวลความน่าจะเป็นของการแจกแจงทวินามจะมีลักษณะสมมาตร

**การแจกแจงที่มีคุณสมบัติโอเวอร์ดิซเพอซันของตัวแปรสุ่ม** วิทยุที่ขยายจากการแจกแจงปัวซองผสมการแจกแจงแกมมา (gamma distribution) คือการแจกแจงทวินามลบ แทนด้วย  $X \sim NB(r, p)$  มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

$$P(x; r, p) = \frac{\Gamma(x+r)}{x! \Gamma(r)} p^r (1-p)^x \quad (3)$$

สำหรับ  $x = 0, 1, 2, \dots$ ;  $r > 0$ ,  $\Gamma(\cdot)$  คือ ฟังก์ชันแกมมา,  $0 < p < 1$  และที่อื่นๆ  $P(x; r, p) = 0$  ที่มีค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ  $r(1-p)/p$  ค่า  $\mu_2 = r(1-p)p^{-2}$  และ  $\mu_3 = r(1-p)(2-p)p^{-3}$  จะมีคุณลักษณะของค่าดัชนีวัดการกระจาย คือ  $p^{-1}$  และค่าสัมประสิทธิ์ความเบ้คือ  $(2-p)(r(1-p))^{-1/2}$  นั่นคือฟังก์ชันมวลความน่าจะเป็นของการแจกแจงทวินามลบจะมีลักษณะเบ้ขวาและจะมีลักษณะสมมาตรเมื่อ  $r$  มีค่าเพิ่มขึ้น ( $r \rightarrow \infty$ )

ในกรณีที่เหตุการณ์ที่สนใจไม่ได้เกิดขึ้นบ่อยๆ ในช่วงเวลาหรือขอบเขตที่สนใจศึกษา เช่น จำนวนอุบัติเหตุบนถนนสายหลัก เชียงใหม่-เชียงใหม่ในแต่ละเดือน ซึ่งในบางเดือนอาจไม่มีอุบัติเหตุเกิดขึ้นเลยก็ได้ ดังนั้นเมื่อเก็บข้อมูลในระยะยาวจะมีจำนวนศูนย์หลายๆ เกิดขึ้น จึงได้มีการให้ความสำคัญกับเหตุการณ์ที่ไม่เกิดขึ้น เรียกรูปแบบการแจกแจงที่มีผลกระทบจากศูนย์ (zero inflation) นั่นคือการแจกแจงปัวซองที่มีผลกระทบจากศูนย์ (zero-inflated Poisson distribution, ZIP) แทนด้วย  $X \sim zip(\eta, \theta)$  มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

$$P(x; \eta, \theta) = \begin{cases} \theta + (1-\theta)e^{-\eta}, & x = 0, \\ (1-\theta) \frac{e^{-\eta} \eta^x}{x!}, & x = 1, 2, 3, \dots, 0 \leq \theta < 1, \end{cases} \quad (4)$$

และที่อื่นๆ  $P(x; \eta, \theta) = 0$  ถ้า  $\theta = 0$  การแจกแจงปัวซองที่มีผลกระทบจากศูนย์จะเป็นการแจกแจงปัวซอง ( $\lambda = \eta$ ) ซึ่งการแจกแจง ZIP มีค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ  $\eta(1-\theta)$  และโมเมนต์ศูนย์กลางที่สองและสาม คือ  $\mu_2 = \eta(1-\theta)(1+\theta\eta)$  และ  $\mu_3 = \eta(1-\theta)[1+3\theta\eta-(1-2\theta)\theta\eta^2]$  จะมีคุณลักษณะของค่าดัชนีวัดการกระจาย คือ  $1+\theta\eta$  และค่าสัมประสิทธิ์ความเบ้ คือ  $\eta(1-\theta)[1+3\theta\eta-(1-2\theta)\theta\eta^2] / [\eta(1-\theta)(1+\theta\eta)]^{3/2}$  นั่นคือฟังก์ชันมวลความน่าจะเป็นของการแจกแจง ZIP จะมีลักษณะเบ้ขวาเมื่อ  $1+3\theta\eta > (1-2\theta)\theta\eta^2$  และมีลักษณะเบ้ซ้ายเมื่อ  $1+3\theta\eta < (1-2\theta)\theta\eta^2$  ซึ่งการแจกแจงทวินามลบที่มีผลกระทบจากศูนย์ (zero-inflated

negative binomial distribution, ZINB) จะพัฒนาคล้ายกับการแจกแจงปัวซองที่มีผลกระทบจากศูนย์

