Technische Universität München

Department of Mathematical Statistics

# D-vine Regression Based Bayesian Networks Applied to the Sachs Dataset

Master Thesis

by

Sebastian Scharl

| | |
|---|---|
| Supervisor: | Prof. Claudia Czado, Ph.D |
| Advisor: | Prof. Claudia Czado, Ph.D |
| Submission Date: | 20.04.2021 |

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Munich, 20.04.2021

# Acknowledgments

First of all, I would like to thank Prof. Claudia Czado for all the help in the process of writing this thesis. Explicitly, I would like to thank her for supporting my plan to write parts far away in Siberia. Even though due to that, and Covid-19, most of our conversations had to be held remotely her doors were always open for guidance and assistance.

Further I would like thank Thomas Nagler for the quick and uncomplicated help whenever I had any question regarding his R-packages.

Finally I would like to thank my family, friends, here and and in Russia, and my girlfriend Sophia for all the love and support throughout my life and especially within the past few months.

Большое спасибо всем!

**Abstract**

Appropriately modeling multivariate data is an important and present topic in statistics. Bayesian Networks, following Pearl (1988), have become a useful and intuitive tool to challenge this task. A common approach is to model each conditional density as a conditional normal density receiving a Linear Gaussian Bayesian Network (Koller and Friedman, 2009), which then represents a multivariate Gaussian distribution. This automatically results in disadvantages if the underlying data is not jointly normally distributed.

In this thesis we will present a solution approach to this shortfall where we will model the conditional density of each node as a D-vine on the set of its parents following the work of Kraus and Czado (2017). We illustrate this approach with the biological data set from Sachs et al. (2005) describing different levels of phosphoproteins and phospholipids in individual cells. We allow for different marginals and copulas while modeling and compare the results to a Linear Gaussian Bayesian Network. We compare the models using different goodness of fit measures and their ability to recreate the original data. Further, we analyze how the conditional density of each node, given its parents, behaves when conditioning on specific values.

# Contents

# Chapter 1

# Introduction

These days the computational power is continuously growing, without any signs that this trend might stop in the near future. Despite this development, the at least equally strong growing availability of data has made it more and more necessary to create a balance between the complexity of the models and their calculation time. This is even more important if not only the number of samples but also the amount of covariates grows.

One approach to fit models is the use of copulas based on Sklar's theorem (Sklar, 1959) providing the advantage to model the dependence structure and the margins independent of each other. Resulting in one, depending on the amount of covariates, high dimensional copula modeling the whole dependence structure. Instead Joe (1996) proposed a decomposition into two-dimensional building blocks further developed by Bedford and Cooke (2002). Therefore, increasing the number of covariates does not result in an increased dimension but an excessive increase in building blocks restricting the usability at a certain size.

To overcome this we will utilize a Bayesian Network following Koller and Friedman (2009), dividing the problem of fitting a model to a full data set into a set of problems on smaller data sets. Each of these smaller problems is then tackled following the idea of Kraus and Czado (2017) by performing a D-vine regression.

In the following we will first introduce the necessary theoretical background and later apply the approach to the data set discussed in Sachs et al. (2005) and compare the results to a Linear Gaussian Bayesian Network as defined by Koller and Friedman (2009). As we will find indications that the data might not be identical distributed throughout the whole set, we will repeat our approach only considering parts of the data. We then conclude the thesis stating the main results of the analysis.

# Chapter 2

# Theoretical Background

In this chapter we will, if not cited otherwise, closely follow the book "Analysing Dependent Data with Vine Copulas" by Czado (2019).

## 2.1 Introduction to Copulas

### 2.1.1 General Notation

As we will have to deal a lot with distributions and densities in the following, it is important to fix the notation first. From now on we will denote densities with $f$ and the corresponding distribution function with $F$. If we look at random variables, we will always denote them with capital letters and their outcome with small letters, i.e., $X = x$. More often we will have to deal with several variables at once. In this case we will consider them as a vector.

**Definition 1.** Let $X_1, ..., X_d$ be $d$ random variables where each $X_i$ follows some distribution. Then, we can arrange them in a vector $\mathbf{X} = (X_1, ..., X_d)^\top$ and assign the following properties to them

- Joint density $f(\mathbf{x}) = f(x_1, ..., x_d)$: The density of all elements in $\mathbf{X}$ together
- Marginal density $f_j(x_j)$: The density of the $j$-th element of the vector for $j \in \{1, ..., d\}$
- Conditional density $f_{j|k}(x_j|x_k)$: The density of the $j$-th element of the vector given the $k$-th element of the vector for all $j \neq k$

For all three properties there also exists the corresponding distribution function denoted by $F(x)$, $F_j(x_j)$ and $F_{j|k}(x_j|x_k)$.

One distribution which we will often need in the following is the the normal or Gaussian distribution. It is the probably most used continuous distribution.

**Example 1.** Given a vector $\mathbf{X} = (X_1, ..., X_d)^\top$ where each $X_i$, $i \in \{1, ..., d\}$, follows a Gaussian distribution with parameter $\mu_i$ and variance $\sigma_i^2$. We write $X_i \sim N(\mu_i, \sigma_i^2)$. Hence the density of $X_i$ is given by

$$f(x_i; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(x_i - \mu_i)^2\right\} \tag{2.1}$$

In that case $\mathbf{X}$ is also normally distributed and we write $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = (\mu_1, ..., \mu_d)^\top$ and $\Sigma = (\sigma_{ij})_{i,j\in[1,d]} \in \mathbb{R}^{d\times d}$. Its density function is then given by

$$f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

Following the first part of the example the marginal distribution of each $X_i$ is $N(\mu_i, \sigma_i)$.

In practice we often only have some multivariate dimensional data $\mathbf{X}$ but do not know which kind of distribution the data follows. Therefore, there are two questions arising:

- What is the marginal distribution of each $X_i$?
- How do the $X_i$ interfere with each other?

The copula approach tries to separate these problems and tackle them one after the other. First, we have to find suitable margins which we then use to transform the data to a space where we can use copulas to model dependencies.

### 2.1.2 Marginal Distributions and the Probability Integral Transform

The easiest way to find a marginal distribution for an arbitrary data set would be to use empirical distribution functions.

**Definition 2.** (Empirical distribution function) Let $\mathbf{x_i} = (x_{1i}, ..., x_{di})$ be an independent and identical distributed sample of size $n$ from the $d$-dimensional distribution $F$. The multivariate empirical distribution function is defined as

$$\hat{F}(x_1, ..., x_d) := \sum_{i=1}^{n} 1_{x_{1i}\leq x_1, ..., x_{di}\leq x_d}$$

for all $\mathbf{x} := (x_1, ..., x_d)^\top \in \mathbb{R}^d$. This also holds for the simple case where $d = 1$.

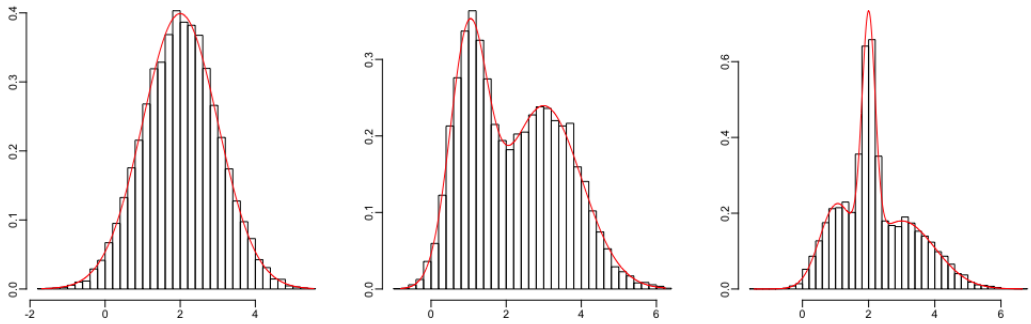


Figure 2.1: Histograms and densities of a normal distribution and two normal mixtures. Left: normal distribution with $\mu = 2$ and $\sigma = 1$. Middle: Two component Gaussian mixture model with $\mu_1 = 1,\ \mu_2 = 3,\ \sigma_1 = 0.5,\ \sigma_2 = 1$ and $w_1 = 0.4,\ w_2 = 0.6$. Right: Three component Gaussian mixture model with $\mu_1 = 1,\ \mu_2 = 2,\ \mu_3 = 3,\ \sigma_1 = 0.5, \sigma_2 = 0.2,\ \sigma_3 = 1$ and $w_1 = 0.25,\ w_2 = 0.3,\ w_3 = 0.45$

A disadvantage of the empirical distribution function is that it is not possible to calculate a measure for the goodness of fit, like the later explained likelihood, for it. Hence, it cannot be used when we compare different models. Therefore, we have to find other solutions.

In our case one approach will be the use of mixtures of normal distributions. In Example 1 we have already seen the normal distribution. However in practice data does not always follow this rather strict shape. Therefore, to be able to model more general kinds of data we define a mixture of several normal distributions following Härdle and Simar (2015).

**Definition 3.** (Gaussian mixture model) The probability density function of a mixture distribution consists of $k$ distributions with density functions $p_i$ and can be written as

$$f(x) := \sum_{i=1}^{k} w_i p_i(x)$$

under the constraints

$$0 \leq w_i \leq 1$$
$$\sum_{i=1}^{k} w_i = 1$$
$$\int p_i(x)dx = 1,\ \forall i = 1, ..., k$$

We speak of a Gaussian mixture model if all $p_i$ are of the form as in Equation (2.1).

For our later used transformations, we do not need the probability density function but the probability distribution function. Thanks to the formulation of the Gaussian mixture

model we can integrate each component separately. This is, knowing that all $p_i$ follow a Gaussian distribution, rather simple. Hence, we end up with a weighted sum of Gaussian probability distribution functions.

Another approach to model complex densities in a flexible way are kernel density estimates as defined in Weglarczyk (2018).

**Definition 4.** (Kernel density estimate) Let $\mathbf{x} = (x_1, ..., x_n)$ be an independent and identically distributed sample of observations of an unknown probability distribution function $f(x)$. Then, the kernel estimate $\hat{f}(x)$ assigns a function $K(x_i, t)$ to each sample data point $x_i$ which is called a kernel function. This is done in the following way

$$\hat{f}(x) := \frac{1}{n} \sum_{i=1}^{n} K(x_i, t) \tag{2.2}$$

with $K(x, t)$ being non-negative and bounded for all real $x_i$ and $t$. Additionally for all real $x$ it holds

$$\int K(x, t)dt = 1$$

If $K(x, t)$ is a symmetric function, we can rewrite the kernel function in a formulation more frequently used

$$K_{sym}(x, t) := \frac{1}{h} K' \left( \frac{x - t}{h} \right)$$

where $h$ is the so called smoothing parameter.

In our case, we will use a Gaussian kernel which satisfies the symmetry requirement as it is the density function of a standard normal distribution.

$$K_{Gauss}(t) := \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{-t^2}{2} \right\}$$

Again, this is important as we later need the probability distribution function, i.e., we need to integrate over the probability density function. This can easily be done over a weighted combination of densities of the standard normal distribution. To later compare fitted kernel density estimates we need to be able to determine the amount of parameters used. For kernel density estimates they are called effective parameters. As it is not crucial for the the work in this thesis, we refer for the calculation procedure to Example 5.1 in Loader (1999).

Figure 2.2 illustrates kernel density estimates fitted to data drawn from the distributions from Figure 2.1. We observe that the kernel density estimates fit very well and it is hard to see any difference compared to the original distributions.

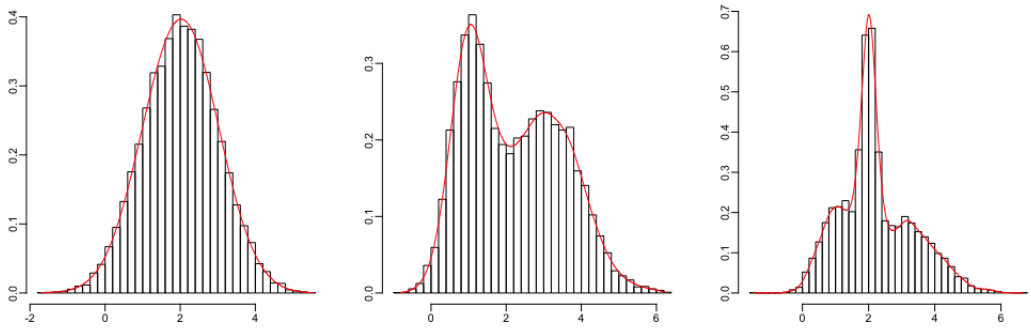


Figure 2.2: Fitted kernel density estimates to the data drawn from the distributions in Figure 2.1

For a more scientific way to compare the fit of different marginals, than graphically comparing histograms, we introduce different goodness of fit measures following Held and Sabanés Bové (2014).

**Definition 5.** (Marginal log-likelihood, $\text{AIC}_M$ and $\text{BIC}_M$) Given independent realizations $\mathbf{x} = (x_1, ..., x_n)$ of the random variable $X$ and assume $X$ follows a density function $f(x_i; \boldsymbol{\theta})$ with parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_r)$. Then, the marginal log-likelihood, Akaike information criterion ($\text{AIC}_M$) and Bayesian information criterion ($\text{BIC}_M$) are defined as

- Log-likelihood:

$$l_M(\boldsymbol{\theta}; \mathbf{x}) := \sum_{i=1}^{n} \ln\left(f(x_i; \boldsymbol{\theta})\right) \tag{2.3}$$

- Akaike information criterion:

$$AIC_M(\boldsymbol{\theta}; \mathbf{x}) := -2 \sum_{i=1}^{n} \ln(f(x_i; \boldsymbol{\theta})) + 2|\boldsymbol{\theta}|$$

- Bayesian information criterion:

$$BIC_M(\boldsymbol{\theta}; \mathbf{x}) := -2 \sum_{i=1}^{n} \ln(f(x_i; \boldsymbol{\theta})) + |\boldsymbol{\theta}| \ln(n)$$

In general a higher marginal log-likelihood, respectively smaller $\text{AIC}_M$ and $\text{BIC}_M$, corresponds to a better fit when comparing two different densities for the same realizations.

Now that we have ways to find suitable margins we want to use them to transform the data to a space where it is easier to investigate dependencies. Therefore, we want to transform each $X_i$ into a uniformly $[0, 1]$ distributed variable. For this we use the probability integral transform (PIT).

**Definition 6.** (Probability integral transform (PIT)) Let $X$ follow a continuous distribution $F$ and let $x$ be the observed value of $X$. Then, the transformation $u := F(x)$ is called the probability integral transform (PIT) at $x$.

This new random variable $U := F(X)$ is now uniformly distributed as

$$P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$$

The random variable $U$ does not contain any more information about the distribution of $X$. Hence, we can now analyze dependencies between uniformly distributed variables $U_1, ..., U_d$ without the influence of the marginal distributions of $X_1, ..., X_d$.

### 2.1.3 Copulas

Being now able to transform a data set to a multivariate uniformly distributed data set defined on a $d$-dimensional $[0, 1]$ hypercube, we want to estimate how the $d$ variables interact with each other. Since we got rid of the information about the marginals by applying the PIT to each variable, we have the suitable data set for this purpose. We will now model this dependence by a special function called a copula.

**Definition 7.** (Copulas) Let $\mathbf{U} = (U_1, ..., U_d)^\top$ be a $d$-dimensional random vector on the hypercube $[0, 1]^d$ with uniform margins. Then, the $d$-dimensional multivariate distribution function $C$ of $\mathbf{U}$ is called a copula. Assuming $C$ is continuous, then the associated copula density is obtained by partially differentiating $C$, i.e.,

$$c(u_1, ..., u_d) := \frac{\partial^d}{\partial u_1, ..., \partial u_d} C(u_1, ..., u_d)$$

For all $\mathbf{u} \in [0, 1]^d$.

Up to here we do not have a result saying that it is possible to separate the estimation of marginals and and dependence structure. However, this is what is stated in Sklar's theorem. For every multivariate distribution function there is a copula $C$ which uses the marginal distribution functions as inputs and allows us to express the associate multivariate distribution function.

**Theorem 1.** (Sklar's theorem part 1) Let $\mathbf{X}$ be a $d$-dimensional random vector with joint distribution function $F$ and marginal distribution functions $F_i$, $i \in [1, d]$. Then, the joint distribution function of $\mathbf{X}$ is given by

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$$

Further its density function can be expressed by

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) \times f_1(x_1) \times ... \times f_d(x_d)$$

for some $d$-dimensional copula $C$ with copula density $c$.

Moreover, Sklar's theorem also states, that if the underlying distribution is absolutely continuous, $C$ is also absolutely continuous. This is important as the data studied later is assumed to be exactly that. What is missing is if we can construct a copula derived from every $d$-dimensional distribution. This is also stated in Sklar's theorem.

**Theorem 2.** (Sklar's theorem part 2) The copula corresponding to a multivariate distribution function $F$ with marginal distribution functions $F_i$, $i \in [1, d]$, can be expressed as

$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)) \tag{2.4}$$

and its copula density by

$$c(u_1, ..., u_d) = \frac{f(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \times ... \times f_d(F_d^{-1}(u_d))}$$

Returning to the previous subsection, it is often required to estimate a density given another variable. Hence, are also interested how to represent conditional densities using copulas. As we will later focus on two-dimensional data, we will in the following only consider conditional densities in the bivariate case.

**Lemma 1.** The conditional density and distribution function of a bivariate variable can be expressed as

$$\begin{aligned} f_{1|2}(x_1|x_2) &= c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \\ F_{1|2}(x_1|x_2) &= \frac{\partial}{\partial u_2} C_{12}(F_1(x_1), u_2)|_{u_2=F_2(x_2)} := \frac{\partial}{\partial F_2(x_2)} C_{12}(F_1(x_1), F_2(x_2)) \end{aligned}$$

Similar to the conditional density of a copula we can define a conditional distribution for a bivariate distribution derived from a copula. In literature, this is often called the $h$-function.

**Definition 8.** ($h$-functions) Given a bivariate copula $C_{12}$ the corresponding $h$-functions for all $(u_1, u_2) \in [0, 1]$ are defined as

$$\begin{aligned} h_{1|2}(u_1|u_2) &:= \frac{\partial}{\partial u_2} C_{12}(u_1, u_2) \\ h_{2|1}(u_2|u_1) &:= \frac{\partial}{\partial u_1} C_{12}(u_2, u_1) \end{aligned}$$

Up to this point we have been talking about modeling dependence structures between variables. Still, we do not know any procedure to quantify the strength of the dependence between the variables. Hence, next we will discuss measures that will help us to quantify dependence.

### 2.1.4 Dependence Measures

The most famous dependence measure between two random variables $X_1$ and $X_2$ might be the Pearson product-moment correlation defined as

$$\mathrm{Cor}(X_1, X_2) := \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}}$$

with Cov denoting the covariance and Var the variance of the variables. In our case the Pearson product-moment correlation is not applicable, as it is not invariant under monotone transformations. By applying the PIT we applied exactly such a transformation, i.e., $\mathrm{Cor}(X_i, X_j) = \mathrm{Cor}(U_i, U_j)$ does not hold in general for all $i, j \in [1, d]$. Hence, even if we measure dependence between $U_i$ and $U_j$, using the Pearson product-moment correlation, we cannot use the results to make assumptions about the dependence between $X_i$ and $X_j$. To overcome this problem we will in the following use Kendall's $\tau$. It is a rank based dependence measure and therefore invariant under monotone transformations.

**Definition 9.** (Kendall's $\tau$) Having two continuous random variables $X_1$ and $X_2$. Then Kendall's $\tau$ of $X_1$ and $X_2$ is defined as

$$\tau(X_1, X_2) := P((X_{11} - X_{21})(X_{12} - X_{22}) > 0) - P((X_{11} - X_{21})(X_{12} - X_{22}) < 0)$$

where $(X_{11}, X_{12})$ and $(X_{21}, X_{22})$ are independent and identical distributed copies of $(X_1, X_2)$.

Again, note that we often do not have explicit knowledge about the distribution of the random vector, but only have its outcomes given. Hence, we have to estimate Kendall's $\tau$ using the observations. For this we have to count concordant, discordant and, depending on allowing ties, extra pairs.

**Definition 10.** (Concordant, discordant and extra pairs) Given two realizations of the random variables $X_1$ and $X_2$ $\mathbf{x_i} = (x_{i1}, x_{i2})$ and $\mathbf{x_j} = (x_{j1}, x_{j2})$ then the pair $(\mathbf{x}_i, \mathbf{x}_j)$ is called

- Concordant if the ordering in $\mathbf{x^1} := (x_{i1}, x_{j1})$ is the same as in $\mathbf{x^2} := (x_{i2}, x_{j2})$, i.e., $x_{i1} < x_{j1}$ and $x_{i2} < x_{j2}$ holds or $x_{i1} > x_{j1}$ and $x_{i2} > x_{j2}$ holds

- Discordant if the ordering in $\mathbf{x^1}$ is opposite to the ordering of $\mathbf{x^2}$ so either $x_{i1} < x_{j1}$ and $x_{i2} > x_{j2}$ or $x_{i1} > x_{j1}$ and $x_{i2} < x_{j2}$ holds
- Extra $x_1$ pair if $x_{i1} = x_{j1}$ holds
- Extra $x_2$ pair if $x_{i2} = x_{j2}$ holds

Given a sample of realization of $X_1$ and $X_2$ of size $n$ and consider all $\binom{n}{2}$ unordered pairs of $(\mathbf{x}_i, \mathbf{x}_j)$. We count all of the pairs where $N_c$ is the number of concordant pairs, $N_d$ the number of discordant pairs, $N_1$ the number of extra $x_1$ pairs and $N_2$ the number of extra $x_2$ pairs. These pairs can then be used to estimate Kendall's $\tau$ between $X_1$ and $X_2$ either by allowing for ties or allowing no ties.

**Definition 11.** (Estimate of Kendall's $\tau$ allowing for ties) When allowing for ties, Kendall's $\tau$ between the variables $X_1$ and $X_2$ on a sample of size $n$ can be estimated by

$$\hat{\tau}_n^* := \frac{N_c - N_d}{\sqrt{N_c + N_d + N_1} \times \sqrt{N_c + N_d + N_2}}$$

**Definition 12.** (Estimate of Kendall's $\tau$ not allowing for ties) An estimate of Kendall's $\tau$ between the variable $X_1$ and $X_2$ on a sample of size $n$ without ties is defined as

$$\hat{\tau}_n := \frac{N_c - N_d}{\binom{n}{2}}$$

In addition we are able to express Kendall's $\tau$ by using copulas.

**Theorem 3.** (Kendall's $\tau$ expressed in terms of a copula) Let $(X_1, X_2)$ be continuous variables, then Kendall's $\tau$ can be expressed as

$$\tau = 4 \int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1$$

However, one dependence measure often is not sufficient to describe the dependence structure between two variables. It often happens that the dependence structure around the mean of the variables strongly differs to the structure in the outer quantiles. For example the data might have a higher dependence between (extreme) high values of $X_1$ and $X_2$ than in the middle of the data. Hence, we introduce the concept of tail dependence.

**Definition 13.** (Upper and lower tail dependence coefficient) The upper tail dependence coefficient of a bivariate distribution with copula $C$ is defined as

$$\lambda^{upper} := \lim_{t \to 1^-} P(X_2 > F_2^{-1}(t) | X_1 > F_1^{-1}(t)) = \lim_{t \to 1^-} \frac{1 - 2t + C(t,t)}{1 - t}$$

while the lower tail dependence coefficient is defined as,

$$\lambda^{lower} := \lim_{t \to 0^+} P(X_2 \leq F_2^{-1}(t) | X_1 \leq F_1^{-1}(t)) = \lim_{t \to 0^+} \frac{C(t,t)}{t}$$

Hence, up to this point we are able to measure dependence between two variables. Often this is not sufficient for example if we have a $d > 2$ dimensional data set and we are interested in how two variables correlate if we remove the effect of the remaining variables. Thereto we introduce the concept of partial correlations.

**Definition 14.** (Partial regression and partial correlation) Let $X_1, ..., X_d$ be random variables with mean zero and variance $\sigma_i^2$. Further, denote by $I^d_{-(i,j)}$ the set $\{1, ..., d\}$ with indices $i$ and $j$, $i \neq j$, removed. Define the partial regression coefficients $b_{i,j;I^d_{-(i,j)}}$ for $i < j$ as the quantities which minimize

$$E\left(\left[X_i - \sum_{j=2, j\neq i}^{d} b_{i,j;I^d_{-(i,j)}} X_j\right]^2\right)$$

The corresponding $\binom{n}{2}$ partial correlations $\rho_{i,j;I^d_{-(i,j)}}$ are defined as

$$\rho_{i,j;I^d_{-(i,j)}} := \operatorname{sgn}\left(\mathrm{b_{i,j;I^d_{-(i,j)}}}\right) \times \sqrt{\mathrm{b_{i,j;I^d_{-(i,j)}}} \times \mathrm{b_{j,i;I^d_{-(i,j)}}}}$$

To simplify the calculation of the partial correlations we can use the following recursion formula.

**Theorem 4.** (Recursion for partial correlation) The partial correlations satisfy the following recursions

$$\rho_{i,j;I^d_{-(i,j)}} = \frac{\rho_{i,j;I^{d-1}_{-(i,j)}} - \rho_{i,d;I^{d-1}_{-(i,j)}} \times \rho_{j,d;I^{d-1}_{-(i,j)}}}{\sqrt{1 - \rho^2_{i,d;I^{d-1}_{-(i,j)}}}\sqrt{1 - \rho^2_{j,d;I^{d-1}_{-(i,j)}}}} \tag{2.5}$$

Now, we are able to quantify dependence in the data. Next, we will see which kind of bivariate copulas are available to model these dependencies.

### 2.1.5 Bivariate Copulas

Later we will find a way to describe $d$-dimensional data by a structure consisting only of two-dimensional building blocks, which we will then be able to model by two-dimensional copulas. Therefore, we will in the following concentrate on these bivariate copulas. In general we distinguish between parametric and non-parametric copulas which we will consider in this section.

The class of parametric copulas can be separated into three different types:

- Elliptical copulas: Recalling Equation (2.4) we see an easy way to construct copulas. We only need a multivariate distribution and its marginals. Well known examples for this are the bivariate Gaussian copula and the bivariate Student's t copula.

**Example 2.** (Bivariate Gaussian copula) Given a bivariate normal distribution with mean zero vector and correlation $\rho$, in the following denoted as $\Phi_2(\cdot, \cdot; \rho)$, and a standard normal distribution, i.e., with zero mean and variance one, denoted by $\Phi(\cdot)$. Then, we can apply the inverse of Sklar's theorem to define the Gaussian copula as

$$C(u_1, u_2; \rho) := \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)$$

**Example 3.** (Multivariate Student's t copula) Instead of using a Gaussian distribution function we can also use a Student's t distribution function. For this let $\mu$ be a zero vector of size two and

$$\Sigma_\rho = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Let us denote by $T_{2,\Sigma_\rho,\nu}$ the two-dimensional Student's t distribution with zero mean vector, scale parameter matrix $\Sigma_\rho$ and $\nu$ degrees of freedom and by $T_\nu^{-1}$ the inverse of the univariate standard Student's t distribution with $\nu$ degrees of freedom. We can then define the bivariate Student's t copula as

$$C(u_1, u_2; \Sigma_\rho, \nu) := T_{2,\Sigma_\rho,\nu}(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2))$$

- Archimedean copulas: The second class are the Archimedean copulas, for its construction we use generator functions $\varphi$.

  **Definition 15.** (Generators and Archimedean copulas) Let $\Omega$ be the set of all continuous, strictly monotone decreasing and convex functions $\varphi : I \to [0, \infty]$ with $\varphi(1) = 0$. Additionally denote the pseudo-inverse of $\varphi$ by $\varphi^{[-1]}$. It is defined as $\varphi^{[-1]} : [0, \infty] \to [0, 1]$ by

$$\varphi^{[-1]}(t) := \begin{cases} \varphi^{-1}(t) & , 0 \leq t \leq \varphi(0) \\ 0 & , \varphi(0) \leq t \leq \infty \end{cases}$$

  If it further holds that $\varphi(0) = \infty$ then the generator is called strict.
  We define an Archimedean copula with generator $\varphi$ by

$$C(u_1, u_2) := \varphi^{[-1]}(\varphi(u_1), \varphi(u_2))$$

  Well known examples of the Archimedean copula class with one parameter are the Clayton, Gumbel, Frank and Joe copula families.

**Example 4.**

– Clayton copula

$$C(u_1, u_2) := (u_1^{-\delta} + u_2^{-\delta} - 1)^{-\frac{1}{\delta}}$$

where $0 < \delta < \infty$ is the dependence parameter denoting full dependence at $\delta \to \infty$ and independence at $\delta \to 0$

– Gumbel copula

$$C(u_1, u_2) := \exp\left[-\left\{(-\ln u_1)^\delta + (-\ln u_2)^\delta\right\}^{\frac{1}{\delta}}\right]$$

where $\delta \geq 1$ is the parameter for dependence. With $\delta \to \infty$ denoting full dependence, while $\delta = 1$ corresponds to independence

– Frank copula

$$C(u_1, u_2) := -\frac{1}{\delta} \ln\left(\frac{1}{1 - e^{-\delta}}\left[(1 - e^{-\delta}) - (1 - e^{-\delta u_1})(1 - e^{-\delta u_2})\right]\right)$$

with $\delta \in [-\infty, \infty] \setminus \{0\}$ where $\delta \to 0^+$ corresponds to independence

– Joe copula

$$C(u_1, u_2) := 1 - \left((1 - u_1)^\delta + (1 - u_2)^\delta - (1 - u_1)^\delta (1 - u_2)^\delta\right)^{\frac{1}{\delta}}$$

with $\delta \geq 1$ and independence corresponding to $\delta = 1$

In addition there are Archimedean copulas with two parameters. For example the copulas from the $BB1$ and $BB7$ family.

**Example 5.**

– $BB1$ copula

$$\begin{aligned} C(u_1, u_2; \theta, \delta) :=& \left\{1 + \left[(u^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta\right]^{\frac{1}{\delta}}\right\}^{-\frac{1}{\theta}} \\ =& \eta\left(\eta^{-1}(u_1) + \eta^{-1}(u_2)\right) \end{aligned}$$

with $\eta$ being the inverse generator , i.e., $\eta(s) = \eta_{\theta,\delta}(s) = \left(1 + s^{\frac{1}{\delta}}\right)^{-\frac{1}{\theta}}$. The parameters are defined as $\theta > 0$ and $\delta \geq 1$ where for $\theta \to 0^+$ and $\delta \to 1^+$ independence arises

– $BB7$ copula

$$\begin{aligned} C(u_1, u_2; \theta, \delta) :=& \left(1 - \left[\left(1 - (1 - u_1)^\theta\right)^{-\delta} + \left(1 - (1 - u_2)^\theta\right)^{-\delta} - 1\right]^{-\frac{1}{\delta}}\right)^{\frac{1}{\theta}} \\ =& \eta\left(\eta^{-1}(u_1) + \eta^{-1}(u_2)\right) \end{aligned}$$

with $\eta(s) = \eta_{\theta,\delta}(s) = 1 - \left[1 - (1 + s)^{-\frac{1}{\delta}}\right]^{\frac{1}{\theta}}$, $\theta \geq 1$ and $\delta > 0$. Independence arises for $\theta = 1$ and $\delta = 0$

- Extreme-value copulas: The third class is often used to study the behavior of extreme events. To characterize the class of bivariate extreme value copulas we use the characterization of Pickands (1981).

  **Theorem 5.** (Characterization of bivariate extreme-value copulas in terms of the Pickands dependence function) A bivariate copula $C$ is an extreme-value copula if and only if

  $$C(u_1, u_2) = \exp\left\{[\ln(u_1) + \ln(u_2)]A\left(\frac{\ln(u_2)}{\ln(u_1 u_2)}\right)\right\}$$

  where $A$ is called the Pickands dependence function $A : [0,1] \to [\frac{1}{2}, 1]$ which is convex and satisfies $\max\{1-t, t\} \leq A(t) \leq 1, \ \forall t \in [0,1]$.

  Note that this characterization is only valid in the bivariate case. An example for extreme value copulas are the Marshall-Olkin and the Tawn copula.

  **Example 6.**

  - Marshall-Olkin copula

    $$A(t) = \max\{1 - \alpha_1(1-t), 1 - \alpha_2 t\}$$

    with parameters $0 \leq \alpha_1, \alpha_2 \leq 1$

  - Tawn copula

    $$A(t) = (1-\psi_1)(1-t) + (1-\psi_2)t + [(\psi_1(1-t))^\theta + (\psi_2 t)^\theta]^{\frac{1}{\theta}}$$

    with $0 \leq \psi_1, \psi_2 \leq 1$ and $\theta \geq 1$

Examples of non-parametric copulas estimations are the Empirical Bernstein copula (Yang et al. (2020)), Bernstein polynomials (Sancetta and Satchell (2004)) or B-splines (Kauermann et al. (2013)). As we will later use the vinereg package by Nagler (2019) which only uses transformation local likelihood kernel estimators (TLL) based copulas these will be the only ones we define. We will here use the definition as in Nagler et al. (2017).

The idea of TLL copulas is not to fit the copula to the uniformly distributed data but instead using a transformation trick so we fit the copula to data where the margins are normally distributed. The reason is, that kernel density estimators work very well if they are calculated on an unbounded set.

**Definition 16.** (Transformation trick) Given a random sample of two uniformly $[0,1]$ distributed random variables $U_1$ and $U_2$. Let $\Phi$ be the standard normal distribution function

and $\phi$ its density. Then, $(Z_1, Z_2) = (\Phi^{-1}(U_1), \Phi^{-1}(U_2))$ has normally distributed margins and is supported on $\mathbb{R}^2$. Using the first part of Sklar's theorem in Theorem 1 we can write the density of $(Z_1, Z_2)$ as

$$f(z_1, z_2) = c(\Phi(z_1), \Phi(z_2) \times \phi(z_1) \times \phi(z_2)$$

for all $z_1, z_2 \in \mathbb{R}$. By rearranging and a change of variables $u_j = \Phi(z_j)$ for $j = 1, 2$ we obtain

$$c(u_1, u_2) = \frac{f(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1)) \times \phi(\Phi^{-1}(u_1))}$$

In a next step we can estimate $f$ with an estimation $\hat{f}$ which will give us an estimation $\hat{c}$ of $c$

$$\hat{c}(u_1, u_2) = \frac{\hat{f}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1)) \times \phi(\Phi^{-1}(u_1))}$$

For calculation purposes we will not approximate $f(z_1, z_2)$ of the random vector $\mathbf{Z} = (Z_1, Z_2)$ but $\log f(z_1, z_2)$. We do this by using a polynomial of degree $q$.

**Example 7.** (Log-quadratic expansion) Letting $q = 2$ we can approximate $\log f(z_1', z_2')$ by

$$\begin{aligned}\log f(z_1', z_2') \approx\ & a_1 + a_2(z_1 - z_1') + a_3(z_2 - z_2') + a_4(z_1 - z_1')^2 \\ & + a_5(z_1 - z_1')(z_2 - z_2') + a_6(z_2 - z_2')^2\end{aligned}$$

for $(z_1', z_2')$ in a neighborhood of $\mathbf{z} = (z_1, z_2)$. To keep it readable we denote the right side of the equation from now on with $P_{\mathbf{a}}(\underline{\mathbf{z}})$ with $\mathbf{a} = (a_1, ..., a_s)^\top$. Additionally we introduce $A(\mathbf{z}) = 1 + (z_1 - z_1') + ... + (z_2 - z_2')^r$ which we will need later.

What is missing is a way to calculate the set $\mathbf{a}$. It can be obtained solving the weighted maximum likelihood problem

$$\begin{aligned}\hat{\mathbf{a}} = \arg\ \max_{\mathbf{a}} \Bigg\{ & \sum_{i=1}^{n} \mathbf{K}\left(B^{-1}\left(\mathbf{z} - \mathbf{Z}^{(i)}\right)\right) P_{\mathbf{a}}\left(\mathbf{z} - \mathbf{Z}^{(i)}\right) \\ & - n \int_{\mathbb{R}^2} \mathbf{K}\left(B^{-1}(\mathbf{z} - \mathbf{s})\right) \times \exp\left\{P_{\mathbf{a}}(\mathbf{z} - \mathbf{s})\right\} d\mathbf{s} \Bigg\}\end{aligned} \tag{2.6}$$

where we denoted the $i$-th sample point of $\mathbf{Z}$ with $Z^{(i)}$ and $\mathbf{K}(\mathbf{z}) = K(z_1)K(z_2)$ is the product kernel with $K(\cdot)$ a kernel function as in Definition 4. The matrix $B \in \mathbb{R}^{2\times 2}$ with $\det(B) > 0$ is called the bandwidth matrix, it therefore plays the role of $h$ in Definition 4. We obtain $\hat{\mathbf{a}}_\mathbf{1}$ as the optimized parameter of $\log f(z_1, z_2)$, so $\exp\{\hat{\mathbf{a}}_\mathbf{1}\}$ is the desired parameter vector needed for the estimate of $f(z_1, z_2)$.

**Definition 17.** (Transformation local likelihood kernel estimator (TLL)) A copula constructed using Definition 16 where $\hat{f}\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$ is approximated using a polynomial of degree $q$, which is solved by solving Equation (2.6), is called a $q$-TLL copula.

While for the earlier defined copulas the number of degrees of freedom can easily be determined this is not the case for the class of TLL copulas. Here, we have to be content with calculating the effective number of parameters as we already did for the kernel density estimates in Definition 4. For the calculation procedure we again refer to Loader (1999).

**Theorem 6.** (Effective number of parameters for TLL copulas) Given a TLL copula as defined in Definition 17, then the effective number of parameters $\nu$ is defined as

$$\nu := \sum_{i=1}^{n} \text{infl}(Z^{(i)})$$

where infl is defined as

$$\text{infl}(x) := n^{-1}\mathbf{K}(B^{-1}(0))e_1^\top \mathbf{M}^{-1}e_1$$

and $\mathbf{e_1}$ is the unit vector $\mathbf{e_1} = (1, 0, ..., 0) \in \mathbb{R}^{q+1}$ and

$$\mathbf{M} := \int \mathbf{K}(B^{-1}(v))A(v)A(v)^\top \exp\left\{P_{\mathbf{a}}(\mathbf{z} - \mathbf{s})\right\} dv$$

To complete our list of the different copula classes we define another very important copula. A copula to model the case where two variables are independent. Here, we use the definition as in Sun et al. (2020).

**Definition 18.** (Independence copula) Given two variables $U_1$ and $U_2$ that are independent we define the following copula to model this independence

$$C(u_1, u_2) := u_1 \times u_2$$

This is the so called independence copula.

If we look at the copulas that we have defined, we can see that some of the copulas are not symmetric. To extend the range of dependencies that can be modeled by these copulas we define a way to counter clockwise rotate rotate them.

**Definition 19.** (Rotated and reflected copulas) Let $c(\cdot, \cdot)$ be a copula density, then its counter clockwise rotation is defined by

- $90° : c_{90}(u_1, u_2) := c(1 - u_2, u_1)$

- $180° : c_{180}(u_1, u_2) := c(1 - u_1, 1 - u_2)$
- $270° : c_{270}(u_1, u_2) := c(u_2, 1 - u_1)$

Sometimes one might not be interested in the dependence between two variable but instead in the dependence of the two variables given a third $X_3 = x_3$. We model this again using a bivariate copula.

**Definition 20.** (Conditional bivariate copula with one-dimensional conditioning set) Let $U_1$ and $U_2$ be two uniformly $[0, 1]$ distributed random variables and $U_3 = v_3$ given. Then, the conditional bivariate copula of $U_1$ and $U_2$ given $U_3 = v_3$ is defined as

$$C_{12,3}(u_1, u_2|v_3) := C_{12|3}(C^{-1}_{1|3}(u_1|v_3), C^{-1}_{2|3}(u_2|v_3)|v_3)$$

where $C_{12|3}$ is the distribution function of $U_1$ and $U_2$ given $U_3 = v_3$

$$C_{12|3}(u_1, u_2|v_3) := \frac{\partial}{\partial u_3} C_{123}(u_1, u_2, u_3)|_{u3=v3}$$

In the following we assume that the conditional bivariate copula does not depend on the specific value $v_3$. This is called the simplifying assumption.

**Definition 21.** (Simplifying assumption in three dimensions) The simplifying assumption of a conditional bivariate copula with one-dimensional conditioning set is satisfied when for any $v_3 \in \mathbb{R}$ it holds that

$$c_{12;3}(u_1, u_2; v_3) = c_{12;3}(u_1, u_2)$$

for all $u_1, u_2 \in [0, 1]$.

In Section 2.1.4 we have seen a way to calculate Kendall's $\tau$. As the following definition shows us there exists a direct connection between some parametric copulas and Kendall's $\tau$.

**Theorem 7.** (Kendall's $\tau$ for elliptical, Archimedean and extreme-value copulas) Let $\rho$ be the association parameter for the bivariate elliptical copula, $\varphi$ the generator of the bivariate Archimedean copula and $A$ the Pickands dependence function of the bivariate extreme-value copula with existing first derivative. Then, for the bivariate elliptical copula it holds that

$$\rho = \sin\left(\frac{\pi}{2}\tau\right)$$

For the bivariate Archimedean copula it holds that

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$$

and for the bivariate extreme-value copula that

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} dA'(t)$$

Hence, to this point, given two random variables we can calculate their dependency with each other in terms of Kendall's $\tau$ and have a wide range of bivariate copulas with, even more parameter sets, that we can use to create a model with the same Kendall's $\tau$. Now is the question which of the model fits best to the given data. For this we again introduce three characteristics to quantify the fit. First, the well known log-likelihood second the Akaike information criterion and then the Bayesian information criterion. The last two penalize the size of the used parameter vector.

**Definition 22.** (Copula log-likelihood, $\text{AIC}_C$ and $\text{BIC}_C$) Given a copula density $c(\cdot, \cdot)$ with parameter vector $\boldsymbol{\theta}$ of size $|\boldsymbol{\theta}|$ and data $U = \begin{pmatrix} u_{11} & u_{12} \\ \vdots & \vdots \\ u_{n1} & u_{n2} \end{pmatrix}$ the copula log-likelihood, copula Akaike information criterion ($\text{AIC}_C$) and copula Bayesian information criterion ($\text{BIC}_C$) are defined as

- Copula log-likelihood:

$$l_C(\boldsymbol{\theta}; U) := \sum_{i=1}^{n} \ln(c(u_{i1}, u_{i2}; \boldsymbol{\theta}))$$

- Copula Akaike information criterion:

$$AIC_C(\boldsymbol{\theta}; U) := -2 \sum_{i=1}^{n} \ln(c(u_{i1}, u_{i2}; \boldsymbol{\theta})) + 2|\boldsymbol{\theta}|$$

- Copula Bayesian information criterion:

$$BIC_C(\boldsymbol{\theta}; U) := -2 \sum_{i=1}^{n} \ln(c(u_{i1}, u_{i2}; \boldsymbol{\theta})) + |\boldsymbol{\theta}| \ln(n)$$

where $\mathbf{u_i} = (u_{i1}, u_{i2})$ denotes the $i$-th observation of the pseudo copula data. In general a higher copula log-likelihood (smaller $\text{AIC}_C$ and $\text{BIC}_C$) indicates a better fit when comparing two different copulas.

Now, we are able to model two-dimensional dependencies using bivariate copulas. In the next sections we will introduce different approaches to also model $d$-dimensional data using bivariate copulas.

## 2.2 D-vine Pair Copula Constructions

As shown by Russell and Norvig (2003) we can model the density of any $d$-dimensional random vector using

$$\begin{aligned} f(x_1, ..., x_d) &= f(x_1) \times f(x_2|x_1) \times ... \times f(x_d|x_1, ..., x_{d-1}) \\ &= f(x_1) \times \prod_{i=2}^{d} f(x_i|x_1, ..., x_{i-1}) \end{aligned} \tag{2.7}$$

The question is, whether there is a way to substitute every conditional density with copula terms and marginal densities. One possibility is to decompose the density to a so called drawable D-vine density. Before doing so we have to adjust our notation to keep the upcoming terms readable.

**Definition 23.** (Copulas associated with bivariate conditional distributions) Let $(X_1, ..., X_d)$ be a set of random variables.

- Let $D$ be a set of indices from $\{1, ..., d\}$ not including $i$ and $j$. The copula associated with the bivariate conditional distribution $(X_i, X_j)$ given $\mathbf{X} = \mathbf{x}_D$ is denoted by $C_{ij;D}(\cdot, \cdot; \mathbf{x}_D)$
- In contrast the conditional distribution function of $(U_i, U_j)$ given $\mathbf{U}_D = \mathbf{u}_D$ is expressed as $C_{ij|D}(\cdot, \cdot; \mathbf{u}_D)$ with bivariate density function $c_{ij|D}(\cdot, \cdot; \mathbf{u}_D)$
- For distinct indices $i, j$ and $D = \{i_1, ..., i_k\}$ with $i < j$ and $i_1 < ... < i_k$ we use the abbreviation

$$c_{i,j;D} := c_{i,j;D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D); \mathbf{x}_D)$$

**Theorem 8.** (Drawable D-vine density) Using the abbreviations from Definition 23 we can decompose every density $f_{1,...,d}$ to

$$f_{1,...,d}(x_1, ..., x_d) = \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j);(i+1),...,(i+j-1)}\right] \left[\prod_{k=1}^{d} f_k(x_k)\right] \tag{2.8}$$

The distribution associated with this density decomposition is called a drawable D-vine.

*Proof.* Using Definition 1 for the conditional distribution of $(X_1, X_t)$ given $X_2, ..., X_{t-1}$ we can express the conditional density $f_{t|1,...,t-1}(x_t|x_2, ..., x_{t-1})$ recursively as

$$\begin{aligned} f_{t|1,...,t-1}(x_t|x_1, ..., x_{t-1}) &= c_{1,t|2,...,t-1} \times f_{t|2,...,t-1}(x_t|x_2, ..., x_{t-1}) \\ &= \left[\prod_{s=1}^{t-2} c_{s,t;s+1,...,t-1}\right] \times c_{(t-1),t} \times f_t(x_t) \end{aligned}$$

Plugging this into Equation (2.7) and setting $s = i$ and $t = i + j$ leads to

$$\begin{aligned} f_{1,\dots,d}(x_1, \dots, x_d) &= \left[\prod_{t=2}^{d}\prod_{s=1}^{t-2} c_{s,t;s+1,\dots,t-1}\right] \times \left[\prod_{k=1}^{d} f_k(x_k)\right] \\ &= \left[\prod_{j=1}^{d-1}\prod_{i=1}^{d-1} c_{i,(i+j);(i+1),\dots,(i+j-1)}\right] \times \left[\prod_{k=1}^{d} f_k(x_k)\right] \end{aligned}$$

□

Note that in Equation (2.8) we only have pair copula densities $c_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$ evaluated at conditional distribution functions $F_{i|D}(x_i|x_1, ..., x_{i_k})$ and $F_{j|D}(x_j|x_i, ..., x_{i_k})$ for specified indices $i, j, i_1, ..., i_k$ and marginal densities $f_k$. That is why this class of decompositions is called a pair copula decomposition. With this specific decomposition class named the class of D-vine compositions.

In the following we will introduce a way to compute the conditional distribution functions which we evaluate in the copula terms.

**Theorem 9.** (Recursion for conditional distribution functions) Let $X$ be a random variable and $\mathbf{Y}$ a random vector which has an absolute continuous joint distribution. Let $Y_j$ be a component of $\mathbf{Y}$ and denote the sub-vector of $\mathbf{Y}$ with $Y_j$ removed by $\mathbf{Y}_{-j}$. In this case the conditional distribution $F_{X|\mathbf{Y}}(\cdot|\mathbf{y})$ of $X$ given $\mathbf{Y} = \mathbf{y}$ satisfies the following recursion

$$F_{X|\mathbf{Y}}(\cdot|\mathbf{y}) = \frac{\partial C_{X,Y_j;\mathbf{Y}_{-j}}(F_{X|\mathbf{Y}_j}(x|\mathbf{y}_{-j}), F_{Y_j|\mathbf{Y}_j}(y|\mathbf{y}_{-j}))}{\partial F_{Y_j|\mathbf{Y}_{-j}}(y_j|\mathbf{y}_{-j})}$$

where $C_{X,Y_j;\mathbf{Y}_{-j}}(\cdot, \cdot|\mathbf{y}_{-j})$ denotes the copula corresponding to $(X, Y_j)$ given $\mathbf{Y}_{-j} = \mathbf{y}_{-j}$.

Hence, while increasing the conditional set in each step we can calculate the necessary conditional distribution function with terms that we already know. These are the copula and conditional distribution functions derived in the last step with a smaller conditioning set.

Note that in Definition 21 we have only defined a simplifying assumption for conditioning on one variable. In Equation (2.8) copulas with whole sets of conditioning variables might appear. Hence, we have to extend Definition 21 on bigger sets. In the following we will assume that Definition 24 holds.

**Definition 24.** (Simplifying assumption for D-vines) Assume that

$$c_{ij,D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D), \mathbf{x}_D) = c_{ij,D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D))$$

holds for all $\mathbf{x_D}$ and $i, j$ and $D$ are chosen to occur as in Equation (2.8) then the corresponding D-vine distribution is called simplified.

Bear in mind that the order in which we aligned the variables in the density function in Equation (2.8) was arbitrary. However, different orders will lead to different copulas being modeled. We will come back to this in Section 2.3 when we have a way to graphically express D-vines.

**Example 8.** (Simplified D-vine density on five dimensions) For $d = 5$ one way to express the simplified D-vine density is

$$\begin{aligned} f_{12345}(x_1, x_2, x_3, x_4, x_5) = &\left[\prod_{i=1}^{5} f_i(x_i)\right] \times c_{12}(x_1, x_2) \times c_{23}(x_2, x_3) \times c_{34}(x_3, x_4) \times c_{45}(x_4, x_5) \\ &\times c_{13,2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \times c_{24,3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\ &\times c_{35,4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)) \\ &\times c_{14,23}(F_{1|23}(x_1|x_2, x_3), F_{4|23}(x_4|x_2, x_3)) \\ &\times c_{25,34}(F_{2|34}(x_2|x_3, x_4), F_{5|34}(x_5|x_3, x_4)) \\ &\times c_{15|234}(F_{1|234}(x_1|x_2, x_3, x_4), F_{5|234}(x_5|x_2, x_3, x_4)) \end{aligned}$$

## 2.3 Graphical Representation of D-vines

In Section 2.2 we have seen how to decompose a density into a drawable D-vine but we did not speak about why this D-vine is called drawable. For this we first need some graph theoretic background.

**Definition 25.** (Graph, node, edge, adjacent, degree)

- A graph $G := (N, E)$ is a pair of sets such that $E \subseteq \{\{x, y\} : x, y \in N\}$
- Elements of $E$ are called edges of the graph $G$ while elements of $N$ are called nodes
- Two nodes are said to be adjacent if there is an edge $e \in E$ connecting them, otherwise they are said to be non-adjacent. Nodes that are adjacent to a node $x \in N$ are called neighbors of $x$
- The number of neighbors of a node $x \in N$ is the degree of $x$, denoted by $d(x)$

**Definition 26.** (Path, cycle, connected) A path from $v_0$ to $v_k$ is a graph $P' = (N', E')$ with node set $N' = \{v_0, v_1, ..., v_k\}$ and edges $E' = \{\{v_0, v_1\}, \{v_1, v_2\}, ..., \{v_{k-1}, v_k\}\}$ for any $k \in \mathbb{N}$. It is called a cycle if it also holds that $v_0 = v_k$. A graph $G := (N, E)$ is called connected if for any two nodes $x,\ y \in N$ there exists a path $P' = (N', E')$, $N' \subset N,\ E' \subset E$ from $x$ to $y$.

**Theorem 10.** (Characterization of trees) The following statements are equivalent for a graph $T = (N, E)$

- T is a tree
- Any two nodes of $T$ are connected by a unique path in $T$
- T is minimally connected, i.e., $T$ is connected but $T - e$ is disconnected for every edge $e$
- T is maximally acyclic, i.e., T contains no cycle but $T + \{x, y\}$ does for any non-adjacent nodes $x, y \in N$

We now have all the necessary basics to define a regular (R-)vine tree sequence. Our D-vine tree sequence will be a special case thereof.

**Definition 27.** (Regular (R-)vine tree sequence) A set of trees $\mathcal{T} = (T_1, ..., T_{d-1})$ is a regular vine tree sequence of $d$ elements if

- Each tree $T_j = (N_j, E_j)$ is connected
- $T_1$ is a tree with node set $N_1 = \{1, ..., d\}$ and edge set $E_1$
- For $j \geq 2,\ T_j$ is tree with node set $N_j = E_{j-1}$ and edge set $E_j$
- For $j = 2, ..., d-1$ and $\{a, b\} \in E_j$ it must hold that $|a \cap b| = 1$

The last property of Definition 27 ensures that if two nodes are connected by an edge $e$ in $T_j$, $j \geq 2$, then these two nodes, which are edges in $T_{j-1}$ share a common node. This is called the proximity condition.

As already mentioned, a D-vine sequence is a special case of Definition 27.

**Definition 28.** (D-vine tree sequence) A regular vine tree sequence $\mathcal{T} = (T_1, ..., T_{d-1})$ is called a D-vine sequence if for each node $n \in N_i$ we have $|\{e \in E_i | n \in e\}| \leq 2$, i.e., each node has at most two neighbors.

Following the proximity condition this means that once $T_1$ is fixed all other trees $T_2, ..., T_{d-1}$, are already uniquely determined.

**Example 9.** (Four-dimensional D-vine tree sequence) Given the first tree $T_1 = (N_1 = \{1, 2, 3, 4\},\ E_1 = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\})$ of a D-vine tree sequence. Then, the second and third tree are defined as

$$T_2 : N_2 = E_1 = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$$

$$E_2 = \{\{\{12\},\{23\}\},\{\{23\},\{34\}\}\}$$

$$\begin{aligned} T_3 : N_3 &= E_2 = \{\{\{12\},\{23\}\},\{\{23\},\{34\}\}\} \\ E_3 &= \{\{\{\{12\},\{23\}\},\{\{23\},\{34\}\}\}\} \end{aligned}$$

Already in the case with only four nodes the notation gets complicated. Therefore, we introduce a more simple one by using conditioning sets.

**Definition 29.** (Complete union and conditioning sets) Given a regular vine tree sequence $\mathcal{T}$. For any edge $e \in E_i$ we define the set

$$A_e := \{j \in N_1 | \exists e_1 \in E_1, ..., e_{i-1} \in E_{i-1} \text{ such that } j \in e_1 \in, ..., \in e_{i-1} \in e\}$$

The set $A_e$ is called the complete union of the edge $e$. The conditioning set $D_e$ of an edge $e = \{a, b\}$ is defined as

$$D_e := A_1 \cap A_b$$

and the conditioned sets $C_{e,a}$ and $C_{e,b}$ are given by

$$C_{e,a} := A_a \setminus D_e,\ C_{e,b} := A_b \setminus D_e \text{ and } C_e := C_{e,a} \cup C_{e,b}$$

Often the edges $e = (C_{e,a}, C_{e,b}; D_e)$ in the vine tree sequence are abbreviated by

$$e = (e_a, e_b; D_e)$$

**Example 10.** (Continuation of Example 9) Given the tree structure from Example 9 using the notation from Definition 29 the edge sets simplify to

$$\begin{aligned} E_1 &= \{(1,2),(2,3),(3,4)\} \\ E_2 &= \{(1,3;2),(2,4;3)\} \\ E_3 &= \{(1,4;2,3)\} \end{aligned}$$

Figure 2.3 displays a graphical representation of the structure.

Up to this point the R-vine sequence, with the special case of a D-vine sequence, is only a graph theoretic object. Next we want to link it with a stochastic component.

