การกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มในอัลกอริทึมนอร์ทาโดยวิธีการจำลอง

Determining an Initial Correlation Matrix in NORTA Algorithm Using Simulation Approach

ประพันธ์ กล่อมพร (Praphan Klompon)* ดร.พัทธ์ชนก ศรีสุรเดชชัย (Dr.Patchanok Srisuradetchai)**

บทคัดย่อ

นอร์ทา (NORTA) เป็นอัลกอริทึมที่นิยมใช้สร้างตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงและเมทริกซ์สหสัมพันธ์ตามที่ต้องการ $(\mathbf{\Sigma}_X)$ แต่ในขั้นต้นของอัลกอริทึม ผู้ใช้จะกำหนดเมทริกซ์สหสัมพันธ์ของตัวแปรสุ่มที่มีการแจกแจงปรกติหลายตัวแปร $(\mathbf{\Sigma}_Z)$ แรกเริ่มที่เหมาะสม ในงานวิจัย ได้พัฒนาอัลกอริทึมเพื่อกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มโดยใช้แนวทางการจำลอง สองอัลกอริทึมที่นำเสนอ ได้แก่ อัลกอริทึมที่ใช้วิธีโมเมนต์ในการประมาณเมทริกซ์สหสัมพันธ์ (เรียกว่า อัลกอริทึม N) และอัลกอริทึมที่ใช้วิธีโมเมนต์ในการประมาณเมทริกซ์สหสัมพันธ์ที่มีการคัดเลือกเมทริกซ์เริ่มต้น (เรียกว่า อัลกอริทึม NK) อัลกอริทึม N และ NK ที่นำมาเปรียบเทียบได้สร้างเมทริกซ์สหสัมพันธ์แรกเริ่มโดยการนำเมทริกซ์สหสัมพันธ์ตามที่ต้องการบวกกับเมทริกซ์สุ่มสมมาตรที่มีสมาชิกบนแนวทแยงมุมหลักมีค่าเท่ากับ 0 ในขณะที่สมาชิกอื่นเป็นตัวแปรสุ่มที่มีการแจกแจงเลขชี้กำลัง หรือการแจกแจงครึ่งปรกติ นำเมทริกซ์ที่ได้ไปสร้างตัวแปรสุ่มตามอัลกอริทึมนอร์ทาจนกระทั่งได้ตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงและเมทริกซ์สหสัมพันธ์ตามที่ต้องการ จากการเปรียบเทียบ 2 อัลกอริทึม พบว่า อัลกอริทึม NK ให้ผลบวกกำลังสองของความคลาดเคลื่อนใกล้เคียงกับอัลกอริทึม N แต่ใช้เวลาในการประมวลผลน้อยกว่าอัลกอริทึม N

ABSTRACT

The NORTA is a popular algorithm to generate multivariate random vectors with desired distributions and correlation matrix. However, in the beginning step of NORTA algorithm, users need to determine a suitable correlation matrix for the multivariate normal random vector. In this study, two algorithms are developed in order to find an initial correlation matrix by simulation approach. Two proposed algorithms are (1) algorithm using the moment method to estimate correlation matrix (called algorithm N) and (2) algorithm using the moment method to estimate correlation matrix with selected initial correlation matrices (called algorithm NK). Both algorithms determine an input correlation matrix by using summation of desired correlation matrix and a random symmetric matrix that all elements on a main diagonal equal to zero and other elements are random numbers generated from exponential or half-normal distributions. The resulting matrix is used in NORTA algorithm until multivariate random vectors have desired distributions and correlation matrix. The results of simulation show that an input correlation matrix using algorithm N and NK gives about the same sums of squared errors. However, algorithm NK takes time less than algorithm N.

คำสำคัญ: อัลกอริทึมนอร์ทา ตัวแปรสุ่มหลายตัวแปร เมทริกซ์สหสัมพันธ์แรกเริ่ม

Keywords: NORTA Algorithm, Multivariate random variable, Initial Correlation Matrix

---

** นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์*

*** ผู้ช่วยศาสตราจารย์ สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์*

บทนำ

เหตุการณ์ส่วนใหญ่ในชีวิตจริงมักมีความสัมพันธ์กัน เช่น เหตุการณ์ที่จัดอยู่ในกระบวนการสโตแคสติก โดยตัวอย่างของเหตุการณ์ คือ ระยะเวลาในกระบวนการผลิตสินค้าโดยผ่านเครื่องจักรหลายขั้นตอน โดยที่เวลาจะมีความสัมพันธ์กับลักษณะเฉพาะของสินค้าแต่ละชิ้นที่ผลิต ตัวอย่างของงานวิจัยที่มีการประยุกต์ใช้ความสัมพันธ์ เช่น ในด้านวิศวกรรมการเงิน Das et al. (2001) พบว่าการประมาณความเสี่ยงของสินเชื่ออาจต่ำกว่าความเป็นจริงหากไม่สนใจสหสัมพันธ์ ในผังโครงสร้างงาน พบว่าต้นทุนในแต่ละส่วนมีความสัมพันธ์กัน (Lurie, Goldberg, 1998) ในตัวแบบการคืนทุนมีตัวแปรอธิบายคือขนาดของตลาดและราคาขาย โดยที่ตัวแปรอธิบายมีความสัมพันธ์กัน (Hull, 1997) ในงานของ Corner, Kirkwood (1996) พบว่าค่าคลาดเคลื่อนของการวิเคราะห์การตัดสินใจมีค่าเพิ่มขึ้นเมื่อสหสัมพันธ์ที่มีค่าเป็นบวกมีค่าเพิ่มขึ้น ในด้านการวิเคราะห์การตัดสินใจและการวิเคราะห์ความเสี่ยง Clemen, Reilly (1999) ใช้ตัวแบบคอพิวลาโดยที่ตัวแปรมีความสัมพันธ์กัน กับข้อมูลของการบิน ในงานของ Hill, Reilly (2000) สหสัมพันธ์ของสัมประสิทธิ์ของปัญหากระสอบทรายถูกใช้ในการทดสอบประสิทธิภาพ Yahav, Shmueli (2012) ใช้การจำลองตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงปัวซงในตัวแบบการกำหนดราคาผลิตภัณฑ์ไม่คงทน และใช้ในการตรวจสอบการแจ้งเตือนที่ผิดพลาดของตัวแบบการตรวจตราโรคระบาด จากสถานการณ์ที่กล่าวมาข้างต้น ถ้าหากจำเป็นต้องใช้การจำลองในการทดสอบตัวแบบ วิธีที่มีประสิทธิภาพที่ใช้สร้างตัวแปรสุ่มหลายตัวแปรจึงมีความสำคัญ

