บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษา “การพัฒนาการแจกแจงความน่าจะเป็นของตัวแปรสุ่มทวินามลบ-ลินเลย์และการประยุกต์”ได้ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องมาเป็นแนวทางในการวิจัยซึ่งสามารถนำมาเสนอตามลำดับ ดังต่อไปนี้

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง
การทดสอบภาวะสารูปสนิทดี
แนวคิดเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์
งานวิจัยที่เกี่ยวข้อง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มคือกฎหรือวิธีการที่กำหนดความน่าจะเป็นให้แก่ค่าตัวแปรสุ่ม อาจจะอยู่ในรูปของตาราง กราฟ หรือสูตร มี 2 ชนิด คือ การ แจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องและการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่องถ้าอยู่ในรูปของสูตร f(x) ที่ปรากฏในสูตร จะเรียกว่าฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม มี 2 ชนิด คือ การ แจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องและการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง ในที่นี้จะกล่าวถึงการแจกแจงความน่าจะเป็นไม่ต่อเนื่องซึ่งเกี่ยวข้องกับการศึกษาครั้งนี้เท่านั้น

บทนิยาม 2.1 ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องโดยมี R_xเป็นเรนจ์ และ $x \in R_x$จะเรียก $f(x) = P(X = x)$ ว่าฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องก็ต่อเมื่อ

1. $f(x) \geq 0$ สำหรับทุก ๆ ค่าของ x
2. $\sum_{\text{allx}} f(x) = 1$

การแจกแจงปัวซง

การแจกแจงปัวซงเป็นการแจกแจงไม่ต่อเนื่องที่เกิดจากการทดลองทีทำให้ได้ค่าของตัวแปรสุ่ม X ซึ่ง X เป็นจำนวนครั้งของการเกิดเหตุการณ์ที่สนใจในช่วงเวลาหนึ่ง หรืออาณาบริเวณหนึ่งที่กำหนดให้ เรียกการทดลองลักษณะนี้ว่าการทดลองปัวซง (poisson experiment) เรียกตัวแปรสุ่มนี้ว่าตัวแปรสุ่มปัวซง และเรียกการแจกแจงความน่าจะเป็นของตัวแปรสุ่มนี้ว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่มปัวซง ประกอบด้วยพารามิเตอร์หนึ่งตัวคือ λ และสามารถเขียนฟังก์ชันความน่าจะเป็นได้ดังนี้ (จิรัชย์ สุขะเกตุ, 2548).

$$f(x)=\frac{e^{(-\lambda)}\lambda^{x}}{x!}, \quad x=0,\ 1,\ 2,\ldots \text{for } \lambda > 0$$

ทฤษฎีบท 2.1 ถ้า X มีการแจกแจงปัวซงที่มี λ เป็นพารามิเตอร์แล้วจะได้

1. ค่าคาดหวังของ X คือ $E(X) = \lambda$
2. ค่าความแปรปรวนของ X คือ $V(X) = \lambda$
3. ฟังก์ชันก่อกำเนิดโมเมนต์ของ X คือ $M_x(t) = e^{\lambda(e^t - 1)}, t \in IR.$

การแจกแจงทวินามลบ

การแจกแจงทวินามลบเป็นการแจกแจงแบบไม่ต่อเนื่องที่มีฟังก์ชันความน่าจะเป็นที่เกิดจากการผสมระหว่างการแจกแจงแบบปัวซงกับการแจกแจงแกมม่า เกิดขึ้นเมื่อมีการทดลองที่ซ้ำ ๆ กัน โดยในการทดลองแต่ละครั้งมีสิ่งที่เกิดขึ้นเพียงสองอย่าง คือ ความสำเร็จกับความล้มเหลวด้วยความน่าจะเป็นp และ 1-p ตามลำดับ ถ้าให้ X เป็นจำนวนครั้งของความล้มเหลวก่อนที่จะได้ความสำเร็จครบ r ครั้ง จะได้ว่า X เป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นแบบทวินามลบ ประกอบด้วยพารามิเตอร์สองตัวคือ p และ r สามารถเขียนฟังก์ชันความน่าจะเป็น ได้ดังนี้ (จิรัชย์ สุขะเกตุ, 2548)