**สำหรับการแจกแจงที่มีคุณสมบัติมิคซิดิซเพอซัน** ที่พัฒนาต่อจากการแจกแจงปัวซองที่จะนำเสนอในบทความนี้ ได้แก่ การแจกแจงปัวซองวางนัยทั่วไป (generalized Poisson distribution, GP) แทนด้วย  $X \sim GP(\eta, \theta)$  มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

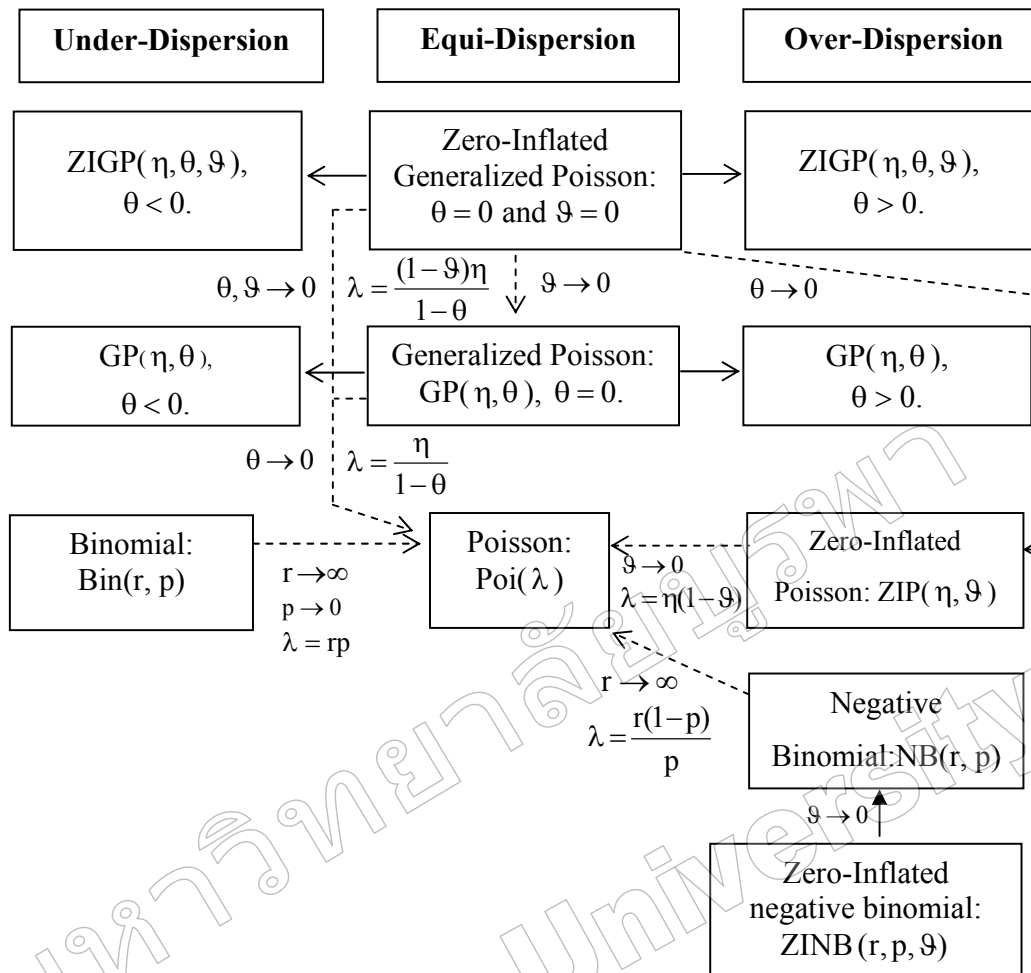
$$P(x; \eta, \theta) = \begin{cases} \frac{\eta(\eta+\theta x)^{x-1} e^{-(\eta+\theta x)}}{x!}, & x = 0, 1, 2, \dots \\ 0, & x > m \text{ when } \theta < 0 \end{cases} \quad (5)$$

