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

มานัดถุ์ คำกอง*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

Manad Khamkong*

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

บทคัดย่อ

บทความนี้มุ่งเสนอแนะเกี่ยวกับการเปรียบเทียบแบบจำลองของการแจกแจงปัวซงและการแจกแจงวิยุตที่เกี่ยวข้องที่ควรพิจารณาคุณลักษณะต่างๆ ของการแจกแจงและความสัมพันธ์ระหว่างการแจกแจง ทั้งนี้เพื่อเป็นแนวทางให้นักวิจัยเลือกใช้การแจกแจงวิยุตที่เหมาะกับลักษณะของข้อมูลในแต่ละสถานการณ์

คำสำคัญ : การแจกแจงปัวซง 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

ปัจจุบันมีงานวิจัยหลายสาขาที่ตัวแปรสุ่มที่สนใจศึกษาเป็นตัวแปรสุ่มวิยุต (discrete random variable) ซึ่งข้อมูลมีลักษณะเป็นข้อมูลจำนวนนับ (count data) เช่น จำนวนอุบัติเหตุบนถนนสายหลักเชียงใหม่-เชียงรายในแต่ละเดือน จำนวนการขอรับสินไหมทดแทนจากอุบัติเหตุทางรถยนต์ของบริษัทประกันภัยแห่งหนึ่งในรอบไตรมาส จำนวนผู้ป่วยที่เป็นโรคไข้หวัดนกที่มารับการรักษาในแต่ละเดือน จำนวนสินค้าที่ผลิตไม่ได้มาตรฐานตามที่กำหนดในรอบของการผลิต เป็นต้น เมื่อ ตัวแปรสุ่มที่สนใจศึกษามีการเก็บรวบรวมข้อมูลที่เป็นจำนวนนับในช่วงเวลาหรือขอบเขตที่สนใจศึกษา นักวิจัยส่วนใหญ่มีข้อสมมุติให้ตัวแปรสุ่มวิยุตที่สนใจศึกษามีการแจกแจงปัวซง (Poisson distribution, Poi) ซึ่งนำเสนอโดยนักคณิตศาสตร์ชาวฝรั่งเศส Siméon Denis Poisson ในระหว่างปี ค.ศ. 1781 -1840 โดยสมบัติของการแจกแจงปัวซง คือ ค่าเฉลี่ยเท่ากับค่าความแปรปรวนซึ่งเท่ากับ λ, $E[X]=Var[X]=\lambda$ เรียกว่า อิควีพดิซเพอชั่น (equi-dispersion) แต่บ่อยครั้งที่ค่าความแปรปรวนของตัวแปรสุ่มมากกว่าค่าเฉลี่ย เรียกว่า โอเวอร์ดิซเพอชั่น (over-dispersion) หรือค่าความแปรปรวนของตัวแปรสุ่มน้อยกว่าค่าเฉลี่ย เรียกว่า อันเดอร์ดิซเพอชั่น (under-dispersion) ซึ่งนักวิจัยได้ประยุกต์ใช้การแจกแจงทวินามลบ (negative binomial distribution, NB) และการแจกแจงทวินาม (binomial distribution, Bin) เมื่อเกิดปัญหาโอเวอร์และอันเดอร์ดิซเพอชั่น ตามลำดับ (Haight, 1967; Lawless, 1992; Cameron & Trivedi, 1998) และบ่อยครั้งที่ข้อมูลของตัวแปรสุ่มวิยุตที่สนใจศึกษาที่เป็นจำนวนนับในช่วงเวลาหรือขอบเขตที่สนใจศึกษานั้นเกิดขึ้นได้น้อยมากซึ่งเป็นอีกสาเหตุหนึ่งที่ทำให้เกิดปัญหาโอเวอร์ดิซเพอชั่นได้ ในปี ค.ศ. 1992 Lambert ได้ประยุกต์ใช้การแจกแจงปัวซงกรณีที่มีผลกระทบจากศูนย์ (zero-inflated Poisson distribution, ZIP) ในการหาปัจจัยที่ส่งผลต่อคุณภาพในกระบวนการผลิต ที่ไม่ได้มาตรฐาน และต่อมาได้มีนักสถิติหลายๆ ท่านพยายามพัฒนาการแจกแจงวิยุตให้มีความเหมาะสมกับข้อมูลเชิงจำนวนนับโดยมุ่งเน้นการขยายขอบเขตของการแจกแจงปัวซงให้มีลักษณะที่เป็นแบบทั่วไปที่ประกอบด้วยทั้งสามลักษณะคือ อิควีพ-โอเวอร์และอันเดอร์ดิซเพอชั่น เรียกว่า มิคซดิซเพอชั่น (mixed-dispersion) เช่น การแจกแจงปัวซงวางนัยทั่วไป (generalized Poisson distribution, GP) การแจกแจงปัวซงวางนัยทั่วไปที่มีผลกระทบจากศูนย์ (zero-inflated generalized Poisson distribution, ZIGP) ในกระบวนการอนุมานเชิงสถิติ เมื่อตัวอย่างสุ่ม X1, X_2, ..., X_n ขนาด n ที่สุ่มมาศึกษา ไม่ทราบการแจกแจงที่แท้จริง แต่จะสมมุติให้ตัวอย่างสุ่มนั้นถูกสุ่มมาศึกษาอย่างเป็นอิสระต่อกันและมีรูปแบบการแจกแจงอย่างเดียวกัน (independent and identically distributed, iid) และพยายามสมมุติให้มีการแจกแจงที่เหมาะสมกับลักษณะการเกิดขึ้นของข้อมูลตัวอย่างสุ่มที่สุ่มมาศึกษาชุดนั้นๆ ซึ่งไม่ว่าจะเป็นการประมาณค่าและการทดสอบสมมุติฐานของแบบจำลองที่สร้างขึ้นจะมีความยากในการเลือก การแจกแจงวิยุตที่เหมาะสมกับลักษณะของข้อมูลจำนวนนับเนื่องจากแต่ละการแจกแจงที่กล่าวมา มีความสัมพันธ์และคล้ายคลึงกันในบางสถานการณ์