การสร้างตัวแปรสุ่มหลายตัวแปรสามารถทำได้หลายวิธี Niaki, Abbasi (2008) ได้จัดวิธีการสร้างตัวแปรสุ่มหลายตัวแปรที่พบได้โดยทั่วไปออกเป็น 3 ประเภท ได้แก่ วิธีเชิงวิเคราะห์ (Analytical approach) เป็นวิธีที่สร้างตัวแปรสุ่มหลายตัวแปรโดยใช้ฟังก์ชันการแจกแจงตามขอบ (Marginal distribution function) และฟังก์ชันการแจกแจงแบบมีเงื่อนไข (Conditional distribution function) วิธีเชิงตัวเลข (Numerical approach) เป็นวิธีที่ใช้วิธีการยอมรับและปฏิเสธ (acceptance/rejection method) ซึ่งต้องเลือกใช้ฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม (Joint probability density function) ที่คลอบคลุมฟังก์ชันความหนาแน่นความน่าจะเป็นร่วมที่ต้องการ และวิธีเชิงการจำลอง (Simulation approach) เป็นวิธีที่ใช้คุณลักษณะที่เฉพาะเจาะจงในการแปลง เช่น การแปลงนอร์ทา (NORTA transformation ซึ่งย่อมาจาก NORmal-To Anything transformation) หรือมีอีกชื่อหนึ่งคือ การแปลงนาทาฟ (Nataf transformation) ใช้ฟังก์ชันความหนาแน่นความน่าจะเป็นตามขอบและเมทริกซ์สหสัมพันธ์ซึ่งเป็นคุณลักษณะที่เฉพาะเจาะจงในการสร้างตัวแปรสุ่มหลายตัวแปร โดยการแปลงนอร์ทาถูกเสนอโดย Cario, Nelson (1997) ซึ่งมีจุดเริ่มต้นของแนวคิดมาจากงานของ Mardia (1970) และ Li, Hammond (1975)

การแปลงนอร์ทามีจุดประสงค์เพื่อใช้ในการสร้างเวกเตอร์ตัวแปรสุ่มหลายตัวแปรโดยการแปลงเวกเตอร์ตัวแปรสุ่มหลายตัวแปรที่มีเมทริกซ์สหสัมพันธ์ $(\mathbf{\Sigma}_Z)$ ซึ่งแต่ละตัวแปรสุ่มมีการแจกแจงตามขอบคือการแจกแจงปรกติมาตรฐาน ให้เป็นเวกเตอร์ตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงตามที่ต้องการและมีเมทริกซ์สหสัมพันธ์ตามที่ต้องการ $(\mathbf{\Sigma}_X)$ การแปลงนอร์ทามีขั้นตอนดังต่อไปนี้ เริ่มจากทำการสร้างเวกเตอร์ตัวแปรสุ่มที่มีการแจกแจงปรกติมาตรฐาน m ตัวแปร $\mathbf{Y}=(Y_1,...,Y_m)^T$ ที่เป็นอิสระกัน จากนั้นทำการแยกส่วนประกอบแบบโชเลสกี (Cholesky decomposition) $\mathbf{\Sigma}_Z$ โดยที่ $\mathbf{\Sigma}_Z=\mathbf{LL}^T$ เมื่อ $\mathbf{L}$ เป็นเมทริกซ์สามเหลี่ยมล่าง แล้วสร้างเวกเตอร์ตัวแปรสุ่มที่มีการแจกแจงปรกติมาตรฐาน m ตัวแปร $\mathbf{Z}=(Z_1,...,Z_m)^T$ ที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\mathbf{\Sigma}_Z$ โดยการแปลง $\mathbf{Z}=\mathbf{LY}$ แล้วทำการสร้างเวกเตอร์ตัวแปรสุ่ม $\mathbf{X}=(X_1,...,X_m)^T$ ที่มีการแจกแจงและเมทริกซ์สหสัมพันธ์ที่ต้องการ โดยการใช้ฟังก์ชันการแจกแจงสะสมในการแปลง $X_i=F_i^{-1}(Z_i)$ เมื่อ F_i^{-1} เป็นฟังก์ชันการแจกแจงสะสมผกผันของ X_i จะได้ว่า $\mathbf{X}=(X_1,...,X_m)^T$ เป็นเวกเตอร์ตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงตามที่ต้องการ และมีเมทริกซ์สหสัมพันธ์เท่ากับ $\mathbf{\Sigma}_X$

ปัญหาสำคัญของการแปลงนอร์ทา คือ การกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มของเวกเตอร์ตัวแปรสุ่มที่มีการแจกแจงปรกติหลายตัวแปรเพื่อใช้ในการแปลงให้เป็นเวกเตอร์ตัวแปรสุ่มที่มีการแจกแจงหลายตัวแปรและเมทริกซ์สหสัมพันธ์ตามที่ต้องการ เนื่องจากเมทริกซ์สหสัมพันธ์ของตัวแปรสุ่มที่เราต้องการสร้างมีค่าเท่ากับ $\mathbf{\Sigma}_X$ ในกรณีทั่วไป เมื่อสร้างเวกเตอร์ตัวแปรสุ่มที่มีการแจกแจงปรกติมาตรฐาน m ตัวแปร ที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\mathbf{\Sigma}_X$ แล้วนำไปแปลงเป็นเวกเตอร์ตัวแปรสุ่มมีการแจกแจงที่ต้องการ จะพบว่าเมทริกซ์สหสัมพันธ์ของเวกเตอร์ตัวแปรสุ่มที่สร้างมีค่าไม่เท่ากับ $\mathbf{\Sigma}_X$ ดังนั้นต้องมีการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มเพื่อให้ได้ตัวแปรสุ่มหลายตัวแปรที่มีเมทริกซ์สหสัมพันธ์ตามที่ต้องการ Cario, Nelson (1997) ได้เสนอให้ใช้การค้นหาแบบทวิภาค (Bisection search) ในการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่ม ในเวลาต่อมา งานส่วนใหญ่จะมีการมุ่งเน้นไปที่การหาความสัมพันธ์ระหว่างสหสัมพันธ์แรกเริ่ม ρ_Z และสหสัมพันธ์ที่ต้องการ ρ_X โดยพิจารณาสมการ

$$\rho_X(i,j)=\frac{E\left[X_iX_j\right]-E\left[X_i\right]E\left[X_j\right]}{\sqrt{Var\left(X_i\right)Var\left(X_j\right)}}=\frac{E\left[X_iX_j\right]}{\sigma_i\sigma_j}-\frac{\mu_i\mu_j}{\sigma_i\sigma_j} \tag{1}$$

โดยที่ $\mu_i,\ \mu_j$ แทนค่าเฉลี่ยของ $X_i,\ X_j$ ตามลำดับ $\sigma_i,\ \sigma_j$ แทนส่วนเบี่ยงเบนมาตรฐานของ $X_i,\ X_j$ ตามลำดับโดยในที่นี้จะพิจารณาเฉพาะการแจกแจงที่ทราบค่าเฉลี่ยและความแปรปรวน เนื่องจาก $\left(Z_i,Z_j\right)$ มีการแจกแจงปรกติมาตรฐานสองตัวแปรที่มีสหสัมพันธ์ $Corr\left(Z_i,Z_j\right)=\rho_Z(i,j)$ ทำให้ได้ว่า $E\left[X_iX_j\right]$ ในสมการ (1) สามารถเขียนให้อยู่ในรูป

$$E\left[X_iX_j\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F_i^{-1}\left[\Phi\left(z_i\right)\right]F_j^{-1}\left[\Phi\left(z_j\right)\right]\phi\left[z_i,z_j,\rho_Z(i,j)\right]dz_idz_j$$