สำหรับ  $\eta > 0$ ,  $\max(-1, -\eta/m) \leq \theta \leq 1$  และ  $m \geq 4$  เป็นจำนวนจริงบวกที่มากที่สุดสำหรับ  $\theta < 0$  และที่อื่นๆ  $P(x; \eta, \theta) = 0$  ถ้า  $\theta = 0$  การแจกแจง GP จะเป็นการแจกแจงปัวซอง ( $\lambda = \eta$ ) ซึ่งการแจกแจง GP จะมีค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ  $\eta(1-\theta)^{-1}$  ค่า  $\mu_2 = \eta(1-\theta)^{-3}$  และ  $\mu_3 = \eta(1-2\theta)(1-\theta)^{-5}$  ค่าคุณลักษณะของค่าดัชนีวัดการกระจายคือ  $(1-\theta)^{-2}$  นั่นคือ เมื่อ  $\theta < 0$  จะเป็นกรณีที่ค่าเฉลี่ยมากกว่าค่าความแปรปรวน (อันเดอร์ดิซเพอซัน) และ  $\theta > 0$  จะเป็นกรณีที่ค่าเฉลี่ยน้อยกว่าค่าความแปรปรวน (โอเวอร์ดิซเพอซัน) สำหรับค่าสัมประสิทธิ์ความเบ้คือ  $(1+2\theta)(\eta(1-\theta))^{-1/2}$  ดังนั้นฟังก์ชันมวลความน่าจะเป็นของการแจกแจง GP จะมีขึ้นอยู่กับค่าพารามิเตอร์  $\eta$  และ  $\theta$  เมื่อนำหนดค่า  $\theta$  ใดๆ ค่าสัมประสิทธิ์ความเบ้จะเข้าใกล้ศูนย์เมื่อ  $\eta$  มีค่าเพิ่มขึ้น นั่นคือจะมีลักษณะสมมาตรเมื่อ  $\eta$  มีค่ามากๆ ( $\eta \rightarrow \infty$ ) ส่วนฟังก์ชันมวลความน่าจะเป็นและคุณลักษณะที่สำคัญของการแจกแจงปัวซองวางนัยทั่วไปที่มีผลกระทบจากศูนย์ (zero-inflated generalized Poisson distribution, ZIGP) จะพัฒนาคล้ายกับการแจกแจงปัวซองที่มีผลกระทบจากศูนย์

Joe และ Zhu (Joe & Zhu, 2005) พบว่าภายใต้การกำหนดสถานการณ์ที่ให้ค่าเฉลี่ยและความแปรปรวนของการแจกแจงเท่ากัน การแจกแจงปัวซองวางนัยทั่วไปจะมีปลายหางที่หนากว่า (heavy tail) การแจกแจงทวินามลบ แต่การแจกแจงปัวซองวางนัยทั่วไปจะมีฟังก์ชันมวลความน่าจะเป็นที่จุด  $x$  เท่ากับศูนย์น้อยกว่าการแจกแจงทวินามลบ ซึ่งสามารถสรุปความสัมพันธ์ระหว่างการแจกแจงปัวซองและการแจกแจงวิทยุที่กล่าวมาได้ดังรูปภาพที่ 1

#### การทดสอบความเหมาะสมของการแจกแจงวิทยุ

สมมติตัวแปรสุ่มวิทยุ  $X_1, X_2, \dots, X_n$  เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันมวลความน่าจะเป็น,  $P(x; \theta)$ ,  $\theta \in \Theta$  สำหรับการทดสอบความเหมาะสมของตัวอย่างสุ่มที่สุ่มมาศึกษาสอดคล้องกับการแจกแจงวิทยุแบบใดมีสมมุติฐานสำหรับการทดสอบ คือ