ดังนั้นบทความวิชาการนี้ผู้ศึกษาจึงมีความสนใจที่จะเสนอแนวทางในการเลือก การแจกแจงวิยุตที่เหมาะสมกับการวิเคราะห์ข้อมูลของแต่ละสถานการณ์ เพื่อให้ผู้สนใจเลือกใช้การแจกแจงวิยุตที่เหมาะสมกับงานวิจัยที่มีตัวแปรสุ่มที่สนใจศึกษาเป็นตัวแปรสุ่มวิยุตที่เกิดขึ้น ในช่วงเวลาหรือขอบเขตที่สนใจศึกษา

การแจกแจงปัวซงและการแจกแจงวิยุตที่เกี่ยวข้อง

สำหรับตัวแปรสุ่มวิยุตที่สนใจศึกษามีการบันทึกข้อมูลในช่วงเวลาหรือขอบเขตที่สนใจศึกษา นักวิจัยส่วนใหญ่มีข้อสมมุติให้ตัวแปรสุ่มที่สนใจศึกษามีการแจกแจงปัวซง แทนด้วย $X \sim Poi(\lambda)$ มีฟังก์ชันมวลความน่าจะเป็น (probability mass function, pmf) ดังนี้

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

สำหรับ x = 0, 1, 2, ... ; $\lambda > 0$ 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 $\mu_3\mu_2^{-3/2} = \lambda^{-1/2}$ นั่นคือฟังก์ชันมวลความน่าจะเป็นของการแจกแจงปัวซงมีลักษณะเบ้ขวาและจะมีลักษณะสมมาตรเมื่อค่าเฉลี่ยมีค่ามากๆ ($\lambda \to \infty$)

สำหรับการแจกแจงที่มีคุณสมบัติอันเดอร์ดิซเพอชั่นของตัวแปรสุ่มวิยุตที่ใกล้เคียงกับการแจกแจงปัวซง คือการแจกแจงทวินาม แทนด้วย $X \sim Bin(r,p)$ มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

