หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

ในการวิจัยทั่ว ๆไป มักจะมีการกำหนดสมมุติฐานเพื่อแสดงความเชื่อของผู้วิจัยเกี่ยวกับเรื่องที่สนใจศึกษา การพิสูจน์ความเชื่อดังกล่าวนี้ผู้วิจัยจะต้องเลือกตัวสถิติทดสอบให้เหมาะสมกับข้อมูล และปัญหาที่ต้องการทดสอบจึงจะได้ผลสรุปที่เชื่อถือได้ แต่ตัวสถิติที่เลือกมาใช้ในการทดสอบนี้ ส่วนใหญ่จะมีข้อจำกัดเกี่ยวกับการแจกแจงของประชากรที่นำมาทดสอบและการกระจายของข้อมูลที่สนใจศึกษาในประชากรนั้น ๆ

การแจกแจงเอกซโปเนนเชียล (Exponential Distribution)พบมากในข้อมูลที่เกี่ยวกับ เวลารอคอย (waiting time)เช่น การแจกแจงของเวลารอคอยในการมารับบริการของลูกค้า การแจกแจงของเวลาที่ให้บริการของร้านค้า ข้อมูลเกี่ยวกับอายุ (life data) เช่น อัตราการเสีย (hazard rate) หรือ อัตราการตาย (mortality-rate) ในทางทฤษฎีได้พิสูจน์แล้วว่าข้อมูลดังกล่าวมีการแจกแจงแบบเอกซโปเนนเชียลจริง แต่ในทางปฏิบัตินั้น ก่อนจะนำเอาข้อมูลไปวิเคราะห์ทางสถิติ ต้องทำการทดสอบการแจกแจงของข้อมูลก่อนว่ามีการแจกแจงเอกซโปเนนเชียลจริงหรือไม่ ทั้งนี้เนื่องจากถ้าข้อมูลไม่เป็นไปตามข้อกำหนด ผลที่ได้จากการวิเคราะห์ก็จะไม่ถูกต้องและขาดความเชื่อถือ

อนึ่งถ้าหากข้อมูลที่นำมาวิเคราะห์มีการขาดหาย (censored data) เกิดขึ้นซึ่งอาจส่งผลต่อการแจกแจงของข้อมูลที่เหลืออยู่ และนำข้อมูลเหล่านั้นไปวิเคราะห์ เพื่อหาค่าประมาณของพารามิเตอร์หรือการทดสอบสมมุติฐานทางสถิติ อาจทำให้ผลการวิเคราะห์ที่ได้ขาดความเชื่อถือ เช่น การประมาณค่าที่สูงกว่าความเป็นจริง หรือต่ำกว่าความเป็นจริง หรืออำนาจการทดสอบต่ำ ดังนั้นจึงจำเป็นต้องทำการทดสอบการแจกแจงของข้อมูลที่มีลักษณะดังกล่าวนี้

ดังนั้นงานวิจัยเรื่องนี้ผู้วิจัยจึงได้มุ่งประเด็นความสนใจไปที่ การทดสอบการแจกแจงเอกซโปเนนเชียลกรณีที่ข้อมูลขาดหาย นักสถิติหลายท่านได้สนใจพัฒนาตัวสถิติทดสอบการแจกแจงข้อมูลแบบเอกซโปเนนเชียลขึ้นมาหลายวิธี ซึ่งสามารถแบ่งได้เป็น 3 ประเภท คือ

1. การทดสอบโดยใช้การถดถอยของตัวสถิติลำดับบนพารามิเตอร์บางตัวของตัวสถิติลำดับนั้น (Regression of the order statistics on some parameter of the distribution of the order statistics) ได้แก่ Regression Test, Shapiro and Wilk Test และ Probability Plot Correlation Coefficience

2. ทดสอบโดยใช้ฟังก์การแจกแจงเอมไพริคอล (Empirical Distribution Function) ได้แก่ Kolmogorov Smirnov Test, Cremer-Von-Mises Test ฯลฯ

3. ทดสอบโดยใช้ผลต่างของสถิติลำดับที่อยู่ติดกันโดยถ่วงน้ำหนัก (Weighted-Difference Sucessive Order Statistics) ได้แก่ Gnedenko F Test และ Bivariate F Test.

