บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

โดยทั่วไปแล้วรูปแบบฟังก์ชันการแจกแจงของตัวแปรสุ่มจะถูกกำหนดด้วยพารามิเตอร์ตั้งแต่หนึ่งตัวขึ้นไป และค่าพารามิเตอร์ดังกล่าวจะเป็นตัวที่ไม่ทราบค่า การประมาณค่าของพารามิเตอร์จึงเป็นเรื่องหนึ่งที่ควรจะศึกษา การประมาณค่าพารามิเตอร์นั้น วิธีที่นิยมใช้กันอย่างแพร่หลายมีอยู่ด้วยกันหลายวิธี เช่น วิธีโมเมนต์ (Method of Moment) ตัวประมาณที่ได้จากวิธีนี้หาได้ง่าย แต่จะเป็นตัวประมาณที่ไม่พึงปรารถนาเพราะจะมีความแปรปรวนสูง วิธีความควรจะเป็นสูงสุด (Method of Maximum Likelihood) ตัวประมาณความควรจะเป็นสูงสุดจะเป็นตัวประมาณที่คงเส้นคงวาและเป็นฟังก์ชันของสถิติที่มีความเพียงพอ วิธีไคกำลังสองต่ำสุด (Method of Minimum Chi-Square) ตัวประมาณที่ได้จากวิธีนี้จะมีคุณสมบัติที่คล้ายคลึงกับตัวประมาณความควรจะเป็นสูงสุด แต่วิธีความควรจะเป็นสูงสุดดำเนินการได้ง่ายกว่าและเป็นวิธีที่นิยมใช้กันมากกว่าวิธีไคกำลังสองต่ำสุด และวิธีของเบย์ (Method of Bayes) วิธีนี้จะใช้ฟังก์ชันการเสี่ยงช่วยในการตัดสินใจ การหาตัวประมาณจากวิธีนี้จะต้องอาศัยความรู้บางอย่างเกี่ยวกับพารามิเตอร์ที่ต้องการจะประมาณ การประมาณด้วยวิธีนี้ค่อนข้างจะยุ่งยาก เป็นต้น

Jain และ Consul (1971) ได้เสนอฟังก์ชันความน่าจะเป็นของการแจกแจงทวินามลบแบบทั่วไป (The Generalized Negative Binomial Distribution , GNBD) ซึ่งเกิดจากการทดลองแบบแบร์นูลลี (Bernoulli trials)ซ้ำๆอย่างอิสระกัน โดยที่ X แทนจำนวนครั้งของผลไม่สำเร็จ x ครั้งก่อนที่จะได้ผลสำเร็จเป็นครั้งที่ $m+\beta x-x$ การทดลองในลักษณะนี้จะเรียกว่าเป็นการทดลองทวินามลบแบบทั่วไป ซึ่งจะได้ว่าตัวแปรสุ่ม X เป็นตัวแปรสุ่มที่มีการแจกแจงทวินามลบแบบทั่วไปซึ่งมีรูปแบบฟังก์ชันความน่าจะเป็นดังนี้

$$P(X=x) = \frac{m}{m+\beta x}\binom{m+\beta x}{x}\theta^{x}(1-\theta)^{m+\beta x-x}$$

เมื่อ $x = 0,1,2,3,4,...$

$0 < \theta < 1 \quad , \quad 1 \leq \beta < \theta^{-1} \quad , \quad m > 0$

โดยที่ค่าคาดหวัง (Expectation) ค่าความแปรปรวน (Variance) และค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation) สามารถคำนวนได้จากสูตร

$$E(X) = \frac{m\theta}{(1-\beta\theta)}$$

$$V(X) = \frac{m\theta(1-\theta)}{(1-\beta\theta)^3}$$

$$CV(X) = \left\{\frac{(1-\theta)}{m\theta(1-\beta\theta)}\right\}^{1/2} \times 100\,\%$$

การแจกแจงทวินามลบแบบทั่วไปจะลดรูปเป็น การแจกแจงทวินาม(Binomial Distribution) เมื่อค่า $\beta = 0$ และจะลดรูปเป็น การแจกแจงทวินามลบ (Negative Binomial Distribution) เมื่อค่า $\beta = 1$

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ของข้อมูลที่มีการแจกแจงทวินามลบแบบทั่วไป โดยใช้วิธีการประมาณค่าพารามิเตอร์ดังต่อไปนี้

