บทที่ 2

ทฤษฎีและสถิติที่เกี่ยวข้อง

วัตถุประสงค์หลักในการทำวิจัยนี้ เพื่อทำการศึกษาและเปรียบเทียบค่าความสัมพันธ์ของค่าสุดปลาย (Extreme Values) ของตัวแปรสุ่ม 2 ตัว โดยเรียกส่วนนี้ว่า Tail Dependence ของตัวแปรสุ่ม 2 ตัว ที่มีความสัมพันธ์กันและนำผลที่ได้มาประยุกต์ใช้กับการประเมินมูลค่าของตราสารอนุพันธ์ (Collateralized Dept Obligations : CDO) โดยทำการศึกษาและเปรียบเทียบการแจกแจงทั้ง 3 แบบ คือ การแจกแจงแบบปกติ, การแจกแจงแบบที และ การแจกแจงแบบดับเบิ้ลที ภายใต้สถานการณ์ต่าง ๆ ดังที่ได้กล่าวไว้ในขอบเขตการวิจัยในบทที่ 1

ในบทนี้จะกล่าวถึงรายละเอียดของทฤษฎีและการแจกแจงที่เกี่ยวข้องกับการวิจัย โดยมีทฤษฎี ที่เกี่ยวข้องดังนี้

2.1 Correlation

สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ระหว่าง X และ Y

$$\rho[X,Y]=\frac{Cov[X,Y]}{\sigma_X\sigma_Y}$$

โดยที่ $Cov[X,Y]$ เป็น Covariance ระหว่าง X และ Y

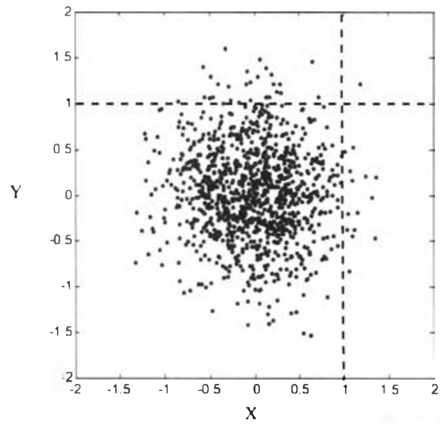
$$Cov[X,Y]=E[XY]-E[X]E[Y]$$

และ $\sigma^2[X]$, $\sigma^2[Y]$ เป็น Variance ของ X และ Y

สัมประสิทธิ์สหสัมพันธ์เชิงเส้นเป็นหน่วยวัดสัมประสิทธิ์ความสัมพันธ์เชิงเส้น

ในกรณีที่ตัวแปรสุ่มไม่มีความสัมพันธ์กัน $\rho[X,Y]=0$ ซึ่งค่า $Cov[X,Y]=0$ โดยที่การแจกแจงที่ไม่มีความสัมพันธ์กันนั้น เราจะเรียกการแจกแจงนี้ว่า Spherical Distribution

ตัวอย่าง เช่น การแจกแจงแบบ Standard Multivariate Normal โดยที่ X และ Y มีการแจกแจงส่วนริมเป็น Standard Normal $N(0,1)$ ที่มีค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0 โดยมีภาพแสดงการกระจายของจุดตัวอย่างจำนวน 1,000 จุด ดังภาพ 1.3



ภาพ 2.1 Standard Multivariate Normal ซึ่งมีค่า ρ เป็น 0

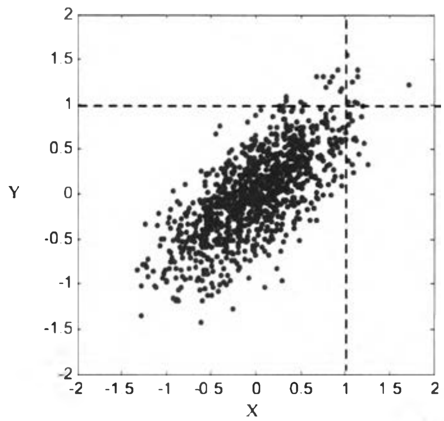
ในกรณีที่ตัวแปรสุ่มมีความสัมพันธ์กันอย่างสมบูรณ์โดยมีความสัมพันธ์กันเป็นเชิงเส้น เช่น $Y = aX + b$ หรือ $P[Y = aX + b] = 1$, $\forall a \in \Re \setminus \{0\}$, $b \in \Re$