โดยที่ $\phi\left[z_i,z_j,\rho_z(i,j)\right]$ แทนฟังก์ชันการแจกแจงความน่าจะเป็นร่วมของการแจกแจงปรกติมาตรฐานสองตัวแปร โดยมีสหสัมพันธ์เท่ากับ $\rho_Z(i,j)$ ซึ่งจะมีฟังก์ชันคือ

$$\phi\left[z_i,z_j,\rho_Z(i,j)\right]=\frac{1}{2\pi\sqrt{1-\rho_Z^2(i,j)}}\exp\left\{-\frac{z_i^2-2\rho_Z(i,j)z_iz_j+z_j^2}{2\left[1-\rho_Z^2(i,j)\right]}\right\}$$

ในการแก้สมการ (1) เพื่อหาสหสัมพันธ์แรกเริ่มทำได้ยาก และต้องแก้หาคำตอบจำนวน $m(m-1)$ สมการ ซึ่งมีหลายงานวิจัยที่เสนออัลกอริทึมที่ให้ในการหาเมทริกซ์สหสัมพันธ์แรกเริ่มเมื่อ X_i และ X_j เป็นตัวแปรสุ่มต่อเนื่องทั้งคู่ เช่น สูตรเชิงประจักษ์ (Empirical formulae) ถูกเสนอโดย Kiureghian, Liu (1986) วิธีการหาราก (Root finding method) ถูกเสนอโดย Chen (2001) และวิธีการค้นหาเชิงเส้น (Linear search method) ถูกเสนอโดย Li et al. (2008) ในกรณีที่ X_i หรือ X_j เป็นตัวแปรสุ่มไม่ต่อเนื่องไม่สามารถใช้วิธีที่กล่าวมาข้างต้นในการหาคำตอบของสมการได้ ต่อมา Avramidis et al. (2009) ได้ศึกษาเกี่ยวกับฟังก์ชันความสัมพันธ์ระหว่าง ρ_Z และ ρ_X และพัฒนาวิธีที่ใช้ในการกำหนดค่า ρ_Z ของตัวแปรสุ่มแบบไม่ต่อเนื่อง 2 ตัวแปร Avramidis (2013) ได้ศึกษาเพิ่มในกรณีผสมระหว่างตัวแปรสุ่มแบบต่อเนื่องและตัวแปรสุ่มแบบไม่ต่อเนื่อง Xiao (2017) ได้เสนอวิธีการหาเมทริกซ์สหสัมพันธ์แรกเริ่มโดยใช้วิธีการวางตัวผิดที่ (False position method) และเขียนสมการที่ (1) ในรูปแบบใหม่โดยการแทนค่า $Z_i = Y_i$, $Z_j = \rho_Z Y_i + \sqrt{1-\rho_Z^2} Y_i$, $u_i = \Phi(Y_i)$, และ $u_j = \Phi(Y_j)$ ได้สมการดังนี้

$$\rho_X(i,j) = -\frac{\mu_i \mu_j}{\sigma_i \sigma_j} + \frac{1}{\sigma_i \sigma_j} \int_0^1 \int_0^1 H\left[u_i, u_j; \rho_Z(i,j)\right] du_i du_j \tag{2}$$

โดยที่ $H\left[u_i, u_j; \rho_Z(i,j)\right] = F_i^{-1}(u_i) \cdot F_i^{-1}\left\{\Phi\left[\rho_Z(i,j)\Phi^{-1}(u_i) + \sqrt{1-\rho_Z^2(i,j)}\Phi^{-1}(u_j)\right]\right\}$ โดยใช้วิธีเสมือนมอนติคาร์โล (Quasi Monte Carlo method) โดยใช้ลำดับฮัลตันในการประมาณปริพันธ์สองชั้นในสมการที่ (2) Niavarani, Smith (2013) ได้เสนออัลกอริทึมในการสร้างตัวแปรสุ่มหลายตัวแปรโดยวิธีนอทาร์ ซึ่งหลีกเลี่ยงการแก้สมการที่ (1) จำนวน $m(m-1)$ สมการ ซึ่งขั้นตอนมีดังนี้ เริ่มจากการสร้างเมทริกซ์สุ่มสมมาตร $\mathbf{D}$ โดยมีสมาชิกบนแนวทแยงมุมหลักเท่ากับ 0 และสมาชิกอื่นๆเป็นเลขสุ่มบนช่วง $(-1-\rho_X(i,j), 1-\rho_X(i,j))$ จากนั้นสร้างเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ โดยการนำเมทริกซ์ $\mathbf{D}$ มาบวกกับเมทริกซ์เริ่มต้นโดยแทนค่าด้วยเมทริกซ์สหสัมพันธ์ที่ต้องการ $\mathbf{\Sigma}_X$ จะได้ $\mathbf{\Sigma}_Z = \mathbf{\Sigma}_X + \mathbf{D}$ และตรวจสอบว่าเมทริกซ์ $\mathbf{\Sigma}_Z$ เป็นเมทริกซ์บวกแน่นอน (Positive definite matrix) หรือไม่ ในกรณีที่ $\mathbf{\Sigma}_Z$ ไม่เป็นเมทริกซ์บวกแน่นอน ให้ทำการสร้างเมทริกซ์ $\mathbf{D}$ ใหม่จนกว่าจะได้ว่า $\mathbf{\Sigma}_Z$ เป็นเมทริกซ์บวกแน่นอน เมื่อได้ว่า $\mathbf{\Sigma}_Z$ เป็นเมทริกซ์บวกแน่นอน จะนำเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ ไปสร้างเวกเตอร์ตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงตามที่ต้องการ จำนวน 5000 เวกเตอร์ จากนั้นคำนวณผลบวกกำลังสองของความคลาดเคลื่อน $SSE = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\left[\rho_X(i,j) - \hat{\rho}_X(i,j)\right]^2$ และตรวจสอบว่ามีค่าน้อยกว่าผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับได้หรือไม่ ในกรณีที่มีค่ามากกว่าหรือเท่ากับ จะทำการกลับไปยังขั้นตอนการสร้างเมทริกซ์ $\mathbf{D}$ ใหม่จนกว่าจะสามารถสร้างเวกเตอร์ตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงตามที่ต้องการที่มีผลบวกกำลังสองของความคลาดเคลื่อนน้อยกว่าผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับ ซึ่งจะได้ว่าเวกเตอร์ตัวแปรสุ่มหลายตัวแปรจำนวน 5000 เวกเตอร์นั้น เป็นเวกเตอร์ตัวแปรสุ่มหลายตัวแปรที่มีการแจกแจงและเมทริกสหสัมพันธ์ตามที่ต้องการ ผลการศึกษาของอัลกอริทึมนี้ได้ถูกนำไปเปรียบเทียบกับวิธีโครงข่ายประสาทเทียม (Artificial neural networks) ที่ถูกเสนอโดย Niaki, Abbasi (2008) และวิธีนิวตัน (Newton method) ผลที่ได้พบว่าวิธีที่ Niavarani, Smith (2013) เสนอสามารถสร้างเวกเตอร์ตัวแปนสุ่มหลายตัวแปรที่มีผลบวกกำลังสองของความคลาดเคลื่อนที่น้อยกว่าอีก 2 วิธี