**Definition 30.** (Regular vine distribution) The joint distribution $F$ for the $d$-dimensional random vector $\mathbf{X} = (X_1, ..., X_d)$ has a regular vine distribution if we can specify the following triplet $(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{B}})$:

- **Marginal distributions**: $\boldsymbol{\mathcal{F}} = (F_1, ..., F_d)$ is a vector of continuous invertible marginal distribution functions representing the marginal distribution function of the random variables $X_i$, $i = 1, ..., d$
- **Regular vine tree sequence**: $\boldsymbol{\mathcal{T}}$ is a R-vine tree sequence on $d$ elements
- **Bivariate copulas**: $\boldsymbol{\mathcal{B}} = \{C_e | e \in E_i, i = 1, ..., d-1\}$ is the set of symmetric bivariate copula families $C_e$ with densities $c_e$ and $\boldsymbol{\theta} = \{\theta_{C_e} | e \in E_i, i = 1, ..., d-1\}$ the set of the corresponding parameters. Here, $E_i$ is the edge set of tree $T_i$ in the R-vine tree sequence $\boldsymbol{\mathcal{T}}$
- **Relationship between R-vine tree sequence $\boldsymbol{\mathcal{T}}$ and set $\boldsymbol{\mathcal{B}}$ of bivariate copulas**: For each $e \in E_i$, $i = 1, ..., d$, $e = \{a, b\}$, $C_e$ is the copula associated with the conditional distribution of $X_{C_{e,a}}$ and $X_{C_{e,b}}$ given $\mathbf{X}_{D_e} = \mathbf{x}_{D_e}$. Further $C_e(\cdot, \cdot)$ does not depend on the specific value of $\mathbf{x}_{D_e}$

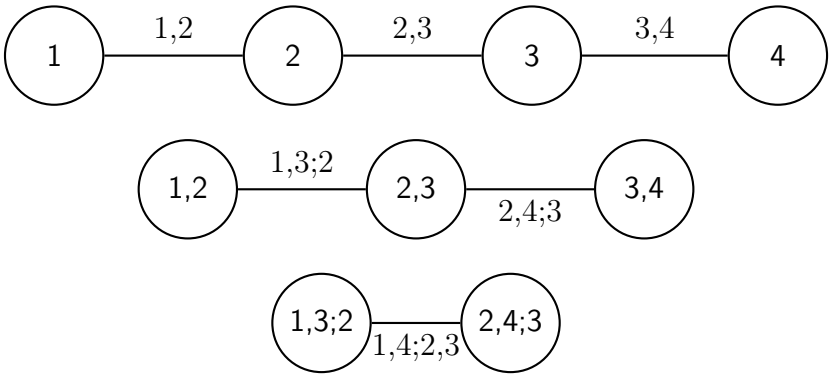



Figure 2.3: Representation of the D-vine tree structure from Example 10

As shown in Bedford and Cooke (2002) every R-vine triplet $(\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{B}})$ which satisfies the first three properties of Definition 30 can be uniquely connected to a $d$-dimensional distribution $F$. As we will solely use D-vines, for which we have already seen a representation of a $d$-dimensional distribution in Theorem 8, this is of no interest in the following. Instead we only needed a way to graphically represent D-vines.

What we still require is a procedure to denote which variable $X_1, ..., X_d$ is assigned to which node. For this purpose we define the order of a D-vine.

**Definition 31.** (Order of a D-vine) Given the first tree of a D-vine tree sequence as in Definition 28 on $d$ nodes with node set $N_1 = \{1, ...d\}$ and edge set $E_1 = \{(j, j+1),\ j = 1, ..., d-1\}$ as well as $d$ random variables, $X_1, ..., X_d$. We denote $\mathbf{l} = (l_1, ..., l_d)$ as the order of the D-vine if it is a bijection of the variables $X_i$ to the nodes in $N_1$. We say that $X_i^{\mathbf{l}}$ is the variable $X_j$ assigned to the node $i$ according to $\mathbf{l}$.

Referring to the variables assigned to the nodes greatly simplifies the notation, as the nodes and the respective variables assigned to them have the same index. This is something we will especially need in Section 2.5 and Section 2.6 as different orderings come in line with a different goodness of fit when we have to estimate the copulas.

Note that for each order $\mathbf{l}$ a second order $\mathbf{l}'$ exists such that if a variable $X_i$ is assigned to node $j$ in $\mathbf{l}$, it is assigned to node $d-j$ in $\mathbf{l}'$. However, both orderings correspond to the same D-vine and specify the same set of copulas. Instead the D-vine is only viewed at from the other direction.

## 2.4 Estimating d-Dimensional Densities Using Directed Acyclic Graphs (DAGs)

We have already seen in Equation (2.7) that we can write the density of any $d$-dimensional random vector as

$$f(x_1,...,x_d) = f(x_1) \times f(x_2|x_1) \times ... \times f(x_d|x_1,...,x_{d-1}) = f(x_1) \times \prod_{i=2}^{d} f(x_i|x_1,...,x_{i-1})$$

or in terms of a drawable non-simplified D-vine

$$f_{1,...,d}(x_1,...,x_d) = \left[\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,(i+j);(i+1),...,(i+j-1)}\right] \times \left[\prod_{k=1}^{d} f_k\right]$$

The problem is that already for an 11-dimensional density 55 copulas need to be modeled. This would grow to 190 copulas if a 20-dimensional density would be considered instead. To cope with that problem we need to find a way to simplify some of the terms by reducing the number of conditioning variables.

### 2.4.1 Conditional Independence

Given $d$ random variables the question arises: does each of the $d$ variables have a dependency with all of the others. And even more important does this also hold if all other variables are given. To this end, we now introduce the concept of conditional independence.

**Definition 32.** (Conditional density and conditional independence (Mathias Drton, 2009)) Let $\mathbf{X} = (X_1,...,X_d)$ be an $d$-dimensional random vector with joint density $f(\mathbf{x}) = f(x_1,...,x_d)$. For each subset $A \subseteq \{1,...,d\}$ let $\mathbf{X_A} = (X_a)_{a\in A}$ be the subvector of $\mathbf{X}$ indexed by $A$. The density $f_A(\mathbf{x_A})$ of $\mathbf{X_A}$ is the marginal density of the subset $A$.

Now, let $A, B \subseteq \{1, ..., d\}$ be two disjoint subsets. If $f_B(\mathbf{x_B}) > 0$, the conditional density of $\mathbf{X_A}$ given $\mathbf{X_B} = \mathbf{x_B}$ is defined as

$$f_{A|B}(\mathbf{x_A}|\mathbf{x_B}) := \frac{f_{A,B}(\mathbf{x_A}, \mathbf{x_B})}{f_B(\mathbf{x_B})} \tag{2.9}$$

where $f_{A,B}(\mathbf{x_A}, \mathbf{x_B})$ is the marginal density of the union of the subsets $A$ and $B$. The conditional density is undefined if $f_B(\mathbf{x_B}) = 0$. Now, let $A, B, C \subseteq \{1, ..., d\}$ be pairwise disjoint. We say the random vector $\mathbf{X_A}$ is conditionally independent of $\mathbf{X_B}$ given $\mathbf{X_C}$ if and only if

$$f_{A\cup B|C}(\mathbf{x_A}, \mathbf{x_B}|\mathbf{x_C}) = f_{A|C}(\mathbf{x_A}|\mathbf{x_C}) f_{B|C}(\mathbf{x_B}|\mathbf{x_C})$$

for all $\mathbf{x_A}, \mathbf{x_B}$ and $\mathbf{x_C}$ such that $f(\mathbf{x_C}) > 0$. In that case we write $\mathbf{X_A} \coprod \mathbf{X_B}|\mathbf{X_C}$ which we abbreviate in the following by $A \coprod B|C$.

From this definition there directly follow some properties, the so called conditional independence axioms following Mathias Drton (2009). We refer to the same work for the proof of the following proposition.

**Proposition 1.** Let $A, B, C, D \subseteq [d]$ be pairwise disjoint sets. Then,

- $A \coprod B|C \Rightarrow B \coprod A|C$ (symmetry)
- $A \coprod B \cup D|C \Rightarrow A \coprod B|C$ (decomposition)
- $A \coprod B \cup D|C \Rightarrow A \coprod B|C \cup D$ (weak union)
- $A \coprod B|C \cup D$ and $A \coprod D|C \Rightarrow A \coprod B \cup D|C$(contraction)

### 2.4.2 Modelling Conditional Dependencies Using DAGs

Now knowing which of our variables are conditional dependent or independent, we would like to represent these relationships in some way. For this we will use directed acyclic graphs, a special case of the graphs in Definition 25. To be able to distinguish between the two definitions we use a slightly different notation as well as a different style when plotting. Nodes in directed acyclic graphs are colored in black with white text whereas graphs that describe vine tree structures have white nodes and black text.

**Definition 33.** (Directed acyclic graphs (DAGs), (Wasserman, 2004)) A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consist of a set of vertices $\mathcal{V}$ and an edge set $\mathcal{E}$ of ordered pairs of vertices. A directed graph is called acyclic if there does not exist a directed path, i.e., a set of arrows all pointing in the same direction linking one node to another, for any node to itself.

In our case each vertex corresponds to a random variable. If $(X, Y)$ is part of the edge set $\mathcal{E}$ we display it by an arrow going from the vertex $X$ to the vertex $Y$.

**Definition 34.** (Parents) Given a directed acyclic graph. We denote with $\pi(X)$ all nodes which have an edge to a node $X$, i.e.,

$$\pi(X) := \{Y \in \mathcal{V} \text{ with } (Y, X) \in \mathcal{E}\}$$

We say the vertices in the set $\pi(X)$ are the parents of $X$.

The characterization of parents can now be used to give a DAG a topological order. For this we follow the definition of Koller and Friedman (2009).

**Definition 35.** (Topological order of a DAG) Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a DAG. An order $X_1, ..., X_d$ on the vertices of $\mathcal{G}$ is a topological order relative to $\mathcal{G}$ if $i < j$, whenever $X_i \to X_j \in \mathcal{E}$.

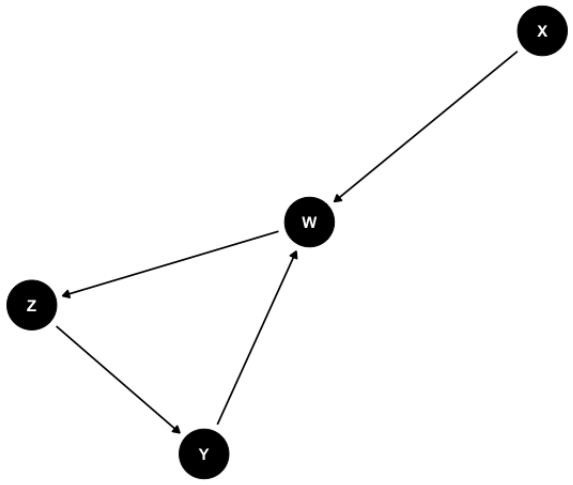



Figure 2.4: Example of a directed cyclic graph on four vertices

**Example 11.** Figure 2.5 shows a directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{W, X, Y, Z\}$ and edge set $\mathcal{E} = \{(X, W), (Y, W), (Z, W), (Z, Y)\}$. It follows, e.g., $\pi(W) = \{(Y, Z)\}$.

**Example 12.** Figure 2.4 shows a directed cyclic graph as there exists a directed path $W \to Z \to Y \to W$.

We further want to consider a procedure to model conditional dependencies using DAGs. For this we follow such a procedure proposed by Russell and Norvig (2003).

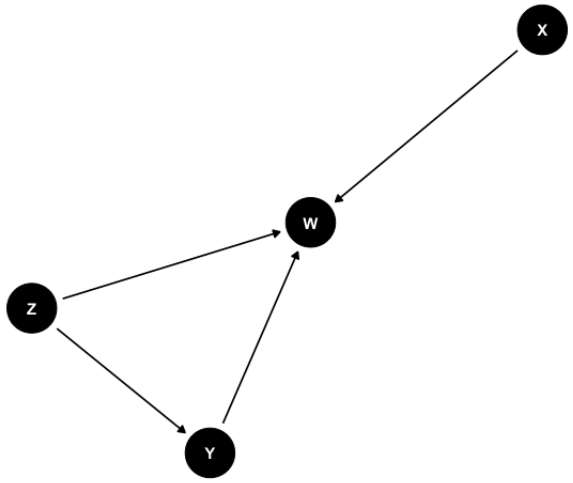



Figure 2.5: Example of a directed acyclic graph (DAG) on four vertices

**Definition 36.** (Modelling dependency structures using a DAG)

- Nodes: Given $d$ variables $X_1, ..., X_d$ in a topological order and assign each of them to one vertex
- Edges: FOR $i = 2 : d$ DO
  - Choose for $X_I$ the minimal set of parents $\pi(X_i) \subset \{X_1, ..., X_{i-1}\}$ such that

$$P(X_i|X_1, ..., X_{i-1}) = P(X_i|Parents(X_i)) \tag{2.10}$$

    In accordance with Definition 34 the right hand side can be written as $P(X_i|\pi(X_i))$
  - For each parent insert an edge from the parent to $X_i$

More precisely Equation (2.10) can be understood as: Find the biggest subset $\mathbf{X_{B_i}}$ of $\mathbf{X_{i-1}} = (X_1, ..., X_{i-1})$ such that

$$P(X_i|\mathbf{X_{i-1}}) = P(X_i|\mathbf{X_{i-1}} \setminus \mathbf{X_{B_i}}) \text{ for } i = 2, ..., d$$

After some calculations we can see that this subset $\mathbf{X_{B_i}}$ can easily be interpreted. For this we define $\mathbf{X_{A_i}} := \mathbf{X_{i-1}} \setminus \mathbf{X_{B_i}}$ which corresponds to the parents of $X_i$. We obtain

$$\begin{aligned} P(X_i|\mathbf{X_{i-1}}) &= P(X_i|\mathbf{X_{i-1}} \setminus \mathbf{X_{B_i}}) \\ \Leftrightarrow P(X_i|\mathbf{X_{i-1}}) &= P(X_i|\mathbf{X_{A_i}}) \\ \Leftrightarrow \frac{P(X_1, ..., X_i)}{P(X_1, ..., X_{i-1})} &= \frac{P(X_i, \mathbf{X_{A_i}})}{P(\mathbf{X_{A_i}})} \end{aligned}$$

$$
\begin{aligned}
&\Leftrightarrow \frac{P(X_i, \mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(\mathbf{X_{A_i}}, \mathbf{X_{B_i}})} = \frac{P(X_i, \mathbf{X_{A_i}})}{P(\mathbf{X_{A_i}})} \\
&\Leftrightarrow \frac{P(X_i, \mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(X_i, \mathbf{X_{A_i}})} = \frac{P(\mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(\mathbf{X_{A_i}})} \\
\Leftrightarrow &\frac{P(X_i, \mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(\mathbf{X_{A_i}})} \frac{P(\mathbf{X_{A_i}})}{P(X_i, \mathbf{X_{A_i}})} = \frac{P(\mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(\mathbf{X_{A_i}})} \\
&\Leftrightarrow \frac{P(X_i, \mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(\mathbf{X_{A_i}})} = \frac{P(X_i, \mathbf{X_{A_i}})}{P(\mathbf{X_{A_i}})} \frac{P(\mathbf{X_{A_i}}, \mathbf{X_{B_i}})}{P(\mathbf{X_{A_i}})} \\
&\Leftrightarrow P(X_i, \mathbf{X_{B_i}} | \mathbf{X_{A_i}}) = P(X_i | \mathbf{X_{A_i}}) P(\mathbf{X_{B_i}} | \mathbf{X_{A_i}})
\end{aligned}
$$

Recalling Equation (2.9) this means $X_i$ is conditionally independent of all elements in $\mathbf{X_{B_i}}$ given $\mathbf{X_{A_i}}$. Or in other words: a vertex $X_j$ with $j < i$ exactly is a parent of $X_i$ if it is not conditional independent from $X_i$ given the other parents.

Hence, if we are able to construct a DAG using the procedure from Definition 36, it is enough to only condition every variable on the set of its parents instead of all variables that appear before it in the topological order. Inserting this into Equation (2.7) we observe that it simplifies to

$$
f(x_1, ..., x_d) = \prod_{i=1}^{d} f(x_i | \pi(X_i)) \tag{2.11}
$$

leaving us with fewer dependencies to model. Following Pearl (1988) this structure is now called a Bayesian Network.

### 2.4.3 Linear Gaussian Bayesian Networks (LGBNs)

Section 2.4.2 still leaves us with the question on how to model the $f(x_i|\pi(X_i))$, to which we will refer to as conditional probability distributions (CPDs). Before coming up with an approach involving copulas, a first possibility is to use conditional linear Gaussian probability distributions. In this section we will closely follow Koller and Friedman (2009).

**Definition 37.** (Linear Gaussian conditional probability distribution) Let $Y, X_1, ..., X_k$ be random variables. Using the notation from Example 1, we say $Y$ follows a linear Gaussian model of $X_1, ..., X_k$, if there are parameters $\boldsymbol{\beta} = (\beta_0, ..., \beta_k)^\top$ such that

$$
\begin{aligned}
(Y | X_1, ..., X_k) &\sim N(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k; \sigma^2) \\
&= N(\beta_0 + \boldsymbol{\beta}^\top \mathbf{x}; \sigma^2)
\end{aligned}
$$

It shows that, $Y$ is a linear function of the parameters $X_1, ..., X_k$ with addition of Gaussian noise $\epsilon$ with mean 0 and variance $\sigma^2$, i.e.,

$$Y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \epsilon$$

If this holds for all variables in a network on the set of its parents, this defines a Linear Gaussian Bayesian Network.

**Definition 38.** (Linear Gaussian Bayesian Network (LGBN)) Given a DAG as in Definition 33 where every conditional probability distribution is of the form given in Definition 37 then the network is called a Linear Gaussian Bayesian Network (LGBN).

We now have a way to model all the CPDs necessary to calculate Equation (2.11) which describes a Linear Gaussian Bayesian Network. Even more the LGBN will have a density of a well known form.

**Theorem 11.** Let $Y$ follow a Linear Gaussian conditional probability distribution of its parents $X_1, ..., X_k$ as in Definition 37. If $X_1, ..., X_k$ are jointly Gaussian distributed $N(\boldsymbol{\mu}; \Sigma)$ with $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)^\top$ and $\Sigma = (\Sigma_{ij})_{i,j \in [1,k]}$, then

- The joint distribution of $(X_1, ..., X_k, Y)$ is a normal distribution with

$$\mathrm{Cov}(X_i; Y) = \sum_{j=1}^{k} \beta_j \Sigma_{i,j}$$

- The unconditional distribution of $Y$ is a normal distribution with
  - $\mu_Y = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{\mu}$
  - $\sigma_Y^2 = \sigma^2 + \boldsymbol{\beta}^\top \Sigma \boldsymbol{\beta}$
- Or the other way around, given a multivariate normal distribution the conditional distribution of $Y$ given $\mathbf{X} = (X_1, ..., X_k)$ is

$$(Y|\mathbf{X}) \sim N(\beta_0 + \boldsymbol{\beta}^\top \mathbf{X}; \sigma^2)$$

  with
  - $\beta_0 = \mu_Y - \Sigma_{Y\mathbf{X}} \Sigma_{\mathbf{XX}}^{-1} \boldsymbol{\mu}_{\boldsymbol{X}}$
  - $\boldsymbol{\beta} = \Sigma_{\mathbf{XX}}^{-1} \Sigma_{Y\mathbf{X}}$
  - $\sigma^2 = \Sigma_{YY} - \Sigma_{Y\mathbf{X}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{X}Y}$

where we decompose the joint distribution over $\mathbf{X}$ and $Y$ to

$$(\mathbf{X}, Y) \sim N\left(\begin{pmatrix}\boldsymbol{\mu_X}\\ \mu_Y\end{pmatrix}; \begin{bmatrix}\Sigma_{\mathbf{XX}} & \Sigma_{\mathbf{X}Y}\\ \Sigma_{Y\mathbf{X}} & \Sigma_{YY}\end{bmatrix}\right)$$

with $\boldsymbol{\mu_X} \in \mathbb{R}^k$, $\mu_Y \in \mathbb{R}$, $\Sigma_{\mathbf{XX}} \in \mathbb{R}^{k\times k}$, $\Sigma_{Y\mathbf{X}} = \Sigma_{\mathbf{X}Y}^\top \in \mathbb{R}^{1\times k}$ and $\Sigma_{YY} \in \mathbb{R}^{1\times 1}$.

**Theorem 12.** Given a Linear Gaussian Bayesian Network the specified joint distribution follows a multivariate Gaussian distribution.

*Proof.* Proof by induction:

- Base case: Given any node without parents. Following Definition 37 it is Gaussian distributed as it follows a $N(\beta_0; \sigma^2)$ distribution
- Induction step: Assume all parents of a node follows a Gaussian distribution. Then, following Theorem 11 the child node together with the parent nodes follow a multivariate Gaussian distribution

□

### 2.4.4 Modelling Conditional Densities Using D-vines

We now want to introduce a way to model the conditional densities using copulas. We start from Theorem 1 but this time assume that the D-vine is simplified as in Definition 24. Using the abbreviations from Definition 23 we follow the idea of the proof of Theorem 8 and recursively obtain

$$f_{t|1,\dots,t-1}(x_t|x_1,\dots,x_{t-1}) = \left[\prod_{s=1}^{t-2} c_{s,t;s+1,\dots,t-1}\right] \times c_{(t-1),t} \times f_t(x_t)$$

Reversing the order such that we have the density of the first variable given the rest, we derive

$$f_{1|2,\dots,t}(x_1|x_2,\dots,x_t) = \left[\prod_{s=2}^{t-1} c_{1,s+1;2,\dots,s}\right] \times c_{1,2} \times f_1(x_1) \tag{2.12}$$

This is the desired result, i.e., we can express any conditional density of a variable given its parents only using its marginal density and copula terms.

**Example 13.** (Conditional density in a D-vine on four nodes) Given four variables $X_1$, $X_2$, $X_3$ and $X_4$ which we model using the D-vine tree structure from Figure 2.3 and its associated regular D-vine distribution. Assigning $X_i$ to node $i$ we can write the conditional density of $X_1$ given $X_2$, $X_3$ and $X_4$ by

$$f_{1|2,\dots,t}(x_1|x_2,x_3,x_4) = c_{1,4;2,3} \times c_{1,3;2} \times c_{1,2} \times f_1(x_1) \tag{2.13}$$

Looking at the conditional density in Example 13 we observe that the copulas appearing in the expression of the density are exactly the first copulas in each tree of the associated D-vine structure in Figure 2.3. We can see this behavior for any length of the D-vine. Hence, to express the density of a variable given the set of its parents, it is sufficient to model them using a D-vine where the desired variable is modeled as the first node. Then, only the most left copula of each tree is needed to express the conditional density.

### 2.4.5 Special Case: Gaussian D-vines

We observed in Section 2.4.3 that we can interpret the Linear Gaussian Bayesian Network as a set of conditional normal distributions. For the D-vine approach there is a similar conclusion for the special case of a Gaussian D-vine following Morales et al. (2008). It gives a convenient specification of a multivariate normal distribution.

**Definition 39.** (Gaussian D-vine) Given a $d$-dimensional D-vine such that every modeled copula is a Gaussian copula and every marginal follows a Gaussian distribution. Then, the D-vine is called a Gaussian D-vine. Furthermore, each $d$-dimensional D-vine specifies a $d$-dimensional normal distribution $N(\boldsymbol{\mu};\Sigma)$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_1^2 & p_{1,2}\sigma_1\sigma_2 & \cdots & p_{1,d}\sigma_1\sigma_d \\ p_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_{d-1,d}\sigma_{d-1}\sigma_d \\ p_{1,d}\sigma_1\sigma_d & \cdots & p_{d-1,d}\sigma_{d-1}\sigma_d & \sigma_d^2 \end{pmatrix}.$$

The $\mu_i$ and $\sigma_i^2$ $i = 1, ..., d$ are given by the mean and the variance of the $i$-th marginal and the correlations $p_{i,j}$ are specified by the copulas.

This leaves us with the translation of the copula parameters to the entries of the correlation matrix. Each copula $c_{i,i+1}$ of the first tree defines the $i, i+1$ entry of the correlation matrix by its parameter with $p_{i,i+1} = \rho_{i,i+1}$. Since we only have $d-1$ copulas in the first tree but need $\frac{d^2-d}{2}$ correlations this is not sufficient.

To overcome this problem we follow Kendall and Stuart (1961) proposing that in multivariate normal distributions conditional correlations, which we have given by the parameters defined by the copulas in the lower trees, equal partial correlations. Hence, we can use Equation (2.5) to calculate the missing correlations.

**Example 14.** (Correlation calculation in a four-dimensional Gaussian D-vine) Given a D-vine on four variables $U_1, U_2, U_3$ and $U_4$, in this order, which have been transformed from $X_1, X_2, X_3$ and $X_4$ using Gaussian margins, i.e., each variable $X_i$ follows a $N(\mu_i; \sigma_i)$ distribution. Further assume that it each copula of the D-vine is a Gaussian copula. Then, $p_{1,2}, p_{2,3}$ and $p_{3,4}$ are already defined by the parameters of the $c_{1,2}, c_{2,3}$ and $c_{3,4}$ copula, i.e., $p_{1,2} = \rho_{1,2}$ , $p_{2,3} = \rho_{2,3}$ and $p_{3,4} = \rho_{3,4}$. Apart from this the parameters of the copulas from the deeper trees define conditional correlations, i.e., $\rho_{1,3;2}$, $\rho_{2,4;3}$ and $\rho_{1,4;2,3}$. Rearranging Equation (2.5) we can calculate $p_{1,3}$ and $p_{2,4}$ by

$$p_{1,3} = \rho_{1,3} = \rho_{1,3;2} \times \sqrt{((1-\rho_{1,2}^2) \times (1-\rho_{2,3}^2))} + \rho_{1,2} \times \rho_{2,3}$$
$$p_{2,4} = \rho_{2,4} = \rho_{2,4;3} \times \sqrt{((1-\rho_{2,3}^2) \times (1-\rho_{3,4}^2))} + \rho_{2,3} \times \rho_{3,4}$$

Now, only the $p_{1,4}$ correlation is missing. To calculate it we need to repeat the application of Equation (2.5) until we have all the necessary values to calculate $p_{1,4}$.

$$p_{3,4;2} = \rho_{3,4;2} = \frac{\rho_{3,4} - \rho_{2,3} \times \rho_{2,4}}{\sqrt{(1-\rho_{2,3}^2)(1-\rho_{2,4}^2)}}$$
$$p_{1,4;2} = \rho_{1,4;2} = \rho_{1,4;2,3} \times \sqrt{((1-\rho_{1,3;2}^2) \times (1-\rho_{3,4;2}^2))} + \rho_{1,3;2} \times \rho_{3,4;2}$$

and then finally

$$p_{1,4} = \rho_{1,4} = \rho_{1,4;2} \times \sqrt{((1-\rho_{1,2}^2) \times (1-\rho_{2,4}^2))} + \rho_{1,2} \times \rho_{2,4}$$

This gives us all the needed correlations to specify $\Sigma$ and therefore the multivariate normal distribution.

As we have already seen in Theorem 11 the conditional distribution of a multivariate normal distribution is again a normal distribution. Equation (2.13) therefore also specifies a conditional normal distribution. Hence, we can represent the Gaussian D-vine as a Gaussian distribution. The only difference to fitting the Linear Gaussian Bayesian Network directly, as in Section 2.4.3, is that there the optimization is performed in one step. For the D-vine first the marginals have to be estimated to then be able to estimate the corresponding copulas. Therefore we expect very similar outcomes in a LGBN and a Gaussian D-vine, with the D-vine possibly fitting a little bit worse.

**Example 15.** (Linear Gaussian conditional probability distribution specified by a four-dimensional D-vine) Given the D-vine of the four variables $X_1, X_2, X_3$ and $X_4$ as in Example 14. Then, it specifies a four-dimensional Gaussian distribution $N(\boldsymbol{\mu}; \Sigma)$ as defined in Definition 39. Following Theorem 11 we can calculate the conditional distribution of $X_1$ given $X_2, X_3$ and $X_4$. By decomposing the joint density we obtain

$$\mu_{\mathbf{X}} = \begin{pmatrix} \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \ \mu_Y = \mu_1, \ \Sigma_{\mathbf{XX}} = \begin{pmatrix} \sigma_2^2 & p_{2,3}\sigma_2\sigma_3 & p_{2,4}\sigma_2\sigma_4 \\ p_{2,3}\sigma_2\sigma_3 & \sigma_3^2 & p_{3,4}\sigma_3\sigma_4 \\ p_{2,4}\sigma_2\sigma_4 & p_{3,4}\sigma_3\sigma_4 & \sigma_4^2 \end{pmatrix}, \ \Sigma_{\mathbf{X}Y} = \begin{pmatrix} p_{1,2}\sigma_1\sigma_2 \\ p_{1,3}\sigma_1\sigma_3 \\ p_{1,4}\sigma_1\sigma_4 \end{pmatrix},$$

$\Sigma_{Y\mathbf{X}} = \Sigma_{\mathbf{X}Y}^\top$ and $\Sigma_{YY} = \sigma_4^2$

Plugging the above into the equations from Theorem 11

- $\beta_0 = \mu_Y - \Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\boldsymbol{\mu}_{\boldsymbol{X}}$
- $\boldsymbol{\beta} = \Sigma_{\mathbf{XX}}^{-1}\Sigma_{Y\mathbf{X}}$
- $\sigma^2 = \Sigma_{YY} - \Sigma_{Y\mathbf{X}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{X}Y}$

we obtain the necessary parameters to define the conditional distribution of $X_1$ given $X_2 = x_2, \ X_3 = x_3$ and $X_4 = x_4$. Denoting $\mathbf{x} = (x_2, x_3, x_4)$ we can write it as

$$(X_1 | X_2, X_3, X_4) \sim N(\beta_0 + \boldsymbol{\beta}^\top \mathbf{x}; \sigma^2)$$

Apart from knowing that the node corresponding to the variable we want to model has to be the first node in the D-vine structure, we do not know in which order the other variables appear in the structure. As the explanatory power depends on the specific order we want to find a way to order these variables in a way that gives us the best fit without having to test all possible combinations.

## 2.5 Structure Selection

Section 2.5 is based on Kraus and Czado (2017).

Given the variables $Y, X_1, ..., X_d$ and we want to model the conditional density of $Y$ given $X_1, ..., X_d$ using a D-vine. Then, $Y$ has to be assigned to the first node but for the $X_1, ..., X_d$ there are $d!$ possible orderings on the remaining nodes of the D-vine. It does not seem reasonable to try out all these possibilities to see which results in the ordering $\mathbf{l}$ that maximizes the conditional goodness of fit. Instead we consider a procedure where we stepwise improve the fit. We do this such that in the $k$-th step, $k = 1, ..., d-1$, the $k-1$

nodes right of $Y$ are known and we only find the best node to be at the $k$-th position right of $Y$. Hence, after $d-1$ steps we have an order $\mathbf{l}$ of the variable $X_1, ..., X_d$ that provides us with the best goodness of fit. To formally introduce this we first need to define how we measure this goodness of fit.

**Definition 40.** (Conditional copula log-likelihood) Given an estimated D-vine on the data set $(\mathbf{y}, X)$ with ordering $\mathbf{l}$, pair copula families $\boldsymbol{\mathcal{B}}$, corresponding parameters $\boldsymbol{\theta}$ and pseudo copula data $(\mathbf{v}, U)$ constructed by applying the PIT using the marginals $\boldsymbol{\mathcal{F}}$. Then the conditional copula log-likelihood is defined as

$$cll_C(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}; \mathbf{v}, U) := \sum_{i=1}^{n} \ln c_{V|U}(v^{(i)}|\mathbf{u}^{(i)}; \mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}) \tag{2.14}$$

with $c_{V|U}$ being the right-hand side of Equation (2.12) without the marginal density so

$$c_{V|U} := \left[\prod_{s=2}^{t-1} c_{1,s+1;2,...,s}\right] \times c_{1,2}$$

Similar to Definition 22 we extend Definition 40 to penalize the use of parameters.

**Definition 41.** (AIC- and BIC-corrected conditional copula log-likelihood) Given the same setting as in Definition 40 the Akaike information criterion ($cll_C^{AIC}$) and Bayesian information criterion ($cll_C^{BIC}$) corrected conditional copula log-likelihood are defined as

$$cll_C^{AIC}(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}; \mathbf{v}, U) := -2\ cll_C(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}; \mathbf{v}, U) + 2|\boldsymbol{\theta}|$$
$$cll_C^{BIC}(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}; \mathbf{v}, U) := -2\ cll_C(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}; \mathbf{v}, U) + \log(n)|\boldsymbol{\theta}|$$

In the following we will often refer to the contribution of each separate copula in $c_{V|U}$. We then speak of the copula log-likelihood of a specific copula even though, if we do not speak of the $c_{1,2}$ copula, it is technically a conditional copula log-likelihood. The same holds for the $cll_C^{AIC}$ and $cll_C^{BIC}$ to which we, in correspondence with Definition 22, refer to as $AIC_C$ and $BIC_C$.

After having found a measure to quantify the fit of a D-vine we continue by introducing a procedure to sequentially order the $d$ predictor nodes in the D-vine. Assume that at the beginning of the $k$-th step of the algorithm the current optimal D-vine contains $k-1$ predictors. For each of the remaining variables $X_j$ that have not been chosen yet, we fit the pair copulas that are needed to extend the model to a D-Vine with $X_j$ as the next node. Then, the current model is updated by adding the variable corresponding to the highest $cll_C$, $cll_C^{AIC}$ or $cll_C^{BIC}$ depending on the measure of choice, concluding step $k$. This way, step by step, the covariates are ordered regarding their power to predict the response. If at some step $k$ it is not possible to choose a covariate to increase the measure of choice, then the algorithm terminates and the model only contains $k-1$ predictors.

## 2.6 Model Comparison

To this point, we have found two fundamentally different ways to model our data with the help of graphical models by either modeling each conditional density as Linear Gaussian CPD or using a D-Vine. In the second approach we can additionally vary the marginals and copula families. Now, we are interested in a measure to compare these models with each other. We will again use versions of the log-likelihood, and later the Akaike information criterion and Bayesian information criterion for this. First however, we introduce a way to decompose the log-likelihood of the whole model into the log-likelihoods of each conditional density following Koller and Friedman (2009).

**Theorem 13.** (Global log-likelihood decomposition) Given a Bayesian Network on the nodes $\mathbf{X} = (X_1, ..., X_d)$ which is used to model a data set $X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix}$ with parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_r)$. Then, the global log-likelihood of the model can be decomposed into

$$l_F(\boldsymbol{\theta}, X) = \sum_{i=1}^{d} \ln(l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)}))$$

where $l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)})$ is the local likelihood function, i.e., the conditional density function of the node $X_i$ given the parents $\pi(X_i)$ of $X_i$ and the parameters defining the CDF. In our case the set $\boldsymbol{\theta}$ is a disjoint set for the different CDFs, i.e., each conditional density is modeled with its own set of parameters.

The remaining question is how the $l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)})$ look like in the different models. For the LGBN we have already seen this in Section 2.4.3.

**Theorem 14.** (Local likelihood function in the Linear Gaussian Bayesian Network) In the Linear Gaussian Bayesian Network the local likelihood function is a conditional Gaussian density following Theorem 11

$$l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)}) \sim N(\beta_0 + \boldsymbol{\beta}^\top \pi(X_i); \sigma^2)$$

with $\boldsymbol{\theta}_{X_i|\pi(X_i)} = (\beta_0, \boldsymbol{\beta})$.

Hence, in the Linear Gaussian Bayesian Network the global log-likelihood is a sum over log-likelihoods of conditional normal densities. For the D-vine models it is even more simple as the local log-likelihood can be expressed by a sum over two measures which we have already defined.

**Theorem 15.** (Local log-likelihood in the D-vine model) In the D-vine model on the data set $X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix}$ the local log-likelihood of the node $i$ with ordering $\mathbf{l}$, pair copula families $\boldsymbol{\mathcal{B}}$ and parameters $\boldsymbol{\theta}_{X_i|\pi(X_i)}$ can be expressed as

$$l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)}) = \; cll_C(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}_{C,X_i|\pi(X_i)}; \mathbf{u_i}, U) \times l_M(\boldsymbol{\theta}_{M,X_i|\pi(X_i)}; x_i)$$

Here $U = \begin{pmatrix} u_{11} & \cdots & u_{1d} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nd} \end{pmatrix}$ denotes the data set after applying the PIT with the chosen marginals $f_1, ..., f_d$ to the data set $X$. Furthermore, $\boldsymbol{\theta}_{C,X_i|\pi(X_i)}$ and $\boldsymbol{\theta}_{M,X_i|\pi(X_i)}$ denote the parameters of the copulas and the marginals conditional on the parents of $X_i$ in the model.

*Proof.* For D-vine models we have already seen a representation of the conditional densities in Equation (2.12). For readability reasons we will refer to the variables assigned to the nodes according to $\mathbf{l}$ as introduced in Definition 31, i.e., $X_j^{\mathbf{l}}$ is the variable assigned to node $j$ and especially $X_1^{\mathbf{l}} = X_i$. Hence, we can write the conditional density as

$$f(x_i|\pi(X_i)) = f_{1|2,...,t}(x_1^{\mathbf{l}}|x_2^{\mathbf{l}}, ..., x_t^{\mathbf{l}}) = \left[\prod_{s=2}^{t-1} c_{1,s+1;2,...,s}\right] \times c_{1,2} \times f_1(x_1^{\mathbf{l}})$$

Now, we take the logarithm and use that we model the copulas and the marginals on distinct parameter sets. Therefore, $\boldsymbol{\theta}_{X_i|\pi(X_i)}$ decomposes to $\boldsymbol{\theta}_{X_i|\pi(X_i)} = (\boldsymbol{\theta}_{C,X_i|\pi(X_i)}, \boldsymbol{\theta}_{M,X_i|\pi(X_i)})$, and we can see that this splits up in into terms that we have already defined

$$\begin{aligned}
&\ln(l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)})) = \ln(f_{1|2,...,t}(\mathbf{x}_1^{\mathbf{l}}|\mathbf{x}_2^{\mathbf{l}}, ..., \mathbf{x}_t^{\mathbf{l}}, \boldsymbol{\theta}_{X_i|\pi(X_i)})) \\
&= \ln\left(\prod_{s=2}^{t-1} c_{1,s+1;2,...,s}(F_{1|2,...,s}(\mathbf{x}_1^{\mathbf{l}}|x_{2,...,s}^{\mathbf{l}}), F_{s+1|2,...,s}(\mathbf{x}_{s+1}^{\mathbf{l}}|x_{2,...,s}^{\mathbf{l}})); x_{2,...,s}^{\mathbf{l}}, \boldsymbol{\theta}_{C,X_i|\pi(X_i)})\right) \\
&\quad \times \ln\left(c_{1,2}(F_1(\mathbf{x}_1^{\mathbf{l}}), F_2(\mathbf{x}_2^{\mathbf{l}}))\boldsymbol{\theta}_{C,X_i|\pi(X_i)} \times f_1(\mathbf{x_1}^{\mathbf{l}}; \boldsymbol{\theta}_{M,X_i|\pi(X_i)})\right) \\
&= \ln\left(\prod_{j=1}^{n}\prod_{s=2}^{t-1} c_{1,s+1;2,...,s}(F_{1|2,...,s}(x_{j,1}^{\mathbf{l}}|\mathbf{x}_{j,2,...,s}^{\mathbf{l}}), F_{s+1|2,...,s}(x_{j,s+1}^{\mathbf{l}}|\mathbf{x}_{j,2,...,s}^{\mathbf{l}}); \mathbf{x}_{j,2,...,s}^{\mathbf{l}}, \boldsymbol{\theta}_{C,X_i|\pi(X_i)})\right) \\
&\quad \times \ln\left(\prod_{j=1}^{n} c_{1,2}(F_1(x_{j,1}^{\mathbf{l}}), F_2(x_{j,2}^{\mathbf{l}}); \boldsymbol{\theta}_{C,X_i|\pi(X_i)}) \times f_1(x_{j,1}^{\mathbf{l}}, \boldsymbol{\theta}_{M,X_i|\pi(X_i)})\right) \\
&= \sum_{j=1}^{n}\sum_{s=2}^{t-1} \ln\left(c_{1,s+1;2,...,s}(F_{1|2,...,s}(x_{j,1}^{\mathbf{l}}|\mathbf{x}_{2,...,s}^{\mathbf{l}}), F_{s+1|j,2,...,s}(x_{j,s+1}^{\mathbf{l}}|\mathbf{x}_{j,2,...,s}^{\mathbf{l}}); \mathbf{x}_{j,2,...,s}^{\mathbf{l}}, \boldsymbol{\theta}_{C,X_i|\pi(X_i)})\right) \\
&\quad + \sum_{j=1}^{n} \ln\left(c_{1,2}(F_1(x_{j,1}^{\mathbf{l}}), F_2(x_{j,2}^{\mathbf{l}}); \boldsymbol{\theta}_{C,X_i|\pi(X_i)}) + f_1(x_{j,1}^{\mathbf{l}}, \boldsymbol{\theta}_{M,X_i|\pi(X_i)})\right)
\end{aligned}$$

Following Equation (2.14) and Equation (2.3) we can rewrite this as

$$= cll_C(\mathbf{l}, \boldsymbol{\mathcal{B}}, \boldsymbol{\theta}_{C,X_i|\pi(X_i)}; \mathbf{u_i}, U) \times l_M(\boldsymbol{\theta}_{M,X_i|\pi(X_i)}; \mathbf{x}_i)$$

with $\boldsymbol{\mathcal{B}}$ being the chosen pair copula families in the model and $\mathbf{l}$ the order of the D-vine which we denoted by $X_2, ..., X_t$.

□

Hence, to calculate the local log-likelihood of a conditional density of a node in the D-vine model we can sum up the marginal log-likelihood and the copula log-likelihood of the node. Furthermore, to calculate the global log-likelihood of the model it is sufficient to sum this up over all nodes.

Taking the number of parameters in the model into account, we can again define two different versions of the Akaike- and Bayesian information criterion.

**Definition 42.** (Global and local Akaike- and Bayesian information criterion) Given the global and local log-likelihood as in Theorem 13, then the local Akaike- and Bayesian information criterion are defined as

- Local Akaike information criterion:

$$AIC_{F,i}(\boldsymbol{\theta}_{X_i|\pi(X_i)}, X) := -2l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)}) + 2|\boldsymbol{\theta}_{X_i|\pi(X_i)}|$$

- Local Bayesian information criterion:

$$BIC_{F,i}(\boldsymbol{\theta}_{X_i|\pi(X_i)}, X) := -2l_{F,i}(X; \boldsymbol{\theta}_{X_i|\pi(X_i)}) + \ln(n) \times |\boldsymbol{\theta}_{X_i|\pi(X_i)}|$$

Additionally, the global Akaike- and Bayesian information criterion are defined as the sum over the local Akaike- and Bayesian information criterion of all nodes, respectively.

Another way to compare models is to compare their ability to recreate the data set on which they have been fitted. When fitting models which only consist of conditional densities it might further be important to see how these conditional densities behave when conditioned on different values. For both it is crucial to be able to sample from the conditional densities given certain conditioning values. We have seen that the conditional densities in the Linear Gaussian Bayesian Network each follow a Gaussian density. Hence, here we only need to simulate from a Gaussian distribution. This can for example be done by following the approach of Box and Muller (1958) and then rescaling according to the parameters.

The tricky part about sampling from D-vine models is that we do not generally want to sample the D-vine. An algorithm for this can for example be found in Czado (2019).

Instead we want to sample from the D-vine given some of the covariates. For this, we refer to the sampling algorithm from Bevacqua et al. (2017).

**Theorem 16.** (Conditional sampling from a D-vine) To create a uniform sample $\mathbf{U} = (U_1, ..., U_{N_{cond}}, U_{N_{cond}+1}, ..., U_d)$ from a $d$-dimensional D-vine given the corresponding pair copulas, parameter matrix $\Theta$ and the values $U_1, ..., U_{N_{cond}} = u_1^{cond}, ..., u_{N_{cond}}^{cond}$ it is sufficient to follow the upcoming steps:

- Sample $w_{N_{cond}+1}, ..., w_d$ independent from $[0, 1]$
- IF $N_{cond} \neq 0$ THEN
  - FOR $i = 1, ..., N_{cond}$ DO
    - $w_i = u_i^{cond}$
  - END FOR
- END IF
- $u_1 = v_{1,1} = w_1$
- IF $N_{cond} < 2$ THEN
  - $u_2 = v_{2,1} = h^{-1}(w_2, v_{1,1}, \theta_{1,1})$
- ELSE
  - $u_2 = v_{2,1} = w_2$
- END IF
- $v_{2,2} = h(v_{1,1}, v_{2,1}, \theta_{1,1})$
- FOR $i = 3, ..., d$ DO
  - $v_{i,1} = w_i$
  - IF $i > N_{cond}$ then
    - FOR $k = i-1, i-2, ..., 2$ DO
      - $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,2k-2}, \theta_{k,i-k})$
    - END FOR
    - $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,1}, \theta_{1,i-1})$
  - END IF
  - $u_i = v_{i,1}$

  - IF $i = d$ THEN
    - STOP
  - END IF
  - $v_{i,2} = h(v_{i-1,1}, v_{i,1}, \theta_{1,i-1})$
  - $v_{i,3} = h(v_{i,1}, v_{i-1,1}, \theta_{1,i-1})$
  - IF $i > 3$
    - FOR $j = 2, ..., i-2$ DO
      - $v_{i,2j} = h(v_{i-1,2j-2}, v_{i,2j-1}, \theta_{j,i-j})$
      - $v_{i,2j+1} = h(v_{i,2j-1}, v_{i-1,2j-2}, \theta_{j,i-j})$
    - END FOR
  - END IF
  - $v_{i,2i-2} = h(v_{i-1,2i-4}, v_{i,2i-3}, \theta_{i-1,1})$
- END FOR

Note that in the algorithm the D-vine is conditioned on the most left variables. Hence, to sample from the first node, given all other nodes, we have to revert the order and set the size of the conditioning set to $d-1$.

**Example 16.** (Sampling on a four-dimensional D-vine) Given a D-vine on four variable $U_1$, $U_2$, $U_3$ and $U_4$ in that order, and the corresponding pair copulas and parameter matrix $\Theta$. Assume that we want to draw a sample from $U_1$ given $U_2 = u_2^*$, $U_3 = u_3^*$ and $U_4 = u_4^*$. To follow the Algorithm from Theorem 16 we first need to revert the order, i.e., $U_1 \to U_4$, $U_2 \to U_3$, $U_3 \to U_2$ and $U_4 \to U_1$. With that notation we derive $N_{cond} = 3$, $U_1, U_2, U_3 = u_1^{cond}, u_2^{cond}, u_3^{cond} = u_4^*, u_3^*, u_2^*$ and $N_{cond} + 1 = d = 4$. Now, we follow the steps of the algorithm introduced in Theorem 16.

- Sample $w_4$ independent from $[0, 1]$
- $w_1 = u_4^*$, $w_2 = u_3^*$ and $w_3 = u_2^*$
- $u_1 = v_{1,1} = w_1$
- $u_2 = v_{2,1} = w_2$
- $v_{2,2} = h(v_{1,1}, v_{2,1}, \theta_{1,1})$
- $i = 3$:

  - $v_{3,1} = w_3$
  - $u_3 = v_{3,1}$
  - $v_{3,2} = h(v_{2,1}, v_{3,1}, \theta_{1,2})$
  - $v_{3,3} = h(v_{3,1}, v_{2,1}, \theta_{1,2})$
  - $v_{3,4} = h(v_{2,2}, v_{3,3}, \theta_{2,1})$

- $i = 4$:
  - $v_{4,1} = w_4$
    - $v_{4,1} = h^{-1}(v_{4,1}, v_{3,4}, \theta_{3,1})$
    - $v_{4,1} = h^{-1}(v_{4,1}, v_{3,2}, \theta_{2,2})$
  - $v_{4,1} = h^{-1}(v_{4,1}, v_{3,1}, \theta_{1,3})$
  - $u_4 = v_{4,1}$

- $i = d \Rightarrow$ STOP

Hence, the output of the algorithm is $(u_1, u_2, u_3, u_4)$. Reverting the order to agree with the original one. Then $U_1 = u_4$, $U_2 = u_3 = u_2^*$, $U_3 = u_2 = u_3^*$ and $U_4 = u_1 = u_4^*$ is the final sample.

# Chapter 3

# Data Exploration of the Sachs Dataset

In the following we will examine the "Sachs Protein Data" data set. The data set, in the following abbreviated as Sachs data set, was first analyzed in "Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data" by Sachs et al. (2005). It consists of logarithmized levels of 11 phosphoproteins and phospholipids in individual cells after the cells have been perturbed. The researchers then used these measurements to reconstruct the classic signaling network using Bayesian Network modeling.

While the original data set includes the results of 14 experiments, we will work with the implementation from the sparsebn package from Aragam et al. (2019) which only takes 9 of the 14 experiments into account. These 9 experiments are namely: $b2camp$, $cd3cd28$, $cd3cd28 + aktinhib$, $cd3cd28 + g007$, $cd3cd28 + ly$, $cd3cd28 + psitect$, $cd3cd28 + u0126$, $cd3cd28icam2$ and $pma$. In the following we consider the different phosphoproteins and phospholipids as nodes, representing a variable, and each level as a sample.

To later transform the data to the copula scale the corresponding marginal distribution for each variable is needed, therefore choosing a suitable marginal distribution is crucial. First though, we clean up the data set. In Table 3.1 we can observe that for the node $pkc$ almost ten percent of the samples are zero. Hence, independent of the marginal distribution that we choose, there will always be a peak in the histogram after applying the PIT. Hence, we will not be able to achieve a uniform distribution which is needed to correctly fit copulas to the data. Thus, we first delete all 1305 rows in the data where at least one variable equals zero. This now leaves us with a smaller data set with only 6161 samples.

| raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 115 | 122 | 134 | 79 | 164 | 3 | 16 | 698 | 165 | 159 |

Table 3.1: Quantity of samples that equal zero for each node

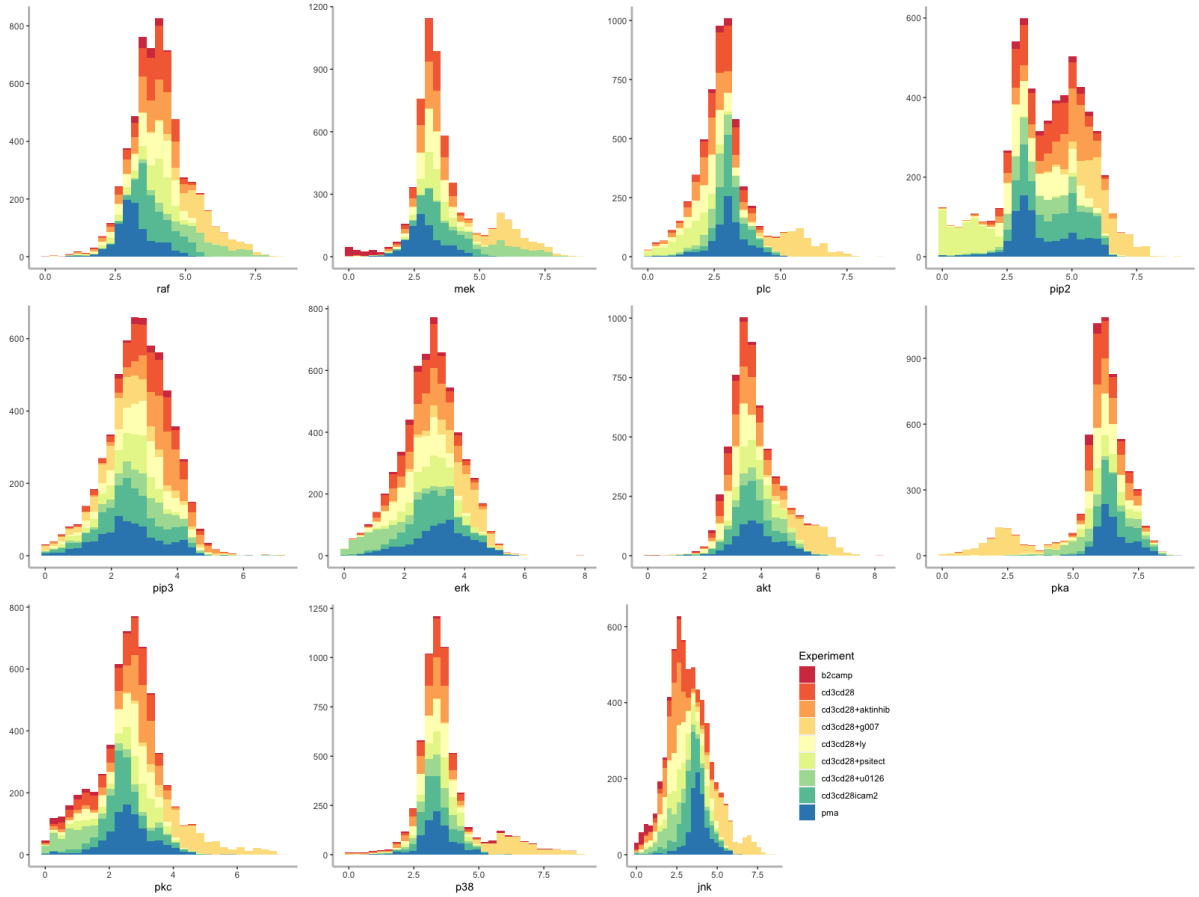



Figure 3.1: Marginal histograms of each of the eleven nodes colored according to the respective experiment the data is from

Table 3.2 shows us how the formerly described clean-up affected each of the experiments. We observe that as most of the samples from the $b2camp$ data set included the value zero such that it now only contains 155 data points. Apart from that, also the $cd3cd28 + u0126$ experiment and the $cd3cd28 + psitect$ experiment contained many samples including zero values and therefore shrunk notable.

| Exp. | b2camp | cd3cd28 | cd3cd28+akt. | cd3cd28+g. | cd3cd28+ly | cd3cd28+psi. | cd3cd28+u. | cd3cd28icam2 | pma |
|---|---|---|---|---|---|---|---|---|---|
| Before | 707 | 853 | 911 | 723 | 848 | 810 | 799 | 902 | 913 |
| After | 155 | 817 | 845 | 695 | 801 | 604 | 508 | 863 | 873 |

Table 3.2: Number of samples in each of the nine experiments before and after the samples containing at least one zero value have been deleted

Figure 3.1 shows the histograms of this new data set before transforming them to the $[0, 1]^{11}$-hypercube. Additionally the data is colored according to the respective experiment. In the following we will first assume that the data in the Sachs data set is independent and identically distributed throughout the whole data set, even though the colored histograms

indicate that this might not hold. We observe notable differences throughout the different experiments. The biggest differences in the histograms can be seen for the $cd3cd28 + g007$ experiment. Its data points lie mostly in the tails of the distributions, for example for the nodes $pka$ and $p38$.

Later we will repeat our analysis on some of the selected experiments only assuming that the data within these experiments is independent and identically distributed. In addition to that Figure 3.1 indicates that all nodes other than $pip3$ and $erk$ follow a non-normal marginal distribution.

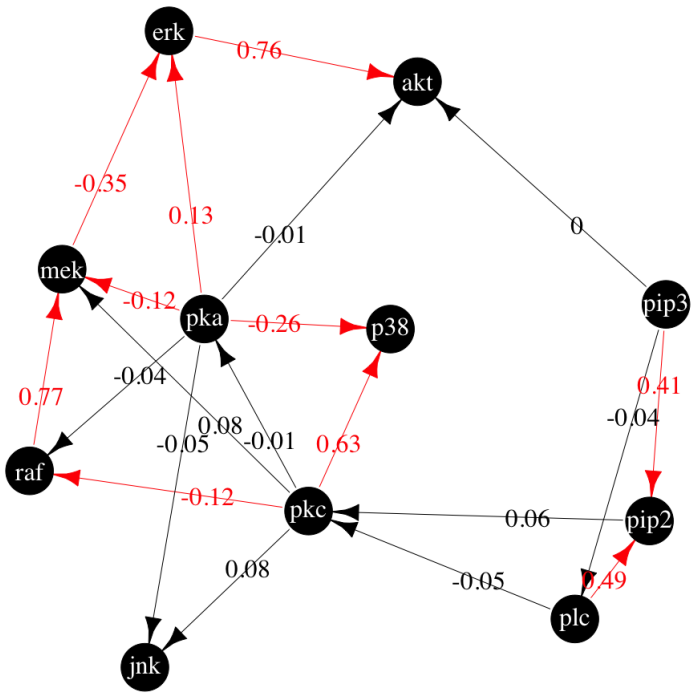



Figure 3.2: Illustration of the consent DAG with partial correlations, conditioned on all other nodes, as edge weights. Edges are colored in red if the absolute value of the empirical partial correlations is bigger than 0.1

To be able to fit both types of our models, the D-vine copula approach and the Linear Gaussian Bayesian Network, we need to make assumptions on which variables are dependent and model these relationships as a graph. Instead of calculating this on our own we instead use the already known dependence graph from Aragam et al. (2019). It is a directed acyclic graph on 11 nodes with 20 edges. We refer to it in the following as the consent DAG or only DAG. Figure 3.2 gives an illustration of the structure and Table 3.3 a topological order.

