

Gibbs Sampling in Inference of Copula Gaussian Graphical Model Adapted to Biological Networks

V. PURUTÇUOĞLU* AND H. FARNOUDKIA

Middle East Technical University, Department of Statistics, Ankara, Turkey

Markov chain Monte Carlo methods (MCMC) are iterative algorithms that are used in many Bayesian simulation studies, where the inference cannot be easily obtained directly through the defined model. Reversible jump MCMC methods belong to a special type of MCMC methods, in which the dimension of parameters can change in each iteration. In this study, we suggest Gibbs sampling in place of RJMCMC, to decrease the computational demand of the calculation of high dimensional systems. We evaluate the performance of the suggested algorithm in three real benchmark datasets, by comparing the accuracy and the computational demand with its strong alternatives, namely, birth-death MCMC, RJMCMC and QUIC algorithms. From the comparative analyses, we detect that Gibbs sampling improves the computational cost of RJMCMC without losing the accuracy.

DOI: [10.12693/APhysPolA.132.1112](https://doi.org/10.12693/APhysPolA.132.1112)

PACS/topics: 02.50–R, 02.50–TT

1. Introduction

The conditional dependence between variables is one of the most important issues that has been investigated within last twenty years through the Bayesian approach. The major challenge in this field is that, in some cases, the number of parameters resulting in the true model is unknown. In order to solve the underlying uncertainty, the Bayesian approach is used and Markov chain Monte Carlo (MCMC) methods are the most popular approaches to infer the associated parameter estimation.

In general, MCMC methods are iterative algorithms which generate, at each iteration, a random variable from the posteriors of the parameters, yielding a Markov chain, from which an estimator is computed [1]. There are several ways to overcome the problem of the model selection and to infer associated model parameters simultaneously, if these methods are implemented in the construction of the biological networks, modelled by the Gaussian copula graphical model.

The reversible jump Markov chain Monte Carlo (RJMCMC) method is one of the well-known approaches to unravel both challenges at the same time [2]. RJMCMC is the modified version of the Metropolis-Hasting method, which provides jumps between spaces of different dimensionality. In this algorithm, the move from (k, θ^k) to $(k', \theta^{k'})$ is not always possible because of the modality of the Metropolis-Hasting calculation. Here, k and k' refer to the number of edges, i.e., the dimension of the precision matrix, and the proposal number of edges, respectively, and θ^k and $\theta^{k'}$ indicate the associated precision matrix under the k and proposal k' edges, in order. Thereby, the acceptance probability for this movement is computed as

$$R_{k,k'} = \frac{P(k', \theta^{k'} | y) \tilde{q}(k, \theta^k | k', \theta^{k'})}{P(k, \theta^k | y) \tilde{q}(k', \theta^{k'} | k, \theta^k)}, \quad (1)$$

where y refers to the normal random variables of the current position. Finally, P shows the likelihood function and \tilde{q} presents the kernel density. Under the “dimension matching” condition, i.e., $\dim(\theta^{k'}, x) = \dim(\theta^k, y)$, where x and y are variables drawn from the proposal distribution \tilde{q}_1 , the acceptance probability is equivalent to

$$R_{k,k'} = \frac{P(k', \theta^{k'} | x)}{P(k, \theta^k | y)} \times \frac{\tilde{q}_1(k | k') \tilde{q}_2(x)}{\tilde{q}_1(k' | k) \tilde{q}_2(y)} \times \left| \frac{\partial(\theta^{k'}, x)}{\partial(\theta^k, y)} \right|. \quad (2)$$

In Eq. (2), $\tilde{q}_2(\cdot)$ refers to the kernel for the given random variable and $\left| \frac{\partial(\theta^{k'}, x)}{\partial(\theta^k, y)} \right|$ represents the determinant of the Jacobian matrix. In the study of Dobra and Len [3], this method is used to infer the parameters of the copula Gaussian graphical model (CGGM). In the application of CGGM for the construction of biological networks, it is found that RJMCMC has a long burn-in period [3, 4] and additionally, it needs to calculate the Jacobian term for each associated iteration, as in Eq. (2).