ในงานของ Niavarani, Smith (2013) ซึ่งได้เสนอวิธีที่เข้าใจง่าย แต่เนื่องจากการสร้างสมาชิกในเมทริกซ์ $\mathbf{D}$ เป็นการสุ่มเลขสุ่มบนช่วง $(-1-\rho_x(i,j), 1-\rho_x(i,j))$ ซึ่งอาจใช้เวลาในการประมวลผลนาน ในงานวิจัยนี้จึงมีแนวคิดที่จะพัฒนาการสร้างสมาชิกในเมทริกซ์ $\mathbf{D}$ ใหม่ โดยใช้คุณสมบัติที่ Xiao (2017) ได้กล่าวว่าถ้าฟังก์ชันการแจกแจงตามขอบของ X_i และ X_j เป็นฟังก์ชันไม่เสื่อมสภาพ (Non-degenerate function) ρ_X และ ρ_Z จะมีคุณสมบัติ 3 ข้อดังนี้

1. ρ_X เป็นฟังก์ชันเพิ่มโดยแท้ต่อเนื่อง (Continuous strictly increasing function) ของ ρ_Z (Avramidis, 2013)
2. ถ้า $\rho_Z \geq 0 (\leq 0)$ จะได้ว่า $\rho_X \geq 0 (\leq 0)$ (Cario, Nelson, 1997)
3. $|\rho_X| \leq |\rho_Z|$ (Liu, Kiureghian, 1986)

จะเห็นได้ว่าขนาดของสมาชิกของเมทริกซ์สหสัมพันธ์แรกเริ่มมีขนาดมากกว่าสมาชิกในตำแหน่งเดียวกันของเมทริกซ์สหสัมพันธ์ที่ต้องการ ดังนั้นจึงได้เสนอการสร้างสมาชิกในเมทริกซ์ $\mathbf{D}$ ใหม่ โดยการใช้ตัวแปรสุ่มที่มีการแจกแจงเลขชี้กำลังหรือการแจกแจงครึ่งปรกติ และเนื่องจากในแต่ละรอบของการสร้างเมทริกซ์ $\mathbf{D}$ มีแนวคิดจะเปลี่ยนเมทริกซ์ที่นำมาบวกกับเมทริกซ์ $\mathbf{D}$ เมื่อเมทริกซ์นั้นให้ผลรวมของความคลาดเคลื่อนกำลังสองที่น้อยกว่าในรอบก่อนหน้า

จากเมทริกซ์ $\mathbf{\Sigma}_X$ ไปเป็นเมทริกซ์ $\mathbf{\Sigma}_Z$ แต่เนื่องจากการเปลี่ยนเมทริกซ์เริ่มต้นทำให้ต้องมีการกำหนดเครื่องหมายให้กับเมทริกซ์ $\mathbf{D}$ ก่อนที่จะนำมาสร้างเมทริกซ์สหสัมพันธ์แรกเริ่ม โดยอัลกอริทึมที่นำมาเปรียบเทียบมี 2 อัลกอริทึม ได้แก่ อัลกอริทึมที่ใช้วิธีโมเมนต์ในการประมาณเมทริกซ์สหสัมพันธ์ โดยจะเรียกว่า อัลกอริทึม N และอัลกอริทึมที่ใช้วิธีโมเมนต์ในการประมาณเมทริกซ์สหสัมพันธ์ที่มีการคัดเลือกเมทริกซ์เริ่มต้น โดยจะเรียกว่า อัลกอริทึม NK

วัตถุประสงค์การวิจัย

เพื่อพัฒนาอัลกอริทึมที่ใช้ในการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มสำหรับอัลกอริทึม NORTA และทำการเปรียบเทียบอัลกอริทึมทั้ง 2 อัลกอริทึม ได้แก่ อัลกอริทึม N และอัลกอริทึม NK

วิธีการวิจัย

ขอบเขตการศึกษา

ใช้กรณีศึกษา 3 กรณี ดังนี้

กรณีที่ 1 ตัวแปรสุ่ม 2 ตัวแปร $(X_1, X_2)^T$ โดยที่ X_1 มีการแจกแจงเอกรูปไม่ต่อเนื่อง $(1,10)$ และ X_2 มีการแจกแจงเลขชี้กำลัง $(\beta = 10)$ โดยที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ และกำหนดผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับได้ (SSE_{target}) เท่ากับ 3×10^{-6}

กรณีที่ 2 ตัวแปรสุ่ม 3 ตัวแปร $(X_1, X_2, X_3)^T$ โดยที่ X_1, X_2 และ X_3 มีการแจกแจงทวินาม $(n = 3,\ p = 0.5)$ โดยที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\begin{bmatrix} 1 & 0.2 & -0.8 \\ 0.2 & 1 & 0.2 \\ -0.8 & 0.2 & 1 \end{bmatrix}$ และกำหนดผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับได้ (SSE_{target}) เท่ากับ 6×10^{-5}

กรณีที่ 3 ตัวแปรสุ่ม 4 ตัวแปร $(X_1, X_2, X_3, X_4)^T$ โดยที่ X_1, X_2, X_3 และ X_4 มีการแจกแจงแกมมา $(\alpha = 14.4,\ \beta = 0.03424)$ โดยที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\begin{bmatrix} 1 & 0.7 & 0.5 & -0.9 \\ 0.7 & 1 & 0.7 & -0.6 \\ 0.5 & 0.7 & 1 & -0.3 \\ -0.9 & -0.6 & -0.3 & 1 \end{bmatrix}$ และกำหนดผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับได้ (SSE_{target}) เท่ากับ 3.5×10^{-4}

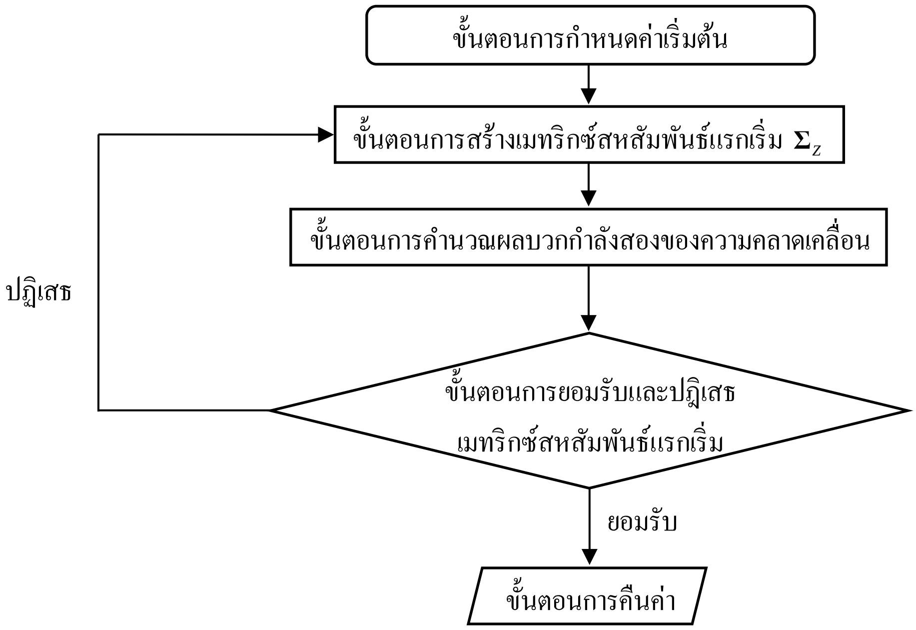
ในส่วนของการกำหนดพารามิเตอร์ของการแจกแจงเลขชี้กำลังและการแจกแจงครึ่งปรกติจะกำหนดให้มีค่าเฉลี่ยเท่ากับ 0.075, 0.05, 0.025, 0.01 และ 0.005

