บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การแจกแจงทวินามนิเสธ (negative binomial distribution) เป็นการแจกแจงวิยุต (discrete distribution) แบบหนึ่งที่มีการศึกษามาอย่างยาวนานในทฤษฎีความน่าจะเป็น และสถิติ เช่นเดียวกับการแจกแจงทวินาม (binomial distribution) โดยทั่วไปรูปแบบของการแจกแจงทวินามนิเสธมีสองรูปแบบ ซึ่งทั้งสองรูปแบบสามารถนำไปประยุกต์ใช้ในเรื่องที่แตกต่างกัน ขึ้นอยู่กับว่าผู้ที่จะนำไปใช้จะเลือกใช้รูปแบบที่สอดคล้องกับความต้องการอย่างไร อย่างไรก็ตามรูปแบบหนึ่งของการแจกแจงที่มีการศึกษาและวิจัยกันมาก คือ รูปแบบที่คล้ายกับการแจกแจงทวินาม ให้ Y เป็นตัวแปรสุ่มทวินามนิเสธที่มีพารามิเตอร์ $r > 0$ และ $0 < p < 1$ ซึ่งมีฟังก์ชันความน่าจะเป็นดังนี้

$$p_Y(k) = \frac{\Gamma(r+k)}{k!\Gamma(r)} p^r q^k \qquad , k = 0,1,2,\ldots \tag{1.1}$$

โดยที่ Γ คือ ฟังก์ชันแกมมา (gamma function) ค่าคาดหมายและความแปรปรวนของตัวแปรสุ่ม Y คือ $E(Y) = \frac{rq}{p}$ และ $Var(Y) = \frac{rq}{p^2}$ ตามลำดับ ในกรณีที่ r เป็นจำนวนเต็มบวกเราอาจพิจารณาการแจกแจงนี้ในรูปของการแจกแจงของจำนวนความล้มเหลว (failure) ก่อนเกิดความสำเร็จ (success) ครั้งที่ r ในลำดับของการทดลองย่อยแบร์นูลลี (bernouli trial) ที่เป็นอิสระต่อกัน โดยที่ความสำเร็จและความล้มเหลวที่เกิดขึ้นในแต่ละการทดลองย่อยมีความน่าจะเป็นเท่ากับ p และ $q = 1 - p$ ตามลำดับ และในกรณีที่ $r = 1$ เราจะเรียกการแจกแจงทวินามนิเสธว่า การแจกแจงเรขาคณิต (geometric distribution) ที่มีพารามิเตอร์ $0 < p < 1$

การประยุกต์ของการแจกแจงทวินามนิเสธ เราพบว่าการแจกแจงนี้สามารถนำไปประยุกต์ใช้ในหลาย ๆ ด้าน ได้แก่ การวิเคราะห์ข่ายงานโทรคมนาคม การวิเคราะห์สินค้าคงคลัง และวิเคราะห์ทางด้านพันธุศาสตร์ของประชากร เป็นต้น นอกจากการประยุกต์ดังกล่าว เรายังพบว่ามีการใช้การแจกแจงทวินามนิเสธที่มีฟังก์ชันความน่าจะเป็นดังสมการ (1.1) ในการประมาณการแจกแจงของตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบ (non-negative integer-valued random variable) บางการแจกแจงได้เช่นเดียวกับการประมาณด้วยการแจกแจงที่สำคัญอื่น ๆ ตัวอย่างเช่น Brown and Phillips (1999) ได้ใช้การแจกแจงทวินามนิเสธไปประมาณการแจกแจงของผลรวมของตัวแปรสุ่มแบร์นูลลี และนำผลลัพธ์ที่ได้ไปประยุกต์ใช้ในการประมาณการแจกแจงโพลยา (Pólya distribution)