ภาพที่ 1 ความสัมพันธ์ระหว่างการแจกแจงปัวซองและการแจกแจงวิตซ์ที่เกี่ยวข้อง

$H_0$  : ตัวอย่างสุ่มมีการแจกแจงวิตซ์ที่ต้องการทดสอบ เทียบกับ  $H_1$  : ตัวอย่างสุ่มไม่มีการแจกแจงวิตซ์ที่ต้องการทดสอบ สำหรับสถิติที่ทดสอบความเหมาะสม (goodness of fit test) ของตัวอย่างสุ่มสามารถประยุกต์ใช้การทดสอบด้วยไคกำลังสอง (chi-square test) การทดสอบโคลโมโกรอฟ-สเมอร်นอฟ (Kolmogorov-Smirnov test) และการทดสอบคราเมอ์-วอนมิเชส (Cramer-von Mises test) เป็นต้น โดยการทดสอบการแจกแจงปัวซองที่พัฒนาจากสถิติทดสอบด้วยไคกำลังสองที่อาศัยอัตราส่วนระหว่างค่าความแปรปรวนและค่าเฉลี่ยจะมีประสิทธิภาพของการทดสอบเมื่อค่าดัชนีวัดการกระจาย (ID) มีค่าห่างจาก 1 ไปมากๆ ทั้งกรณีโอเวอร์และอันเดอร์ดิซเพอชั่น แต่เมื่อค่าดัชนีวัดการกระจายเข้าใกล้ 1 จะไม่สามารถแยกความแตกต่างระหว่างการแจกแจงปัวซองและการแจกแจงอื่น ๆ ที่เกี่ยวข้องได้ดังตารางที่ 1 (Karlis & Xekalaki, 2000; Gurtler & Henze, 2000; Meintanis & Nikitin, 2008) โดยส่วนใหญ่ในการเก็บรวบรวมข้อมูลจำนวนนับ

ในขอบเขตที่สนใจศึกษาจะมีปัญหาที่ค่าความแปรปรวนของตัวแปรสุ่มวิตซ์มากกว่าค่าเฉลี่ย ดังนั้นงานวิจัยส่วนใหญ่จึงมุ่งไปที่การทดสอบและการประมาณค่าพารามิเตอร์ของตัวแบบในกรณีที่เกิดปัญหาโอเวอร์ดิซเพอชั่น ซึ่งจะแบ่งการทดสอบออกเป็น 2 ส่วนใหญ่ๆ (Yang *et al.*, 2010; Garay *et al.*, 2011) ดังนี้

1) การทดสอบความเหมาะสมระหว่างการแจกแจงปัวซองกับการแจกแจงทวินามลบและการแจกแจงปัวซองวางนัยทั่วไป โดยใช้การทดสอบสคอ์ (score test) ในการทดสอบค่าพารามิเตอร์ดิซเพอชั่น ( $\phi$ ) ของแต่ละการแจกแจง เช่น การแจกแจงทวินาม  $\phi = r^{-1}$  สมมุติฐานในการทดสอบคือ  $H_0 : \phi = 0$  เทียบกับ  $H_1 : \phi > 0$  และการแจกแจงปัวซองวางนัยทั่วไป  $\phi = (1-\theta)^2$  มีสมมุติฐานในการทดสอบคือ  $H_0 : \phi = 1$  เทียบกับ  $H_1 : \phi > 1$

2) การทดสอบความเหมาะสมระหว่างการแจกแจงปัวซองกับการแจกแจงที่มีผลกระทบจากศูนย์ของการแจกแจงทวินามลบ การแจกแจงปัวซองและการแจกแจงปัวซองวางนัยทั่วไปโดยใช้การ