1. วิธีความควรจะเป็นสูงสุด (Method of Maximum Likelihood)
2. วิธีสองโมเมนต์แรกและสัดส่วนศูนย์ (Method of First Two Moments and Proportion of Zeros)
3. วิธีไคกำลังสองต่ำสุด (Method of Minimum Chi-Square)
4. วิธีระยะห่างต่ำสุด (Method of Minimum Distance)

ทั้งนี้จะใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) เป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ทั้ง 4 วิธี

1.3 สมมติฐานของการวิจัย

การประมาณค่าพารามิเตอร์ด้วยวิธีไคกำลังสองต่ำสุด จะให้ค่าประมาณใกล้เคียงกับค่าจริงมากกว่าวิธีการประมาณค่าพารามิเตอร์วิธีอื่น ๆ

1.4 ขอบเขตของการวิจัย

1. ศึกษาความเอนเอียงและความแปรปรวนของวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด วิธีสองโมเมนต์แรกและสัดส่วนศูนย์ วิธีไคกำลังสองต่ำสุด และวิธีระยะห่างต่ำสุด

2. การวิจัยครั้งนี้จะจำลองตัวอย่างสุ่มที่มีขนาด 30 , 50 , 70 , 100 และ 200 โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique)

3. การวิจัยครั้งนี้ศึกษาภายใต้ลักษณะการแจกแจงทวินามลบแบบทั่วไปที่กำหนดค่าพารามิเตอร์ให้สอดคล้องกับค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation : CV) ระดับต่าง ๆ ดังต่อไปนี้

3.1 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 50 % ศึกษาที่ $\theta = 0.3$, $\beta = 1.5$, $m = 16.9697$ และ $\theta = 0.4$, $\beta = 1.1$, $m = 10.7143$

3.2 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 75 % ศึกษาที่ $\theta = 0.2$, $\beta = 1.5$, $m = 10.1587$ และ $\theta = 0.3$, $\beta = 1.5$, $m = 7.5421$

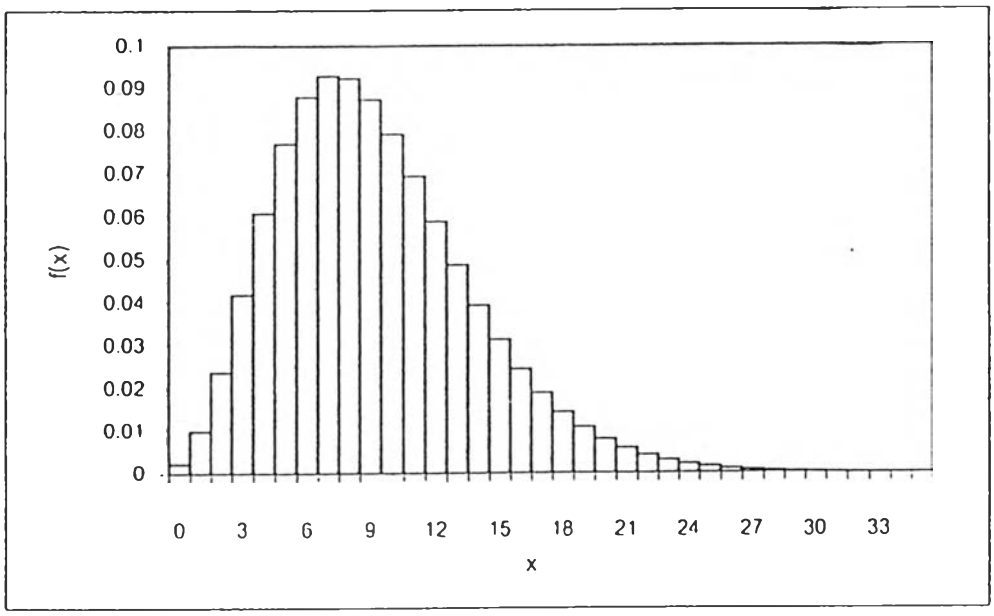
3.3 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 100 % ศึกษาที่ $\theta = 0.3$, $\beta = 1.5$, $m = 4.2424$ และ $\theta = 0.4$, $\beta = 1.1$, $m = 2.6786$

3.4 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 125 % ศึกษาที่ $\theta = 0.1$, $\beta = 2.0$, $m = 7.1999$ และ $\theta = 0.3$, $\beta = 1.5$, $m = 2.7152$

