บทที่ 2
ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

วัตถุประสงค์หลักในการทำวิจัยนี้ เพื่อทำการศึกษาและเปรียบเทียบค่าโดยทำการศึกษาและเปรียบเทียบการแจกแจงทั้ง 2 แบบ คือ การแจกแจงแบบปกติ และ การแจกแจงแบบที ภายใต้สถานการณ์ต่าง ๆ ดังที่ได้กล่าวไว้ในขอบเขตการวิจัยในบทที่ 1

ในบทนี้จะกล่าวถึงรายละเอียดของทฤษฎีและการแจกแจงที่เกี่ยวข้องกับการวิจัย โดยมีทฤษฎี ที่เกี่ยวข้องดังนี้

2.1 การแจกแจงแบบปกติ(กัลยา วานิชย์บัญชา, 2542)

นิยาม การแจกแจงแบบปกติหนึ่งตัวแปร (Univariate Normal Distribution)

ตัวแปรสุ่ม X มีการแจกแจงปกติ หรือ การแจกแจงแบบเกาส์ ด้วยพารามิเตอร์ μ และ σ^2 เขียนแทนด้วย $X \sim N\left(\mu, \sigma^2\right)$ ถ้า X มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function) เป็น

$$f(x) = f\left(x; \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , -\infty < x < \infty$$

และ $-\infty < \mu < \infty, 0 < \sigma^2 < \infty$

โดยที่ $\pi = 3.14159\ldots$ และ $e = 2.71828\ldots$

เมื่อ μ เป็นพารามิเตอร์แสดงตำแหน่งของการแจกแจง

σ^2 เป็นพารามิเตอร์แสดงขนาดของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบปกติ จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \mu$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \sigma^2$$

3. สัมประสิทธิ์ความเบ้

$$\gamma_1 = 0$$

4. สัมประสิทธิ์ความโด่ง

$$\gamma_3 = 0$$

นิยาม การแจกแจงแบบปกติสองตัวแปร (bivariate normal distribution)

ตัวแปรสุ่ม $(X,Y)^T$ มีการแจกแจงปกติสองตัวแปร ด้วยพารามิเตอร์ μ_1,μ_2,σ_1^2 และ σ_2^2 เขียนแทนด้วยสัญลักษณ์ $(X,Y)^T \sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ ถ้า $(X,Y)^T$ มีฟังก์ชันความหนาแน่นร่วมเป็น

$$f(x,y)=f(x,y;\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{q}{2}}\ ,-\infty<x<\infty,-\infty<y<\infty$$

เมื่อ $$q=\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

และ $-\infty<\mu_1<\infty,-\infty<\mu_2<\infty,0<\sigma_1^2<\infty,0<\sigma_2^2<\infty,-1<\rho<1$

2.2 การแจกแจงแบบที

นิยาม การแจกแจงแบบทีหนึ่งตัวแปร (Univariate t Distribution)

ตัวแปรสุ่ม X มีการแจกแจงความน่าจะเป็นแบบที ด้วยระดับขั้นความเสรี v เขียนแทนด้วย $X \sim t_{(v)}$ ถ้า X มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function) เป็น

$$f(x)=f(x;v)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\cdot\frac{1}{\left(1+\frac{x^2}{v}\right)^{(v+1)/2}}\ ,-\infty<x<\infty$$

และ $v=1,2,3,\ldots$

เมื่อ v เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบที จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

   $$E(X)=0,\qquad v>1$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

   $$Var(X)=\frac{v}{v-2},\qquad v>2$$

3. สัมประสิทธิ์ความเบ้

   $$\gamma_1=0,\qquad v>3$$

4. สัมประสิทธิ์ความโด่ง

   $$\gamma_2=\frac{3(v-2)}{v-4},\qquad v>4$$

นิยาม การแจกแจงแบบทีสองตัวแปร (Bivariate t Distribution)

