Breaking a Gaussian Graphical Model

Supervisors: Dr Jack Jewson, Dr Ruben Loaiza-Maya

Mashaal Chowdhry Yusufi

Notation for the Problem I

Denote *i*th data vector as

$$y_i = \begin{pmatrix} y_{i1} \\ .. \\ .. \\ .. \\ .. \\ y_{ip} \end{pmatrix}$$

where the *i* th observation of the *j* th variable is y_{ij} and each Y_{ij} follows a multivariate normal distribution:

$$Y_{ij} \sim \mathcal{N}_p(\mu, \Omega^{-1})$$

where $\Sigma = \Omega^{-1}$

Notation for the Problem II

Complete information from the data is defined as:

$$Y = \begin{pmatrix} y_1^{'} \\ \vdots \\ y_n^{'} \end{pmatrix}$$

Likelihood Function

$$L(y_i|\Omega) = \prod_{i=1}^{n} \pi(y_i|\mu = 0, \Omega)$$

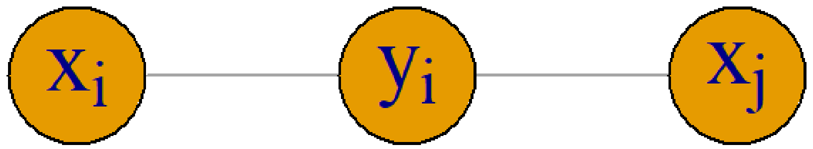
$$L(y_i|\Omega) = \frac{\det(\Omega)^{n/2}}{(2\pi)^{n/2}} \times \exp\left(-\sum_{i=1}^{n} \frac{y_i^T \Omega y_i}{2}\right)$$

Ωij Introduction

$X_i \perp\!\!\!\perp X_j \mid X_{V\setminus\{i,j\}}$

$\Omega_{ij} = 0$

x_i : Ice Cream Sales
y_i : Temperature
x_j : Number of Boats



Reasons for Bayesian Approach to Precision Matrix Estimation

- Focusing on dependence of pairs of variables conditional on the remaining variables
- Control weight given to sparsity in Precision Matrix via prior
- Selects best posterior model i.e model with highest posterior probability via a Sampling Algorithm

Precision Matrix Prior

$$\pi(\Omega) \propto \prod_{j=1}^{p} \pi(\omega_{jj}) \times \prod_{j<k} \pi(\omega_{jk})$$

subject to $1_{\Omega \succ 0}$

Off-Diagonals Prior: Spike and Slab

Given v $\sim Bernoulli(w)$:

if $v = 1$, then $\omega_{jk} \sim N(0, s^2)$

if $v = 0$, then $\omega_{jk} = 0$

$\pi(\omega_{jk}) = (1 - w)\delta_0 + wN(\omega_{jk}, 0, s^2)$

Diagonals Prior: Exponential Distribution

Posterior Distribution

posterior $\propto$ Prior $\times$ Likelihood

$$\pi(\Omega|Y) \propto L(Y|\Omega) \times \pi(\Omega)$$

Particularly interested in $\pi(\omega_{jk} = 0|Y)$

Results I: Assumptions

MCMC Gibbs Sampling Algorithm

MCMC Specifications: $p = 3, n = 1000, iterations = 2000, burn-in = 1000$

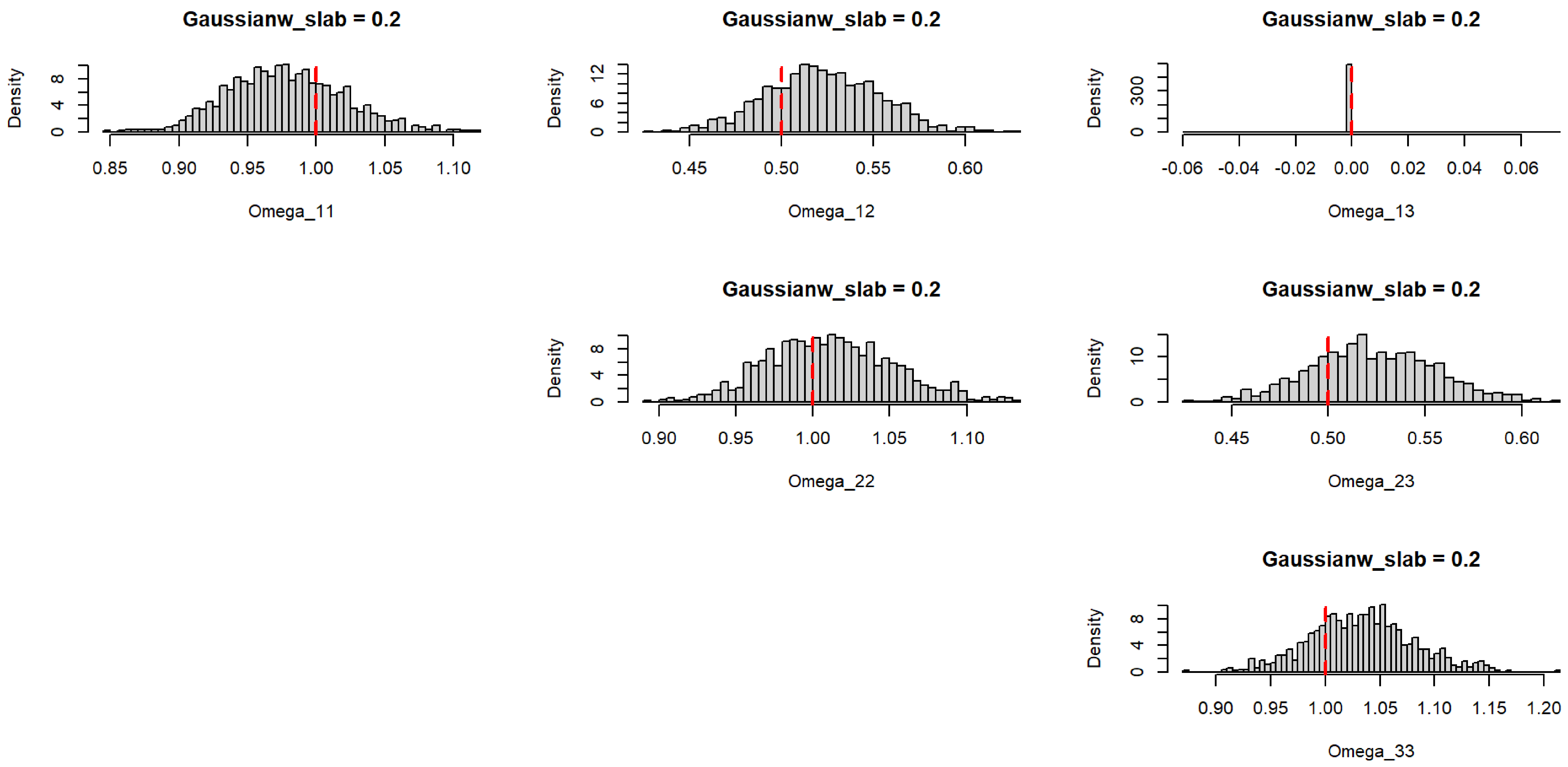
$$Y_{ij} \sim \mathcal{N}_p(\mu, \Omega)$$

$$\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Omega = \Sigma^{-1} = \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}$$

Results II: Posterior Edge Inclusion Probabilities of Off-diagonals with Gaussian Data

