บทที่ 3
วิธีดำเนินการวิจัย

การประมาณค่าพารามิเตอร์สำหรับการแจกแจงที่ถูกตัดปลายทางซ้ายเมื่อไม่ทราบจุดตัดปลาย โดยประมาณค่าพารามิเตอร์ด้วยวิธีโมเมนต์คัดแปลง วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุดคัดแปลงและวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด ในแต่ละวิธีการจะประมาณพารามิเตอร์ทำซ้ำจนกระทั่งได้ค่าประมาณพารามิเตอร์ ในแต่ละวิธีการจะประมาณพารามิเตอร์ ภายใต้สถานการณ์เมื่อขนาดตัวอย่างมี 5 ขนาด คือ 10, 30, 50, 70 และ 100 โดยที่จุดตัดปลายมี 5 ระดับคือ 500, 1000, 1500, 2000, 3000 และการแจกแจงที่ถูกตัดปลายทางซ้าย 3 แบบ คือ แบบลอกนอร์มอลตัดปลายทางซ้าย แบบเอกซโพเนนเชียลตัดปลายทางซ้าย และแบบพาเรโต้ตัดปลายทางซ้าย โดยที่ในแต่ละสถานการณ์มีการทำซ้ำ 1,000 ครั้ง

การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติ คาร์โล (Monte Carlo Simulation Technique) สร้างสถานการณ์ต่างๆ ดังนี้ เริ่มด้วยการกล่าวถึงวิธีการจำลองโดยใช้ เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของขั้นตอนการวิจัย และโปรแกรมที่ใช้ในการวิจัยตามลำดับ

วิธีจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าว จะใช้เลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้คือ

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคนี้ ทั้งนี้เพราะว่าหลักการของการจำลองแบบมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่ม มีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของเลขสุ่มที่ถูกสร้างขึ้นจะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) ตัวเลขสุ่มแต่ละตัวต้องเป็นอิสระต่อกัน และมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ (มีวัฏจักรยาว)

2. การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะนำไปผลิตเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3. การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องากรศึกษาได้แล้วขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำ ๆ กันหลายๆ ครั้ง เพื่อหาคำตอบที่ต้องการ

แผนการทดลอง

การวิจัยครั้งนี้ต้องการประมาณค่าพารามิเตอร์สำหรับการแจกแจงที่ถูกตัดปลายทางซ้ายเมื่อไม่ทราบจุดตัดปลาย การประมาณค่าพารามิเตอร์ ทั้ง 3 วิธี การแจกแจงที่ถูกตัดปลายทางซ้าย 3 แบบ ขนาดตัวอย่าง 5 ขนาด และจุดตัดปลายทั้ง 5 ระดับ รวมทั้งสิ้น 75 สถานการณ์ และค่าพารามิเตอร์ที่ประมาณได้ในแต่ละวิธีการ นำทำการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าประมาณกับค่าจริง ในรูปของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของทั้ง 3 วิธีเพื่อหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ต่อไป รายละเอียดแผนการทดลองมีดังนี้

1. การแจกแจงที่ถูกตัดปลายทางซ้ายเมื่อไม่ทราบจุดตัดปลาย 3 แบบ คือ แบบลอกนอร์มอลตัดปลายทางซ้าย แบบเอกซโพเนนเชียลตัดปลายทางซ้าย และแบบพาเรโต้ตัดปลายทางซ้าย ด้วยวิธีการประมาณพารามิเตอร์ 3 วิธีคือ วิธีโมเมนต์คัดแปลง วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีกำลังสองต่ำสุดเทียม

2. ขนาดตัวอย่าง 5 ขนาด คือ 10,30,50,70 และ 100

3. จุดตัดปลาย 5 ระดับ คือ 500,1000,1500,2000 และ 3000

ขั้นตอนในการวิจัย

ขั้นตอนของการวิจัย แบ่งเป็น 3 ขั้นตอนหลัก คือ

1. จำลองตัวแปรอิสระ X_i จากการแจกแจง 3 แบบ คือ แบบลอกนอร์มอลตัดปลายทางซ้าย แบบเอกซโพเนนเชียลตัดปลายทางซ้าย และแบบพาเรโต้ตัดปลายทางซ้าย

2. นำค่า X_i จากการแจกแจงแต่ละแบบมาหาค่าประมาณพารามิเตอร์ ด้วยวิธีการทั้ง 3 วิธี จะได้ค่าประมาณพารามิเตอร์แต่ละวิธี

3. หาค่าความคลาดเคลื่อนจากการประมาณค่าพารามิเตอร์ ในรูปของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) และทำการเปรียบเทียบซึ่งมีแต่ละขั้นตอนมีรายละเอียดดังนี้

3.1 จำลองตัวแปรอิสระ X_i จากการแจกแจง 3 แบบ คือ

3.1.1 แบบลอกนอร์มอลตัดปลายทางซ้าย (d,μ,σ)(รายละเอียดในภาคผนวก ก)

3.1.2 แบบเอกซโพเนนเชียลตัดปลายทางซ้าย(d,θ)(รายละเอียดในภาคผนวก ก)

3.1.3 แบบพาเรโต้ตัดปลายทางซ้าย(d,α,λ)(รายละเอียดในภาคผนวก ก)

รูปการแจกแจงต่างๆ ที่ใช้ในการวิจัยครั้งนี้ได้แสดงไว้ใน รูปที่ 3.1-3.3

3.2 นำค่า X_i จากการแจกแจงแต่ละแบบมาหาค่าประมาณพารามิเตอร์ ด้วยวิธีการทั้ง 3 วิธี คือ วิธีโมเมนต์ตัดแปลง วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีกำลังสองต่ำสุดเทียม จะได้ค่าประมาณพารามิเตอร์แต่ละวิธี การหาค่าประมาณพารามิเตอร์แต่ละวิธีการมีรายละเอียดดังนี้

3.2.1 วิธีโมเมนต์คัดแปลง มีขั้นตอนดังนี้

3.2.1.1 กรณีการแจกแจงแบบลอกนอร์มอลตัดปลายทางซ้าย และการแจกแจงแบบพาเรโต้ตัดปลายทางซ้าย ที่ประมาณพารามิเตอร์ 3 พารามิเตอร์มีขั้นตอนดังนี้คือ