การเก็บข้อมูลและวิเคราะห์ข้อมูล

การเปรียบเทียบประสิทธิภาพของอัลกอริทึมจะพิจารณาจากเวลาและผลบวกกำลังสองของความคลาดเคลื่อน เวลาวัดค่าจากเวลาที่ใช้ในการประมวลผลตั้งแต่เริ่มอัลกอริทึมไปจนถึงเวลาที่อัลกอริทึมประมวลผลเสร็จ โดยหน่วยของเวลาที่วัดค่ามีหน่วยเป็นวินาที และผลบวกกำลังสองของความคลาดเคลื่อนจะคำนวณหลังจากที่อัลกอริทึมประมวลผลเสร็จ โดยจะนำ $\mathbf{\Sigma}_Z$ ที่ได้จากอัลกอริทึม มาสร้างเวกเตอร์ตัวแปรสุ่ม $\mathbf{X} = (X_1, \ldots, X_m)^T$ จำนวน 10^6 เวกเตอร์ โดยที่ m แทนจำนวนตัวแปร จากนั้นนำมาประมาณเมทริกซ์สหสัมพันธ์ $\hat{\mathbf{\Sigma}}_X$ โดยใช้วิธีโมเมนต์ แล้วทำการคำนวณผลบวกกำลังสองของความคลาดเคลื่อนโดย $SSE = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \left[\rho_X(i,j) - \hat{\rho}_X(i,j)\right]^2$ โดยจะประมวลผลกรณีละ 10 ครั้ง แล้วนำผลที่ได้มาหาค่าเฉลี่ยแล้วทำการเปรียบเทียบ

อัลกอริทึม

อัลกอริทึมทั้ง 4 อัลกอริทึมที่นำมาเปรียบเทียบมีขั้นตอนหลัก ๆ ได้แก่ ขั้นตอนการกำหนดค่าเริ่มต้น ขั้นตอนการสร้างเมทริกซ์ $\mathbf{D}$ และเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ ขั้นตอนการคำนวณผลบวกกำลังสองของความคลาดเคลื่อน ขั้นตอนการยอมรับและปฏิเสธ และขั้นตอนการคืนค่า ดังภาพที่ 1.



ภาพที่ 1 ขั้นตอนการทำงานโดยสรุปของอัลกอริทึมทั้ง 2 อัลกอริทึม

ในขั้นตอนการกำหนดค่าเริ่มต้นให้อัลกอริทึมทั้ง 2 อัลกอริทึม จะมีการกำหนดผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับได้ $\left(SSE_{target}\right)$ กำหนดให้เมทริกซ์เริ่มต้น $\mathbf{\Sigma}_{initial} = \mathbf{\Sigma}_X$ เพื่อใช้ในการสร้างเมทริกซ์สหสัมพันธ์แรกเริ่มและกำหนดเมทริกซ์ $\mathbf{S}_{m\times m}$ โดยที่มี $s_{ij} = \begin{cases} sign\left(\rho_X\left(i,j\right)\right), & i \neq j \\ 0, & i = j \end{cases}$ เป็นสมาชิก และในอัลกอริทึม NK กำหนดให้ผลบวกกำลังสองของความคลาดเคลื่อนที่ใช้ในการคัดเลือกเมทริกซ์เริ่มต้น $SSE_{compared} = m\left(m-1\right)$

ขั้นตอนการสร้างเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ ในอัลกอริทึมทั้ง 2 อัลกอริทึม มีขั้นตอนเหมือนกันได้แก่ เริ่มจากการสร้างเมทริกซ์ $\mathbf{D}_{m\times m}$ โดยที่สมาชิกบนแนวทแยงมุมหลักมีค่าเท่ากับ 0 และสมาชิกอื่นๆ เป็นเลขสุ่มที่สุ่มมาจากตัวแปรสุ่มที่มีการแจกแจงเลขชี้กำลัง หรือการแจกแจงครึ่งปรกติ ที่มีพารามิเตอร์ที่เหมาะสม จากนั้นหาผลคูณในระดับสมาชิก (Element-wise multiplication) ของเมทริกซ์ $\mathbf{S}$ และเมทริกซ์ $\mathbf{D}$ แล้วนำมาบวกกับเมทริกซ์เริ่มต้น $\mathbf{\Sigma}_{initial}$ จะได้เมทริกซ์สหสัมพันธ์แรกเริ่มดังนี้

$$\mathbf{\Sigma}_Z = \mathbf{\Sigma}_{initial} + \mathbf{S} \circ \mathbf{D}$$

จากนั้นทำการตรวจสอบว่าเมทริกซ์ $\mathbf{\Sigma}_Z$ เป็นเมทริกซ์บวกแน่นอนหรือไม่ และตรวจสอบว่า $-1 < \rho_Z\left(i,j\right) < 1,\ i \neq j$ หรือไม่ ถ้ามีอย่างน้อย 1 เงื่อนไขที่ไม่เป็นจริง ต้องกลับไปเริ่มต้นใหม่ตั้งแต่การสร้างเมทริกซ์ $\mathbf{D}_{m\times m}$

ในขั้นตอนการคำนวณผลบวกกำลังสองของความคลาดเคลื่อน เริ่มจากการคำนวณเมทริกซ์สหสัมพันธ์ของตัวแปรสุ่มหลายตัวแปร $\mathbf{\Sigma}_X$ ที่จะสร้างได้จากการใช้เมทริกซ์สหสัมพันธ์เริ่มต้น $\mathbf{\Sigma}_Z$ ที่ได้จากขั้นตอนก่อนหน้า อัลกอริทึมทั้ง 2 อัลกอริทึม จะใช้การสร้างเวกเตอร์สุ่มด้วยการแปลงนอร์ทาจำนวน 50000 เวกเตอร์ โดยใช้ฟังก์ชันการแจกแจงสะสม F_i ของแต่ละตัวแปร และใช้เมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ ที่คำนวณได้จากขั้นตอนก่อนหน้า เพื่อนำมาประมาณเมทริกซ์สหสัมพันธ์ $\hat{\mathbf{\Sigma}}_X$ โดยวิธีโมเมนต์ จากนั้นการคำนวณผลบวกกำลังสองของความคลาดเคลื่อนของทั้ง 2 อัลกอริทึม โดยคำนวณเหมือนกันดังนี้

$$SSE = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \left[\rho_X(i,j) - \hat{\rho}_X(i,j)\right]^2$$