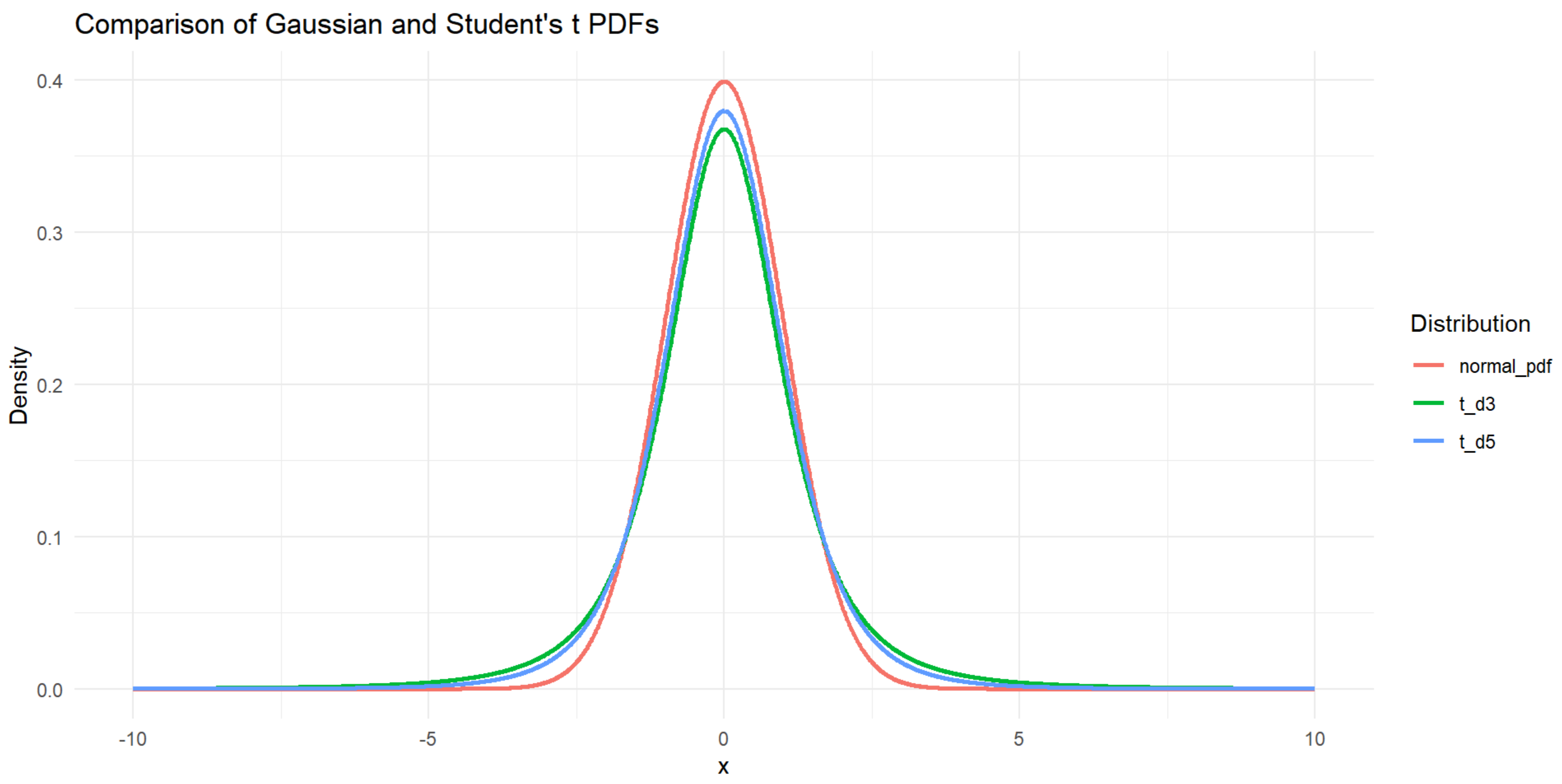
w_{slab}	ω_{12}	ω_{13}	ω_{23}
0.2	1	0.008	1
0.5	1	0.049	1
0.8	1	0.124	1

Results III: Posterior Histogram of Off-Diagonals with Gaussian Data (slab weight = 0.2)



Caveats about Simulation Results

- However, much real data is not Normally Distributed
- What if the data has Student-t margins?



Introducing a Gaussian Copula

$$Y_{ij} \sim \mathcal{N}_p(\mu, \Omega^{-1})$$

which corresponds to the i-th observation and j-th variable

transform the observed data into a uniform distribution: $U_{ij} = \Phi(Y_{ij})$ for $j = 1...p$ and $i = 1...n$

$$U_{ij} \sim Uniform(0, 1)$$

Transform the uniforms into Student-t margins using the quantile function of student-t at ν
$Z_{ij} = T_{\nu}^{-1}(U_{ij})$

Construct a joint distribution that is Gaussian with Student-t margins: $C(U_1 \ldots U_p)$

Results IV: Posterior Graph Inclusion Probabilities for Student-t(df = 5)