3.2.1.1.1 ประมาณค่าพารามิเตอร์บอกขนาดและพารามิเตอร์บอกรูปร่างด้วย

$$E(X) = \overline{X} \tag{3.1}$$

$$E(X^2) = \sum_{i=1}^{n} \frac{X_i^2}{n} \tag{3.2}$$

3.2.1.1.2 ประมาณค่าพารามิเตอร์บอกตำแหน่ง (จุดตัดปลาย) ด้วย

$$E(Y_1) = y_1 \tag{3.3}$$

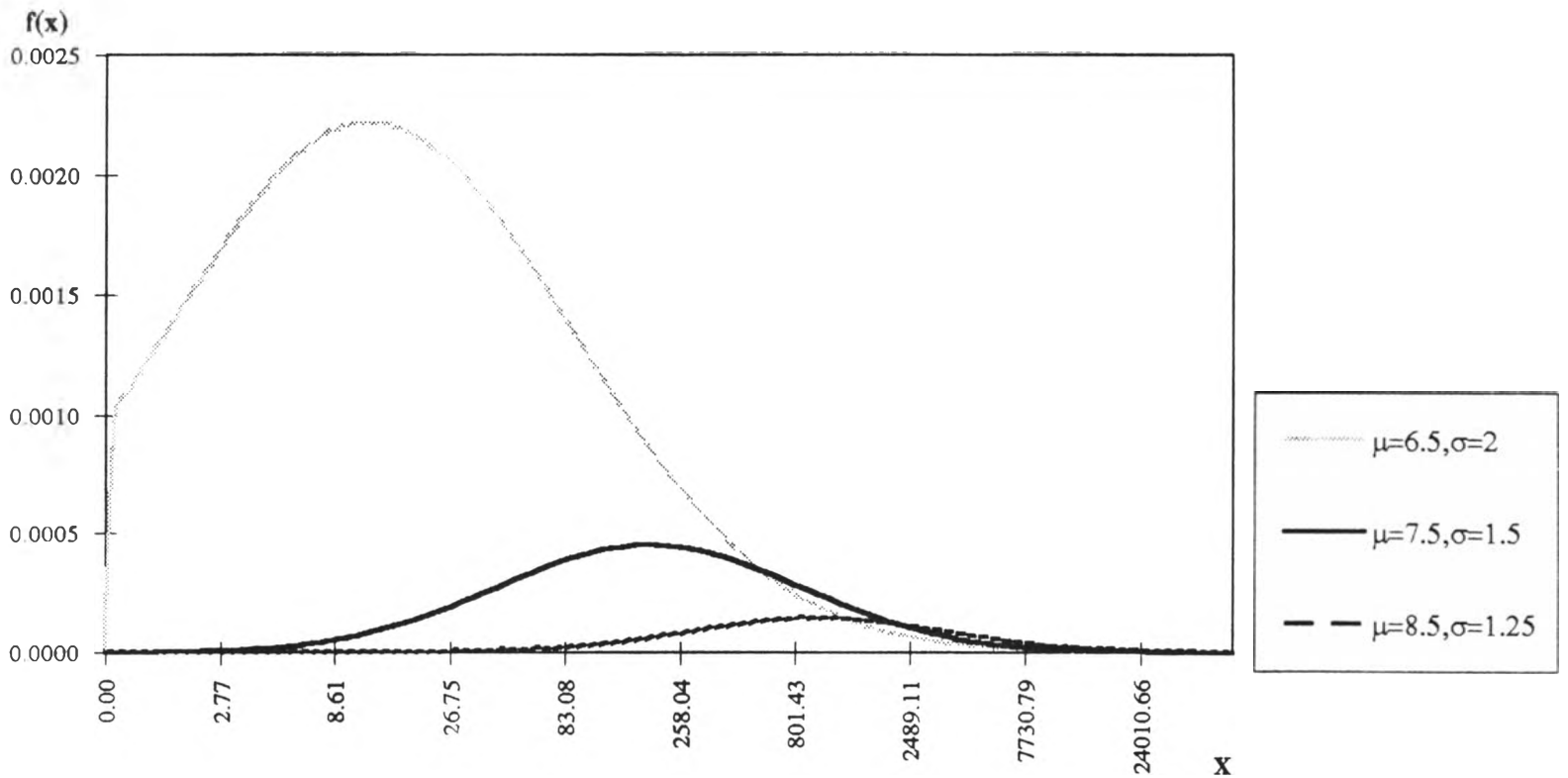
$$E(F(Y_1)) = F(y_1) \tag{3.4}$$

สมการที่ 3.3 ใช้ในการแจกแจงแบบพาเรโต้ตัดปลายทางซ้าย และสมการที่ 3.4 ใช้ในการแจกแจงแบบลอกนอร์มอล โดยที่ $Y_1=\min(X(1),...X(n))$