ตารางที่ 1 ค่าคุณลักษณะที่สำคัญของการแจกแจงปัวซองและการแจกแจงวิฤตที่เกี่ยวข้อง

Type of discrete Distribution	Mean	Variance	Index of Dispersion	Skewness
1. Under-dispersion Binomial: $\text{Bin}\left(r, \frac{\lambda}{r}\right)$	$\lambda$	$\lambda - \frac{\lambda^2}{r}$	$1 - \frac{\lambda}{r}$	$\frac{1 - \frac{2\lambda}{r}}{\sqrt{\lambda\left(1 - \frac{\lambda}{r}\right)}}$
2. Over-dispersion Negative binomial: $\text{NB}\left(r, \frac{r}{r+\lambda}\right)$	$\lambda$	$\lambda + \frac{\lambda^2}{r}$	$1 + \frac{\lambda}{r}$	$\frac{r + 2\lambda}{\sqrt{r\lambda(r + \lambda)}}$
Zero-inflated negative binomial: $\text{ZINB}\left(r, \frac{r}{r+\lambda}, \theta\right)$	$(1 - \theta)\lambda$	$(1 - \theta)\lambda(1 + (r^{-1} + \theta)\lambda)$	$1 + (r^{-1} + \theta)\lambda$	$\frac{1 + (r^{-1} + \theta)3\lambda + (2r^{-1} + 3\theta)r^{-1}\lambda^2 - (1 - 2\theta)\theta\lambda^2}{\sqrt{\lambda(1 - \theta)(1 + (r^{-1} + \theta)\lambda)^3}}$
Zero-inflated Poisson: ZIP $(\lambda, \theta)$	$(1 - \theta)\lambda$	$(1 - \theta)\lambda(1 + \theta\lambda)$	$1 + \theta\lambda$	$\frac{1 + (3 + (2\theta - 1)\lambda)\theta\lambda}{\sqrt{(1 - \theta)\lambda(1 + \theta\lambda)^3}}$
3. Equi- dispersion Poisson: $\text{Poi}(\lambda)$	$\lambda$	$\lambda$	1.00	$\frac{1}{\sqrt{\lambda}}$
4. Mixed-dispersion Generalized Poisson: $\text{GP}(\lambda(1-\theta), \theta)$	$\lambda$	$\frac{\lambda}{(1-\theta)^2}$	$\frac{1}{(1-\theta)^2}$	$\frac{1 + 2\theta}{(1-\theta)\sqrt{\lambda}}$
Zero-inflated Generalized Poisson: $\text{ZIGP}\left(\frac{(1-\theta)\lambda}{(1-\theta)}, \theta, \theta\right)$	$(1 - \theta)\lambda$	$(1 - \theta)\lambda((1-\theta)^{-2} + \theta\lambda)$	$(1 - \theta)^{-2} + \theta\lambda$	$\frac{(3(1-\theta)^{-1} - 2)(1-\theta)^{-3} + (3(1-\theta)^{-2} + (2\theta - 1)\lambda)\theta\lambda}{\sqrt{(1 - \theta)\lambda((1-\theta)^{-2} + \theta\lambda)^3}}$

ทดสอบสคออร์ (score test) ในการทดสอบค่าพารามิเตอร์ผลกระทบจากศูนย์ ( $\theta$ ) ของแต่ละการแจกแจง ซึ่งมีสมมติฐานในการทดสอบคือ  $H_0 : \theta = 0$  เทียบกับ  $H_1 : \theta > 0$

เมื่อค่าพารามิเตอร์ดิซเพอชั่น ( $\theta$ ) มีค่าเข้าใกล้ 1 ค่าพารามิเตอร์ผลกระทบจากศูนย์ ( $\theta$ ) เข้าใกล้ 0 และขนาดตัวอย่าง  $n$  ที่มีค่าไม่ใหญ่พอ การทดสอบสคออร์จะมีประสิทธิภาพต่ำในการจำแนกระหว่างการแจกแจงปัวซองและการแจกแจงวิฤตที่ต้องการทดสอบ จึงได้มีการพิจารณาคคุณลักษณะของค่าสัมประสิทธิ์

ความเบ้ของการแจกแจงวิฤตมาช่วยในการคัดเลือกการแจกแจงที่เหมาะสมกับข้อมูลจำนวนนับที่สนใจศึกษา (Joe & Zhu, 2005; Puig & Valero, 2006; Khamkong, 2010; Nikoloulopoulos & Karlis, 2012) รวมถึงสามารถใช้เกณฑ์การคัดเลือกแบบจำลองด้วยค่า AIC (Akaike information criterion; Akaike, 1973) ซึ่งเป็นการใช้สารสนเทศ Kullback-Leibler ในการประเมินแบบจำลองทางสถิติที่ประมาณการแจกแจงที่แท้จริงของข้อมูลและคุณสมบัติของการแจกแจงนั้นๆ โดยมีเกณฑ์การเลือกแบบจำลอง