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | $X_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| pip3 | plc | pip2 | pkc | pka | p38 | jnk | raf | mek | erk | akt |

Table 3.3: Topological order of the consent graph

As we will model each node dependent on the set of its parents it would be good to have high dependencies between nodes that are connected via an edge and even better to have low dependencies between nodes that are not connected. Looking at the values of Kendall's $\tau$ for all pairs of nodes in the upper triangle in Figure 5.2, to be found in Chapter 5.1, we can see that the highest dependencies exist between the variables: $raf \leftrightarrow mek$, $erk \leftrightarrow akt$, $pkc \leftrightarrow p38$, $plc \leftrightarrow pip2$ and $p38 \leftrightarrow jnk$. Of these only the nodes $p38$ and $jnk$ with a value of Kendall's Tau of 0.31 miss an edge in the graph. This dependence might instead also be related to the fact that both have the same set of parents and have at least a high dependence to one of them, namely the to the node $pkc$.

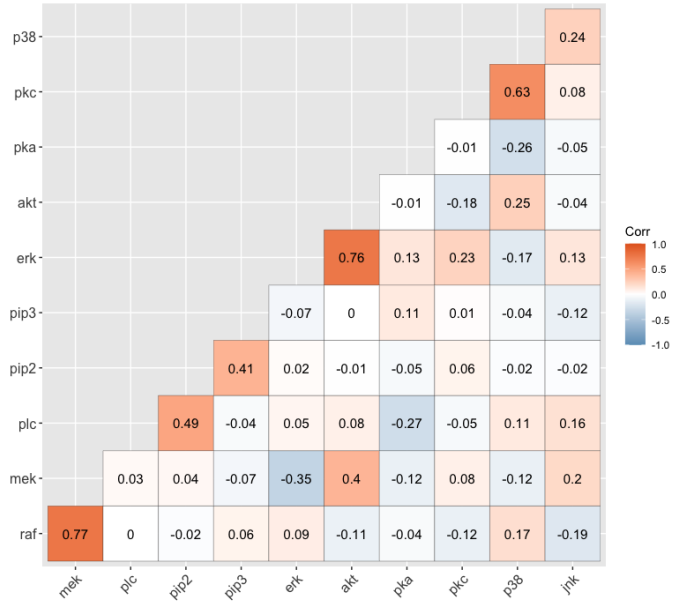



Figure 3.3: Partial correlations, conditioned on all other nodes, between each set of nodes in the Sachs data set

Further, we examine the partial correlations between the variables, i.e., how variable are correlated given all other nodes, displayed in Figure 3.3. We observe that the five pairs of nodes with the highest partial correlation each have an edge between them in the consent DAG. In total ten of twenty node pairs that are connected in the consent graph by an edge have a partial correlation bigger than 0.1. What also needs to be mentioned is the edge $pip3 \rightarrow akt$ with a partial correlation of zero indicating independence.

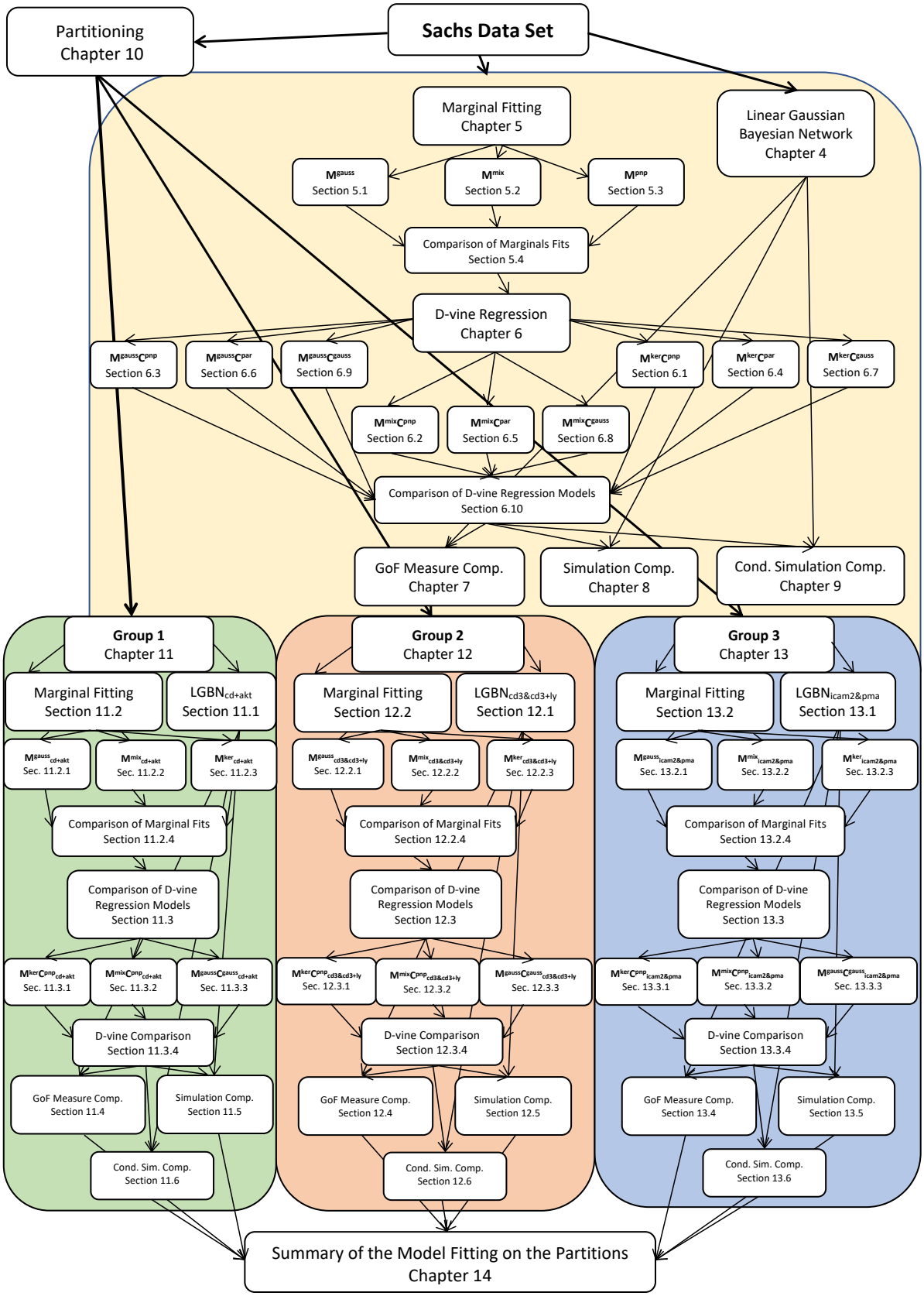



Figure 3.4: Flowchart of the upcoming chapters. Colors indicate the different underlying data sets

Now, we can start to fit our models. As already mentioned we will first assume an identical distribution throughout the Sachs data set. There, we will fit a Linear Gaussian Bayesian Network as described in Section 2.4.3. After that, we will fit various D-vine regression models where we vary the chosen marginals and copulas. To shorten the notation we denote the models as **MC** models where the super scripted text of **M** denotes the chosen marginals and the super scripted text of **C** denotes the chosen copulas with the abbreviations following in the brackets. We will order the models such that in each step we restrict the copulas or margins even more to see how the results worsen. For this, we start by allowing for parametric and non-parametric copulas (**pnp**) then only for parametric (**par**) copulas and finally for only Gaussian (**gauss**) copulas. Each time using the three different types of margins which we will fit in Chapter 5 descending in the goodness of fit, i.e., kernel density margins (**ker**), Gaussian mixture margins (**mix**) and Gaussian margins (**gauss**).

After this we check if it is possible to fit models to the data of certain experiments such that the assumption of identical distributed data is less violated. The crucial part here is to end up with big enough data sets, such that is reasonable to fit copula based models, while it is still justifiable to assume the data is identical distributed. The notations in this cases stay the same only a subscript is added to denote the part of the data on which the models are fitted. Figure 3.4 shows a flowchart of the upcoming model fitting and comparing.

# Chapter 4

# Linear Gaussian Bayesian Network Fitting

The first model we fit to the Sachs data set is a Linear Gaussian Bayesian Network. Given the graph in Figure 3.2 we can decompose its joint density, on the original scale, using Equation (2.11)

$$\begin{aligned} f(Sachs) &= \prod_{node\ in\ Sachs} f(node|\pi(node)) \\ &= f(akt|erk, pka, pip3) * f(erk|mek, pka) * f(mek|raf, pka, pkc) \\ &\quad * f(raf|pka, pkc) * f(jnk|pka, pkc) * f(p38|pka, pkc) * f(pka|pkc) \\ &\quad * f(pkc|pip2, plc) * f(pip2|plc, pip3) * f(plc|pip3) * f(pip3) \end{aligned} \tag{4.1}$$

Following Definition 38 we model every conditional density as a linear function of its parents with additional Gaussian noise.

To fit the Linear Gaussian Bayesian Network we use the *glm* function from the stats package implemented by the R Core Team (2020). Note that in the *glm* function the log-likelihood and not the $\text{AIC}_F$ is maximized. Therefore, we can assume that in the Linear Gaussian Bayesian Network every possible dependency is modeled, even if the influence on the fit of the model is marginal. As every modeled dependency has the same number of parameters, the only influence on the amount of parameters is if a dependency is modeled or not. Since each added dependency results in only one added parameter, the absolute amount of parameters stays low for every conditional density and is limited by five, as the maximum amount of parents of a node is four. Therefore, even if we later compare the models based on the $\text{AIC}_F$ the influence of the misspecification of the LGBN due to the maximization of the log-likelihood when fitting is vanishing.

| Node | Global log-likelihood | # Parameters | $AIC_F$ | $BIC_F$ |
|---|---|---|---|---|
| raf | -8373.91 | 4 | 16755.81 | 16782.72 |
| mek | -6521.22 | 5 | 13052.44 | 13086.07 |
| plc | -10160.95 | 3 | 20327.90 | 20348.08 |
| pip2 | -9618.88 | 4 | 19245.76 | 19272.66 |
| pip3 | -8513.61 | 2 | 17031.22 | 17044.67 |
| erk | -8743.38 | 4 | 17494.77 | 17521.67 |
| akt | -5811.81 | 5 | 11633.62 | 11667.25 |
| pka | -10459.41 | 3 | 20924.83 | 20951.73 |
| pkc | -9009.44 | 4 | 18026.88 | 18053.79 |
| p38 | -6306.69 | 4 | 12621.37 | 12648.27 |
| jnk | -9362.29 | 4 | 18732.58 | 18759.48 |
| $\sum$: | -92881.59 | 42 | 185847.18 | 186136.39 |

Table 4.1: Global log-likelihood, number of parameters, $AIC_F$ and $BIC_F$ of the fitted linear models for each node given the set of its parents in the Linear Gaussian Bayesian Network

| Node | $\hat{\beta}_0$ | $\hat{\boldsymbol{\beta}}$ | $\hat{\sigma}$ |
|---|---|---|---|
| raf | 6.31 | (-0.36, -0.03) | 0.89 |
| mek | 0.53 | (0.98, -0.16, 0.05) | 0.49 |
| plc | 2.92 | 0.02 | 1.59 |
| pip2 | 0.09 | (0.80, 0.53) | 1.33 |
| pip3 | 2.84 | | 0.93 |
| erk | 3.28 | (-0.02, -0.06) | 1.00 |
| akt | 3.28 | (0.67, -0.22, -0.01) | 0.39 |
| pka | 7.40 | -0.59 | 1.75 |
| pkc | 1.48 | (0.01, 0.38) | 1.09 |
| p38 | 3.84 | (-0.31, 0.62) | 0.45 |
| jnk | 3.62 | (-0.28, 0.47) | 1.22 |

Table 4.2: Estimated parameters of the conditional normal distributions $N(\beta_0 + \boldsymbol{\beta}^\top x; \sigma^2)$ of each node on the set of its parents modeled in the Linear Gaussian Bayesian Network

Table 4.1 shows the resulting log-likelihood, $\text{AIC}_F$ and $\text{BIC}_F$ of each conditional distribution. Overall, fitting a Linear Gaussian Network results in a log-likelihood of **-92881.59**, an $\text{AIC}_F$ of **185847.18** and a $\text{BIC}_F$ of **186136.39** of the whole model. As the node *pip*3 does not have any parents we here fit a univariate normal distribution to fulfill the conditions of the model.

Following Definition 37 in the Linear Gaussian Bayesian Network each conditional distribution is a conditional normal distribution of the form $N(\beta_0 + \boldsymbol{\beta}^\top x; \sigma^2)$. The estimated parameters $\hat{\beta}_0$, $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ can be seen in Table 4.2 where the order of the covariates is the same as in Equation (4.1).

# Chapter 5

# Marginal Fitting

In D-vine regression based networks the chosen marginals not only appear in the modeled conditional densities but also the copulas are modeled on the with the marginals transformed data. Hence, choosing suitable marginals is a crucial part here.

In the following we will vary between three different approaches to estimate the margins. First, we estimate Gaussian margins $\mathbf{M^{gauss}}$ using the fitdistrplus package from Delignette-Muller and Dutang (2015). In a second approach, to be able to model more complex marginal densities, we fit Gaussian mixtures margins $\mathbf{M^{mix}}$ instead using the R-packages mixtools from Benaglia et al. (2009) and nor1mix from Maechler (2019). Finally, we use a non-parametric approach and fit kernel density estimates $\mathbf{M^{ker}}$ as implemented in kde1d from Nagler and Vatter (2019).

The measure that we optimize here is the $\text{AIC}_M$. Taking the $\text{BIC}_M$ instead would not result in significantly different results as the number of parameters is fixed when fitting Gaussian margins or Gaussian mixture margins and only varies fractionally when using kernel density margins on this data set.

To see if the chosen marginals are correct we use a graphical analysis on three types of plots. We consider a density plot to see if the density matches the observed histograms, a Q-Q plot to see if the quantiles of the fitted distribution match the empirical ones and finally we plot the histogram after applying the probability integral transform. If the fitted marginal is suitable, the data should then be approximately uniformly distributed.

## 5.1 Gaussian Margin Fitting ($M^{gauss}$)

Overall, assuming independence, fitting Gaussian margins results in a log-likelihood of the marginals of **-109358.56**, an $AIC_M$ of **218761.13** and a $BIC_M$ of **218909.09**. The detailed results for each node can be found in Table 5.1.

As we have seen in Figure 3.1 the marginal distributions of the nodes, apart from *pip*3 and *erk*, do not seem like they follow a Gaussian distribution. Looking at the graphical analysis of the fitted Gaussian margins in Figure 5.1 it seems that our assumption for these two nodes is correct. Especially the data of these two nodes after applying the PIT, using the fitted Gaussian margins, seems to be uniformly distributed. This does not hold for any other node. Therefore technically the requirements to fit a D-vine copula model on this data set in combination with Gaussian margins are not fulfilled.

We observe that in the Linear Gaussian Bayesian Network the node *pip*3 is modeled as an unconditional Gaussian distribution. It therefore coincides with the Gaussian margin we fitted here.

| Node | Marginal log-likelihood | $AIC_M$ | $BIC_M$ |
|---|---|---|---|
| raf | -9204.76 | 18413.51 | 18426.97 |
| mek | -10789.50 | 21583.01 | 21596.46 |
| plc | -10161.41 | 20326.82 | 20340.27 |
| pip2 | -11724.33 | 23452.65 | 23466.10 |
| pip3 | -8513.61 | 17031.22 | 17044.67 |
| erk | -8765.56 | 17535.13 | 17548.58 |
| akt | -8710.61 | 17425.22 | 17438.67 |
| pka | -11182.07 | 22368.13 | 22381.58 |
| pkc | -9633.31 | 19270.62 | 19284.07 |
| p38 | -9960.84 | 19925.69 | 19939.14 |
| jnk | -10712.56 | 21429.13 | 21442.58 |
| $\sum$: | -109358.56 | 218761.13 | 218909.09 |

Table 5.1: Marginal log-likelihood, $AIC_M$ and $BIC_M$ of the fitted Gaussian margins

We then applied the PIT using the chosen marginals in order to receive the data on the copula scale which is needed to fit the copulas. Figure 5.2 shows the normalized pairwise contour and scatter plots as well as again the histograms of the data on the copula scale.

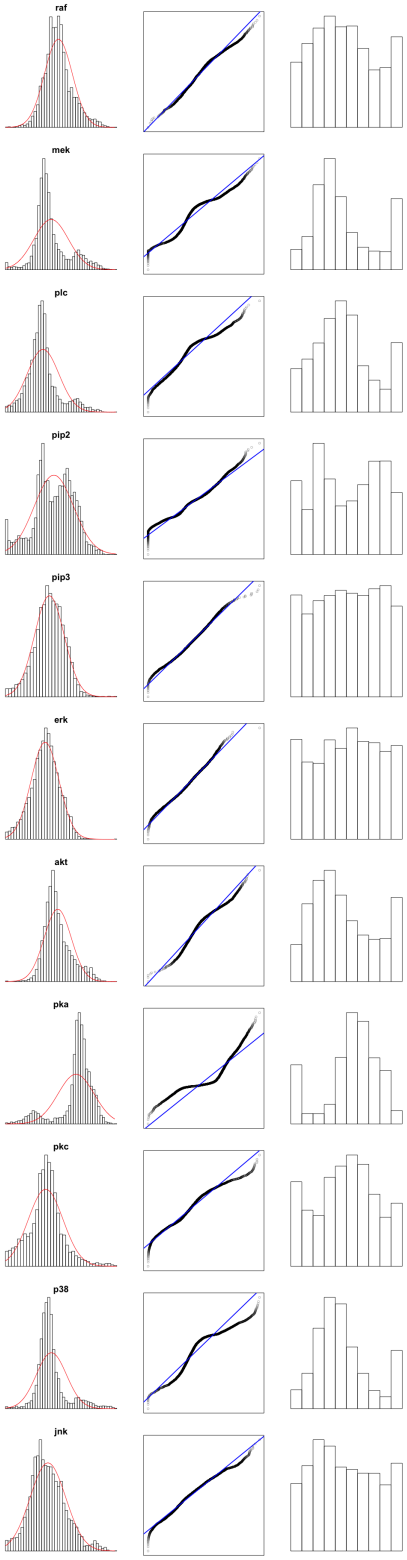



Figure 5.1: Density plot and Q-Q plot of the fitted Gaussian margins (in the first two columns) and histogram of the data after applying the distribution function of the fitted Gaussian margins for all nodes in the third column

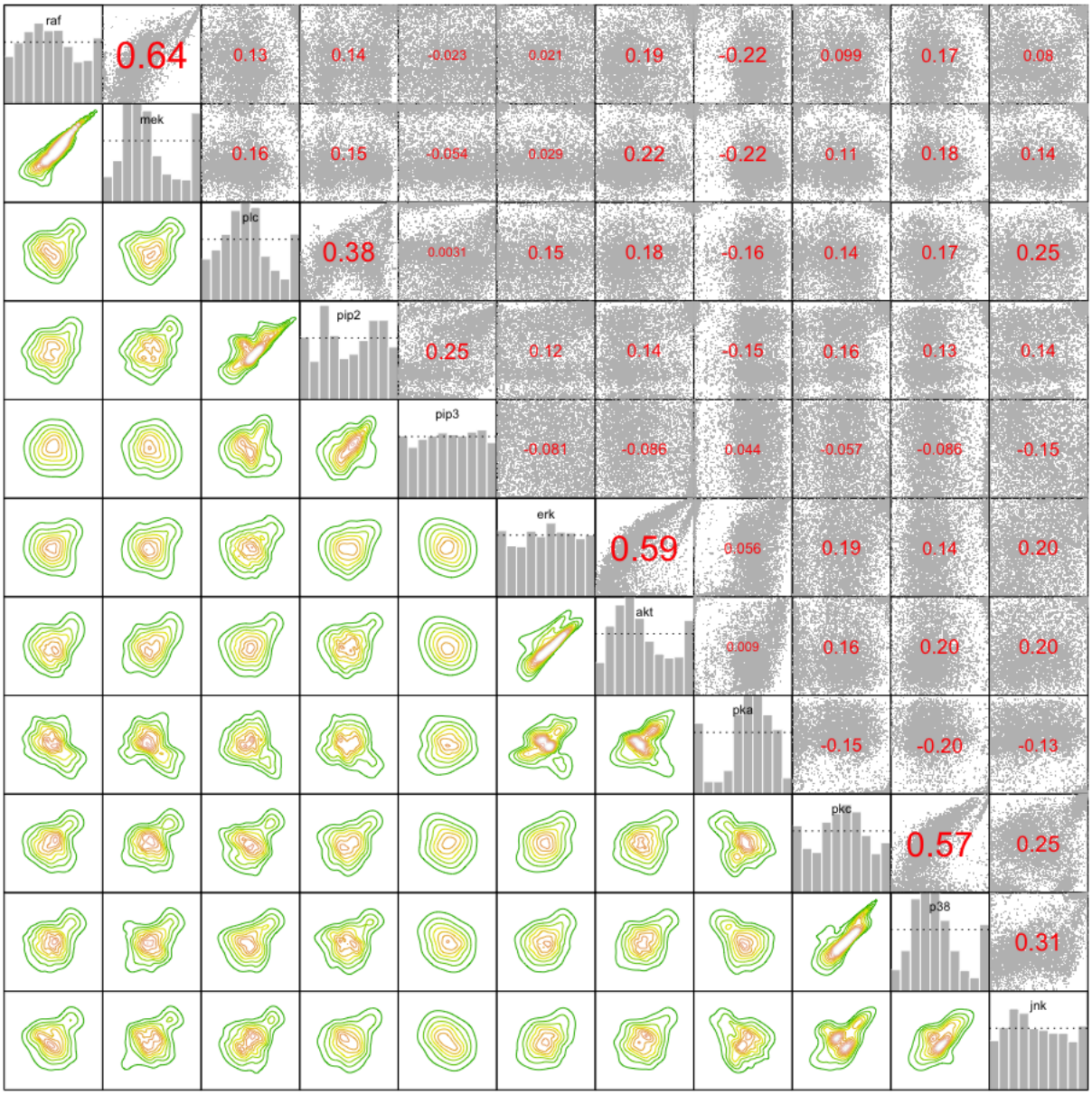



Figure 5.2: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian margins

## 5.2 Gaussian Mixture Margin Fitting ($M^{mix}$)

We have already found suitable Gaussian margins for the nodes *pip*3 and *erk*, which are a special case of Gaussian mixture margins, i.e., with one component, in Section 5.1. Therefore, we also use them in this section. To quantify the number of components of the Gaussian mixture margins for the remaining nodes we again look at Figure 3.1.

- $raf$: While $raf$ almost seems to follow a normal distribution, looking closer we can see that it is positively skewed, indicating it does not follow a simple normal distribution but a mixture of two normal distributions
- *mek*: This node has two modes, one at around 3 and one at approximately 6. Additionally, it seems it has a small peak close to zero. As the modes do not appear to be symmetric an additional component is needed for the model. Hence, we assume it follows a mixture of four normal distributions
- *plc*: Similar to the node *mek* we can see two modes at 3 and 6. As the first one seems to be negatively skewed, we add one more component to model the distribution correctly and end up with a mixture of three normal distributions
- *pip*2: As in the two previous nodes we can see two peaks at 3 and 6. In addition, we need one more component to model the small peak at zero. Hence, we again assume a mixture of three normal distributions
- *akt*: For this node we only see one mode at 3 which is positively skewed indicating the need of a second component. Therefore, we assume it follows a mixture of two normal distributions
- *pka*: We can see two modes here, at 2 and at 6. As the second one is positively skewed we need a third component to model the distribution correctly. Hence, we assume it follows a mixture of three normal distributions
- *pkc*: Even though *pkc* has only one peak at 3 a simple normal distribution will not suffice. Not only that is not symmetrical, it also has high values around zero. Therefore, we assume a mixture of three normal distributions to model the node
- *p*38: As in some of the previous nodes we see two modes at 3 and 6, both being non-symmetric. Hence, we assume a normal mixture distribution with four components
- $jnk$: Here, we can see one clear peak at 3, but as the distribution is not symmetric we need a second component. Additionally, we observe a second mode at around 7. Hence, we assume a mixture of three normal distributions

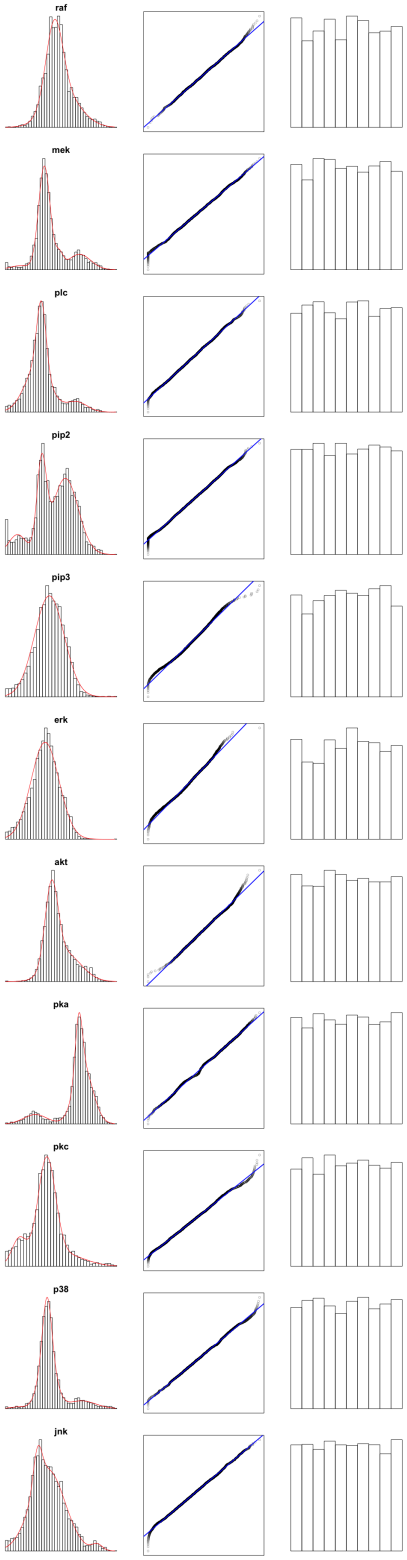



Figure 5.3: Density plot and Q-Q plot of the fitted Gaussian mixture margins (in the first two columns) and histogram of the data after applying the distribution function of the fitted Gaussian mixture margins for all nodes in the third column

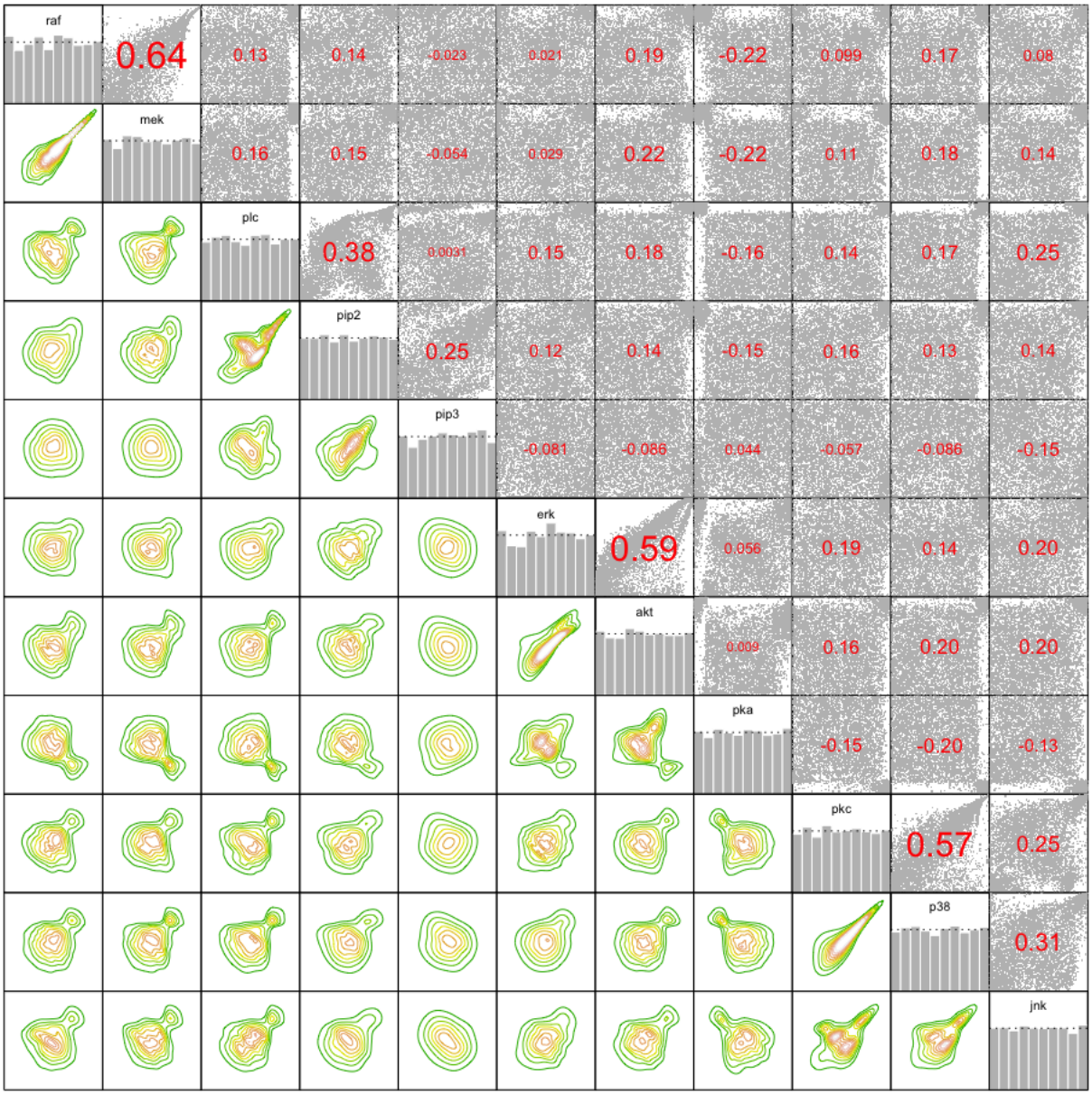



Figure 5.4: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian mixture margins

To see if our assumptions are correct we again use a graphical analysis. Looking at the results in Figure 5.3 we can see that all the proposed distributions fit quite good, where only for very small or big values the Q-Q plots show small inaccuracies. Overall, it seems the estimated Gaussian mixture margins fit well, among others, as the data after applying the PIT with Gaussian mixture margins seems to be approximately uniformly distributed. It therefore fulfills the necessary condition to fit copulas to it. The resulting normalized pairwise contour and scatter plots as well as again the histograms of the data on the copula scale can be seen in Figure 5.4.

Assuming independence in all nodes, this results in an overall log-likelihood of the marginals of **-101244.90**, an $AIC_M$ of **202641.80** and a $BIC_M$ of **203153.00**.

| Node | Distribution | Marginal log-lik. | # Parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | two component normal | -8972.36 | 5 | 17954.72 | 17988.35 |
| mek | four component normal | -9361.27 | 11 | 18744.54 | 18818.52 |
| plc | three component normal | -9277.45 | 8 | 18570.89 | 18624.70 |
| pip2 | three component normal | -11332.45 | 8 | 22680.90 | 22734.71 |
| pip3 | normal | -8513.61 | 2 | 17031.22 | 17044.67 |
| erk | normal | -8765.56 | 2 | 17535.13 | 17548.58 |
| akt | two component normal | -8152.43 | 5 | 16314.87 | 16348.50 |
| pka | three component normal | -8909.55 | 8 | 17835.10 | 17888.91 |
| pkc | three component normal | -9196.36 | 8 | 18408.73 | 18462.53 |
| p38 | four component normal | -8224.14 | 11 | 16470.29 | 16544.27 |
| jnk | three component normal | -10539.72 | 8 | 21095.43 | 21149.24 |
| $\sum$: | | -101244.90 | 76 | 202641.80 | 203153.00 |

Table 5.2: Marginal log-likelihood, number of parameters, $AIC_M$ and $BIC_M$ of the fitted Gaussian mixture margins

## 5.3 Kernel Density Margin Fitting ($M^{ker}$)

As a third approach we fitted kernel density estimates for the margins. Looking at the results in Figure 5.5 we can see almost the same results compared to the use of Gaussian mixture margins in Figure 5.3, only the Q-Q plots seem to fit a little bit better. Therefore, we also accept the proposed kernel density margins.

Assuming independence for all nodes this leads to an overall log-likelihood of the marginals of **-100636.10**, an $AIC_M$ of **201575.00** and a $BIC_M$ of **202593.20**. The detailed results can be found in Table 5.3.

| Node | Distribution | Marginal log-likelihood | Effective # parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | kernel density | -8943.97 | 14.87 | 17917.68 | 18017.66 |
| mek | kernel density | -9286.63 | 19.65 | 18612.56 | 18744.74 |
| plc | kernel density | -9229.39 | 14.61 | 18487.99 | 18586.23 |
| pip2 | kernel density | -11247.11 | 10.30 | 22514.81 | 22584.06 |
| pip3 | kernel density | -8416.39 | 10.87 | 16854.53 | 16927.66 |
| erk | kernel density | -8740.15 | 4.11 | 17488.52 | 17516.18 |
| akt | kernel density | -8110.79 | 18.70 | 16258.99 | 16384.80 |
| pka | kernel density | -8850.50 | 20.44 | 17741.87 | 17879.34 |
| pkc | kernel density | -9129.07 | 11.42 | 18280.98 | 18357.79 |
| p38 | kernel density | -8179.29 | 16.96 | 16392.50 | 16506.58 |
| jnk | kernel density | -10502.81 | 9.46 | 21024.52 | 21088.12 |
| $\sum$: | | -100636.10 | 151.39 | 201575.00 | 202593.20 |

Table 5.3: Marginal log-likelihood, number of effective parameters, $AIC_M$ and $BIC_M$ of the fitted kernel density margins

Now, using these margins to transform our data to the copula scale, as displayed in Figure 5.6, we can see almost the same pattern as in the approach with Gaussian mixture margins.

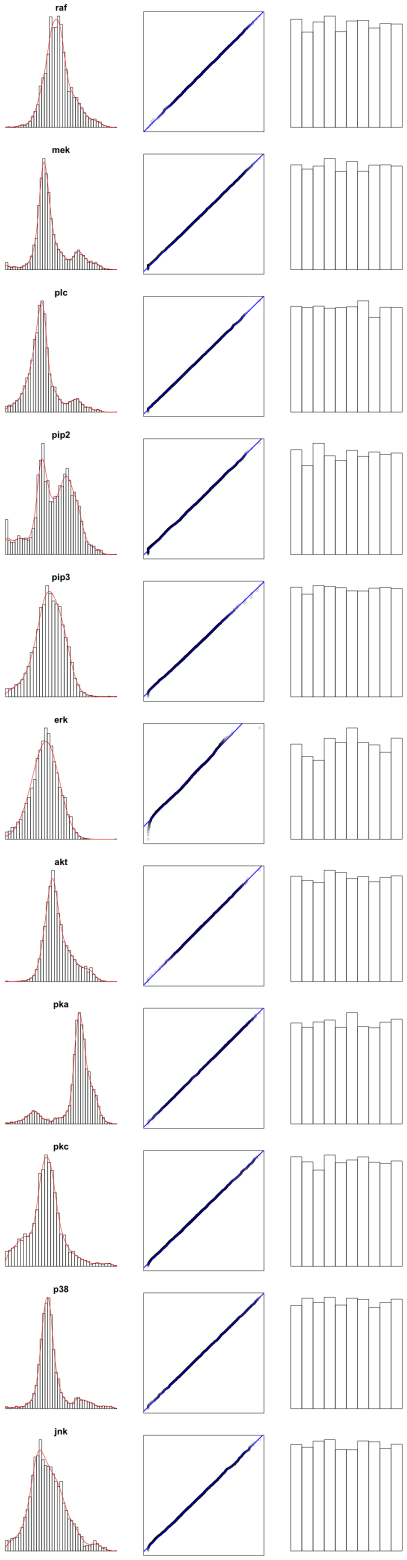



Figure 5.5: Density plot and Q-Q plot of the fitted kernel density margins (in the first two columns) and histogram of the data after applying the distribution function of the fitted kernel density margins for all nodes in the third column

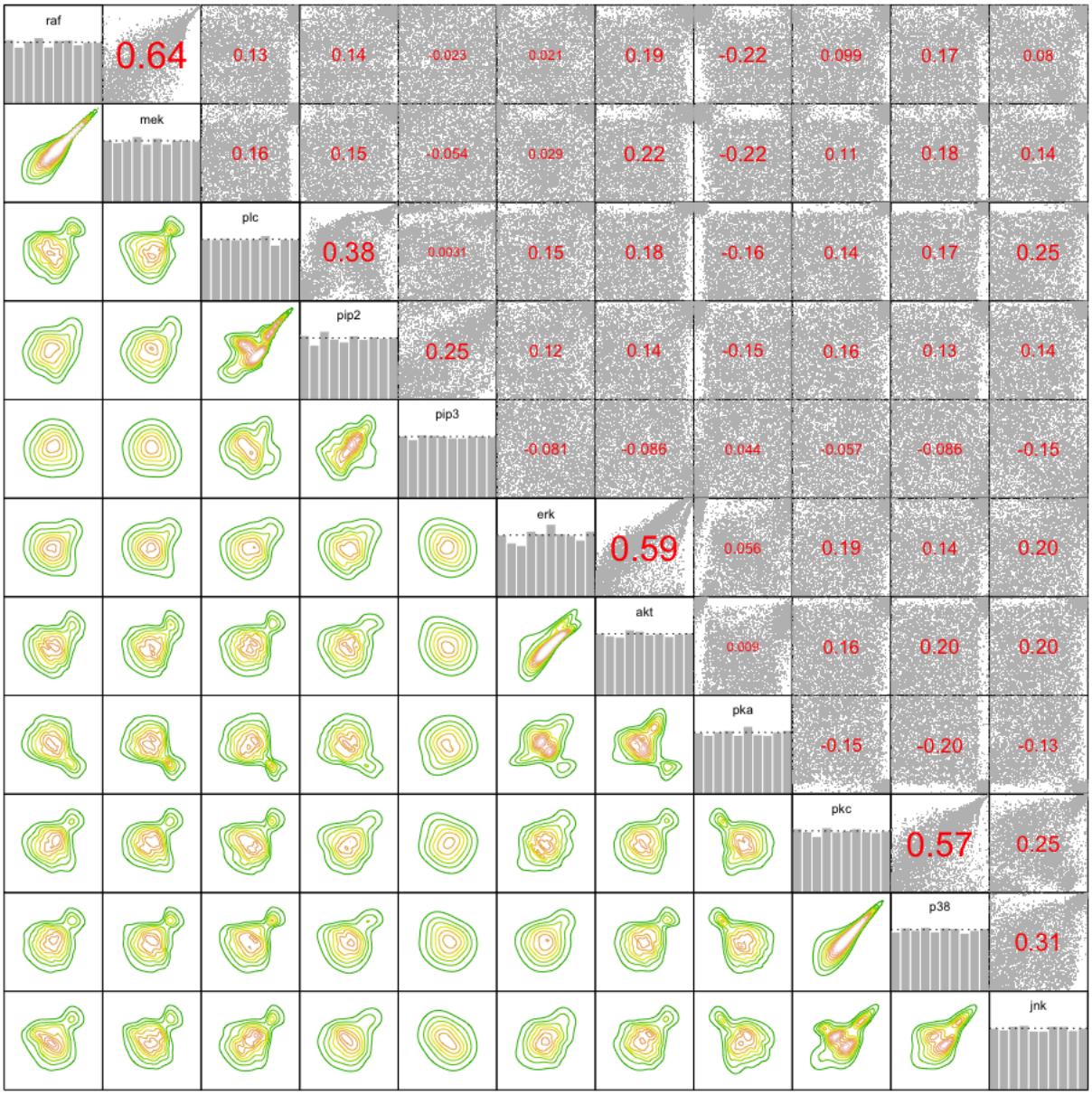



Figure 5.6: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted kernel density margins

## 5.4 Comparison of Marginal Fits

Comparing the overall log-likelihood of the three different approaches to find the best suiting margins, we observe that using Gaussian margins results in a significantly worse fit than the other two. Further, looking at the log-likelihood of the Gaussian mixture margins and kernel density margins, we observe that the second one is only a little bit better. Taking the absolute value of the log-likelihood into account, this advantage is almost negligible. Considering the log-likelihood of the individual nodes in these two approaches, we can see that this holds for all of them, i.e., their marginals have an almost equal log-likelihood when choosing Gaussian mixture margins or kernel density margins.

Comparing the $\text{AIC}_M$, instead of the log-likelihood, does not change the argument. This holds due to the size of the log-likelihood and the fact that in both models the marginals have, at least speaking of the order of magnitude, a similar number of parameters. In total, in the kernel density margin approach twice as many effective parameters are used as in the Gaussian mixture margin approach. The same holds when comparing the $\text{BIC}_M$. Here, the parameters are only stressed approximately four times as strong as in the $\text{AIC}_M$. Again due to the similar number of (effective) parameters there are no notable changes in the outcome.

Comparing the by-node log-likelihood of these two similar approaches with the Gaussian margin approach, we observe a strong worsening in the Gaussian approach for all nodes except *pip*3, *erk* and *jnk*. For the first two nodes, the result is not surprising as *pip*3 and *erk* are also modeled as a Gaussian distribution in the Gaussian mixture margin approach. For the node *jnk* we have already seen in its histogram in Figure 3.1 that it almost follows a normal distribution if we ignore the small peak in the right tail. Hence, it does seem reasonable that the difference for this node is rather small when comparing the three approaches. Again, if we switch to any other goodness of fit measure, either $\text{AIC}_M$ or $\text{BIC}_M$, the arguments will not change. As before, this is due to the similar amount of (effective) parameters, the size of the data set and in this case the notable differences in the log-likelihood.

To highlight even more how similar the chosen Gaussian mixture margins and the kernel density margins are we look at the scatter and contour plots after applying the PIT in Figure 5.4 and Figure 5.6. In both cases we observe very similar patterns in the scatter plots and shapes of the contour plots up to a level where it is hard to see any differences. This is not the case when comparing them to the same plots for the fitted Gaussian margins for the PIT in Figure 5.2. Especially in some of the scatter plots extremely different patterns can be observed. Nevertheless, we can see in the scatter plots of the three chosen margins that tail dependencies are present between almost all pairs of nodes.

In this context node *pka* is to be mentioned particularly. Small values for this node seem to come in line with only very high values of all other nodes except *pip*3. While this behavior is present when using Gaussian margins, in Figure 5.2, it can be observed more clearly when using Gaussian mixture margins or kernel density margins in Figure 5.4 and Figure 5.6.

Overall, the pseudo copula data when using Gaussian mixture margins or kernel density margins for the PIT seems to be very similar. This does not hold when instead using Gaussian margins for the probability integral transform. Therefore, we expect similar results when fitting D-vine models with either Gaussian mixture margins or kernel density margins. On the other hand using Gaussian margins might result in different copulas. If the fit of the copulas is not superior in these models the fitted conditional densities will automatically be worse than in the models with Gaussian mixture margins or kernel density margins. This is due to the significantly worse fit of the Gaussian margins.

# Chapter 6

# D-vine Regression

To fit the D-vine regressions we again decompose the distribution in the same way as for the Linear Gaussian Bayesian Network. In a next step we model the conditional densities using copulas following Equation (2.12). Therefore, each node and its parents will be modeled as a simplified D-vine.

The selected node whose conditional density is modeled will then be assigned to the first, i.e., most left, node of the D-vine. If a node has more than one parent, the order of the D-vines might change depending on the chosen marginals and copulas. In this case we use numbers to denote the position in the order of the D-vine instead of the explicit nodes. Hence, "1" denotes the response variable, i.e., the modeled node, "2" the second node in the D-Vine and so on.

Therefore, we can write the joint density as

$$
\begin{aligned}
f(Sachs) &= \prod_{node\ in\ Sachs} f(node|\pi(node)) \\
&= \prod_{node\ in\ Sachs} \prod_{\#\ of\ trees\ in\ D-Vine} c_{(first\ copula\ in\ tree)} * f(node) \\
&= c_{(akt\ 1)} * c_{(akt\ 2;\ 1)} * c_{(akt\ 3;\ 1,\ 2)} * f(akt) * c_{(erk\ 1)} * c_{(erk\ 2;\ 1)} * f(erk) \\
&\quad * c_{(mek\ 1)} * c_{(mek\ 2;\ 1)} * c_{(mek\ 3;\ 1,\ 2)} * f(mek) * c_{(raf\ 1)} * c_{(raf\ 2;\ 1)} * f(raf) \\
&\quad * c_{(jnk\ 1)} * c_{(jnk\ 2;\ 1)} * f(jnk) * c_{(p38\ 1)} * c_{(p38\ 2;\ 1)} * f(p38) \\
&\quad * c_{(pka\ pkc)} * f(pka) * c_{(pkc\ 1)} * c_{(pkc\ 2;\ 1)} * f(pkc) \\
&\quad * c_{(pip2\ 1)} * c_{(pip2\ 2;\ 1)} * f(pip2) * c_{(plc\ pip3)} * f(plc) * f(pip3)
\end{aligned}
$$

In the following, when we speak of the goodness of fit of a node this refers to the goodness of fit of its conditional density given the parents of the node.

First, we will only look at the fit of the copulas per node and ignore the margin terms in the conditional densities. We will speak here of the fit "of the copula terms". Later, we will include the different choices of the marginals such that we are able to finalize the models and compare them with the Linear Gaussian Bayesian Network modeled in Chapter 4.

As a measure to optimize we use the AIC-corrected conditional log-likelihood $\text{AIC}_C$. If instead we would take the BIC-corrected version $\text{BIC}_C$, similar to when fitting the margins, this would not significantly change the results. This is due to the fact that in the cases with only parametric or Gaussian copulas the number of parameters is fixed. Also, when we allow for parametric and non-parametric copulas the number of parameters is not that much larger that there are notable difference.

It is important to note that if one wants to use any other method than kernel density margins to transform the input data to the copula scale before fitting the D-vine models using the built-in regression function in the vinereg package changes have to be applied to the source code. Otherwise, the regression function of the package automatically applies kernel density estimates to the input data even if it is already approximately uniformly distributed in $[0, 1]^d$ and performs a second PIT.

## 6.1 Kernel Density Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{ker}C^{pnp}}$)

Transforming the data to the copulas scale using kernel density margins and then fitting the copulas using the D-vine regression allowing for parametric and non-parametric copulas results in a log-likelihood of **27699.15**, an $AIC_C$ of **-54193.44** and a $BIC_C$ of **-50141.44** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 1239.90 | -2325.20 | -1805.29 |
| mek | raf, pkc, pka | 5618.52 | -11051.44 | -10427.22 |
| plc | pip3 | 525.89 | -1001.89 | -834.06 |
| pip2 | plc, pip3 | 4341.34 | -8572.24 | -8200.81 |
| erk | pka, mek | 1558.95 | -2985.48 | -2540.14 |
| akt | erk, pka, pip3 | 5148.76 | -10157.28 | -9685.63 |
| pka | pkc | 1202.51 | -2338.71 | -2115.73 |
| pkc | plc, pip2 | 1263.94 | -2410.94 | -2017.66 |
| p38 | pkc, pka | 4634.90 | -9131.93 | -8668.29 |
| jnk | pkc, pka | 2164.43 | -4218.33 | -3846.61 |
| $\sum$: | | 27699.15 | -54193.44 | -50141.44 |

Table 6.1: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{ker}C^{pnp}}$ model

The detailed results for each node can be seen in Table 6.1. We observe that in the model all of the twenty edges in the DAG are modeled using a copula. A list of these copulas can be found in Table 6.2 where we can see that nineteen of them are modeled as a non-parametric copula and only one, i.e., the $c_{akt,pip3;erk,pka}$ copula, as a parametric one. Contour plots of the fitted copulas are illustrated in Section 6.3 in Figure 6.1 together with the ones from the $\mathbf{M^{mix}C^{pnp}}$ model and the $\mathbf{M^{gauss}C^{pnp}}$ model.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | tll | 36.97 | 1116.67 | -2159.41 | -1910.78 | -0.22 |
| | raf pkc; pka | tll | 40.33 | 123.23 | -165.79 | 105.49 | 0.10 |
| mek | mek raf | tll | 35.51 | 5373.82 | -10676.63 | -10437.81 | 0.64 |
| | mek pkc; raf | tll | 27.42 | 181.41 | -307.99 | -123.58 | 0.11 |
| | mek pka; raf pkc | tll | 29.88 | 63.29 | -66.82 | 134.17 | -0.22 |
| plc | plc pip3 | tll | 24.95 | 525.89 | -1001.89 | -834.06 | 0 |
| pip2 | pip2 plc | tll | 26.72 | 3228.61 | -6403.79 | -6224.10 | 0.38 |
| | pip2 pip3; plc | tll | 28.51 | 1112.73 | -2168.45 | -1976.71 | 0.25 |
| erk | erk pka | tll | 33.97 | 1433.57 | -2799.20 | -2570.7 | 0.06 |
| | erk mek; pka | tll | 32.24 | 125.38 | -186.28 | 30.56 | 0.03 |
| akt | akt erk | tll | 29.44 | 4376.55 | -8694.21 | -8496.17 | 0.59 |
| | akt pka; erk | tll | 39.68 | 770.75 | -1462.14 | -1195.26 | 0.01 |
| | akt pip3; erk pka | frank | 1.00 | 1.47 | -0.93 | 5.80 | -0.09 |
| pka | pka pkc | tll | 33.15 | 1202.51 | -2338.71 | -2115.73 | -0.15 |
| pkc | pkc plc | tll | 29.76 | 1228.73 | -2397.94 | -2197.78 | 0.14 |
| | pkc pip2; plc | tll | 28.71 | 35.21 | -13.00 | 180.12 | 0.16 |
| p38 | p38 pkc | tll | 36.68 | 4428.45 | -8783.52 | -8536.78 | 0.57 |
| | p38 pka; pkc | tll | 32.25 | 206.45 | -348.41 | -131.51 | -0.20 |
| jnk | jnk pkc | tll | 28.30 | 2029.11 | -4001.62 | -3811.31 | 0.25 |
| | jnk pka; pkc | tll | 26.97 | 135.33 | -216.71 | -35.30 | -0.13 |

Table 6.2: Summary of all copulas fitted in the $\mathbf{M^{ker}C^{pnp}}$ model

## 6.2 Gaussian Mixture Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{mix}C^{pnp}}$)

Instead transforming the data to the copulas scale using Gaussian mixture margins and then fitting the copulas using the D-vine regression allowing for parametric and non-parametric copulas results in a log-likelihood of **27585.26**, an $AIC_C$ of **-54046.21** and a $BIC_C$ of **-50265.02** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 1246.03 | -2356.53 | -1900.71 |
| mek | raf, pkc, pka | 5582.16 | -10987.23 | -10391.68 |
| plc | pip3 | 516.07 | -959.16 | -713.72 |
| pip2 | plc, pip3 | 4297.59 | -8484.32 | -8111.51 |
| erk | pka, mek | 1560.08 | -2992.08 | -2561.31 |
| akt | erk, pka, pip3 | 5138.75 | -10131.24 | -9639.35 |
| pka | pkc | 1199.86 | -2340.47 | -2141.19 |
| pkc | plc, pip2 | 1224.98 | -2394.34 | -2207.31 |
| p38 | pkc, pka | 4665.35 | -9201.86 | -8768.53 |
| jnk | pkc, pka | 2154.39 | -4198.98 | -3829.71 |
| $\sum$: | | 27585.26 | -54046.21 | -50265.02 |

Table 6.3: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{mix}C^{pnp}}$ model

Table 6.3 shows the results for each node. We observe that in the model again all twenty edges in the DAG are modeled using a copula. The list of copulas modeled in the $\mathbf{M^{mix}C^{pnp}}$ model can be found in Table 6.4 where we can see that eighteen of them are modeled as a non-parametric copula and two, i.e., the $c_{akt,pip3;erk,pka}$ and the $c_{pkc,pip2;plc}$ copula, as a parametric one. Contour plots of the fitted copulas can be found in Section 6.3 in Figure 6.1.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | tll | 34.14 | 1119.28 | -2170.28 | -1940.67 | -0.22 |
| | raf pkc; pka | tll | 33.63 | 126.75 | -186.25 | 39.96 | 0.10 |
| mek | mek raf | tll | 34.81 | 5348.02 | -10626.43 | -10392.33 | 0.64 |
| | mek pkc; raf | tll | 23.92 | 178.61 | -309.37 | -148.46 | 0.11 |
| | mek pka; raf pkc | tll | 29.81 | 55.53 | -51.43 | 149.11 | -0.22 |
| plc | plc pip3 | tll | 36.49 | 516.07 | -959.16 | -713.72 | 0 |
| pip2 | pip2 plc | tll | 23.92 | 3205.13 | -6362.40 | -6201.49 | 0.38 |
| | pip2 pip3; plc | tll | 31.50 | 1092.46 | -2121.92 | -1910.02 | 0.25 |
| erk | erk pka | tll | 31.86 | 1435.08 | -2806.43 | -2592.11 | 0.06 |
| | erk mek; pka | tll | 32.18 | 125.01 | -185.65 | 30.80 | 0.03 |
| akt | akt erk | tll | 32.17 | 4336.67 | -8608.99 | -8392.59 | 0.59 |
| | akt pka; erk | tll | 39.96 | 800.90 | -1521.88 | -1253.12 | 0.01 |
| | akt pip3; erk pka | frank | 1.00 | 1.19 | -0.37 | 6.36 | -0.09 |
| pka | pka pkc | tll | 29.63 | 1199.86 | -2340.47 | -2141.19 | -0.15 |
| pkc | pkc plc | tll | 26.81 | 1223.12 | -2392.63 | -2212.33 | 0.14 |
| | pkc pip2; plc | clayton | 1.00 | 1.85 | -1.71 | 5.02 | 0.16 |
| p38 | p38 pkc | tll | 34.87 | 4454.75 | -8839.77 | -8605.25 | 0.57 |
| | p38 pka; pkc | tll | 29.56 | 210.60 | -362.09 | -163.28 | -0.20 |
| jnk | jnk pkc | tll | 25.25 | 2021.56 | -3992.61 | -3822.75 | 0.25 |
| | jnk pka; pkc | tll | 29.65 | 132.83 | -206.37 | -6.96 | -0.13 |

Table 6.4: Summary of all copulas fitted in the $\mathbf{M^{mix}C^{pnp}}$ model

## 6.3 Gaussian Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{gauss}C^{pnp}}$)

As a third approach to transform the data to the copula scale we use Gaussian margins and then allow for parametric and non-parametric copulas in the D-vine regression. This results in a log-likelihood of **23993.53**, an $AIC_C$ of **-46336.37** and a $BIC_C$ of **-40785.13** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 1170.45 | -2227.48 | -1846.08 |
| mek | raf, pka, pkc | 5884.02 | -11586.52 | -10976.06 |
| plc | pip3 | 165.01 | -222.22 | 140.28 |
| pip2 | plc, pip3 | 2416.08 | -4739.43 | -4427.60 |
| erk | pka, mek | 1080.37 | -2021.97 | -1555.29 |
| akt | erk, pka | 4488.86 | -8477.06 | -6793.34 |
| pka | pkc | 1207.66 | -2269.58 | -1779.43 |
| pkc | plc, pip2 | 1356.81 | -2517.94 | -1859.86 |
| p38 | pkc, pka | 4058.33 | -8024.75 | -7715.65 |
| jnk | pkc, pka | 2165.94 | -4249.42 | -3972.10 |
| $\sum$: | | 23993.53 | -46336.37 | -40785.13 |

Table 6.5: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{gauss}C^{pnp}}$ model

To examine the fit of each node, Table 6.5 shows the results for each of them. Note that the dependency of *akt* from *pip*3 is not modeled. This time six of nineteen copulas are modeled as a parametric one. A list of all copulas is displayed in Table 6.6. Contour plots of the fitted copulas can be found in Figure 6.1.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | bb8 | 2.00 | 1013.06 | -2022.11 | -2008.66 | -0.22 |
| | raf pkc; pka | tll | 54.71 | 157.39 | -205.37 | 162.58 | 0.10 |
| mek | mek raf | bb6 | 2.00 | 4895.28 | -9786.55 | -9773.10 | 0.64 |
| | mek pka; raf | tll | 42.34 | 807.46 | -1530.24 | -1245.43 | -0.22 |
| | mek pkc; raf pka | tll | 46.42 | 181.28 | -269.73 | 42.47 | 0.11 |
| plc | plc pip3 | tll | 53.90 | 165.01 | -222.22 | 140.28 | 0 |
| pip2 | pip2 plc | t | 2.00 | 1572.98 | -3141.95 | -3128.50 | 0.38 |
| | pip2 pip3; plc | tll | 44.36 | 843.10 | -1597.48 | -1299.10 | 0.25 |
| erk | erk pka | tll | 67.38 | 941.11 | -1747.46 | -1294.23 | 0.06 |
| | erk mek; pka | bb8 | 2.00 | 139.26 | -274.51 | -261.06 | 0.03 |
| akt | akt erk | tll | 53.15 | 2982.76 | -5859.21 | -5501.71 | 0.59 |
| | akt pka; erk | tll | 197.18 | 1506.11 | -2617.85 | -1291.63 | 0.01 |
| pka | pka pkc | tll | 72.87 | 1207.66 | -2269.58 | -1779.43 | -0.15 |
| pkc | pkc plc | tll | 52.00 | 1211.87 | -2319.73 | -1969.95 | 0.14 |
| | pkc pip2; plc | tll | 45.84 | 144.94 | -198.21 | 110.09 | 0.16 |
| p38 | p38 pkc | gumbel | 1.00 | 3081.91 | -6161.82 | -6155.09 | 0.57 |
| | p38 pka; pkc | tll | 44.96 | 976.42 | -1862.93 | -1560.56 | -0.20 |
| jnk | jnk pkc | tll | 39.23 | 2005.32 | -3932.19 | -3668.32 | 0.25 |
| | jnk pka; pkc | bb7 | 2.00 | 160.61 | -317.23 | -303.78 | -0.13 |