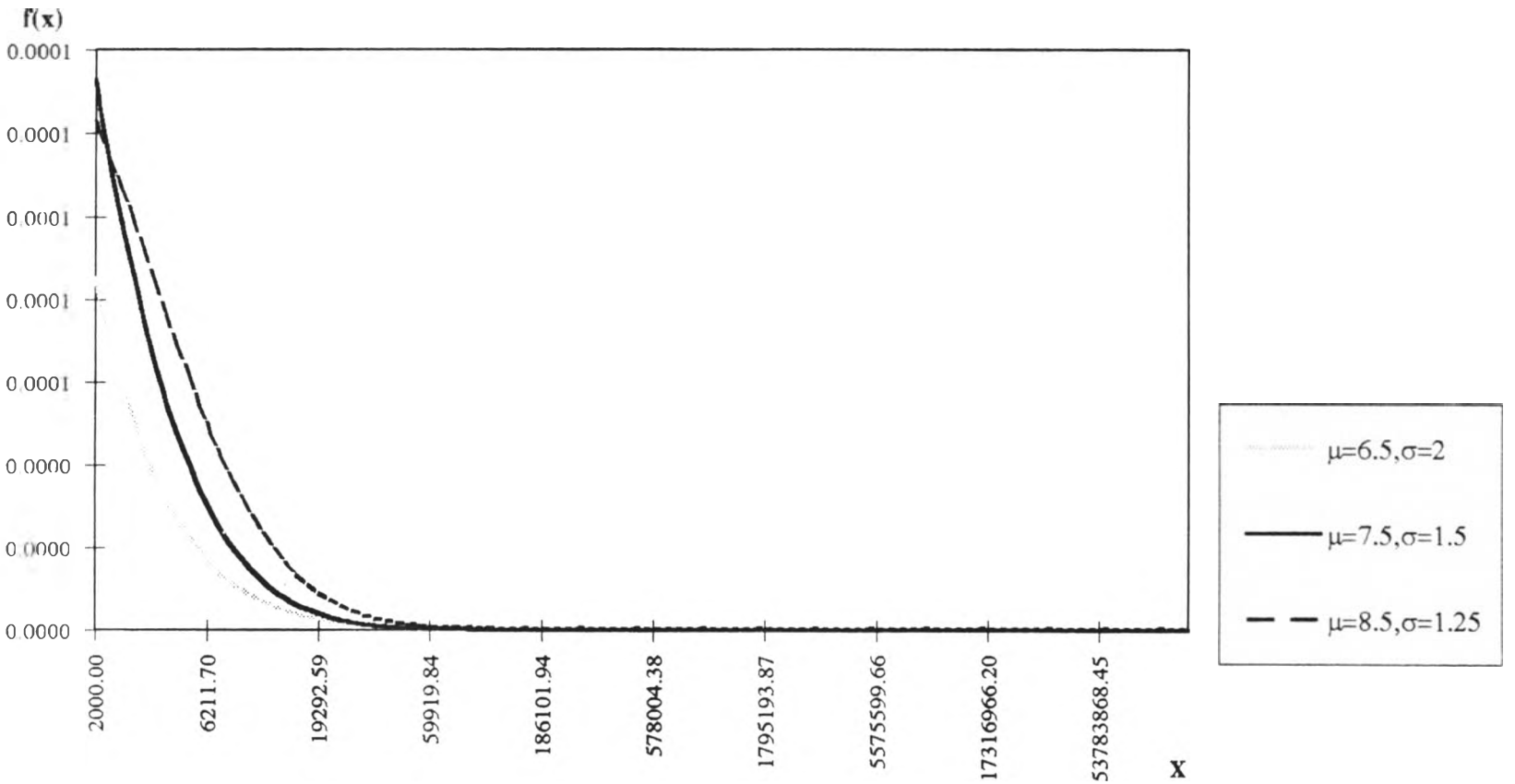
3.2.1.1.3 ทำการแก้สมการทั้ง 3 สมการออกมาเพื่อให้ได้ค่าประมาณพารมิเตอร์ถ้าสมการมีความยุ่งยากใช้วิธีการเชิงตัวเลข (Numerical Method) เข้าช่วยในงานวิจัยนี้ใช้วิธีการนิวตันราฟสัน (Newton-Raphson) ช่วยในการหาค่าประมาณพารามิเตอร์

3.2.1.2 กรณีการแจกแจงแบบเอกซโพเนนเชียลตัดปลายทางซ้าย ที่ประมาณพารามิเตอร์ 2 พารามิเตอร์มีขั้นตอนดังนี้คือ