โดยที่ $\rho_X(i,j)$ และ $\hat{\rho}_X(i,j)$ เป็นสมาชิกแถวที่ i หลักที่ j ของ $\mathbf{\Sigma}_X$ และ $\hat{\mathbf{\Sigma}}_X$ ตามลำดับ

ในขั้นตอนการยอมรับและปฏิเสธเมทริกซ์สหสัมพันธ์แรกเริ่ม อัลกอริทึมทั้ง 2 อัลกอริทึม จะทำการตรวจสอบว่าผลบวกกำลังสองของความคลาดเคลื่อนที่คำนวณได้ (SSE) มีค่าน้อยกว่าผลบวกกำลังสองของความคลาดเคลื่อนที่ยอมรับได้ (SSE_{target}) หรือไม่ ถ้า $SSE < SSE_{target}$ จะดำเนินการในขั้นตอนการคืนค่า แต่ถ้าเงื่อนไขไม่เป็นจริง ในอัลกอริทึม N จะกลับไปดำเนินการใหม่ในขั้นตอนการสร้างเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ ในขณะที่อัลกอริทึมที่ NK จะมีการคัดเลือกเมทริกซ์เริ่มต้น โดยการตรวจสอบเงื่อนไข $SSE < SSE_{compared}$ ถ้าเงื่อนไขไม่เป็นจริง จะกลับไปดำเนินการใหม่ในขั้นตอนการสร้างเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$ แต่ถ้าเงื่อนไขเป็นจริง จะทำการกำหนดเมทริกซ์เริ่มต้นใหม่ โดยให้เมทริกซ์เริ่มต้น $\mathbf{\Sigma}_{initial} = \mathbf{\Sigma}_Z$ กำหนดให้ผลบวกกำลังสองของความคลาดเคลื่อนที่ใช้ในการคัดเลือกเมทริกซ์เริ่มต้น $SSE_{compared} = SSE$ และกำหนดเมทริกซ์ $\mathbf{S}$ ใหม่ โดยที่มีสมาชิก $s_{ij} = sign\left(\rho_X(i,j) - \hat{\rho}_X(i,j)\right)$ จากนั้นจึงกลับไปดำเนินการใหม่ในขั้นตอนการสร้างเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$

ในขั้นตอนการคืนค่ามีการคำนวณเวลาที่ใช้ในการประมวลผลตั้งแต่ขั้นตอนแรกจนเข้าสู่ขั้นตอนการคืนค่า จากนั้นทำการคืนค่าเวลาในการประมวลผล และเมทริกซ์สหสัมพันธ์แรกเริ่ม $\mathbf{\Sigma}_Z$

ผลการวิจัย

กรณีศึกษาที่ 1 การสร้างตัวแปรสุ่มที่มีการแจกแจงสองตัวแปร ได้แก่ ตัวแปรสุ่มที่มีการแจกแจงเอกรูปไม่ต่อเนื่อง $(1,10)$ และตัวแปรสุ่มที่มีการแจกแจงเลขชี้กำลัง $(\beta = 10)$ โดยที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ ผลปรากฏว่า อัลกอริทึม NK มีแนวโน้มที่จะใช้เวลาในการประมวลผลน้อยว่าอัลกอริทึม N โดยเมื่อใช้การแจกแจงเลขชี้กำลังที่มีค่าเฉลี่ยเท่ากับ 0.075, 0.05, 0.025, 0.01 และ 0.005 ใช้เวลาในการประมวลผลเฉลี่ยประมาณ 1.7, 1.8, 1.1, 1.2 และ 2.2 วินาที ตามลำดับ และเมื่อใช้การแจกแจงครึ่งปรกติที่มีค่าเฉลี่ยเท่ากับ 0.075, 0.05, 0.025, 0.01 และ 0.005 ใช้เวลาในการประมวลผลเฉลี่ยประมาณ 3.2, 2.2, 1.4, 1.3 และ 2.4 วินาที ตามลำดับ ดังแสดงในตารางที่ 1 ซึ่งการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มโดยการใช้อัลกอริทึม N และอัลกอริทึม NK มีผลบวกกำลังสองของความคลาดเคลื่อนไม่แตกต่างกัน ดังแสดงในตารางที่ 4

กรณีศึกษาที่ 2 การสร้างตัวแปรสุ่มที่มีการแจกแจงทวินามสามตัวแปร $(n = 3,\ p = 0.5)$ โดยที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\begin{bmatrix} 1 & 0.2 & -0.8 \\ 0.2 & 1 & 0.2 \\ -0.8 & 0.2 & 1 \end{bmatrix}$ ผลปรากฏว่า อัลกอริทึม NK มีแนวโน้มที่จะใช้เวลาในการประมวลผลน้อยกว่า

อัลกอริทึม N เมื่อใช้การแจกแจงเลขชี้กำลังที่มีค่าเฉลี่ยเท่ากับ 0.075, 0.05, 0.025, 0.01 และ 0.005 ใช้เวลาในการประมวลผลเฉลี่ยประมาณ 59.7, 8.8, 3.6 และ 4.7 วินาที ตามลำดับ และเมื่อใช้การแจกแจงครึ่งปรกติที่มีค่าเฉลี่ยเท่ากับ 0.075, 0.05, 0.025, 0.01 และ 0.005 ใช้เวลาในการประมวลผลเฉลี่ยประมาณ 140.8, 115.6, 11.8, 2.9 และ 4.7 วินาที ตามลำดับ ดังแสดงในตารางที่ 2 ซึ่งการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มโดยการใช้อัลกอริทึม N และอัลกอริทึม NK มีผลบวกกำลังสองของความคลาดเคลื่อนไม่แตกต่างกัน ดังแสดงในตารางที่ 5

กรณีศึกษาที่ 3 การสร้างตัวแปรสุ่มที่มีการแจกแจงแกมมาสี่ตัวแปร $(\alpha = 14.4,\ \beta = 0.03424)$ โดยที่มีเมทริกซ์สหสัมพันธ์เท่ากับ $\begin{bmatrix} 1 & 0.7 & 0.5 & -0.9 \\ 0.7 & 1 & 0.7 & -0.6 \\ 0.5 & 0.7 & 1 & -0.3 \\ -0.9 & -0.6 & -0.3 & 1 \end{bmatrix}$ ผลปรากฏว่า อัลกอริทึม NK มีแนวโน้มที่จะใช้เวลาในการประมวลผลน้อยกว่าอัลกอริทึม N โดยเมื่อใช้การแจกแจงเลขชี้กำลังที่มีค่าเฉลี่ยเท่ากับ 0.05, 0.025, 0.01 และ 0.005 ใช้เวลาในการประมวลผลเฉลี่ยประมาณ 553.4, 42.1, 3.4 และ 2.8 วินาที ตามลำดับ ส่วนในกรณีที่ใช้ค่าเฉลี่ยเท่ากับ 0.075 ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม. และเมื่อใช้การแจกแจงครึ่งปรกติที่มีค่าเฉลี่ยเท่ากับ 0.025, 0.01 และ 0.005 ใช้เวลาในการประมวลผลเฉลี่ยประมาณ 100.7, 2.9 และ 2.8 วินาที ตามลำดับ ส่วนในกรณีที่ใช้ค่าเฉลี่ยเท่ากับ 0.075 และ 0.05 ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม. ดังแสดงในตารางที่ 3 ซึ่งการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มโดยการใช้อัลกอริทึม N และอัลกอริทึม NK มีผลบวกกำลังสองของความคลาดเคลื่อนไม่แตกต่างกัน ดังแสดงในตารางที่ 6