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

สำหรับ x = 0, 1, 2, ... , 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

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

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

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

สำหรับ $x = 0, 1, 2, \ldots$; $r > 0$, $\Gamma(.)$ 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 จำนวนอุบัติเหตุบนถนนสายหลักเชียงใหม่-เชียงรายในแต่ละเดือน ซึ่งในบางเดือนอาจไม่มีอุบัติเหตุเกิดขึ้นเลยก็ได้ ดังนั้นเมื่อเก็บข้อมูลในระยะยาวจะมีจำนวนศูนย์มากๆ เกิดขึ้น จึงได้มีการให้ความสำคัญกับเหตุการณ์ที่ไม่เกิดขึ้นเรียกการแจกแจงที่มีผลกระทบจากศูนย์ (zero inflation) นั้นว่าการแจกแจงปัวซงกรณีที่มีผลกระทบจากศูนย์ (zero-inflated Poisson distribution, ZIP) แทนด้วย $X \sim zip(\eta,\vartheta)$ มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

$$P(x; \eta, \vartheta) = \begin{cases} \vartheta + (1-\vartheta)e^{-\eta}, & x = 0, \\ (1-\vartheta)\dfrac{e^{-\eta}\eta^x}{x!}, & x = 1, 2, 3, \ldots, 0 \leq \vartheta < 1, \end{cases} \qquad (4)$$

และที่อื่นๆ $P(x; \eta, \vartheta) = 0$ ถ้า $\vartheta = 0$ การแจกแจงปัวซงกรณีที่มีผลกระทบจากศูนย์จะเป็นการแจกแจงปัวซง ($\lambda = \eta$) ซึ่งการแจกแจง ZIP มีค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ $\eta(1-\vartheta)$ และโมเมนต์ศูนย์กลางที่สองและสาม คือ $\mu_2 = \eta(1-\vartheta)(1+\vartheta\eta)$ และ $\mu_3 = \eta(1-\vartheta)\left[1+3\vartheta\eta-(1-2\vartheta)\vartheta\eta^2\right]$ จะมีคุณลักษณะของค่าดัชนีวัดการกระจาย คือ $1+\vartheta\eta$ และค่าสัมประสิทธิ์ความเบ้ คือ $\eta(1-\vartheta)\left[1+3\vartheta\eta-(1-2\vartheta)\vartheta\eta^2\right]\left[\eta(1-\vartheta)(1+\vartheta\eta)\right]^{-3/2}$ นั่นคือฟังก์ชันมวลความน่าจะเป็นของการแจกแจง ZIP จะมีลักษณะเบ้ขวาเมื่อ $1+3\vartheta\eta>(1-2\vartheta)\vartheta\eta^2$ และมีลักษณะเบ้ซ้ายเมื่อ $1+3\vartheta\eta<(1-2\vartheta)\vartheta\eta^2$ ซึ่งการแจกแจงทวินามลบที่มีผลกระทบจากศูนย์ (zero-inflated negative binomial distribution, ZINB) จะพัฒนาคล้ายกับการแจกแจงปัวซงกรณีที่มีผลกระทบจากศูนย์

สำหรับการแจกแจงที่มีคุณสมบัติมิคซดิซเพอชั่นที่พัฒนาต่อจากการแจกแจงปัวซงที่จะนำเสนอในบทความนี้ ได้แก่ การแจกแจงปัวซงวางนัยทั่วไป (generalized Poisson distribution, GP) แทนด้วย $X \sim GP(\eta,\theta)$ มีฟังก์ชันมวลความน่าจะเป็น ดังนี้