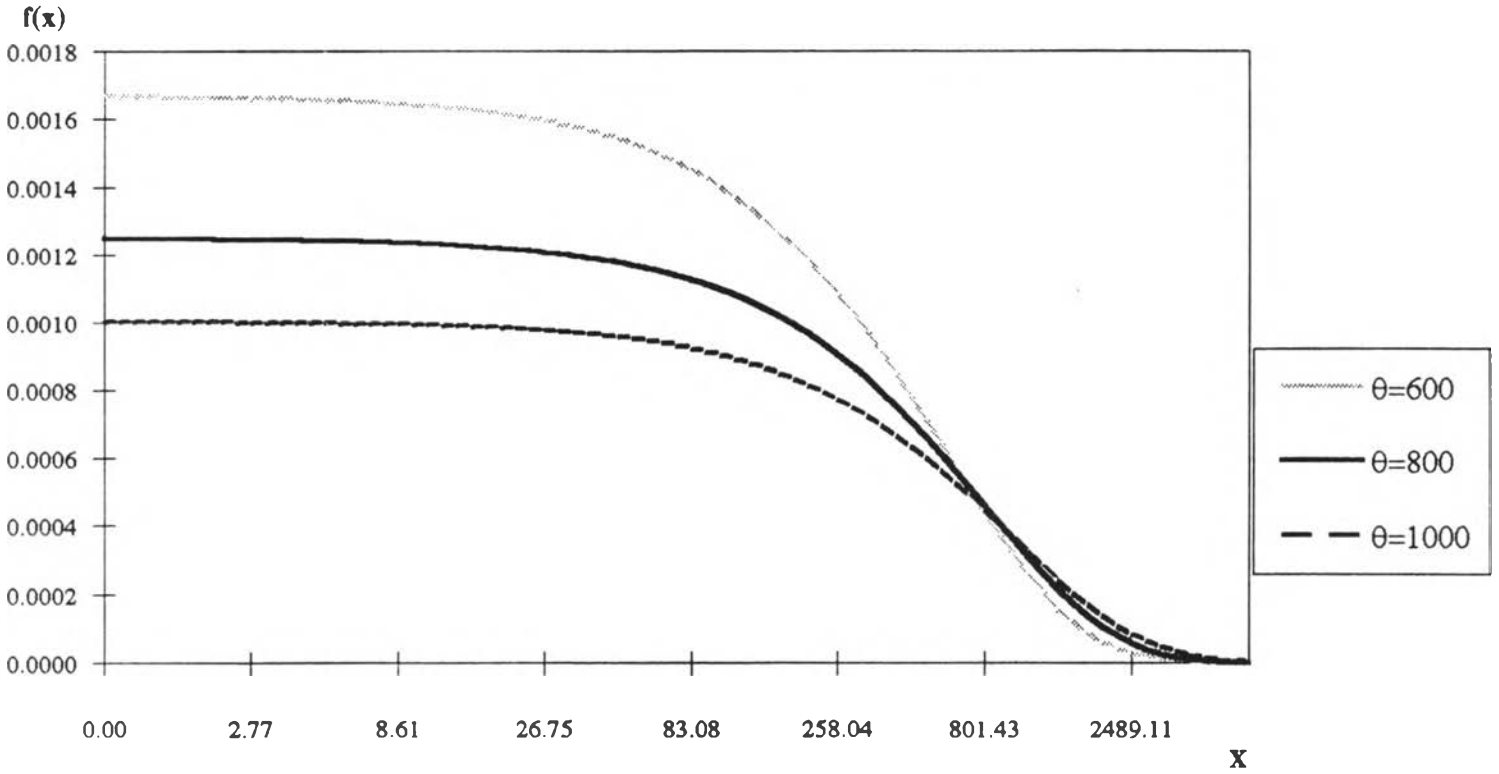
รูปที่ 3.1 แสดงการแจกแจงแบบลอกนอร์มอล



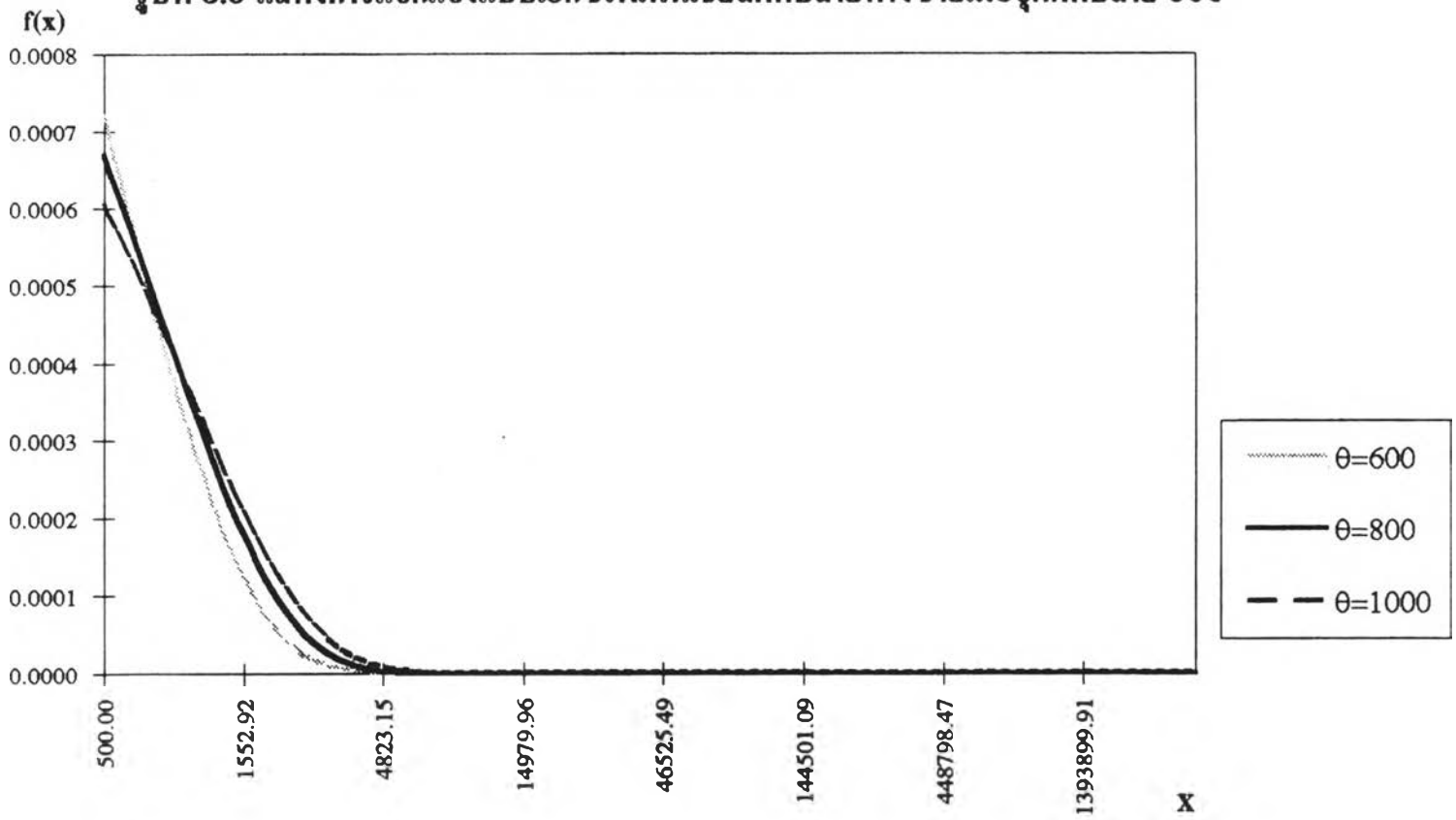
รูปที่ 3.3 แสดงการแจกแจงแบบลอกนอร์มอลตัดปลายทางซ้ายเมื่อจุดตัดปลาย 2000



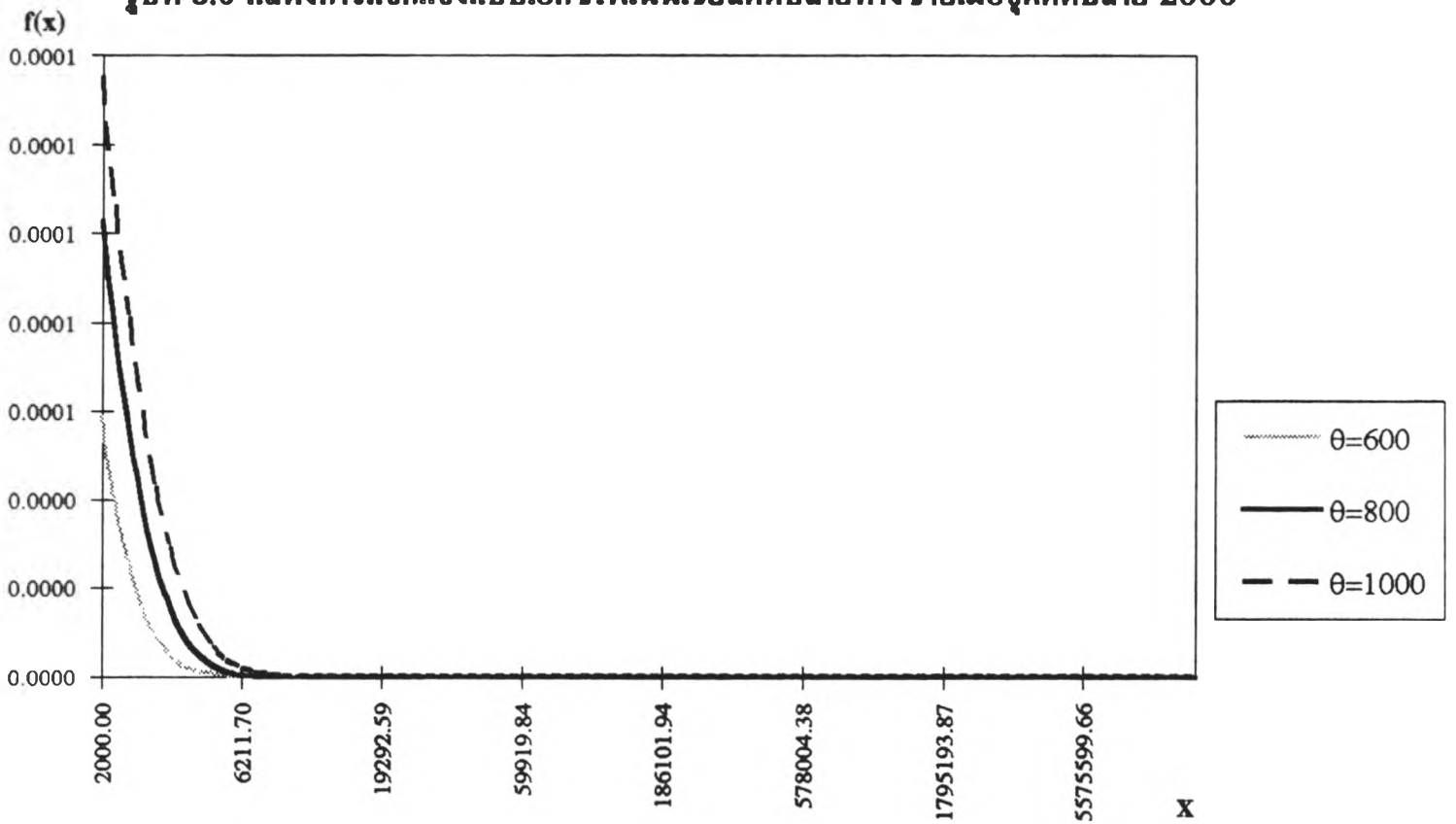
รูปที่ 3.4 แสดงการแจกแจงแบบเอกซโพเนนเชียล



