หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3
วิธีดำเนินการวิจัย

ในการศึกษาครั้งนี้สิ่งที่ต้องการคือการประมาณค่าฟังก์ชันการอยู่รอด (S(t)) และเปรียบเทียบวิธีการประมาณค่าฟังก์ชันสำหรับข้อมูลที่มีค่าถูกตัดทิ้งประเภทที่ 1 ที่ได้เสนอไว้ 4 วิธีด้วยกันคือ วิธีลิมิตผลคูณ วิธีทางคณิตศาสตร์ประกันชีวิต วิธีประมาณแบบคลาสสิค และวิธีนอนพาราเมตริกแบบเบส์ ที่ได้กำหนดการแจกแจงก่อนเป็นแบบกระบวนการดีริชเลต์ และแบบกระบวนการโฮโมจีเนียสอย่างง่าย โดยจะพิจารณาความคลาดเคลื่อนสัมบูรณ์ และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ที่ได้จากการประมาณในแต่ละวิธีมาเปรียบเทียบกัน สำหรับข้อมูลที่มีการแจกแจง 3 แบบ คือ ไวบูลล์ ลอกนอร์มอล และ พาเรโต โดยมีขนาดตัวอย่าง 10,20,30,50 และ 100 เปอร์เซ็นต์ของข้อมูลที่ถูกตัดทิ้งเป็น 20%,30%,40%และ50% เวลาสิ้นสุดการเก็บรวบรวมข้อมูลที่ถูกกำหนดไว้ล่วงหน้า (Tc) มีค่าเท่ากับ น้อยกว่าและมากกว่าค่าเฉลี่ยของการแจกแจงของระยะเวลาอยู่รอด โดยให้น้อยกว่าและมากกว่าเป็นระยะห่าง 25%, 50%, 75% และ 100%

การวิจัยครั้งนี้ใช้เทคนิคการจำลองข้อมูลแบบมอนติคาร์โล (Monte Carlo Simulation Technique) สร้างสถานการต่างๆ ในการทดลองโดยใช้โปรแกรมภาษาฟอร์แทน 77 (Fortran 77) ทำการทดลองซ้ำ 1,000 ครั้ง ดังนั้นในตอนแรกของบทจะกล่าวถึงวิธีการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของแผนการทดลอง ขั้นตอนการวิจัย และโปรแกรมที่ใช้ในการวิจัยตามลำดับ

3.1 แผนการทดลอง

ในการศึกษาครั้งนี้ต้องการเปรียบเทียบการประมาณค่าฟังก์ชันการอยู่รอด(S(t)) และเปรียบเทียบวิธีการประมาณค่าฟังก์ชันสำหรับข้อมูลที่มีค่าถูกตัดทิ้งประเภทที่ 1ที่ได้เสนอไว้ 4 วิธีด้วยกันคือ วิธี ลิมิตผลคูณ วิธีทางคณิตศาสตร์ประกันชีวิต วิธีประมาณแบบคลาสสิค และวิธีนอนพาราเมตริกแบบเบส์ ที่มีการแจกแจงก่อนเป็นกระบวนการดีริชเลต์ และกระบวนการโฮโมจีเนียสอย่างง่าย สำหรับข้อมูลที่มีการแจกแจง 3 แบบคือ ไวบูลล์ ลอกนอร์มอลและพาเรโต โดยมีขนาดตัวอย่าง 10,20,30,50 และ 100 เปอร์เซ็นต์ของข้อมูลที่ถูกตัดทิ้งเป็น 20%,30%,40% และ 50% เวลาสิ้นสุดการเก็บรวบรวมข้อมูลที่ถูกกำหนดไว้ล่วงหน้า (Tc) มีค่าเป็น 2, 3, 4, 5, 6, 7, 8, 9 และ 10 รวมทั้งสิ้น 540 สถานการณ์ และทำการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์ (APE) และค่าเฉลี่ยเปอร์เซ็นต์

ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ที่ได้จากการประมาณในแต่ละวิธีมาเปรียบเทียบกันเพื่อหาวิธีที่ดีที่สุดของแต่ละสถานการณ์ โดยมีข้อกำหนดดังนี้

1) N คือจำนวนค่าสังเกต โดยที่ X_{i1} เป็นค่าสังเกตที่ได้จากการจำลองข้อมูลจากการแจกแจงแบบไวบูลล์ (Weibull Distribution) $W(c,\tau)$ เมื่อ c มีค่าเป็น 0.13 และ τ มีค่าเป็น1.25 X_{i2} เป็นค่าสังเกตที่ได้จากการจำลองข้อมูลจากการแจกแจงแบบลอกนอร์มอล (Lognormal Distribution) $LN(\mu,\delta)$ เมื่อ μ มีค่าเป็น 1.45 และ δ มีค่าเป็น 1 X_{i3} เป็นค่าสังเกตที่ได้จากการจำลองข้อมูลจากการแจกแจงแบบพาเรโต (Pareto Distribution) $P(\alpha,\lambda)$ เมื่อ α มีค่าเป็น 17.50 และ λ มีค่าเป็น 3.50