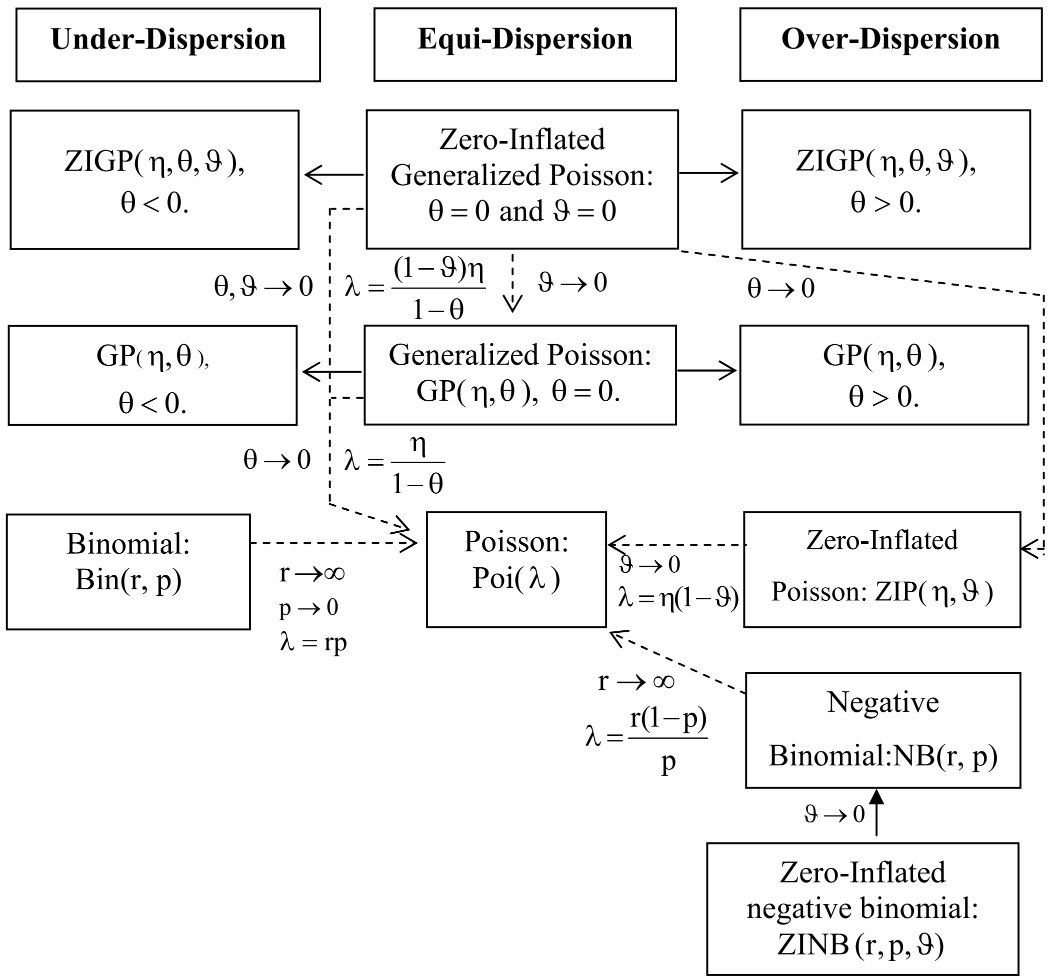
$$P(x; \eta, \theta) = \begin{cases} \dfrac{\eta(\eta+\theta x)^{x-1}e^{-(\lambda+\theta x)}}{x!}, & x = 0, 1, 2, \ldots \\ 0, & x > m \text{ when } \theta < 0 \end{cases} \qquad (5)$$