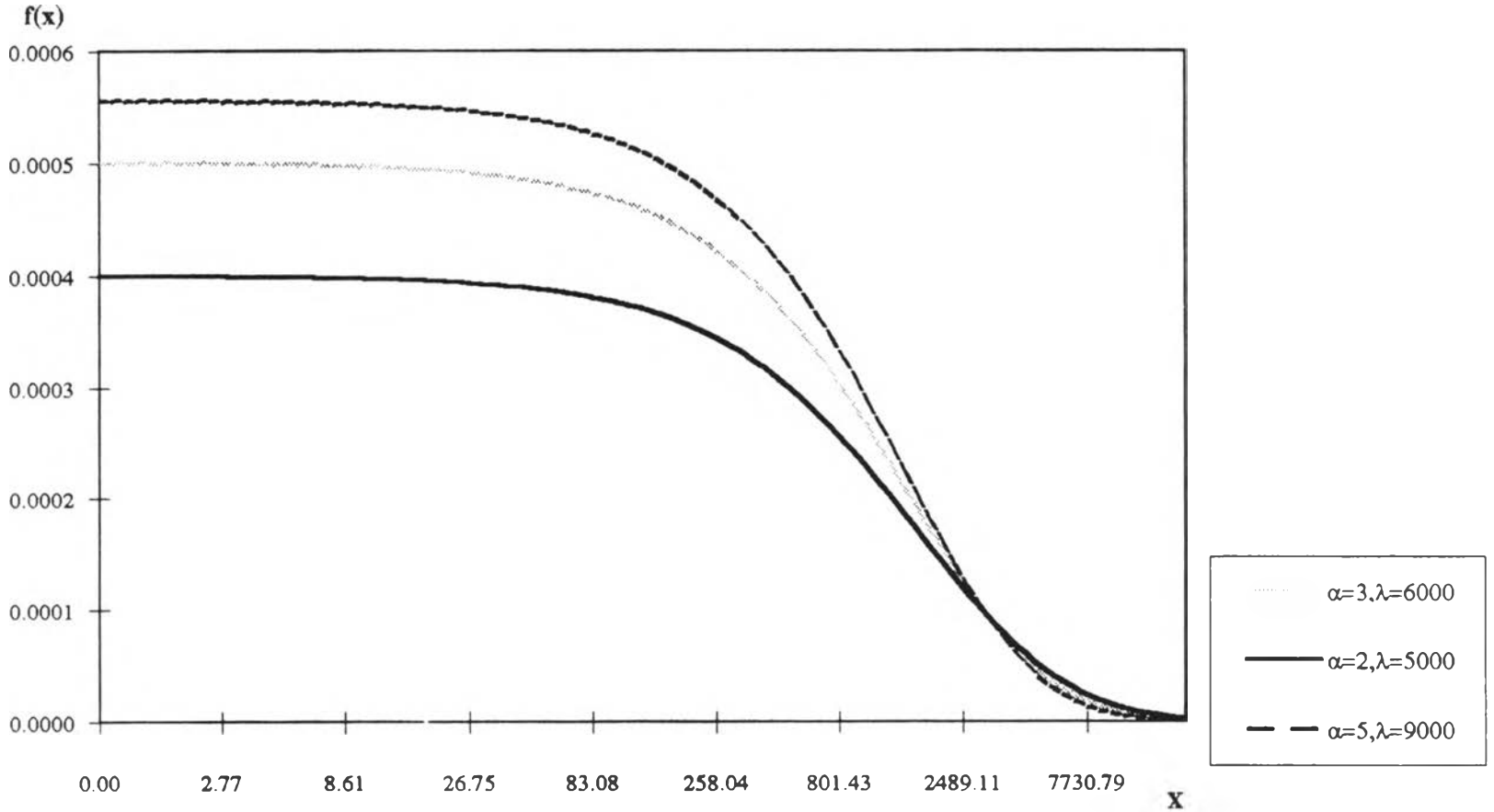
รูปที่ 3.5 แสดงการแจกแจงแบบเอกซโพเนนเชียลตัดปลายทางซ้ายเมื่อจุดตัดปลาย 500



