BAYESIAN ANALYSIS FOR A GRAPHICAL T-MODEL

A. Bekker[1,3], J.T. Ferreira[1,3], J. Pillay[1,3] , and M. Arashi [2]

[1] Department of Statistics, University of Pretoria, South Africa, (e-mails: andriette.bekker@up.ac.za, u18067434@tuks.co.za johan.ferreira@up.ac.za)

[2] Department of Statistics, Ferdowsi University of Mashhad, Iran (e-mail: arashi@um.ac.ir)

[3] Centre of Excellence in Mathematical and Statistical Sciences, Johannesburg, South Africa

ABSTRACT: Modelling noisy data in a network context remains an unavoidable obstacle; fortunately, random matrix theory may comprehensively describe network environments effectively. Thus it necessitates the probabilistic characterisation of these networks (and accompanying noisy data) using matrix variate models. Denoising network data using a Bayes approach is not common in surveyed literature. Thus we briefly introduce a new matrix-variate t model in a prior sense for the noise process following the Gaussian graphical network, for the cases when the assumption of normality is violated in the model and cases when Gaussian distributions is no longer sufficient to explain variation in the data. We investigate the performance of this matrix-variate t distribution applied to a network setting within a Bayesian context. Calculation and approximation of the resulting posterior are of interest to assess the considered model's network centrality measures, which is illustrated using real-life stock price data.

KEYWORDS: adjacency matrix, Bayesian estimation, Gaussian graphical model, matrix-variate t, stock price data.

1 Introduction

Let G_t be a sequence of directed networks for $t = 1, \ldots, T$ for $T \in \mathbb{N}$. Assume that the number of nodes do not change with respect to t, but the number of edges can. Assume that each of the nodes bears a stationary time series of variables that estimates a sequence of networks G_t at time t. Then an adjacency matrix is estimated for G_t at each time index t, say $\boldsymbol{Y}_t$. A stationary time series implies that network structure itself at time t is nothing more than a deviation from an underlying adjacency matrix $\boldsymbol{B}$ independent of time t. In other words, the true graphical network structure is stationary. $\boldsymbol{Y}_t$ is thus viewed as 'noisy copy' of $\boldsymbol{B}$ given by:

$$\boldsymbol{Y}_t = \boldsymbol{B} + \boldsymbol{E}_t \quad \text{for } t = 1, \ldots, T. \tag{1}$$

$\boldsymbol{E}_t : n \times n$ is a random error term, independent and identically distributed for all $t = 1, \ldots, T$. The matrix-variate Gaussian distribution is fundamental for inference, but is sometimes inadequate for modelling populations where the matrix variate-t distribution may be a better fit. There is extensive literature around a multivariate Gaussian distribution of errors. Articles that date back as early as the classical linear models (Arnold, 1979) to relatively recent ones on engineering processes (Amiri *et al.*, 2018), with recent contributions including the work by Billio *et al.*, 2021. Instead, a t distribution seems a suitable choice to characterise error. Thus, consider $\boldsymbol{E}_t$ as matrix-variate t distributed with corresponding probability density function (pdf), then $\boldsymbol{E}_t \sim t_{n,n}(\boldsymbol{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ and the pdf of $\boldsymbol{E}_t$ is given by,

$$f(\boldsymbol{E}_t|\nu, \boldsymbol{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \frac{\Gamma_n\left(\frac{\nu+2n-1}{2}\right)}{\pi^{\left(\frac{n^2}{2}\right)}\Gamma_n\left(\frac{\nu+n-1}{2}\right)} |\boldsymbol{\Sigma}_1|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_2|^{-\frac{n}{2}} |\boldsymbol{I}_n + \boldsymbol{\Sigma}_1^{-1}\boldsymbol{E}_t\boldsymbol{\Sigma}_2^{-1}\boldsymbol{E}_t'|^{-\frac{\nu+2n-1}{2}}, \tag{2}$$

where $\Gamma_n(\cdot)$ is the multivariate gamma function. By the linearity property of a matrix-variate t distribution, (1) implies that $\boldsymbol{Y}_t \sim t_{n,n}(\nu, \boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ and is consequently called the matrix-variate t model. Since $\boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ and ν are unknown, they must be estimated. Bayesian methodology for estimating the unknown parameters is followed and implementing the matrix-variate Gamma and inverse matrix-variate Gamma as priors for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ respectively, and a new graphical t-model as a result. Applying the methodology reveals a clear discrepancy between estimates from raw data and the Bayesian approach, which highlights the misleading impact that noise in data has and how it may lead to more grave consequences for any analysis built upon said noise.

2 A new graphical t-model construction

Assume that the prior density functions are mutually independent. The joint pdf $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \nu)$ is then proportional to :

$$\gamma^{a_\gamma - 1}\nu^{a_\nu - 1}\exp\left[-\frac{1}{2}\text{tr}\left(\text{vec}(\boldsymbol{B})'(\boldsymbol{\Omega}_2 \otimes \boldsymbol{\Omega}_1)^{-1}\text{vec}(\boldsymbol{B})\right) - \frac{\gamma}{b_\gamma} - \frac{\nu}{b_\nu}\right] f(\boldsymbol{\Sigma}_1|\gamma) f(\boldsymbol{\Sigma}_2|\gamma), \tag{3}$$

where $f(\mathbf{\Sigma}_1|\gamma), f(\mathbf{\Sigma}_2|\gamma)$ are some conditional prior pdfs of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively. It follows $\boldsymbol{B}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \gamma, \nu$ has likelihood function equal to:

$$\prod_{t=1}^{T} \frac{\Gamma_n(\frac{\nu+2n-1}{2})}{\pi^{(\frac{n^2}{2})}\Gamma_n(\frac{\nu+n-1}{2})} |\mathbf{\Sigma}_1|^{-\frac{n}{2}} |\mathbf{\Sigma}_2|^{-\frac{n}{2}} |\boldsymbol{I}_n + \mathbf{\Sigma}_1^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})\mathbf{\Sigma}_2^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})'|^{-\frac{\nu+2n-1}{2}}. \tag{4}$$