Table 1: Gaussian {#tbl-second}

(a) t_5

w_{slab}	ω_{12}	ω_{13}	ω_{23}
0.2	1	0.009	1
0.5	1	0.037	1
0.8	1	0.135	1

(b) Gaussian

w_{slab}	ω_{12}	ω_{13}	ω_{23}
0.2	1	0.008	1
0.5	1	0.049	1
0.8	1	0.124	1

Results V: Graph Inclusion Probabilities for Student-t(df = 3) vs Gaussian

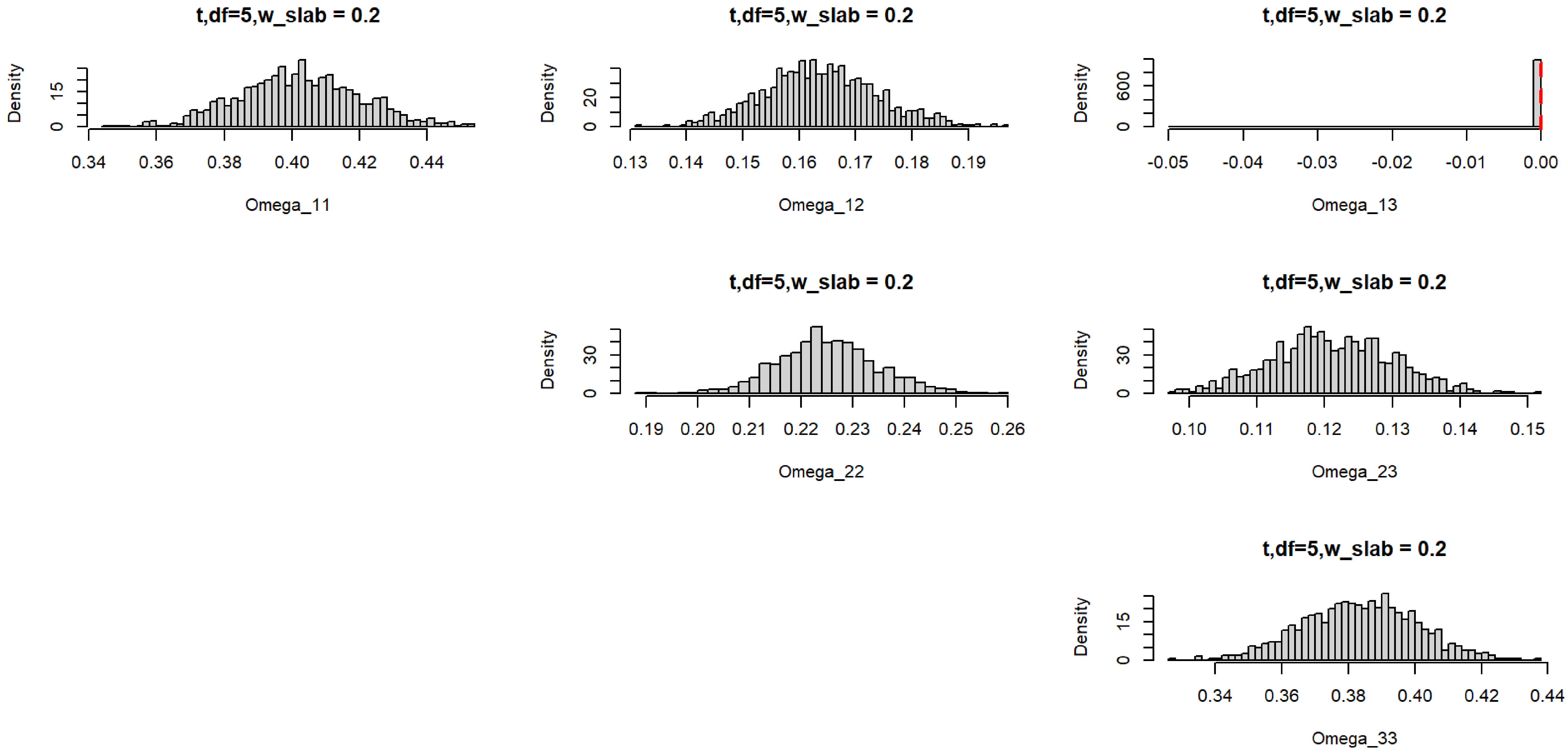
Table 2: t_3

w_{slab}	ω_{12}	ω_{13}	ω_{23}
0.2	1	0.506	0.577
0.5	1	0.610	0.622
0.8	1	0.760	0.729