เราจะได้ว่า $\rho[X,Y] = \pm 1$

และในกรณีที่ตัวแปรสุ่มมีความสัมพันธ์กันอย่างไม่สมบูรณ์ $-1 < \rho[X,Y] < 1$

ซึ่งทั้งสองกรณีหลังนี้มีการแจกแจงที่เรียกว่า Elliptical Distribution

ตัวอย่าง เช่น การแจกแจงแบบ Standard Multivariate Normal โดยที่ X และ Y มีการแจกแจงส่วนริมเป็น Standard Normal $N(0,1)$ ที่มีค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.7โดยมีภาพแสดงการกระจายของจุดตัวอย่างจำนวน 1,000 จุด ดังภาพนี้



ภาพ 2.2 Standard Multivariate Normal ซึ่งมีค่า ρ เป็น 0.7

2.2 Copula

Copula เป็นการแจกแจงร่วมของตัวแปรสุ่ม n ตัว โดยแต่ละตัวมีการแจกแจงส่วนริม (Marginal Distribution) เป็นการแจกแจงแบบเอกรูป (Uniform Distribution) บนช่วง [0,1] คือ $C:[0,1]^n \rightarrow [0,1]$ ซึ่งมีคุณสมบัติ ดังนี้

1. $C(x_1,...,x_n)$ เป็นฟังก์ชันเพิ่มขึ้นในทุก ๆ x_i
2. $C(1,...,1,x_i,1,...,1) = x_i$ ทุก I
3. สำหรับทุก ๆ $(a_1,...,a_n),(b_1,...,b_n) \in [0,1]^n$ โดย $a_i \le b_i$ เราจะได้ว่า

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} C\left(x_{(1,i_1)},\ldots,x_{(n,i_n)}\right) \ge 0$$

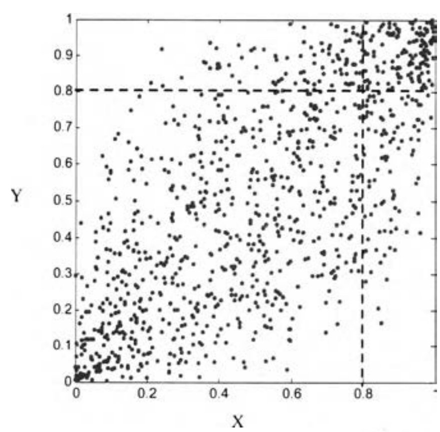
โดย $x_{(j,1)} = a_j$ และ $x_{(j,2)} = b_j$ สำหรับทุก ๆ $j \in \{1,...,n\}$

ในที่นี้เราจะศึกษา 2 Distribution ดังนี้

2.2.1 Gaussian Copula or Normal Copula สำหรับตัวแปรสุ่ม 2 ตัวมีรูปแบบ ดังนี้

$$C_\rho^{Ga}(x,y) = \int_{-\infty}^{\Phi^{-1}(x)} \int_{-\infty}^{\Phi^{-1}(y)} \frac{1}{2\pi\left(1-\rho^2\right)^{1/2}} \exp\left\{\frac{-\left(s^2-2\rho st+t^2\right)}{2\left(1-\rho^2\right)}\right\} ds\, dt$$

เมื่อ $-1 < \rho[X,Y] < 1$ และ Φ เป็น Univariate Standard Normal Distribution Function โดยมีภาพแสดงการกระจายของจุดตัวอย่าง จำนวน 1,000 จุด ดังภาพนี้

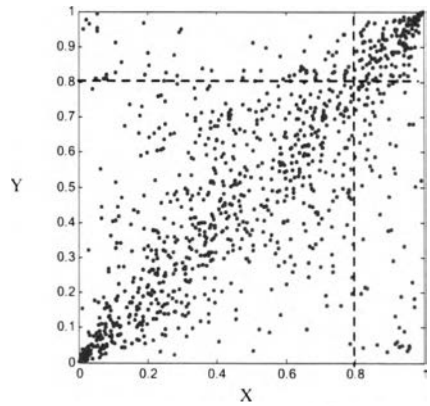


ภาพ 2.3 Gaussian Copula ที่มีค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.7

2.2.2 t-Distribution Copula สำหรับตัวแปรสุ่ม 2 ตัวมีรูปแบบ ดังนี้

$$C^{t}_{\upsilon,\rho}(x.y) = \int_{-\infty}^{t_{\upsilon}^{-1}(x)} \int_{-\infty}^{t_{\upsilon}^{-1}(y)} \frac{1}{2\pi\left(1-\rho^{2}\right)^{1/2}} \left\{1+\frac{\left(s^{2}-2\rho st+t^{2}\right)}{\upsilon\left(1-\rho^{2}\right)}\right\}^{-(\upsilon+2)/2} ds\, dt$$