ที่เหมาะสมที่ให้ค่า AIC ต่ำที่สุด ทั้งนี้มีความเชื่อว่าแบบจำลองนั้นได้เก็บสารสนเทศที่สำคัญจากข้อมูลไว้ครบถ้วนแล้ว ในกรณีที่มีการแจกแจงเป็นเครือข่ายกัน (nested) เช่น การทดสอบความเหมาะสมของการแจกแจง ZIGP GP และ Poi แต่ถ้าการแจกแจงไม่เป็นเครือข่ายกันจะประยุกต์ใช้การทดสอบ Vuong (Vuong, 1989) เช่น การทดสอบความเหมาะสมระหว่างการแจกแจง GP และ NB เนื่องจากสถิติทดสอบ Vuong ได้พัฒนาการทดสอบจากหลักการทดสอบอัตราส่วนควรจะเป็น (likelihood ratio test) ภายใต้เงื่อนไขทั่วไปที่สามารถทดสอบได้ทั้งแบบจำลองที่เป็นเครือข่ายและไม่เป็นเครือข่าย ซึ่งจะให้กำลังการทดสอบ (power of test) ที่สูงในการจำแนกระหว่างแบบจำลอง

### ข้อเสนอแนะ

ตัวแปรสุ่มวิฤตที่เกิดขึ้นในขอบเขตที่สนใจศึกษาที่เป็นข้อมูลจำนวนนับมีการประยุกต์ใช้ในหลายๆ สาขาวิชา เช่น วิทยาศาสตร์ สิ่งแวดล้อม วิศวกรรมศาสตร์ เศรษฐศาสตร์ ระบาดวิทยา เป็นต้น ซึ่งนักวิจัยมีข้อสมมุติให้ตัวแปรสุ่มวิฤตที่สนใจศึกษาภายใต้ขอบเขตที่สนใจศึกษา มีการแจกแจงปัวซองและบ่อยครั้งที่ตัวแปรสุ่มวิฤตที่สนใจศึกษานั้นไม่เป็นไปตามข้อสมมุติของการแจกแจงปัวซองโดยจะเกิดปัญหาค่าเฉลี่ยของตัวแปรสุ่มมากกว่าค่าความแปรปรวน (อันเดอร์ดิซเพอชั่น) นักวิจัยจะหลีกเลี่ยงไปใช้การแจกแจงทวินามหรือการแจกแจงปัวซองวงนัยทั่วไป แต่สำหรับกรณีที่ค่าเฉลี่ยของตัวแปรสุ่มน้อยกว่าค่าความแปรปรวน (โอเวอร์ดิซเพอชั่น) นักวิจัยจะหลีกเลี่ยงไปใช้การแจกแจงทวินามลบ การแจกแจงปัวซองวงนัยทั่วไป ซึ่งในบางกรณีที่ตัวแปรสุ่มวิฤตที่สนใจภายใต้ขอบเขตที่สนใจศึกษาเกิดขึ้นได้น้อยจะส่งผลให้มีค่าที่เป็นศูนย์เป็นจำนวนมากซึ่งเป็นอีกสาเหตุหนึ่งที่ทำให้ค่าความแปรปรวนมากกว่าค่าเฉลี่ยนักวิจัยจะประยุกต์ใช้การแจกแจงวิฤตที่มีผลกระทบจากศูนย์ของการแจกแจงปัวซอง การแจกแจงทวินามลบ และการแจกแจงปัวซองวงนัยทั่วไป โดยมีวิธีการในการทดสอบความเหมาะสมของการแจกแจงด้วยการทดสอบไคกำลังสองและการทดสอบโคลโมโกรอฟ-สมอร์นอฟ ซึ่งถ้าตัวอย่างสุ่มชุดหนึ่งมีการแจกแจงที่เหมาะสมมากกว่าหนึ่งการแจกแจงนักวิจัยควรพิจารณาเลือกการแจกแจงด้วยคุณลักษณะอื่นๆ ของการแจกแจง ตัวอย่างเช่น หากการแจกแจงที่เป็นเครือข่ายกันจะเลือกการแจกแจงที่ให้ค่า AIC ต่ำที่สุด เช่น การทดสอบความเหมาะสมระหว่างการแจกแจงทวินามลบที่มีผลกระทบจากศูนย์และการแจกแจงทวินามลบ หรือการทดสอบความเหมาะสมของการแจกแจงปัวซองวงนัยทั่วไปที่มีผลกระทบจากศูนย์ การแจกแจงปัวซองวงนัยทั่วไป การแจกแจงปัวซองที่มีผลกระทบจากศูนย์ และ