ต่อมา Vellaisamy and Upadhye (2009) ได้ใช้การแจกแจงทวินามนิเสธประมาณการแจกแจงของผลรวมของตัวแปรสุ่มเรขาคณิตที่เป็นอิสระต่อกัน และในกรณีที่ X เป็นตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบที่มีฟังก์ชันความน่าจะเป็น $p_X(x)>0$ สำหรับทุก $x \in S(x)$ เมื่อ $S(x)$ คือ เซตของค่าของตัวแปรสุ่ม X หรือเรียกว่าเซตค้ำจุน (support) ของ X และให้ μ และ $\sigma^2 (0<\sigma^2<\infty)$ เป็นค่าคาดหมายและความแปรปรวนของ X ตามลำดับ ในกรณีนี้ Teerapabolarn and Boondirek (2010) ได้ประมาณการแจกแจงของตัวแปรสุ่ม X ด้วยการแจกแจงทวินามนิเสธในรูปต่อไปนี้

$$\left|P(X \in A) - P(Y \in A)\right| \le \frac{p(1-p^r)}{rq} \mathrm{E}\left|\frac{(r+X)q}{p} - z(X)\right| + (1-p_X(0))\left|\frac{rq}{p} - \mu\right| \tag{1.2}$$

โดยที่ $A \subseteq \mathbb{N} \cup \{0\}$, $P(X \in A) = \sum_{k \in A} p_X(k)$, $P(Y \in A) = \sum_{k \in A} p_Y(k)$ และ

$$z(x) = \sigma^2 w(x) = \frac{\sum_{k=0}^{x} (\mu - k) p_X(k)}{p_X(x)}$$

และในกรณีที่ $r=1$ Teerapabolarn (2011) ได้ประมาณการแจกแจงของตัวแปรสุ่ม X ด้วยการแจกแจงเรขาคณิตดังผลลัพธ์ต่อไปนี้

$$\left|P(X \in A) - P(Y \in A)\right| \le \sum_{k \in S(x) \setminus \{0\}} \frac{\left|(k+1)q - p\sigma^2 w(k)\right| p_X(k)}{k} + \left|\frac{q}{p} - \mu\right| \left[1 - p_X(0)q\right] \tag{1.3}$$

และสำหรับ $A = \{0, \dots, x_0\}$ โดยที่ $x_0 \in S(x)$ ผลลัพธ์ที่ได้อยู่ในรูปของผลต่างระหว่างฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่ม X และตัวแปรสุ่มเรขาคณิตพร้อมด้วยขอบเขตบนทั้งแบบเอกรูปและไม่เอกรูปดังอสมการต่อไปนี้

$$\begin{aligned} \left|F(0) - G(0)\right| \le & \sum_{k \in S(x) \setminus \{0\}} \left|\frac{q}{k} - \frac{p\sigma^2 w(k)}{k(k+1)}\right| p_X(k) \\ & + \left|q - p\mu\right| \left\{ p_X(0) + \sum_{k \in S(x) \setminus \{0\}} \frac{p_X(k)}{k} \right\} \end{aligned} \tag{1.4}$$

และเมื่อ $x_0 > 0$ จะได้

$$\left|F(x_0) - G(x_0)\right| \le \sum_{k \in S(x)} \left|q - \frac{p\sigma^2 w(k)}{k+1}\right| p_X(k) + \left|\frac{q}{p} - \mu\right| \sum_{k \in S(x) \setminus \{0\}} \frac{p_X(k)}{k} \tag{1.5}$$

และในกรณีที่ $0 < q < 1/2$ จะได้ผลลัพธ์ที่มีขอบเขตบนแบบไม่เอกรูปเป็นดังนี้

$$\left|F(x_0)-G(x_0)\right| \le \frac{1}{x_0+1}\left\{\sum_{k\in S(x)}\left|(k+1)q-p\sigma^2 w(k)\right| p_X(k)+\left|\frac{q}{p}-\mu\right|\left[1-p_X(0)\right]\right\} \tag{1.6}$$

โดยที่ $F(x_0)=\sum_{k=0}^{x_0} p_X(k)$ และ $G(x_0)=\sum_{k=0}^{x_0} pq^k$ คือ ฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่ม X และตัวแปรสุ่มเรขาคณิตตามลำดับ และ $w(x)=z(x)/\sigma^2$

พิจารณาในกรณีที่ $A=\{x_0\}$ เมื่อ $x_0 \in S(x)$ และให้ $\mathcal{NB}(x_0)=\frac{\Gamma(r+x_0)}{x_0!\Gamma(r)}p^r q^{x_0}$ และ $\mathcal{G}(x_0)=pq^{x_0}$ เป็นฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มทวินามนิเสธ และตัวแปรสุ่มเรขาคณิตที่ x_0 ตามลำดับ แล้วผลลัพธ์ในอสมการ (1.2) และ (1.3) ที่อยู่ในรูปของเมตริกแบบจุด (point metric) หรือผลต่างของฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X และตัวแปรสุ่มทวินามนิเสธ พร้อมด้วยขอบเขตบนแบบเอกรูปเป็นดังนี้

$$\left|p_X(x_0)-\mathcal{NB}(x_0)\right| \le \frac{p(1-p^r)}{rq}\mathrm{E}\left|\frac{(r+X)q}{p}-\sigma^2 w(X)\right|+(1-p_X(0))\left|\frac{rq}{p}-\mu\right| \tag{1.7}$$