In order to decrease the computational demand and optimize the calculation, Mohammadi and Wit [4] suggest the birth-and-death method. RJMCMC with the split-merge approach have been also intensively studied in the work of Richardson and Green [5]. On the other hand, there are other alternatives of this method which have not been proposed for the biological networks. For instance, Carlin and Chib algorithm [6] and the Gibbs algorithm [7] are in this group. Hereby, in this study, we propose the Gibbs sampling to improve the computational demand of the estimates, without losing the accuracy. We compare the performance of all listed approaches in three benchmark real datasets and compare their outputs, based on both accuracy and computational efficiency.

Hence, in the following part, we initially describe CGGM and RJMCMC. Then, in Section 3, we explain their alternatives, together with the Gibbs sampling,

*corresponding author; e-mail: vpurutcu@metu.edu.tr

which we introduce first time, and present our comparative analysis with its strong alternatives. In Section 4, we conclude our findings and discuss the future works.

2. Methods

2.1. Copula Gaussian graphical model

Let we have a data matrix with p variables and n samples and we are interested in obtaining the relationship between these variables. In this kind of networks, which is common in social surveys and biological aspects, each variable is shown by a node in the graph and the conditional dependence between two nodes is shown by an undirected edge. Now, assume that the vector Y follows a p -dimensional multivariate normal distribution $N_p(0, K^{-1})$, where K is the inverse of covariance matrix. With n samples, the likelihood function is proportional to

$$p(Y^{1:n} | K) \propto |K|^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (K^T U) \right\}, \quad (3)$$

where $|\cdot|$ shows the determinant of the given matrix and U is the trace of the Y'' matrix. So a graphical model with V having E number of edges, i.e., (V, E) , and p number of nodes from a p -dimensional multivariate normal distribution with a zero mean vector and K^{-1} variance-covariance matrix, i.e., $N_p(0, K^{-1})$, is called the Gaussian graphical model (GGM).

This model can be also seen as the probabilistic alternative of the ordinary differential equation model [8], the extended version of the probabilistic Boolean model [9] and the lasso-type of the generalized additive model [10]. In GGM, if the normality assumption does not hold for the data matrix, the *copula* can solve the problem by combining data in a way that their joint distribution is Gaussian with the same covariance matrix. For binary and ordinal categorical data, Muthen [11] introduces a continuous latent variable Z , by defining some thresholds $\tau_v = (\tau_{v,0}, \tau_{v,1}, \dots, \tau_{v,\omega_v})$ with $-\infty = \tau_{v,0} < \tau_{v,1} < \dots < \tau_{v,\omega_v} = \infty$. So, the state of the node under the j th model can be described by

$$y_v^j = \sum_{l=1}^{\omega_v} l \times 1_{\tau_{v,l-1} < z_v^j \leq \tau_{v,l}}, \quad j = 1, 2, \dots, n. \quad (4)$$

For an n -dimensional system, the relationship between Y_{ij} and Z_{ij} under the i th and j th node satisfies the constraint $y_{ij} < y_{ik} \rightarrow z_{ij} < z_{ik}, z_{ij} < z_{ik} \rightarrow y_{ij} \leq y_{ik}$. Thus, the correlation matrix between the i th and the j th node is defined as

$$Y_{i,j}(K) = \frac{(k^{-1})_{i,j}}{\sqrt{(k^{-1})_{i,i}(k^{-1})_{j,j}}}, \quad (5)$$

and $Z_V \sim N_p(0, K^{-1})$, which has a one-to-one correspondence with observed data.

2.2. RJMCMC with birth-and-death steps

In this kind of RJMCMC, the change in the dimension of the model parameters, resulting in the model itself, is done by births and deaths steps, which are proposed

with probabilities P_B and P_D , respectively. Given $Y_s = (k, \theta_{1:k})$ showing the state of the system under the s th iteration of the MCMC algorithm, when the system is k -dimensional and has the precision matrix, Green [2] suggests the following calculation:

- The birth step is computed with a probability P_B by drawing the graph G from the density $G \sim q$ as θ_{k+1} . So the dimension of the state Y becomes $\dim(Y_s) : k \rightarrow (k + 1)$.
- The death step is found with a probability P_D by selecting one of θ_j ($j = 1, \dots, k$) and dropping the corresponding parameter from $\theta_{1:k}$. So the dimension of Y becomes $\dim(Y_s) : k \rightarrow k - 1$.

Hence, the acceptance probability for the birth move in a k -component configuration is $\min(1, A)$ via

$$A = \frac{\pi(k + 1, \theta^k)}{\pi(k, \theta^k)} \times \frac{P_D}{P_B q(\theta)}, \quad (6)$$

in which π is a stationary distribution, q presents the proposal kernel. Thus, the acceptance probability for the death is found from $\min(1, 1/A)$.