การแจกแจงปัวซอง แต่สำหรับการแจกแจงที่ไม่เป็นเครือข่ายกัน การคัดเลือกการแจกแจงที่เหมาะสมควรเลือกการทดสอบ Vuong เช่น การเลือกความเหมาะสมระหว่างการแจกแจงปัวซองวงนัยทั่วไปกับการแจกแจงทวินามลบ

### กิตติกรรมประกาศ

ผู้เขียนขอขอบพระคุณผู้ประเมินบทความทุกท่านที่ได้ให้คำแนะนำที่เป็นประโยชน์และข้อคิดดีๆ ในการเขียนบทความครั้งนี้

### เอกสารอ้างอิง

- Akaike, H. (1973). Information theory and extension of the maximum likelihood principle. In *Proceeding the second international symposium on information theory*. (pp. 267-281) B.N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest.
- Balakrishnan, N. & Nevzorov, V.B. (1956). *A Primer on Statistical Distributions*. New York: John Wiley & Sons.
- Cameron, A.C. & Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- Consul, P.C. (1989). *Generalized Poisson Distributions: Properties and applications*. New York: Marcel Dekker.
- Evans, M., Hastings, N. & Peacock, B. (2000). *Statistical Distributions*. (3<sup>rd</sup> Ed.). New York: John Wiley & Sons.
- Garay, A.M., Hashimoto, E.M., Ortega, M.M. & Lachos, V.H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics and Data Analysis*, 55, 1304-1318.
- Gordon, H. (1997). *Discrete Probability*. New York: Springer-Verlag.
- Gupta, P.L., Gupta, R.C., & Tripathi, R.C. (1996). Analysis of zero-adjusted count data. *Computational Statistics & Data Analysis*, 23, 207-218.

- Gurtler, N. & Henze, N. (2000). Recent and classical goodness-of-fit tests the Poisson distribution. *Journal of Statistical Planning and Inference*, 90, 207-225.
- Haight, F.A. (1967). *Handbook of the Poisson Distribution*. New York: John Wiley & Sons.
- Joe, H. & Zhu, R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal*, 47, 219-229.
- Johnson, N.L., Kotz, S. & Kemp, A. W. (1992). *Univariate Discrete Distributions*. (2<sup>m</sup> Ed.). New York: John Wiley & Sons.
- Karlis, D. & Xekalaki, E. (2000). A Simulation comparison of several procedures for testing the Poisson assumption. *The Statistician*, 49, 355-382.
- Khamkong, M. (2010). Comparing models for fitting zero-inflated data. In *Proceeding the 6th IMT-GT Conference on Mathematics and its Applications*. (pp. 362-366). Kuala Lumpur: Universiti Tunku Abdul Rahman.
- Lambert, D. (1992). Zero-inflated Poisson Regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- Lawless, J.F. (1992). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15, 209-225.
- Meintanis, S.G. & Nikitin, Y.Y. (2008). A class of count models and a new consistent test for the Poisson distribution. *Journal of Statistical Planning and Inference*, 138, 3722-3732.
- Nikoloulopoulos, A.K. & Karlis, D. (2008). On modeling count data: a comparison of some well-known discrete distributions. *Journal of Statistical Computation and Simulation*, 78, 437-457.
- Puig, P. & Valero, J. (2006). Count data distribution: some characterizations with applications. *Journal of American Statistical Association*, 101, 332-340.
- Vuong, Q.H. (1988). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-333.
- Yang, Z., Hardin, J.W. & Addy, C.L. (2010). Score tests for zero-inflation in overdispersed count data. *Communications in Statistics -Theory Methods*, 39, 2008-2030.
- Zelterman, D. (2004). *Discrete Distribution: Applications in the Health Sciences*. New York: John Wiley & Sons.