Table 6.6: Summary of all copulas fitted in the $\mathbf{M^{gauss}C^{pnp}}$ model


(a) $\mathbf{M^{ker}C^{pnp}}$ (b) $\mathbf{M^{mix}C^{pnp}}$ (c) $\mathbf{M^{gauss}C^{pnp}}$

Figure 6.1: Contour plots of the respective copulas in the models allowing for parametric and non-parametric pair copulas where kernel density margins, Gaussian mixture margins or Gaussian margins were used for the PIT

## 6.4 Kernel Density Margins with Parametric Copulas ($\mathbf{M^{ker}C^{par}}$)

Transforming the data to the copulas scale using kernel density margins and then fitting the copulas using the D-vine regression allowing only for parametric copulas results in a log-likelihood of **19042.41**, an $AIC_C$ of **-38020.84** and a $BIC_C$ of **-37805.57** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 794.14 | -1580.28 | -1553.36 |
| mek | raf, pka, pkc | 5260.60 | -10515.20 | -10495.01 |
| plc | pip3 | 32.77 | -63.54 | -56.81 |
| pip2 | plc, pip3 | 2675.20 | -5342.39 | -5315.49 |
| erk | pka, mek | 316.10 | -626.21 | -606.03 |
| akt | erk, pka, pip3 | 3249.16 | -6486.32 | -6445.97 |
| pka | pkc | 674.83 | -1345.66 | -1332.21 |
| pkc | plc, pip2 | 698.03 | -1390.06 | -1369.87 |
| p38 | pkc, pka | 4006.29 | -8006.58 | -7986.41 |
| jnk | pkc, pka | 1335.30 | -2664.60 | -2644.41 |
| $\sum$: | | 19042.41 | -38020.84 | -37805.57 |

Table 6.7: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{ker}C^{par}}$ model

The detailed results for each node are displayed in Table 6.7. As in the previous model where kernel density margins were used all dependencies from the DAG are modeled. A list of the copulas can be found in Table 6.8. Note that not a single copula is modeled as a Gaussian copula. The contour plots of the fitted copulas can be found in Section 6.6 in Figure 6.2 together with the ones from the $\mathbf{M^{mix}C^{par}}$ model and the $\mathbf{M^{gauss}C^{par}}$ model.

| Node | Pair copula | Family | Rotation | Copula log-lik. | $AIC_C$ | $BIC_C$ | 1. Par. | 2. Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|---|---|
| raf | raf pka | bb8 | 270 | 769.87 | -1535.74 | -1522.28 | 1.78 | 0.97 | -0.22 |
| | raf pkc; pka | t | 0 | 24.27 | -44.54 | -31.08 | 0 | 10.84 | 0.10 |
| mek | mek raf | joe | 0 | 5157.49 | -10312.99 | -10306.26 | 5.00 | | 0.64 |
| | mek pka; raf | clayton | 90 | 67.32 | -132.64 | -125.91 | 0.17 | | 0.11 |
| | mek pkc; raf pka | joe | 180 | 35.79 | -69.57 | -62.84 | 1.05 | | -0.22 |
| plc | plc pip3 | clayton | 180 | 32.77 | -63.54 | -56.81 | 0.10 | | 0 |
| pip2 | pip2 plc | bb7 | 0 | 2027.09 | -4050.18 | -4036.73 | 2.14 | 0.35 | 0.38 |
| | pip2 pip3; plc | bb8 | 0 | 648.11 | -1292.21 | -1278.76 | 4.09 | 0.56 | 0.25 |
| erk | erk pka | t | 0 | 269.77 | -535.54 | -522.08 | 0.07 | 2.67 | 0.06 |
| | erk mek; pka | clayton | 270 | 46.34 | -90.67 | -83.95 | 0.11 | | 0.03 |
| akt | akt erk | bb8 | 0 | 2905.74 | -5807.48 | -5794.03 | 4.51 | 0.89 | 0.59 |
| | akt pka; erk | bb8 | 270 | 337.05 | -670.10 | -656.65 | 1.45 | 0.96 | 0.01 |
| | akt pip3; erk pka | bb8 | 90 | 6.37 | -8.74 | 4.71 | 1.16 | 0.76 | -0.09 |
| pka | pka pkc | bb8 | 90 | 674.83 | -1345.66 | -1332.21 | 1.54 | 0.99 | -0.15 |
| pkc | pkc plc | bb8 | 0 | 691.19 | -1378.38 | -1364.92 | 1.53 | 1.00 | 0.14 |
| | pkc pip2; plc | frank | 0 | 6.84 | -11.68 | -4.95 | 0.27 | | 0.16 |
| p38 | p38 pkc | joe | 0 | 3841.30 | -7680.59 | -7673.87 | 3.84 | | 0.57 |
| | p38 pka; pkc | bb8 | 270 | 164.99 | -325.99 | -312.54 | 1.40 | 0.93 | -0.20 |
| jnk | jnk pkc | bb7 | 0 | 1280.26 | -2556.53 | -2543.07 | 1.78 | 0.14 | 0.25 |
| | jnk pka; pkc | joe | 270 | 55.03 | -108.07 | -101.34 | 1.10 | | -0.13 |

Table 6.8: Summary of all copulas fitted in the $\mathbf{M^{ker}C^{par}}$ model

## 6.5 Gaussian Mixture Margins with Parametric Copulas ($\mathbf{M^{mix}C^{par}}$)

If we use Gaussian mixture margins to transform the data to the copula scale and then fit the copulas using the D-vine regression allowing only for parametric copulas this results in a log-likelihood of **18933.93**, an $AIC_C$ of **-37799.87** and a $BIC_C$ of **-37571.19** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 800.24 | -1592.48 | -1565.58 |
| mek | raf, pka, pkc | 5198.59 | -10389.18 | -10362.28 |
| plc | pip3 | 28.55 | -55.10 | -48.37 |
| pip2 | plc, pip3 | 2653.53 | -5299.06 | -5272.16 |
| erk | pka, mek | 278.28 | -550.56 | -530.38 |
| akt | erk, pka, pip3 | 3273.49 | -6534.98 | -6494.63 |
| pka | pkc | 671.30 | -1338.60 | -1325.14 |
| pkc | plc, pip2 | 682.87 | -1359.73 | -1339.55 |
| p38 | pkc, pka | 4004.82 | -8001.63 | -7974.73 |
| jnk | pkc, pka | 1342.27 | -2678.55 | -2658.37 |
| $\sum$: | | 18933.93 | -37799.87 | -37571.19 |

Table 6.9: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{mix}C^{par}}$ model

If one wants to look closer at the results for each node they can be seen in Table 6.9. There, it is displayed that again all dependencies are modeled. The individual copulas which are modeled are shown in Table 6.10. Here, again no single copula is modeled as a Gaussian copula. The contour plots of the fitted copulas are given in Section 6.6 in Figure 6.2.

| Node | Pair copula | Family | Rotation | Copula log-lik. | $AIC_C$ | $BIC_C$ | 1. Par. | 2. Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|---|---|
| raf | raf pka | bb8 | 270 | 776.54 | -1549.07 | -1535.62 | 1.77 | 0.97 | -0.22 |
| | raf pkc; pka | t | 0 | 23.71 | -43.41 | -29.96 | 0 | 10.52 | 0.10 |
| mek | mek raf | bb8 | 0 | 5109.24 | -10214.48 | -10201.03 | 4.96 | 1.00 | 0.64 |
| | mek pka; raf | clayton | 90 | 54.39 | -106.77 | -100.05 | 0.15 | | 0.11 |
| | mek pkc; raf pka | joe | 180 | 34.96 | -67.93 | -61.20 | 1.06 | | -0.22 |
| plc | plc pip3 | clayton | 180 | 28.55 | -55.10 | -48.37 | 0.10 | | 0 |
| pip2 | pip2 plc | bb7 | 0 | 2016.64 | -4029.28 | -4015.83 | 2.14 | 0.37 | 0.38 |
| | pip2 pip3; plc | bb8 | 0 | 636.89 | -1269.78 | -1256.33 | 3.99 | 0.57 | 0.25 |
| erk | erk pka | t | 0 | 230.29 | -456.58 | -443.13 | 0.07 | 2.84 | 0.06 |
| | erk mek; pka | clayton | 270 | 47.99 | -93.98 | -87.25 | 0.11 | | 0.03 |
| akt | akt erk | bb8 | 0 | 2884.33 | -5764.67 | -5751.22 | 4.59 | 0.88 | 0.59 |
| | akt pka; erk | bb8 | 270 | 382.00 | -760.00 | -746.55 | 1.45 | 0.97 | 0.01 |
| | akt pip3; erk pka | bb8 | 90 | 7.15 | -10.31 | 3.14 | 1.16 | 0.79 | -0.09 |
| pka | pka pkc | bb8 | 90 | 671.30 | -1338.60 | -1325.14 | 1.54 | 0.99 | -0.15 |
| pkc | pkc plc | bb8 | 90 | 675.27 | -1346.54 | -1333.09 | 1.54 | 0.99 | 0.14 |
| | pkc pip2; plc | frank | 0 | 7.60 | -13.19 | -6.46 | 0.29 | | 0.16 |
| p38 | p38 pkc | bb7 | 0 | 3839.17 | -7674.34 | -7660.89 | 3.81 | 0.06 | 0.57 |
| | p38 pka; pkc | bb8 | 270 | 165.65 | -327.29 | -313.84 | 1.40 | 0.93 | -0.20 |
| jnk | jnk pkc | bb7 | 0 | 1284.02 | -2564.04 | -2550.59 | 1.78 | 0.15 | 0.25 |
| | jnk pka; pkc | joe | 270 | 58.25 | -114.51 | -107.78 | 1.10 | | -0.13 |

Table 6.10: Summary of all copulas fitted in the $\mathbf{M^{mix}C^{par}}$ model

## 6.6 Gaussian Margins with Parametric Copulas ($\mathbf{M^{gauss}C^{par}}$)

Instead using Gaussian margins to transform the data to the copula scale and then fit the copulas using the D-vine regression allowing only for parametric copulas results in a log-likelihood of **20722.79**, an $AIC_C$ of **-41373.59** and a $BIC_C$ of **-41131.46** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 1097.47 | -2186.95 | -2160.05 |
| mek | raf, pka, pkc | 5605.06 | -11198.12 | -11157.77 |
| plc | pip3 | 9.05 | -16.10 | -9.37 |
| pip2 | plc, pip3 | 2292.29 | -4576.59 | -4549.68 |
| erk | mek, pka | 202.24 | -396.49 | -369.58 |
| akt | erk, pka, pip3 | 3618.07 | -7224.14 | -7183.79 |
| pka | pkc | 1023.43 | -2042.87 | -2029.42 |
| pkc | plc | 953.83 | -1903.65 | -1890.20 |
| p38 | pkc, pka | 4293.46 | -8580.92 | -8560.74 |
| jnk | pkc, pka | 1627.88 | -3247.76 | -3220.86 |
| $\sum$: | | 20722.79 | -41373.59 | -41131.46 |

Table 6.11: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{gauss}C^{par}}$ model

The goodness of fit measures of the individual nodes can be found in Table 6.11. Note that in this model the dependency of *pkc* from *pip*2 is not modeled. The individual copulas are displayed in Table 6.12. Here, as well as in the other two models where only parametric copulas are used, no single copula is modeled as a Gaussian copula. The contour plots of the fitted copulas are displayed in Figure 6.2.

| Node | Pair copula | Family | Rotation | Copula log-lik. | $AIC_C$ | $BIC_C$ | 1. Par. | 2. Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|---|---|
| raf | raf pka | bb8 | 270 | 1013.06 | -2022.11 | -2008.66 | 2.45 | 0.88 | -0.22 |
| | raf pkc; pka | bb8 | 270 | 84.42 | -164.84 | -151.39 | 1.13 | | 0.10 |
| mek | mek raf | bb6 | 0 | 4895.28 | -9786.55 | -9773.10 | 1.99 | 1.95 | 0.64 |
| | mek pka; raf | bb8 | 270 | 644.12 | -1284.24 | -1270.79 | 1.79 | 0.95 | 0.11 |
| | mek pkc; raf pka | t | 0 | 65.67 | -127.33 | -113.88 | -0.01 | 6.64 | -0.22 |
| plc | plc pip3 | clayton | 180 | 9.05 | -16.10 | -9.37 | 0.04 | | 0 |
| pip2 | pip2 plc | t | 0 | 1572.98 | -3141.95 | -3128.50 | 0.62 | 14.81 | 0.38 |
| | pip2 pip3; plc | bb8 | 0 | 719.32 | -1434.64 | -1421.18 | 5.52 | 0.47 | 0.25 |
| erk | erk mek | t | 0 | 171.61 | -339.21 | -325.76 | 0.06 | 4.36 | 0.03 |
| | erk pka; mek | t | 0 | 30.64 | -57.28 | -43.82 | -0.04 | 11.12 | 0.06 |
| akt | akt erk | bb8 | 0 | 2487.68 | -4971.35 | -4957.90 | 5.57 | 0.75 | 0.59 |
| | akt pka; erk | bb8 | 270 | 1123.02 | -2242.05 | -2228.60 | 2.17 | 0.92 | 0.01 |
| | akt pip3; erk pka | bb8 | 90 | 7.37 | -10.74 | 2.71 | 1.08 | 0.99 | -0.09 |
| pka | pka pkc | bb8 | 90 | 1023.43 | -2042.87 | -2029.42 | 1.76 | 0.98 | -0.15 |
| pkc | pkc plc | bb8 | 90 | 953.83 | -1903.65 | -1890.20 | 1.76 | 0.98 | 0.14 |
| p38 | p38 pkc | gumbel | 0 | 3081.91 | -6161.82 | -6155.09 | 2.23 | | 0.57 |
| | p38 pka; pkc | bb8 | 270 | 1211.55 | -2419.10 | -2405.65 | 3.98 | 0.74 | -0.20 |
| jnk | jnk pkc | bb7 | 0 | 1448.77 | -2893.53 | -2880.08 | 1.74 | 0.20 | 0.25 |
| | jnk pka; pkc | bb7 | 90 | 179.12 | -354.23 | -340.78 | 1.06 | 0.18 | -0.13 |

Table 6.12: Summary of all copulas fitted in the $\mathbf{M^{gauss}C^{par}}$ model


(a) $\mathbf{M^{ker}C^{par}}$ (b) $\mathbf{M^{mix}C^{par}}$ (c) $\mathbf{M^{gauss}C^{par}}$

Figure 6.2: Contour plots of the respective copulas in the models with only parametric copulas where kernel density margins, Gaussian mixture margins or Gaussian margins were used for the PIT

## 6.7 Kernel Density Margins with Gaussian Copulas ($\mathbf{M^{ker}C^{gauss}}$)

Restricting the set of copula families to only Gaussian copulas in the D-vine regression and using kernel density margins to transform the data to the copula scale results in a log-likelihood of **11874.60**, an $AIC_C$ of **-23709.19** and a $BIC_C$ of **-23574.66** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 412.65 | -821.30 | -807.85 |
| mek | raf, pkc, pka | 3271.45 | -6536.90 | -6516.71 |
| plc | pip3 | 4.39 | -6.78 | -0.06 |
| pip2 | plc, pip3 | 1910.02 | -3816.05 | -3802.59 |
| erk | pka, mek | 38.61 | -73.23 | -59.78 |
| akt | erk, pka, pip3 | 2353.47 | -4700.93 | -4680.75 |
| pka | pkc | 251.68 | -501.36 | -494.63 |
| pkc | plc, pip2 | 320.93 | -637.87 | -624.41 |
| p38 | pkc, pka | 2455.79 | -4907.58 | -4894.14 |
| jnk | pkc, pka | 855.60 | -1707.19 | -1693.74 |
| $\sum$: | | 11874.60 | -23709.19 | -23574.66 |

Table 6.13: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{ker}C^{gauss}}$ model

Detailed results for each node are displayed in Table 6.13. As in the previous models where kernel density margins were used for the PIT all dependencies from the DAG are modeled. A list of copulas can be found in Table 6.14. The contour plots of the fitted copulas are displayed in Section 6.9 in Figure 6.3 together with the ones from the $\mathbf{M^{mix}C^{gauss}}$ model and the $\mathbf{M^{gauss}C^{gauss}}$ model.

| Node | Pair copula | Family | Copula log-lik. | $AIC_C$ | $BIC_C$ | Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | gaussian | 398.67 | -795.33 | -788.61 | -0.35 | -0.22 |
| | raf pkc; pka | gaussian | 13.98 | -25.97 | -19.24 | 0.07 | 0.10 |
| mek | mek raf | gaussian | 3198.08 | -6394.16 | -6387.43 | 0.80 | 0.64 |
| | mek pkc; raf | gaussian | 51.39 | -100.79 | -94.06 | 0.13 | 0.11 |
| | mek pka; raf pkc | gaussian | 21.98 | -41.95 | -35.22 | -0.08 | -0.22 |
| plc | plc pip3 | gaussian | 4.39 | -6.78 | -0.06 | 0.04 | 0 |
| pip2 | pip2 plc | gaussian | 1405.94 | -2809.87 | -2803.14 | 0.60 | 0.38 |
| | pip2 pip3; plc | gaussian | 504.09 | -1006.18 | -999.45 | 0.39 | 0.25 |
| erk | erk pka | gaussian | 28.80 | -55.60 | -48.88 | 0.10 | 0.06 |
| | erk mek; pka | gaussian | 9.81 | -17.63 | -10.90 | 0.06 | 0.03 |
| akt | akt erk | gaussian | 2224.36 | -4446.71 | -4439.99 | 0.72 | 0.59 |
| | akt pka; erk | gaussian | 123.62 | -245.24 | -238.51 | -0.20 | 0.01 |
| | akt pip3; erk pka | gaussian | 5.49 | -8.98 | -2.25 | -0.04 | -0.09 |
| pka | pka pkc | gaussian | 251.68 | -501.36 | -494.63 | -0.28 | -0.15 |
| pkc | pkc plc | gaussian | 261.19 | -520.38 | -513.65 | 0.28 | 0.14 |
| | pkc pip2; plc | gaussian | 59.74 | -117.49 | -110.76 | 0.14 | 0.16 |
| p38 | p38 pkc | gaussian | 2298.39 | -4594.78 | -4588.06 | 0.72 | 0.57 |
| | p38 pka; pkc | gaussian | 157.40 | -312.80 | -306.08 | -0.22 | -0.20 |
| jnk | jnk pkc | gaussian | 776.30 | -1550.60 | -1543.87 | 0.47 | 0.25 |
| | jnk pka; pkc | gaussian | 79.30 | -156.59 | -149.87 | -0.16 | -0.13 |

Table 6.14: Summary of all copulas fitted in the $\mathbf{M^{ker}C^{gauss}}$ model

## 6.8 Gaussian Mixture Margins with Gaussian Copulas ($\mathbf{M^{mix}C^{gauss}}$)

Using Gaussian mixture margins for the PIT to transform the data to the copula scale and then using only Gaussian copulas in the D-vine regression results in a log-likelihood of **11767.73**, an $AIC_C$ of **-23495.45** and a $BIC_C$ of **-23360.93** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 417.38 | -830.75 | -817.30 |
| mek | raf, pkc, pka | 3261.15 | -6516.30 | -6496.12 |
| plc | pip3 | 2.79 | -3.58 | 3.14 |
| pip2 | plc, pip3 | 1889.09 | -3774.17 | -3760.72 |
| erk | pka, mek | 36.19 | -68.38 | -54.93 |
| akt | erk, pka, pip3 | 2243.36 | -4480.71 | -4460.53 |
| pka | pkc | 256.32 | -510.65 | -503.92 |
| pkc | plc, pip2 | 324.13 | -644.26 | -630.81 |
| p38 | pkc, pka | 2467.63 | -4931.26 | -4917.80 |
| jnk | pkc, pka | 869.70 | -1735.39 | -1721.94 |
| $\sum$: | | 11767.73 | -23495.45 | -23360.93 |

Table 6.15: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{mix}C^{gauss}}$ model

Results for each node are displayed in Table 6.15. Again as in the previous models where Gaussian mixture margins were used for the PIT all dependencies from the DAG are modeled. The copulas that are estimated in the $\mathbf{M^{mix}C^{gauss}}$ model can be found in Table 6.16. Additionaly the contour plots of the fitted copulas are displayed in Section 6.9 in Figure 6.3.

| Node | Pair copula | Family | Copula log-lik. | $AIC_C$ | $BIC_C$ | Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | gaussian | 403.60 | -805.20 | -798.48 | -0.35 | -0.22 |
| | raf pkc; pka | gaussian | 13.77 | -25.55 | -18.82 | 0.07 | 0.10 |
| mek | mek raf | gaussian | 3187.30 | -6372.61 | -6365.88 | 0.80 | 0.64 |
| | mek pkc; raf | gaussian | 51.01 | -100.03 | -93.30 | 0.13 | 0.11 |
| | mek pka; raf pkc | gaussian | 22.83 | -43.66 | -36.94 | -0.09 | -0.22 |
| plc | plc pip3 | gaussian | 2.79 | -3.58 | 3.14 | 0.03 | 0 |
| pip2 | pip2 plc | gaussian | 1402.82 | -2803.64 | -2796.91 | 0.60 | 0.38 |
| | pip2 pip3; plc | gaussian | 486.27 | -970.53 | -963.81 | 0.38 | 0.25 |
| erk | erk pka | gaussian | 29.55 | -57.11 | -50.38 | 0.10 | 0.06 |
| | erk mek; pka | gaussian | 6.64 | -11.27 | -4.55 | 0.05 | 0.03 |
| akt | akt erk | gaussian | 2113.49 | -4224.97 | -4218.24 | 0.70 | 0.59 |
| | akt pka; erk | gaussian | 124.47 | -246.94 | -240.21 | -0.20 | 0.01 |
| | akt pip3; erk pka | gaussian | 5.40 | -8.80 | -2.08 | -0.04 | -0.09 |
| pka | pka pkc | gaussian | 256.32 | -510.65 | -503.92 | -0.28 | -0.15 |
| pkc | pkc plc | gaussian | 260.53 | -519.06 | -512.33 | 0.28 | 0.14 |
| | pkc pip2; plc | gaussian | 63.60 | -125.20 | -118.48 | 0.14 | 0.16 |
| p38 | p38 pkc | gaussian | 2314.38 | -4626.77 | -4620.04 | 0.73 | 0.57 |
| | p38 pka; pkc | gaussian | 153.24 | -304.49 | -297.76 | -0.22 | -0.20 |
| jnk | jnk pkc | gaussian | 788.25 | -1574.51 | -1567.78 | 0.48 | 0.25 |
| | jnk pka; pkc | gaussian | 81.44 | -160.88 | -154.16 | -0.16 | -0.13 |

Table 6.16: Summary of all copulas fitted in the $\mathbf{M^{mix}C^{gauss}}$ model

## 6.9 Gaussian Margins with Gaussian Copulas ($\mathbf{M^{gauss}C^{gauss}}$)

Finally using Gaussian margins together with Gaussian copulas in the D-vine regression results in a log-likelihood of **16144.37**, an $AIC_C$ of **-32254.75** and a $BIC_C$ of **-32140.40** of the copula terms.

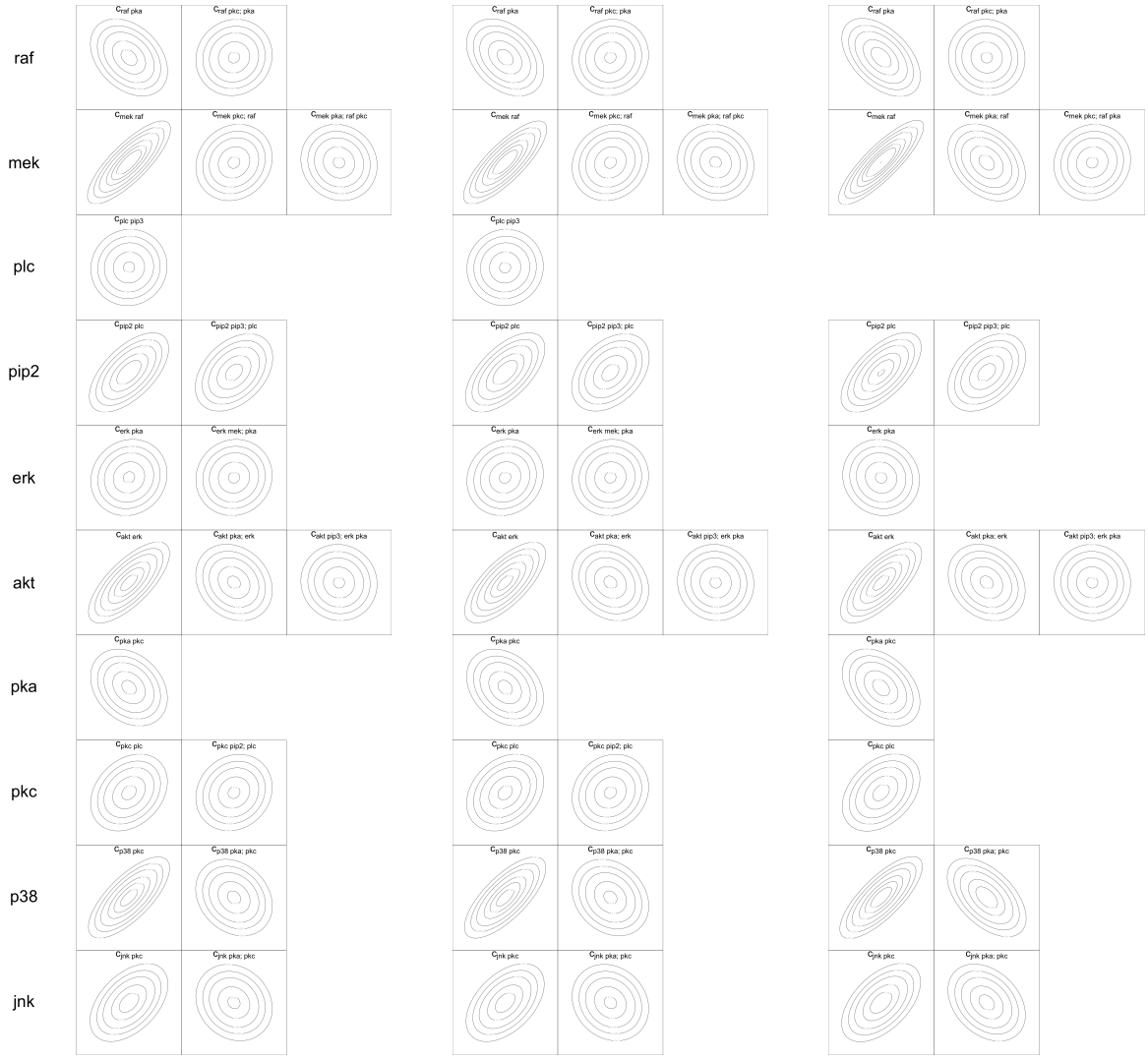
| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 830.82 | -1657.63 | -1644.18 |
| mek | raf, pka, pkc | 4268.49 | -8530.99 | -8510.81 |
| plc | | | | |
| pip2 | plc, pip3 | 2105.36 | -4206.71 | -4193.25 |
| erk | pka | 20.58 | -39.15 | -32.43 |
| akt | erk, pka, pip3 | 2625.44 | -5244.88 | -5224.71 |
| pka | pkc | 690.23 | -1378.45 | -1371.73 |
| pkc | plc | 601.84 | -1201.67 | -1194.94 |
| p38 | pkc, pka | 3652.30 | -7300.60 | -7287.14 |
| jnk | pkc, pka | 1349.33 | -2694.67 | -2681.21 |
| $\sum$: | | 16144.37 | -32254.75 | -32140.40 |

Table 6.17: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{gauss}C^{gauss}}$ model

Table 6.17 shows the results for each node. In this model three dependencies of the graph are not modeled. Namely $mek \to erk$, $pip2 \to pkc$ and $pip3 \to plc$. Especially the last one is interesting as it is the only dependency of the node $plc$, meaning that in this model $plc$ is modeled independent of all other nodes. The remaining copulas which are estimated can be found in Table 6.18 and the respective contour plots in Figure 6.3.

| Node | Pair copula | Family | Copula log-lik. | $AIC_C$ | $BIC_C$ | Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | gaussian | 826.44 | -1650.89 | -1644.16 | -0.49 | -0.22 |
| | raf pkc; pka | gaussian | 4.37 | -6.74 | -0.02 | -0.04 | 0.10 |
| mek | mek raf | gaussian | 3915.99 | -7829.99 | -7823.26 | 0.85 | 0.64 |
| | mek pka; raf | gaussian | 336.79 | -671.59 | -664.86 | -0.32 | 0.11 |
| | mek pkc; raf pka | gaussian | 15.71 | -29.41 | -22.69 | 0.07 | -0.22 |
| pip2 | pip2 plc | gaussian | 1545.35 | -3088.70 | -3081.97 | 0.63 | 0.38 |
| | pip2 pip3; plc | gaussian | 560.01 | -1118.01 | -1111.28 | 0.41 | 0.25 |
| erk | erk pka | gaussian | 20.58 | -39.15 | -32.43 | -0.07 | 0.06 |
| akt | akt erk | gaussian | 2125.42 | -4248.85 | -4242.12 | 0.71 | 0.59 |
| | akt pka; erk | gaussian | 498.70 | -995.39 | -988.67 | -0.21 | 0.01 |
| | akt pip3; erk pka | gaussian | 1.32 | -0.64 | 6.08 | -0.02 | -0.09 |
| pka | pka pkc | gaussian | 690.23 | -1378.45 | -1371.73 | -0.38 | -0.15 |
| pkc | pkc plc | gaussian | 601.84 | -1201.67 | -1194.94 | 0.36 | 0.14 |
| p38 | p38 pkc | gaussian | 2697.70 | -5393.40 | -5386.67 | 0.76 | 0.57 |
| | p38 pka; pkc | gaussian | 954.60 | -1907.20 | -1900.47 | -0.53 | -0.20 |
| jnk | jnk pkc | gaussian | 1032.88 | -2063.77 | -2057.04 | 0.52 | 0.25 |
| | jnk pka; pkc | gaussian | 316.45 | -630.90 | -624.17 | -0.32 | -0.13 |

Table 6.18: Summary of all copulas fitted in the $\mathbf{M^{gauss}C^{gauss}}$ model




(a) $\mathbf{M^{ker}C^{gauss}}$ (b) $\mathbf{M^{mix}C^{gauss}}$ (c) $\mathbf{M^{gauss}C^{gauss}}$

Figure 6.3: Contour plots of the respective copulas in the models with only Gaussian copulas where kernel density margins, Gaussian mixture margins or Gaussian margins were used for the PIT

As we have already stated in Section 5.1 the conditions to fit a D-vine after using Gaussian margins for the PIT are not fulfilled. Here, we observe this by noting that the relationship between Kendall's $\tau$ and the parameter of the Gaussian copulas, as derived in Theorem 7, is not fulfilled for all copulas.

Furthermore, we have observed in Section 2.4.5 that a Gaussian D-vine specifies a Gaussian distribution. Conditioning on the D-vine therefore results in a conditional Gaussian distribution. The respective parameters of the conditional distributions can be found in Table 6.19.

| Node | $\hat{\beta}_0$ | $\hat{\boldsymbol{\beta}}$ | $\hat{\sigma}$ |
|---|---|---|---|
| raf | 6.29 | ( -0.36, -0.03) | 0.89 |
| mek | 0.55 | (0.98, -0.17, 0.05) | 0.49 |
| plc | 2.97 | | 1.59 |
| pip2 | 0.08 | (0.81, 0.53) | 1.33 |
| pip3 | 2.84 | | 0.93 |
| erk | 3.12 | -0.05 | 1.00 |
| akt | 2.52 | (0.69, -0.10, -0.02) | 0.47 |
| pka | 7.13 | -0.49 | 1.89 |
| pkc | 1.69 | 0.33 | 1.16 |
| p38 | 3.71 | (-0.30, 0.66) | 0.45 |
| jnk | 3.58 | (-0.28, 0.49) | 1.23 |

Table 6.19: Estimated parameters of the conditional normal distributions $N(\beta_0+\boldsymbol{\beta}^\top x;\sigma^2)$ of each node on the set of its parents modeled in the Gaussian D-vine in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model. The order of the parents is the same as in Table 4.2

## 6.10 Comparison of D-vine Regression Models

After fitting nine different D-vine regression models with different copula families and margins we can observe that on this data set the choice of copula families has a huge influence. None of the models where only parametric copulas are used has a comparable fit close to the goodness of fit of any model where parametric and non-parametric copulas are used. The same holds when only Gaussian copulas are used compared to the use of parametric copulas. Hence, in this section we will only compare how the models within the use of a certain class of copulas behave with the change of margins used.

When fitting kernel density margins and Gaussian mixture margins in Chapter 5 we have already seen that both are very similar. Hence, we assumed that the data set after using both of them for the PIT is very similar as well. Now, after fitting several D-vine models to both data sets, our assumption can be confirmed. For each of the three choices of copula families, $\mathbf{C^{pnp}}$, $\mathbf{C^{par}}$ and $\mathbf{C^{gauss}}$, the models have a similar log-likelihood of the copula terms, $\text{AIC}_C$ and $\text{BIC}_C$ when using kernel density margins or Gaussian mixture margins.

This does not only hold when looking at the whole model, but also when looking at each node alone. Even the order of the D-vines are the same and many similarities in the individual modeled copulas are present. When allowing for parametric and non-parametric copulas only the $c_{plc,pip3}$ copula and the $c_{pkc,pip2;plc}$ copula are modeled differently in the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model. All other copulas are from the same family and, if they are modeled as TLL-copula, have a similar number of effective parameters. In the $c_{plc,pip3}$ copula even only the number of parameters differs. Comparing the contour plots in Figure 6.1 it is hard to observe any difference.

When instead restricting to only parametric copulas we can see that seventeen of twenty of the modeled copulas are from the the same family and have the same rotation and similar parameters in the $\mathbf{M^{ker}C^{par}}$ model and the $\mathbf{M^{mix}C^{par}}$ model. Note that the Joe copula corresponds to a BB8 copula with a second parameter that equals one, as in Cheng et al. (2020). Therefore, a BB8 copula with a second parameter near one almost defines a Joe copula and hence, even 18 of the copulas almost agree. Only the copulas $c_{pkc,plc}$ and $c_{p38,pkc}$ are modeled differently in both models where for the $c_{pkc,plc}$ copula only the rotation does not coincide. Comparing the contour plots of the copulas in both models in Figure 6.2 it is again almost impossible to note any differences, even in the $c_{p38,pkc}$ copula.

In the approach with only Gaussian copulas all copulas have a similar parameter in both the $\mathbf{M^{ker}C^{gauss}}$ model and the $\mathbf{M^{mix}C^{gauss}}$ model. Here, once more it is hardly possible to see differences in the contour plots for both models in Figure 6.3. Hence, to summarize, in regards to the copula terms there are no big differences whether kernel density margins

or Gaussian mixture margins are used in the model. Instead the fit only depends on the choice of copulas in the D-vine regression.

Comparing these models with the models where we used Gaussian margins for the data transformation the bigger variations in the orders of the D-vine is what stands out first. For the $\mathbf{M^{ker}C^{pnp}}$ model, the $\mathbf{M^{ker}C^{gauss}}$ model, the $\mathbf{M^{mix}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{gauss}}$ model all D-vines are ordered in the same way. The $\mathbf{M^{ker}C^{par}}$ model and the $\mathbf{M^{mix}C^{par}}$ model differ only from these models for the order of the D-vine for the node $mek$. In each of these models all dependencies from the DAG are modeled.

This is not the case when using Gaussian margins. There, for each choice of copulas at least one dependency is not modeled and orders might not coincide. This is namely for $pip3 \to akt$ in the $\mathbf{M^{gauss}C^{pnp}}$ model, $pip2 \to pkc$ in the $\mathbf{M^{gauss}C^{par}}$ model and the $\mathbf{M^{gauss}C^{gauss}}$ model, $pip3 \to plc$ and $mek \to erk$ in the $\mathbf{M^{gauss}C^{gauss}}$ model. While not modeling the dependencies $pip3 \to akt$, $pip2 \to pkc$ and $pip3 \to plc$ seems reasonable due to their low partial correlations of 0, 0.06 and $-0.04$ this does not hold for $mek \to erk$. It has a partial correlation of $-0.35$ which is the seventh highest of all pairs of nodes. Figure 6.4 graphically displays the order of the D-vines in each model.

Speaking of the goodness of fit, we observe that the $\mathbf{M^{gauss}C^{pnp}}$ model in total has a worse fit speaking of log-likelihood, $\text{AIC}_C$ and $\text{BIC}_C$ than the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model. This worsening does not affect every node similar. While some like $akt$ and $p38$ suffer a modest worsening, the fit of $pip2$ shrinks heavily. On the other hand there are nodes that have a similar of even better fit, like e.g. $mek$, than in the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model. While when using kernel density margins or Gaussian mixture margins only one or two copulas are modeled as a parametric copula, in the $\mathbf{M^{gauss}C^{pnp}}$ model six of nineteen copulas are parametric ones. Comparing the contour plots of the $\mathbf{M^{gauss}C^{pnp}}$ model to the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model in Figure 6.1 several notable differences are present for example in the $c_{mek,raf}$ or the $c_{pip2,plc}$ copula.

When instead using Gaussian margins and only allowing for parametric copulas the $\mathbf{M^{gauss}C^{par}}$ model in total outperforms the $\mathbf{M^{ker}C^{par}}$ model and the $\mathbf{M^{mix}C^{par}}$ model in all three measures. Looking at the individual nodes, we observe that only the D-vines at $plc$, $pip2$ and $erk$ are modeled better in the $\mathbf{M^{ker}C^{par}}$ model and the $\mathbf{M^{mix}C^{par}}$ model. All other nodes have a better goodness of fit in regards of the log-likelihood of the copula terms, $\text{AIC}_C$ and $\text{BIC}_C$ when using Gaussian margins. As already eight of nineteen copulas are modeled with a different copula family in the $\mathbf{M^{gauss}C^{par}}$ model than in the $\mathbf{M^{ker}C^{par}}$ model and the $\mathbf{M^{mix}C^{par}}$ model it does not surprise that we observe many differences in the contour plots in Figure 6.2.

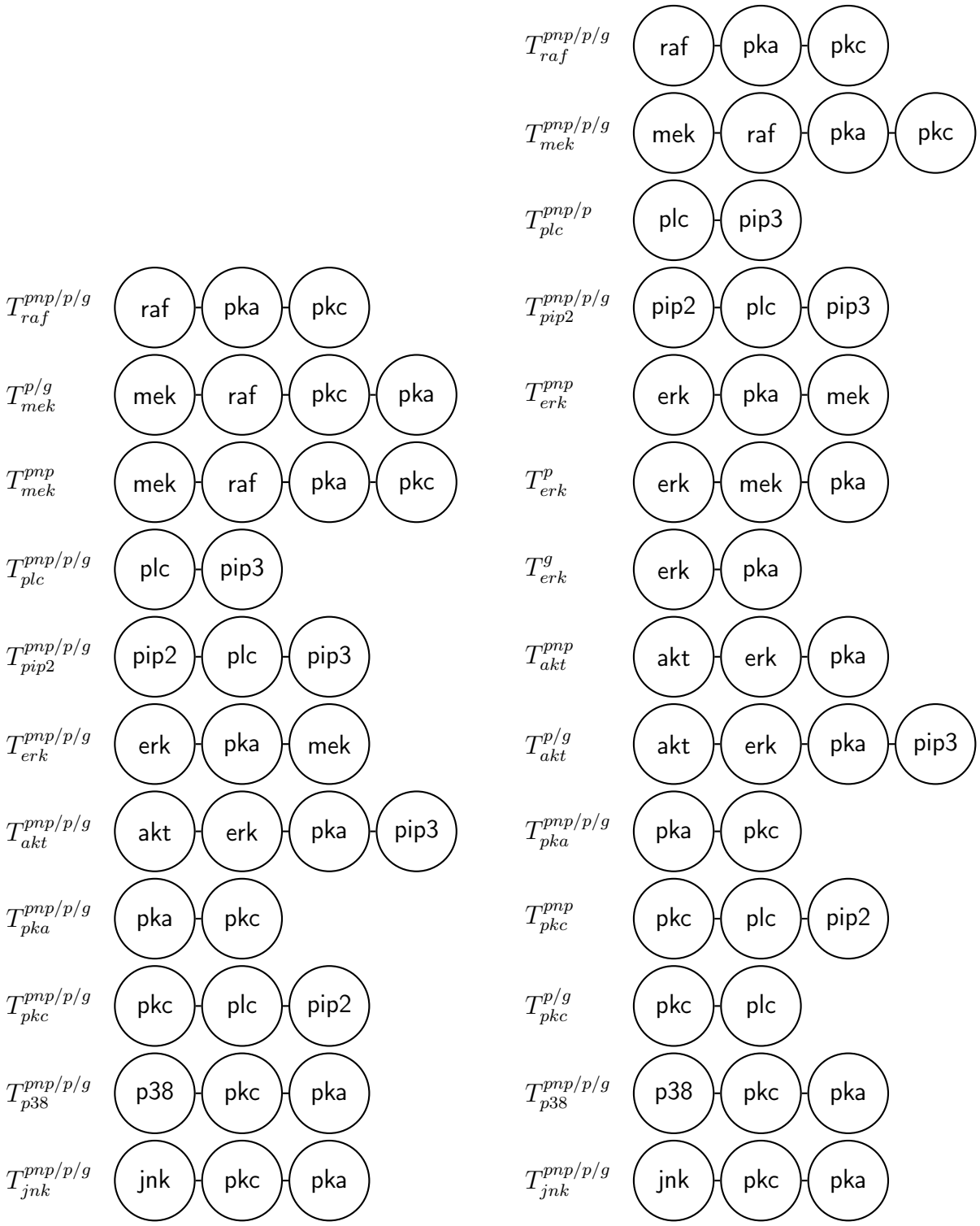



Figure 6.4: Fitted D-vines of each node on the set of its parents when using kernel density margins $\mathbf{M^{ker}}$ or Gaussian mixture margins $\mathbf{M^{mix}}$ for the PIT (left) or Gaussian margins $\mathbf{M^{gauss}}$ (right). Text in superscript indicating which kind of copulas are used, i.e., pnp = $\mathbf{C^{pnp}}$, p = $\mathbf{C^{par}}$ or g = $\mathbf{C^{gauss}}$

We can see a similar behavior comparing the models where we restricted to Gaussian copulas only. There, the copula terms of the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model outperform the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{gauss}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{gauss}}$ model in all three measures for all nodes except *plc* and *erk*. However, for *plc* and *erk* the fit is generally quite poor in all three models which reduces the significance here.

We have already seen that in all three models where Gaussian margins are used, at least one dependency in the DAG is not modeled. This does not play a big role in the overall fit when compared to the other models as the missing dependencies only contribute slightly to the fit of the D-vine in the models where it is present. Generally one might think about truncating some of the D-vines. Some nodes, and therefore dependencies, modeled have a very low Kendall's $\tau$ such that their influence can almost be neglected. Examples are $erk \to pka$ with a Kendall's $\tau$ of 0.06 or $pip3 \to plc$ with a Kendall's $\tau$ of 0.

Overall, we observe that using kernel density margins or Gaussian mixture margins results in almost the same copula terms. Comparing them to the models where Gaussian margins are used the Gaussian margins models resulted in a better fit when using only parametric or Gaussian copulas but in a worse fit when using parametric and non-parametric copulas.

# Chapter 7

# Goodness of Fit Measures Model Comparison

While fitting the D-vine models in Chapter 6, we have observed that restrictions on copula families or margins always come in line with a worsening of the goodness of fit measure that we optimize. In this chapter we will now analyze this in more depth. More precisely, we want to compare the fit of the conditional densities calculated for each node. For this we add up the goodness of fit measures of the marginals and the copulas. The question then is, whether this worsening, caused by the restriction on the copulas or margins, in each model affects every node equally. Further, we compare the goodness of fit of the D-vine models to the Linear Gaussian Bayesian Network.

In the following, we will, if not stated otherwise, only compare the models using the $\text{AIC}_F$ as this is the measure we used to optimize in the copula models. Note that the Linear Gaussian Bayesian Network was fitted maximizing the log-likelihood. However, as we have already described in Chapter 4, if we would have instead maximized the $\text{AIC}_F$ a possible increase in the $\text{AIC}_F$ would be, taking the absolute value into account, almost vanishing and only less dependencies would have been modeled.

A summary of the log-likelihood, the $\text{AIC}_F$, the $\text{BIC}_F$ and the number of parameters of the conditional densities of the fitted models are displayed in the Tables 7.1, 7.2, 7.3 and 7.4. We already derived that in general it does not play a role if Gaussian mixture margins or kernel density margins were used for the PIT on this data set. In both models the $\text{AIC}_F$ only depends on the chosen copulas, with the model where kernel density margins are used being slightly better. Using Gaussian margins results in different outcomes which are in general worse than in the other two models. This is mostly because of the bad fit of the margins which prevails the better fit of the copulas in the models with only parametric or Gaussian copulas.

| Copulas | Margins | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LGBN | | -8374 | -6521 | -10161 | -9619 | -8514 | -8743 | -5812 | -10459 | -9009 | -6307 | -9362 | -92882 |
| $\mathbf{C^{pnp}}$ | $\mathbf{M^{ker}}$ | -7704 | -3668 | -8704 | -6906 | -8416 | -7181 | -2962 | -7648 | -7865 | -3544 | -8338 | -72937 |
| | $\mathbf{M^{mix}}$ | -7726 | -3779 | -8761 | -7035 | -8514 | -7205 | -3014 | -7710 | -7971 | -3559 | -8385 | -73660 |
| | $\mathbf{M^{gauss}}$ | -8034 | -4905 | -9996 | -9308 | -8514 | -7685 | -4222 | -9974 | -8276 | -5903 | -8547 | -85365 |
| $\mathbf{C^{par}}$ | $\mathbf{M^{ker}}$ | -8150 | -4026 | -9197 | -8572 | -8416 | -8424 | -4862 | -8176 | -8431 | -4173 | -9168 | -81594 |
| | $\mathbf{M^{mix}}$ | -8172 | -4163 | -9249 | -8679 | -8514 | -8487 | -4879 | -8238 | -8513 | -4219 | -9197 | -82311 |
| | $\mathbf{M^{gauss}}$ | -8107 | -5184 | -10152 | -9432 | -8514 | -8563 | -5093 | -10159 | -8679 | -5667 | -9085 | -88636 |
| $\mathbf{C^{gauss}}$ | $\mathbf{M^{ker}}$ | -8531 | -6015 | -9225 | -9337 | -8416 | -8702 | -5757 | -8599 | -8808 | -5724 | -9647 | -88762 |
| | $\mathbf{M^{mix}}$ | -8555 | -6100 | -9275 | -9443 | -8514 | -8729 | -5909 | -8653 | -8872 | -5757 | -9670 | -89477 |
| | $\mathbf{M^{gauss}}$ | -8374 | -6521 | -10161 | -9619 | -8514 | -8745 | -6085 | -10492 | -9031 | -6309 | -9363 | -93214 |

Table 7.1: Global log-likelihood of the models fitted on the Sachs data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Copulas | Margins | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LGBN | | 16756 | 13052 | 20328 | 19246 | 17031 | 17495 | 11634 | 20925 | 18027 | 12621 | 18733 | 185847 |
| $\mathbf{C^{pnp}}$ | $\mathbf{M^{ker}}$ | 15592 | 7561 | 17486 | 13943 | 16855 | 14503 | 6102 | 15403 | 15870 | 7261 | 16806 | 147382 |
| | $\mathbf{M^{mix}}$ | 15598 | 7757 | 17612 | 14197 | 17031 | 14543 | 6184 | 15495 | 16014 | 7268 | 16896 | 148596 |
| | $\mathbf{M^{gauss}}$ | 16186 | 9996 | 20105 | 18713 | 17031 | 15513 | 8948 | 20099 | 16753 | 11901 | 17180 | 172425 |
| $\mathbf{C^{par}}$ | $\mathbf{M^{ker}}$ | 16337 | 8097 | 18424 | 17172 | 16855 | 16862 | 9773 | 16396 | 16891 | 8386 | 18360 | 163554 |
| | $\mathbf{M^{mix}}$ | 16362 | 8355 | 18516 | 17382 | 17031 | 16985 | 9780 | 16496 | 17049 | 8469 | 18417 | 164842 |
| | $\mathbf{M^{gauss}}$ | 16227 | 10385 | 20311 | 18876 | 17031 | 17139 | 10201 | 20325 | 17367 | 11345 | 18181 | 177388 |
| $\mathbf{C^{gauss}}$ | $\mathbf{M^{ker}}$ | 17096 | 12076 | 18481 | 18699 | 16855 | 17415 | 11558 | 17241 | 17643 | 11485 | 19317 | 177866 |
| | $\mathbf{M^{mix}}$ | 17124 | 12228 | 18567 | 18907 | 17031 | 17467 | 11834 | 17324 | 17764 | 11539 | 19360 | 179146 |
| | $\mathbf{M^{gauss}}$ | 16756 | 13052 | 20327 | 19246 | 17031 | 17496 | 12180 | 20990 | 18069 | 12625 | 18734 | 186506 |

Table 7.2: $\mathrm{AIC}_F$ of the models fitted on the Sachs data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Copulas | Margins | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | ∑: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LGBN | | 16783 | 13086 | 20348 | 19273 | 17045 | 17522 | 11667 | 20952 | 18054 | 12648 | 18759 | 186136 |
| $\mathbf{C^{pnp}}$ | $\mathbf{M^{ker}}$ | 16212 | 8318 | 17752 | 14383 | 16928 | 14976 | 6699 | 15764 | 16340 | 7838 | 17242 | 152452 |
| | $\mathbf{M^{mix}}$ | 16088 | 8427 | 17911 | 14623 | 17045 | 14987 | 6709 | 15748 | 16255 | 7776 | 17320 | 152888 |
| | $\mathbf{M^{gauss}}$ | 16581 | 10620 | 20481 | 19038 | 17045 | 15993 | 10645 | 20602 | 17424 | 12223 | 17470 | 178124 |
| $\mathbf{C^{par}}$ | $\mathbf{M^{ker}}$ | 16464 | 8250 | 18529 | 17269 | 16928 | 16910 | 9939 | 16547 | 16988 | 8520 | 18444 | 164788 |
| | $\mathbf{M^{mix}}$ | 16423 | 8456 | 18576 | 17463 | 17045 | 17018 | 9854 | 16564 | 17123 | 8570 | 18491 | 165582 |
| | $\mathbf{M^{gauss}}$ | 16267 | 10439 | 20331 | 18916 | 17045 | 17179 | 10255 | 20352 | 17394 | 11378 | 18222 | 177778 |
| $\mathbf{C^{gauss}}$ | $\mathbf{M^{ker}}$ | 17210 | 12228 | 18586 | 18781 | 16928 | 17456 | 11704 | 17385 | 17733 | 11612 | 19394 | 179018 |
| | $\mathbf{M^{mix}}$ | 17171 | 12322 | 18628 | 18974 | 17045 | 17494 | 11888 | 17385 | 17832 | 11626 | 19427 | 179792 |
| | $\mathbf{M^{gauss}}$ | 16783 | 13086 | 20340 | 19273 | 17045 | 17516 | 12214 | 21010 | 18089 | 12652 | 18761 | 186769 |

Table 7.3: $\text{BIC}_F$ of the models fitted on the Sachs data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Copulas | Margins | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LGBN | | 4 | 5 | 3 | 4 | 2 | 4 | 5 | 3 | 4 | 4 | 4 | 42 |
| $\mathbf{C^{pnp}}$ | $\mathbf{M^{ker}}$ | 92 | 112 | 40 | 66 | 11 | 70 | 89 | 54 | 70 | 86 | 65 | 754 |
| | $\mathbf{M^{mix}}$ | 73 | 100 | 44 | 63 | 2 | 66 | 78 | 38 | 36 | 75 | 63 | 638 |
| | $\mathbf{M^{gauss}}$ | 59 | 93 | 56 | 48 | 2 | 71 | 252 | 75 | 100 | 48 | 43 | 847 |
| $\mathbf{C^{par}}$ | $\mathbf{M^{ker}}$ | 19 | 23 | 16 | 14 | 11 | 7 | 25 | 22 | 14 | 20 | 12 | 183 |
| | $\mathbf{M^{mix}}$ | 9 | 15 | 9 | 12 | 2 | 5 | 11 | 10 | 11 | 15 | 11 | 110 |
| | $\mathbf{M^{gauss}}$ | 6 | 8 | 3 | 6 | 2 | 6 | 8 | 4 | 4 | 5 | 6 | 58 |
| $\mathbf{C^{gauss}}$ | $\mathbf{M^{ker}}$ | 17 | 23 | 16 | 12 | 11 | 6 | 22 | 21 | 13 | 19 | 11 | 171 |
| | $\mathbf{M^{mix}}$ | 7 | 14 | 9 | 10 | 2 | 4 | 8 | 9 | 10 | 13 | 10 | 96 |
| | $\mathbf{M^{gauss}}$ | 4 | 5 | 2 | 4 | 2 | 3 | 5 | 3 | 3 | 4 | 4 | 39 |

Table 7.4: Number of (effective) parameters of the models fitted on the Sachs data set for each node

For the node *pip*3 we observe only differences in the way its marginal density is modeled, as the node has no parents in the consent graph. For *pip*3 the estimated Gaussian mixture margin is a normal distribution. Hence, it coincides with the one in the model with only Gaussian margins. Using kernel density margins here, gives only a minimal advantage as we have already stated when comparing the marginals in Section 5.4.

For most of the other nodes modeling them using a D-vine copula approach provides a significant advantage when using Gaussian mixture margins or kernel density margins compared to the LGBN. This especially holds when allowing for parametric and non-parametric copulas. We observe that in the models with parametric and non-parametric copulas with other margins than Gaussian margins, all nodes have a better $\text{AIC}_F$ than the Linear Gaussian Bayesian Network, most of them even more than 10% better. If we use Gaussian margins instead the overall $\text{AIC}_F$ is still better than in the LGBN but for some nodes the advantage shrinks heavily. Partly even that much that there is almost no difference any more such as for example the node *pka*.

The worsening that occurs through restricting on parametric copulas has different effects on each node. While for example *mek* and *p*38 only suffer a small worsening, keeping their good fit, the $\text{AIC}_F$ of *erk* increases strongly compared to the models with parametric and non-parametric copulas. This holds for all three models. We further observe that the reaction on the restrictions on copulas is different depending on the chosen margins. The node *plc* worsens significantly when the model is fitted using Gaussian mixture margins or kernel density margins but almost does not react to the restrictions when using Gaussian margins. Compared to the LGBN model especially *mek*, *pip*2, *akt*, *pka* and *p*38 still have a superior fit. Especially interesting are the nodes $raf$ and $jnk$. There, the very good fit of the copulas in the $\mathbf{M^{gauss}C^{par}}$ model exceeds the bad fit of the marginals compared to the $\mathbf{M^{pnp}C^{par}}$ model and the $\mathbf{M^{gauss}C^{par}}$ model and therefore has the best $\text{AIC}_F$ of the three models. However, looking at the overall fit using Gaussian mixture margins or kernel densities margins in the model together with parametric copulas still results in a better $\text{AIC}_F$ than the reference model, i.e., the Linear Gaussian Bayesian Network. If we instead assume Gaussian margins, the resulting model ranges almost on the same level as the LGBN.