สำหรับ $\eta>0$, $\max(-1,-\eta/m)\leq\theta<1$ และ $m\geq4$ เป็นจำนวนจริงบวกที่มากที่สุดสำหรับ $\theta<0$ และที่อื่นๆ $P(x; \eta, \theta)=0$ ถ้า $\theta=0$ การแจกแจง GP จะเป็นการแจกแจงปัวซง ($\lambda=\eta$) ซึ่งการแจกแจง GP จะมีค่าเฉลี่ยของตัวแปรสุ่มเท่ากับ $\eta(1-\theta)^{-1}$ ค่า $\mu_2 = \eta(1-\theta)^{-3}$ และ $\mu_3 = \eta(1-2\theta)(1-\theta)^{-5}$ ค่าคุณลักษณะของค่าดัชนีวัดการกระจาย คือ $(1-\theta)^{-2}$ นั่นคือ เมื่อ $\theta<0$ จะเป็นกรณีที่ค่าเฉลี่ยมากกว่าค่าความแปรปรวน (อันเดอร์ดิซเพอชั่น) และ $\theta>0$ จะเป็นกรณีที่ค่าเฉลี่ยน้อยกว่าค่าความแปรปรวน (โอเวอร์ดิซเพอชั่น) สำหรับค่าสัมประสิทธิ์ความเบ้ คือ $(1+2\theta)(\eta(1-\theta))^{-1/2}$ ดังนั้นฟังก์ชันมวลความน่าจะเป็นของการแจกแจง GP จะมีขึ้นอยู่กับค่าพารามิเตอร์ η และ θ เมื่อกำหนดค่า θ ใดๆ ค่าสัมประสิทธิ์ความเบ้จะเข้าใกล้ศูนย์เมื่อ η มีค่าเพิ่มขึ้น นั่นคือจะมีลักษณะสมมาตรเมื่อ η มีค่ามากๆ ($\eta \rightarrow \infty$) ส่วนฟังก์ชันมวลความน่าจะเป็นและคุณลักษณะที่สำคัญของการแจกแจงปัวซงวางนัยทั่วไปที่มีผลกระทบจากศูนย์ (zero-inflated generalized Poisson distribution, ZIGP) จะพัฒนาคล้ายกับการแจกแจงปัวซงกรณีที่มีผลกระทบจากศูนย์

Joe และ Zhu (Joe & Zhu, 2005) พบว่าภายใต้การกำหนดสถานการณ์ที่ให้ค่าเฉลี่ยและความแปรปรวนของการแจกแจงเท่ากัน การแจกแจงปัวซงวางนัยทั่วไปจะมีปลายหางที่หนากว่า (heavy tail) การแจกแจงทวินามลบ แต่การแจกแจงปัวซงวางนัยทั่วไปจะมีฟังก์ชันมวลความน่าจะเป็นที่จุด x เท่ากับศูนย์น้อยกว่าการแจกแจงทวินามลบ ซึ่งสามารถสรุปความสัมพันธ์ระหว่างการแจกแจงปัวซงและการแจกแจงวิยุตที่กล่าวมาได้ดังรูปภาพที่ 1

การทดสอบความเหมาะสมของการแจกแจงวิยุต

สมมุติตัวแปรสุ่มวิยุต $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันมวลความน่าจะเป็น, $P(x; \theta)$, $\theta \in \Theta$ สำหรับการทดสอบความเหมาะสมของตัวอย่างสุ่มที่สุ่มมาศึกษาสอดคล้องกับการแจกแจงวิยุตแบบใดมีสมมุติฐานสำหรับการทดสอบ คือ



ภาพที่ 1 ความสัมพันธ์ระหว่างการแจกแจงปัวซงและการแจกแจงวิยุตที่เกี่ยวข้อง

H_0 : ตัวอย่างสุ่มมีการแจกแจงวิยุตที่ต้องการทดสอบ

เทียบกับ H_1 : ตัวอย่างสุ่มไม่มีการแจกแจงวิยุตที่ต้องการทดสอบ สำหรับสถิติที่ทดสอบความเหมาะสม (goodness of fit test) ของตัวอย่างสุ่มสามารถประยุกต์ใช้การทดสอบด้วยไคกำลังสอง (chi-square test) การทดสอบโคลโมโกรอฟ-สเมอร์นอฟ (Kolmogorov-Smirnov test) และการทดสอบคราเมอร์-วอนมิเซส (Cramer-von Mises test) เป็นต้น โดยการทดสอบการแจกแจงปัวซงที่พัฒนาจากสถิติทดสอบด้วยไคกำลังสองที่อาศัยอัตราส่วนระหว่างค่าความแปรปรวนและค่าเฉลี่ยจะมีประสิทธิภาพของการทดสอบเมื่อค่าดัชนีวัดการกระจาย (ID) มีค่าห่างจาก 1 ไปมากๆ ทั้งกรณีโอเวอร์และอันเดอร์ดิซเพอชั่น แต่เมื่อค่าดัชนีวัดการกระจายเข้าใกล้ 1 จะไม่สามารถแยกความแตกต่างระหว่างการแจกแจงปัวซงและการแจกแจงอื่นๆ ที่เกี่ยวข้องได้ดังตารางที่ 1 (Karlis & Xekalaki, 2000; Gurtler & Henze, 2000; Meintanis & Nikitin, 2008) โดยส่วนใหญ่ในการเก็บรวบรวมข้อมูลจำนวนนับในขอบเขตที่สนใจศึกษาจะมีปัญหาที่ค่าความแปรปรวนของตัวแปรสุ่มวิยุตมากกว่าค่าเฉลี่ย ดังนั้นงานวิจัยส่วนใหญ่จึงมุ่งไปที่การทดสอบและการประมาณค่าพารามิเตอร์ของตัวแบบในกรณีที่เกิดปัญหาโอเวอร์ดิซเพอชั่น ซึ่งจะแบ่งการทดสอบออกเป็น 2 ส่วนใหญ่ๆ (Yang *et al*., 2010; Garay *et al*., 2011) ดังนี้