แต่สำหรับกรณีที่ข้อมูลขาดหาย ตัวสถิติที่ใช้ในการทดสอบการแจกแจงเอกซโปเนนเชียล ที่กล่าวถึงได้แก่ Regression test (Z), Gnedenko F test (F), Bivariate F Test และ Shapiro and Wilk Test

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบเอกซโปเนนเชียล 3 ตัวคือ

1. Shapiro and Wilk Test (W)
2. Regression Test (Z)
3. Bivariate F Test (BF)

เมื่อข้อมูลที่นำมาวิเคราะห์มีการขาดหาย โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

1.3 สมมุติฐานของการวิจัย

โดยทั่ว ๆ ไป ตัวสถิติทดสอบ W จะให้อำนาจการทดสอบสูงสุด เมื่อเทียบกับตัวสถิติทดสอบ Z และตัวสถิติทดสอบ BF

1.4 ขอบเขตของการวิจัย

1.4.1 ศึกษาความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจของการทดสอบ โดยประชากรมีการแจกแจงดังนี้

ก. การแจกแจงเอกซโปเนนเชียล (Exponential Distribution)

ฟังก์ชั่นความหนาแน่นอยู่ในรูป

$$f(x) = \begin{cases} \exp(-(x-\alpha_0)/\beta) & ;\beta > 0 \\ & x > \alpha_0 \\ 0 & ;\text{อื่น ๆ} \end{cases}$$

เมื่อ β เป็น Scale parameter ซึ่งแสดงขนาดของการแจกแจงเท่ากับ 1

ข. การแจกแจงแบบแกมม่า (Gamma distribution)

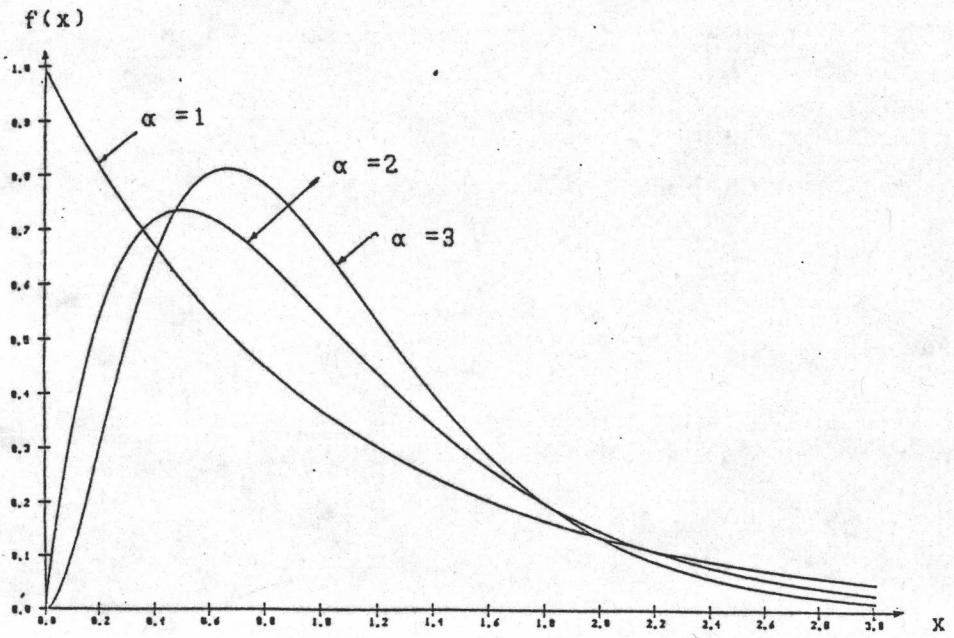
ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{(x^{\alpha-1}\exp(-x/\beta))}{\beta^{\alpha}\cdot\Gamma(\alpha)} & ;\ x > 0 \\ & ;\ \alpha > 0\ ,\beta > 0 \\ 0 & \text{อื่นๆ} \end{cases}$$