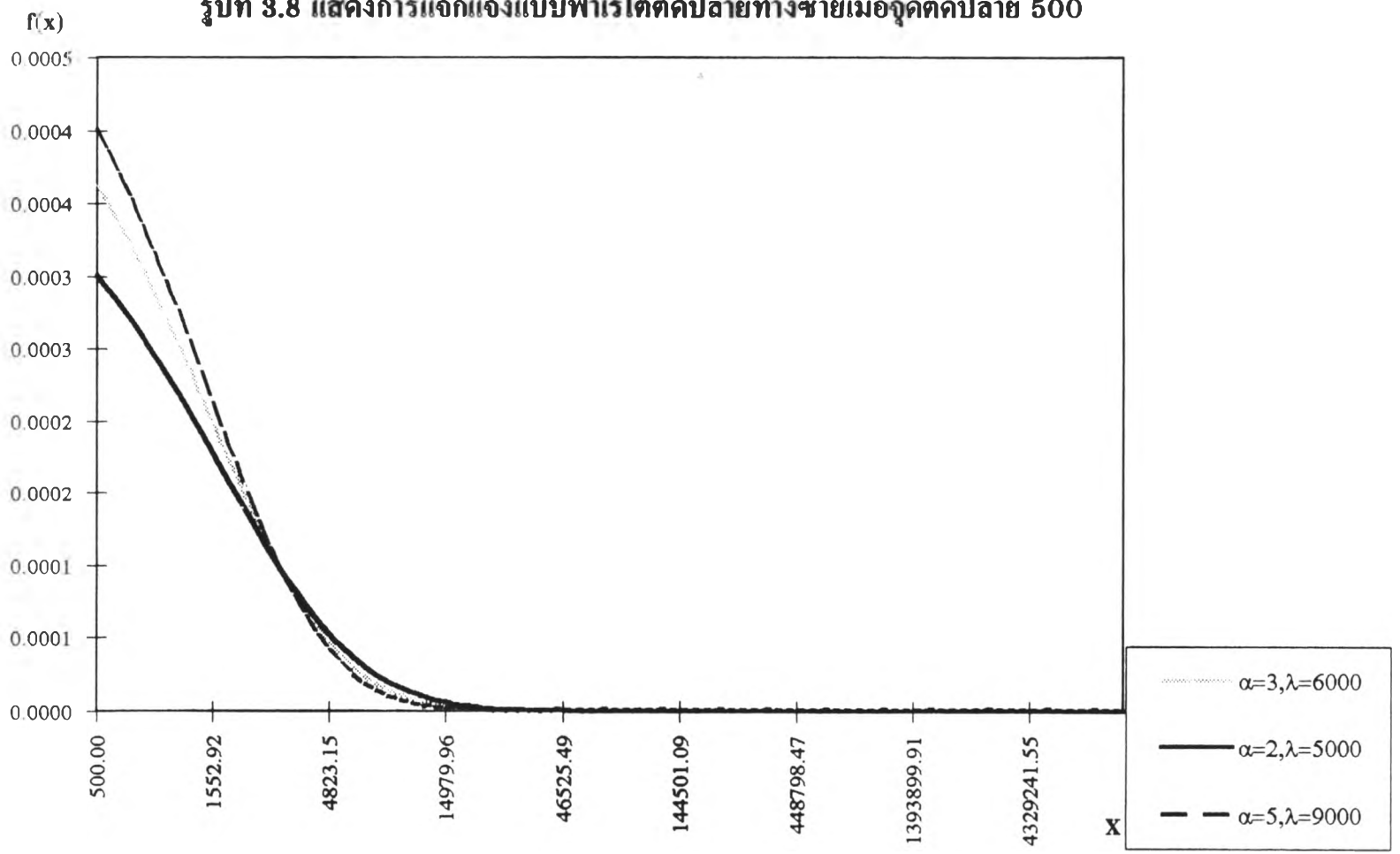
รูปที่ 3.6 แสดงการแจกแจงแบบเอกซโพเนนเชียลตัดปลายทางซ้ายเมื่อจุดตัดปลาย 2000