When restricting only on Gaussian copulas the node *pka* still has a 10% better $\text{AIC}_F$ than the LGBN when using Gaussian mixture margins or kernel density margins. While this restriction compared to the models with parametric copulas has a moderate impact on most nodes in these two models, for some it does almost not have any impact at all, e.g. for the nodes *plc* and *erk*. On the contrary the restriction has the biggest impact on the nodes *mek* and *akt*. Additionally we observe that the nodes $raf$ and $jnk$ are again the only nodes which are modeled better when using Gaussian margins than when using

Gaussian mixture margins or kernel density margins. Altogether, when using Gaussian mixture margins or kernel density margins the models with only Gaussian copulas still have a slightly better fit than the LGBN.

Looking at the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model we observe that the $\text{AIC}_F$ is almost equal to the fit of the LGBN for all nodes except *akt*. Knowing that both models describe a conditional Gaussian distribution this is what we would expect. For the node *akt* it surprises that the LGBN, even though it was fitted optimizing the log-likelihood, has a better $\text{AIC}_F$ than the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model. This is due to two reasons. First, optimizing the log-likelihood instead of the $\text{AIC}_F$ for the Linear Gaussian Bayesian Network does not have a big influence. Second the parameters of the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model are not modeled in one step but two. In the first step we model the margins, which define the mean and variance of the distribution, and in a second step the dependence structure which defines the correlation. This can result in a worse fit than when fitting all parameters in one step.

As we know that both models specify a conditional Gaussian distribution, we can compare the corresponding parameters in Table 4.2 and Table 6.19. There, we can see that they are mostly similar. Note that the conditional normal specified by the Gaussian D-vine does not model the dependencies $pip3 \to plc$, $mek \to erk$ and $pip2 \to pkc$.

In general, it is hard to predict a scheme of how the fit of a node behaves compared to the reference model, when its copula families are being more and more restricted. For most nodes the $\text{AIC}_F$ decreases in similar steps for each added restriction. On the other hand, some nodes stay almost unaffected by the restrictions, e.g., $raf$. For other nodes restricting to parametric has almost no impact but restricting to Gaussian copulas has, e.g., *mek* and *akt*.

If instead we would compare the log-likelihood the results stay the same, i.e., if a node has a better log-likelihood in one model than in another one, then it also has a better $\text{AIC}_F$ in that one. This also holds when looking at the $\text{BIC}_F$ apart from when comparing a model with kernel density margins to a model with Gaussian mixture margins on the same set of copula. There, in some cases a node has a better $\text{BIC}_F$ when using Gaussian mixture margins in the model.

In summary, modeling the Sachs data set using a D-vine copula regression on the consent graph using Gaussian mixture margins or kernel density margins provides a significant improvement compared to the Linear Gaussian Bayesian Network. This hold especially when allowing for non-parametric and parametric or only parametric copulas. Allowing only for Gaussian copulas ends up with a similar result as the LGBN with better fits for some nodes in both models. Using Gaussian margins takes away most of the advantages of the copulas models, such that only allowing for parametric and non-parametric copulas

in the D-vine regression gives a reason to use a copula approach instead of the Linear Gaussian Bayesian Network. Already when using only parametric copulas the D-vine model with Gaussian margins and the LGBN have a similar goodness of fit. Restricting to only Gaussian copulas results in a conditional Gaussian distribution and therefore in almost exactly the same fit as the Linear Gaussian Bayesian Network.

# Chapter 8

# Simulation Based Comparison

A second approach to compare the models is to compare the strength of each model to recreate the Sachs data set. For this we selected four of the ten models. The four models we want to compare more closely are, the two models with parametric and non-parametric copulas with kernel density margins or Gaussian mixture margins $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ and $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$, the model with only Gaussian copulas and Gaussian margins $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ and as a reference model the Linear Gaussian Bayesian Network. From each model we sampled 6161 times by starting in the node *pip*3 and sampling it using its marginal density. We then moved on by following the topological order, in each step simulating the data of the node given the data we have already simulated for its parents.

For the Linear Gaussian Bayesian Network we used that we can express the conditional distribution as a conditional normal distribution which is easy to sample from. To sample from the D-vine models we followed the algorithm from Theorem 16. We then fitted kernel density estimates to this data. If the model fits well we expect similar results as when fitting kernel density estimates to the Sachs data set after deleting all samples that contain a zero value. The kernel density estimates fitted to the Sachs data set can be found in Section 5.3.

The kernel density estimates fitted to the simulated data sets are displayed in Figure 8.1. There, we observe very similar results between the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model. The same holds for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model and the LGBN where only notable differences for the node *akt* are present. Note that Figure 8.1 does not display any indications of the goodness of fit of the dependence structure, but only information about the fit of the marginals. The reason is that we are not looking at each sample point for each node but instead at the whole set of 6161 samples. In particular for the marginals we already know that kernel density margins and Gaussian mixture margins outperform Gaussian margins and have analyzed their fit in detail in Section 5.4.

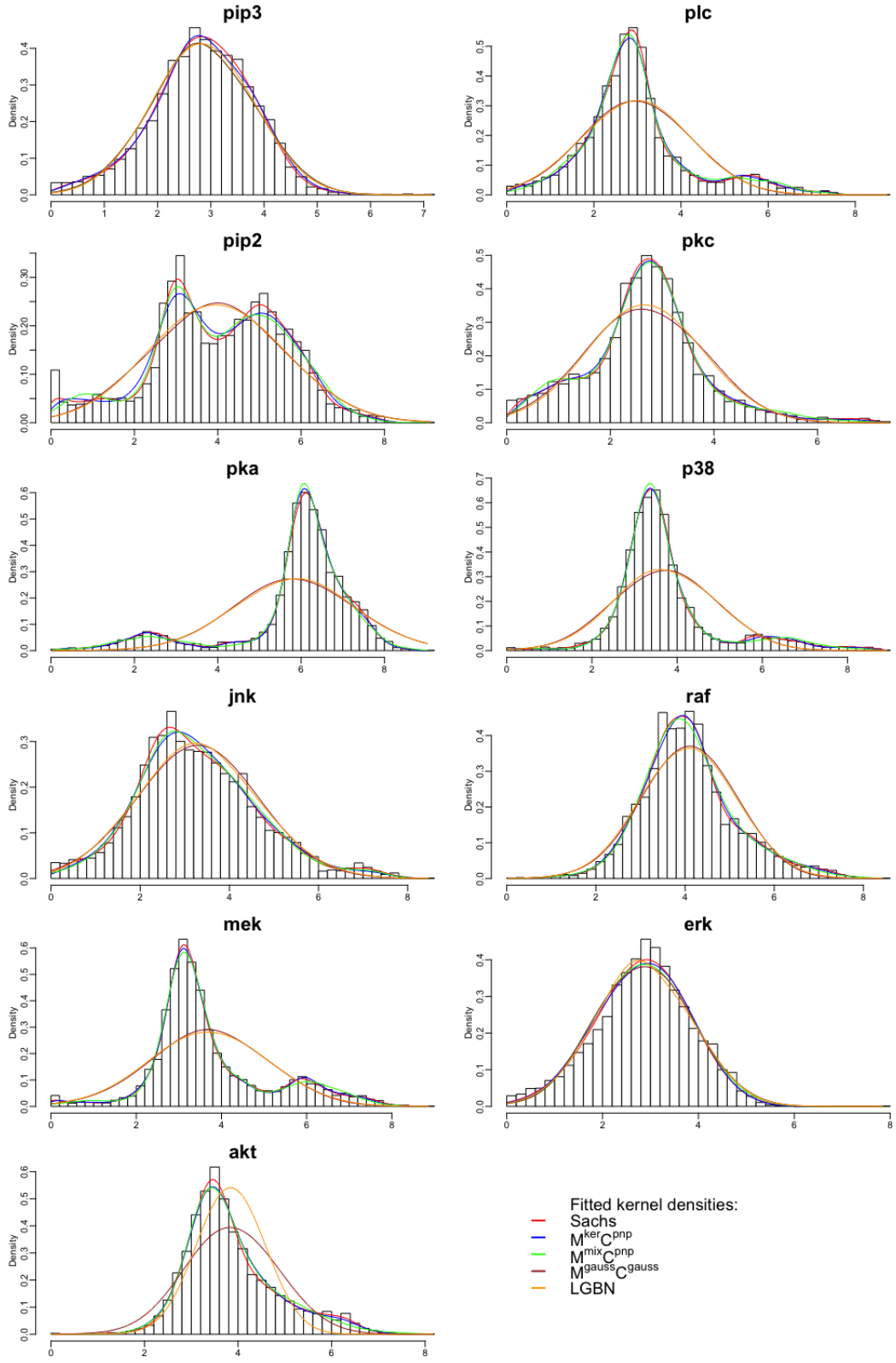



Figure 8.1: Marginal histograms of the Sachs data set of each of the eleven nodes in topological order after samples containing a zero value have been deleted and kernel density estimates have been fitted to the simulated data from the different models

What we can instead do, is to compare the pairs plots of the simulated data sets with the pairs plots of the Sachs data set. The results for the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model are displayed in Figure 8.2 and the results for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model and the Linear Gaussian Bayesian Network in Figure 8.3. Note that the order of the variables in the upper panel is the same as in the lower panel so the plots are directly comparable and do not need to be reflected.

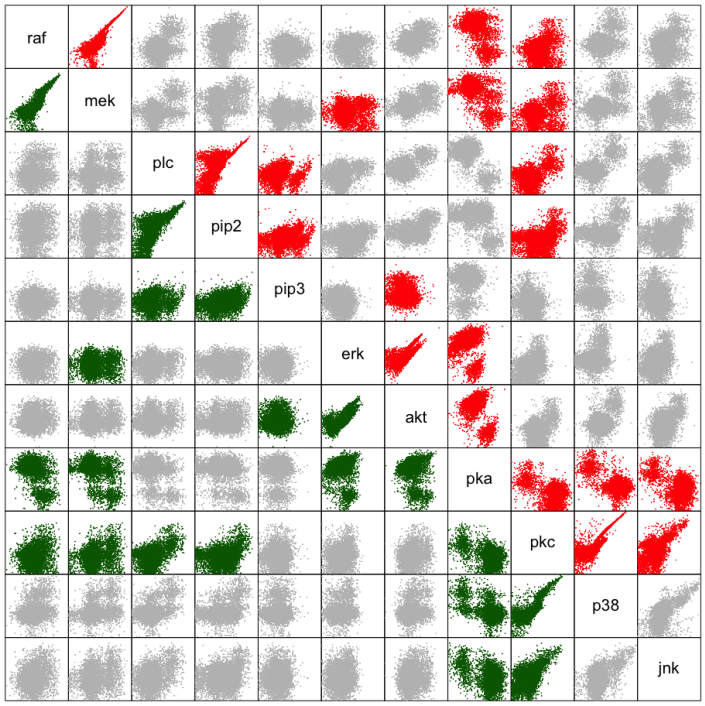
Again, we observe almost indistinguishable results comparing the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model. Therefore, we can compare the pairs plots of both models together with the pairs plots of the Sachs data set. Looking at the pairs plots directly modeled, i.e., the nodes between which a modeled edge exists, colored in green in Figure 8.2, we derive that for all plots at least the general shape agrees between the simulated data and the Sachs data. This even holds for the more complicated looking shapes, e.g., between the node pairs $raf \leftrightarrow mek$ or $plc \leftrightarrow pip2$ which are very well replicated. Only small dissimilarities can be observed for example between the nodes $pkc \leftrightarrow p38$ or $akt \leftrightarrow pka$.

Now, on the other hand, comparing the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model and the LGBN in Figure 8.3 we observe again many similarities. Only two dependencies, $erk \leftrightarrow akt$ and $akt \leftrightarrow pka$, show visible differences. Nevertheless, comparing the pairs plots of both models with the ones in the Sachs data set we observe almost no similarities. Hence, again the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model outperform the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model and the Linear Gaussian Bayesian Network.

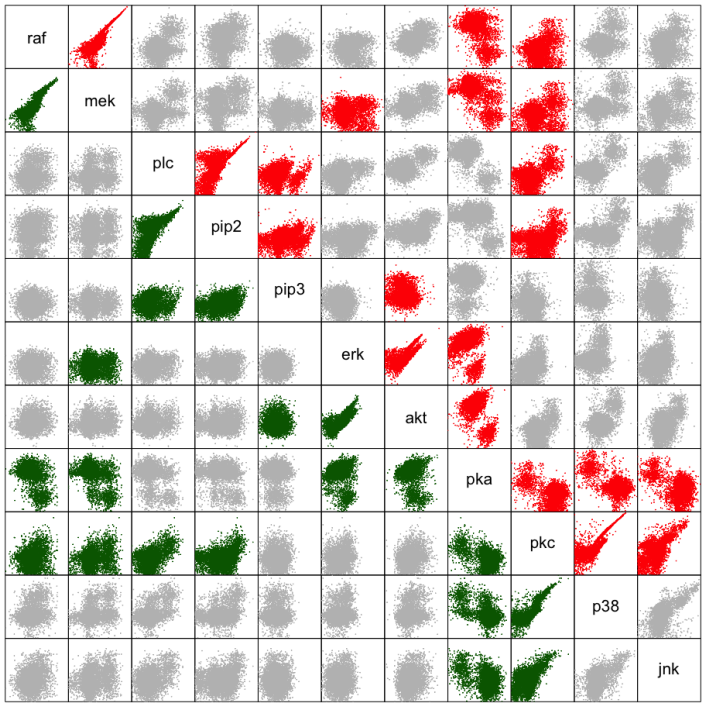
Further, for a suitable model one would expect that in the simulated data set samples with high/low values in several nodes together appear in a similar frequency as in the original data set. To see if this is the case one can sum up over the data of several nodes, in both the model one wants to check and the Sachs data set, and analyze the histograms of the sums. If they are similar, this is an indication for a good fit of the model.

In both, the D-vine models and the LGBN, we modeled a set of 10 conditional densities on different nodes. In the following, we first sum according to each of these, i.e the node and the set of its parents, giving us 10 different sums to compare. Note that this time we strictly define the parents over the edges in the DAG and not by whether a specific dependency is actually modeled. This is important for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model as there dependencies are missing.

Looking at the results in Figure 8.4 we can again observe similar results for both the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model. For both models all histograms are very close to the ones calculated on the Sachs data set.

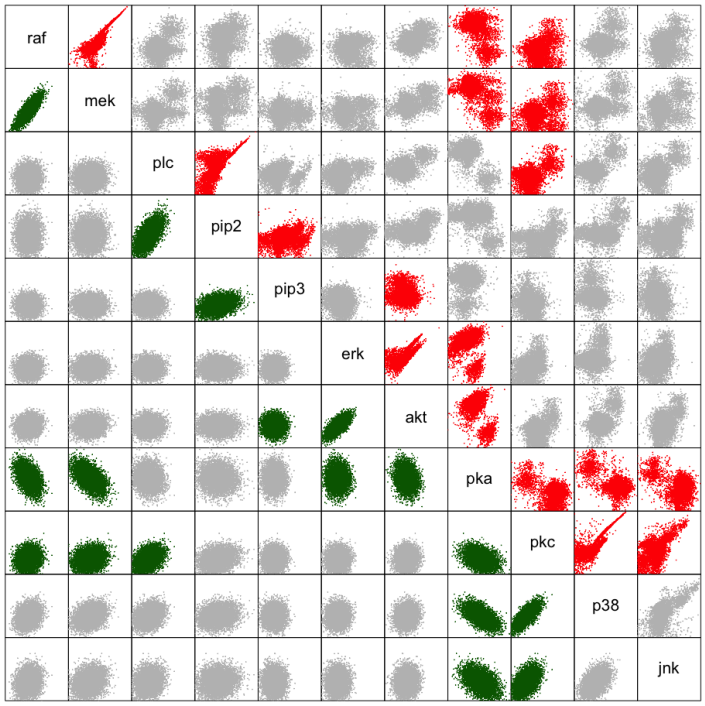



(a) $\mathbf{M^{ker}C^{pnp}}$

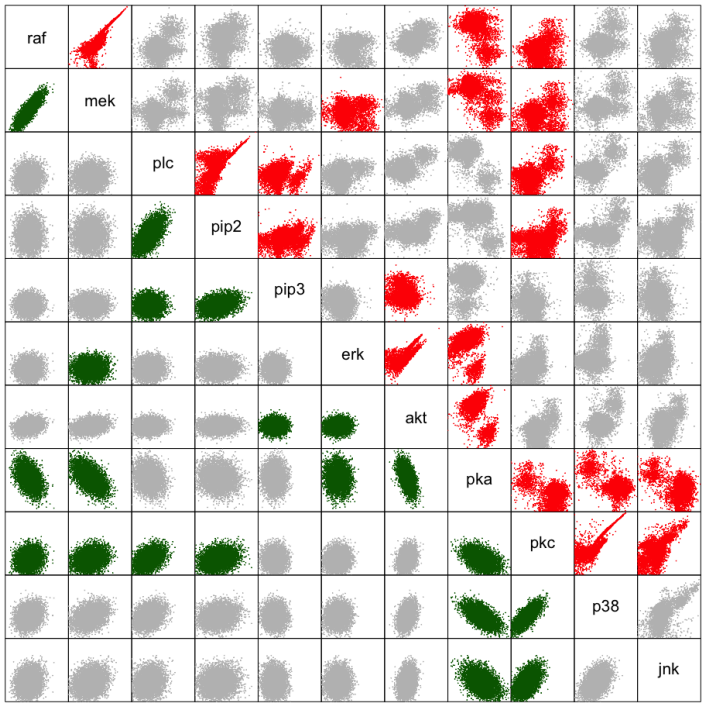



(b) $\mathbf{M^{mix}C^{pnp}}$

Figure 8.2: Pairs plots for each pair of nodes of the respective model in the lower panel and of the Sachs data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel




(a) $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$




(b) Linear Gaussian Bayesian Network

Figure 8.3: Pairs plots for each pair of nodes of the respective model in the lower panel and of the Sachs data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel

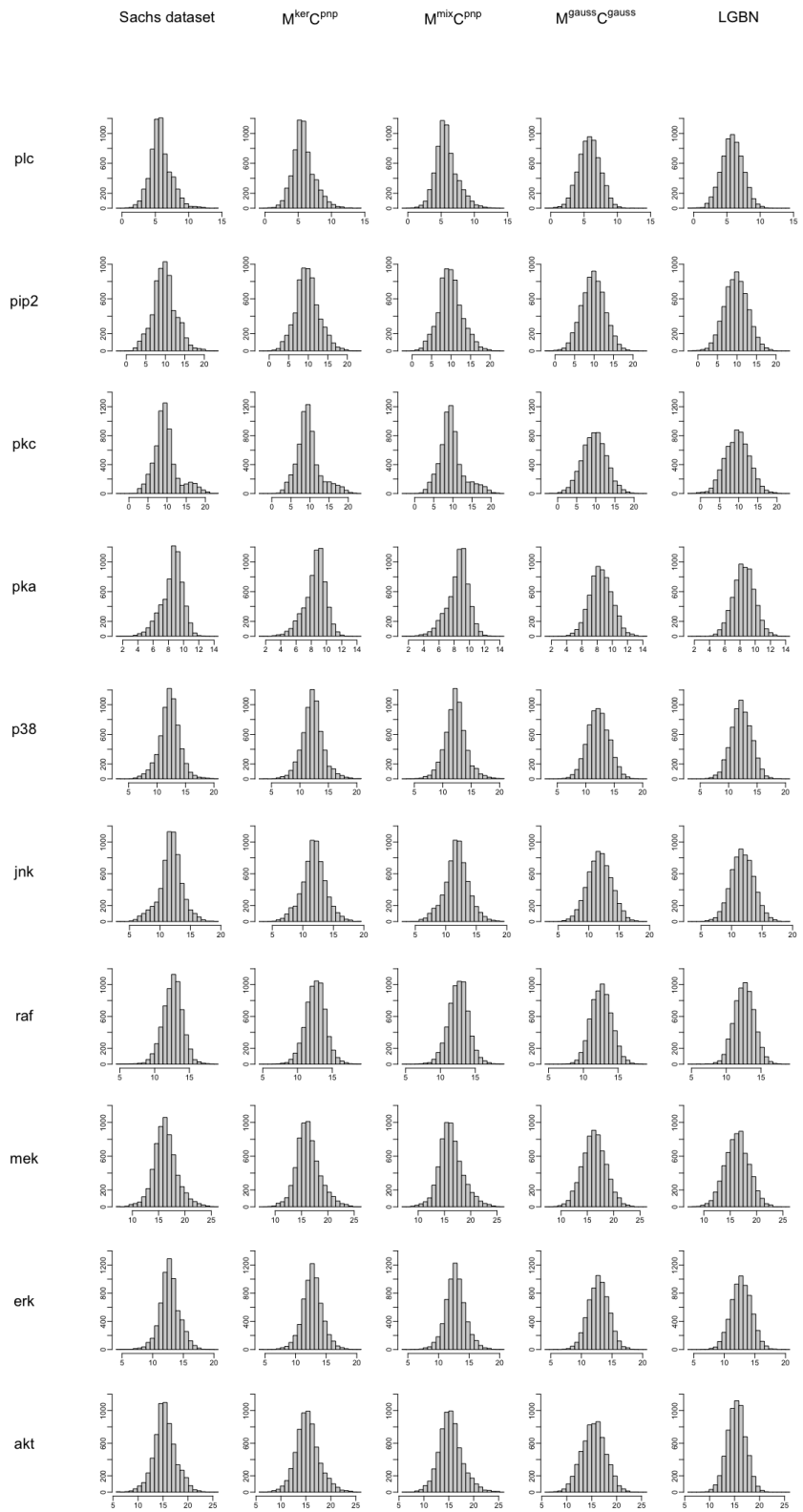



Figure 8.4: Histograms of the sum of each node on the set of its parents according to the consent DAG in the Sachs data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}}$ model, $\mathbf{M^{mix}C^{pnp}}$ model, $\mathbf{M^{gauss}C^{gauss}}$ model and the Linear Gaussian Bayesian Network

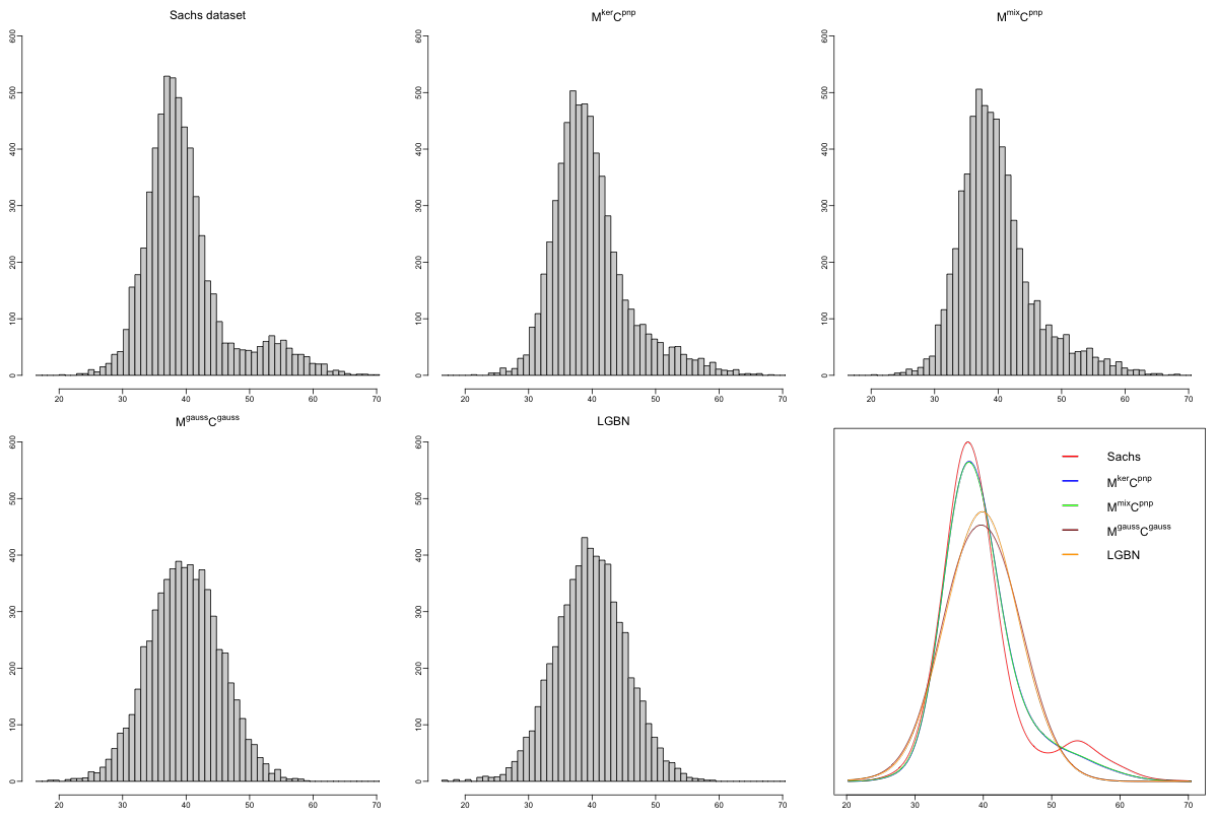



Figure 8.5: Histograms of the sum over all nodes in the Sachs data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}}$ model, $\mathbf{M^{mix}C^{pnp}}$ model, $\mathbf{M^{gauss}C^{gauss}}$ model and the Linear Gaussian Bayesian Network in the first five plots and fitted kernel density estimates to each of the data sets in the sixth plot.

The same holds for the histograms of the $\mathbf{M^{gauss}C^{gauss}}$ model and the Linear Gaussian Bayesian Network which are very similar for all nodes except *akt*. However, it seems that both models have problems to recreate the correct height of the peaks of the Sachs data set, e.g., for the nodes *pkc*, *mek* and *akt*. For the node *akt* the $\mathbf{M^{gauss}C^{gauss}}$ model struggles more than the LGBN. The histograms of the samples from the $\mathbf{M^{gauss}C^{gauss}}$ model and the LGBN therefore again seem way less similar to the ones from the Sachs data set than the ones from the samples from the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model.

As we are not only interested in how good the modeled conditional densities fit compared to the Sachs data set but also in how good the whole model fits we sum up over the data of all nodes and compare the resulting histograms. The results are displayed in Figure 8.5. We observe that all of the four models have a problem recreating the second peak, at about 55, in the Sachs data set. Even more all of them struggle to achieve the height of the first peak, at around 38, of the Sachs data set. Again, the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model are significantly better than the $\mathbf{M^{gauss}C^{gauss}}$ model and the LGBN.

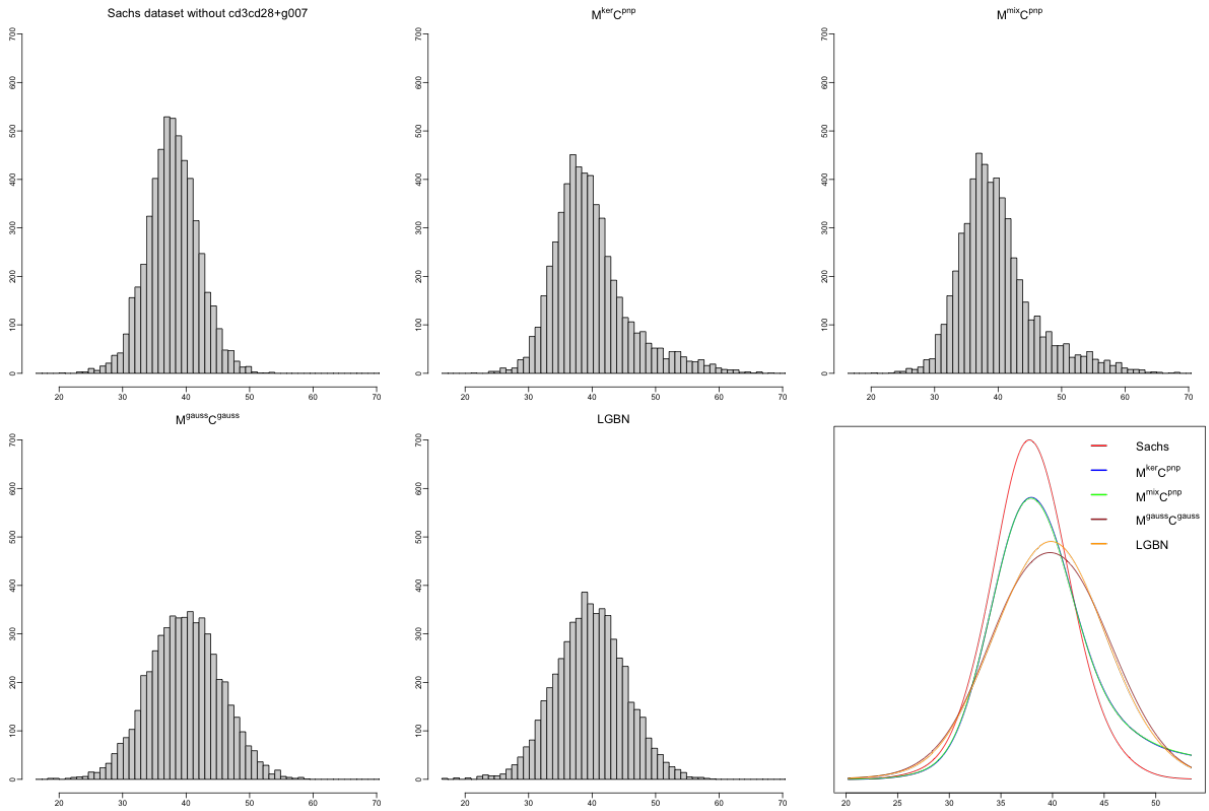



Figure 8.6: Histograms of the sum over all nodes in the Sachs data set without the samples from the $cd3cd28{+}g007$ experiment and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}}$ model, $\mathbf{M^{mix}C^{pnp}}$ model, $\mathbf{M^{gauss}C^{gauss}}$ model and the Linear Gaussian Bayesian Network in the first five plots and fitted kernel density estimates to each of the data sets in the sixth plot

Searching for reasons why the fitted models have a problems with the second peak we go back to the normalized pairs plots, as for example given in Figure 5.6. There, we observed that low values of the node *pka* coincide with high values in nearly all other nodes. Going back to the histogram of the Sachs data set distinguished by the underlying experiment in Figure 3.1, we observed that the node *pka* has a small first peak where the data almost only originates from the $cd3cd28 + g007$ experiment. Combining this, lets us assume that this second peak, which is present in the histogram of the sum over all nodes, comes exactly from the samples from the $cd3cd28 + g007$ experiment.

To confirm the assumption we again calculate the sum over all nodes in the Sachs data set but ignore the samples from the $cd3cd28 + g007$ experiment. This leaves us with 5466 samples. Taking samples of the same size from the four models and calculating the sum over all nodes, we can again compare the resulting histograms. As Figure 8.6 shows the second peak in the Sachs data set is gone. Nevertheless, the problem that the fitted models are not able to reach the height of the first peak is still present.

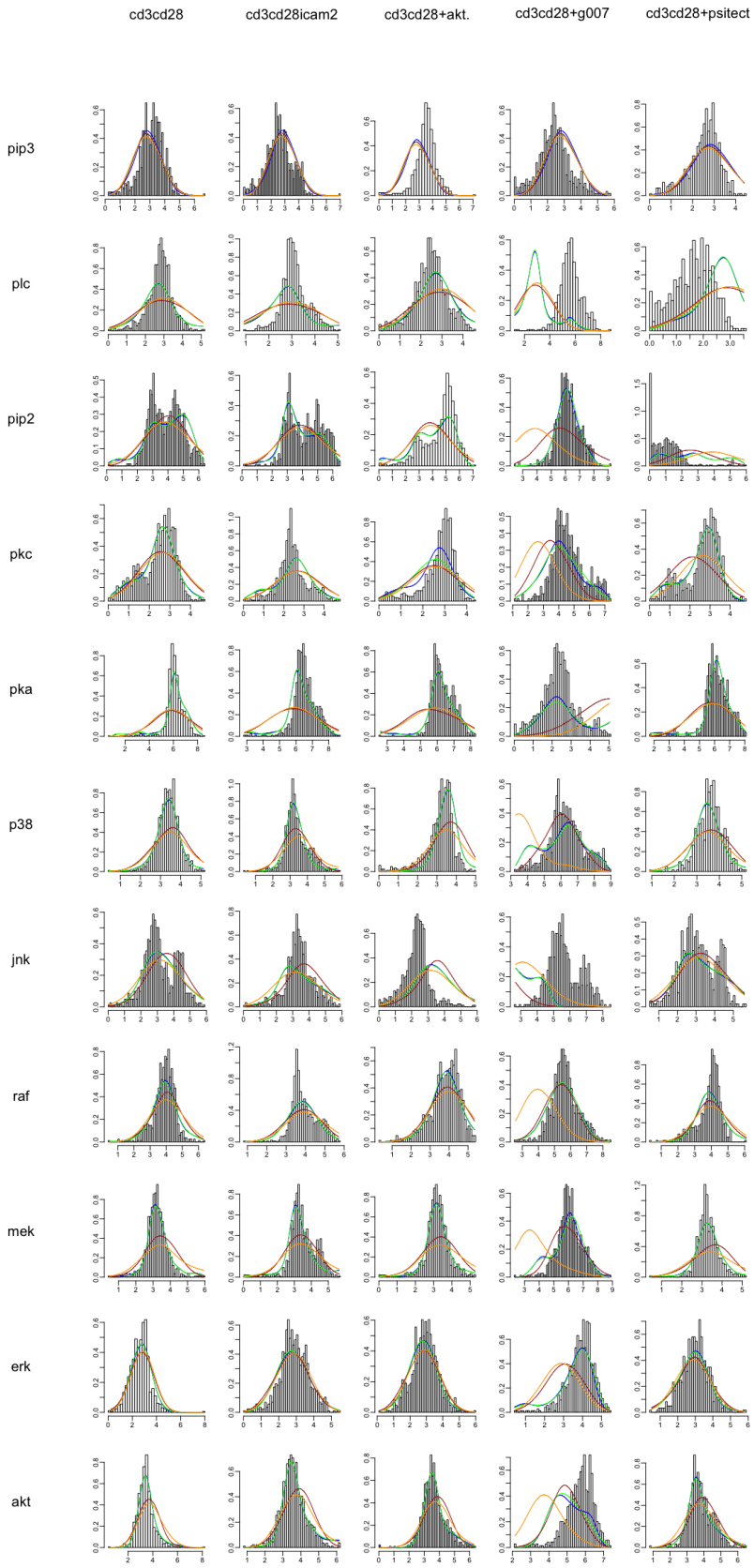



Figure 8.7: Marginal histograms of each node split by experiments and kernel density estimates fitted to the data simulated from the conditional densities of the different models given the parents as in the data set from the specific experiment. Blue: $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$. Green: $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$. Brown: $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$. Orange: LGBN

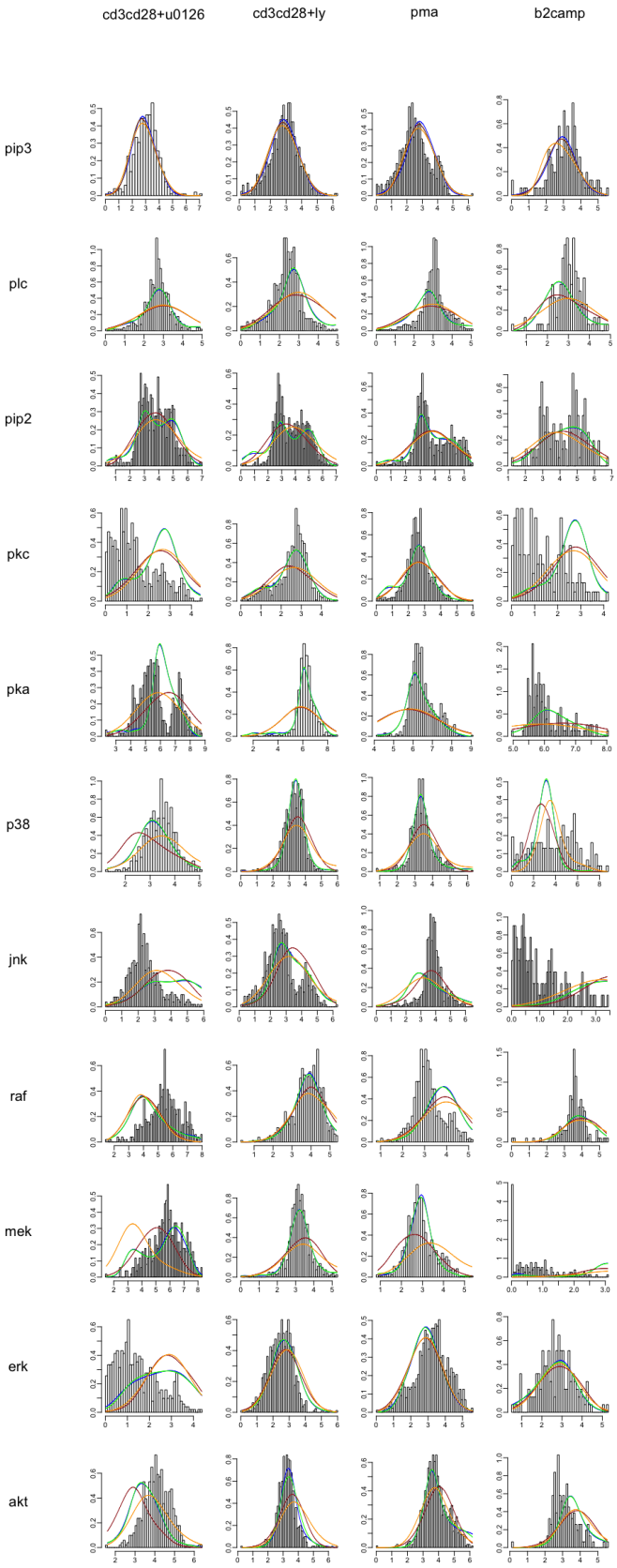



Figure 8.8: Marginal histograms of each node split by experiments and kernel density estimates fitted to the data simulated from the conditional densities of the different models given the parents as in the data set from the specific experiment. Blue: $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$. Green: $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$. Brown: $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$. Orange: LGBN.

Hence, it looks like the models are not able to recreate the data from the $cd3cd28 + g007$ experiment properly. Therefore, it would be interesting to examine if this also holds for other experiments. For this we partition the Sachs data set according to the nine experiments. On each of these smaller data sets we then sample from each conditional density of the four models, given the actual values for the parents and again fit kernel density estimates to the resulting output. We can then compare the fitted kernel density estimates with the histograms of the node of the different experiment to check if we can find any pattern throughout the nodes and data sets.

First, we observe something we already mentioned while performing the data exploration, no single node seems to be identically distributed over all experiments. The one closest to an identical distribution is the node $pip3$. Here, only for the $cd3cd28 + psitect$ experiment fewer values lie in the right tail as in the other experiments. Looking at the different experiments we observe that for the $cd3cd28 + g007$ ($plc$), $cd3cd28 + psitect$ ($pip2$), $cd3cd28 + u0126$ ($erk$) and the $b2camp$ ($mek$) experiment at least for one node none of the fitted models is able to properly model its distribution. For the $b2camp$ experiment one needs to take into account that after deleting the zero rows it only consists of 155 samples and therefore here the analysis is less meaningful. Looking at the individual nodes it seems that for the nodes $pip3$, $jnk$, $raf$ and $erk$ the differences between the four models are rather small. On the other hand, for the other nodes the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model in general result in way better fits than the $\mathbf{M^{gauss}C^{gauss}}$ model or the Linear Gaussian Bayesian Network.

# Chapter 9

# Conditional Simulation Based Comparison

In the last chapter we have simulated data from the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model, the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model, the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model and the Linear Gaussian Bayesian Network and compared the results to the Sachs data set. Next, we want to examine how the conditional densities of these models behave and if we can see differences between them.

If a node only has one parent in the Sachs data set, we chose the 10%, 50% and 90% quantile of the empirical distribution of the parent node as conditioning values. This was done for the nodes *plc* and *pka* and can be found in Figure 9.1. As there is no exact equivalent of quantiles in two or three dimensional distributions a different approach had to be used for all other nodes.

If a node has two parents, we fitted a two-dimensional kernel density estimate to the joint distribution of the parent nodes using the Sachs data. We then manually choose one point close to the mode of the fitted density, two in the tails and two more which are neither close to the mode nor in the tails. This is illustrated in Figure 9.2. Note that while six nodes have two parents, $raf$, $pip2$, $erk$, $pkc$, $p38$ and $jnk$, three of them, $raf$, $p38$ and $jnk$, have the same set of parents, i.e., $pka$ and $pkc$, so we only obtain four plots. A similar approach was performed when a node has three parents, i.e., for the nodes $mek$ and $akt$. Then, a three-dimensional kernel density estimate was fitted to the joint data of the parents and the same procedure was used to identify five points to condition on. The results are given in Figure 9.3.

The list of points on which we condition on can be found in Table 9.1. In the following we will headline the points which we chose close to the the mode of the fitted distribution as "Mode", the points in the tails as "Tail" and the points between them as "Middle". If a

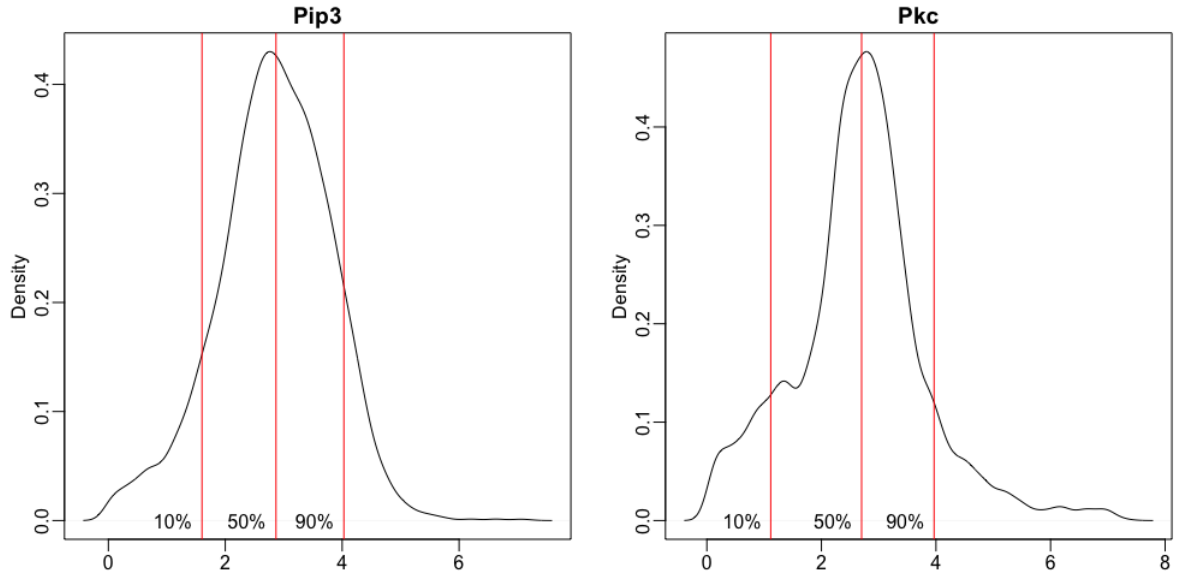



Figure 9.1: Empirical density of the nodes *pip*3 and *pkc* with horizontal lines at the 10%, 50% and 90% quantile

| Node | Parent | Mode | Middle | | Tail | |
|---|---|---|---|---|---|---|
| raf, p38, jnk | pka | 6.20 | 6.20 | 7.00 | 5.50 | 2.50 |
| | pkc | 2.50 | 2.00 | 3.00 | 0.75 | 4.50 |
| mek | raf | 4.52 | 3.86 | 2.78 | 4.52 | 7.15 |
| | pka | 5.79 | 6.92 | 6.76 | 2.96 | 4.61 |
| | pkc | 2.50 | 3.51 | 2.73 | 2.62 | 3.14 |
| plc | pip3 | 1.60 | | | 2.87 | 4.03 |
| pip2 | plc | 2.75 | 2.75 | 2.40 | 3.75 | 5.50 |
| | pip3 | 2.75 | 1.50 | 3.50 | 4.20 | 2.00 |
| erk | mek | 3.10 | 3.70 | 2.90 | 6.25 | 2.00 |
| | pka | 6.10 | 7.10 | 6.40 | 2.50 | 6.30 |
| akt | erk | 2.63 | 3.08 | 3.38 | 4.49 | 5.06 |
| | pka | 6.00 | 6.24 | 7.47 | 2.80 | 5.81 |
| | pip3 | 3.19 | 1.20 | 3.45 | 4.76 | 2.54 |
| pka | pkc | 1.12 | | | 2.70 | 3.97 |
| pkc | plc | 3.00 | 2.60 | 2.60 | 5.50 | 2.00 |
| | pip2 | 3.00 | 2.60 | 4.80 | 6.00 | 1.25 |

Table 9.1: Parent nodes used for conditioning of each density

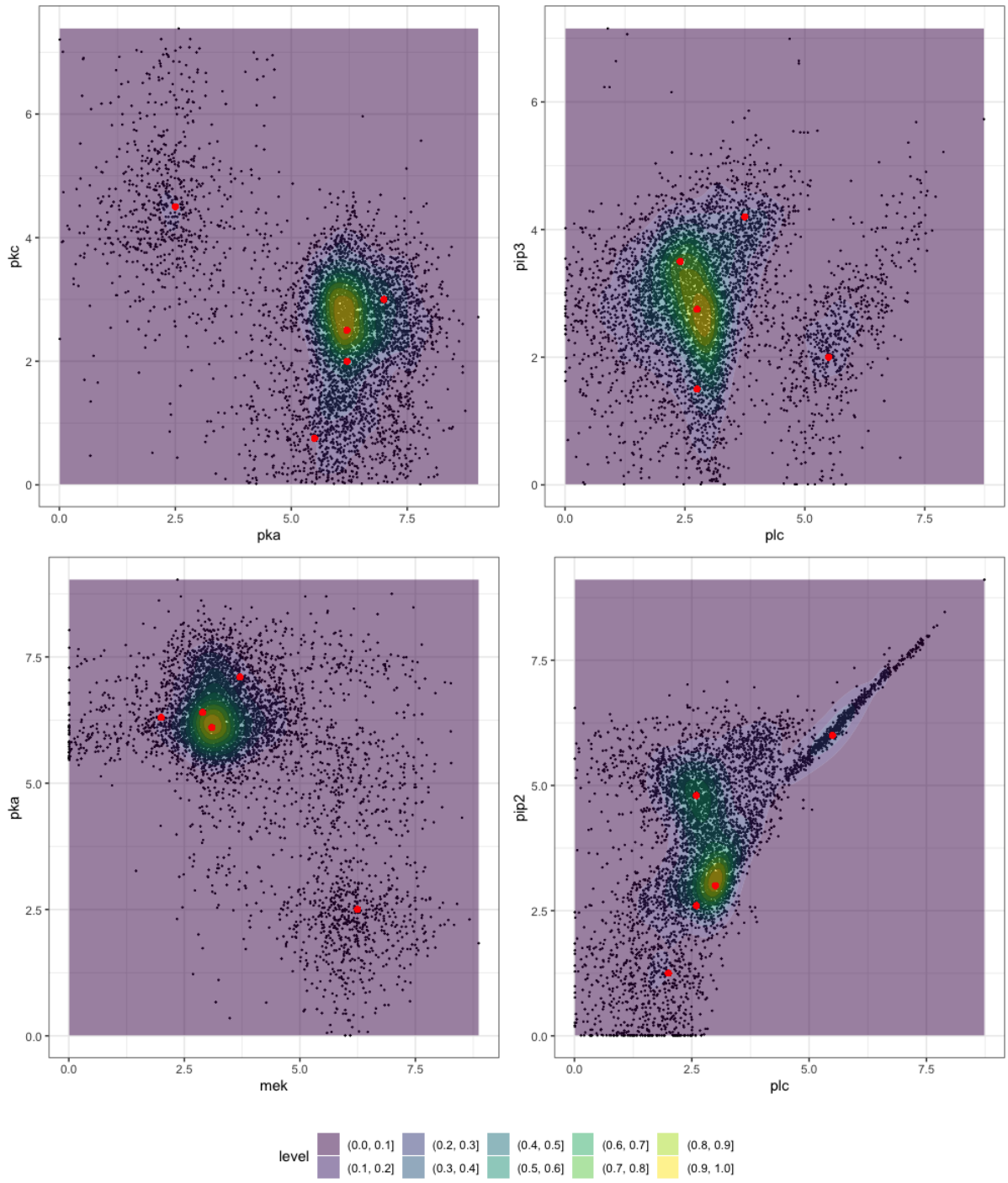



Figure 9.2: Contour plots of the two-dimensional kernel density estimates of the node pairs $pka \leftrightarrow pkc$, $plc \leftrightarrow pip3$, $mek \leftrightarrow pka$ and $plc \leftrightarrow pip2$. Red points correspond to the points chosen for conditioning

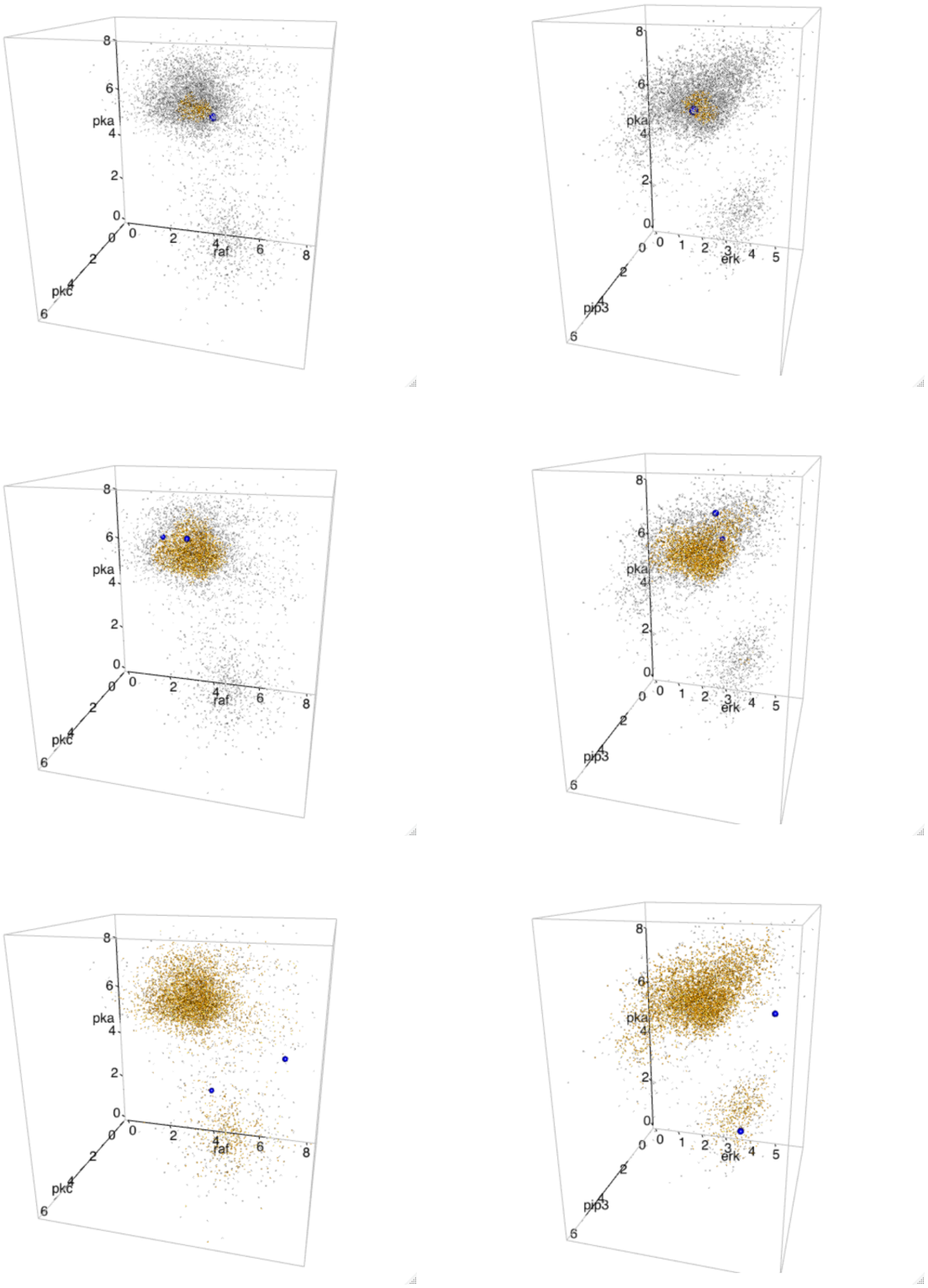



Figure 9.3: Scatter plots of the parents of the nodes *mek* (left) and *akt* (right). Points in yellow are points within the 10%, 50% and 90% quantile of the fitted three-dimensional kernel density estimates (from top to bottom). Blue points are the points chosen for conditioning

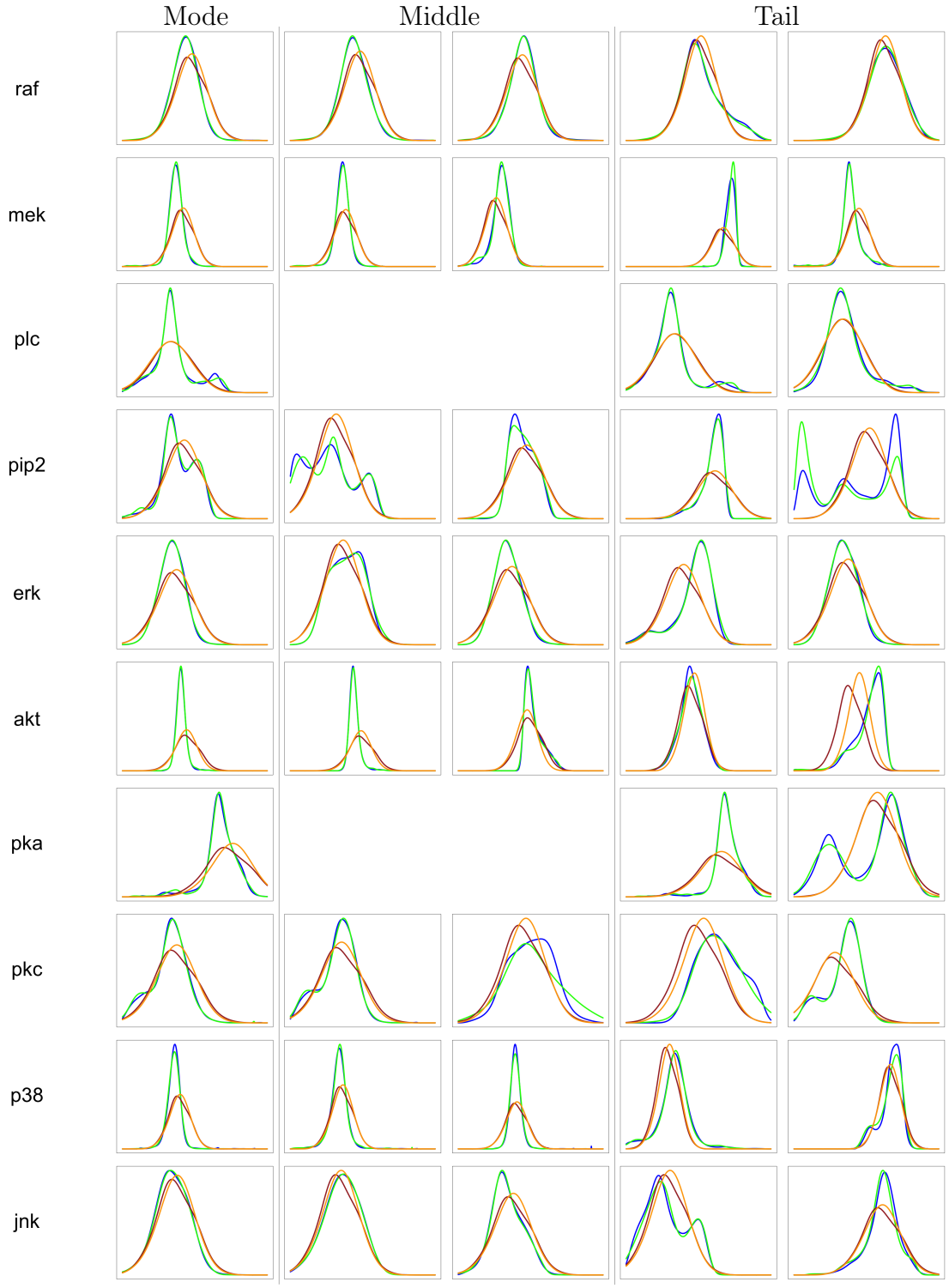



Figure 9.4: Density plots of the kernel density estimates fitted to the simulated data from the different models conditioned on the nodes in Table 9.1 in the same order. Blue: $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$. Green: $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$. Brown: $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$. Orange: LGBN

node only has one parent, we will assign the points from the 10% and the 90% quantile to "Tail" and the point from the 50% quantile to "Mode". Note that as in the $\mathbf{M^{gauss}C^{gauss}}$ model not all dependencies have been modeled, in this model sometimes the number of parents differs from the other models.