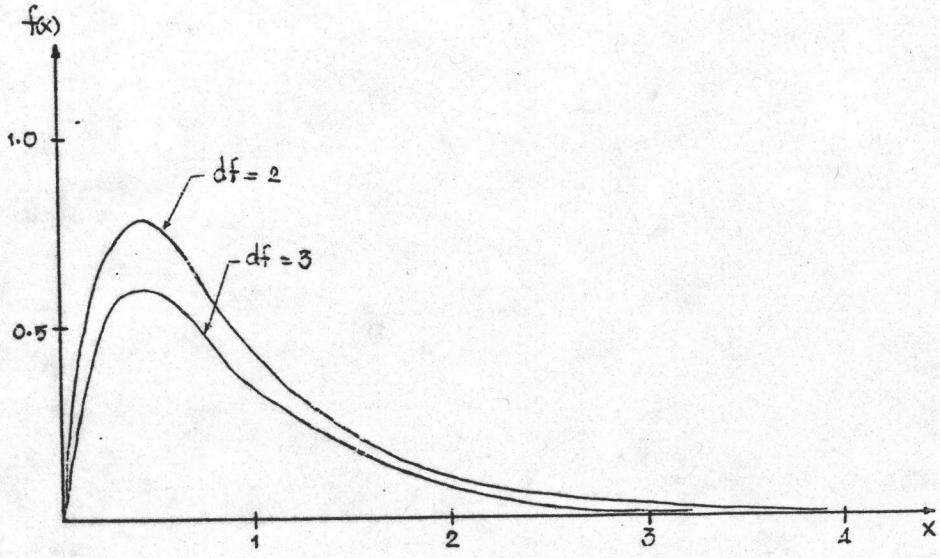
เมื่อ β เป็น Scale parameter ซึ่งแสดงขนาดของการแจกแจงเท่ากับ 1

α เป็น Shape parameter ซึ่งแสดงรูปร่างของการแจกแจงเท่ากับ 1,2 และ 3

เมื่อ $\beta=2$, $\alpha=n/2$ การแจกแจงแบบแกมม่า คือการแจกแจงแบบไคสแควร์ ณ ระดับความเป็นอิสระ n



รูปที่ 1.1 แสดงการแจกแจงแบบแกมม่าที่มี Shape parameter(α) เท่ากับ 1,2,3 และ Scale parameter เท่ากับ 1



รูปที่ 1.2 แสดงการแจกแจงแบบไคสแควร์ที่มีองศาแห่งความเป็นอิสระ(df) เท่ากับ 2 และ 3 ตามลำดับ

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

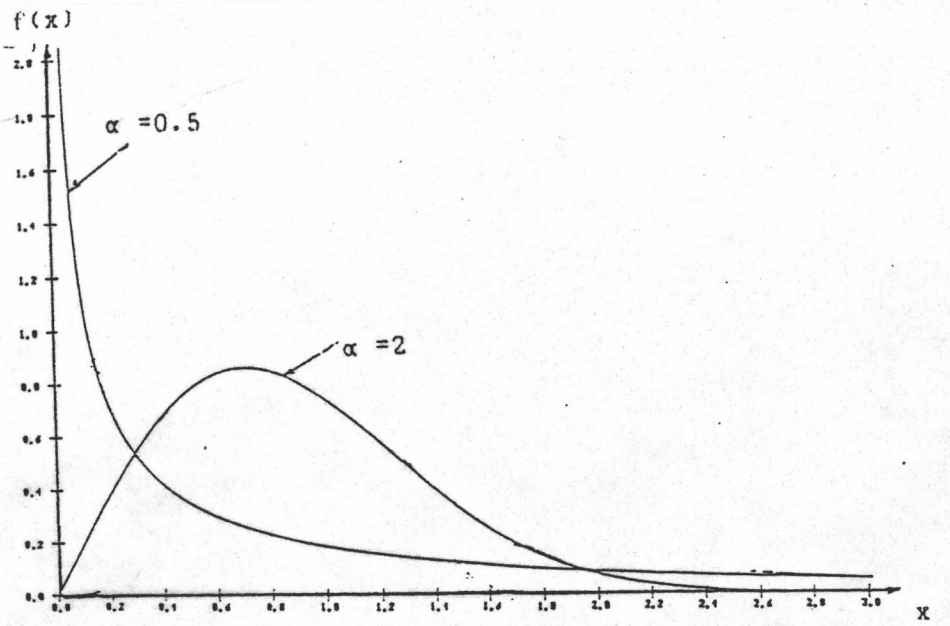
ค. การแจกแจงแบบไวบูลล์ (weibull distribution)