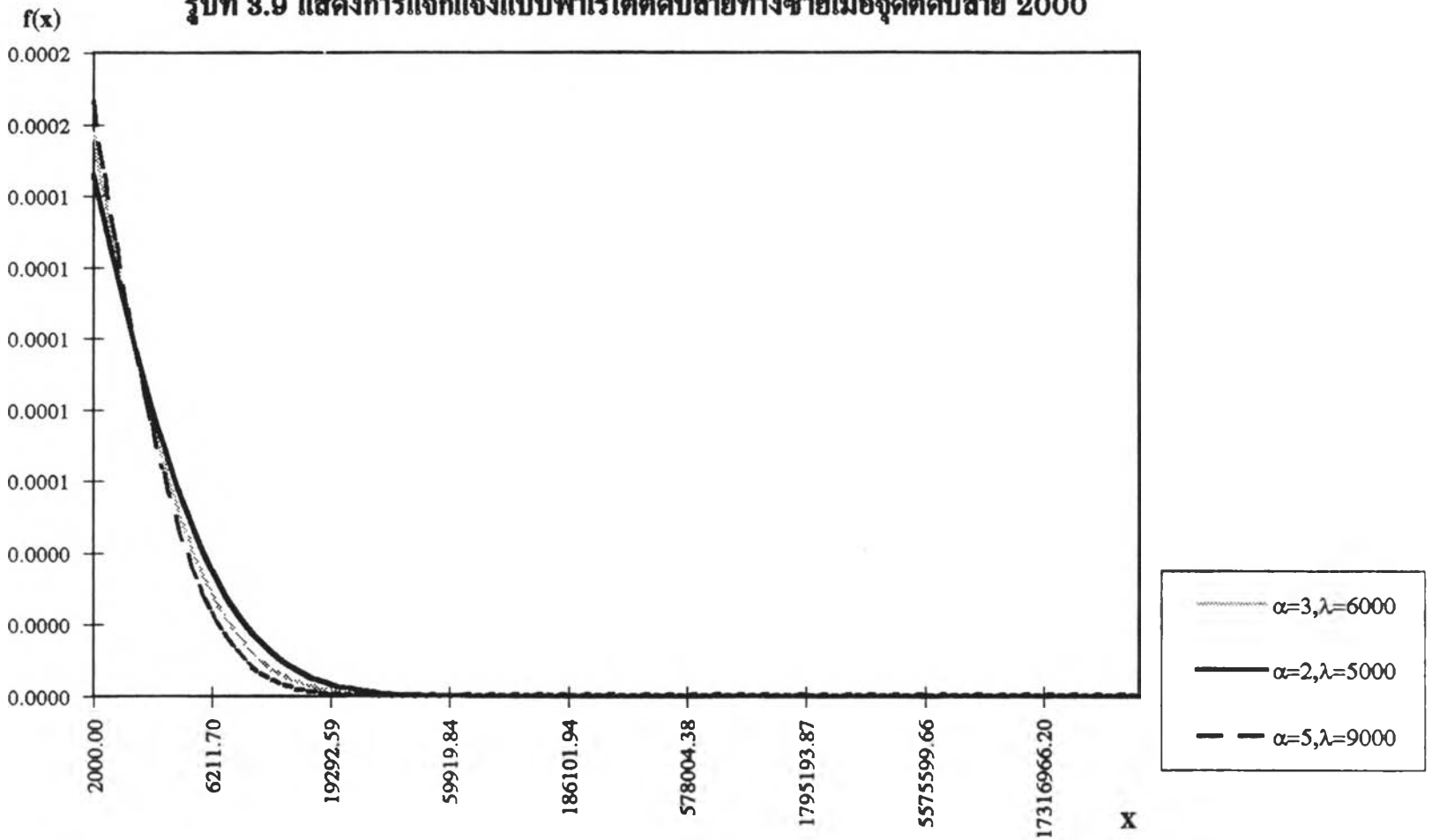
รูปที่ 3.7 แสดงการแจกแจงแบบพาเรโต้



รูปที่ 3.8 แสดงการแจกแจงแบบพาเรโต้ตัดปลายทางซ้ายเมื่อจุดตัดปลาย 500



รูปที่ 3.9 แสดงการแจกแจงแบบพาเรโต้ตัดปลายทางซ้ายเมื่อจุดตัดปลาย 2000



3.2.1.2.1 ประมาณค่าพารามิเตอร์บอกขนาดและพารามิเตอร์บอกรูปร่างด้วย

$$E(X) = \overline{X} \tag{3.5}$$

3.2.1.2.2 ประมาณค่าพารามิเตอร์บอกตำแหน่ง (จุดตัดปลาย) ด้วย

$$E(Y_1) = y_1 \tag{3.6}$$

โดยที่ y_1=min(X(1),..X(n))

3.2.1.3 ทำการแก้สมการทั้ง 2 สมการ ออกมา เพื่อให้ได้ค่าประมาณพารามิเตอร์ถ้าสมการมีความยุ่งยากใช้วิธีการเชิงตัวเลข (Numerical Method) เข้าช่วยในงานวิจัยนี้ใช้วิธีการนิวตันราฟสัน (Newton-Raphson) ช่วยในการหาค่าประมาณพารามิเตอร์

3.2.2 วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด มีขั้นตอนดังนี้

3.2.2.1 สำหรับการประมาณ ค่าพารามิเตอร์สำหรับพารามิเตอร์บอกขนาด และบอกรูปร่างมีขั้นตอนในการประมาณดังนี้ ฟังก์ชันภาวะน่าจะเป็นของการแจกแจงที่ถูกตัดปลายทางซ้าย เมื่อไม่ทราบจุดตัดปลายเป็นดังนี้

$$L = \prod_{i=1}^{n} f_w(x_i) \tag{3.7}$$

3.2.2.2 ประมาณค่าพารามิเตอร์ โดยการหาอนุพันธ์บางส่วนของลอกของฟังก์ชันภาวะน่าจะเป็น เทียบกับพารามิเตอร์ และให้อนุพันธ์บางส่วนเท่ากับ 0 $\frac{\partial \ln(L)}{\partial \mu}$=0 , $\frac{\partial \ln(L)}{\partial \sigma}$=0 สำหรับการแจกแจงแบบลอกนอร์มอลตัดปลายทางซ้าย $\frac{\partial \ln(L)}{\partial \alpha}$=0 , $\frac{\partial \ln(L)}{\partial \lambda}$=0 สำหรับการแจกแจงแบบพาเรโต้ตัดปลายทางซ้าย และ $\frac{\partial \ln(L)}{\partial \theta}$=0 สำหรับการแจกแจงแบบเอกซโพเนนเชียลตัดปลายทางซ้าย

3.2.2.3 สำหรับการประมาณค่าพารามิเตอร์บอกตำแหน่ง(จุดตัดปลาย) ประมาณด้วยค่าข้อมูลที่มีค่าต่ำสุดเนื่องจากเป็นค่าที่ใกล้เคียงจุดตัดปลายที่สุด

$$\hat{d} = y_1 = \min(X(1),...X(n))$$

3.2.3 วิธีกำลังสองต่ำสุด มีขั้นตอนดังนี้

3.2.3.1 สำหรับการประมาณค่าพารามิเตอร์สำหรับพารามิเตอร์บอกขนาด และบอกรูปร่างมีขั้นตอนในการประมาณโดย การหาค่าผลบวกกำลังสองของผลต่างระหว่างค่า $\ln\left(\left[1-F_n(x_i)\right]^{-1}\right)$ กับค่า $\ln\left(\left[1-F_w(x_i)\right]^{-1}\right)$ โดยที่ $F_n(x_i)$ เป็นฟังก์ชันการแจกแจงของตัวอย่าง และ $F_w(x_i)$ เป็นฟังก์ชันการแจกแจงสะสมแบบตัดปลายทางซ้ายเมื่อไม่ทราบจุดตัดปลาย เป็นดังนี้

I17262624

$$LQ = \sum_{i=1}^{n}\left(y_i - \ln\left(\left[1 - F_n(x_i)\right]^{-1}\right)\right)^2 \qquad (3.8)$$

โดยที่ $y_i = \left[1 - F_n(x_i)\right]^{-1}$, $F_n(x_i) = \frac{k_i}{n+1}$ และ k_i คือจำนวนข้อมูลที่มีค่าน้อยหรือเท่ากับ x_i

3.2.3.2 ประมาณค่าพารามิเตอร์ โดยการหาอนุพันธ์บางส่วนของฟังก์ชัน LQ เทียบกับพารามิเตอร์ และให้อนุพันธ์บางส่วนเท่ากับ 0 $\frac{\partial \ln(LQ)}{\partial \mu}=0$, $\frac{\partial \ln(LQ)}{\partial \sigma}=0$ สำหรับการแจกแจงแบบลอกนอร์มอลตัดปลายทางซ้าย $\frac{\partial \ln(LQ)}{\partial \alpha}=0$, $\frac{\partial \ln(LQ)}{\partial \lambda}=0$ สำหรับการแจกแจงแบบพาเรโต้ตัดปลายทางซ้าย และ $\frac{\partial \ln(LQ)}{\partial \theta}=0$ สำหรับการแจกแจงแบบเอกซโพเนนเชียลตัดปลายทางซ้าย