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, x = 0,1,2,\ldots \text{ for } r > 0 \text{ and } 0 < p < 1$$

ทฤษฎีบท 2.2ให้ X มีการแจกแจงทวินามลบ ที่มี r และ p เป็นพารามิเตอร์ โดย r > 0 และ 0 < p < 1 จะได้

1. ค่าคาดหวังของ X คือ $E(X) = \dfrac{r(1-p)}{p}$
2. ค่าความแปรปรวนของ X คือ $Var(X) = \dfrac{r(1-p)}{p^2}$
3. ฟังก์ชันก่อกำเนิดโมเมนต์ของ X คือ $M_x(t) = p^r(1-\exp(1-p))^{-r}$ เมื่อt เป็นจำนวนอตรรกยะ

การแจกแจงลินเลย์

การแจกแจงลินเลย์เป็นการแจกแจงแบบต่อเนื่องที่เกิดจากการผสมระหว่างการแจกแจงเอกโปเนนเชียลกับการแจกแจงแกมมา ประกอบด้วยพารามิเตอร์หนึ่งตัวคือ θ ซึ่งสามารถเขียนฟังก์ชันความน่าจะเป็น ดังนี้ (Zakerzadeh, 2009)

$$f(x) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta} \quad ; \quad x > 0, \theta > 0$$

ทฤษฎีบท 2.3ให้ X มีการแจกแจงลินเลย์ ที่มี θ เป็นพารามิเตอร์ โดย $\theta > 0$ จะได้

1. ค่าคาดหวังของ X คือ $E(X) = \dfrac{\theta+2}{\theta(\theta+1)}$
2. ค่าความแปรปรวนของ X คือ $Var(X) = \dfrac{2(\theta+3)}{\theta^2(\theta+1)} - \left(\dfrac{\theta+2}{\theta(\theta+1)}\right)^2$

3. ฟังก์ชันก่อกำเนิดโมเมนต์ของ X คือ $M_x(t) = \dfrac{\theta^2(\theta - t + 1)}{(\theta + 1)(\theta - t)^2}$

การแจกแจงลินเลย์วางนัยทั่วไป

การแจกแจงลินเลย์วางนัยทั่วไปเป็นการแจกแจงไม่ต่อเนื่องที่พัฒนาขึ้นจากการแจกแจงลินเลย์ ประกอบด้วยสามพารามิเตอร์คือ α θ และ γ ซึ่งสามารถเขียนฟังก์ชันความน่าจะเป็น ดังนี้ (Zakerzadeh, 2009)

$$f(x) = \frac{\theta^2(\theta x)^{\alpha-1}(\alpha + \gamma x)e^{-\theta x}}{(\gamma + \theta)\Gamma(\alpha + 1)} \quad ; \quad \alpha, \theta, \gamma, x > 0$$

ทฤษฎีบท 2.4 ให้ X มีการแจกแจงลินเลย์ทั่วไป ที่มี α θ และ γ เป็นพารามิเตอร์ โดย $\theta >$ 0 จะได้

1. ค่าคาดหวังของ X คือ $E(X) = \dfrac{\theta(\alpha\theta) + \gamma(\alpha + 1)\theta}{\theta + \gamma}$.

2. ค่าความแปรปรวนของ X คือ $\mathrm{Var} = \dfrac{\theta\left[\alpha\theta^2 + (\alpha\theta)^2\right] + \gamma\left[(\alpha + 1)\theta^2 + ((\alpha + 1)\theta)^2\right]}{\theta + \gamma}$

$$- \frac{\theta(\alpha\theta) + \gamma(\alpha + 1)\theta}{\theta + \gamma}$$

3. ฟังก์ชันก่อกำเนิดโมเมนต์ของ X คือ $M_x(t) = \left(\dfrac{\theta}{\theta - t}\right)^{\alpha+1} \dfrac{\theta - t + \gamma}{\theta + \gamma}$