ตัวแปรสุ่ม $(X,Y)^T$ มีการแจกแจงความน่าจะเป็นแบบทีสองตัวแปร ด้วยระดับขั้นความเสรี v ถ้า $(X,Y)^T$ มีฟังก์ชันความหนาแน่นร่วมเป็น

$$f(x,y)=f(x,y;v,\rho)=\frac{\Gamma\left(\frac{v+2}{2}\right)}{(v\pi)\Gamma\left(\frac{v}{2}\right)\sqrt{1-\rho^2}}\cdot\frac{1}{\left(1+\frac{x^2-2\rho xy+y^2}{v\left(1-\rho^2\right)}\right)^{(v+2)/2}}$$

เมื่อ $-\infty<x<\infty,-\infty<y<\infty$ และ $-1<\rho<1$, $v=1,2,3,\ldots$

นิยาม การแจกแจงแบบทีแบบปรับค่าความแปรปรวนเป็น 1 (t-adjust Distribution)

ตัวแปรสุ่ม X มีการแจกแจงความน่าจะเป็นแบบที ด้วยระดับขั้นความเสรี v เขียนแทนด้วย $X\sim t_{(v)}$ จะทำการปรับการแจกแจงของทีโดยกำหนดให้ $Y=\sqrt{\frac{\nu-2}{\nu}}X$ ซึ่งจะทำให้ Y มีฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function) เป็น

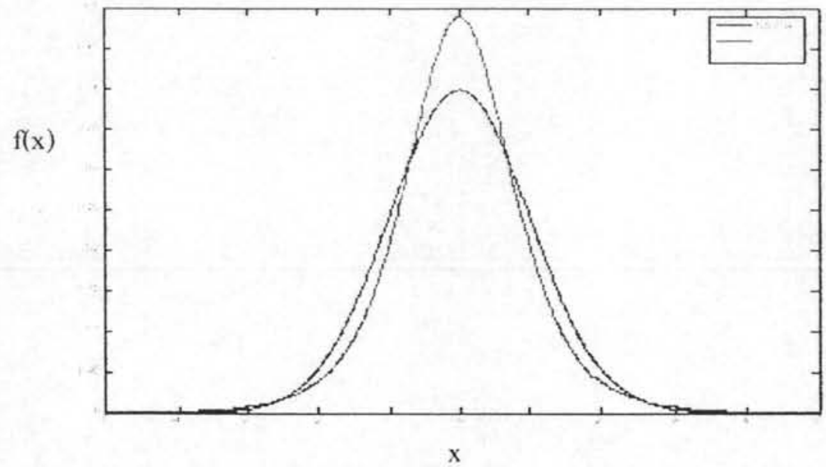
$$f(y)=f(y;v)=\sqrt{\frac{\nu}{\nu-2}}\cdot\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\cdot\frac{1}{\left(1+\frac{y^2}{v-2}\right)^{(v+1)/2}},-\infty<y<\infty$$

และ $v=1,2,3,\ldots$

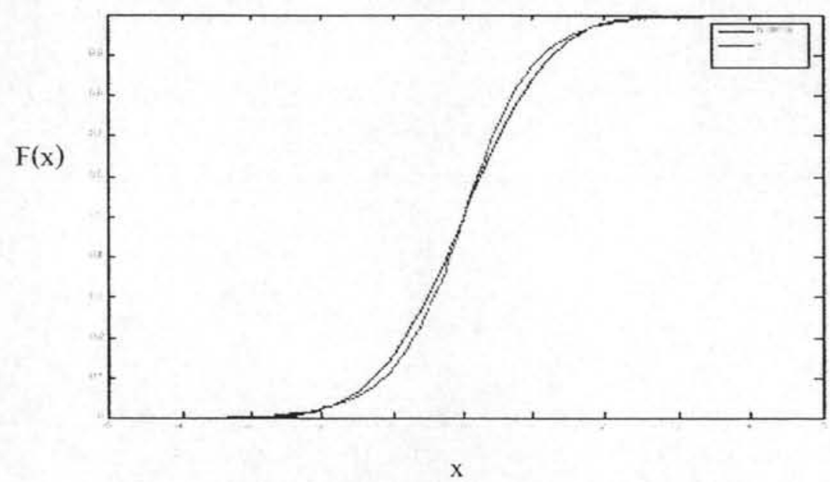
เมื่อ v เป็นพารามิเตอร์แสดงรูปร่างของการแจกแจง