2.3. RJMCMC with split-merge steps

In this method, there is a deterministic invertible transformation function $t : R^2 \rightarrow R^2$, playing the main role as below:

- The split step is calculated with a probability P_S by initially choosing one of the indices $i, j = 1, \dots, k$ and a rough G from the proposal kernel q , i.e., $G \otimes q$. Then, we put them into the t function $R^2 \rightarrow R^2$. The result has two components formed by the combination of j and G under t . So $\dim(Y_s) : k \rightarrow (k + 1)$.
- The merge is obtained with a probability P_M by initially selecting 2 indices $i, j = 1, \dots, k$, and then by putting them into the invertible t function $R^2 \rightarrow R^2$. Finally, the first component of $t^{-1}(\theta_i, \theta_j)$ is assigned in place of two parameters, so that the dimension of the state Y under the s th MCMC iteration moves from k to $k + 1$, i.e., $\dim(Y_s) : k \rightarrow (k - 1)$.

Hence, the proposal probability density function depends on both $P_S P_M, q$ and the Jacobian of t .

3. RJMCMC alternatives

As discussed shortly in the introduction, there are also some alternative approaches for RJMCMC. These are the Carlin Chib algorithm [6], the birth-death MCMC (BD-MCMC), a special case of the Gibbs sampling, and finally, the quadratic approximation for the sparse inverse covariance estimation, shortly called as the (QUIC) method, which does not use the Bayesian approach. Below, we present the mathematical details of each method.

3.1. Birth-and-death MCMC

This method is based on the continuous time approach, where the dimension of the parameter is not fixed. In this approach, new components are born according to the Poisson process with a rate λ_B and the i th component in a k -component configuration, which dies with a rate $\lambda_D(i) = \frac{\pi(k-1, \theta_{1:i-1}, \theta_{i+1:k})}{\pi(k, \theta_{1:k})}$. In this expression, $\pi(\cdot)$ is the density kernel, when $\theta_{1:k}$ implies the first k parameters and $q(\theta_i)$ represents the proposal kernel for the i th component of the parameter θ , as used beforehand. The choice of the birth and death rates determines the birth-death process and is made in such a way, that the stationary distribution is precisely the posterior distribution of interest. Contrary to the RJMCMC approach, the moves between models are always accepted, which makes BDMCMC extremely efficient.

3.2. Carlin-Chib algorithm

In the application of the MCMC technology to any problem involving a choice between K competing the Bayesian model specification, M is defined an integer-value parameter which indexes the model collection. The Carlin-Chib algorithm [6] shows how the Gibbs sampling methodology may be a special method to choose across finite collections of models, without destroying the convergence.

Suppose that $f(y|\theta_j, M=j)$ for $j=1, \dots, k$ is the corresponding likelihood of the model j and $P(\theta_j, |M=j)$ is the prior distribution of the parameter under the model j . Here, y is independent on $\theta_{i \neq j}$, given that $M=j$. As it is mentioned before, M is a model indicator and for the given M , various θ'_j s are assumed to be completely independent. By defining $\pi_i = P(M=j)$, such that $\sum_{j=1}^k \pi_j = 1$, the joint distribution of y and θ , when $M=j$ is as below.

$$P(y, \theta, M=j) =$$

$$f(y, \theta_j, M=j) \times \pi_j \times \left\{ \prod_{i=1}^k P(\theta_i | M=j) \right\}. \quad (7)$$

The following equation shows the full conditional independence of each θ_j and M via

$$P(\theta_j | \theta_{i \neq j}, M, y) \propto \begin{cases} f(y | \theta_j, M=j) & \text{for } M=j, \\ P(\theta_j | M \neq j) & \text{for } M \neq j, \end{cases} \quad (8)$$

where $P(\theta_j | M \neq j)$ is called ‘‘pseudoprior’’. When $M=j$, we generate the graph from the usual model of the full conditional distribution and when $M \neq j$, we generate it from the linking density.

Hence for the model M , we have $P(M=j | \theta, y) = \frac{f(y|\theta_j, M=j) \prod_{i=1}^k p(\theta_i | M=j) \pi_j}{\sum_{n=1}^k f(y|\theta_n, M=n) \prod_{i=1}^k p(\theta_i | M=n) \pi_n}$.