1) การทดสอบความเหมาะสมระหว่างการแจกแจงปัวซงกับการแจกแจงทวินามลบและการแจกแจงปัวซงวางนัยทั่วไป โดยใช้การทดสอบสคอร์ (score test) ในการทดสอบค่าพารามิเตอร์ดิซเพอชั่น (φ) ของแต่ละการแจกแจง เช่น การแจกแจงทวินาม $\varphi = r^{-1}$ สมมุติฐานในการทดสอบคือ $H_0 : \varphi = 0$ เทียบกับ $H_1 : \varphi > 0$ และการแจกแจงปัวซงวางนัยทั่วไป $\varphi = (1-\theta)^{-2}$ มีสมมุติฐานในการทดสอบคือ $H_0 : \varphi = 1$ เทียบกับ $H_1 : \varphi > 1$

2) การทดสอบความเหมาะสมระหว่างการแจกแจงปัวซงกับการแจกแจงที่มีผลกระทบจากศูนย์ของการแจกแจงทวินามลบ การแจกแจงปัวซงและการแจกแจงปัวซงวางนัยทั่วไปโดยใช้การ

ตารางที่ 1 ค่าคุณลักษณะที่สำคัญของการแจกแจงปัวซงและการแจกแจงวิยุตที่เกี่ยวข้อง

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

ทดสอบสคอร์ (score test) ในการทดสอบค่าพารามิเตอร์ผลกระทบจากศูนย์ (ϑ) ของแต่ละการแจกแจง ซึ่งมีสมมุติฐานในการทดสอบคือ $H_0 : \vartheta = 0$ เทียบกับ $H_1 : \vartheta > 0$

เมื่อค่าพารามิเตอร์ดิซเพอชั่น (φ) มีค่าเข้าใกล้ 1 ค่าพารามิเตอร์ผลกระทบจากศูนย์ (ϑ) เข้าใกล้ 0 และขนาดตัวอย่างสุ่ม n ที่มีค่าไม่ใหญ่พอ การทดสอบสคอร์จะมีประสิทธิภาพต่ำในการจำแนกระหว่างการแจกแจงปัวซงและการแจกแจงวิยุตที่ต้องการทดสอบ จึงได้มีการพิจารณาคุณลักษณะของค่าสัมประสิทธิ์ความเบ้ของการแจกแจงวิยุตมาช่วยในการคัดเลือกการแจกแจงที่เหมาะสมกับข้อมูลจำนวนนับที่สนใจศึกษา (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 ในกรณีที่การแจกแจงเป็นเครือข่ายกัน (nested) เช่น การทดสอบความเหมาะสมของการแจกแจง ZIGP GP และ Poi แต่ถ้าการแจกแจงไม่เป็นเครือข่ายกันจะประยุกต์ใช้การทดสอบ Vuong (Vuong, 1989) เช่น การทดสอบความเหมาะสมระหว่างการแจกแจง GP และ NB เนื่องจากสถิติทดสอบ Vuong ได้พัฒนาการทดสอบจากหลักการทดสอบอัตราส่วนควรจะเป็น (likelihood ratio test) ภายใต้เงื่อนไขทั่วไปที่สามารถทดสอบได้ทั้งแบบจำลองที่เป็นเครือข่ายและไม่เป็นเครือข่ายกัน ซึ่งจะให้กำลังการทดสอบ (power of test) ที่สูงในการจำแนกระหว่างแบบจำลอง