From (3) and (4) the posterior pdf follows as

$$\begin{aligned} &\prod_{t=1}^{T} \frac{\Gamma_n(\frac{\nu+2n-1}{2})}{\pi^{(\frac{n^2}{2})}\Gamma_n(\frac{\nu+n-1}{2})} |\mathbf{\Sigma}_1|^{-\frac{n}{2}} |\mathbf{\Sigma}_2|^{-\frac{n}{2}} |\boldsymbol{I}_n + \mathbf{\Sigma}_1^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})\mathbf{\Sigma}_2^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})'|^{-\frac{\nu+2n-1}{2}} \\ &\times \gamma^{a_\gamma - 1}\nu^{a_\nu - 1}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathrm{vec}(\boldsymbol{B})'(\mathbf{\Omega}_2 \otimes \mathbf{\Omega}_1)^{-1}\mathrm{vec}(\boldsymbol{B})\right) - \frac{\gamma}{b_\gamma} - \frac{\nu}{b_\nu}\right] \\ &\times f(\mathbf{\Sigma}_1|\gamma) f(\mathbf{\Sigma}_2|\gamma). \end{aligned} \tag{5}$$

For this paper, consider the matrix-variate and inverse matrix-variate Gamma distributions, i.e., $\mathbf{\Sigma}_1 \sim MG_n(\delta_1, \beta, (\gamma\mathbf{\Phi}_1)^{-1})$ and $\mathbf{\Sigma}_2 \sim IMG_n(\delta_2, \beta, (\gamma\mathbf{\Phi}_2)^{-1})$ as priors. Notice that the scalar shape parameter β can be fixed, or have a prior imposed on it also. Either way the estimation procedure unaffected. As is usual with Bayesian estimation, an observed matrix $\boldsymbol{B}_i$ from the posterior distribution is an estimate of the true adjacency matrix $\boldsymbol{B}$ - thus, the average of a sample estimates $\boldsymbol{B}$. To simulated a sample, the Gibbs sampling algorithm is used.

3 Application and evaluation

The methodology is applied to the weekly stock prices of 70 European firms, resulting in 105 observations. Granger causality hypothesis tests are applied pairwise for week t. The resulting test statistics belong in a matrix that is an observed $\boldsymbol{Y}_t$ [*]. We employ well-known centrality measures, such as a graph's degree, closeness, eigen centrality, and betweenness, to evaluate a matrix variate estimator. These measures are univariate scores that measure a node's influence in a graph.

The results from the application are shown in Figure 1[†]. It is observed that there are clear discrepancies between the different estimators, with particular

[*] Data provided by Prof. M. Billio, University of Venice, Italy.

[†] The simulations were run on MATLAB R2022b on University of Pretoria server with 501Gb of RAM and 48 cores. Runtime for simulations was 16h excluding time to compute Granger causality test statistics.

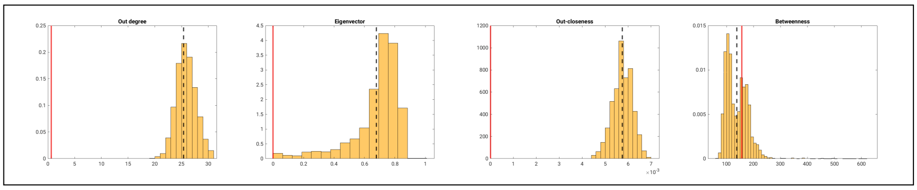

Figure 1: Estimated centrality measures: The solid red line and dashed black lines represent the averages of the raw data, and methodology respectively.

attention to the out-degree, out-closeness, and eigencentrality. The raw data seems to underestimate the centrality measures. In other words, the discrepancies highlight how noise left in data may jeopardise the validity and reliability of analysis built on data.

4 Acknowledgements

This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa (SA), grant RA201125576565 & RA211204653274, nr 145681 & 151035; NRF ref. SRUG2204203865, the Centre of Excellence in Mathematical and Statistical Sciences, based at the University of the Witwatersrand (SA). The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

References

AMIRI, AMIRHOSSEIN, GHASHGHAEI, REZA, & MALEKI, MOHAMMAD REZA. 2018. On the effect of measurement errors in simultaneous monitoring of mean vector and covariance matrix of multivariate processes. *Transactions of the Institute of Measurement and Control*, **40**(1), 318–330.

ARNOLD, STEVEN F. 1979. Linear models with exchangeably distributed errors. *Journal of the American statistical Association*, **74**(365), 194–199.

BILLIO, MONICA, CASARIN, ROBERTO, COSTOLA, MICHELE, & IACOPINI, MATTEO. 2021. A matrix-variate t model for networks. *Frontiers in Artificial Intelligence*, 49.