In the usual condition, the algorithm produces samples from the correct joint posterior distribution. In particular, the ratio is a simple estimate to compute the Bayes factor between any two of models, while g denotes the number of samples. Thus, $M^{(g)} = j$ means the j th model for the g th sample.

3.3. Gibbs sampling

By using the Bayes theorem, a complete model for a joint density for $j=1, 2, \dots$ can be written as

$$p(y, \theta^j, k) = p(y, \theta^k, k) p(\theta^1 | \theta^2) \dots$$

$$p(\theta^{k-1} | \theta^k) p(\theta^{k+1} | \theta^k) p(\theta^{k+2} | \theta^{k+1}) \dots \quad (9)$$

If we denote π_k as the prior distribution for the unknown dimension of a parameter k and $\pi_k(\theta^k)$ as the prior distribution for $\theta^k | k$, we can represent the joint distribution of the state y with a model parameter under the k dimension via $p(y, \theta^k, k) = p(y | \theta^k, k) \pi_k(\theta^k) \pi_k$.

Here, to move between dimensions, we have infinity choices that cause the precise probabilities, which cannot be found. To solve the problem, Walker [7] introduces an auxiliary variable u which helps us to have finite choices in order to move between dimensions. On the other hand, the latent variable u has a distribution in which $u=k$ with a probability q and $u=k+1$ with a probability $1-q$. Since u depends only on k , the complete model can be stated as $p(u, y, \theta^j, k) = p(u | k) p(y, \theta^j, k)$. Thereby, the steps of the algorithm can be listed as below:

1. Sample $\theta^{(k)}$ from $\pi_k(\theta^k | y, k)$, sample $\theta^{(k+1)}$ from $p(\theta^{(k+1)} | \theta^{(k)})$ and sample $\theta^{(k-1)}$ from $p(\theta^{(k-1)} | \theta^{(k)})$.
2. Sample u from some kind of a binomial distribution in which $p(u=k+1) = q$ and $p(u=k) = 1-q$.
3. For the given k , sample j , which will be the next k , from the distribution below:

$$j = k | u = k + 1 \propto$$

$$(1-q)p(y, \theta^{(k+1)}, k+1) p(\theta^{(k)} | \theta^{(k+1)}),$$

$$j = k + 1 | u = k + 1 \propto qp(y, \theta^{(k)}, k) p(\theta^{(k+1)} | \theta^{(k)}),$$

$$j = k | u = k \propto (1-q)p(y, \theta^{(k)}, k) p(\theta^{(k-1)} | \theta^{(k)}),$$

$$j = k - 1 | u = k \propto$$

$$qp(y, \theta^{(k-1)}, k-1) p(\theta^{(k)} | \theta^{(k-1)}).$$

In these expressions, the sampling strategy is simplified by the following equality.

$$P(\theta^{(k)} | \theta^{(k+1)}) \times \pi_{k+1}(\theta^{(k+1)}) =$$

$$P(\theta^{(k+1)} | \theta^{(k)}) \times \pi_k(\theta^{(k)}). \quad (10)$$

Thus we show its validity under the Gaussian Copula graphical model (see Supplementary Material for the proof). So the simplified version of the third step of the algorithm can be described as follows.

$$j = k | u = k + 1 \propto$$

$$(1-q)p(y, \theta^{(k+1)}, k+1) \pi(\theta^{(k+1)}),$$

$$j = k + 1 | u = k + 1 \propto qp(y, \theta^{(k)}, k) \pi(\theta^{(k)}),$$

$$j = k | u = k \propto (1-q)p(y, \theta^{(k)}, k) \pi(\theta^{(k-1)} | \theta^{(k)}),$$

$$j = k - 1 | u = k \propto qp(y, \theta^{(k-1)}, k - 1) \pi(\theta^{(k)} | \theta^{(k-1)}),$$

3.4. Quadratic approximation for sparse inverse covariance estimation

This algorithm [12] is suggested to estimate the inverse of a sparse co-variance matrix, where the data are Gaussian. In this calculation, there is a penalty term in the general formula, related to the sparsity of the related graph. By increasing the underlying term, the precision matrix, which shows the structure of the conditional dependence structure between nodes, becomes sparser.

Hereby, in the algorithm, let Y be a $(n \times p)$ -dimensional data matrix and the sample covariance matrix is denoted by $S = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{\mu})(y_k - \hat{\mu})^T$, where $\hat{\mu} = \frac{1}{n} \sum_{k=1}^n y_k$ and $(.)^T$ refers to the transpose of the given vector.