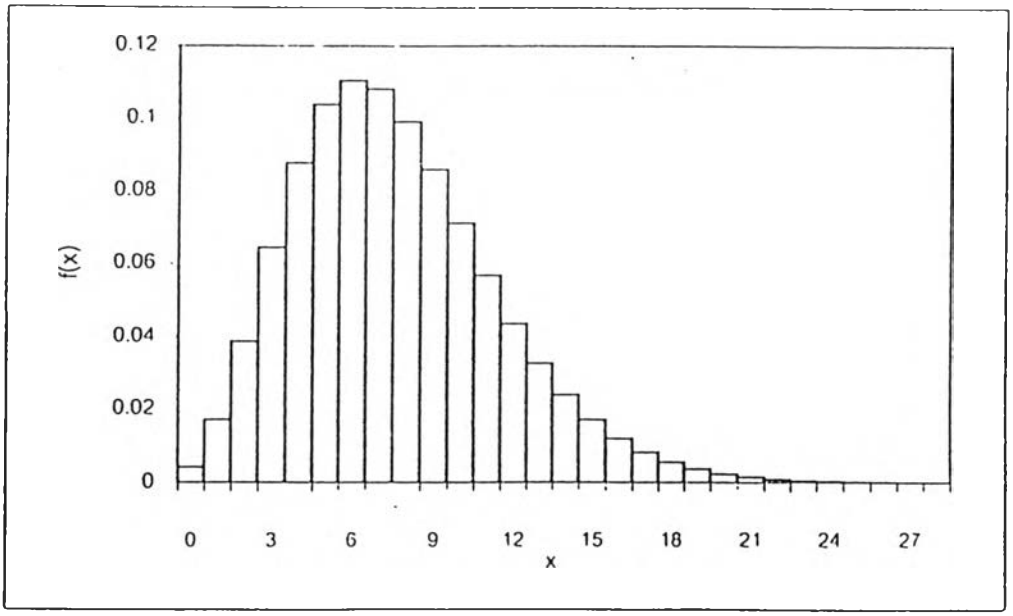
3.5 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 150 % ศึกษาที่ $\theta = 0.2$, $\beta = 1.5$, $m = 2.5397$ และ $\theta = 0.3$, $\beta = 1.5$, $m = 1.8855$

3.6 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 175 % ศึกษาที่ $\theta = 0.1$, $\beta = 2.0$, $m = 3.6735$ และ $\theta = 0.3$, $\beta = 1.5$, $m = 1.3853$

ซึ่งแผนภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ได้ทำการศึกษาในครั้งนี้ แสดงในแผ่นภูมิแท่งรูปที่ 1.1 – 1.12

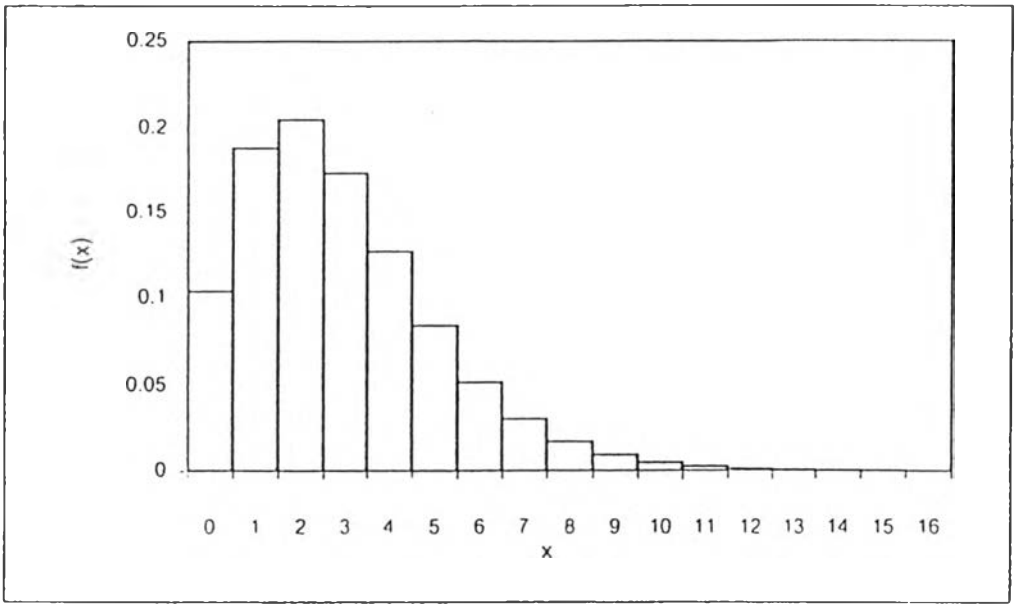


รูปที่ 1.1 แสดงแผนภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.3,\beta=1.5,m=16.9697$