2) เวลาสิ้นสุดการเก็บรวบรวมข้อมูลที่ถูกกำหนดไว้ล่วงหน้า (Tc) มีค่าเป็น 2, 3, 4, 5, 6, 7, 8, 9 และ 10

3) ขนาดตัวอย่าง 10,20,30,50 และ 100

4) เปอร์เซ็นต์ของข้อมูลที่ถูกตัดทิ้งมี 4 ระดับ คือ 20%,30%,40%และ 50%

3.2 <u>ขั้นตอนในการศึกษาวิจัย</u>

1) จำลองข้อมูลของระยะเวลาจากการแจกแจงแบบไวบูลล์ การแจกแจงแบบลอกนอร์มอล และการแจกแจงแบบพาเรโต โดยรายละเอียดพร้อมทั้งโปรแกรมที่ใช้ในการจำลองข้อมูลได้อธิบายไว้ในภาคผนวก ก.

2) กำหนดข้อมูลให้มีลักษณะเป็นข้อมูลที่ถูกตัดทิ้งทางขวาประเภทที่ 1 โดยกำหนด $T_i = T_c$ เมื่อ $T_i > T_c$ ในการวิจัยครั้งนี้จะกำหนดเวลาสิ้นสุดการเก็บข้อมูลไว้ล่วงหน้า ดังนั้น ถ้าข้อมูลที่ได้จากการจำลองข้อมูลมีค่าเกินกว่าเวลาที่ถูกกำหนดไว้ล่วงหน้าให้ถือว่าข้อมูลนั้นเป็นตัวที่ถูกตัดทิ้ง

3) นำข้อมูลที่ได้จากการดำเนินการตามข้อ 1 และ 2 มาเรียงลำดับจากน้อยไปมาก นั่นคือจะนำข้อมูลทั้งที่ถูกตัดทิ้งและไม่ถูกตัดทิ้งมาเรียงลำดับเพื่อใช้ในการคำนวณขั้นต่อไป

4) คำนวณหาค่าฟังก์ชันการอยู่รอด ด้วยวิธีประมาณค่าทั้ง 4 วิธี โดยแต่ละวิธีจะประมาณค่าดังต่อไปนี้

4.1 วิธีลิมิตผลคูณ

1. นำข้อมูลมาเรียงลำดับค่าจากน้อยไปมาก เพื่อหาลำดับที่ของข้อมูลสำหรับข้อมูลที่ไม่มีค่าสังเกตซ้ำ หรือนับจำนวนตัวอย่างที่มีค่าอยู่ในแต่ละช่วงที่ต้องการประมาณค่าเมื่อข้อมูลมีค่าสังเกตซ้ำ

2. หาค่าประมาณฟังก์ชันการอยู่รอด $\hat{S}(t)$ จากสมการ (2.2)

จะได้ว่า $\hat{S}(t_i) = \prod_{i=1}^{i} \left(\frac{n_i - \delta_i}{n_i} \right)$

4.2 วิธีทางคณิตศาสตร์ ประกันชีวิต

1. นำข้อมูลมาเรียงลำดับค่าจากน้อยไปมาก เพื่อหาลำดับที่ของข้อมูลสำหรับข้อมูลที่ไม่มีค่าสังเกตซ้ำ หรือนับจำนวนตัวอย่างที่มีค่าอยู่ในแต่ละช่วงที่ต้องการประมาณค่าเมื่อข้อมูลมีค่าสังเกตซ้ำ

2. หาค่าประมาณอัตราค่ากลาง $\hat{m}_i$

3. หาค่าประมาณฟังก์ชันการอยู่รอด $\hat{S}(t)$ จากสมการ (2.3)

จะได้ว่า $\hat{S}(t_i) = \prod_{i=1}^{i} e^{-\hat{m}_i}$

4.3 วิธีประมาณแบบคลาสสิค

1. นำข้อมูลมาเรียงลำดับค่าจากน้อยไปมาก เพื่อนับจำนวนตัวอย่างที่มีค่ามากกว่าหรือเท่ากับแต่ละช่วงที่ต้องการประมาณค่า

2. หาค่าประมาณของระยะเวลาของกรมธรรม์ที่อยู่ในแต่ละช่วง $\hat{E}_i$

3. หาค่าประมาณฟังก์ชันการอยู่รอด $\hat{S}(t)$ จากสมการ (2.4)