ตารางที่ 1 ระยะเวลาเฉลี่ย(วินาที)ในการหาเมทริกสหสัมพันธ์แรกเริ่มของแต่ละอัลกอริทึม กรณีศึกษาที่ 1

Algorithm	Distribution	Mean of the probability distribution				
		0.075	0.05	0.025	0.01	0.005
N	Exponential	2.0	2.8	6.5	293.2	*
	Half-Normal	5.0	3.3	20.7	*	*
NK	Exponential	1.8	1.7	1.1	1.2	2.2
	Half-Normal	3.2	2.2	1.4	1.3	2.4

* ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม.

ตารางที่ 2 ระยะเวลาเฉลี่ย(วินาที)ในการหาเมทริกสหสัมพันธ์แรกเริ่มของแต่ละอัลกอริทึม กรณีศึกษาที่ 2

Algorithm	Distribution	Mean of the probability distribution				
		0.075	0.05	0.025	0.01	0.005
N	Exponential	172.8	715.2	*	*	*
	Half-Normal	132.2	145.0	*	*	*
NK	Exponential	59.7	26.4	8.8	3.6	4.7
	Half-Normal	140.8	115.6	11.8	2.9	4.7

* ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม.

ตารางที่ 3 ระยะเวลาเฉลี่ย(วินาที)ในการหาเมทริกสหสัมพันธ์แรกเริ่มของแต่ละอัลกอริทึม กรณีศึกษาที่ 3

Algorithm	**Distribution**	**Mean of the probability distribution**				
		0.075	**0.05**	**0.025**	**0.01**	**0.005**
N	Exponential	*	549.0	50.0	3.9	9.1
	Half-Normal	*	*	63.9	6.9	21.4
NK	Exponential	*	553.4	42.1	3.4	2.8
	Half-Normal	*	*	100.7	2.9	2.8

* ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม.

ตารางที่ 4 ผลบวกกำลังสองของความคลาดเคลื่อนเฉลี่ย $\left(\times 10^{-6}\right)$ เมื่อนำเมทริกสหสัมพันธ์แรกเริ่มไปสร้างเวกเตอร์ตัวแปรสุ่มจำนวน 10^6 เวกเตอร์ ของแต่ละอัลกอริทึม กรณีศึกษาที่ 1

Algorithm	**Distribution**	**Mean of the probability distribution**				
		0.075	**0.05**	**0.025**	**0.01**	**0.005**
N	Exponential	5.0157	18.2278	9.2302	7.1595	*
	Half-Normal	18.1576	8.0579	11.1187	*	*
NK	Exponential	16.2831	7.6364	6.8699	13.6039	5.1191
	Half-Normal	4.6969	8.4089	1.5117	5.2397	7.2546

* ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม.

ตารางที่ 5 ผลบวกกำลังสองของความคลาดเคลื่อนเฉลี่ย $\left(\times 10^{-5}\right)$ เมื่อนำเมทริกสหสัมพันธ์แรกเริ่มไปสร้างเวกเตอร์ตัวแปรสุ่มจำนวน 10^6 เวกเตอร์ ของแต่ละอัลกอริทึม กรณีศึกษาที่ 2

Algorithm	**Distribution**	**Mean of the probability distribution**				
		0.075	**0.05**	**0.025**	**0.01**	**0.005**
N	Exponential	9.4382	11.6214	*	*	*
	Half-Normal	8.2558	6.7145	*	*	*
NK	Exponential	5.3826	4.9964	7.9258	4.5455	2.8280
	Half-Normal	6.0170	6.9817	4.3827	6.7347	7.2907

* ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม.

ตารางที่ 6 ผลบวกกำลังสองของความคลาดเคลื่อนเฉลี่ย $\left(\times 10^{-4}\right)$ เมื่อนำเมทริกสหสัมพันธ์แรกเริ่มไปสร้างเวกเตอร์ตัวแปรสุ่มจำนวน 10^6 เวกเตอร์ ของแต่ละอัลกอริทึม กรณีศึกษาที่ 3

Algorithm	Distribution	Mean of the probability distribution				
		0.075	0.05	0.025	0.01	0.005
N	Exponential	*	2.9117	2.2204	2.0004	2.5008
	Half-Normal	*	*	3.1875	2.0856	2.9897
NK	Exponential	*	2.4860	2.9609	1.8329	2.2696
	Half-Normal	*	*	2.6811	2.7266	2.3392

* ไม่สามารถหาเมทริกซ์สหสัมพันธ์แรกเริ่มได้ภายในเวลา 1 ชม.

ตารางที่ 7 เปรียบเทียบเมทริกซ์สหสัมพันธ์แรกเริ่ม

กรณี	การแจกแจงความน่าจะเป็นตามขอบ	เมทริกซ์สหสัมพันธ์ที่ต้องการ Σ_X	เมทริกซ์สหสัมพันธ์แรกเริ่ม Σ_Z	
			Niavarani, Smith's method	NK
1	$X_1 \sim Discrete\ U(1,10)$ $X_2 \sim Exp(10)$	$\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5710 \\ -0.5710 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5740 \\ -0.5740 & 1 \end{bmatrix}$
2	$X_i \sim Binomial(3,0.5)$ $i=1,2,3$	$\begin{bmatrix} 1 & 0.2 & -0.8 \\ 0.2 & 1 & 0.2 \\ -0.8 & 0.2 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2278 & -0.8880 \\ 0.2278 & 1 & 0.2226 \\ -0.8880 & 0.2226 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2231 & -0.8923 \\ 0.2231 & 1 & 0.2323 \\ -0.8923 & 0.2323 & 1 \end{bmatrix}$
3	$X_i \sim \Gamma(14.4,0.034424)$ $i=1,2,3,4$	$\begin{bmatrix} 1 & 0.7 & 0.5 & -0.9 \\ 0.7 & 1 & 0.7 & -0.6 \\ 0.5 & 0.7 & 1 & -0.3 \\ -0.9 & -0.6 & -0.3 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.6934 & 0.4977 & -0.9054 \\ 0.6934 & 1 & 0.6946 & -0.6054 \\ 0.4977 & 0.6946 & 1 & -0.3054 \\ -0.9054 & -0.6054 & -0.3054 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.7027 & 0.5063 & -0.9093 \\ 0.7027 & 1 & 0.7066 & -0.6220 \\ 0.5063 & 0.7066 & 1 & -0.3089 \\ -0.9093 & -0.6220 & -0.3089 & 1 \end{bmatrix}$