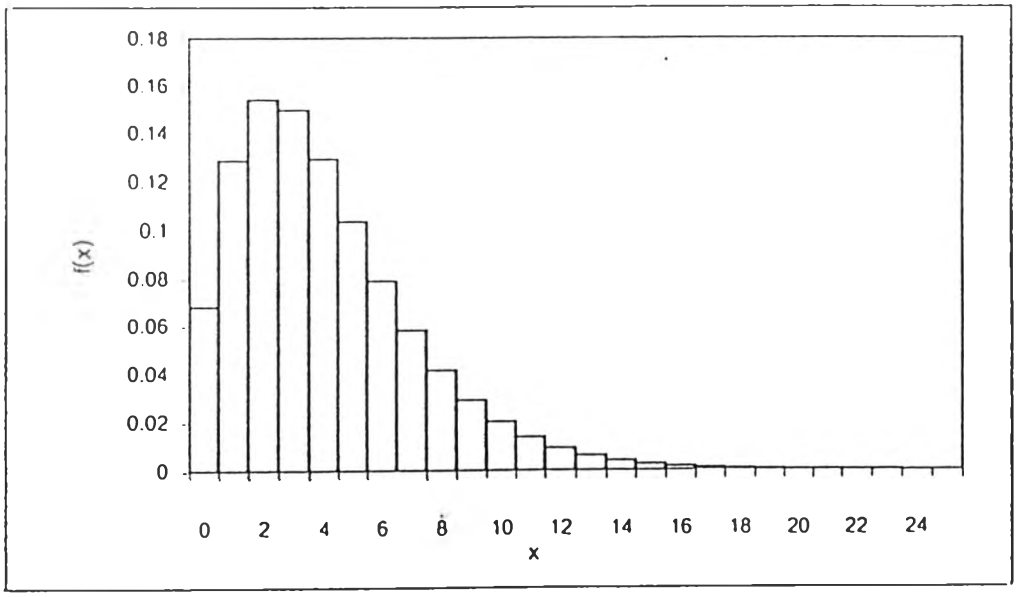


รูปที่ 1.2 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.4,\beta=1.1,m=10.7143$

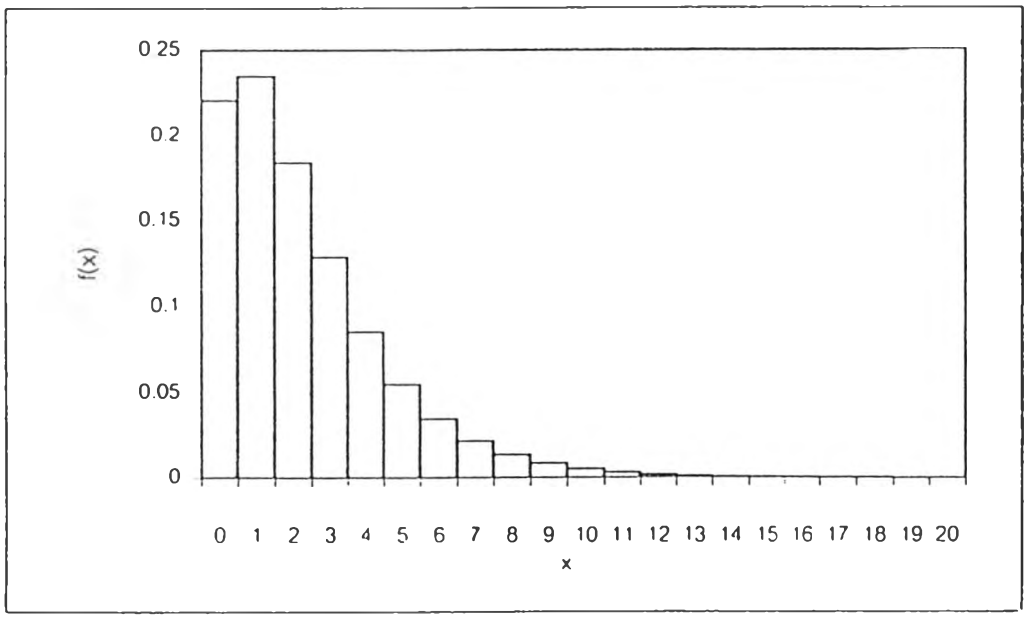
หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