ข้อเสนอแนะ

ตัวแปรสุ่มวิยุตที่เกิดขึ้นในขอบเขตที่สนใจศึกษาที่เป็นข้อมูลจำนวนนับมีการประยุกต์ใช้ในหลายๆ สาขาวิชา เช่น วิทยาศาสตร์สิ่งแวดล้อม วิศวกรรมศาสตร์ เศรษฐศาสตร์ ระบาดวิทยา เป็นต้น ซึ่งนักวิจัยมีข้อสมมุติให้ตัวแปรสุ่มวิยุตที่สนใจศึกษาภายใต้ขอบเขตที่สนใจศึกษา มีการแจกแจงปัวซงและบ่อยครั้งที่ตัวแปรสุ่มวิยุตที่สนใจศึกษานั้นไม่เป็นไปตามข้อสมมุติของการแจกแจงปัวซงโดยจะเกิดปัญหาค่าเฉลี่ยของตัวแปรสุ่มมากกว่าค่าความแปรปรวน (อันเดอร์ดิซเพอชั่น) นักวิจัยจะหลีกเลี่ยงไปใช้การแจกแจงทวินามหรือการแจกแจกปัวซงวางนัยทั่วไป แต่สำหรับกรณีที่ค่าเฉลี่ยของตัวแปรสุ่มน้อยกว่าค่าความแปรปรวน (โอเวอร์ดิซเพอชั่น) นักวิจัยจะหลีกเลี่ยงไปใช้การแจกแจงทวินามลบ การแจกแจงปัวซงวางนัยทั่วไป ซึ่งในบางกรณีที่ตัวแปรสุ่มวิยุตที่สนใจภายใต้ขอบเขตที่สนใจศึกษาเกิดขึ้นได้น้อยจะส่งผลให้มีค่าที่เป็นศูนย์เป็นจำนวนมากซึ่งเป็นอีกสาเหตุหนึ่งที่ทำให้ค่าความแปรปรวนมากกว่าค่าเฉลี่ยนักวิจัยจะประยุกต์ใช้การแจกแจงวิยุตที่มีผลกระทบจากศูนย์ของการแจกแจงปัวซง การแจกแจงทวินามลบ และการแจกแจงปัวซงวางนัยทั่วไป โดยมีวิธีการในการทดสอบความเหมาะสมของการแจกแจงด้วยการทดสอบไคกำลังสองและการทดสอบโคลโมโกรอฟ-สเมอร์นอฟ ซึ่งถ้าตัวอย่างสุ่มชุดหนึ่งมีการแจกแจงที่เหมาะสมมากกว่าหนึ่งการแจกแจงนักวิจัยควรพิจารณาเลือกการแจกแจงด้วยคุณลักษณะอื่นๆ ของการแจกแจง ตัวอย่างเช่น หากการแจกแจงที่เป็นเครือข่ายกันจะเลือกการแจกแจงที่ให้ค่า AIC ต่ำที่สุด เช่น การทดสอบความเหมาะสมระหว่างการแจกแจงทวินามลบที่มีผลกระทบจากศูนย์และการแจกแจงทวินามลบ หรือการทดสอบความเหมาะสมของการแจกแจงปัวซงวางนัยทั่วไปที่มีกระทบจากศูนย์ การแจกแจงปัวซงวางนัยทั่วไป การแจกแจงปัวซงที่มีผลกระทบจากศูนย์ และการแจกแจงปัวซง แต่สำหรับการแจกแจงที่ไม่เป็นเครือข่ายกัน การคัดเลือกการแจกแจงที่เหมาะสมควรเลือกการทดสอบ Vuong เช่น การเลือกความเหมาะสมระหว่างการแจกแจงปัวซงวางนัยทั่วไปกับการแจกแจงทวินามลบ

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*. (3rd 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[rn] 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.

Nikoloulopulos, 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.