3.2.3.3 สำหรับการประมาณค่าพารามิเตอร์บอกตำแหน่ง(จุดตัดปลาย) ประมาณด้วยค่าข้อมูลที่มีค่าต่ำสุดเนื่องจากเป็นค่าที่ใกล้เคียงจุดตัดปลายที่สุด

$$\hat{d} = y_1 = \min(X(1),...X(n))$$

3.3 หาค่าความคลาดเคลื่อนจากการประมาณค่าพารามิเตอร์สำหรับการแจกแจงที่ถูกตัดปลายทางซ้ายเมื่อไม่ทราบจุดตัดปลาย เทียบกับค่าพารามิเตอร์จริง เพื่อคำนวณหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนของความคลาดเคลื่อนกำลังสอง (RMSE) เนื่องจากในการทดลองได้จำลองข้อมูลซ้ำๆ กัน จำนวน 1,000 ครั้ง ในแต่ละสถานการณ์ให้ j แทนรอบที่ทำซ้ำ j=1,...,1000 ดังนั้น การหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองหาได้จากสูตรดังนี้

$$MSE = \frac{1}{1000 * n}\sum_{j=1}^{1000}\sum_{i=1}^{n}\left(\theta_{ij} - \hat{\theta}_{ij}\right)^2$$

$$RMSE = \sqrt{MSE}$$

โดยที่ n แทน จำนวนพารามิเตอร์ ; i=1,...,n

j แทน จำนวนรอบที่ทำซ้ำ ; j=1,...,1000

MSE แทน ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง

RMSE แทน ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง

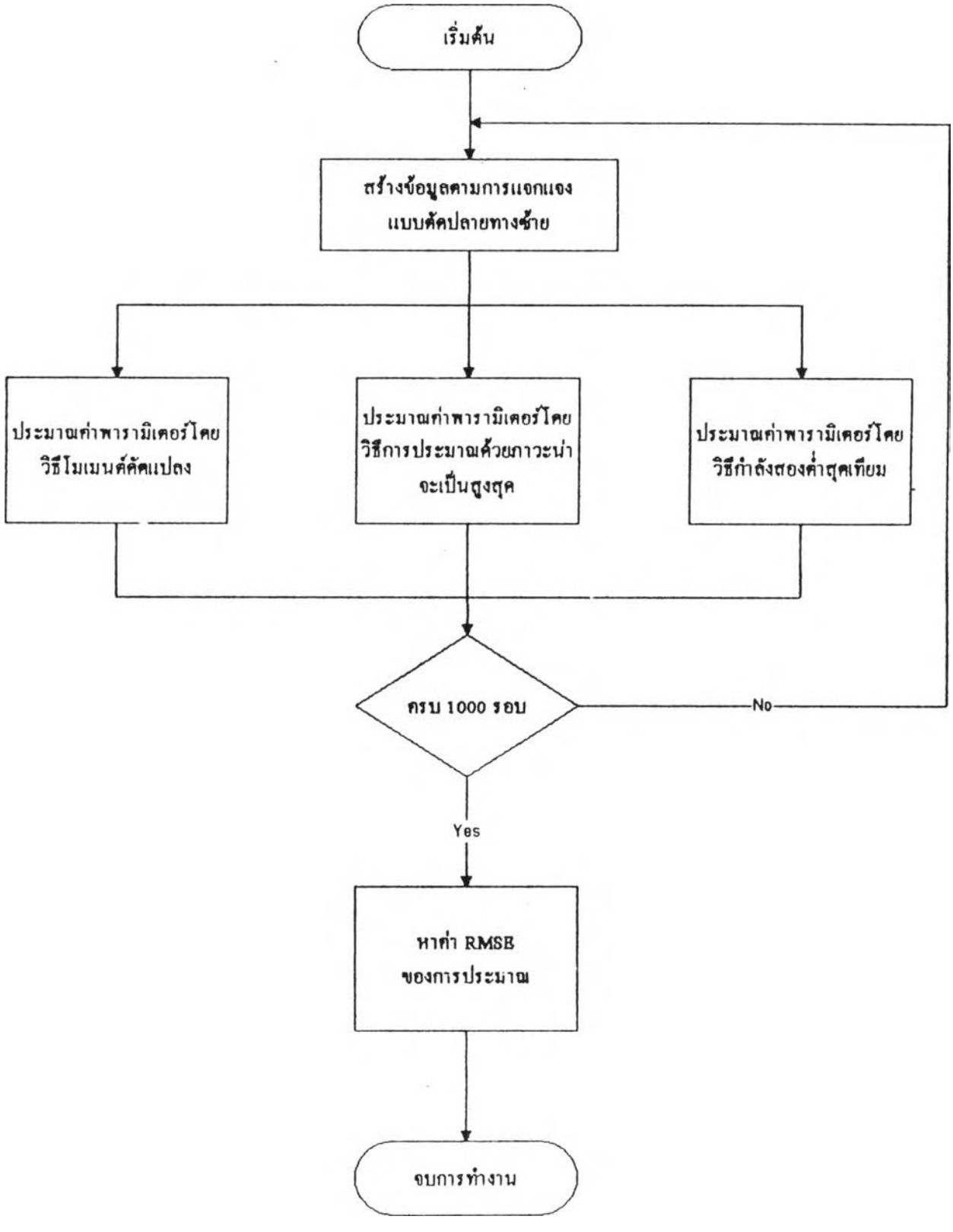
จากนั้นนำค่า RMSE ของการประมาณค่าตัวแปรตาม ของวิธีการทั้ง 3 วิธี มาเปรียบเทียบเพื่อหาว่าวิธีการใดให้ค่า RMSE ของการประมาณค่าพารามิเตอร์ต่ำที่สุดจะเป็นวิธีที่ประมาณค่าพารามิเตอร์ของแต่ละสถานการณ์ได้ดีที่สุด

การคำนวณค่า RMSE ของการประมาณค่าพารามิเตอร์ ทั้ง 3 วิธี ในการทดลองที่ขนาดตัวอย่างหนึ่ง จะเปลี่ยนการแจกแจง 3 แบบ เปลี่ยนจุดตัดปลาย 5 ระดับ และจะเปลี่ยนขนาด

ของตัวอย่าง 5 ขนาด โดยแต่ละสถานการณ์จะทดลองซ้ำๆ จำนวน 1,000 ครั้ง จนครบทุกสถานการณ์ ซึ่งขั้นตอนของการทดลองดังกล่าวนี้ สรุปเป็นผังงานได้ดังรูปที่ 3.4

โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมด เขียนด้วยภาษาฟอร์แทรน (FORTRAN) โดยใช้กับเครื่อง AMSAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลอง ลักษณะของการทำงานของโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของโปรแกรมจะแสดงไว้ในภาคผนวก ข ซึ่งจะเป็นโปรแกรมการทำงานของแต่ละการแจกแจง การแจกแจงแบบลอกนอร์มอล การแจกแจงแบบเอกซโพเนนเชียล และการแจกแจงแบบพาเรโต้ ซึ่งแต่ละการแจกแจงก็จะมีแต่ละวิธีคือ วิธีโมเมนต์ คัดแปลง วิธีการประมาณค้วยภาวะน่าจะเป็นสูงสุด และวิธีกำลังสองต่ำสุดเทียม



รูปที่ 3.4 แสดงผังงานสำหรับหาค่าความคลาดเคลื่อนจากการประมาณค้วยวิธีการทั้ง 3 วิธี