เมื่อ t_{υ}^{-1} เป็น Inverse Distribution Function ของ Standard Univariate t ด้วยระดับขั้นความเสรี ν โดยมีภาพแสดงการกระจายของจุดตัวอย่างจำนวน 1,000 จุดดังภาพนี้



ภาพ 2.4 t-Copula ที่มีค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.7

2.3 การแจกแจงแบบปกติ

นิยาม การแจกแจงแบบปกติหนึ่งตัวแปร (Univariate Normal Distribution)

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

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

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

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

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

2.4 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

$$\gamma_3 = \frac{3(v-2)}{v-4}, \qquad v > 4$$

นิยาม การแจกแจงแบบทีสองตัวแปร (Bivariate t Distribution)

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

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

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

2.5 การแจกแจงแบบดับเบิ้ลที

นิยาม การแจกแจงแบบดับเบิ้ลทีหนึ่งตัวแปร

ตัวแปรสุ่ม X มีการแจกแจงความน่าจะเป็นแบบดับเบิ้ลที ถ้ากำหนดให้ M และ Z เป็นตัวแปรสุ่มที่มีการแจกแจงแบบทีที่อิสระต่อกัน และ $X = M + Z$

การวิจัยครั้งนี้ต้องการศึกษา การแจกแจงดับเบิ้ลทีที่มี ค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น 1 จึงกำหนดให้ M และ Z มีการแจกแจงแบบทีที่มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น 1 และมีรูปแบบฟังก์ชันเป็น

$$X = a_1 M + b_1 Z$$

ซึ่ง $0 < a_1 \leq 1$, $0 < b_1 \leq 1$ และ $b_1 = \sqrt{1 - a_1^2}$

นิยาม การแจกแจงแบบดับเบิ้ลทีสองตัวแปร

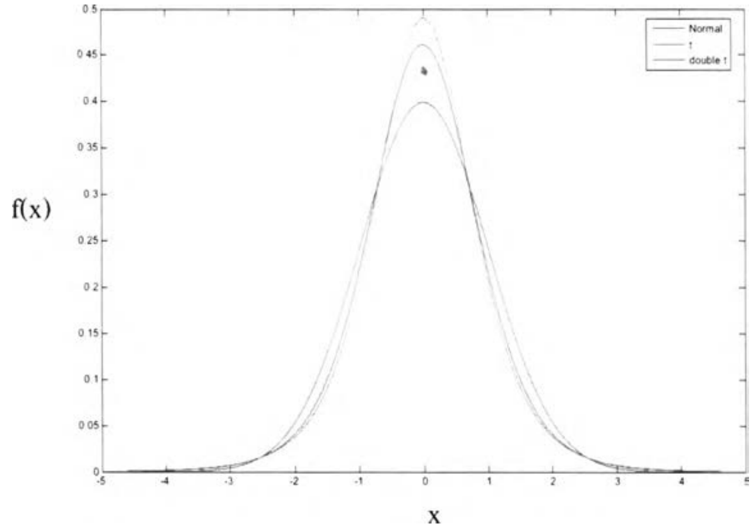
ในการวิจัยครั้งนี้ต้องการจะศึกษา การแจกแจงดับเบิ้ลทีสองตัวแปรที่มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น 1 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X และ Y เป็น $rho = a_1 a_2$ มีรูปแบบฟังก์ชันของ X และ Y เป็น