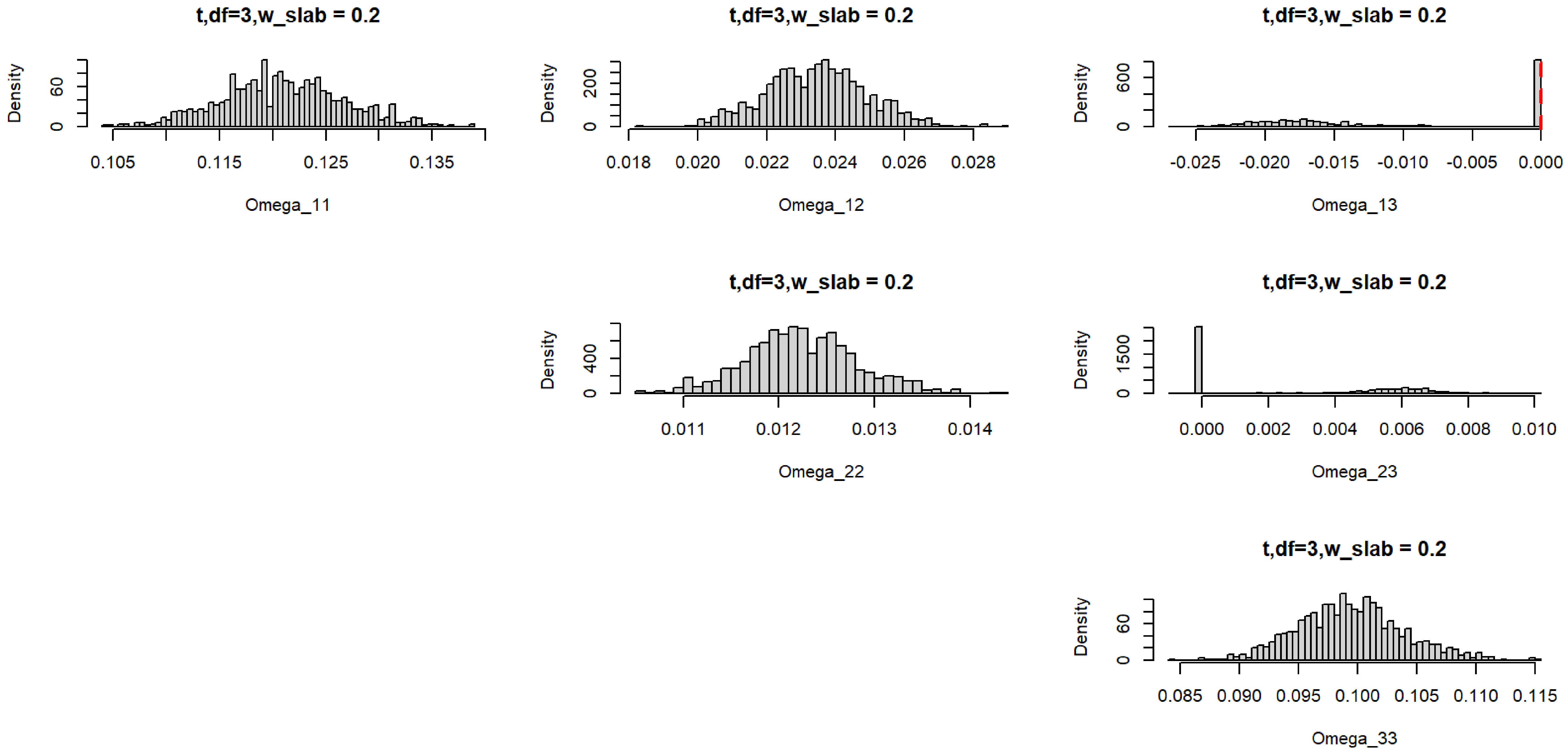
Table 3: Gaussian

w_{slab}	ω_{12}	ω_{13}	ω_{23}
0.2	1	0.008	1
0.5	1	0.049	1
0.8	1	0.124	1

Results VI: Posterior Histograms for Student-t(df = 5) slab weight = 0.2



Results VII: Posterior Histograms for Student-t(df = 3) slab weight = 0.2



Conclusion and Goal

Gaussian Graphical Model breaks when data is not Gaussian

Use a Gaussian Copula to create a more flexible model