As in Chapter 8 we sampled 6161 times from each node of the model given the parent nodes and then fitted kernel density estimates to this sample. The resulting density plots are displayed in Figure 9.4. There, we observe that when comparing the models $\mathbf{M^{ker}C^{pnp}}$ and $\mathbf{M^{mix}C^{pnp}}$, apart from the nodes *pip*2 and *pkc*, only small differences in the general shape of the density plots can be found. More precisely, the differences are normally limited to the conditional density of one of the two models having a higher peak. This holds for the plots of all areas.

In Chapter 8 we have seen that the $\mathbf{M^{gauss}C^{gauss}}$ model and the Linear Gaussian Bayesian Network display similar results when simulating from both. This is here again the case. On the other hand, for all five points we cannot identify a single node for which the $\mathbf{M^{gauss}C^{gauss}}$ model results in similar density plots as the other two D-vine models. Especially, it holds that for at least one of the two points chosen in the tail, the plot barely agrees with the ones from the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model.

As we cannot properly compare the conditional densities to the ones of the Sachs data set it is hard to make a statement about the quality of the plots in Figure 9.4. We can say however that the $\mathbf{M^{ker}C^{pnp}}$ model or $\mathbf{M^{mix}C^{pnp}}$ model are able to recreate more flexible conditional densities than the other two models which is generally a big plus. Later, we will also be able to compare the density plots in this chapter to the ones from the models fitted on partitions of the Sachs data set.

# Chapter 10

# Partitioning of the Sachs Dataset

We already observed in the data exploration in Chapter 3 that the assumption on having identically distributed data throughout the Sachs data set does not seem to be correctly fulfilled. Instead the distribution of the data depends on the experiment through which the data is generated. When we compared the fitted models in Chapter 8, we have seen that the models had problems recreating some parts of the data set and conjectured that this is due to the violation of this requirement. To investigate how the results change if we actually have independent and identically distributed data we will in the following repeat our analysis but this time only consider parts of the Sachs data set.

The crucial part is considering a sufficiently large data set, to be able to properly fit copula models, while still having identically distributed data throughout the data set. It is clear that within the data of any single experiment assuming that the data is identically distributed is reasonable. On the other hand, we observed in Table 3.2 that, after deleting all data points with at least one zero value, about half of the experiments consist of approximately 800 samples, the rest of considerable less. Therefore, we also want to investigate if we can identify pairs of experiments within which we can assume that the data is identically distributed.

To this end, we fitted kernel density estimates to the data of the nine experiments for each node. We assume that within two experiments the data is identically distributed if the fitted kernel density estimates are similar for all eleven nodes in both data sets. We can observe in Figure 10.1 that it is not possible to find such a pair for the data sets from the $b2camp$, $cd3cd28 + g007$, $cd3cd28 + u0126$ and $cd3cd28 + psitect$ experiment. For the $b2camp$ experiment we cannot find a similar distribution for example for the nodes $mek$ or $p38$. For the $cd3cd28 + g007$ experiment, among other things, the distribution of the nodes $plc$ or $pka$ is unique. For the data from the $cd3cd28 + u0126$ experiment no other fitted kernel density estimate of another experiment has a similar distribution for

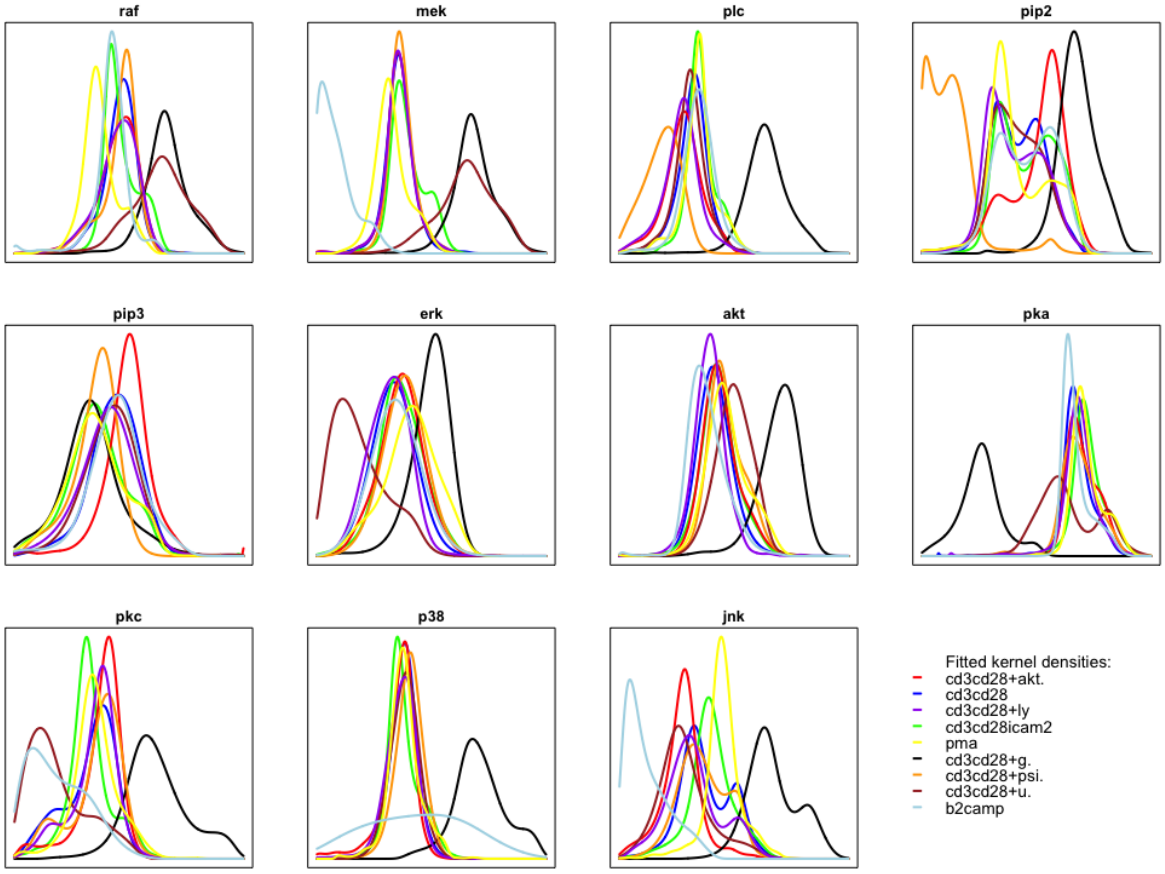



Figure 10.1: Kernel density estimates for each node in the nine experiments the Sachs data set consists of

the node $erk$. Further, for the data from the $cd3cd28+psitect$ experiment the distribution of $pip2$ is different to the distribution of $pip2$ for all other experiments. As all of the four described experiments consist of less than 700 data points (155, 695, 508 604) it does not seem reasonable to fit copula models to one of these data sets alone.

Instead we will focus on the remaining five experiments. For a better visibility Figure 10.2 shows only the kernel density estimates of these experiments. While all of them are mostly similar for the nodes $erk$, $akt$, $pka$ and $p38$, it seems that for the other nodes it is possible to divide them into **three groups**. First considering the data from the $cd3cd28+aktinhib$ experiment we observe that the fitted kernel density estimates look particularly different to the ones from all other data sets for the nodes $pip2$, $pip3$ and $jnk$. However, the $cd3cd28 + aktinhib$ data set consists of a reasonable large number of sample data points (845). We therefore say that **Group 1** consists only of the data from this experiment.

Next, we can observe that the data from the $cd3cd28$ and $cd3cd28+ly$ experiments seems to be similar. This is especially visible when looking at the nodes $pip2$, $pkc$ and $jnk$ where

the kernel density estimates fitted to both data sets each time have a similar shape which is different to all others. Hence, we say these two experiments belong to **Group 2**.

Something similar can be observed between the remaining data sets of the $cd3cd28icam2$ and $pma$ experiments. Here, the shape of the fitted kernel density estimates are especially similar, and different to the ones of all other experiments, for the nodes $raf$, $mek$, $plc$ and $pip3$. Therefore, these two experiments belong to **Group 3**. Now, partitioning the Sachs data set into these **three groups** leaves us with data sets consisting of 845, 1618 and 1736 data points.

To investigate further whether the data within **Group 2** and **Group 3** is identically distributed we check if the data in the experiments the groups consist of is similar dependent. For this we analyze the estimated partial correlations, given all other nodes, calculated for the data of each experiment. We hereby only consider pairs of nodes which are connected by an edge in the DAG and therefore can be directly modeled.

Looking at the results in Table 10.1 we observe that for the data in **Group 2**, i.e., from the $cd3cd28$ and $cd3cd28 + ly$ experiments, almost all connected pairs of nodes have a similar partial correlation in both experiments. The biggest differences are between the node pairs $jnk \leftrightarrow pkc$ and $pip2 \leftrightarrow plc$ which are the only ones higher than 0.1. Hence, no clear indications against the assumption of identically distributed data in **Group 2** were found.

| | **Group 1** | **Group 2** | | **Group 3** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dep. | cd3cd28+akt. | cd3cd28 | cd3cd28+ly | cd3cd28icam2 | pma | cd3cd28+g. | cd3cd28+psi. | cd3cd28+u. | b2camp |
| raf-pka | 0.06 | 0.03 | -0.03 | -0.05 | 0.03 | 0.06 | -0.03 | -0.00 | -0.07 |
| raf-pkc | 0.06 | 0.01 | -0.03 | 0.05 | 0.04 | 0.01 | 0.01 | 0.02 | -0.06 |
| mek-raf | 0.65 | 0.69 | 0.66 | 0.88 | 0.65 | 0.99 | 0.57 | 0.99 | 0.11 |
| mek-pka | -0.00 | -0.03 | 0.03 | 0.06 | -0.02 | -0.06 | -0.00 | 0.01 | 0.02 |
| mek-pkc | -0.04 | -0.03 | 0.02 | -0.06 | 0.00 | -0.01 | -0.05 | -0.02 | 0.14 |
| plc-pip3 | 0.32 | 0.02 | -0.02 | 0.33 | 0.11 | -0.16 | 0.11 | 0.01 | 0.22 |
| pip2-plc | -0.00 | 0.13 | 0.01 | 0.11 | 0.37 | 0.98 | 0.08 | 0.17 | 0.04 |
| pip2-pip3 | 0.41 | 0.35 | 0.40 | 0.60 | 0.61 | 0.26 | 0.22 | 0.47 | 0.67 |
| erk-mek | 0.03 | 0.08 | 0.06 | -0.01 | 0.02 | 0.04 | 0.01 | -0.01 | 0.02 |
| erk-pka | -0.42 | 0.02 | -0.07 | 0.12 | -0.25 | 0.06 | -0.09 | 0.67 | -0.23 |
| akt-erk | 0.87 | 0.79 | 0.73 | 0.83 | 0.88 | 0.96 | 0.82 | 0.01 | 0.83 |
| akt-pka | 0.60 | 0.23 | 0.31 | 0.25 | 0.50 | -0.02 | 0.34 | 0.44 | 0.55 |
| akt-pip3 | 0.02 | -0.00 | -0.03 | 0.02 | 0.04 | -0.00 | 0.02 | 0.02 | 0.00 |
| pka-pkc | -0.09 | 0.07 | -0.03 | 0.03 | -0.00 | 0.02 | 0.05 | 0.02 | 0.09 |
| pkc-pip2 | 0.07 | -0.04 | -0.01 | -0.05 | 0.07 | -0.00 | -0.06 | 0.07 | -0.07 |
| pkc-plc | -0.01 | -0.01 | -0.02 | 0.02 | -0.02 | 0.01 | 0.02 | 0.13 | 0.04 |
| p38-pka | 0.03 | -0.02 | 0.02 | -0.04 | 0.01 | -0.01 | -0.03 | -0.02 | -0.01 |
| p38-pkc | 0.74 | 0.68 | 0.75 | 0.56 | 0.65 | 0.98 | 0.75 | 0.50 | 0.29 |
| jnk-pka | 0.07 | 0.03 | -0.01 | -0.01 | 0.03 | -0.03 | 0.01 | 0.07 | -0.08 |
| jnk-pkc | -0.05 | -0.34 | -0.53 | 0.43 | 0.26 | 0.44 | -0.31 | 0.28 | 0.21 |

Table 10.1: Estimations of the partial correlations between nodes which are connected by an edge in the DAG, given all other nodes, of the different data sets

Slightly bigger differences can be observed between the experiments in **Group 3**, $cd3cd28icam2$ and $pma$, with the biggest difference being between the nodes $erk$ and $pka$ with an absolute value of 0.37. In total in seven of the twenty observed pairs the difference

between the partial correlation in the two experiments is bigger than 0.1. Therefore, still most of the node pairs have a similar partial correlation in the experiments in **Group 3** and it seems reasonable to assume that the data in this group is identically distributed.

In the following chapters, we will fit four models to each of the groups: a Linear Gaussian Bayesian Network and three kinds of D-vine regressions, where in one we will use kernel density margins for the probability integral transform and allow for parametric and non-parametric copulas, one where we will use Gaussian mixture margins instead and one where we use Gaussian margins and only Gaussian copulas. These are the models we have already analyzed further in Chapter 8 and Chapter 9. We will start with the data set from **Group 1**, then repeat the procedure on the data set from **Group 2** and finally fit the models on the data from **Group 3**.

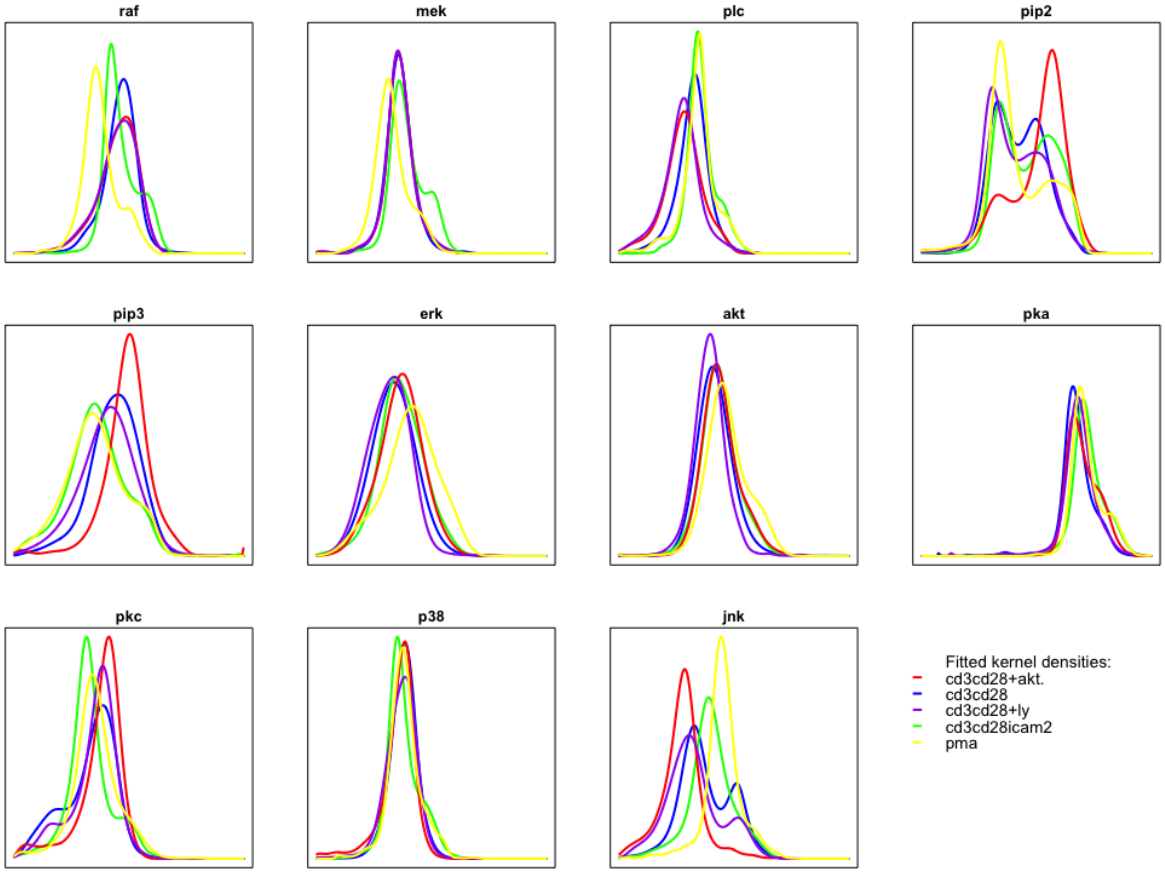



Figure 10.2: Kernel density estimates for each node of the experiments in **Group 1** ($cd3cd28 + aktinhib$), **Group 2** ($cd3cd28$ and $cd3cd28 + ly$ ) and **Group 3** ($cd3cd28icam2$ and $pma$)

# Chapter 11

# Analysis of Group 1: cd3cd28 + aktinhib

In this chapter we will investigate the data of the *cd*3*cd*28 + *aktinhib* experiment using three types of D-vine models and a Linear Gaussian Bayesian Network.

## 11.1 Linear Gaussian Bayesian Network Fitting ($\mathbf{LGBN_{cd3+akt}}$)

Fitting a Linear Gaussian Bayesian Network on the data from the *cd*3*cd*28 + *aktinhib* experiment results in a log-likelihood of **-9189.29**, an $AIC_F$ of **18462.58** and a $BIC_F$ of **18666.37** of the whole model. Detailed results are displayed in Table 11.1.

| Node | Global log-likelihood | # Parameters | $AIC_F$ | $BIC_F$ |
|---|---|---|---|---|
| raf | -969.82 | 4 | 1947.64 | 1966.60 |
| mek | -551.50 | 5 | 1112.99 | 1136.69 |
| plc | -921.77 | 3 | 1849.53 | 1863.75 |
| pip2 | -1249.37 | 4 | 2506.74 | 2525.70 |
| pip3 | -986.90 | 2 | 1977.81 | 1987.29 |
| erk | -968.07 | 4 | 1944.14 | 1963.09 |
| akt | -196.03 | 5 | 402.06 | 425.75 |
| pka | -830.92 | 3 | 1667.84 | 1686.80 |
| pkc | -952.32 | 4 | 1912.64 | 1931.60 |
| p38 | -570.72 | 4 | 1149.45 | 1168.40 |
| jnk | -991.87 | 4 | 1991.74 | 2010.70 |
| $\sum$: | -9189.29 | 42 | 18462.58 | 18666.37 |

Table 11.1: Global log-likelihood, number of parameters, $AIC_F$ and $BIC_F$ of the fitted linear models for each node given the set of its parents in the Linear Gaussian Bayesian Network on the data set from the *cd*3*cd*28 + *aktinhib* experiment

## 11.2 Marginal Fitting ($\mathbf{M_{cd3+akt}}$)

### 11.2.1 Gaussian Margin Fitting ($\mathbf{M^{gauss}_{cd3+akt}}$)

Fitting Gaussian margins to the data from the $cd3cd28 + aktinhib$ experiment results in a log-likelihood of the marginals of **-10690.40**, an $AIC_M$ of **21424.77** and a $BIC_M$ of **21529.04** when assuming independence in the data. Table 11.2 displays detailed results for each node.

| Node | Marginal log-likelihood | $AIC_M$ | $BIC_M$ |
|---|---|---|---|
| raf | -971.98 | 1947.95 | 1957.43 |
| mek | -785.21 | 1574.41 | 1583.89 |
| plc | -979.48 | 1962.97 | 1972.44 |
| pip2 | -1338.108 | 2680.35 | 2689.83 |
| pip3 | -986.90 | 1977.81 | 1987.29 |
| erk | -997.54 | 1999.07 | 2008.55 |
| akt | -912.43 | 1828.85 | 1838.33 |
| pka | -831.22 | 1666.44 | 1675.92 |
| pkc | -954.10 | 1912.21 | 1921.69 |
| p38 | -925.20 | 1854.40 | 1863.88 |
| jnk | -1008.16 | 2020.31 | 2029.79 |
| $\sum$: | -10690.40 | 21424.77 | 21529.04 |

Table 11.2: Marginal log-likelihood, $AIC_M$ and $BIC_M$ of the fitted Gaussian margins $\mathbf{M^{gauss}_{cd3+akt}}$ to the data set from the $cd3cd28 + aktinhib$ experiment

Looking at the plots in the third column of panel (a) in Figure 11.4 we observe that the data does not seem to be approximately uniformly distributed for all nodes except *erk*. This again violates the requirement to fit copulas to the data.

Normalized pairwise contour and scatter plots using the fitted Gaussian margins for the PIT as well as again the histogram of the data on the copula scale are displayed in Figure 11.1.


Figure 11.1: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian margins to the data from the $cd3cd28 + aktinhib$ experiment

### 11.2.2 Gaussian Mixture Margin Fitting ($\mathbf{M^{mix}_{cd3+akt}}$)

Extending the class of marginals to Gaussian mixture margins and fitting these to the data from the $cd3cd28 + aktinhib$ experiment results in a log-likelihood of the marginals of **-9879.15**, an $AIC_M$ of **19892.29** and a $BIC_M$ of **20209.81**. The results for each node are displayed in Table 11.3.

| Node | Distribution | Marginal log-lik. | # Parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | two component normal | -926.74 | 5 | 1863.48 | 1887.18 |
| mek | two component normal | -725.23 | 5 | 1460.47 | 1484.16 |
| plc | three component normal | -951.96 | 8 | 1919.92 | 1957.83 |
| pip2 | three component normal | -1225.04 | 8 | 2466.07 | 2503.99 |
| pip3 | two component normal | -916.83 | 5 | 1843.66 | 1867.36 |
| erk | normal | -997.54 | 2 | 1999.07 | 2008.55 |
| akt | two component normal | -880.65 | 5 | 1771.30 | 1794.99 |
| pka | three component normal | -769.54 | 8 | 1555.08 | 1592.99 |
| pkc | three component normal | -777.10 | 8 | 1570.20 | 1608.11 |
| p38 | two component normal | -778.79 | 5 | 1567.58 | 1591.28 |
| jnk | three component normal | -929.73 | 8 | 1875.46 | 1913.37 |
| $\sum$: | | -9879.15 | 67 | 19892.29 | 20209.81 |

Table 11.3: Marginal log-likelihood, number of parameters, $AIC_M$ and $BIC_M$ of the fitted Gaussian mixture margins $\mathbf{M^{mix}_{cd3+akt}}$ to the data set from the $cd3cd28 + aktinhib$ experiment

Looking at panel (b) of Figure 11.4 we deduce no obvious indications that the estimated margins do not fit. Especially the data after applying the PIT using the fitted Gaussian mixture margins seems to be uniformly distributed. We already derived that the node *erk* seems to follow a normal distribution when fitting Gaussian margins to this data set. Therefore, also when fitting Gaussian mixture margins the node *erk* is modeled as a normal distribution. The histograms of the data after applying the probability integral transform using the fitted Gaussian mixture margins and the normalized scatter and contour plots can be found in Figure 11.2.


Figure 11.2: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian mixture margins to the data from the $cd3cd28 + aktinhib$ experiment

### 11.2.3 Kernel Density Margin Fitting ($\mathbf{M^{ker}_{cd3+akt}}$)

Again as a third approach fitting kernel density estimates to the data results in a log-likelihood of the marginals of **-9780.28**, an $AIC_M$ of **19786.78** and a $BIC_M$ of **20322.79**. Comparing the results in Table 11.4 with the ones from the fitted Gaussian mixture margins in Table 11.3 we observe a very similar log-likelihood and $AIC_M$ for all nodes. Only the number of (effective) parameters is higher, about twice the size, when using kernel density estimates instead of Gaussian mixture margins. Therefore, we can see slightly bigger differences when comparing the $BIC_M$.

| Node | Distribution | Marginal log-likelihood | Effective # parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | kernel density | -921.52 | 11.06 | 1865.16 | 1917.58 |
| mek | kernel density | -710.24 | 14.94 | 1450.37 | 1521.20 |
| plc | kernel density | -952.81 | 10.70 | 1927.03 | 1977.73 |
| pip2 | kernel density | -1214.93 | 9.04 | 2447.95 | 2490.80 |
| pip3 | kernel density | -895.20 | 11.54 | 1813.47 | 1868.14 |
| erk | kernel density | -991.42 | 7.30 | 1997.43 | 2032.01 |
| akt | kernel density | -865.88 | 11.02 | 1753.80 | 1806.03 |
| pka | kernel density | -767.31 | 8.85 | 1552.31 | 1594.24 |
| pkc | kernel density | -771.76 | 7.28 | 1558.08 | 1592.58 |
| p38 | kernel density | -765.47 | 7.94 | 1546.82 | 1584.45 |
| jnk | kernel density | -923.74 | 13.44 | 1874.36 | 1938.03 |
| $\sum$: | | -9780.28 | 113.11 | 19786.78 | 20322.79 |

Table 11.4: Marginal log-likelihood, number of effective parameters $AIC_M$ and $BIC_M$ of the fitted kernel density margins $\mathbf{M^{ker}_{cd3+akt}}$ to the data set from the $cd3cd28 + aktinhib$ experiment

The similarities continue when looking at the graphical analysis of the kernel density estimates in panel (c) of Figure 11.4 and comparing them with the ones from the Gaussian mixture margins in panel (b). Here, again no obvious indication can be found that the fitted kernel density estimates do not fit. Further, all three types of plots in Figure 11.3 strongly resemble the ones from Figure 11.2.

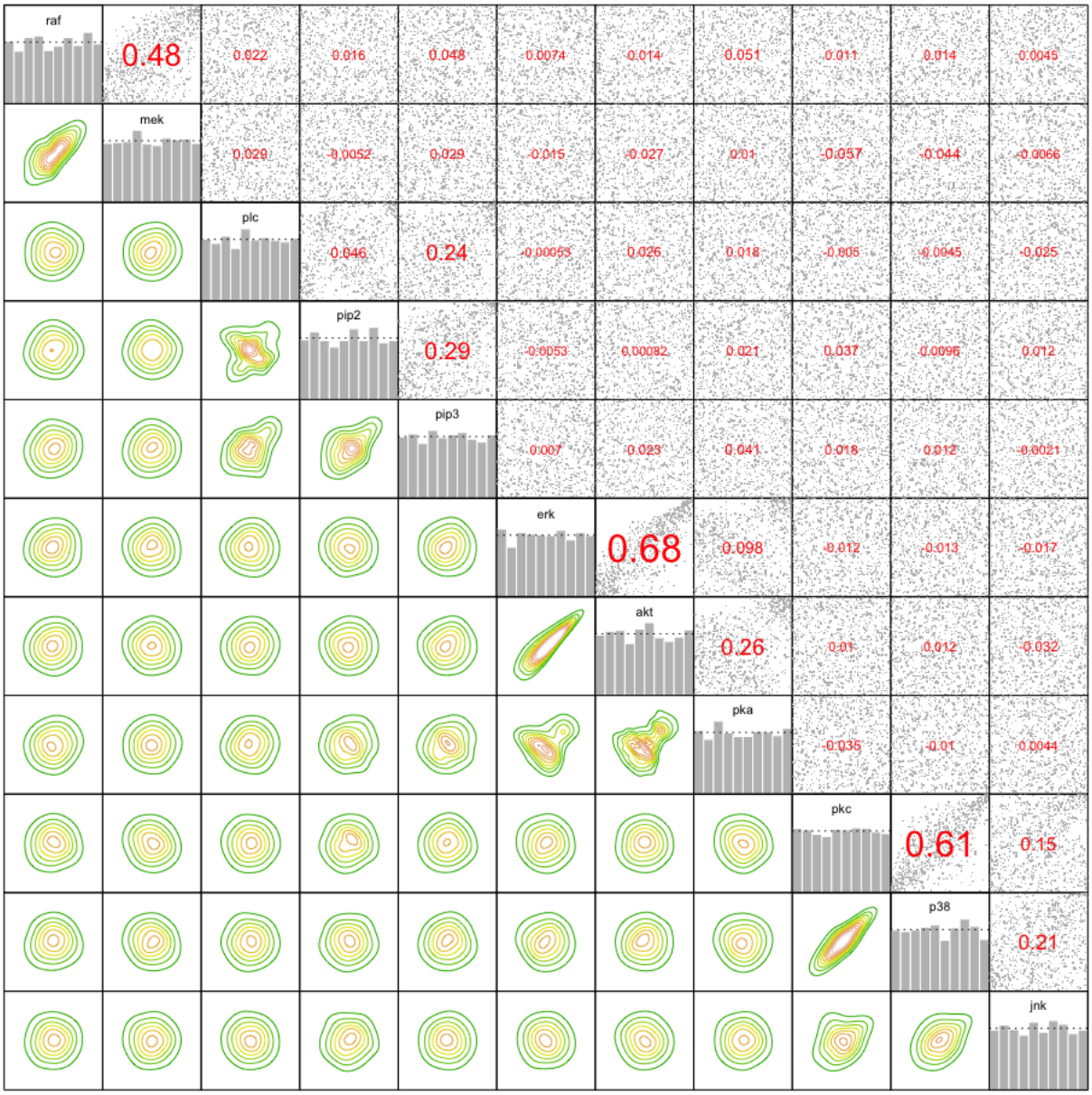
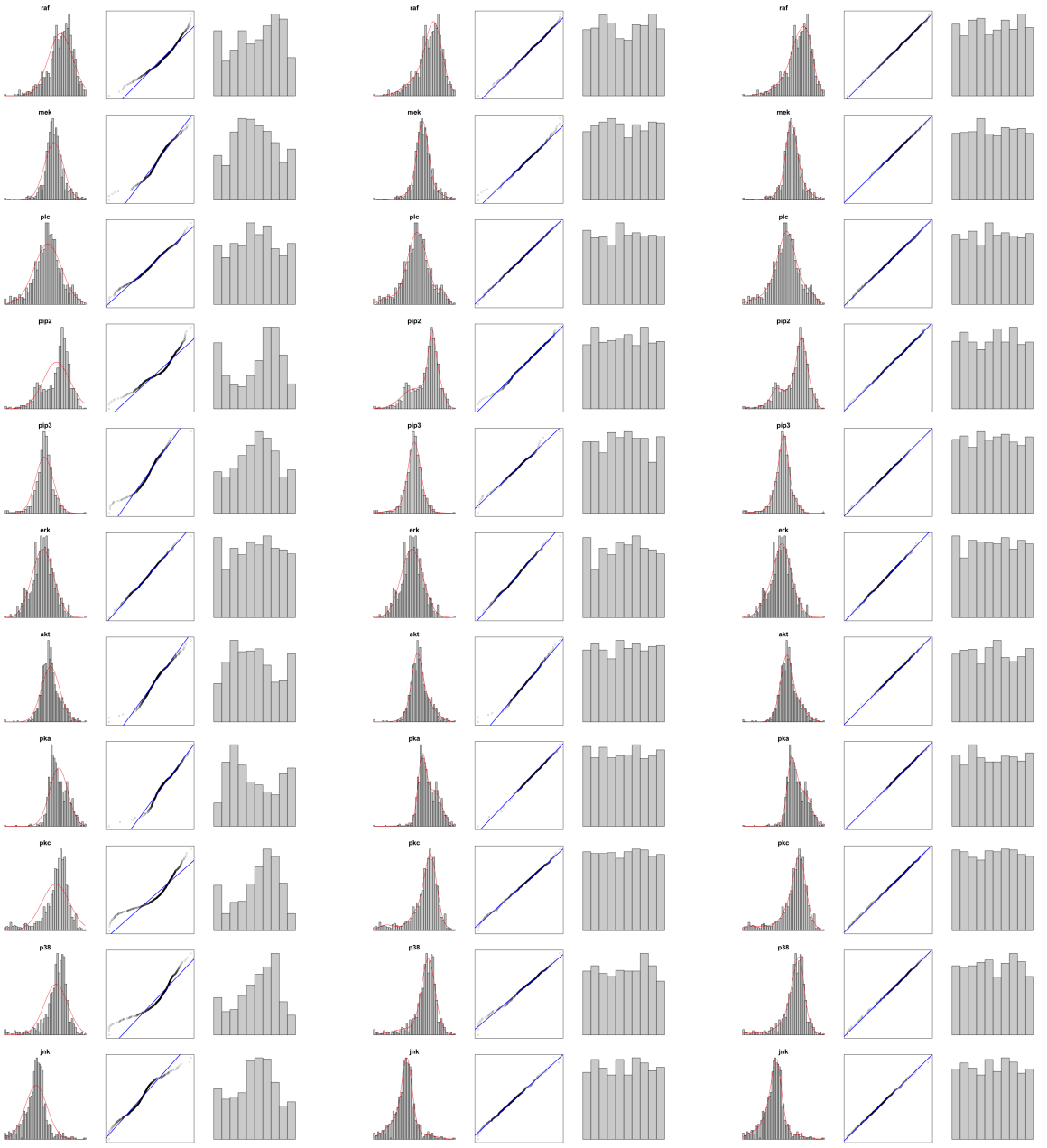



Figure 11.3: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted kernel density margins to the data from the $cd3cd28 + aktinhib$ experiment




(a) $\mathbf{M}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ (b) $\mathbf{M}^{\mathbf{mix}}_{\mathbf{cd3+akt}}$ (c) $\mathbf{M}^{\mathbf{ker}}_{\mathbf{cd3+akt}}$

Figure 11.4: Density plot and Q-Q plot of the chosen margins (in the first two columns) and histogram of the data after applying the distribution function of the fitted marginal distributions, in the third column, for all nodes for the chosen marginals on the *cd*3*cd*28+ *aktinhib* data set

### 11.2.4 Comparison of Marginal Fits

Having fitted three different types of marginals to the $cd3cd28 + aktinhib$ data set we derived that the fitted kernel density estimates and Gaussian mixture margins have an almost identical goodness of fit speaking in terms of the log-likelihood, $\text{AIC}_M$ and $\text{BIC}_M$. This holds even though the kernel density estimates require approximately twice as many (effective) parameters. As the absolute amount of parameters is still very small, together with the size of the data set, this difference has an almost vanishing influence on these three goodness of fit measures.

Compared to these two approaches fitting Gaussian margins results overall in a worse fit. Not only in terms of log-likelihood, $\text{AIC}_M$ or $\text{BIC}_M$, also none of the graphical analysis has shown that Gaussian margins seem suitable. The only exception here is the node $erk$. For this node especially the histogram after using the Gaussian margins seems to be approximately uniformly distributed. Therefore it is also modeled as as Gaussian distribution in the approach where Gaussian mixture margins were estimated.

As we observed that fitting Gaussian mixture margins and kernel density estimates up to now resulted in a similar goodness of fit, it does not surprise that their normalized contour and scatter plots are also almost identical as well. Only about ten of the 55 normalized contour plots have a shape different from an ellipse. Looking at the scatter plots we can see that in most of the pairs of nodes tail dependencies are present, for example between the nodes $akt \leftrightarrow pka$ or $pkc \leftrightarrow p38$. This does not hold for pairs of nodes where the dependence structure follows an elliptical form. There, no clear tail dependencies can be identified.

Instead examining the normalized contour plots with the fitted Gaussian margins many of the shapes are not as smooth as in the other two approaches, for example between the nodes $pip2$ and $pkc$. On the other hand, pairs that follow an elliptical shape in the other two approaches also follow an elliptical shape here and vice versa. Again it needs to be stressed that the data after applying the PIT with Gaussian margins is not uniformly distributed. Therefore, in conclusion the results are very similar to the ones in Section 5.4.

## 11.3 D-vine Regression ($\mathbf{MC_{cd3+akt}}$)

### 11.3.1 Kernel Density Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$)

Using the fitted kernel density margins from Section 11.2.3 to transform the data from the *cd*3*cd*28 + *aktinhib* experiment to the copula scale and then fit a D-vine regression allowing for parametric and non-parametric copulas results in a log-likelihood of **2328.56**, an $AIC_C$ of **-4236.95** and a $BIC_C$ of **-3241.29** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 19.93 | -11.07 | 57.16 |
| mek | raf, pkc, pka | 310.64 | -574.42 | -463.38 |
| plc | pip3 | 125.93 | -207.80 | -103.43 |
| pip2 | pip3, plc | 293.84 | -509.94 | -325.71 |
| erk | pka | 169.48 | -301.30 | -212.06 |
| akt | erk, pka | 786.97 | -1517.86 | -1384.96 |
| pka | pkc | 1.08 | -0.16 | 4.58 |
| pkc | pip2 | 1.82 | -1.64 | 3.09 |
| p38 | pkc, pka | 519.43 | -982.10 | -847.58 |
| jnk | pkc, pka | 99.44 | -130.66 | 31.00 |
| $\sum$: | | 2328.56 | -4236.95 | -3241.29 |

Table 11.5: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model

Detailed results per node are displayed in Table 11.5 and per copula in Table 11.6. In the fitted model three of the twenty edges, namely $mek \to erk$, $pip3 \to akt$ and $plc \to pkc$, in the DAG are not modeled. Of the remaining edges seven are modeled as a parametric copula. Contour plots of the fitted copulas can be found in Section 11.3.3 in Figure 11.5 together with the ones from the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | tll | 13.40 | 18.66 | -10.53 | 52.96 | 0.05 |
| | raf pkc; pka | clayton | 1.00 | 1.27 | -0.54 | 4.20 | 0.01 |
| mek | mek raf | tll | 21.43 | 303.88 | -564.91 | -463.35 | 0.48 |
| | mek pkc; raf | gaussian | 1.00 | 4.52 | -7.03 | -2.29 | -0.06 |
| | mek pka; raf pkc | gumbel | 1.00 | 2.24 | -2.48 | 2.26 | 0.01 |
| plc | plc pip3 | tll | 22.02 | 125.93 | -207.80 | -103.43 | 0.24 |
| pip2 | pip2 pip3 | tll | 19.76 | 175.66 | -311.80 | -218.16 | 0.05 |
| | pip2 plc; pip3 | tll | 19.11 | 118.19 | -198.14 | -107.55 | 0.29 |
| erk | erk pka | tll | 18.83 | 169.48 | -301.30 | -212.06 | 0.10 |
| akt | akt erk | tll | 26.04 | 664.74 | -1277.39 | -1153.96 | 0.67 |
| | akt pka; erk | bb8 | 2.00 | 122.24 | -240.47 | -231.00 | 0.26 |
| pka | pka pkc | frank | 1.00 | 1.08 | -0.16 | 4.58 | -0.03 |
| pkc | pkc pip2 | clayton | 1.00 | 1.82 | -1.64 | 3.09 | -0.01 |
| p38 | p38 pkc | tll | 27.38 | 517.73 | -980.69 | -850.91 | 0.60 |
| | p38 pka; pkc | joe | 1.00 | 1.71 | -1.41 | 3.33 | -0.01 |
| jnk | jnk pkc | tll | 21.09 | 85.61 | -129.04 | -29.08 | 0.15 |
| | jnk pka; pkc | tll | 13.02 | 13.83 | -1.62 | 60.08 | 0.00 |

Table 11.6: Summary of all copulas fitted in the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model

### 11.3.2 Gaussian Mixture Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$)

If we use the fitted Gaussian mixture margins from Section 11.2.2 instead of the fitted kernel density margins for the probability integral transform, the fitted D-vine model has a log-likelihood of **2294.97**, an $AIC_C$ of **-4187.84** and a $BIC_C$ of **-3235.00** of the copula terms if we allow for parametric and non-parametric copulas.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 19.33 | -9.65 | 59.08 |
| mek | raf, pkc, pka | 308.88 | -572.12 | -463.99 |
| plc | pip3 | 125.93 | -210.15 | -111.31 |
| pip2 | pip3, plc | 292.01 | -501.05 | -304.45 |
| erk | pka | 167.43 | -296.76 | -206.45 |
| akt | erk, pka | 778.05 | -1496.78 | -1356.20 |
| pka | pkc | 1.13 | -0.26 | 4.48 |
| pkc | pip2 | 1.84 | -1.68 | 3.06 |
| p38 | pkc, pka | 515.64 | -977.80 | -851.09 |
| jnk | pkc, pka | 84.74 | -121.59 | -8.13 |
| $\sum$: | | 2294.97 | -4187.84 | -3235.00 |

Table 11.7: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model

The results per node are given in Table 11.7 and the results per copula in Table 11.8. There, we observe that the same seventeen edges as in the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model are fitted. Furthermore, eight of them are modeled using a parametric copula. All but the $c_{jnk,pka;pkc}$ copula have also been modeled as a parametric copula in the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model. The respective contour plots can be found in Figure 11.5 in Section 11.3.3.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | tll | 13.50 | 18.12 | -9.24 | 54.75 | 0.05 |
| | raf pkc; pka | clayton | 1.00 | 1.20 | -0.41 | 4.33 | 0.01 |
| mek | mek raf | tll | 20.82 | 302.03 | -562.43 | -463.78 | 0.48 |
| | mek pkc; raf | gaussian | 1.00 | 4.68 | -7.35 | -2.61 | -0.06 |
| | mek pka; raf pkc | gaussian | 1.00 | 2.17 | -2.34 | 2.40 | 0.01 |
| plc | plc pip3 | tll | 20.85 | 125.93 | -210.15 | -111.31 | 0.24 |
| pip2 | pip2 pip3 | tll | 20.25 | 175.48 | -310.46 | -214.50 | 0.05 |
| | pip2 plc; pip3 | tll | 21.23 | 116.53 | -190.59 | -89.95 | 0.29 |
| erk | erk pka | tll | 19.05 | 167.43 | -296.76 | -206.45 | 0.10 |
| akt | akt erk | tll | 27.66 | 653.59 | -1251.86 | -1120.76 | 0.67 |
| | akt pka; erk | bb8 | 2.00 | 124.46 | -244.92 | -235.44 | 0.26 |
| pka | pka pkc | frank | 1.00 | 1.13 | -0.26 | 4.48 | -0.03 |
| pkc | pkc pip2 | clayton | 1.00 | 1.84 | -1.68 | 3.06 | -0.01 |
| p38 | p38 pkc | tll | 25.74 | 513.87 | -976.27 | -854.29 | 0.60 |
| | p38 pka; pkc | joe | 1.00 | 1.77 | -1.53 | 3.20 | -0.01 |
| jnk | jnk pkc | tll | 22.94 | 83.66 | -121.44 | -12.72 | 0.15 |
| | jnk pka; pkc | gumbel | 1.00 | 1.08 | -0.15 | 4.59 | 0.00 |

Table 11.8: Summary of all copulas fitted in the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model

### 11.3.3 Gaussian Margins with Gaussian Copulas ($\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$)

As a third approach to transform the data to the copula scale we used Gaussian margins and then allowed for only Gaussian copulas in the D-vine regression. This results in a log-likelihood of **1497.84**, an $AIC_C$ of **-2971.68** and a $BIC_C$ of **-2914.80** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 2.15 | -0.30 | 9.18 |
| mek | raf, pkc | 232.86 | -461.72 | -452.24 |
| plc | pip3 | 57.72 | -113.43 | -108.69 |
| pip2 | pip3 | 88.80 | -175.61 | -170.87 |
| erk | pka | 29.16 | -56.33 | -51.59 |
| akt | erk, pka | 715.74 | -1427.47 | -1417.99 |
| pka | | | | |
| pkc | pip2 | 1.73 | -1.46 | 3.28 |
| p38 | pkc | 354.14 | -706.28 | -701.54 |
| jnk | pkc | 15.54 | -29.08 | -24.34 |
| $\sum$: | | 1497.84 | -2971.68 | -2914.80 |

Table 11.9: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model

Table 11.9 shows the results for each D-vine while Table 11.10 displays the results for each copula. There, we can see that of the twenty relationships given by the DAG only twelve are modeled. Additional to the ones missing in the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model, this time also the dependencies $pka \to mek$, $plc \to pip2$, $pkc \to pka$, $pka \to p38$ and $pka \to jnk$ are missing. The contour plots are displayed in Figure 11.5.

| Node | Pair copula | Family | Copula log-lik. | $AIC_C$ | $BIC_C$ | Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | gaussian | 1.08 | -0.15 | 4.59 | 0.05 | 0.05 |
| | raf pkc; pka | gaussian | 1.08 | -0.15 | 4.59 | 0.05 | 0.01 |
| mek | mek raf | gaussian | 227.58 | -453.17 | -448.43 | 0.65 | 0.48 |
| | mek pkc; raf | gaussian | 5.27 | -8.55 | -3.81 | -0.11 | -0.06 |
| plc | plc pip3 | gaussian | 57.72 | -113.43 | -108.69 | 0.36 | 0.24 |
| pip2 | pip2 pip3 | gaussian | 88.80 | -175.61 | -170.87 | 0.44 | 0.05 |
| erk | erk pka | gaussian | 29.16 | -56.33 | -51.59 | 0.26 | 0.10 |
| akt | akt erk | gaussian | 530.82 | -1059.63 | -1054.89 | 0.85 | 0.67 |
| | akt pka; erk | gaussian | 184.92 | -367.84 | -363.10 | 0.60 | 0.26 |
| pkc | pkc pip2 | gaussian | 1.73 | -1.46 | 3.28 | 0.06 | -0.01 |
| p38 | p38 pkc | gaussian | 354.14 | -706.28 | -701.54 | 0.75 | 0.60 |
| jnk | jnk pkc | gaussian | 15.54 | -29.08 | -24.34 | 0.19 | 0.15 |

Table 11.10: Summary of all copulas fitted in the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model


(a) $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ (b) $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ (c) $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$

Figure 11.5: Contour plots of the respective copulas in the models allowing for parametric and non-parametric pair copulas where kernel density margins or Gaussian mixture margins were used for the PIT and the model where Gaussian margins and only Gaussian copulas were used on the *cd*3*cd*28 + *aktinhib* data set

### 11.3.4 Comparison of D-vine Regression Models

Comparing the results of the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model it seems that the decision between using kernel density margins and Gaussian mixture margins for the probability integral transform does not have a notable influence on the outcome.

All of the D-vines have the same order in both models with all but two copulas, the $c_{mek,pka;raf,pkc}$ and the $c_{jnk,pka;pkc}$ copula, having the same family and similar number of effective parameters. This also holds for the goodness of fit, speaking in terms of log-likelihood, $\text{AIC}_C$ and $\text{BIC}_C$. Additionally it holds that the $c_{mek,pka;raf,pkc}$ and the $c_{jnk,pka;pkc}$ copula have an almost vanishing influence in both models. Looking at the contour plots in both models in Figure 11.5 it is hard to derive any differences, even in the $c_{mek,pka;raf,pkc}$ copula. The only visible differences can be observed in the contour plot of the $c_{jnk,pka;pkc}$ copula.

In total leading to an almost identical fit of each node in both models and therefore an almost identical fit of the whole model. This is something we have also seen when we fitted these two types of models to the complete Sachs data set in Chapter 6.

When we fitted models with kernel density margins or Gaussian mixture margins and allowing for parametric and non-parametric copulas on the complete Sachs data set all twenty edges of the DAG have been modeled. This is not the case when fitting the same models on the $cd3cd28 + aktinhib$ data set. In both the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model only seventeen of these edges are modeled.

Instead using Gaussian margins and only Gaussian copulas on the $cd3cd28 + aktinhib$ data set does even more decrease the number of modeled dependencies to twelve. One of these additional missing dependencies is the $pkc \to pka$ dependency meaning that in this model the node $pka$ is modeled independent of all other nodes. On the contrary, compared to that in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model still seventeen of the edges were modeled.

Hence, it does not surprise that the goodness of fit of the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ model is way worse than the fit of the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model.

## 11.4 Goodness of Fit Measures Model Comparison

To be able to compare the D-vine regression model and the Linear Gaussian Bayesian Network fitted on the $cd3cd28 + aktinhib$ data set in terms of log-likelihood, $\text{AIC}_F$ and $\text{BIC}_F$ we add up the respective measures of the marginals and of the copula terms. The results for the log-likelihood can be found in Table 11.11, for the $\text{AIC}_F$ in Table 11.12, for the $\text{BIC}_F$ in Table 11.13 and for the (effective) number of parameters used in the models in Table 11.14.

Following Section 2.4.3 and Section 2.4.5 we know that both the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ model and the $\text{LGBN}_{\mathbf{cd3+akt}}$ specify a Gaussian distribution. It therefore does not surprise that the goodness of fit measures are extremely similar up to the point where they are almost identical in regards to the log-likelihood and the $\text{AIC}_F$.

Only for the $\text{BIC}_F$ small differences can be seen. This is due to the fact that we maximized the log-likelihood instead of the $\text{AIC}_F$ when fitting the $\text{LGBN}_{\mathbf{cd3+akt}}$. It therefore models each possible dependency even if it only provides a small increase in the log-likelihood. Hence, in generally we expect this model to have a higher number of parameters

Looking at the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model instead, we observe that both provide a way better fit than the other two models due to the relaxations on the copula families and the use of margins with a better fit. Comparing the log-likelihood and the $\text{AIC}_F$ we derive that all nodes apart from $plc$ have a slightly better fit in the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model than in the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model. Therefore, the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ is overall the best model when looking at the log-likelihood or the $\text{AIC}_F$.

Instead considering the $\text{BIC}_F$, all nodes except $pip2$, $akt$, $pkc$ and $p38$ have a better fit in the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model. This is due to the increased amount of effective parameters required for the kernel density margins compared to the Gaussian mixture margins. Therefore, considering the overall $\text{BIC}_F$ the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model is the best of the four models analyzed. Taking the absolute value of the measures into account the differences between the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model are not significant.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ | -970 | -552 | -922 | -1249 | -987 | -968 | -196 | -831 | -952 | -571 | -992 | -9189 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ | -902 | -400 | -827 | -921 | -895 | -822 | -79 | -766 | -770 | -246 | -824 | -7452 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ | -907 | -416 | -826 | -933 | -917 | -830 | -103 | -768 | -775 | -263 | -845 | -7584 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ | -970 | -552 | -922 | -1249 | -987 | -968 | -197 | -831 | -952 | -571 | -993 | -9193 |

Table 11.11: Global log-likelihood of the model fitted on the $cd3cd28 + aktinhib$ data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ | 1948 | 1113 | 1850 | 2507 | 1978 | 1944 | 402 | 1668 | 1913 | 1149 | 1992 | 18463 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ | 1854 | 876 | 1719 | 1938 | 1813 | 1696 | 236 | 1552 | 1556 | 565 | 1744 | 15550 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ | 1854 | 888 | 1710 | 1965 | 1844 | 1702 | 275 | 1555 | 1569 | 590 | 1754 | 15704 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ | 1948 | 1113 | 1850 | 2505 | 1978 | 1943 | 401 | 1666 | 1911 | 1148 | 1991 | 18453 |

Table 11.12: $\mathrm{AIC}_F$ of the models fitted on the $cd3cd28 + aktinhib$ data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{LGBN_{cd3+akt}}$ | 1967 | 1137 | 1864 | 2526 | 1987 | 1963 | 426 | 1687 | 1932 | 1168 | 2011 | 18666 |
| $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ | 1975 | 1058 | 1874 | 2165 | 1868 | 1820 | 421 | 1599 | 1596 | 737 | 1969 | 17082 |
| $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ | 1946 | 1020 | 1847 | 2200 | 1867 | 1802 | 439 | 1597 | 1611 | 740 | 1905 | 16975 |
| $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ | 1967 | 1132 | 1864 | 2519 | 1987 | 1957 | 420 | 1676 | 1925 | 1162 | 2005 | 18614 |

Table 11.13: $BIC_F$ of the models fitted on the $cd3cd28 + aktinhib$ data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{LGBN_{cd3+akt}}$ | 4 | 5 | 3 | 4 | 2 | 4 | 5 | 3 | 4 | 4 | 4 | 42 |
| $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ | 25 | 38 | 33 | 48 | 12 | 26 | 39 | 10 | 8 | 36 | 48 | 323 |
| $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ | 20 | 28 | 29 | 49 | 5 | 21 | 35 | 9 | 9 | 32 | 32 | 268 |
| $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ | 4 | 4 | 3 | 3 | 2 | 3 | 4 | 2 | 3 | 3 | 3 | 34 |

Table 11.14: Number of (effective) parameters of the models fitted on the $cd3cd28 + aktinhib$ data set for each node

## 11.5 Simulation Based Comparison

To compare the models fitted to the data set of the $cd3cd28+aktinhib$ experiment in more detail we simulated from all of them. To be able to compare the results with the original data set we sampled 845 times from all four models following the procedure as described in Chapter 8. Figure 11.6 shows the histograms of the $cd3cd28+aktinhib$ data set for each node and the fitted kernel density estimates to the data simulated from the four models. For the **MC** models the fit of the kernel density estimates to the simulated data only depends on the fit of the marginals in the models. Hence, it does not surprise that the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model outperform the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model and the $\text{LGBN}_{\mathbf{cd3+akt}}$. This holds as we derived that both the kernel density margins and the Gaussian mixture margins fit quite good to the data. Only for the node $erk$ the fit is similar as it is also modeled as a Gaussian distribution in the approach with Gaussian mixture margins with the fitted kernel density margins being very similar to that. Note that all four models seem to have problems to recreate the correct height of the peak at the nodes $plc$ and $jnk$. Still it looks like in general the kernel density estimates fitted to the data simulated from the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model are quite close to the kernel density estimates fitted to the data from the $cd3cd28+aktinhib$ experiment.

To compare the dependence structures we instead consider the pairs plots of the $cd3cd28+$ $aktinhib$ data set and the simulated data sets in Figure 11.7 and Figure 11.8. Looking at the pairs plots from the data sets simulated from the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model in Figure 11.7 we observe almost no differences between their plots. This does not surprise as we have already seen that their copulas are almost identical. Comparing them to the ones from the $cd3cd28+aktinhib$ data set almost only similarities are visible as well. For all pairs plots the shapes, even more complicated ones like between the node pairs $raf \leftrightarrow mek$ or $erk \leftrightarrow akt$, are close between the simulated data of the two models and the original data.

Comparing the pairs plots of the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model and the $\text{LGBN}_{\mathbf{cd3+akt}}$ instead, it is clear that both of them have problems modeling more complicated shapes as they are only able to model elliptical shaped dependencies. It seems that the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model does this a little bit better looking at the pairs plots of $raf \leftrightarrow mek$ and $erk \leftrightarrow akt$ which are closer to the ones from the original data set than the ones from the $\text{LGBN}_{\mathbf{cd3+akt}}$.

Next, we want to determine if the models create samples in which accumulations of high/low values appear in a similar frequency as in the original data. Hence, we first sum up according to the dependencies given by the DAG and then over all nodes. The results of the first analysis are illustrated in Figure 11.9 and of the second in Figure 11.10.