จะได้ว่าค่าความแปรปรวนของการแจกแจงแบบทีที่ปรับค่าแล้วนั้นจะมีค่าเท่ากับ 1

สามารถแสดงรูปแบบการแจกแจงที่ต่าง ๆ กันได้ดังรูปภาพต่อไปนี้



ภาพ แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติและ การแจกแจงแบบที ตามลำดับ เมื่อมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1



ภาพ แสดงฟังก์ชันความหนาแน่นสะสมของการแจกแจงแบบปกติและ การแจกแจงแบบที เมื่อมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1

2.3 การจำลองแบบด้วยเทคนิคมอนติคาร์โล (เสกสรร เกียรติสุไพบูลย์, 2548)

เทคนิคมอนติคาร์โล เป็นการทดลองโดยใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหาที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น เพราะเลขสุ่มมีประโยชน์หลายประการ คือ

1. ทำให้การเลือกตัวอย่างไม่มีความเอนเอียงในการสำรวจ หรือทดลองในเรื่องนั้น ๆ ทั้งนี้เพราะเลขสุ่มมาจากแนวคิดเกี่ยวกับการคำนวณความน่าจะเป็น
2. เลขสุ่มจะทำให้ได้มาซึ่งรูปแบบต่าง ๆ หรือวิธีการที่สลับซับซ้อน โดยการสร้างสถานการณ์จำลอง (Simulation)

3. การใช้เลขสุ่มอาจทำเพื่อศึกษาคุณสมบัติทางทฤษฎีของกระบวนการทางสถิติ ที่มีความสำคัญสำหรับการประมาณค่าตลอดจนนำไปสู่คำอธิบายเกี่ยวกับอำนาจการทดสอบทางสถิติ (Power of statistic test)
4. เพื่อหาคำตอบในปัญหาทางคณิตศาสตร์ โดยจะพิจารณาจากการแจกแจงความน่าจะเป็นของปัญหานั้น ๆ

หลักการของเทคนิคมอนติคาร์โล คือ การนำเลขสุ่มมาประยุกต์ใช้ในการแก้ปัญหาต่าง ๆ ที่สนใจศึกษาถึงผลสรุปของปัญหานั้น ๆ โดยมีขั้นตอนสำคัญ 3 ขั้นตอนดังนี้

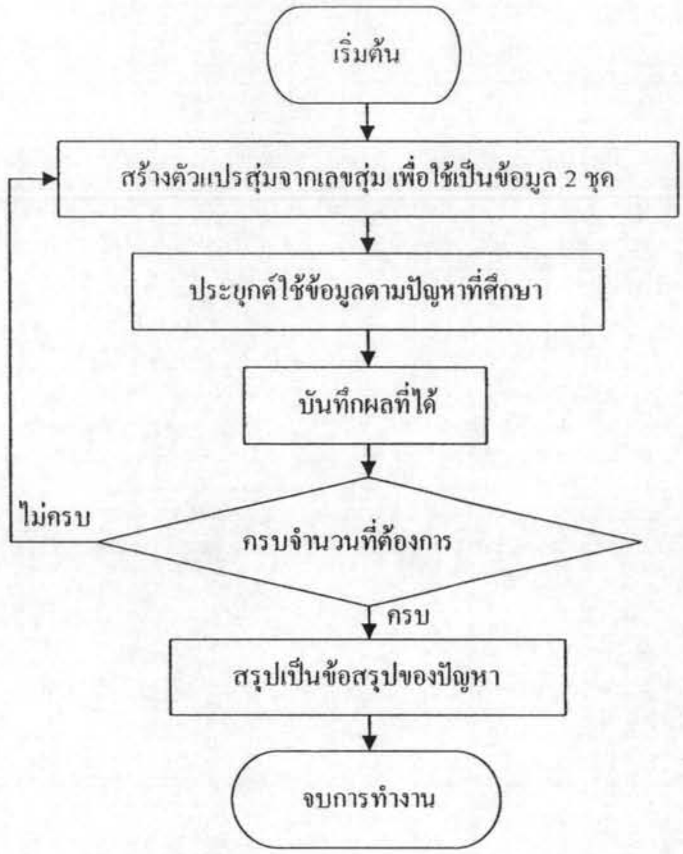
ขั้นตอนที่ 1 การสร้างเลขสุ่ม (Generate random number) การสร้างเลขสุ่มจะกำหนดให้มีการแจกแจงแบบยูนิฟอร์มในช่วง [0,1] และเป็นอิสระซึ่งกันและกัน จากนั้นนำเลขสุ่มนี้ไปสร้างตัวแปรตามลักษณะการแจกแจงที่ต้องการในปัญหาที่ศึกษา เพื่อเป็นข้อมูลปัญหานั้น ๆ