ฟังก์ชั่นความหนาแน่นอยู่ในรูปของ

$$f(x)=\begin{cases}\dfrac{\alpha \cdot x^{\alpha-1} \cdot \exp[-(x/\beta)^{\alpha}]}{\beta^{\alpha}} & ;0< x <\infty \\ & \alpha > 0 \\ 0 & ; \text{อื่น ๆ}\end{cases}$$

เมื่อ Scale parameter (β) เท่ากับ 1

shape parameter (α) เท่ากับ 0.5,2.0



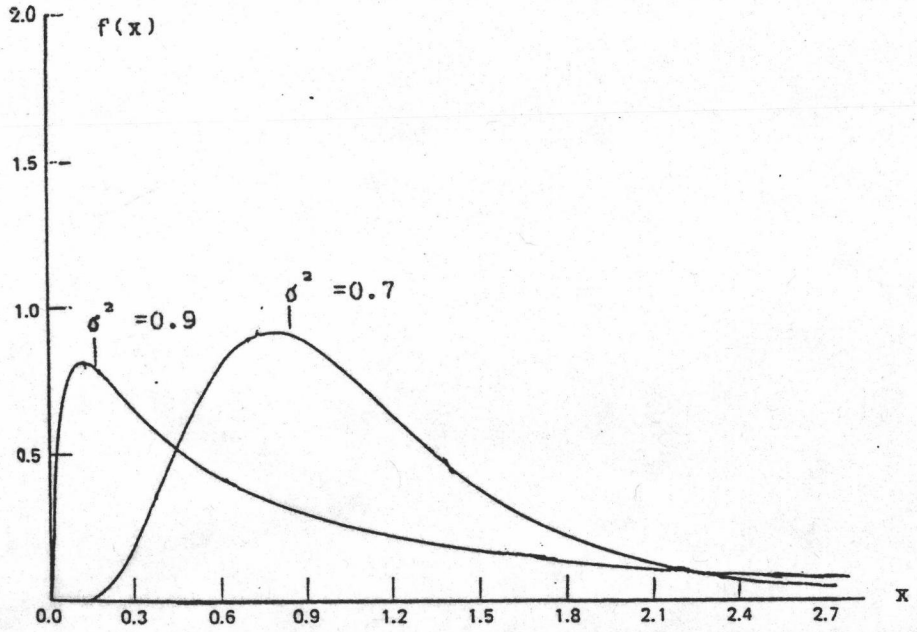
รูปที่ 1.3 แสดงการแจกแจงแบบไวบูลล์ที่มี scale parameterเท่ากับ 1 และ shape psarameter เท่ากับ 0.5 และ 2.0

ง. <u>การแจกแจงแบบลอกนอร์มอล</u> (lognormal distribution)