การแจกแจงทวินามลบ-ลินเลย์

การแจกแจงทวินามลบ-ลินเลย์เป็นการแจกแจงไม่ต่อเนื่องที่เกิดจากการผสมระหว่างการแจกแจงทวินามและการแจกแจงลินเลย์ ประกอบด้วย พารามิเตอร์สองตัวคือ θ และ r ซึ่งสามารถเขียนฟังก์ชันความน่าจะเป็น ดังนี้ (Zamani, 2010)

$$f(x) = \frac{\theta^2}{\theta + 1}\binom{r + x - 1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j \frac{\theta + r + j - 1}{(\theta + r + j)^2}; x = 0,1,2,...$$

ทฤษฎีบท 2.5 ให้ X มีการแจกแจงทวินามลบ-ลินเลย์ ที่มี θ และ r เป็นพารามิเตอร์ จะได้

1. ค่าคาดหวังของ X คือ $E(x) = r\left[\dfrac{\theta^3}{(\theta + 1)(\theta - 1)^2} - 1\right]$

2. ค่าความแปรปรวนของ X คือ

$$\mathrm{Var} = (r+r^2)\frac{\theta^2(\theta-1)}{(\theta+1)(\theta-2)^2} - (r+2r^2)\frac{\theta^3}{(\theta+1)(\theta-1)^2} + r^2 - r^2\left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1\right]^2$$

การทดสอบภาวะสารูปสนิทดี(Test for Goodness of Fit)

การทดสอบภาวะสารูปสนิทดี เป็นกระบวนการทางสถิติที่ใช้เพื่อทำการทดสอบการแจกแจงของประชากรว่ามีการแจกแจงตามที่คาดหวังหรือไม่(อารียา สุดสุข, 2555)

โดยวิธีการทดสอบภาวะสารูปสนิทดีสำหรับการแจกแจง มีขั้นตอนดังนี้กำหนดสมมติฐาน

H_0 :ข้อมูลมีการแจกแจงที่คาดหวังไว้

H_1: ข้อมูลไม่มีการแจกแจงที่คาดหวังไว้

1. กำหนดระดับนัยสำคัญทางสถิติ(α)
2. เลือกสถิติทดสอบและคำนวณค่าสถิติทดสอบ ซึ่งการทดสอบภาวะสารูปสนิทดีในงานวิจัยครั้งนี้เลือกใช้ตัวสถิติทดสอบไคสแควร์
3. หาค่าวิกฤตของสถิติทดสอบที่ระดับนัยสำคัญที่กำหนดไว้
4. ตัดสินใจว่าจะปฏิเสธ H_0 หรือไม่ จะปฏิเสธสมมติฐานหลัก H_0 เมื่อค่าสถิติทดสอบไคสแควร์ที่คำนวณได้มีค่ามากกว่าค่าวิกฤตในตาราง
5. สรุปผลการทดสอบสมมติฐาน

แนวคิดเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์

วิธีภาวะน่าจะเป็นสูงสุด

นิยาม 2.1 ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากรที่มีฟังก์ชันความหนาแน่นน่าจะเป็น $f(x;\theta),\ \theta \in \Omega$ ฟังก์ชันภาวะน่าจะเป็น (likelihood function)ของตัวอย่างสุ่ม คือ ฟังก์ชันความหนาแน่นน่าจะเป็นร่วม $L = L(\theta; x_1, x_2, ..., x_n) = L(\theta)$ ของตัวอย่างสุ่มที่เป็นฟังก์ชันของพารามิเตอร์ θ นั่นคือ $L(\theta; x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i;\theta)$ (ประชุม,2545)

วิธีหาตัวประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด ดังต่อไปนี้

1) หาฟังก์ชันความน่าจะเป็น

$$L(\theta_1,\theta_2,\ldots,\theta_k) = L(\theta_1,\theta_2,\ldots,\theta_k; x_1,x_2,\ldots,x_n)$$
$$= f(x_1,x_2,\ldots,x_n;\theta_1,\theta_2,\ldots,\theta_k)$$
$$= f(x_1;\theta_1,\theta_2,\ldots,\theta_k).f(x_2;\theta_1,\theta_2,\ldots,\theta_k)\ldots f(x_n;\theta_1,\theta_2,\ldots,\theta_k)$$