จะได้ว่า $\hat{S}(t_i) = \prod_{i=1}^{i} \frac{D_i}{E_i}$

4.4 วิธีวิธีนอนพาราเมตริกแบบเบส์

4.4.1 เมื่อกำหนดการแจกแจงเป็นกระบวนการดีริชเลต์

โดยการกำหนดการแจกแจงคาดเดาก่อนการทดลอง (Prior guess) ของฟังก์ชันการอยู่รอด $(S(t))$ สำหรับงานวิจัยครั้งนี้ได้กำหนดการแจกแจงคาดเดาก่อนการทดลอง $(S_0(t))$ มีการแจกแจงแบบเอกซโพเนนเชียล (Exponential Distribution) ดังนี้

$$S_0(t) = e^{-t\theta}$$

$$\hat{\theta} = \frac{r}{\sum_{i=1}^{n} t_i}$$

เมื่อ r เป็นจำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง

3. ประมาณค่า $\alpha(t)$ โดยกำหนดให้ $\alpha(R) = 1$ จากสมการที่ (2.24)

จะได้ว่า $\alpha(t) = S_0(t).\alpha(R)$

$= e^{-t\hat{\theta}}$

4. หาค่าประมาณฟังก์ชันการอยู่รอด $\hat{S}(t)$ จากสมการ (2.23)

จะได้ว่า $\hat{S}(t) = \frac{\alpha(R) - \alpha(t) + h_{j(t)}}{\alpha(R) + n} \prod_{i=1}^{j(t)} \left(\frac{\alpha(R) - \alpha(u_i) + h_i + \lambda_i}{\alpha(R) - \alpha(u_i) + h_i} \right)$

4.4.2 เมื่อกำหนดการแจกแจงเป็นกระบวนการโฮโมจีเนียสอย่างง่าย

โดยการกำหนดการแจกแจงคาดเดาก่อนการทดลอง (Prior guess) ของฟังก์ชันการอยู่รอด $(S(t))$ สำหรับงานวิจัยครั้งนี้ได้กำหนดการแจกแจงคาดเดาก่อนการทดลอง $(S_0(t))$ มีการแจกแจงแบบเอกซโพเนนเชียล (Exponential Distribution) ดังนี้

$$S_0(t) = e^{-t\theta}$$

1. นำข้อมูลมาเรียงลำดับค่าจากน้อยไปมาก เพื่อหาลำดับที่ของข้อมูลสำหรับข้อมูลที่ไม่มีค่าสังเกตซ้ำ หรือนับจำนวนตัวอย่างที่มีค่าอยู่ในแต่ละช่วงที่ต้องการประมาณค่าเมื่อข้อมูลมีค่าสังเกตซ้ำ

2. ประมาณค่า θ ของการแจกแจงคาดเดาก่อนการทดลอง ซึ่งจะประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) โดยจะประมาณค่าจากสมการดังนี้

$$\hat{\theta} = \frac{r}{\sum_{i=1}^{n} t_i}$$

เมื่อ r เป็นจำนวนค่าสังเกตที่ไม่ถูกตัดทิ้ง

3. ประมาณค่า $\gamma(t)$ โดยกำหนดให้ $\tau = 1$ จากสมการที่ (2.32)

5) คำนวณหาค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์จากการประมาณค่าฟังก์ชันการอยู่รอด และนำมาเปรียบเทียบเพื่อที่จะหาวิธีการประมาณที่ดีที่สุด โดยในการทดลองได้ทำการจำลองข้อมูลซ้ำกันจำนวน 1,000 ครั้ง ในแต่ละสถานการณ์ที่กำหนด เมื่อให้ i แทนรอบที่ทำซ้ำ โดยที่ i = 1,2,3,...,1000 ดังนั้น

$$\hat{S}(t) = \frac{\sum_{i=1}^{1000} \hat{S}(t)^{(i)}}{1000}$$

โดยที่ $\hat{S}(t)^{(i)}$ เป็นค่าประมาณ $S(t)$ สำหรับการทำซ้ำรอบที่ i

และคำนวณค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (APE) และค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) ของแต่ละวิธีตามสมการต่อไปนี้