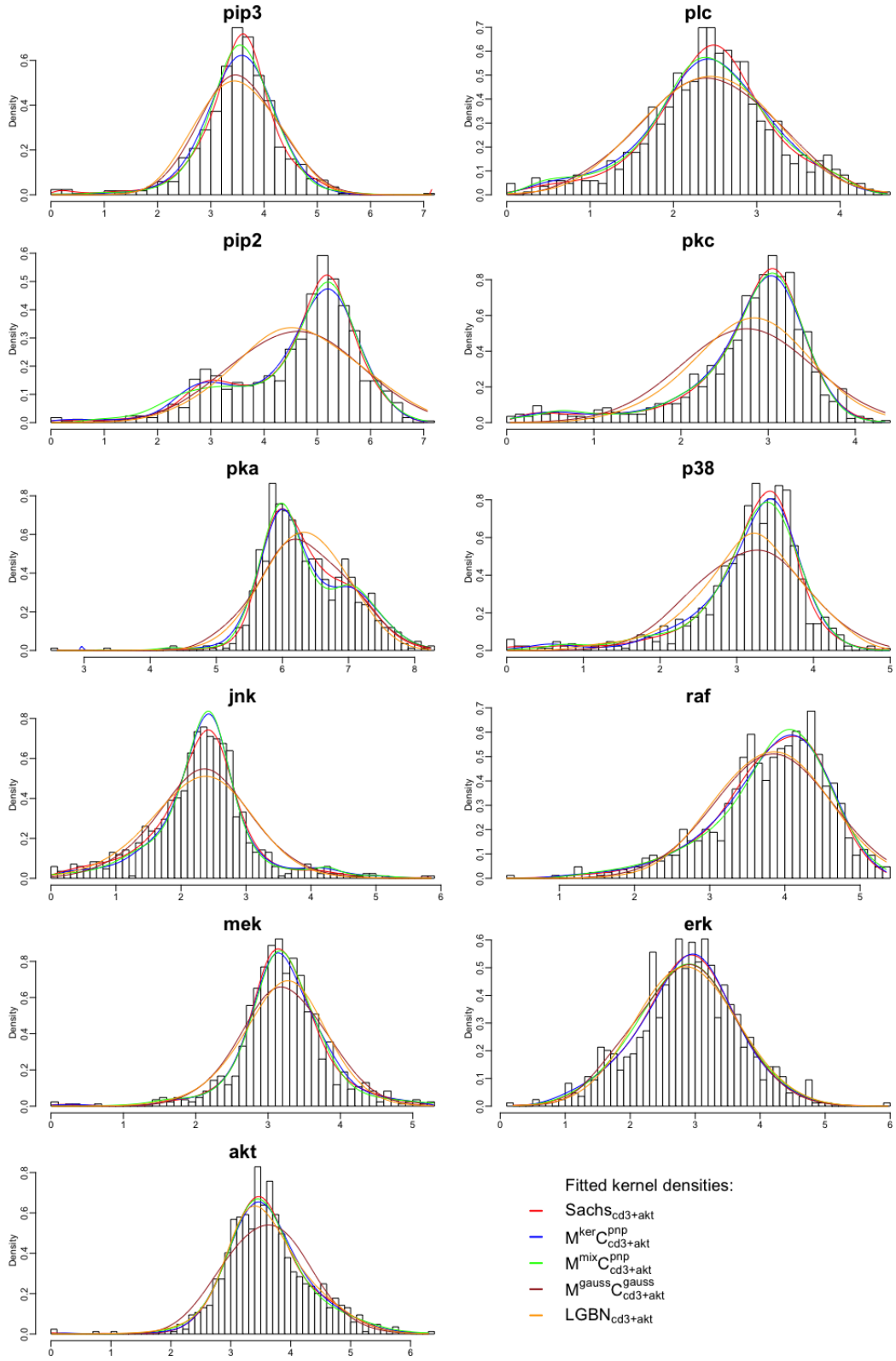
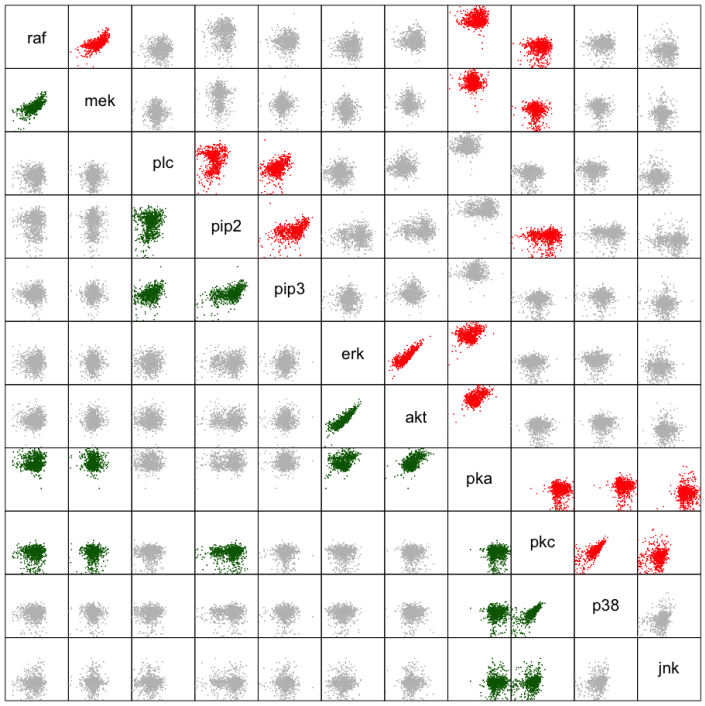
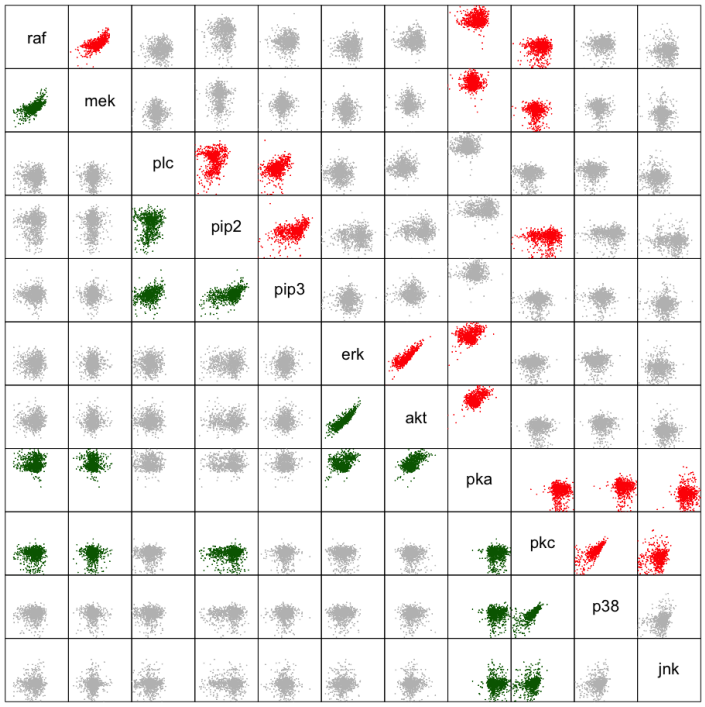



Figure 11.6: Marginal histograms of the $cd3cd28 + aktinhib$ data set of each of the eleven nodes in topological order after samples containing a zero value have been deleted and kernel density estimates have been fitted to the simulated data from the different models

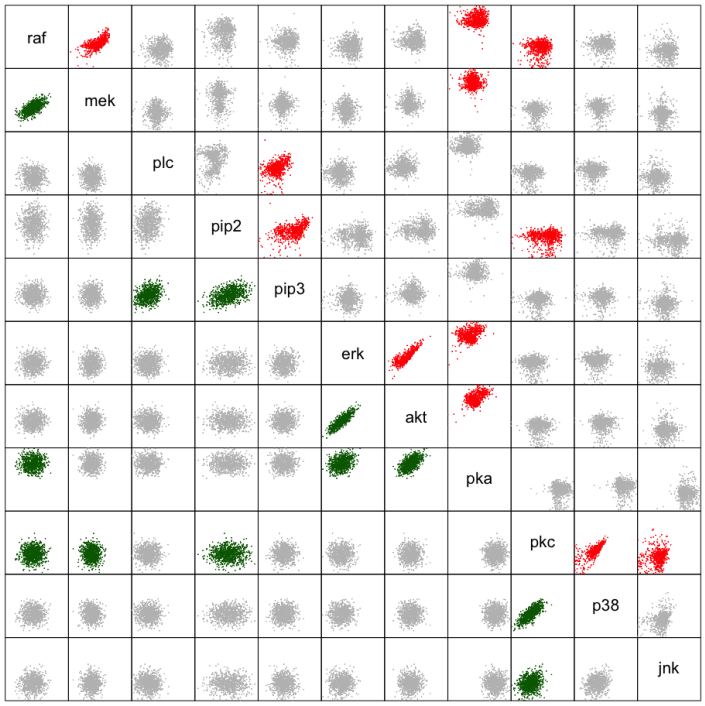



(a) $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$

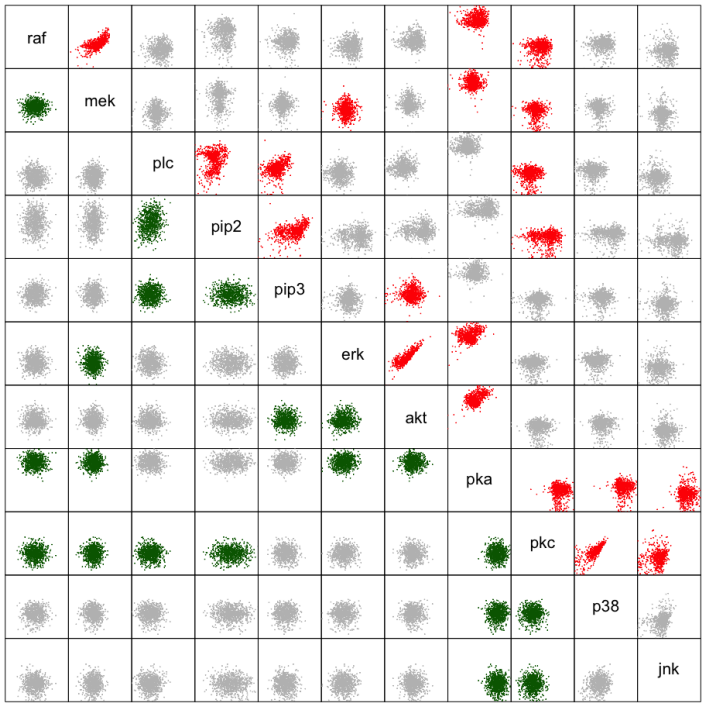



(b) $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$

Figure 11.7: Pairs plots for each pair of nodes of the respective model in the lower panel and of the *cd3cd28 + aktinhib* data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel




(a) $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$




(b) $\text{LGBN}_{\mathbf{cd3+akt}}$

Figure 11.8: Pairs plots for each pair of nodes of the respective model in the lower panel and of the *cd3cd28 + aktinhib* data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel

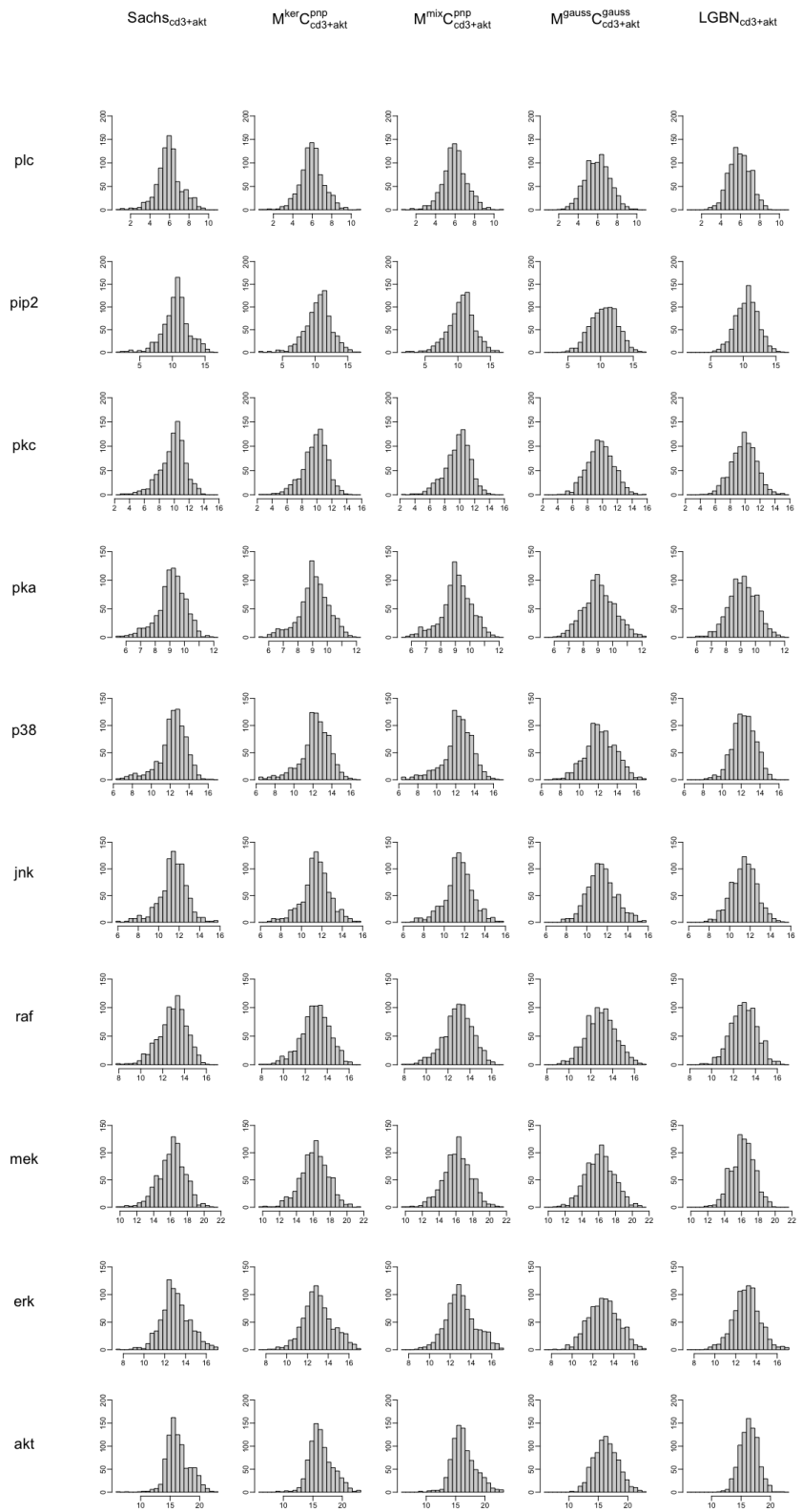



Figure 11.9: Histograms of the sum of each node on the set of its parents according to the consent DAG in the $cd3cd28 + aktinhib$ data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model, $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model, $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model and the $\mathrm{LGBN}_{\mathbf{cd3+akt}}$

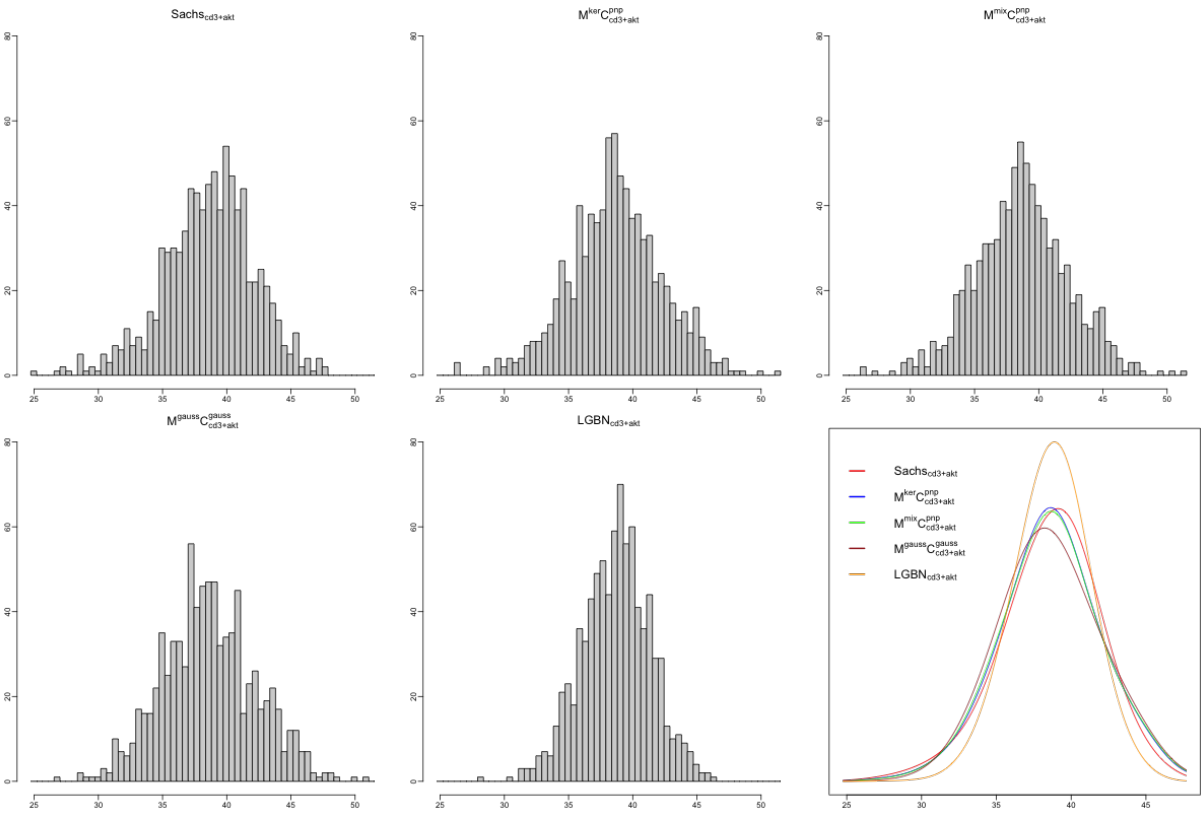



Figure 11.10: Histograms of the sum over all nodes in the *cd3cd28 + aktinhib* data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model, $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model, $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model and the $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ in the first five plots and fitted kernel density estimates to each of the data sets in the sixth plot.

Looking at the histograms in Figure 11.9 we observe that the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model are able to recreate the histograms of the *cd3cd28 + aktinhib* data set quite good. There is only one discrepancy visible for the node *pip2* where both models are not able to recreate the height of the peak. Considering the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model and the $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ we can see that for these models similarities of the histograms to the original data seem to be the exception.

Finally, comparing the histograms of the sum over all nodes in Figure 11.10 we derive that especially the data set simulated from the $\mathbf{M^{ker}C^{pnp}_{cd3+akt}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3+akt}}$ model seem to be very close to the original data set. This holds in particular when considering the fitted kernel density estimates which seem to be only slightly shifted to the left. On the other hand, we observe that the histogram of the sum of the $\mathbf{M^{gauss}C^{gauss}_{cd3+akt}}$ model does not reach the height of the peak compared to the original data set whereas the $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ overestimates the height of the peak.

## 11.6 Conditional Simulation Based Comparison

Now we want to investigate how the conditional densities of the four models behave if the values of the parents are given. This does not show us how well the models are fitted to the data set but gives us the possibility to compare how the models change if we fit them on other data sets.

As conditioning values we take the values we have already determined in Chapter 9 in Table 9.1 in order to compare the plots with the ones of the models fitted on the other data sets.

The results are displayed in Figure 11.11. There, we observe something we have seen throughout the model comparison. This is, the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model in general create very similar results. Here it seems that the peaks of the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model tend to be a little bit higher comparing these two. However, only in some chosen plots clear differences can be seen. This is mostly the case when we condition on points from the tails of the distribution of the parents, e.g., for the nodes $p38$ and $jnk$.

Furthermore, note that for none of the nodes the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ model is close to the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model for all three, respectively five, points on which the node is conditioned on.

Comparing the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ model with the $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ we can now observe more differences than when both models where fitted on the complete Sachs data set. This is due to the fact that in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ model only twelve of the twenty dependencies are modeled while in the $\mathrm{LGBN}_{\mathbf{cd3+akt}}$ all of them are modeled.

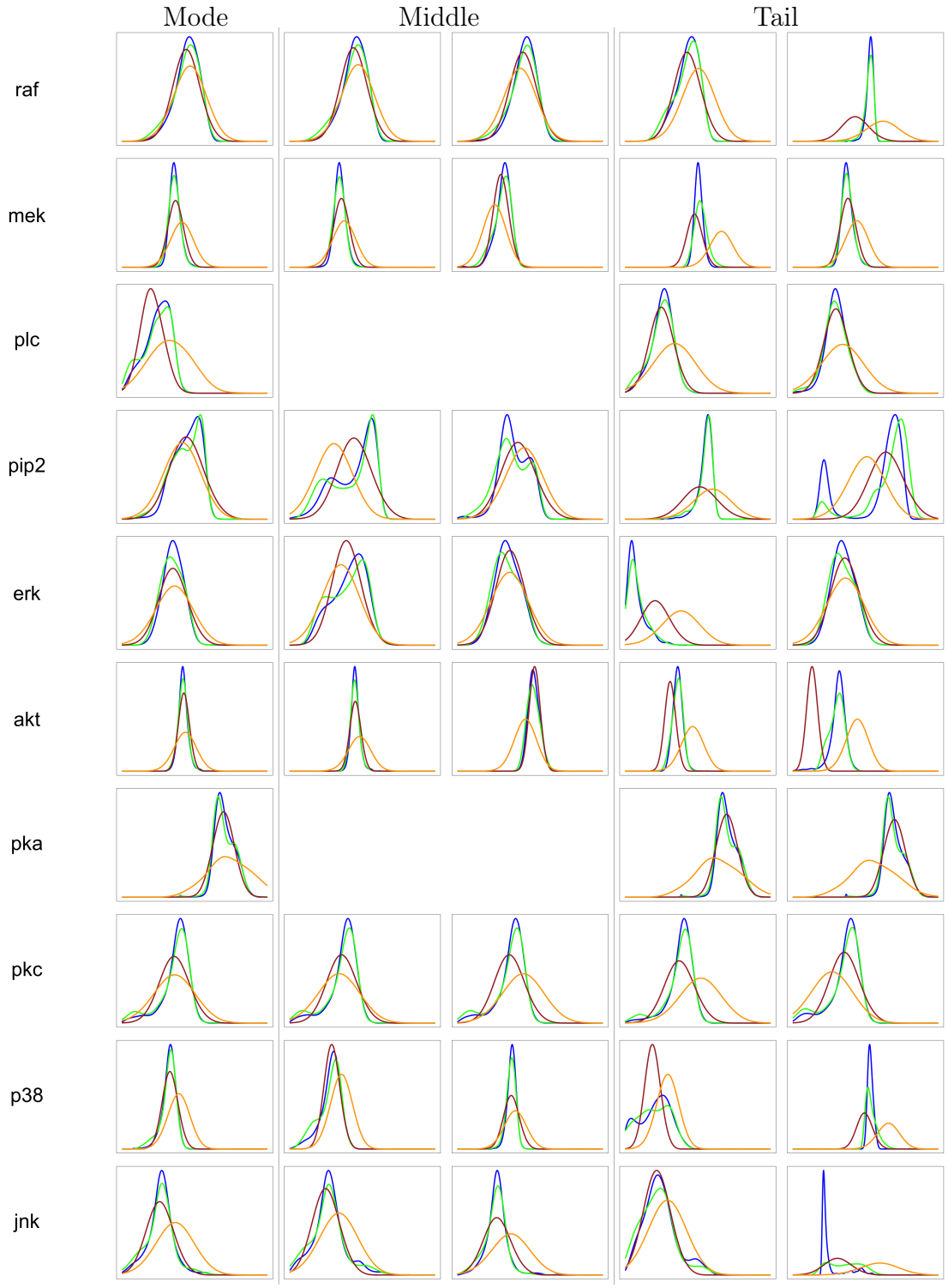



Figure 11.11: Density plots of the kernel density estimates fitted to the simulated data from the different models conditioned on the nodes in Table 9.1 in the same order. Blue: $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$. Green: $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$. Brown: $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$. Orange: $\mathrm{LGBN}_{\mathbf{cd3+akt}}$

# Chapter 12

# Analysis of Group 2: cd3cd28 and cd3cd28 + ly

In this chapter we will investigate the data of the $cd3cd28$ and $cd3cd28 + ly$ experiments using three types of D-vine models and a Linear Gaussian Bayesian Network.

## 12.1 Linear Gaussian Bayesian Network Fitting ($\mathbf{LGBN_{cd3\&cd3+ly}}$)

Fitting a Linear Gaussian Bayesian Network on the data from the $cd3cd28$ and $cd3cd28{+}ly$ experiments results in a log-likelihood of **-18460.01**, an $AIC_F$ of **37003.99** and a $BIC_F$ of **37235.73** of the whole model. The results are shown in Table 12.1.

| Node | Global log-likelihood | # Parameters | $AIC_F$ | $BIC_F$ |
|---|---|---|---|---|
| raf | -1749.70 | 4 | 3507.39 | 3528.95 |
| mek | -985.17 | 5 | 1980.35 | 2007.29 |
| plc | -1793.30 | 3 | 3592.59 | 3608.76 |
| pip2 | -2247.66 | 4 | 4503.31 | 4524.87 |
| pip3 | -2070.70 | 2 | 4145.39 | 4156.17 |
| erk | -1819.81 | 4 | 3647.62 | 3669.17 |
| akt | -728.84 | 5 | 1467.68 | 1494.62 |
| pka | -1781.51 | 3 | 3569.02 | 3590.58 |
| pkc | -1964.41 | 4 | 3936.82 | 3958.38 |
| p38 | -1017.78 | 4 | 2043.56 | 2065.12 |
| jnk | -2301.13 | 4 | 4610.26 | 4631.82 |
| $\sum$: | -18460.01 | 42 | 37003.99 | 37235.73 |

Table 12.1: Global log-likelihood, number of parameters, $AIC_F$ and $BIC_F$ of the fitted linear models for each node given the set of its parents in the Linear Gaussian Bayesian Network on the data set from the $cd3cd28$ and $cd3cd28 + ly$ experiments

## 12.2 Marginal Fitting ($\mathbf{M_{cd3\&cd3+ly}}$)

### 12.2.1 Gaussian Margin Fitting ($\mathbf{M^{gauss}_{cd3\&cd3+ly}}$)

Assuming independence in the data and fitting Gaussian margins to the data from the $cd3cd28$ and $cd3cd28 + ly$ experiments results in a log-likelihood of the marginals of **-20480.04**, an $AIC_M$ of **41004.04** and a $BIC_M$ of **41122.61**. The results for each node are displayed in Table 12.2.

| Node | Marginal log-likelihood | $AIC_M$ | $BIC_M$ |
|---|---|---|---|
| raf | -1750.67 | 3505.33 | 3516.11 |
| mek | -1454.68 | 2913.35 | 2924.13 |
| plc | -1796.23 | 3596.45 | 3607.23 |
| pip2 | -2381.50 | 4767.01 | 4777.79 |
| pip3 | -2070.70 | 4145.39 | 4156.17 |
| erk | -1889.11 | 3782.22 | 3792.99 |
| akt | -1576.09 | 3156.18 | 3166.96 |
| pka | -1783.47 | 3570.94 | 3581.72 |
| pkc | -1967.68 | 3939.35 | 3950.13 |
| p38 | -1443.18 | 2890.36 | 2901.14 |
| jnk | -2366.73 | 4737.46 | 4748.24 |
| $\sum$: | -20480.04 | 41004.04 | 41122.61 |

Table 12.2: Marginal log-likelihood, $AIC_M$ and $BIC_M$ of the fitted Gaussian margins $\mathbf{M^{gauss}_{cd3\&cd3+ly}}$ to the data set from the $cd3cd28$ and $cd3cd28 + ly$ experiments

Analyzing the fit of the margins in panel (a) of Figure 12.3 we observe that for none of the nodes all three types of plots indicate that a Gaussian margin is suitable. Therefore, as when we fitted Gaussian margins on other data sets, technically the requirements to fit copulas to the data using the estimated Gaussian margins for the PIT are not fulfilled. Further, Figure 12.1 displays the normalized contour and scatter plots of each pair of nodes.


Figure 12.1: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian margins to the data from the $cd3cd28$ and $cd3cd28 + ly$ experiments

### 12.2.2 Gaussian Mixture Margin Fitting ($\mathbf{M^{mix}_{cd3\&cd3+ly}}$)

In the next step fitting Gaussian mixture margins instead of Gaussian margins results in a log-likelihood of the marginals of **-19344.55**, an $AIC_M$ of **38847.09** and a $BIC_M$ of **39272.82**. We observe in Table 12.3 that no single node is modeled a Gaussian distribution. Instead all nodes except the node *pip*2, which is modeled as a four component Gaussian mixture, are modeled as a two or three component Gaussian mixture.

| Node | Distribution | Marginal log-Lik. | # Parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | two component normal | -1667.51 | 5 | 3345.02 | 3371.96 |
| mek | two component normal | -1357.54 | 5 | 2725.09 | 2752.03 |
| plc | two component normal | -1744.75 | 5 | 3499.50 | 3526.44 |
| pip2 | four component normal | -2280.28 | 11 | 4582.55 | 4641.83 |
| pip3 | three component normal | -2034.97 | 8 | 4085.94 | 4129.05 |
| erk | three component normal | -1845.22 | 8 | 3706.45 | 3749.56 |
| akt | two component normal | -1487.83 | 5 | 2985.65 | 3012.60 |
| pka | three component normal | -1497.51 | 8 | 3011.02 | 3054.14 |
| pkc | three component normal | -1797.63 | 8 | 3611.26 | 3654.38 |
| p38 | three component normal | -1365.08 | 8 | 2746.15 | 2789.26 |
| jnk | three component normal | -2266.23 | 8 | 4548.46 | 4591.57 |
| $\sum$: | | -19344.55 | 79 | 38847.09 | 39272.82 |

Table 12.3: Marginal log-likelihood, number of parameters, $AIC_M$ and $BIC_M$ of the fitted Gaussian mixture margins $\mathbf{M^{mix}_{cd3\&cd3+ly}}$ to the data set from the $cd3cd28$ and $cd3cd28+ly$ experiments

Again looking at panel (b) of Figure 12.3 we derive that all three types of plots indicate a good fit for all nodes. Especially the histograms of the data after applying the PIT using the fitted Gaussian mixture margins seem to be uniformly distributed. Further information is displayed in Figure 12.2. There the normalized scatter and contour plots as well a again the histograms of the data after applying the PIT using the fitted Gaussian mixture margins can be found.

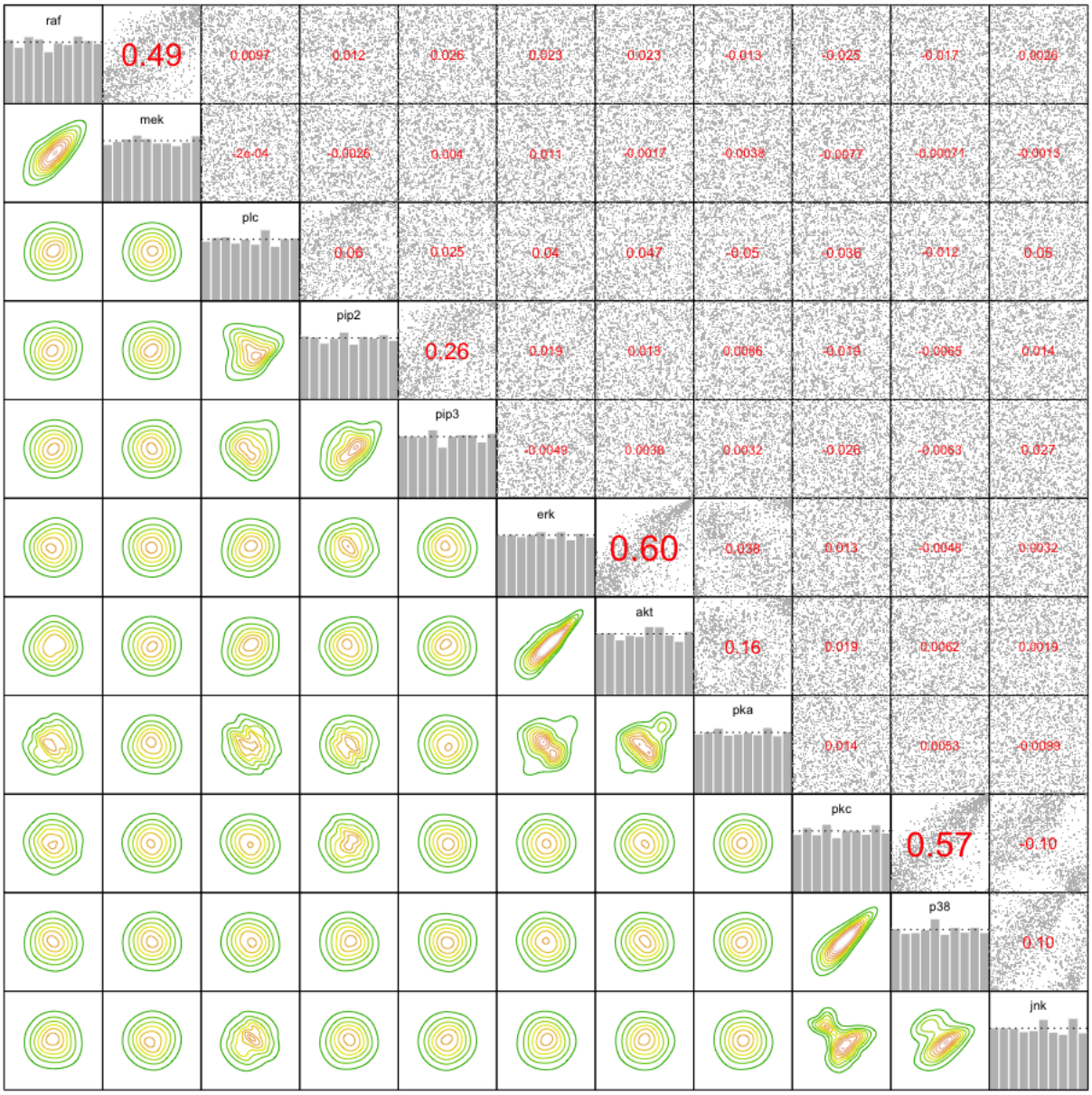



Figure 12.2: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian mixture margins to the data from the $cd3cd28$ and $cd3cd28 + ly$ experiments

### 12.2.3 Kernel Density Margin Fitting ($\mathbf{M^{ker}_{cd3\&cd3+ly}}$)

Instead fitting kernel density estimates to the data, assuming independence, results in a log-likelihood of the marginals of **-19299.83**, an $AIC_M$ of **38962.12** and a $BIC_M$ of **39938.74**. Similar to the earlier approaches on the other data sets, we observe that the results when fitting kernel density estimates are very close to the results when fitting Gaussian mixture margins.

| Node | Distribution | Marginal log-likelihood | Effective # parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | kernel density | -1660.43 | 9.84 | 3340.55 | 3393.60 |
| mek | kernel density | -1348.14 | 49.50 | 2795.28 | 3062.05 |
| plc | kernel density | -1739.36 | 12.51 | 3503.75 | 3571.18 |
| pip2 | kernel density | -2279.52 | 10.53 | 4580.09 | 4636.81 |
| pip3 | kernel density | -2027.86 | 9.56 | 4074.82 | 4126.32 |
| erk | kernel density | -1842.67 | 35.51 | 3756.36 | 3947.71 |
| akt | kernel density | -1495.11 | 7.89 | 3006.01 | 3048.54 |
| pka | kernel density | -1489.91 | 16.56 | 3012.95 | 3102.19 |
| pkc | kernel density | -1795.41 | 6.90 | 3604.63 | 3641.83 |
| p38 | kernel density | -1362.81 | 12.23 | 2750.08 | 2815.98 |
| jnk | kernel density | -2258.61 | 10.19 | 4537.60 | 4592.53 |
| $\sum$: | | -19299.83 | 181.22 | 38962.12 | 39938.74 |

Table 12.4: Marginal log-likelihood, number of effective parameters, $AIC_M$ and $BIC_M$ of the fitted kernel density margins $\mathbf{M^{ker}_{cd3\&cd3+ly}}$ to the data set from the $cd3cd28$ and $cd3cd28 + ly$ experiments

This can be seen in the detailed results for each node in Table 12.4 which are very similar to the ones in Table 12.3 as well as in the graphical analysis in panel (c) of Figure 12.3. There only slightest differences to panel (b) are visible. Furthermore, the normalized contour and scatter plots in Figure 12.4 strongly resemble the ones from Figure 12.2.

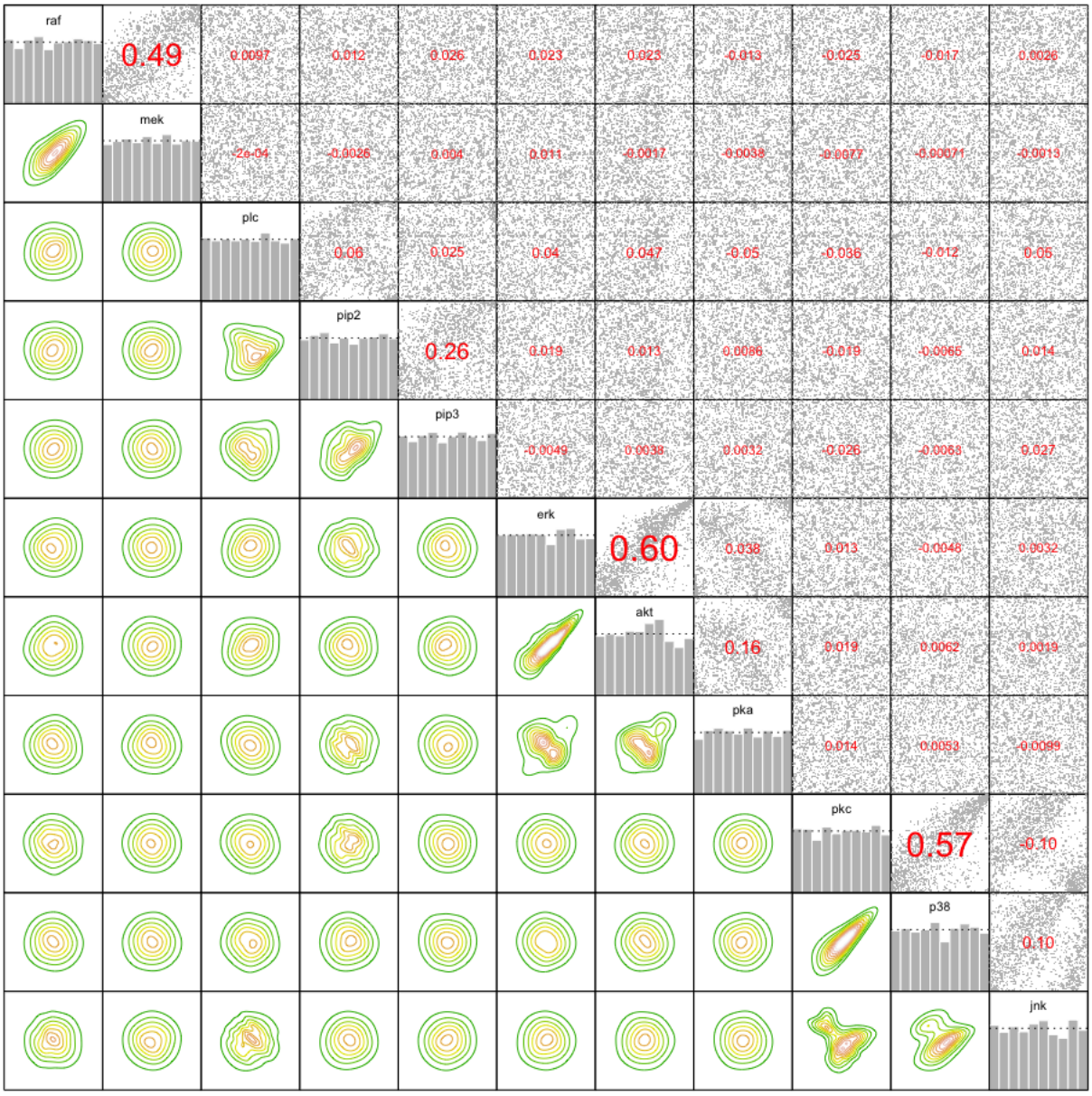
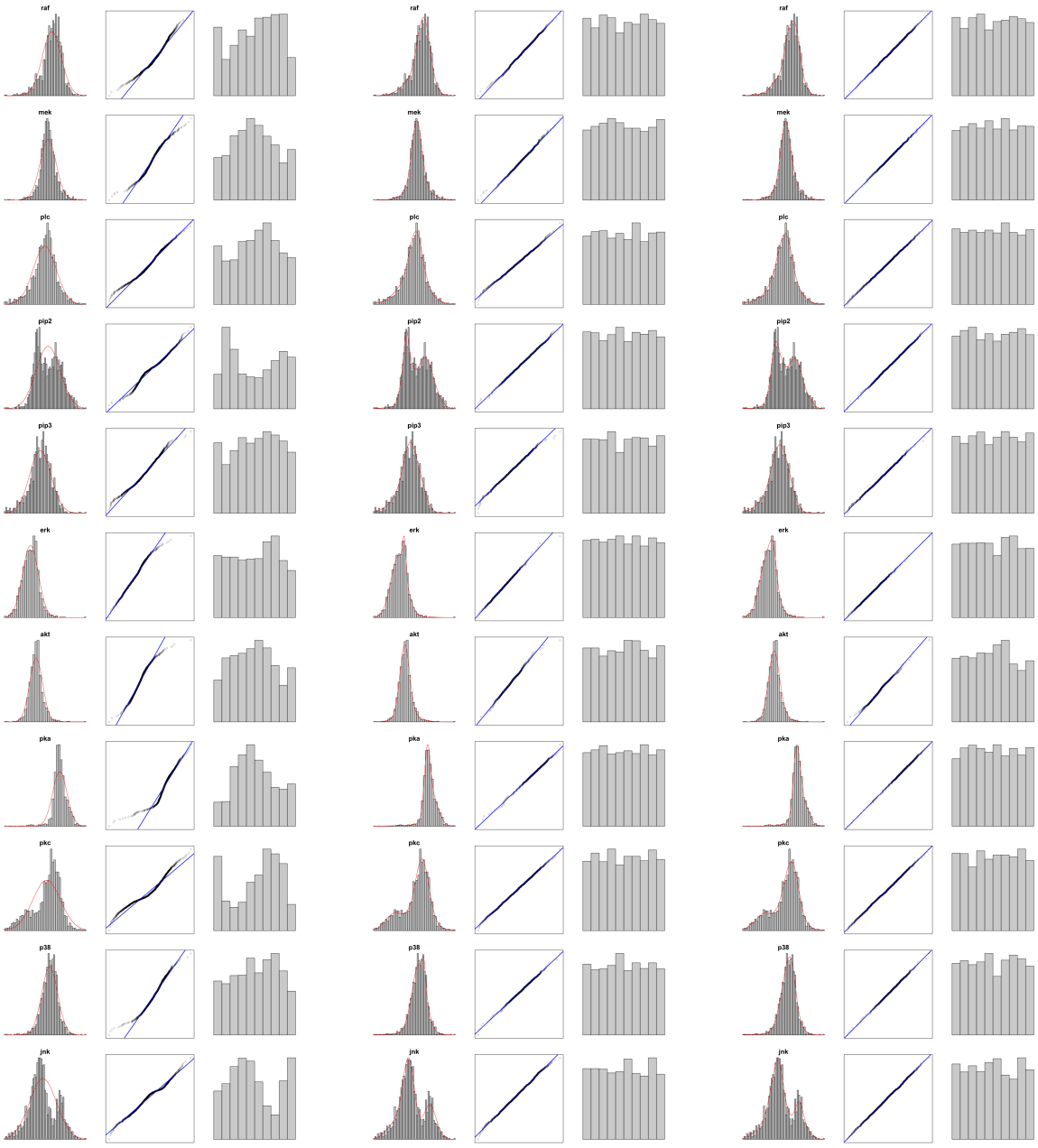



Figure 12.4: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted kernel density margins to the data from the $cd3cd28$ and $cd3cd28 + ly$ experiments




(a) $\mathbf{M}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ (b) $\mathbf{M}^{\mathbf{mix}}_{\mathbf{cd3\&cd3+ly}}$ (c) $\mathbf{M}^{\mathbf{ker}}_{\mathbf{cd3\&cd3+ly}}$

Figure 12.3: Density plot and Q-Q plot of the chosen margins (in the first two columns) and histogram of the data after applying the distribution function of the fitted marginal distributions, in the third column, for all nodes for the chosen marginals on the $cd3cd28$ and $cd3cd28 + ly$ data set

### 12.2.4 Comparison of Marginal Fits

As when we fitted the same three types of margins to the Sachs data set or the the data from the $cd3cd28 + aktinhib$ data set the results when fitting Gaussian mixture margins and kernel density estimates are very similar. Log-likelihood and $\text{AIC}_M$ almost coincide as well as the density plots and Q-Q plots. This holds for all nodes. What is notable here is the bigger difference in the number of (effective) parameters required for the two approaches. While overall fitting kernel density estimates required around double the amount of effective parameters as fitting Gaussian mixture margins we observe strong differences from node to node.

For example for the nodes $p38$ and $jnk$ fitting kernel density estimates requires only around three more effective parameters. For some nodes, e.g., $pip2$ and $pkc$, fitting Gaussian mixture margins even requires more parameters than fitting kernel density estimates. On the other hand the fitted kernel density estimates for the nodes $mek$ and $erk$ require 49.50 and 35.51 effective parameters while Gaussian mixture margins only need 5 and 8 parameters. Therefore, we observe notable differences in the $\text{BIC}_M$ for these two approaches. For both nodes the fitted Gaussian mixture margins provide a better $\text{BIC}_M$ than the fitted kernel density estimates. Further, when we look at the $\text{BIC}_M$ of all nodes together, we can see that the Gaussian mixture margins outperform the kernel density estimates.

As none of the nodes is modeled as a normal distribution when fitting Gaussian mixture margins it does not surprise that the fitted Gaussian margins have a worse fit in terms of the log-likelihood, $\text{AIC}_M$ and $\text{BIC}_M$ than the fitted Gaussian mixture margins and therefore also the fitted kernel density estimates. Looking at the graphical analysis we can see that especially the Q-Q plots show problems in the tails of most nodes. Overall leading to the assumption that fitting Gaussian margins does not seem reasonable.

Considering the normalized contour and scatter plots for the fitted Gaussian mixture margins and kernel density estimates we observe mostly the same shapes. However, compared to the marginal fitting, for example in Section 11.2 where the results were almost identical, this time more differences are present.

It seems that the fitted kernel density estimates result in smoother contour plots for example between the nodes $raf \leftrightarrow pka$ or $plc \leftrightarrow pka$. Even less smooth normalized contour plots can be observed when using the fitted Gaussian margins. Overall, if a pair of nodes has a non-elliptical shape using one of the three approaches to normalize it, it also has a non-elliptical shape using the other two. The only thing changing is the smoothness of the contours.

## 12.3 D-vine Regression ($\mathbf{MC_{cd3\&cd3+ly}}$)

### 12.3.1 Kernel Density Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$)

Transforming the data from the $cd3cd28$ and $cd3cd28 + ly$ experiments to the copula scale using the fitted kernel density margins and then fit a D-vine regression allowing for parametric and non-parametric copulas results in a log-likelihood of **4127.36**, an $AIC_C$ of **-7706.08** and a $BIC_C$ of **-6227.74** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pkc | 1.12 | -0.25 | 5.14 |
| mek | raf, pka | 622.28 | -1147.74 | -886.83 |
| plc | pip3 | 82.14 | -113.07 | 24.93 |
| pip2 | pip3, plc | 539.74 | -963.38 | -650.56 |
| erk | pka, mek | 269.16 | -502.23 | -404.98 |
| akt | erk, pka | 1234.55 | -2366.62 | -2090.52 |
| pka | pkc | 15.85 | -7.24 | 58.63 |
| pkc | plc, pip2 | 4.93 | -5.87 | 4.91 |
| p38 | pkc | 907.17 | -1755.05 | -1595.29 |
| jnk | pkc | 450.42 | -844.63 | -693.17 |
| $\sum$: | | 4127.36 | -7706.08 | -6227.74 |

Table 12.5: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model

Results per node are displayed in Table 12.5 and per copula in Table 12.6. There, we observe that of the twenty possible dependencies given by the DAG only fifteen are modeled. The five missing ones are $pka \to raf$, $pkc \to mek$, $pip3 \to akt$, $pka \to p38$ and $pka \to jnk$. Of the remaining fifteen eleven are modeled as a non-parametric copula. Contour plots of the fitted copulas can be found in Section 12.3.3 in Figure 12.5 together with the ones from the $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model and the $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pkc | frank | 1.00 | 1.12 | -0.25 | 5.14 | -0.01 |
| mek | mek raf | tll | 24.66 | 597.81 | -1146.30 | -1013.41 | 0.49 |
| | mek pka; raf | tll | 23.76 | 24.47 | -1.44 | 126.58 | -0.01 |
| plc | plc pip3 | tll | 25.61 | 82.14 | -113.07 | 24.93 | 0.02 |
| pip2 | pip2 pip3 | tll | 21.99 | 260.59 | -477.20 | -358.70 | 0.06 |
| | pip2 plc; pip3 | tll | 36.06 | 279.15 | -486.18 | -291.86 | 0.26 |
| erk | erk pka | tll | 17.05 | 267.94 | -501.79 | -409.93 | 0.04 |
| | erk mek; pka | gumbel | 1.00 | 1.22 | -0.44 | 4.95 | 0.01 |
| akt | akt erk | tll | 30.18 | 1041.44 | -2022.52 | -1859.87 | 0.60 |
| | akt pka; erk | tll | 21.05 | 193.10 | -344.10 | -230.65 | 0.16 |
| pka | pka pkc | tll | 12.22 | 15.85 | -7.24 | 58.63 | 0.01 |
| pkc | pkc plc | frank | 1.00 | 2.40 | -2.81 | 2.58 | -0.04 |
| | pkc pip2; plc | gumbel | 1.00 | 2.53 | -3.06 | 2.33 | -0.02 |
| p38 | p38 pkc | tll | 29.65 | 907.17 | -1755.05 | -1595.29 | 0.57 |
| jnk | jnk pkc | tll | 28.11 | 450.42 | -844.63 | -693.17 | -0.10 |

Table 12.6: Summary of all copulas fitted in the $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model

### 12.3.2 Gaussian Mixture Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$)

If we use the fitted Gaussian mixture margins from Section 12.2.2 for the probability integral transform instead, the fitted D-vine model has a log-likelihood of **4112.81**, an $AIC_C$ of **-7698.11** and a $BIC_C$ of **-6276.70** of the copula terms if we allow for parametric and non-parametric copulas.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pkc | 1.12 | -0.24 | 5.15 |
| mek | raf, pka | 621.25 | -1149.70 | -899.62 |
| plc | pip3 | 81.51 | -113.43 | 20.22 |
| pip2 | pip3, plc | 532.63 | -956.51 | -663.50 |
| erk | pka, mek | 268.99 | -502.19 | -405.72 |
| akt | erk, pka | 1238.08 | -2379.35 | -2118.47 |
| pka | pkc | 14.85 | -12.50 | 33.84 |
| pkc | plc, pip2 | 4.61 | -5.22 | 5.56 |
| p38 | pkc | 906.87 | -1755.51 | -1598.63 |
| jnk | pkc | 442.89 | -823.46 | -655.53 |
| $\sum$: | | 4112.81 | -7698.11 | -6276.70 |

Table 12.7: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model

Detailed results for each D-vine are displayed in Table 12.7 and the detailed results for each copula in Table 12.8. We observe that compared to using kernel density margins for the probability integral transform and then fit a D-vine with parametric and non-parametric copulas the same fifteen edges are present in the model with the same eleven being modeled as a non-parametric copula. The respective contour plots can be found in Figure 12.5 in Section 12.3.3.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pkc | frank | 1.00 | 1.12 | -0.24 | 5.15 | -0.01 |
| mek | mek raf | tll | 25.68 | 597.47 | -1143.58 | -1005.21 | 0.49 |
| | mek pka; raf | tll | 20.73 | 23.79 | -6.12 | 105.59 | -0.01 |
| plc | plc pip3 | tll | 24.80 | 81.51 | -113.43 | 20.22 | 0.02 |
| pip2 | pip2 pip3 | tll | 21.39 | 257.92 | -473.06 | -357.80 | 0.06 |
| | pip2 plc; pip3 | tll | 32.98 | 274.71 | -483.45 | -305.70 | 0.26 |
| erk | erk pka | tll | 16.90 | 267.58 | -501.35 | -410.27 | 0.04 |
| | erk mek; pka | gumbel | 1.00 | 1.42 | -0.84 | 4.55 | 0.01 |
| akt | akt erk | tll | 28.02 | 1044.72 | -2033.40 | -1882.41 | 0.60 |
| | akt pka; erk | tll | 20.39 | 193.36 | -345.95 | -236.06 | 0.16 |
| pka | pka pkc | tll | 8.60 | 14.85 | -12.50 | 33.84 | 0.01 |
| pkc | pkc plc | frank | 1.00 | 2.39 | -2.78 | 2.61 | -0.04 |
| | pkc pip2; plc | gumbel | 1.00 | 2.22 | -2.44 | 2.95 | -0.02 |
| p38 | p38 pkc | tll | 29.11 | 906.87 | -1755.51 | -1598.63 | 0.57 |
| jnk | jnk pkc | tll | 31.16 | 442.89 | -823.46 | -655.53 | -0.10 |

Table 12.8: Summary of all copulas fitted in the $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model

### 12.3.3 Gaussian Margins with Gaussian Copulas ($\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$)

Again as third approach using Gaussian margins for the probability integral transform and only allowing for Gaussian copulas in the D-vine regression results in a log-likelihood of **2005.63**, an $AIC_C$ of **-3989.28** and an $BIC_C$ of **-3929.99** of the copula terms.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | | | | |
| mek | raf | 469.32 | -936.63 | -931.24 |
| plc | pip3 | 2.93 | -3.86 | 1.53 |
| pip2 | pip3, plc | 133.85 | -263.70 | -252.92 |
| erk | pka | 63.61 | -125.23 | -119.84 |
| akt | erk, pka | 844.16 | -1684.32 | -1673.54 |
| pka | pkc | 1.87 | -1.74 | 3.65 |
| pkc | plc | 2.63 | -3.27 | 2.12 |
| p38 | pkc | 421.66 | -841.33 | -835.94 |
| jnk | pkc | 65.60 | -129.20 | -123.81 |
| $\sum$: | | 2005.63 | -3989.28 | -3929.99 |

Table 12.9: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model

In Table 12.9 more detailed results for each node can be found. As we observe in Table 12.10 only eleven of the twenty possible dependencies are modeled. Additional to the ones missing in the $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model also the dependencies $pkc \to raf$, $pka \to mek$, $mek \to erk$ and $pip2 \to pkc$ are missing. The contour plots are displayed in Figure 12.5.

| Node | Pair copula | Family | Copula log-lik. | $AIC_C$ | $BIC_C$ | Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| mek | mek raf | gaussian | 469.32 | -936.63 | -931.24 | 0.66 | 0.49 |
| plc | plc pip3 | gaussian | 2.93 | -3.86 | 1.53 | 0.06 | 0.02 |
| pip2 | pip2 pip3 | gaussian | 129.93 | -257.87 | -252.48 | 0.39 | 0.06 |
| | pip2 plc; pip3 | gaussian | 3.92 | -5.83 | -0.44 | 0.07 | 0.26 |
| erk | erk pka | gaussian | 63.61 | -125.23 | -119.84 | 0.21 | 0.04 |
| akt | akt erk | gaussian | 785.42 | -1568.85 | -1563.46 | 0.79 | 0.60 |
| | akt pka; erk | gaussian | 58.74 | -115.47 | -110.08 | 0.27 | 0.16 |
| pka | pka pkc | gaussian | 1.87 | -1.74 | 3.65 | 0.05 | 0.01 |
| pkc | pkc plc | gaussian | 2.63 | -3.27 | 2.12 | -0.06 | -0.04 |
| p38 | p38 pkc | gaussian | 421.66 | -841.33 | -835.94 | 0.67 | 0.57 |
| jnk | jnk pkc | gaussian | 65.60 | -129.20 | -123.81 | -0.28 | -0.10 |

Table 12.10: Summary of all copulas fitted in the $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model


(a) $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ (b) $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ (c) $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$

Figure 12.5: Contour plots of the respective copulas in the models allowing for parametric and non-parametric pair copulas where kernel density margins or Gaussian mixture margins were used for the PIT and the model where Gaussian margins and only Gaussian copulas were used on the $cd3cd28$ and $cd3cd28 + ly$ data set

### 12.3.4 Comparison of D-vine Regression Models

Comparing the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model we observe the same strong similarities as when comparing the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3+akt}}$ model. The same fifteen edges are modeled, all of the D-vines have the same order and all copulas have the same family, a similar number of (effective) parameters and goodness of fit (log-likelihood, $\text{AIC}_C$ and $\text{BIC}_C$). Hence, it does not surprise that there are no obvious differences in the contour plots in Figure 12.5.

Again restricting to Gaussian margins and Gaussian copulas unsurprisingly results in a notable decrease of the goodness of fit compared to the other two models. In the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model only eleven of the twenty possible dependencies are modeled. Most importantly both the $pka \to raf$ and $pkc \to raf$ dependencies are missing meaning the node $raf$ is modeled independent of all other nodes in this case. What is striking that in the DAG the node $pka$ is the one with the highest number of children, six, but in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model only one of these dependencies is modeled.

## 12.4 Goodness of Fit Measures Model Comparison

To calculate the goodness of fit measures for the whole D-vine regression models fitted to the $cd3cd28$ and $cd3cd28 + ly$ data set we sum up the measures of the copula terms and of the marginals. Only then we are able to compare the D-vine models to the Linear Gaussian Bayesian Network. The results for the log-likelihood are displayed in Table 12.11, for the $\text{AIC}_F$ in Table 12.12, for the $\text{BIC}_F$ in Table 12.13 and for the number of (effective) parameters used in the models in Table 12.14.

Comparing the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ we observe very similar results. This is again something we expected as both specify a normal distribution. Since we optimized the log-likelihood when fitting the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ it has a better log-likelihood than the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model for every node. The reverse does not hold. Not all nodes have a better $\text{AIC}_F$ in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model than in the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ even though the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model was fitted optimizing the $\text{AIC}_F$.