ฟังก์ชั่นความหนาแน่นอยู่ในรูป

$$f(x) = \begin{cases} \dfrac{\exp[-(\ln x-\mu)^2/2}{x\cdot\sqrt{2\pi\sigma}} & ;x > 0 \\ & \sigma > 0 \\ & -\alpha < \mu < \alpha \\ 0 & ; \text{อื่นๆ} \end{cases}$$

เมื่อ ค่าเฉลี่ย (μ) = 0 และความแปรปรวน (σ^2)= 0.7,0.9



รูปที่ 1.4 แสดงการแจกแจงแบบลอกนอร์มอล เมื่อพารามิเตอร์ μ = 0 และ σ^2 = 0.7 และ 0.9 ตามลำดับ

1.4.2 กำหนดขนาดของตัวอย่างที่ใช้ศึกษา เท่ากับ 10,20,30,50 และ 100

1.4.3 กำหนดระดับนัยสำคัญ เท่ากับ 0.05 และ 0.10

1.4.4 กรณีข้อมูลขาดหายในการวิเคราะห์ จะพิจารณาข้อมูลขาดหายทางซ้าย (Left Censored data = r1) และจำนวนข้อมูลขาดหายทางขวา (Right Censored data = r2) โดยพิจารณาจากกรณีต่อไปนี้

<u>กรณีที่ 1</u> ข้อมูลขาดหายทางซ้าย

r1 = 10% ,20%,30% ของจำนวนข้อมูลทั้งหมด

r2 = 0

<u>กรณีที่ 2</u> ข้อมูลขาดหายทางขวา

r1 = 0

r2 = 10% ,20%,30% ของจำนวนข้อมูลทั้งหมด

<u>กรณีที่ 3</u> ข้อมูลขาดหายทางซ้ายและขวาไม่เท่ากัน

กำหนดให้

r1 = 10% r2 = 20%

r1 = 20% r2 = 10%

<u>กรณีที่ 4</u> ข้อมูลขาดหายทางซ้ายและขวาเท่ากัน

กำหนดให้

r1 = 0% r2 = 0%

r1 = 10% r2 =10%

1.5 <u>ประโยชน์ที่คาดว่าจะได้รับ</u>

เพื่อเป็นแนวทางให้ผู้วิจัยเลือกใช้ตัวสถิติสำหรับการทดสอบที่เหมาะสม สำหรับการแจกแจงเอกซโปเนนเชียล เมื่อมีข้อมูลขาดหาย

1.6 คำจำกัดความ

ก. ความคลาดเคลื่อนประเภทที่ 1 (type I error) หมายถึงความคลาดเคลื่อนซึ่งเกิดจากการปฏิเสธสมมุติฐานว่าง (H_0) เมื่อสมมุติฐานว่างจริง

ข. อำนาจการทดสอบ (power of the test) หมายถึงความน่าจะเป็นที่จะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อสมมุติฐานว่างไม่จริง

ค. ความคลาดเคลื่อนประเภทที่ 2 (type II error) หมายถึงความคลาดเคลื่อนที่เกิดการยอมรับสมมุติฐานว่าง (H_0) เมื่อสมมุติฐานว่างไม่จริง

1.7 เกณฑ์ที่ใช้ในการเปรียบเทียบ

การเปรียบเทียบตัวสถิติที่ใช้ทดสอบการแจกแจงเอกซโปเนนเชียล พิจารณาเป็น 2 ขั้นตอนดังนี้

1.7.1 พิจารณาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 โดยพิจารณาจากความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดลองในแต่ละสถานการณ์ ซึ่งผู้วิจัยจะใช้เกณฑ์ของ Bradley ดังรายละเอียดต่อไปนี้

ถ้าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองอยู่ในช่วง (0.025,0.075) และ (0.051,0.150) ณ ระดับนัยสำคัญ 0.05 และ 0.10 ตามลำดับ จะถือว่าการทดสอบนั้น สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้

1.7.2 พิจารณาอำนาจการทดสอบของตัวสถิติทดสอบแต่ละตัวโดยพิจารณาเฉพาะตัวสถิติที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เท่านั้น