และเมตริกแบบจุดระหว่างการแจกแจงของตัวแปรสุ่ม X และตัวแปรสุ่มเรขาคณิต พร้อมด้วยขอบเขตบนแบบเอกรูปเป็นดังนี้

$$\left|p_X(x_0)-\mathcal{G}(x_0)\right| \le \sum_{k\in S(x)\setminus\{0\}} \frac{\left|(k+1)q-p\sigma^2 w(k)\right| p_X(k)}{k}+\left|\frac{q}{p}-\mu\right|\left[1-p_X(0)q\right] \tag{1.8}$$

เราจะสังเกตได้ว่าขอบเขตบนทางด้านขวามือของอสมการ (1.7) และ (1.8) ไม่มีการเปลี่ยนแปลงสำหรับทุกค่าของ $x_0 \in S(x)$ (เรียกว่าขอบเขตบนแบบเอกรูป (uniform upper bound)) ซึ่งทำให้การประมาณการแจกแจงข้างต้นไม่สอดคล้องกับค่าของ x_0 ที่เปลี่ยนไป แต่ถ้าขอบเขตบนทางด้านขวามือของอสมการ (1.7) และ (1.8) สามารถเปลี่ยนแปลงไปตามค่าของ x_0 ที่เพิ่มขึ้น (เรียกว่าขอบเขตบนแบบไม่เอกรูป (non-uniform upper bound)) จะทำให้การประมาณการแจกแจงมีความถูกต้องมากขึ้นหรือมีความเหมาะสมมากขึ้น

วัตถุประสงค์ของการวิจัย

1. เพื่อหาขอบเขตบนแบบไม่เอกรูปสำหรับเมตริกแบบจุดระหว่างการแจกแจงของตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบและการแจกแจงทวินามนิเสธ

2. เพื่อเปรียบเทียบความถูกต้องของการประมาณระหว่างขอบเขตบนแบบเอกรูปและไม่เอกรูป

3. เพื่อศึกษาการประยุกต์ของการประมาณทวินามนิเสธ สำหรับการแจกแจงโพลยา

การแจกแจงโพลยานิเสธ การแจกแจงไฮเพอร์จีออเมตริก และการแจกแจงไฮเพอร์จีออเมตริกนิเสธ

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ได้ผลลัพธ์ของการประมาณการแจกแจงของตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบด้วยการแจกแจงทวินามนิเสธที่อยู่ในรูปของเมตริกแบบจุด

2. ได้ผลลัพธ์ของการประมาณการแจกแจงของตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบด้วยการแจกแจงเรขาคณิตที่อยู่ในรูปของเมตริกแบบจุด

เราสามารถนำผลลัพธ์ดังกล่าวไปประยุกต์ใช้ในการประมาณการแจกแจงที่เกี่ยวข้องอื่น ๆ ด้วยการแจกแจงทวินามนิเสธ หรือการแจกแจงเรขาคณิตในรูปของเมตริกแบบจุด และใช้ขอบเขตบนแบบไม่เอกรูปเป็นอีกเกณฑ์หนึ่งในการวัดความถูกต้องของการประมาณนี้

ขอบเขตของการวิจัย

งานวิจัยครั้งนี้เราจะศึกษาเฉพาะการประมาณการแจกแจงของตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบด้วยการแจกแจงทวินามนิเสธ และการแจกแจงเรขาคณิตที่อยู่ในรูปของเมตริกแบบจุดระหว่างการแจกแจงสองการแจกแจงที่เกี่ยวข้องพร้อมด้วยขอบเขตบนแบบไม่เอกรูปที่สอดคล้องกัน โดยใช้วิธีของสไตน์สำหรับการแจกแจงทวินามนิเสธ (หรือการแจกแจงเรขาคณิต) และฟังก์ชัน w ที่สัมพันธ์กับตัวแปรสุ่มที่มีค่าจำนวนเต็มไม่เป็นลบ