Given the regularization penalty term $\lambda > 0$, the regularized log-determinate is defined as $\arg \min \left\{ -\log |Y| + \text{tr}(SY) + \lambda \sum_{i,j=1}^p |Y_{ij}| \right\}$. Here, log and tr shows the logarithm and the trace of the matrix, respectively, and $|\cdot|$ denotes the determinant, as used previously.

Then, the algorithm computes the optimal λ by taking the $(p \times p)$ -dimensional empirical covariance matrix S , which is positive semi-definite, and the regularization matrix Λ , whose components are λ' , as inputs, and initializes Y on the first iteration via $Y_0 > 0$ in $t = 1, 2, \dots$, until Y_t converges to $\arg \min |Y|$.

4. Application

In order to evaluate the performance of the RJMCMC alternatives, we compare the results of the three benchmark real networks data. The first dataset is the Rochale dataset, shown in Table I. The second dataset is CellSignal data and the third one is related to the simulated MAPK-ERK dataset by Gillespie algorithm. In our analyses, we compare the F_1 - Score and the Matthew's correlation coefficient (MCC), whose expressions are given below, for RJMCMC, BDMCMC, QUIC and Gibbs, via the true network given in the study of Whittaker [13], which is shown in Table II. In this table, 1 indicates the existence of an edge between two related nodes and 0 shows the conditional independence between two nodes.

$$F_1 - \text{Score} = \frac{2TP}{2TP + FP + FN} \text{ and}$$

$$MCC = \frac{[(TP \times TN) - (FP \times FN)]}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (11)$$

In Eq. (12), TP is the number of truly found edges, FP presents the number of falsely found edges and FN denotes the number of edges that exist, but are not recognized. Finally, the perfection level of F_1 - Score is 1 and the range lies from 0 to 1.

On the other hand, the Matthew's correlation coefficient is also known as *phi coefficient* and turns a value between 0 and +1. A coefficient +1 represents a perfect prediction, 0 implies no better than a random prediction and 1 indicates the total disagreement between the prediction and the observation.

4.1. The Rochale dataset

The Rochale data is a binary dataset collected from 665 samples to assess the relationship among eight factors affecting the economical activities of women. The eight variables are named as a to h and the cells appear in the table row by row in a lexicographical order. The true adjacency matrix of the data is presented in Table I. For this dataset, the number of iterations for RJMCMC, BDMCMC and Gibbs is taken as 10^6 and for QUIC it is set to 1000 iterations.

Table II represents the results of accuracy measures for each method. It is seen that after QUIC, Gibbs sampling and RJMCMC have the same highest accuracy among alternatives. Here, although QUIC is the most speedy method, it is completely non-parametric and suggests a numeric solution for the inference of the precision. On the other hand, the remaining approaches are fully parametric and can be grouped in the same class. We further record the computational time of each algorithm in terms of the central processing unit (CPU) time, in the same table. From the results, it is seen that the Gibbs sampling has almost the same computational demand as RJMCMC and BDMCMC has the same efficiency as other methods, although the coding has been done in C and the remaining were coded in R programming language.

TABLE I

The true adjacency matrix for the networks of the Rochale data.

	a	b	c	d	e	f	g	h
b	0	0	0	1	1	0	0	1
a	0	0	1	1	1	0	1	0
c	1	0	0	0	1	1	1	0
d	1	1	0	0	0	0	1	1
e	1	1	1	0	0	1	0	0
f	0	0	1	0	1	0	1	0
g	1	0	1	1	0	1	0	0
h	0	1	0	1	0	0	0	0

TABLE II

The comparison of accuracy and CPU time of 100 iterations of each inference for the Rochale data.

Methods	TP	FP	FN	TN	F_1 - Score	MCC	CPU
True graph	14	0	0	14	1	1	-
RJMCMC	13	1	1	13	0.93	0.86	0.00
BDMCMC	11	8	3	9	0.70	0.59	0.00
Gibbs ($q = 0.5$)	13	1	1	13	0.93	0.86	0.02
QUIC ($\lambda = 0.12$)	14	1	0	13	0.96	0.93	0.00

4.2. The CellSignal data

This dataset is attached to the BDgraph package [14] via 11672 samples, in which each independent measurement consists of quantitative amounts of each of the 11 phosphorylated molecules, simultaneously measured from single cells. Figure 1 shows the true structure of these data via the Markov network. In Table III, we represent the accuracy measures and the computational time of each approach, similar to the previous dataset. From Table III, it is detected that QUIC is the most accurate method and the Gibbs sampling improves the computational demand significantly, without losing accuracy, among the methods in the parametric class.