$$X = a_1 M + b_1 Z_1$$

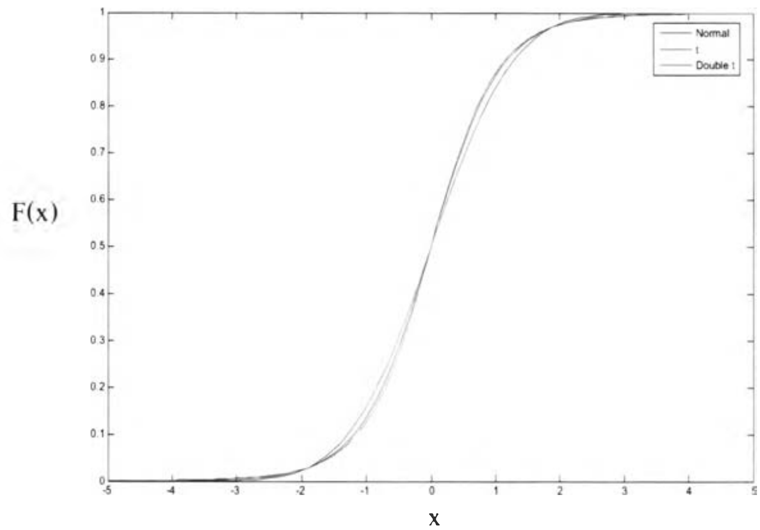
$$Y = a_2 M + b_2 Z_2$$

เมื่อ M , Z_1 และ Z_2 มีการแจกแจงแบบที ซึ่งมีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น 1 และเป็นอิสระซึ่งกันและกัน มีค่าพารามิเตอร์ คือ $0 < a_1 \leq 1$, $0 < b_1 \leq 1, b_1 = \sqrt{1 - a_1^2}$ และ $0 < a_1 \leq 1$, $0 < b_1 \leq 1, b_1 = \sqrt{1 - a_1^2}$

สามารถแสดงรูปแบบการแจกแจงที่ต่าง ๆ กันได้ดังรูปภาพข้างล่างนี้



ภาพ 2.5 แสดงฟังก์ชันความหนาแน่นของการแจกแจงแบบปกติ, การแจกแจงแบบที และการแจกแจงแบบดับเบิ้ลทีตามลำดับ เมื่อมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1



ภาพ 2.6 แสดงฟังก์ชันการแจกแจงสะสมของการแจกแจงแบบปกติ, การแจกแจงแบบที และการแจกแจงแบบดับเบิ้ลที เมื่อมีค่าเฉลี่ย 0 และความแปรปรวน 1

2.6 การแจกแจงแบบเลขชี้กำลัง

นิยาม การแจกแจงแบบเลขชี้กำลัง (Exponential Distribution)

ตัวแปรสุ่มค่าไม่เป็นลบ X มีการแจกแจงแบบเลขชี้กำลัง ที่มีพารามิเตอร์ λ เขียนแทนด้วย $X \sim Exp(\lambda)$ ถ้า X มีฟังก์ชันความหนาแน่นเป็น

$$f(x) = f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & , \quad x \geq 0 \\ 0 & , \quad x \geq 0 \end{cases}$$

หรืออาจนิยามในรูปแบบ

$$f(x) = f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & , \quad 0 \leq x < \infty \; ; \beta > 0 \\ 0 & , \quad x \geq 0 \end{cases}$$

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบเลขชี้กำลัง จะได้ว่า

1. ค่าเฉลี่ยของตัวแปรสุ่ม X

$$E(X) = \frac{1}{\lambda}$$

2. ค่าความแปรปรวนของตัวแปรสุ่ม X

$$Var(X) = \frac{1}{\lambda^2}$$

ตัวแปรสุ่มเลขชี้กำลังมีคุณสมบัติพิเศษ คือ ไม่มีความจำดังทฤษฎีนี้

ทฤษฎีบท คุณสมบัติไม่มีความจำของการแจกแจงแบบเลขชี้กำลัง

ถ้า X มีการแจกแจงแบบเลขชี้กำลัง $X \sim Exp(\lambda)$ จะได้ว่า

$$P(X \geq a + b | X \geq a) = P(X \geq b) \;\; , a,b > 0$$

2.7 Tail Dependence

เป็นการศึกษาความสัมพันธ์ของค่าที่สุดปลาย (Extreme Values) ของตัวแปรสุ่ม เรียกความสัมพันธ์นี้ว่า Tail Dependence โดยมีการแจกแจงส่วนริมของตัวแปรสุ่มเป็นการแจกแจงแบบต่อเนื่อง และสามารถวัดหาความสัมพันธ์ของฟังก์ชันให้อยู่ในรูปแบบ Copula ได้ โดยเป็นฟังก์ชันเพิ่มอย่างแน่นอนภายใต้เงื่อนไขของการแปลงข้อมูล

นิยาม ให้ X และ Y เป็นตัวแปรสุ่มที่มีฟังก์ชันการแจกแจงแบบ F_1 และ F_2 สัมประสิทธิ์ของ Tail Dependence ของ X และ Y มีค่าเป็น

$$\lim_{u \to 1} P\left[Y > F_2^{-1}(u) \middle| X > F_1^{-1}(u)\right] = \tau$$

โดยมีเงื่อนไขว่า $\tau \in [0,1]$

ถ้า $\tau \in (0,1]$ เราจะได้ว่า X และ Y มีความสัมพันธ์กันอย่างแน่นอน

ถ้า $\tau = 0$ เราจะได้ว่า X และ Y อิสระจากกัน

2.8 การจำลองแบบด้วยเทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique)

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