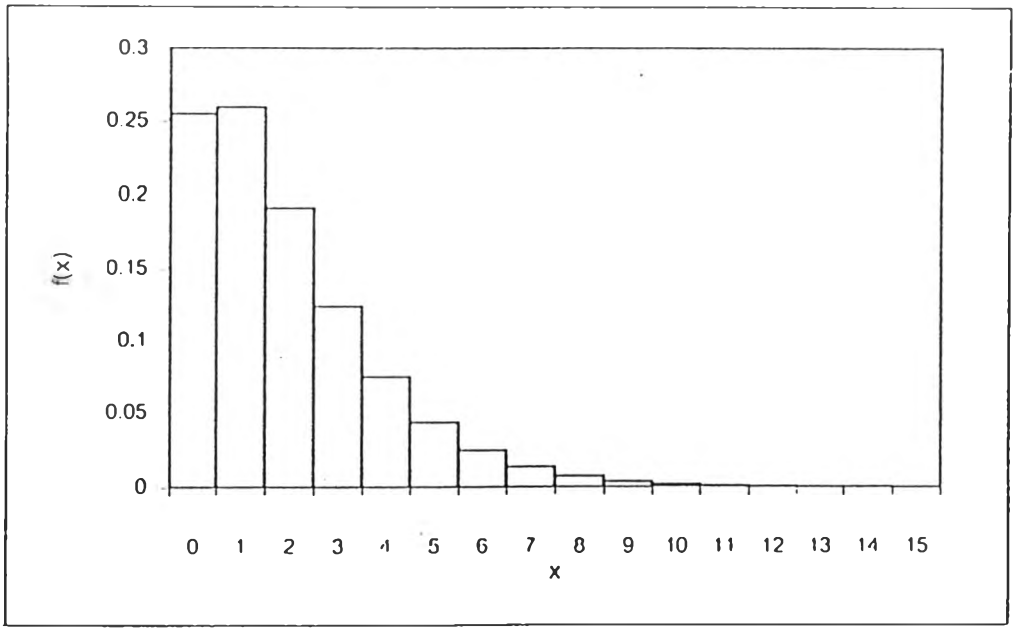
รูปที่ 1.3 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.2,\beta=1.5,m=10.1587$



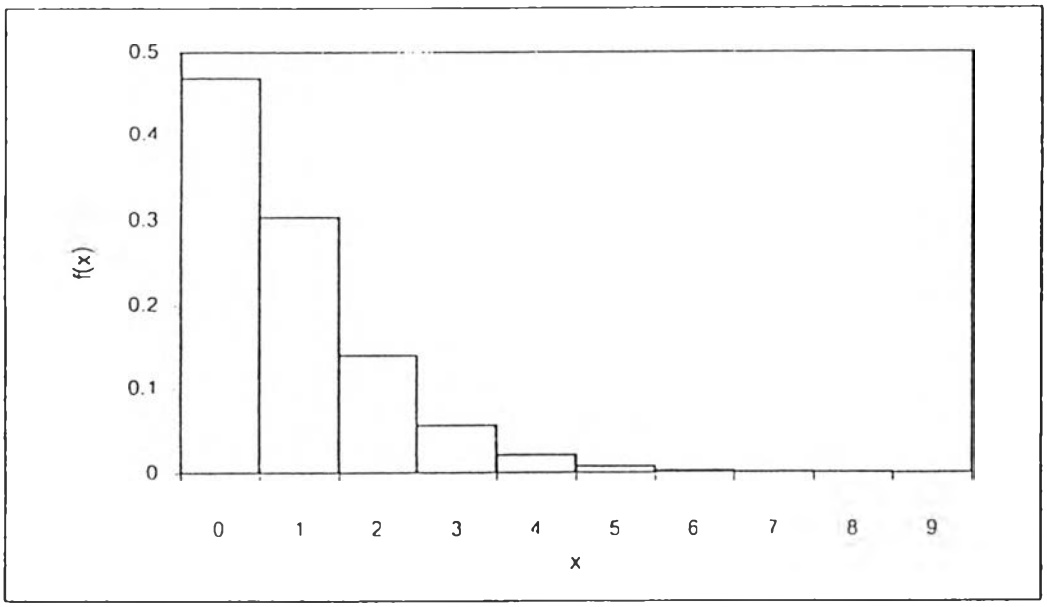
รูปที่ 1.4 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.3,\beta=1.5,m=7.5421$