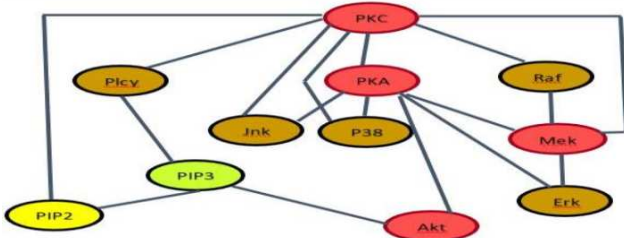


Fig. 1. The structure of the CellSignal data via an undirected graph.

TABLE III

The comparison of accuracy and CPU time of 100 iterations of each inference method for the CellSignal data.

Methods	TP	FP	FN	TN	F_1 - Score	MCC	CPU
True graph	18	0	0	37	1	1	–
RJMCMC	13	27	5	10	0.45	0.00	0.13
BDMCMC	16	32	5	2	0.49	–0.26	0.12
Gibbs ($q = 0.5$)	13	27	5	10	0.45	0.00	0.05
QUIC ($\lambda = 0.2$)	10	22	8	31	0.40	0.12	0.00

4.3. Simulated MAPK-ERK data

The MAPK (mitogen-activated protein kinase) or its synonymous ERK (extracellular signal regulated kinase) pathway is one of the major signal transduction systems which regulates the cellular growth control of all eukaryotes from the reproduction to the death of the cell. For this pathway, since the data have not been coverable for all proteins in the system, we generate a dataset based on the Bayesian inference of the reaction rates by using 5 proteins at 8 time points. The time course data is available in [15]. Then, the system is simulated for a long time via the exact Gillespie algorithm [16], until all proteins reach their steady state positions.

In this simulation, we apply the reaction rate constants listed in Table IV in the study of Purutçuoğlu and Wit [15]. On the other hand, in inference of the system, we eliminated the 27 proteins in 51 total proteins in the systems due to the high correlation and conducted both

algorithm via 24 independent components, where each protein has 1110 observations.

In both pathways, the number of iterations is set to 10000 MCMC and the first 2000 iterations are discarded as the burn-in period. From the results in Table IV, it is seen that RJMCMC, which takes into account the structural dependency in the systems, are more accurate than the findings of the BDMCMC, which cannot be applicable without eliminating those components, in terms of F_1 - score and MCC values.

On the other hand, if we compare the computational demand of both approaches, we see that RJMCMC uses more CPU time. But we think that the advantage of BDMCMC is not due to the plausible high computational demand of the improved RJMCMC. Further, it may be caused by the programming language of each algorithm. The RJMCMC approach is originally written in R, which is an interpreted language, whereas BDMCMC is written in C which is a compiled language.

TABLE IV

The comparison of accuracy and CPU time of each inference method for the simulated MAPK-ERK data.

Methods	TP	TN	FP	FN	F_1 - score	MCC	CPU
True graph	25	251	0	0	1	1	–
RJMCMC	25	213	38	0	0.56	0.58	0.14
BDMCMC	22	219	34	3	0.53	0.53	0.08
Gibbs ($q = 0.5$)	25	213	38	0	0.56	0.58	0.12
QUIC ($\lambda = 0.20$)	20	234	17	5	0.64	0.61	0.00

5. Conclusions

In this study, we have suggested to perform the Gibbs sampling strategy in inference of CGGM, in order to decrease the computational demand and to increase the accuracy of the RJMCMC algorithm. The results of three benchmark datasets have shown, that the Gibbs sampling is promising to improve the accuracy of the estimates, without losing the computational time. Hereby, it can optimize the modeling performance of GGM in time, under different dimensional network.

As the extension of this study, we consider to develop a more efficient algorithm, in particular, when the elements of the network are highly dependent on each other, i.e. they are structurally dependent from the description of the network. Furthermore, we think to extend CGGM, by applying C-D vine approaches [17] and to develop a new update scheme for the MCMC algorithm in inference of this new model.