$$= \prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,\ldots,\theta_k)$$

2) กำหนดให้ $\frac{d}{d\theta_1} L(\theta_1,\theta_2,...,\theta_k) = 0$

$$\frac{d}{d\theta_2} L(\theta_1,\theta_2,...,\theta_k) = 0$$

$$\vdots$$

และ $\frac{d}{d\theta_k} L(\theta_1,\theta_2,...,\theta_k) = 0$

3) แก้สมการเพื่อหาค่าประมาณพารามิเตอร์ $\theta_1,\theta_2,...,\theta_k$ จะได้ $\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_k$ เป็นตัวประมาณภาวะน่าจะเป็นสูงสุดของ $\theta_1,\theta_2,...,\theta_k$ ตามลำดับ

วิธีโมเมนต์

วิธีการหาตัวประมาณค่าของพารามิเตอร์โดยวิธีโมเมนต์เป็นวิธีเก่าแก่ที่สุด ได้ถูกคิดค้นโดย คาร์ล เพียร์สัน (Karl Pearson) เมื่อประมาณ ค.ศ.1894 โดยใช้โมเมนต์ของตัวอย่าง (Sample Moments) เป็นตัวประมาณของโมเมนต์ของประชากร (Population Moments) ที่สมนัยกัน ดังนี้

$E(X^k) = \mu_k^{'} = g_k(\theta_1,\theta_2,\ldots,\theta_k)$ และ $M_k^{'} = \dfrac{\sum_{i=1}^{n} X_i^k}{n}$

ดังนั้นจึงกำหนดให้ $\mu_k^{'} = M_k^{'}$

วิธีหาตัวประมาณด้วยวิธีโมเมนต์ ดังต่อไปนี้

1) กำหนดให้ $E(X) = g_1(\theta_1,\theta_2,\ldots,\theta_k) = \dfrac{\sum_{i=1}^{n} X_i}{n}$

$$E(X^2) = g_2(\theta_1,\theta_2,\ldots,\theta_k) = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

$$\vdots$$

และ $E(X^k) = g_k(\theta_1,\theta_2,\ldots,\theta_k) = \dfrac{\sum_{i=1}^{n} X_i^k}{n}$

2) แก้สมการเพื่อหาค่าประมาณพารามิเตอร์ $\theta_1,\theta_2,...,\theta_k$ จะได้ $\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_k$ เป็นตัวประมาณโดยวิธีโมเมนต์ของ $\theta_1,\theta_2,...,\theta_k$ ตามลำดับ

งานวิจัยที่เกี่ยวข้อง

Vellaisamy and Sankar (2005) ได้ศึกษาถึงการแจกแจงทวินามลบ (negative binomial distribution) และสรุปว่าเป็นการแจกแจงแบบไม่ต่อเนื่องที่มีฟังก์ชันความน่าจะเป็นที่เกิดจากการผสมระหว่างการแจกแจงแบบปัวซงกับการแจกแจงแกมม่า เกิดขึ้นเมื่อมีการทดลองที่ซ้ำ ๆ กัน โดยในการทดลองแต่ละครั้งมีสิ่งที่เกิดขึ้นเพียงสองอย่าง คือ ความสำเร็จกับความล้มเหลวด้วยความน่าจะเป็นp และ 1-p ตามลำดับ ถ้าให้ X เป็นจำนวนครั้งของความล้มเหลวก่อนที่จะได้ความสำเร็จครบ r ครั้ง จะได้ว่า X เป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นแบบทวินามลบ ประกอบด้วยพารามิเตอร์สองตัวคือ p และ r

Ghitany M.E., B. Atieh and S. Nadarajah(2007) ได้ศึกษาเกี่ยวกับการแจกแจงลินเลย์และการประยุกต์ พบว่า การแจกแจการแจกแจงลินเลย์ (Lindley Distribution) เป็นการแจกแจงแบบต่อเนื่องที่เกิดจากการผสมระหว่างการแจกแจงเอกโปเนนเชียลกับการแจกแจงแกมมา ประกอบด้วยพารามิเตอร์หนึ่งตัว