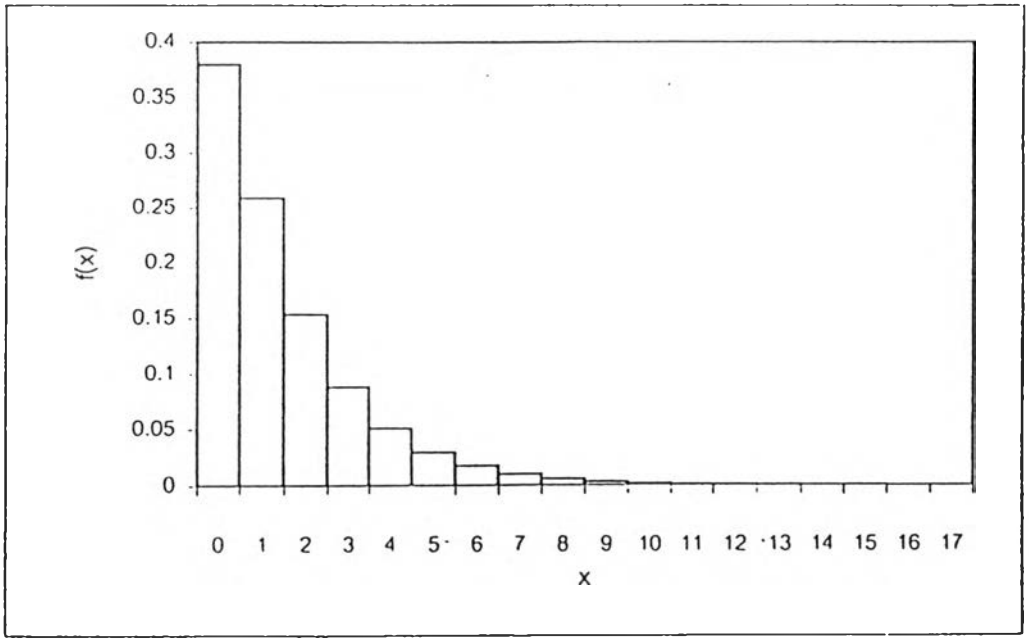
รูปที่ 1.5 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.3,\beta=1.5,m=4.2424$



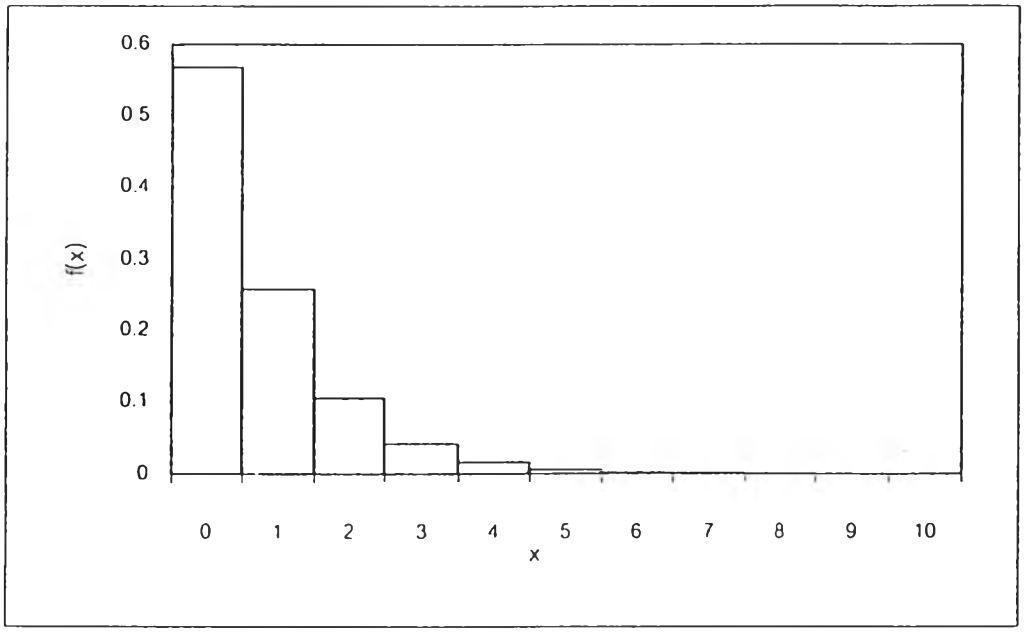
รูปที่ 1.6 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.4,\beta=1.1,m=2.6786$



