

งานวิจัยนี้ได้ศึกษาตัวประมาณเมตริกซ์สหสัมพันธ์ด้วยวิธีเบส เมื่อข้อมูลเกิดการสูญหายแบบเท่ากันโดยมีจุดประสงค์เพื่อแก้ปัญหาการขาดคุณสมบัติไม่เป็นลบแน่นอน (Positive Semi-Definite) ของค่าประมาณเมตริกซ์สหสัมพันธ์ ซึ่งตัวประมาณเมตริกซ์สหสัมพันธ์ด้วยวิธีเบสในงานวิจัยนี้ แบ่งออกเป็น 2 แบบตามการกำหนดการแจกแจงเริ่มแรก (Prior Distribution) คือ การแจกแจงแบบสม่ำเสมอร่วม (Joint Uniform Distribution) และการแจกแจงแบบสม่ำเสมอส่วนริม (Marginal Uniform Distribution) โดยในแต่ละแบบจะอาศัยเทคนิคการจำลองจากวิธีมาร์คอฟเชนมอนติคาร์โล (Markov Chain Monte Carlo: MCMC) ด้วยอัลกอริทึมของเมโทรโพลิซิส-เฮสติ้ง (Metropolis-Hasting Algorithm: MH) ซึ่งจะแบ่งวิธีการจำลองออกเป็น 2 วิธี ตามประเภทของการแจกแจงพรีพอสเชน (Proposal Distribution) คือ อินดิเพนเด้นเชน (Independence Chain) และ บอลวอล์ค (Ball Walk) โดยในงานวิจัยนี้จะศึกษาในกรณีที่ข้อมูลมี 3 มิติ และสมมติให้ข้อมูลมีการแจกแจงปกติหลายตัวแปร ด้วยเวกเตอร์ค่าเฉลี่ย $\mu = 0$ และเมตริกซ์

$$\text{ความแปรปรวนร่วม } \Sigma \text{ ขนาด } 3 \times 3 \text{ ที่มีคุณสมบัติไม่เป็นลบแน่นอนและสมมาตร เมื่อ } \Sigma = R = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix}$$

โดยกำหนดให้เวกเตอร์ขององค์ประกอบเหนือเส้นทแยงมุมของเมตริกซ์สหสัมพันธ์จริง (r) เป็น **0.0, 0.1, 0.5 และ 0.9** ตามลำดับ และกำหนดให้ขนาดตัวอย่างที่ใช้ในการประมาณองค์ประกอบเมตริกซ์สหสัมพันธ์ (m) เป็น 3, 10, 30 และ 100 ตามลำดับ

จากผลการวิจัย เมื่อใช้ระยะทางยูคลิดีียน (Euclidean Distance) เทียบกับเมตริกซ์สหสัมพันธ์จริง เป็นเกณฑ์ในการเปรียบเทียบ สามารถสรุปได้ว่า ค่าประมาณเมตริกซ์สหสัมพันธ์ด้วยวิธีเบส เมื่อกำหนดการแจกแจงเริ่มแรกเป็นการแจกแจงแบบสม่ำเสมอร่วมจะประมาณค่าได้ดี เมื่อข้อมูลมีความสัมพันธ์กันในระดับสูงๆ ส่วนค่าประมาณเมตริกซ์สหสัมพันธ์ด้วยวิธีเบส เมื่อกำหนดให้การแจกแจงเริ่มแรกเป็นการแจกแจงแบบสม่ำเสมอส่วนริม นั้นจะประมาณค่าได้ดีเมื่อข้อมูลไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์กันน้อยๆ

ส่วนการเปรียบเทียบความเร็วในการจำลองจากวิธีมาร์คอฟเชนมอนติคาร์โล ด้วยอัลกอริทึมของเมโทรโพลิซิส-เฮสติ้ง เมื่อกำหนดประเภทของการแจกแจงพรีพอสเชนต่างกันนั้น พบว่าการกำหนดประเภทการแจกแจงพรีพอสเชนเป็นแบบอินดิเพนเด้นเชน และกำหนดการแจกแจงพรีพอสเชนเป็นการแจกแจงแบบสม่ำเสมอในขอบเขตของเมตริกซ์สหสัมพันธ์ที่มีคุณสมบัติไม่เป็นลบแน่นอนจะจำลองได้เร็ว เมื่อขนาดตัวอย่างที่ใช้ในการประมาณองค์ประกอบเมตริกซ์สหสัมพันธ์ มีจำนวนน้อยๆ (วิธีการเคลื่อนที่ตัวอย่างจะกว้าง) ส่วนเมื่อกำหนดประเภทการแจกแจงพรีพอสเชนเป็นแบบบอลวอล์ค และกำหนดการแจกแจงพรีพอสเชนเป็นการแจกแจงแบบสม่ำเสมอในบอล จะจำลองได้เร็ว เมื่อขนาดตัวอย่างที่ใช้ในการประมาณองค์ประกอบเมตริกซ์สหสัมพันธ์ มีจำนวนมากๆ (วิธีการเคลื่อนที่ตัวอย่างจะแคบ)

This research studied Bayes' estimation for a correlation matrix with equal loss data. The main objective of this study was to handle lack of the positive semi-definite property of correlation matrices. The correlation matrices under this study were classified into two types according to their prior distribution, namely a joint uniform distribution and a marginal uniform distribution. The Markov Chain Monte Carlo (MCMC) simulation with Metropolis-Hasting algorithm (MH) was used and was classified into two methods according to the types of proposal distribution, namely independence chain and Ball Walk. This research specifically studied the case where the data are three dimensional and are a multivariate normal distribution with mean vector $\mu = \mathbf{0}$ and variance covariance matrix Σ , which is a

3×3 symmetric and positive semi-definite, where $\Sigma = \mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix}$. We let upper-diagonal elements

vector of true correlation matrix (\mathbf{r}) be **0.0, 0.1, 0.5** and **0.9** respectively. The sample size used to estimate for each elements of correlation matrix (m) were 3, 10, 30 and 100 respectively.

The result of this study can be summarized as follows. Taking Euclidean distance from the true correlation matrix as performance measure, a Bayes' estimate from the joint uniform prior gives a better estimate than that from the marginal uniform prior when the data are highly correlated. On the contrary, a Bayes' estimate from the marginal uniform prior gives a better estimate than that from the joint uniform prior when the data are uncorrelated or weakly correlated.

The result of the comparison between the two types of proposal distribution by MCMC simulation with MH algorithm shows that a Bayes' estimate from independence chain, that is assumed uniform on sets of symmetric positive semi-definite correlation matrices, is faster of the simulation than Ball Walk when small sample size are estimated each elements of correlation matrix (wide sample path). On the contrary, a Bayes' estimate from Ball Walk, that is assumed uniform on ball, is faster of the simulation than independence chain when large sample size (narrow sample path).