Gómez et al.(2008)ได้ศึกษาสร้างตัวแปรสุ่มทวินามลบ -อินเวอร์สเกาส์เซียน (Negative Binomail-inverse Gaussian Distribution)แบบเดียวและแบบหลายตัวแปรสุ่มซึ่งเป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง โดยผสมระหว่างการแจกแจงทวินามลบและการแจกแจงอินเวอร์สเกาส์เซียนและได้ศึกษาคุณสมบัติทางสถิติเชิงคณิตศาสตร์เบื้องต้น ได้แก่ ค่าเฉลี่ย ความแปรปรวน และความเบ้ เป็นต้น พร้อมทั้งได้แสดงการประมาณค่าพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุด

Zakerzadeh H. and A. Dolati (2009) ได้สร้างการแจกแจงลินเลย์วางนัยทั่วไป (Generalized Lindley Distribution) ขึ้นมาใหม่ ซึ่งเป็นการแจกแจงไม่ต่อเนื่องที่พัฒนาขึ้นจากการแจกแจงลินเลย์ ประกอบด้วยสามพารามิเตอร์ ซึ่งการสร้างตัวแปรสุ่มแบบถ่วงน้ำหนักระหว่างการแจกแจงเอกซ์โพแนนเชียล กับการแจกแจงแกมมา และได้ศึกษาคุณสมบัติทางสถิติเชิงคณิตศาสตร์เบื้องต้น ได้แก่ ค่าเฉลี่ย ความแปรปรวน ฟังก์ชันก่อกำเนิดโมเมนต์ (Moment Generating Function) ฟังก์ชันการอยู่รอด (Survival Function) และฟังก์ชันความเสี่ยง (Hazard Function) เป็นต้น และประมาณค่าพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุด

Zamani H. and N. Ismail (2010) ได้ศึกษาการแจกแจงทวินามลบ -ลินเลย์(Negative Binomail-Lindley Distribution) เป็นการแจกแจงไม่ต่อเนื่องที่เกิดจากการผสมระหว่างการแจกแจงทวินามและการแจกแจงลินเลย์ ประกอบด้วย พารามิเตอร์สองตัวคือ θ และ r

Pudprommarat C. et al. (2012) ได้ศึกษาการแจกแจงทวินามลบ-บีต้าเอกซ์โพแนนเชียล (Negative Binomail-Beta Exponential Distribution) เป็นการแจกแจงไม่ต่อเนื่องที่เกิดจากการผสมระหว่างการแจกแจงทวินามลบและการแจกแจงบีต้าเอกซ์โพแนนเชียล ประกอบด้วยพารามิเตอร์สี่ตัว โดยพบว่า การแจกแจงทวินามลบและการแจกแจงบีต้าเอกซ์โพแนนเชียลเป็นการแจกแจงทั่วไปของการแจกแจงวอริงวางนัยทั่วไป (Generalized Waring distribution) การแจกแจงวอริง (Waring distribution) และการแจกแจงยู (Yule distribution) ตามลำดับ

จากงานวิจัยที่กล่าวมาข้างต้น พบว่าได้นำการแจกแจงทวินามลบมาผสมกับการแจกแจงอื่น ๆ เช่น การแจกแจงอินเวอร์สเกาส์เซียน การแจกแจงบีต้าเอกซ์โพแนนเชียล การแจกแจงลินเลย์ เป็นต้น แต่ยังไม่ได้มีการแจกแจงทวินามลบนำมาผสมกับการแจกแจงลินเลย์วางนัยทั่วไป จึงทำให้ผู้วิจัยสนใจที่จะศึกษาการแจกแจงทวินามลบ-ลินเลย์วางนัยทั่วไป พร้อมทั้งศึกษาคุณสมบัติทางสถิติเชิงคณิตศาสตร์ต่าง ๆ เช่น ฟังก์ชันความน่าจะเป็น ฟังก์ชันการแจกแจง ค่าเฉลี่ย และ ความแปรปรวน เป็นต้น พร้อมทั้งการประมาณค่าพารามิเตอร์โดยวิธีภาวะน่าจะเป็นสูงสุดและวิธีโมเมนต์