ขั้นตอนที่ 2 การประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่ม ขั้นตอนนี้ขึ้นอยู่กับปัญหาที่ศึกษา ซึ่งเป็นขั้นตอนที่นำเลขสุ่มมาใช้ในการหาค่าต่าง ๆ ตามปัญหาที่ต้องการตามสูตรการคำนวณในปัญหาที่ศึกษา

ขั้นตอนที่ 3 การทดลอง เมื่อประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่มแล้ว ขั้นตอนต่อไป คือ การทำวิธีการนั้นซ้ำ ๆ กัน (Replication) จำนวนหลายครั้ง โดยถือว่าการทำซ้ำ ๆ กันนั้นเป็นวิธีการเก็บรวบรวมข้อมูลให้มีจำนวนมากเพื่อลดความไม่แน่นอนของคำตอบ

จากหลักการของเทคนิคมอนติคาร์โล การใช้เลขสุ่มเพื่อเป็นพื้นฐานในการหาคำตอบของปัญหา เป็นวิธีการที่จะนำไปสู่แนวคิดในทางทฤษฎีที่เกี่ยวข้องกับการคำนวณ โดยเฉพาะทฤษฎีความน่าจะเป็นที่จะนำไปสู่การอ้างอิงผลสรุปในสถานการณ์ของข้อมูลจริง เพราะไม่มีผลกระทบจากเรื่องอื่น ๆ เข้ามาเกี่ยวข้องในการทดลองเมื่อทำซ้ำ ๆ กันเป็นจำนวนมากแล้ว ความคลาดเคลื่อนอย่างสุ่มที่เกิดขึ้นในการวิเคราะห์หาค่าต่าง ๆ ในแต่ละครั้งจะหมดไป (Counter Balance) จากขั้นตอนเทคนิคมอนติคาร์โล สามารถเขียนแผนผังได้ดังนี้

แผนผังที่ 2.1 แผนผังเทคนิคมอนติคาร์โล



2.4 ผลการประสาน (ธีระพร วีระถาวร, 2539)

ปัญหาบางปัญหาในทางสถิติอาจต้องการทราบการแจกแจงของตัวสถิติที่จะนำไปใช้งาน ทำให้ต้องหาวิธีที่จะทำให้สามารถทราบการแจกแจงตัวสถิติที่ได้จากผลบวกของตัวแปรสุ่มที่ทราบการแจกแจงอยู่แล้ว ซึ่งเทคนิคที่จะนำมาใช้นี้เรียกว่า ผลการประสาน เพื่อนำมาหาการแจกแจงของตัวแปรสุ่มที่ได้จากผลบวกของตัวแปรสุ่มอื่นๆ โดยสำหรับตัวแปรสุ่มประเภทต่อเนื่องมีรายละเอียดดังนี้