$$APE_t = \left| \frac{S(t) - \hat{S}(t)}{S(t)} \right| \times 100\%$$

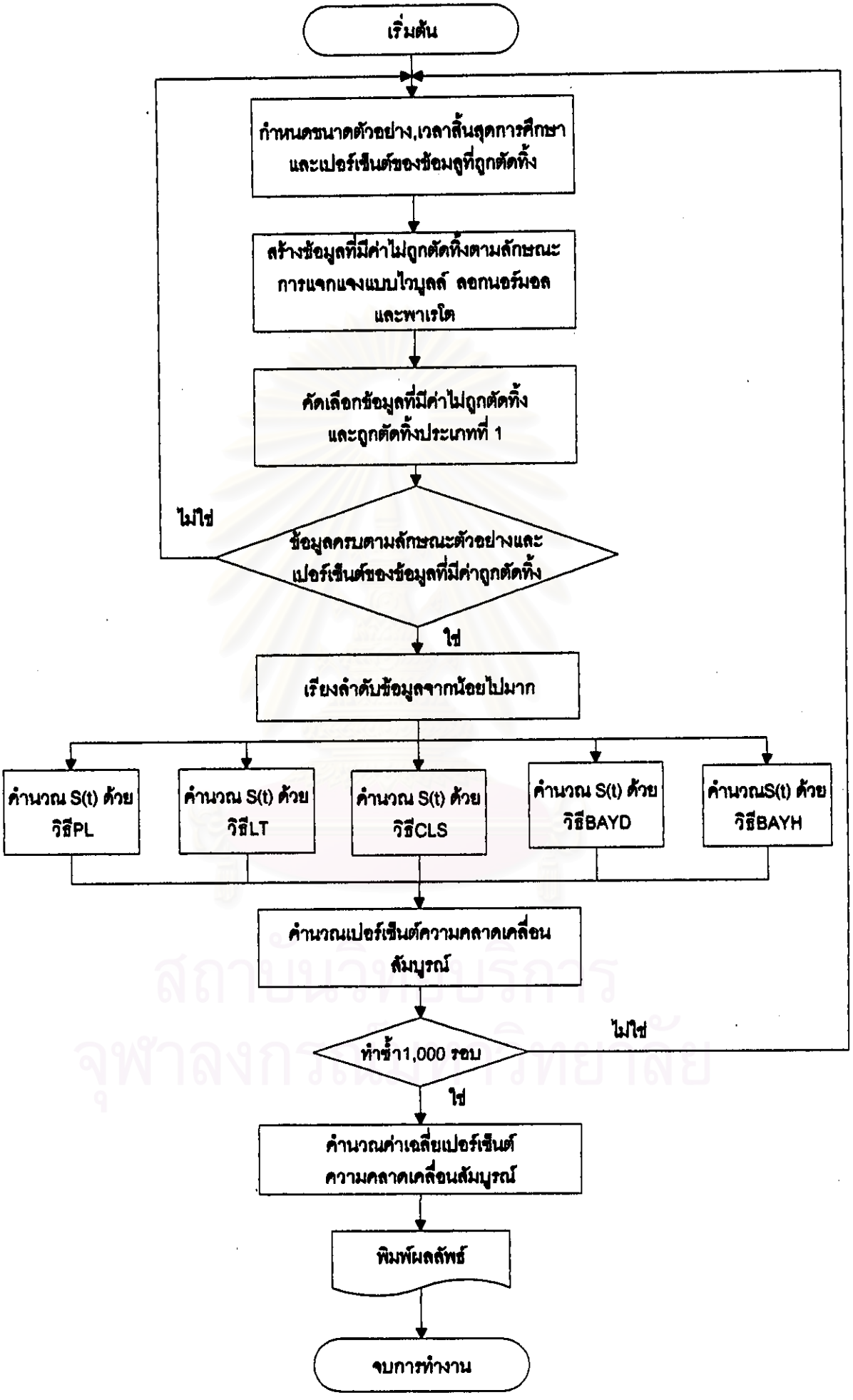
เมื่อ $S(t)$ เป็นค่าฟังก์ชันการอยู่รอด และ $\hat{S}(t)$ เป็นค่าประมาณฟังก์ชันการอยู่รอดด้วยวิธีการประมาณต่างๆ

$$MAPE = \frac{\sum_{t=1}^{T_n} APE_t}{NT}$$

เมื่อ NT เป็นจำนวนจุดเวลาตั้งแต่ 1.0 ถึงเวลาที่ถูกกำหนดไว้ล่วงหน้า โดยเพิ่มทีละ 1.0 แล้วนำค่า MAPE ของการประมาณค่าในแต่ละวิธีมาเปรียบเทียบกัน โดยวิธีใดให้ค่า MAPE ต่ำที่สุดจะเป็นวิธีการประมาณที่ดีที่สุดของแต่ละสถานการณ์

3.3 <u>โปรแกรมที่ใช้ในการศึกษาวิจัย</u>

โปรแกรมที่ใช้ในการศึกษาวิจัยครั้งนี้ เขียนด้วยโปรแกรมภาษาฟอร์แทน77 (FORTRAN 77) โดยใช้กับเครื่อง AMDAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลองลักษณะของการทำงานแต่ละโปรแกรมจะเหมือนกัน สำหรับรายละเอียดของผังงานแสดงการทำงานของโปรแกรม และโปรแกรมที่ใช้ในการประมาณค่า S(t) ด้วยวิธีต่างๆ แสดงไว้ในภาคผนวก ข.



รูปที่ 3.1 แสดงขั้นตอนการวิจัย