ตารางที่ 8 เปรียบเทียบผลบวกกำลังสองของความคลาดเคลื่อน

	$\hat{\Sigma}_X$ ของข้อมูล 10^6 เวกเตอร์ที่สร้างมาจากการใช้ Σ_Z		SSE	
กรณี	Niavarani Smith's method	NK	Niavarani, Smith's method	NK
1	$\begin{bmatrix} 1 & -0.5001 \\ -0.5001 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.4982 \\ -0.4982 & 1 \end{bmatrix}$	0.000021	0.000003
2	$\begin{bmatrix} 1 & 0.1980 & -0.7941 \\ 0.1980 & 1 & 0.2022 \\ -0.7941 & 0.2022 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.1947 & -0.7980 \\ 0.1947 & 1 & 0.2041 \\ -0.7980 & 0.2041 & 1 \end{bmatrix}$	0.000067	0.000049
3	$\begin{bmatrix} 1 & 0.7020 & 0.4984 & -0.8838 \\ 0.7020 & 1 & 0.7009 & -0.5979 \\ 0.4984 & 0.7009 & 1 & -0.3076 \\ -0.8838 & -0.5979 & -0.3076 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.7003 & 0.5031 & -0.8829 \\ 0.7003 & 1 & 0.7037 & -0.6070 \\ 0.5031 & 0.7037 & 1 & -0.3033 \\ -0.8829 & -0.6070 & -0.3033 & 1 \end{bmatrix}$	0.000728	0.000374

อภิปรายและสรุปผลการวิจัย

จากการศึกษาการกำหนดเมทริกซ์สหสัมพันธ์แรกเริ่มโดยใช้อัลกอริทึม N และอัลกอริทึม NK พบว่าการใช้อัลกอริทึม NK ให้ค่าผลบวกกำลังสองของความคลาดเคลื่อนเฉลี่ยที่ใกล้เคียงกับอัลกอริทึม N แต่อัลกอริทึม NK ใช้เวลาในการประมวลผลเฉลี่ยน้อยกว่าอัลกอริทึม N สรุปได้ว่าอัลกอริทึมที่มีการคัดเลือกเมทริกซ์เริ่มต้นใช้เวลาในการประมวลผลน้อยกว่าอัลกอริทึมที่ไม่มีการคัดเลือกเมทริกซ์เริ่มต้น

จากการเปรียบเทียบในตารางที่ 7 พบว่าเมทริกซ์สหสัมพันธ์แรกเริ่มมีค่าใกล้เคียงกันไม่มาก ซึ่งเมื่อนำเมทริกซ์สหสัมพันธ์แรกเริ่มได้ที่ได้มาสร้างเวกเตอร์สุ่มจำนวน 10^6 เวกเตอร์ พบว่าได้ค่าผลบวกกำลังสองของความคลาดเคลื่อนที่น้อยกว่าวิธีเดิม ซึ่งอาจจะเนื่องมาจากการเพิ่มจำนวนเวกเตอร์สุ่มจาก 5000 เวกเตอร์ เป็น 50000 เวกเตอร์ โดยในงานวิจัยนี้เพียงแค่ศึกษาการใช้การแจกแจงเลขชี้กำลังและการแจกแจงครึ่งปรกติในการสร้างสมาชิกในเมทริกซ์ $\mathbf{D}$ ซึ่งอาจมีวิธีอื่นหรืออาจมีการแจกแจงที่ใช้เวลาในการประมวลผลน้อยกว่านี้

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ ผศ.ดร.พัทธ์ชนก ศรีสุรเดชชัย อาจารย์ที่ปรึกษาที่ให้คำปรึกษาและคอยช่วยเหลือตลอดระยะเวลาในการทำวิจัย ขอขอบคุณเพื่อนๆสำหรับคำแนะนำรวมถึงกำลังใจ และขอบคุณคณะวิทยาศาสตร์และเทคโนโลยี สาขาสถิติประยุกต์ มหาวิทยาลัยธรรมศาสตร์ที่ให้ความอนุเคราะห์ด้านสถานที่ในการทำงาน

เอกสารอ้างอิง

Avramidis AN, Channouf N, Ecuyer L. Efficient Correlation Matching for Fitting Discrete multivariate distributions with arbitrary marginal and normal-copula dependence. INFORMS Journal on Computing 2009; 21(1): 88-106.

Avramidis AN. Constructing discrete unbounded distributions with Gaussian-copula dependence and given rank correlation. INFORMS Journal on Computing 2013; 26(2): 269-79.

Cario MC, Nelson BL. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technique Report, Department of Industrial Engineering and Management Science, Northwestern University, Evanston, Illinois, 1997.

Chen HF. Initialization for NORTA: generation of random vectors with specified marginals and correlations. INFORMS Journal on Computing2 001; 13(4): 312-31.

Clemen RT, Reilly T. Correlations and copulas for decision and risk analysis. Management Science 1999; 45(2): 208-24.

Corner JL, Kirkwood CW The magnitude of errors in proximal multiattribute decision analysis with probabilistically dependent attributes. Management Science 1996; 42(7): 1033-42.

Das SR, Fong G, Geng G. Impact of correlated default risk on credit portfolios. Journal of Fixed Income 2001; 11(3): 9-19.

Hull JC. Dealing with dependence in risk simulations. Operational Research Quarterly 1977; 28(1): 201-13.

Hill RR, Reilly CH. The effects of coefficient correlation structure in two-dimensional knapsack problems on solution procedure performance. Management Science 2000; 46(2): 302-17.

Kiureghian AD, Liu P. Structural reliability under incomplete probability information. Journal of Engineering Mechanics 1986; 112(1): 85-104.

Li HS, Lü ZZ, Yuan XK. Nataf transformation based point estimate method. Chinese Science Bulletin 2008; 53(17) :2586-92.

Li ST, Hammond JL. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. IEEE Transactions on Systems, Man and Cybernetics 1975; 5(5): 557-61.

Liu P, Kiureghian AD. Multivariate distribution models with prescribed marginals and covariances. Probabilistic Engineering Mechanics 1986; 1(2): 105-12.

Lurie PM, Goldberg MS. An approximate method for sampling correlated random variables from partially-specified distributions. Management Science 1998; 44(2): 203-18.

Marida KV. A translation family of bivariate distributions and Fréchet's bounds. Sankhya 1970; 32(1): 119-22.

Niaki STA, Abbasi B. Generating correlation matrices for normal random vectors in NORTA algorithm using artificial neural networks. journal of Uncertain Systems 2008; 2(3): 192-201.

Niavarani MR, Smith AJR. Modeling and generating multi-variate-attribute random vectors using a new simulation method combined with NORTA algorithm. Journal of Uncertain Systems 2013; 7(2): 83-91.

Xiao Q. Generating correlated random vector involving discrete variables. Communications in Statistics-Theory and Methods 2017; 46(4): 1594-605.

Yahav I, Shmueli G. On generating multivariate Poisson data in management science applications. Applied Stochastic Models in Business and Industry 2012; 28(1): 91-102.