Additionally, we consider that similar strategies, used in RJMCMC and Gibbs sampling, can be adapted in time series chain graphical model [18], correct inference of which is conducted via the two-stage penalized maximum likelihood method. Finally, different methods suggested for different types of networks can be also adapted for the complex biological systems [19, 20].

Acknowledgments

Authors would like to thank TÜBİTAK (Project No: 114E636) for their support.

References

- [1] Mazet and Brie, *An alternative to the RJMCMC algorithm, Proceeding of IAR Annual Meeting*, Nancy, 2006.
- [2] P.J. Green, *Biometrika* **82**, 711 (1995).
- [3] A. Dobra, A. Lenkoski, *Ann. Appl. Statistics* **5**, 969 (2011).
- [4] A. Mohammadi, E. Wit, *Int. Soc. Bayesian Anal.* **10**, 109 (2015).
- [5] S. Richardson, P.J. Green, *J. Royal Stat. Soc. B* **59**, 731 (1997).
- [6] Carlin, Chibs, *J. Roy. Stat. Soc. B*, **57**, 473 (1995).
- [7] S. Walker, *Electronic Journal of Statistics*, arXiv:0902.4117, 2009.
- [8] K. Ergen, A. Çilli, N. Yahmoğlu, *Acta Phys. Pol. A* **128**, B-273 (2015).
- [9] M. Cevri, D. Üstündağ, *Acta Phys. Pol. A* **130**, 45 (2016).
- [10] N. İyit, H. Yonar, A. Genç, *Acta Phys. Pol. A* **130**, 397 (2016).
- [11] B. Muthen, *Psychometrika* **49**, 115 (1984).
- [12] C.J. Hsieh, M.A. Sustik, I.S. Dhillon, P. Ravikumar, *J. Machine Learning Res.* **15**, 2911 (2014).
- [13] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, John Wiley and Sons, 1990.
- [14] K. Sachs, O. Perez, D. Pe'er, D.A. Lauenburger, G.P. Nolan, *Science* **308**, 523 (2005).
- [15] V. Purutçuoğlu, E. Wit, *Mathemat. Problems Engin.* **2012**, ID: 752631 (2012).
- [16] D. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [17] E.C. Brechmann, U. Schepsmeier, *J. Statistical Software* **52**, 1 (2013).
- [18] F. Abegaz, E. Wit, *Biostatistics* **14**, 586 (2013).
- [19] E. Boutalbi, L. Ait Gougam, F. Mekideche-Chafa, *Acta Phys. Pol. A* **128**, B-271 (2015).
- [20] T. Aydogan, I. Ozcelik, I. Erturk, H. Ekiz, *Acta Phys. Pol. A* **130**, 412 (2016).

Supplementary material

Under the GGM, the conditional probability computed for the RJMCMC inference approach maintains the following equality.

$$P\left(\theta^{(k)} \mid \theta^{(k+1)}\right) \times \pi_{k+1}\left(\theta^{(k+1)}\right) = P\left(\theta^{(k+1)} \mid \theta^{(k)}\right) \times \pi_k\left(\theta^{(k)}\right), \quad (\text{A1})$$

where $P(\cdot)$ denotes the conditional probability and the prior density, respectively. Moreover, θ shows the precision matrix and $(k+1)$ denotes the dimension of θ . Thus, Eq. (A1) is equivalent to the following statement, due to the definition of conditional probability and $\pi_{k+1}\left(\theta^{(k+1)}\right)$.

$$\frac{P\left(\theta^{(k)}, \theta^{(k+1)}\right)}{\theta^{(k+1)}} \times P\left(\theta^{(k+1)} \mid k+1\right) = \frac{P\left(\theta^{(k)}, \theta^{(k+1)}\right)}{\theta^{(k)}} \times P\left(\theta^{(k)} \mid k\right).$$

If we cancel similar expressions on both sides, then,

$$\frac{P\left(\theta^{(k+1)} \mid k+1\right)}{p\left(\theta^{(k+1)}\right)} = \frac{P\left(\theta^{(k)} \mid k\right)}{p\left(\theta^{(k)}\right)}.$$

In CGGM, $P\left(\theta^{(k)} \mid k\right)$ is the G -Wishart distribution, if the dimension is given by k , and $p\left(\theta^{(k)}\right)$ presents the Wishart distribution. So the ratio $\frac{P\left(\theta^{(k)} \mid k\right)}{p\left(\theta^{(k)}\right)}$ is constant and does not depend on k anymore.