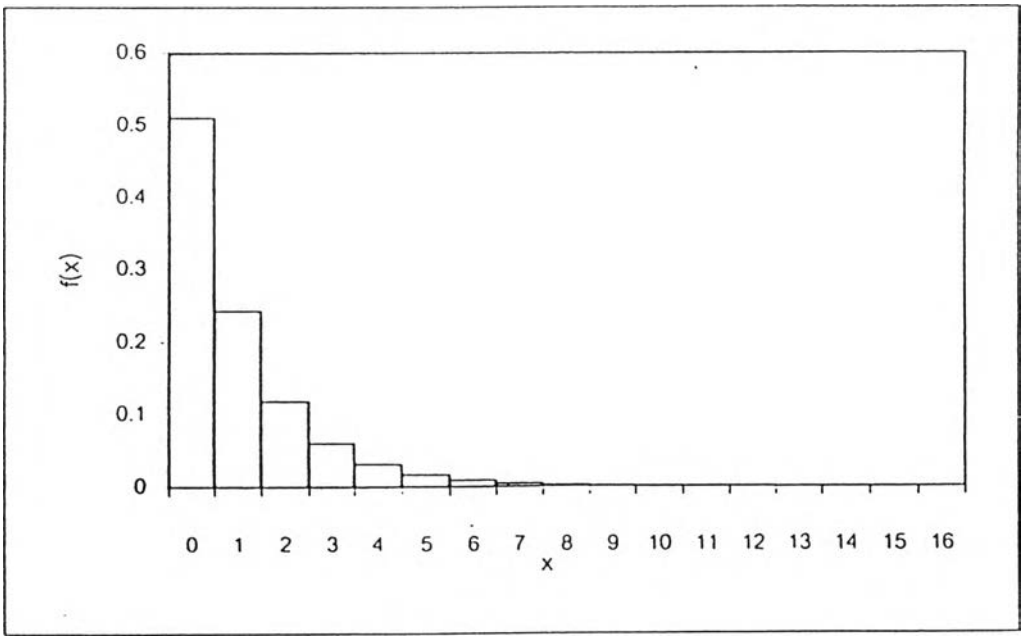
รูปที่ 1.7 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.1,\beta=2.0,m=7.1999$



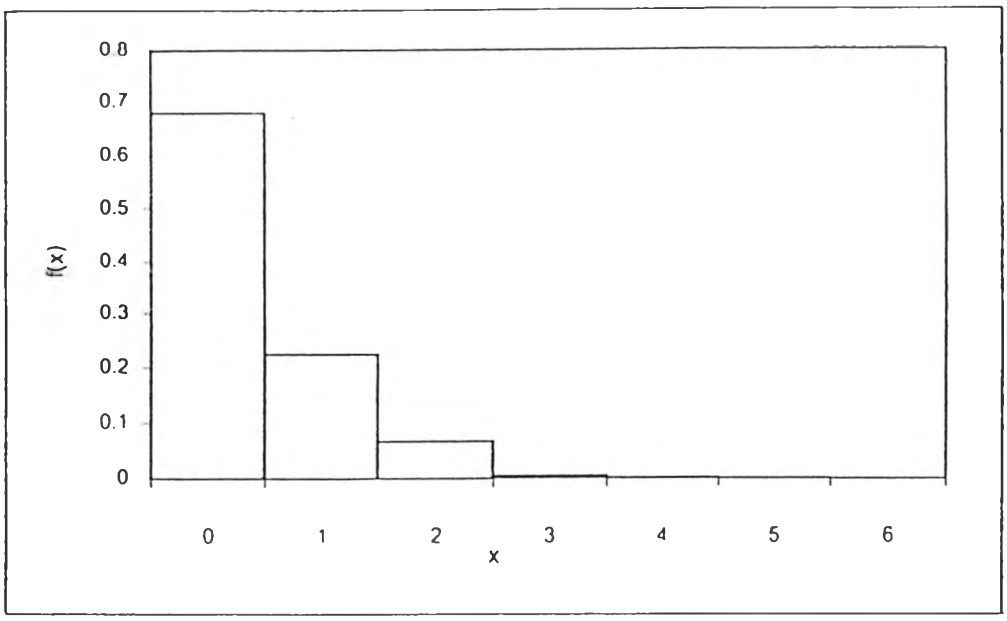
รูปที่ 1.8 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.3,\beta=1.5,m=2.7152$