ให้ X และ Y เป็นตัวแปรสุ่มต่อเนื่องมี f(x,y) เป็นฟังก์ชันความหนาแน่นของ X+Y มีฟังก์ชันการแจกแจงสะสม

$$F_{X+Y}(z) = P(X + Y \le z)$$

$$= \iint\limits_{\{(x,y)|x+y\le z\}} f(x,y)\,dxdy$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f(x,y)\,dy\right]dx$$

$$= \int_{-\infty}^{\infty} f_X(x)F_{Y|X}(z-x|x)\,dx$$

ถ้า X และ Y เป็นอิสระต่อกันจะได้ว่า

$$F_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{z-x} f_Y(y)\,dydx$$

$$= \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)\,dx$$

จากนั้นทำการหาอนุพันธ์ของ $F_{X+Y}(z)$ เทียบกับ Z จะได้ฟังก์ชันความหนาแน่นของ X+Y ดังนี้

$$f_{X+Y}(z) = \frac{d}{dz}F_{X+Y}(z)$$

$$= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$$

เพราะฉะนั้น เมื่อตัวแปรสุ่ม X และ Y เป็นอิสระกัน ฟังก์ชันการแจกแจงสะสม F_{X+Y} ของตัวแปรสุ่ม X+Y คือผลการประสานของฟังก์ชันการแจกแจงสะสม F_X ของ X และฟังก์ชันการแจกแจงสะสม F_Y ของ Y และฟังก์ชันความหนาแน่น f_{X+Y} ของ X+Y คือ ผลการประสานของฟังก์ชันความหนาแน่น f_X ของ X และฟังก์ชันความหนาแน่น f_Y ของ Y

2.5 ค่ามูลความเสี่ยง (Value at Risk: VaR)

เป็นตัวเลขวัดความเสี่ยงของการขาดทุนที่อาจเกิดขึ้นได้ ภายใต้ภาวะตลาดปกติ และภายในช่วงระยะเวลาใดเวลาหนึ่ง ซึ่งในทางสถิติจะประเมินโดยอาศัยความน่าจะเป็น หรือระดับความเชื่อมั่น เช่น ณ ระดับความเชื่อมั่น 95% ซึ่งหมายความว่าการเคลื่อนไหวของราคามีสิทธิที่จะหลุดออกนอกกรอบ ที่กำหนด โดยเฉลี่ย 20 วัน จะมีเหตุการณ์ที่ไม่ปกติปรากฏมาหนึ่งวัน (1 ใน 20 เท่ากับ 5 ใน 100) ระดับความเชื่อมั่นที่จะเลือกใช้นั้น แล้วแต่ว่าต้องการความเชื่อมั่นระดับใด หรือต้องการให้การวัดความเสี่ยงคลอบคลุมการเคลื่อนไหวของราคาในวงกว้างแค่ไหน ซึ่งส่วนใหญ่จะใช้ที่ระดับ 95% โดยวิธีที่นำมาใช้ในการหาค่ามูลค่าความเสี่ยงมี3วิธีคือ

1. การจำลองโดยใช้ข้อมูลจากอดีต (Historical Simulation) วิธีนี้จะทำการเรียงค่าอัตราผลตอบแทนจากน้อยไปหามากและหา percentileที่ i โดยที่ i หมายถึงระดับความเชื่อมั่นที่ต้องการ
2. วิธีเดลต้า โดยใช้การกระจายแบบปกติ (Delta Normal) วิธีนี้ข้อมูลต้องมีการแจกแจงแบบปกติ ซึ่งค่ามูลค่าความเสี่ยงจะคำนวณได้จาก

$$VaR = z_{\alpha} \cdot \sigma$$

โดย z_{α} คือค่าของการแจกแจงแบบปกติ ณ ระดับความเชื่อมั่นที่กำหนด

และ σ คือค่าส่วนเบี่ยงเบนมาตรฐานของอัตราผลตอบแทน

3. วิธีจำลองแบบมอนติ คาร์โล (Monte Carlo Simulation) จะใช้โปรแกรมจำลองข้อมูลอัตราผลตอบแทนแล้วจากนั้นจะใช้วิธีแบบเดียวกับวิธี Historical Simulation

แต่ละวิธีมีข้อดีข้อเสียแตกต่างกันไป การจะเลือกใช้นั้นขึ้นอยู่กับสถานการณ์ซึ่งบางครั้งอาจจะต้องใช้วิธีต่างๆผสมผสานกันไป