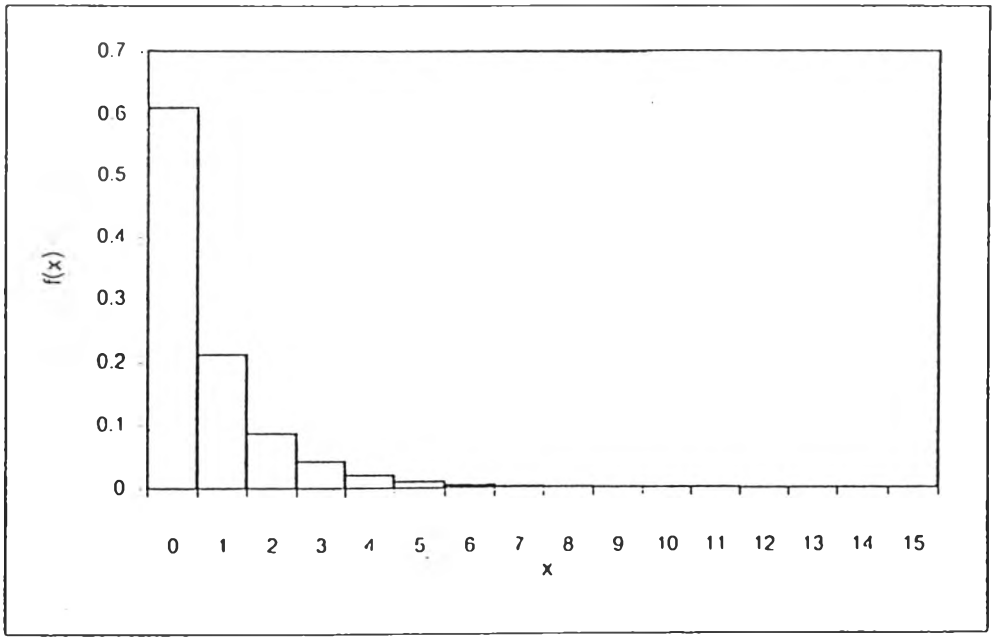
รูปที่ 1.9 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.2,\beta=1.5,m=2.5397$



รูปที่ 1.10 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.3,\beta=1.5,m=1.8855$



รูปที่ 1.11 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.1,\beta=2.0,m=3.6735$



รูปที่ 1.12 แสดงแผ่นภูมิแท่งของการแจกแจงทวินามลบแบบทั่วไปที่ $\theta=0.3,\beta=1.5,m=1.3853$

1.5 เกณฑ์การตัดสินใจ

ในการวัดคุณภาพของตัวประมาณ ในทฤษฎีการอนุมานทางสถิตินั้นคุณสมบัติของตัวประมาณที่ดีมีหลายประการด้วยกัน เช่น ความไม่เอนเอียง (Unbiasedness) ความคงเส้นคงวา (Consistency) ความเพียงพอ (Sufficiency) และความมีประสิทธิภาพ (Efficiency) แต่คุณสมบัติของตัวประมาณที่ดีประการแรกคือ ความไม่เอนเอียงซึ่งหมายความว่าค่าความคาดหวังของตัวประมาณที่ไม่เอนเอียงต้องเท่ากับค่าพารามิเตอร์ที่ต้องการจะประมาณ แต่ความไม่เอนเอียงเพียงอย่างเดียวอาจไม่เป็นการพอเพียงที่จะทำให้เชื่อได้ว่าตัวประมาณนั้นเป็นตัวประมาณที่ดี จึงจำเป็นต้องอาศัยคุณสมบัติอื่นประกอบด้วย คุณสมบัติอื่นที่มักใช้ทั่วไปคือ ค่าความแปรปรวนของตัวประมาณ (Variance) นั่นคือ ในการวิจัยครั้งนี้จะใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE) เป็นเกณฑ์ในการคัดเลือกตัวประมาณที่เหมาะสม เพราะค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจะรวมความเอนเอียงและความแปรปรวนเข้าด้วยกัน ซึ่งเท่ากับค่าความเอนเอียงกำลังสองบวกค่าความแปรปรวนของตัวประมาณ

เกณฑ์การตัดสินใจสำหรับการประมาณค่าพารามิเตอร์ของการแจกแจงทวินามลบแบบทั่วไป จะพิจารณาโดยการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าประมาณพารามิเตอร์กับค่าจริงในรูปของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE)

$$MSE\left(\hat{\theta}_p\right) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_{pi} - \theta_p)^2 \quad ; \quad p = 1,2,3$$

$$MSE = \frac{MSE\left(\hat{\theta}_1\right) + MSE\left(\hat{\theta}_2\right) + MSE\left(\hat{\theta}_3\right)}{3}$$

1.6 คำจำกัดความ

พารามิเตอร์ (Parameter) หมายถึง ค่าที่บ่งบอกลักษณะเฉพาะของประชากร

ความถี่กลุ่มตัวอย่างศูนย์ (Zero Frequency) หมายถึง จำนวนของตัวอย่างที่มีค่าเท่ากับศูนย์

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางให้นักวิจัยมีผลสรุปในการเลือกตัวประมาณค่าพารามิเตอร์สำหรับข้อมูลที่ได้จากการแจกแจงทวินามลบแบบทั่วไป
2. เพื่อเป็นแนวทางในการวิจัยหาตัวประมาณค่าพารามิเตอร์ต่อไป