While most nodes have a better a better $\text{AIC}_F$ in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model compared to the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$, this does not hold for the nodes $erk$, $akt$ and $p38$. The reason is that in the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ the parameters are modeled all together in one step and in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model in two steps which can result in a worse fit. For the node $erk$ the difference is that big that it is the only one with a better $\text{BIC}_F$ in the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$.

As expected both the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model provide a better fit than the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$. Depending on the measure used for the comparison it changes which of the two is slightly better than the other one. When looking at the log-likelihood all nodes except $akt$ have a better fit in the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model. Instead looking at the $\text{AIC}_F$ only five nodes have a better fit in the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model leading to the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model having a slightly better fit when comparing based on the $\text{AIC}_F$. For both measures the differences are, taking the absolute value into account, almost vanishing.

Finally, looking at the $\text{BIC}_F$ all nodes except $pip3$ and $jnk$ have a better fit in the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model. This is due to the excessive amount of effective parameters needed to model the kernel density marginals in this case. Comparing the $\text{BIC}_F$ at least slight differences are present.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ | -1750 | -985 | -1793 | -2248 | -2071 | -1820 | -729 | -1782 | -1964 | -1018 | -2301 | -18460 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | -1659 | -726 | -1657 | -1740 | -2028 | -1574 | -261 | -1474 | -1790 | -456 | -1808 | -15172 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | -1666 | -736 | -1663 | -1749 | -2035 | -1576 | -250 | -1483 | -1793 | -458 | -1823 | -15232 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ | -1751 | -985 | -1793 | -2248 | -2071 | -1826 | -732 | -1782 | -1965 | -1022 | -2301 | -18474 |

Table 12.11: Global log-likelihood of the model fitted on the $cd3cd28$ and $cd3cd28 + ly$ data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ | 3507 | 1980 | 3593 | 4503 | 4145 | 3648 | 1468 | 3569 | 3937 | 2044 | 4610 | 37004 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | 3340 | 1648 | 3391 | 3617 | 4075 | 3254 | 639 | 3006 | 3599 | 995 | 3693 | 31256 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | 3345 | 1575 | 3386 | 3626 | 4086 | 3204 | 606 | 2999 | 3606 | 991 | 3725 | 31149 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ | 3505 | 1977 | 3593 | 4503 | 4145 | 3657 | 1472 | 3569 | 3936 | 2049 | 4608 | 37015 |

Table 12.12: $\text{AIC}_F$ of the models fitted on the $cd3cd28$ and $cd3cd28 + ly$ data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ | 3529 | 2007 | 3609 | 4525 | 4156 | 3669 | 1495 | 3591 | 3958 | 2065 | 4632 | 37236 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | 3399 | 2175 | 3596 | 3986 | 4126 | 3543 | 958 | 3161 | 3647 | 1221 | 3899 | 33711 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | 3377 | 1852 | 3547 | 3978 | 4129 | 3344 | 894 | 3088 | 3660 | 1191 | 3936 | 32996 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ | 3516 | 1993 | 3609 | 4525 | 4156 | 3673 | 1493 | 3585 | 3952 | 2065 | 4624 | 37193 |

Table 12.13: $\text{BIC}_F$ of the models fitted on the $cd3cd28$ and $cd3cd28 + ly$ data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ | 4 | 5 | 3 | 4 | 2 | 4 | 5 | 3 | 4 | 4 | 4 | 42 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | 11 | 98 | 38 | 69 | 10 | 54 | 59 | 29 | 9 | 42 | 38 | 456 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ | 6 | 51 | 30 | 65 | 8 | 26 | 53 | 17 | 10 | 37 | 39 | 343 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ | 2 | 3 | 3 | 3 | 2 | 3 | 4 | 3 | 3 | 3 | 3 | 33 |

Table 12.14: Number of (effective) parameters of the models fitted on the $cd3cd28$ and $cd3cd28 + ly$ data set for each node

## 12.5 Simulation Based Comparison

For a further analysis we again followed the procedure as described in Chapter 8 and sampled 1618 times from the four fitted models to analyze how the models are able to recreate the data set they are fitted on. First, we fit kernel density estimates to the data simulated from each node of the four models and investigate how they behave compared to the kernel density estimates fitted to each node in the $cd3cd28$ and $cd3cd28 + ly$ data set.

The results are displayed in Figure 12.6. We have already stated when we previously analyzed these plots that the quality of the fit almost only depends on the fit of the marginals. Hence, as expected the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model outperform the other two. This holds especially for the nodes where the histograms are not normally distributed, for example $pip2$ and $jnk$. It is not possible to prefer one of the two models over the other, as depending on the specific node it changes which one is closer to the original data.

As we are more interested in the analysis of the dependence structures we analyze the pairs plots next. The results for the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model can be observed in Figure 12.7, as well as for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$ in Figure 12.8.

In Figure 12.7 we can see that the pairs plots of the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model are almost identical. Comparing them to the pairs plots of the $cd3cd28$ and $cd3cd28 + ly$ data set we derive very similar shapes for all dependencies and not only the ones that are directly modeled. This also holds for more complicated shapes like between the nodes $raf$ and $mek$ or $mek$ and $pka$.

As the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model models each dependency as a Gaussian copula the respective pairs plots each have the typical elliptical form. Therefore, it is clear that the model has problems recreating more complicated forms. Looking at Figure 12.8 we can see that the pairs plots from the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model seem to be more similar to the ones from the $cd3cd28$ and $cd3cd28 + ly$ data set than the ones from the $\text{LGBN}_{\mathbf{cd3\&cd3+ly}}$. This can for example be deduced at the plots for the nodes $raf$ and $mek$. Both models are still way worse in recreating the original pairs plots compared to the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model.

Finally, we are looking at the histograms of the sum of each node and the set of its parents as given by the DAG. If the models fit well, high/low values at several nodes at once should appear in a similar frequency as in the original data set. In that case the histograms of the sums should be similar. Figure 12.9 shows the results. It seems that the

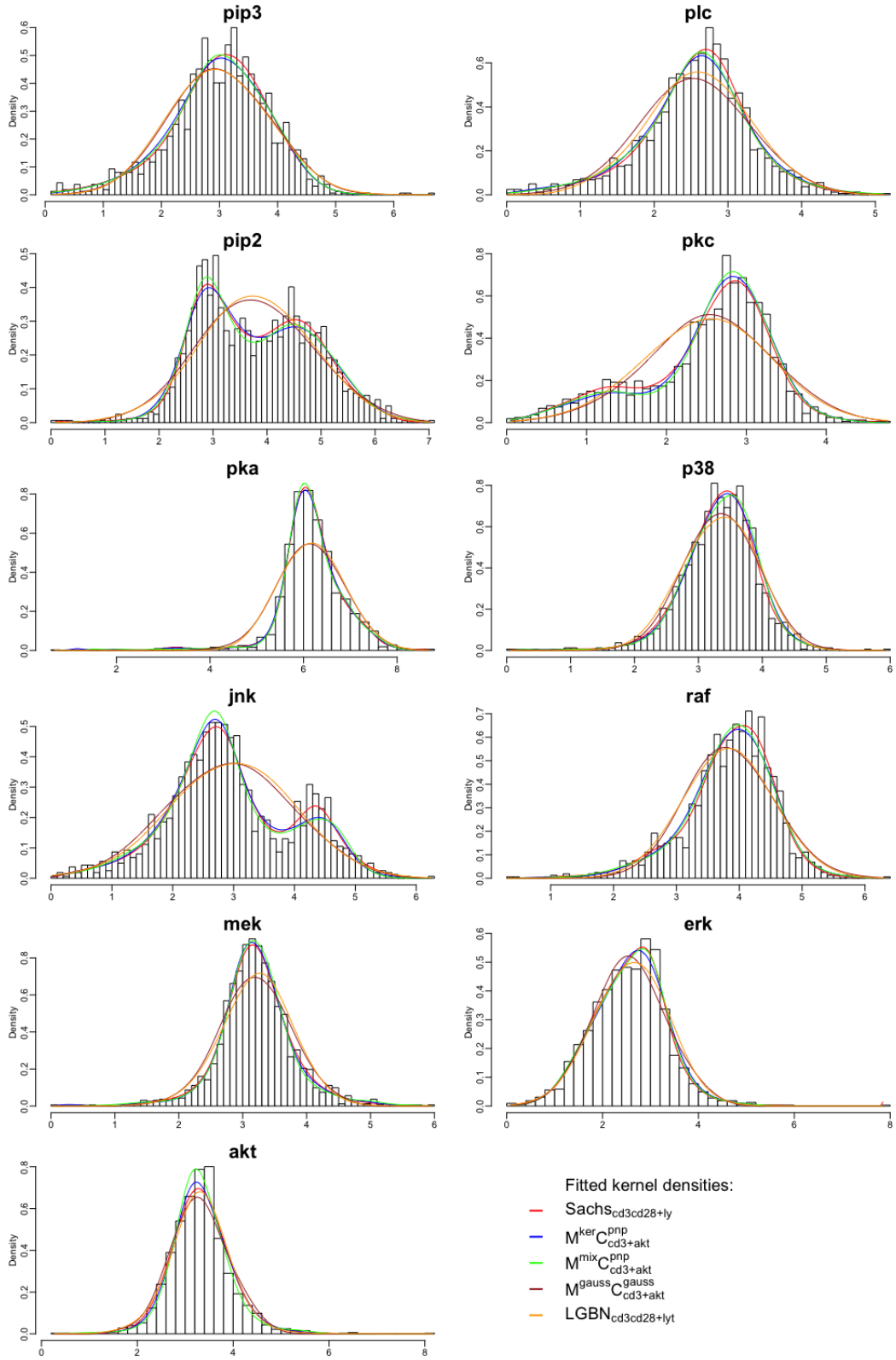
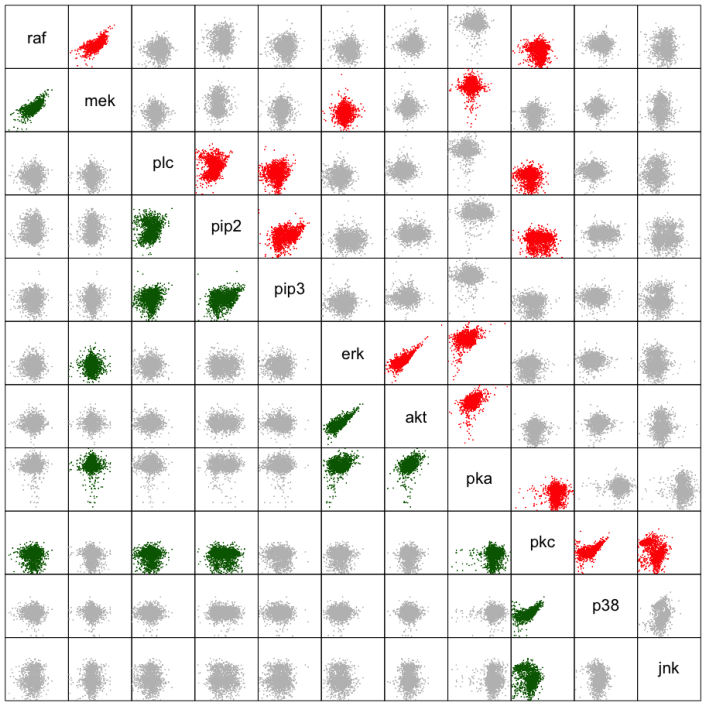
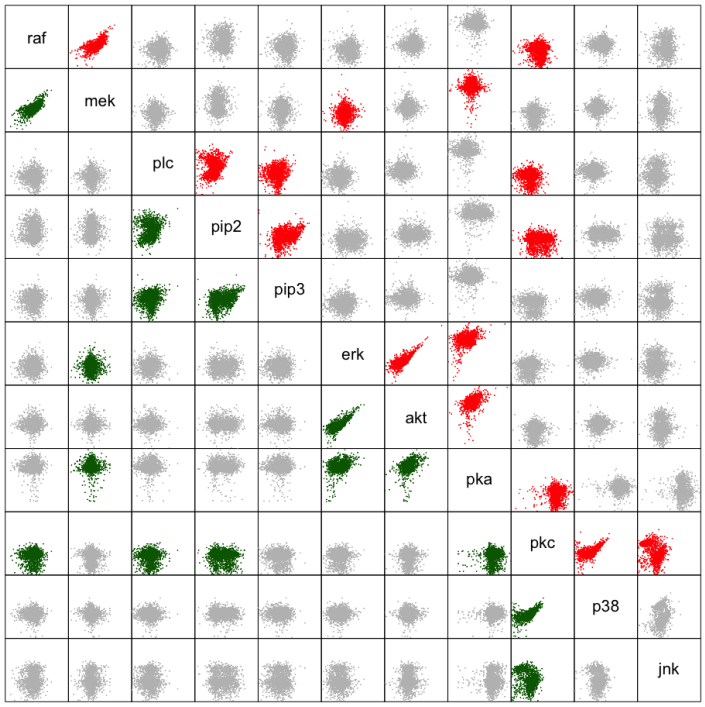



Figure 12.6: Marginal histograms of the $cd3cd28$ and $cd3cd28 + ly$ data set of each of the eleven nodes in topological order after samples containing a zero value have been deleted and kernel density estimates have been fitted to the simulated data from the different models

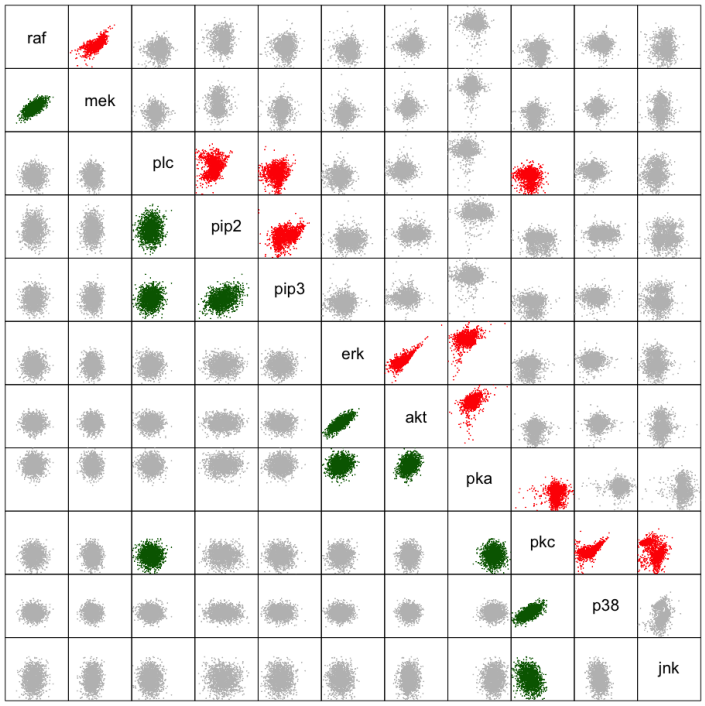



(a) $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$

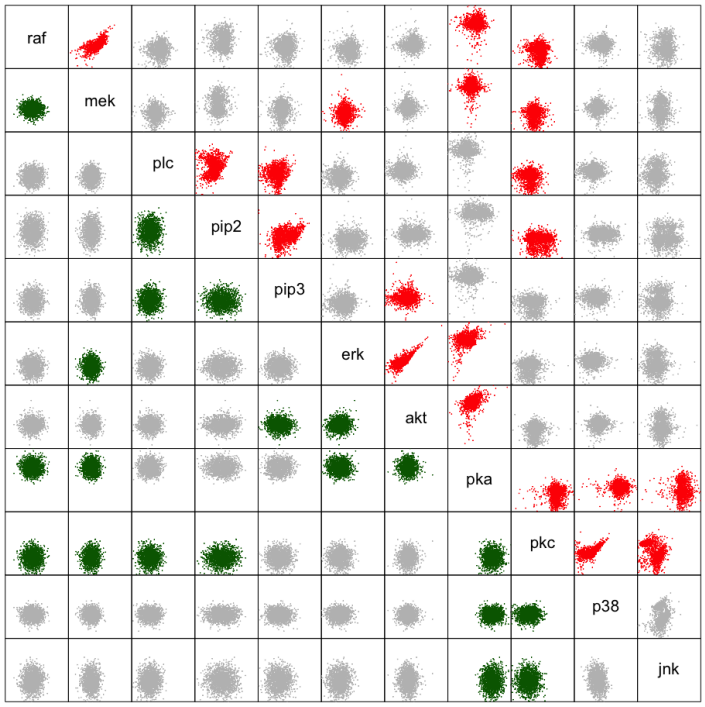



(b) $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$

Figure 12.7: Pairs plots for each pair of nodes of the respective model in the lower panel and of the *cd3cd28* and *cd3cd28* + *ly* data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel




(a) $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$




(b) $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$

Figure 12.8: Pairs plots for each pair of nodes of the respective model in the lower panel and of the *cd*3*cd*28 and *cd*3*cd*28 + *ly* data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel

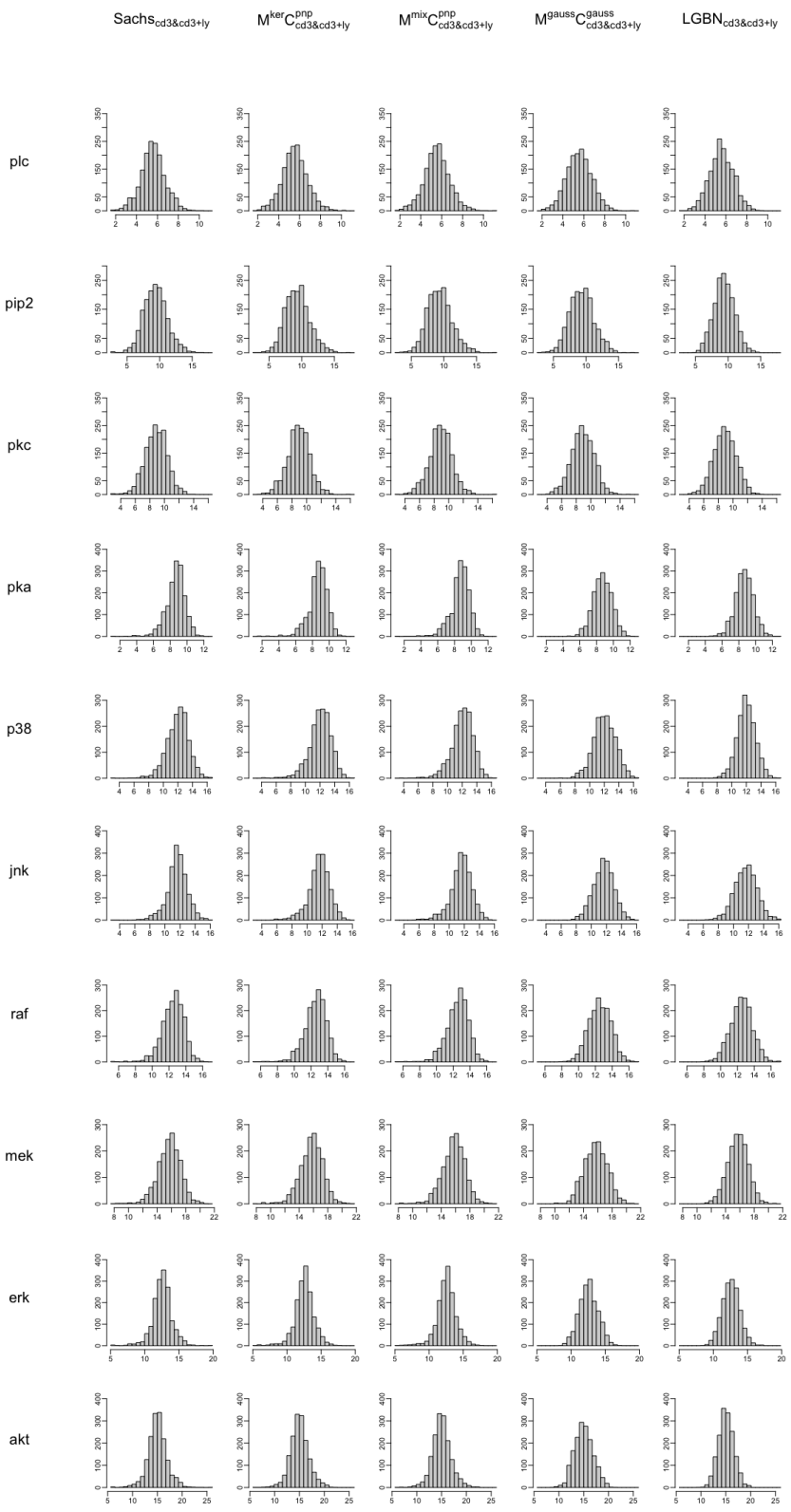



Figure 12.9: Histograms of the sum of each node on the set of its parents according to the consent DAG in the $cd3cd28$ and $cd3cd28 + ly$ data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model, $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model, $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model and the $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$

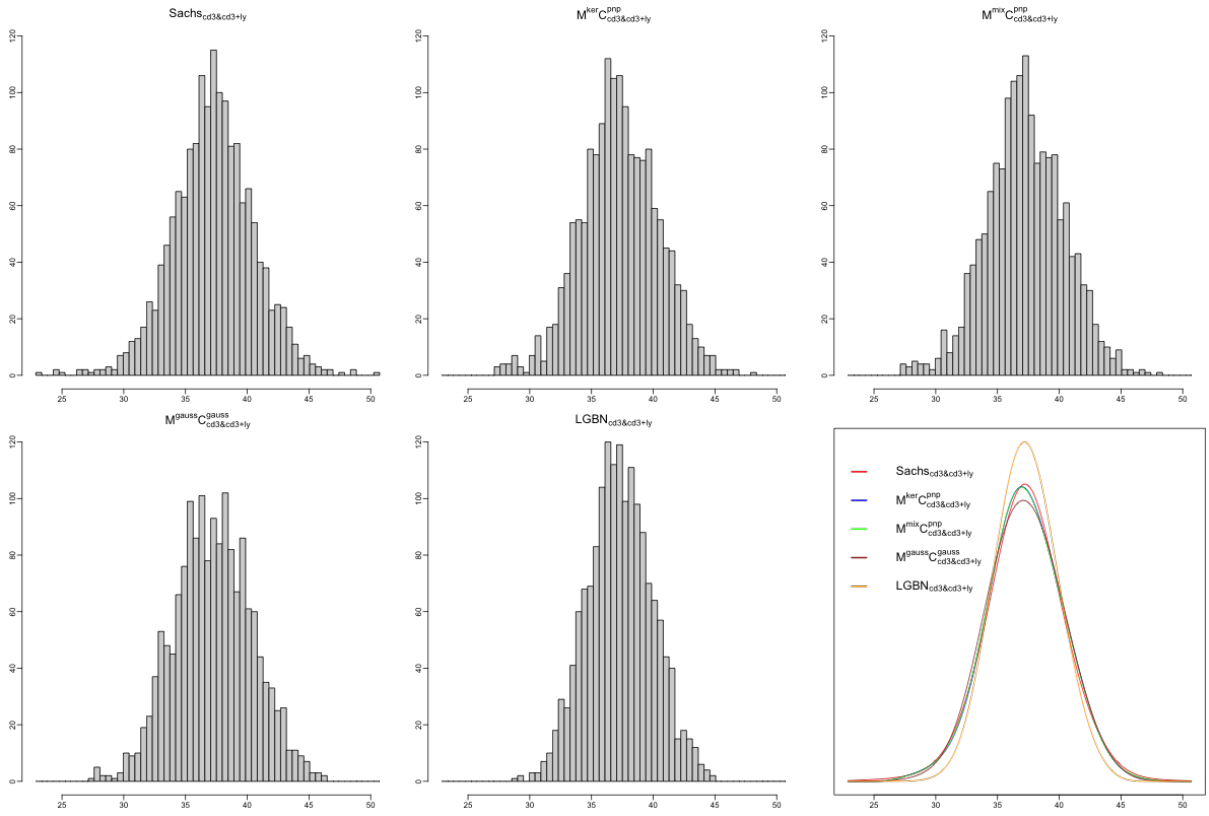



Figure 12.10: Histograms of the sum over all nodes in the $cd3cd28$ and $cd3cd28{+}ly$ data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model, $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model, $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model and the $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$ in the first five plots and fitted kernel density estimates to each of the data sets in the sixth plot

histograms of both the $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model are quite close to the ones from the $cd3cd28$ and $cd3cd28 + ly$ data set for all nodes. Looking at the $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model instead the histograms of some nodes are still close to the original ones, for example for the nodes $plc$ or $pip2$. However, the more non-symmetric the histograms of the underlying data are the more problems the $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model gets at recreating them. For the $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$ almost no similarities with the $cd3cd28$ and $cd3cd28 + ly$ data can be seen.

Looking at the sum over all nodes in Figure 12.10 we observe the same results as when we have been considering the sums according to the set of the parents for each node. The $\mathbf{M^{ker}C^{pnp}_{cd3\&cd3+ly}}$ model and the $\mathbf{M^{mix}C^{pnp}_{cd3\&cd3+ly}}$ model seem to resemble the data from the $cd3cd28$ and $cd3cd28 + ly$ data set quite good. On the other hand, while for the $\mathbf{M^{gauss}C^{gauss}_{cd3\&cd3+ly}}$ model at least some similarities exist, this is not the case for the $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$.

## 12.6 Conditional Simulation Based Comparison

To end the comparison of the models fitted to the $cd3cd28$ and $cd3cd28 + ly$ data set we want to compare how the conditional densities on the different models behave if they are conditioned on different values. This is not done in order to compare the models to the data set they are fitted on, but rather to later be able to compare how the models fitted on different data sets behave. Therefore, we condition the models which we fitted on the $cd3cd28$ and $cd3cd28 + ly$ data set using the nodes we specified in Chapter 9 in Table 9.1.

The results are given in Figure 12.11. As before we observe that the results for the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model are extremely similar. For almost all plots the fitted kernel density estimates on the sampled data from both models are identical. This holds even for not unimodal shapes like for example the plots for the node $pkc$. Only when we condition on points from the tails small differences for the node $pip2$ are visible.

Considering the fitted kernel density estimates of the samples from the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model, we observe that while the results are still comparable to the ones from the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model for the node $raf$ this is not the case for any other node. For all other nodes at least one of the fitted kernel density estimates differs strongly from the ones fitted to the data from the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$ model. Looking at the kernel density estimates fitted to the data simulated from the $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$ we can see no clear similarities to any other model.

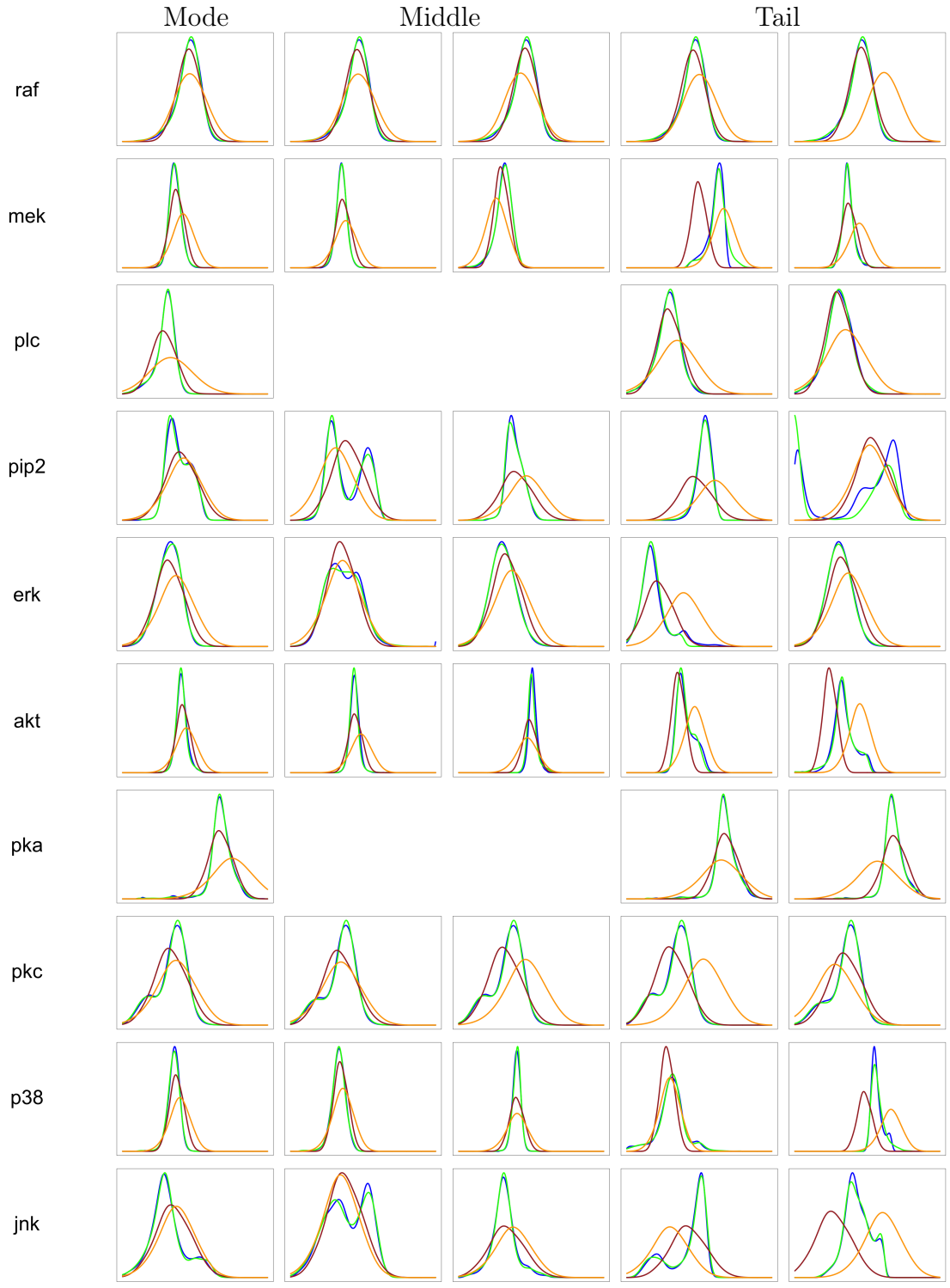



Figure 12.11: Density plots of the kernel density estimates fitted to the simulated data from the different models conditioned on the nodes in Table 9.1 in the same order. Blue: $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$. Green: $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{cd3\&cd3+ly}}$. Brown: $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$. Orange: $\mathrm{LGBN}_{\mathbf{cd3\&cd3+ly}}$

# Chapter 13

# Analysis of Group 3: cd3cd28icam2 and pma

In this chapter we will investigate the data of the *cd*3*cd*28*icam*2 and *pma* experiments using three types of D-vine models and a Linear Gaussian Bayesian Network.

## 13.1 Linear Gaussian Bayesian Network Fitting ($\text{LGBN}_{\text{icam2\&pma}}$)

Fitting a Linear Gaussian Bayesian Network on the data from the *cd*3*cd*28*icam*2 and *pma* experiments results in a log-likelihood of **-18200.39**, an $\text{AIC}_F$ of **36484.75** and a $\text{BIC}_F$ of **36719.50** of the whole model. Detailed results for each node are displayed in Table 13.1.

| Node | Global log-likelihood | # Parameters | $\text{AIC}_F$ | $\text{BIC}_F$ |
|---|---|---|---|---|
| raf | -2005.29 | 4 | 4018.57 | 4040.41 |
| mek | -1070.10 | 5 | 2150.20 | 2177.50 |
| plc | -1481.60 | 3 | 2969.19 | 2985.57 |
| pip2 | -2147.33 | 4 | 4302.66 | 4324.50 |
| pip3 | -2410.15 | 2 | 4824.29 | 4835.21 |
| erk | -2038.28 | 4 | 4084.57 | 4106.40 |
| akt | -419.79 | 5 | 849.59 | 876.88 |
| pka | -1772.69 | 3 | 3551.37 | 3573.21 |
| pkc | -1968.04 | 4 | 3944.07 | 3965.91 |
| p38 | -1105.76 | 4 | 2219.51 | 2241.35 |
| jnk | -1781.36 | 4 | 3570.73 | 3592.56 |
| $\sum$: | -18200.39 | 42 | 36484.75 | 36719.50 |

Table 13.1: Global log-likelihood, number of parameters, $\text{AIC}_F$ and $\text{BIC}_F$ of the fitted linear models for each node given the set of its parents in the Linear Gaussian Bayesian Network on the data set from the *cd*3*cd*28*icam*2 and *pma* experiments

## 13.2 Marginal Fitting ($\mathbf{M_{icam2\&pma}}$)

### 13.2.1 Gaussian Margin Fitting ($\mathbf{M^{gauss}_{icam2\&pma}}$)

Again fitting Gaussian margins, this time to the *cd*3*cd*28*icam*2 and *pma* data set, assuming independence in the data, results in a log-likelihood of the marginals of **-22794.69**, an $AIC_M$ of **45633.37** and a $BIC_M$ of **45753.48**. The detailed results for each node can be found in Table 13.2.

| Node | Marginal log-likelihood | $AIC_M$ | $BIC_M$ |
|---|---|---|---|
| raf | -2006.54 | 4017.08 | 4028.00 |
| mek | -1990.81 | 3985.61 | 3996.53 |
| plc | -1704.21 | 3412.41 | 3423.33 |
| pip2 | -2762.70 | 5529.41 | 5540.33 |
| pip3 | -2410.15 | 4824.29 | 4835.21 |
| erk | -2301.77 | 4607.55 | 4618.46 |
| akt | -1968.98 | 3941.97 | 3952.89 |
| pka | -1772.70 | 3549.39 | 3560.31 |
| pkc | -1968.31 | 3940.62 | 3951.54 |
| p38 | -1753.96 | 3511.92 | 3522.84 |
| jnk | -2154.56 | 4313.12 | 4324.04 |
| $\sum$: | -22794.69 | 45633.37 | 45753.48 |

Table 13.2: Marginal log-likelihood, $AIC_M$ and $BIC_M$ of the fitted Gaussian margins $\mathbf{M^{gauss}_{icam2\&pma}}$ to the data set from the *cd*3*cd*28*icam*2 and *pma* experiments

We observe in panel (a) of Figure 13.3 that assuming an underlying Gaussian distribution for the node *erk* seems reasonable. For this node especially the histogram of the data after applying the PIT with the fitted Gaussian distribution is approximately uniformly distributed. This does not hold for any other node. Therefore, the necessary requirements to fit copulas to the data in combination with Gaussian margins are not fulfilled. Figure 13.1 shows a further graphical analysis of the data.


Figure 13.1: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian margins to the data from the *cd*3*cd*28*icam*2 and *pma* experiments

### 13.2.2 Gaussian Mixture Margin Fitting ($\mathbf{M^{mix}_{icam2\&pma}}$)

Fitting Gaussian mixture margins to the data set from the *cd*3*cd*28*icam*2 and *pma* experiments results in a log-likelihood of the marginals of **-21727.79**, an $AIC_M$ of **43625.60** and a $BIC_M$ of **44089.62** assuming independence in the data. As we can see in Table 13.3 all nodes except of four, i.e., *pip*2, *pip*3, *erk* and *akt*, are fitted as a three component normal distribution. Out of these only the node *erk* is fitted as a Gaussian distribution.

| Node | Distribution | Marginal log-lik. | # Parameters | $AIC_M$ | $BIC_M$ |
|---|---|---|---|---|---|
| raf | three component normal | -1966.20 | 8 | 3948.4 | 3992.07 |
| mek | three component normal | -1937.97 | 8 | 3891.94 | 3935.62 |
| plc | three component normal | -1557.11 | 8 | 3130.22 | 3173.89 |
| pip2 | four component normal | -2500.14 | 11 | 5022.29 | 5082.34 |
| pip3 | four component normal | -2367.66 | 11 | 4757.32 | 4817.37 |
| erk | normal | -2301.77 | 2 | 4607.55 | 4618.46 |
| akt | two component normal | -1922.48 | 5 | 3854.96 | 3882.25 |
| pka | three component normal | -1611.80 | 8 | 3239.60 | 3283.28 |
| pkc | three component normal | -1863.56 | 8 | 3743.11 | 3786.79 |
| p38 | three component normal | -1648.78 | 8 | 3313.57 | 3357.24 |
| jnk | three component normal | -2050.32 | 8 | 4116.64 | 4160.31 |
| $\sum$: | | -21727.79 | 85 | 43625.60 | 44089.62 |

Table 13.3: Marginal log-likelihood, number of parameters, $AIC_M$ and $BIC_M$ of the fitted Gaussian mixture margins $\mathbf{M^{mix}_{icam2\&pma}}$ to the data set from the *cd*3*cd*28*icam*2 and *pma* experiments

Looking at panel (b) of Figure 13.3 we observe no clear evidence that the fitted distributions are incorrect. Especially the histograms of the data after applying the PIT with the fitted Gaussian mixture margins seem to follow a uniform distribution. Normalized contour and scatter plots as well as the histograms of the data after applying the PIT are displayed in Figure 13.2.


Figure 13.2: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted Gaussian mixture margins to the data from the *cd*3*cd*28*icam*2 and *pma* experiments

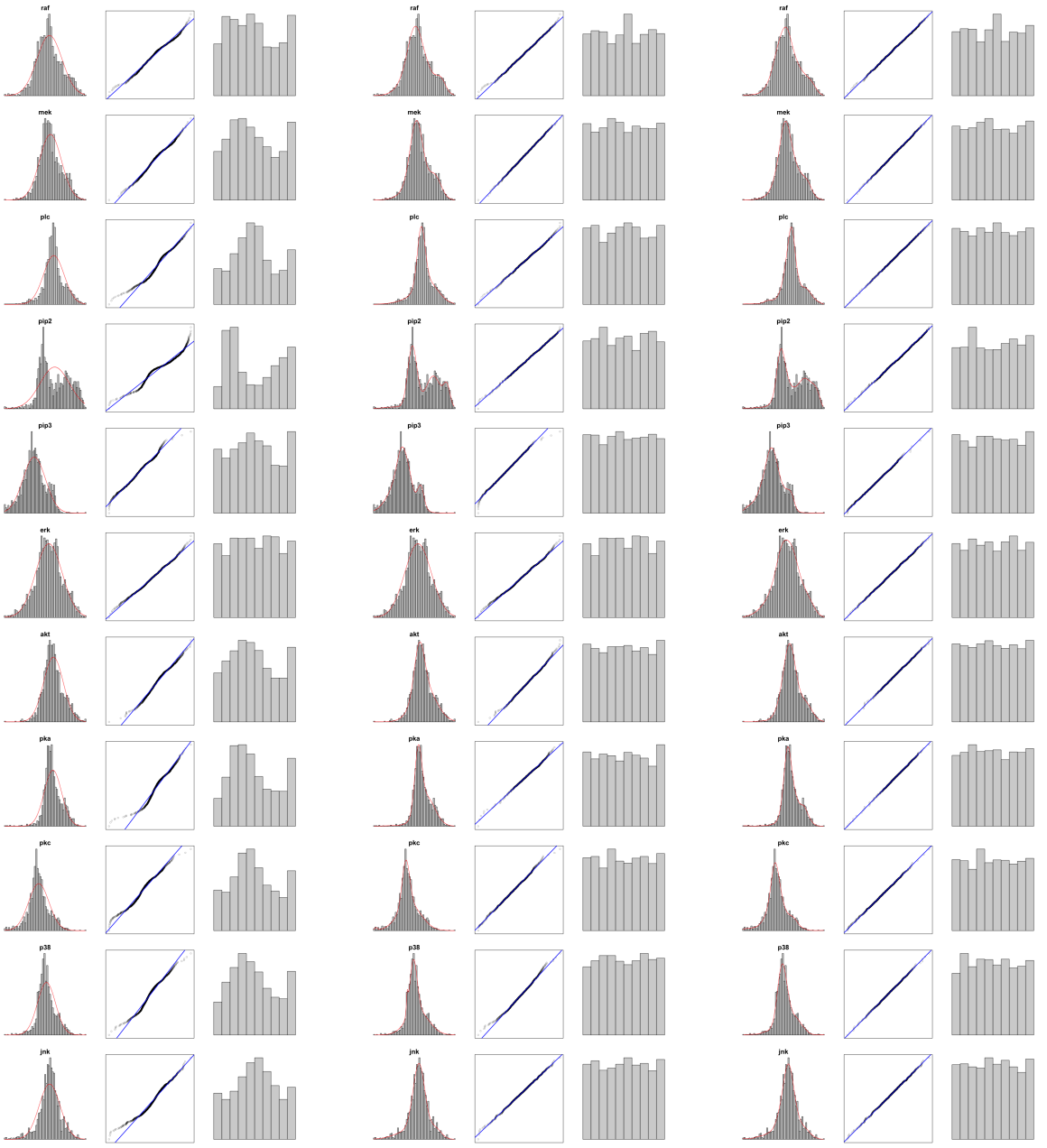
### 13.2.3 Kernel Density Margin Fitting ($\mathbf{M}^{\mathbf{ker}}_{\mathbf{icam2\&pma}}$)

Finally, fitting kernel density estimates to the data from the *cd*3*cd*28*icam*2 and *pma* experiments, assuming independence, results in a log-likelihood of the marginals of **-21630.72**, an $\text{AIC}_M$ of **43574.47** and a $\text{BIC}_M$ of **44428.90**. As we derive in Table 13.4 the log-likelihood and the $\text{AIC}_F$ for each node are almost the same as in the approach where we used Gaussian mixture margins to model the marginals. On the other hand, compared to the Gaussian mixture margins, we observe small differences when considering the $\text{BIC}_F$. This is due to the higher number of effective parameters needed to model the kernel density estimates.

| Node | Distribution | Marginal log-likelihood | Effective # parameters | $\text{AIC}_M$ | $\text{BIC}_M$ |
|---|---|---|---|---|---|
| raf | kernel density | -1958.42 | 13.06 | 3942.95 | 4014.24 |
| mek | kernel density | -1936.74 | 20.98 | 3915.44 | 4029.99 |
| plc | kernel density | -1544.03 | 14.56 | 3117.18 | 3196.69 |
| pip2 | kernel density | -2499.00 | 16.16 | 5030.32 | 5118.54 |
| pip3 | kernel density | -2355.16 | 10.55 | 4731.43 | 4789.04 |
| erk | kernel density | -2286.60 | 12.10 | 4597.41 | 4663.46 |
| akt | kernel density | -1909.99 | 11.36 | 3842.70 | 3904.70 |
| pka | kernel density | -1602.01 | 13.97 | 3231.96 | 3308.23 |
| pkc | kernel density | -1855.62 | 14.55 | 3740.35 | 3819.81 |
| p38 | kernel density | -1636.79 | 15.29 | 3304.16 | 3387.63 |
| jnk | kernel density | -2046.36 | 13.92 | 4120.57 | 4196.57 |
| $\sum$: | | -21630.72 | 156.50 | 43574.47 | 44428.90 |

Table 13.4: Marginal log-likelihood, number of effective parameters, $\text{AIC}_M$ and $\text{BIC}_M$ of the fitted kernel density margins $\mathbf{M}^{\mathbf{ker}}_{\mathbf{icam2\&pma}}$ to the data set from the *cd*3*cd*28*icam*2 and *pma* experiments

Considering the density plots and Q-Q plots in panel (c) of Figure 13.3 we can see no indications that the fitted kernel density estimates are not suitable. The same holds for the histograms of the data after applying the PIT with the fitted kernel density estimates which seem to be approximately uniformly distributed. Additionally normalized contour and scatter plots are displayed in Figure 13.4.




(a) $\mathbf{M}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ (b) $\mathbf{M}^{\mathbf{mix}}_{\mathbf{icam2\&pma}}$ (c) $\mathbf{M}^{\mathbf{ker}}_{\mathbf{icam2\&pma}}$

Figure 13.3: Density plots and Q-Q plot of the chosen margins (in the first two columns) and histogram of the data after applying the distribution function of the fitted marginal distributions, in the third column, for all nodes for the chosen marginals on the *cd*3*cd*28*icam*2 and *pma* data set

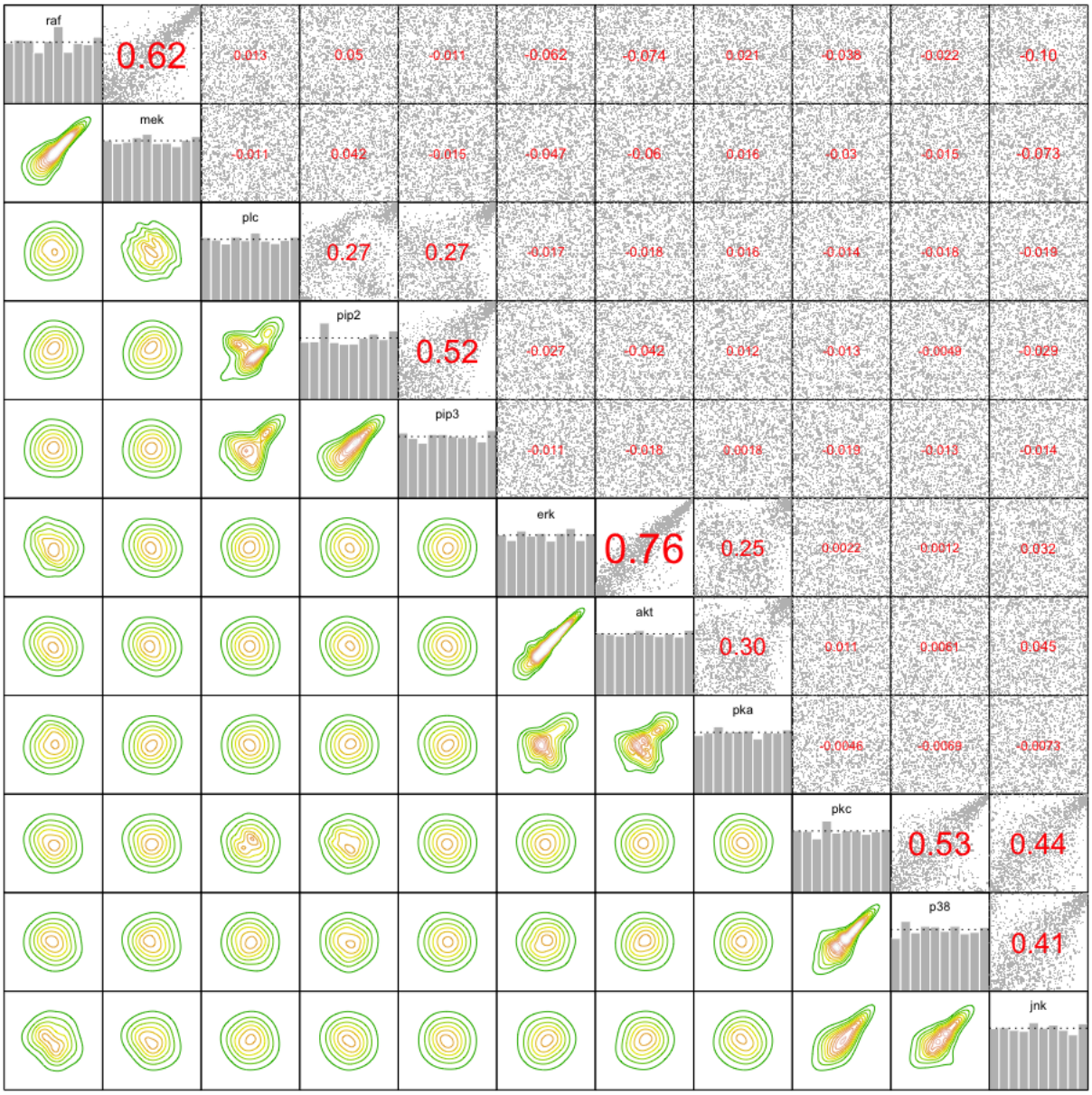



Figure 13.4: Normalized contour plots in the lower left triangle, normalized scatter plots in the upper right triangle, with the value of Kendall's tau displayed in the middle of the plot, and histograms in the diagonal elements after applying the PIT using the fitted kernel density margins to the data from the $cd3cd28icam2$ and $pma$ experiments

### 13.2.4 Comparison of Marginal Fits

Fitting the same three types of margins as before, this time to the $cd3cd28icam2$ and $pma$ data set, we again observe similar results. The fitted kernel density estimates and Gaussian mixture margins have an almost identical log-likelihood, $\text{AIC}_M$ and and a similar $\text{BIC}_M$. Looking at their density plots and Q-Q plots it is hard to spot differences that would point to preferring one over the other. This also holds when looking at the histograms of the data after applying the probability integral transform. For both choices of margins the data is approximately uniformly distributed after performing the PIT.

Fitting kernel density estimates overall requires about twice the amount of effective parameters as fitting Gaussian mixture margins. These numbers strongly depend on the explicit node. For example the fitted kernel density estimate to the node $pip3$ has even less effective parameters than the Gaussian mixture margin fitted to the node. On the other hand when fitting Gaussian mixture margins, the node $erk$ is modeled as a normal distribution and therefore has only two parameters while the fitted kernel density estimate for this node requires 12.10 effective parameters. Due to the still small absolute amount of (effective) parameters and the size of the data set the influence on the $\text{AIC}_C$ is almost vanishing if we take the absolute value into account.

When fitting Gaussian mixture margins the node $erk$ is modeled as a normal distribution. It therefore coincides with how it is fitted when fitting Gaussian margins. For all other nodes restricting to only Gaussian margins comes with a worsening in the results compared to the other two approaches. This holds for the log-likelihood, $\text{AIC}_M$ and $\text{BIC}_M$ as well as for the graphical analysis. Again especially the data after applying the PIT with the fitted Gaussian margins is not uniformly distributed. Therefore, the requirements to fit copulas to the data are not fulfilled.

Looking at the normalized contour and scatter plots in the approaches with Gaussian mixture margins and kernel density estimates it is hard to observe any differences. Comparing them to the ones where Gaussian margins are used, it seems that there the contours are not as smooth as in the other two approaches. However, the general form of the shapes is the same in all three.

## 13.3 D-vine Regression ($\mathbf{MC_{icam2\&pma}}$)

### 13.3.1 Kernel Density Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$)

Using the fitted kernel density margins to transform the data from the *cd*3*cd*28*icam*2 and *pma* experiments to the copula scale using the PIT with the fitted kernel density margins and then fit a D-vine regression model allowing for parametric and non-parametric copulas results in a log-likelihood of **6955.73**, an $AIC_C$ of **-13275.60** and a $BIC_C$ of **-11539.93** of the copula terms. Detailed results for each node are displayed in Table 13.5.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 29.88 | -16.73 | 100.73 |
| mek | raf | 1220.48 | -2372.48 | -2185.56 |
| plc | pip3 | 542.91 | -1033.12 | -889.30 |
| pip2 | pip3, plc | 949.67 | -1838.96 | -1674.15 |
| erk | pka, mek | 452.58 | -792.00 | -483.11 |
| akt | erk, pka | 2040.21 | -3953.44 | -3606.82 |
| pka | | | | |
| pkc | pip2 | 30.74 | -5.99 | 145.48 |
| p38 | pkc | 1063.79 | -2073.06 | -1924.25 |
| jnk | pkc | 625.48 | -1189.82 | -1022.95 |
| $\sum$: | | 6955.73 | -13275.60 | -11539.93 |

Table 13.5: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ model

As we observe in Table 13.6 thirteen of the twenty possible dependencies are modeled, all but two of them as a non-parametric copula. The missing dependencies are the $pka \to mek$, $pkc \to mek$, $pip3 \to akt$, $pkc \to pka$, $plc \to pkc$, $pka \to p38$ and $pka \to jnk$ dependency. Contour plots of the fitted copulas can be found in Section13.3.3 in Figure 13.5 together with the ones from the $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model and the $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ model.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | tll | 20.52 | 27.25 | -13.47 | 98.53 | 0.02 |
| | raf pkc; pka | frank | 1.00 | 2.63 | -3.26 | 2.20 | -0.04 |
| mek | mek raf | tll | 34.24 | 1220.48 | -2372.48 | -2185.56 | 0.62 |
| plc | plc pip3 | tll | 26.34 | 542.91 | -1033.12 | -889.30 | 0.27 |
| pip2 | pip2 pip3 | bb8 | 2.00 | 764.85 | -1525.71 | -1514.79 | 0.26 |
| | pip2 plc; pip3 | tll | 28.19 | 184.81 | -313.25 | -159.36 | 0.52 |
| erk | erk pka | tll | 23.35 | 412.13 | -777.56 | -650.06 | 0.25 |
| | erk mek; pka | tll | 33.23 | 40.44 | -14.44 | 166.95 | -0.05 |
| akt | akt erk | tll | 42.53 | 1858.63 | -3632.19 | -3400.01 | 0.75 |
| | akt pka; erk | tll | 20.96 | 181.59 | -321.25 | -206.81 | 0.30 |
| pkc | pkc pip2 | tll | 27.75 | 30.74 | -5.99 | 145.48 | -0.01 |
| p38 | p38 pkc | tll | 27.26 | 1063.79 | -2073.06 | -1924.25 | 0.53 |
| jnk | jnk pkc | tll | 30.57 | 625.48 | -1189.82 | -1022.95 | 0.43 |

Table 13.6: Summary of all copulas fitted in the $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ model

### 13.3.2 Gaussian Mixture Margins with Parametric and Non-Parametric Copulas ($\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$)

As a second approach on the *cd*3*cd*28*icam*2 and the *pma* data set we used the fitted Gaussian mixture margins from Section 13.2.2 for the probability integral transform. The fitted D-vine model using this transformation has a log-likelihood of **6896.51**, an $AIC_C$ of **-13171.49** and a $BIC_C$ of **-11474.94** of the copula terms if we allow for parametric and non-parametric copulas. Results for each D-vine can be found in Table 13.7 and the detailed results for each copula in Table 13.8.

| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pka, pkc | 28.70 | -19.87 | 82.58 |
| mek | raf | 1208.66 | -2345.68 | -2150.11 |
| plc | pip3 | 536.89 | -1023.96 | -887.94 |
| pip2 | pip3, plc | 941.82 | -1824.66 | -1663.67 |
| erk | pka, mek | 449.19 | -792.71 | -504.26 |
| akt | erk, pka | 2024.51 | -3916.92 | -3556.31 |
| pka | | | | |
| pkc | pip2 | 28.58 | -5.68 | 134.83 |
| p38 | pkc | 1055.53 | -2054.46 | -1899.96 |
| jnk | pkc | 622.62 | -1187.55 | -1030.10 |
| $\sum$: | | 6896.51 | -13171.49 | -11474.94 |

Table 13.7: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model

Again we can see that compared to when using kernel density margins for the probability integral transform and then fit a D-vine with parametric and non-parametric copulas the same edges are present in the model with the same being modeled as a non-parametric copula. The contour plots of the model are displayed in Figure 13.5 in Section 13.3.3.

| Node | Pair copula | Family | Effective # parameters | Copula log-lik. | $AIC_C$ | $BIC_C$ | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pka | tll | 17.77 | 25.95 | -16.38 | 80.61 | 0.02 |
| | raf pkc; pka | frank | 1.00 | 2.75 | -3.49 | 1.97 | -0.04 |
| mek | mek raf | tll | 35.82 | 1208.66 | -2345.68 | -2150.11 | 0.62 |
| plc | plc pip3 | tll | 24.92 | 536.89 | -1023.96 | -887.94 | 0.27 |
| pip2 | pip2 pip3 | bb8 | 2.00 | 758.55 | -1513.09 | -1502.17 | 0.26 |
| | pip2 plc; pip3 | tll | 27.49 | 183.27 | -311.57 | -161.50 | 0.52 |
| erk | erk pka | tll | 22.55 | 409.98 | -774.86 | -651.75 | 0.25 |
| | erk mek; pka | tll | 30.29 | 39.21 | -17.85 | 147.49 | -0.05 |
| akt | akt erk | tll | 44.66 | 1847.79 | -3606.26 | -3362.45 | 0.75 |
| | akt pka; erk | tll | 21.40 | 176.73 | -310.66 | -193.86 | 0.30 |
| pkc | pkc pip2 | tll | 25.74 | 28.58 | -5.68 | 134.83 | -0.01 |
| p38 | p38 pkc | tll | 28.30 | 1055.53 | -2054.46 | -1899.96 | 0.53 |
| jnk | jnk pkc | tll | 28.84 | 622.62 | -1187.55 | -1030.10 | 0.43 |

Table 13.8: Summary of all copulas fitted in the $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model

### 13.3.3 Gaussian Margins with Gaussian Copulas ($\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$)

In a third approach we fitted a Gaussian D-vine, i.e., used Gaussian margins for the probability integral transform and only allowed for Gaussian copulas in the D-vine regression. This resulted in a log-likelihood of **4591.00**, an $AIC_C$ of **-9160.00** and a $BIC_C$ of **-9099.95** of the copula terms. Table 13.9 shows the results for each node and Table 13.10 the results for each copula.

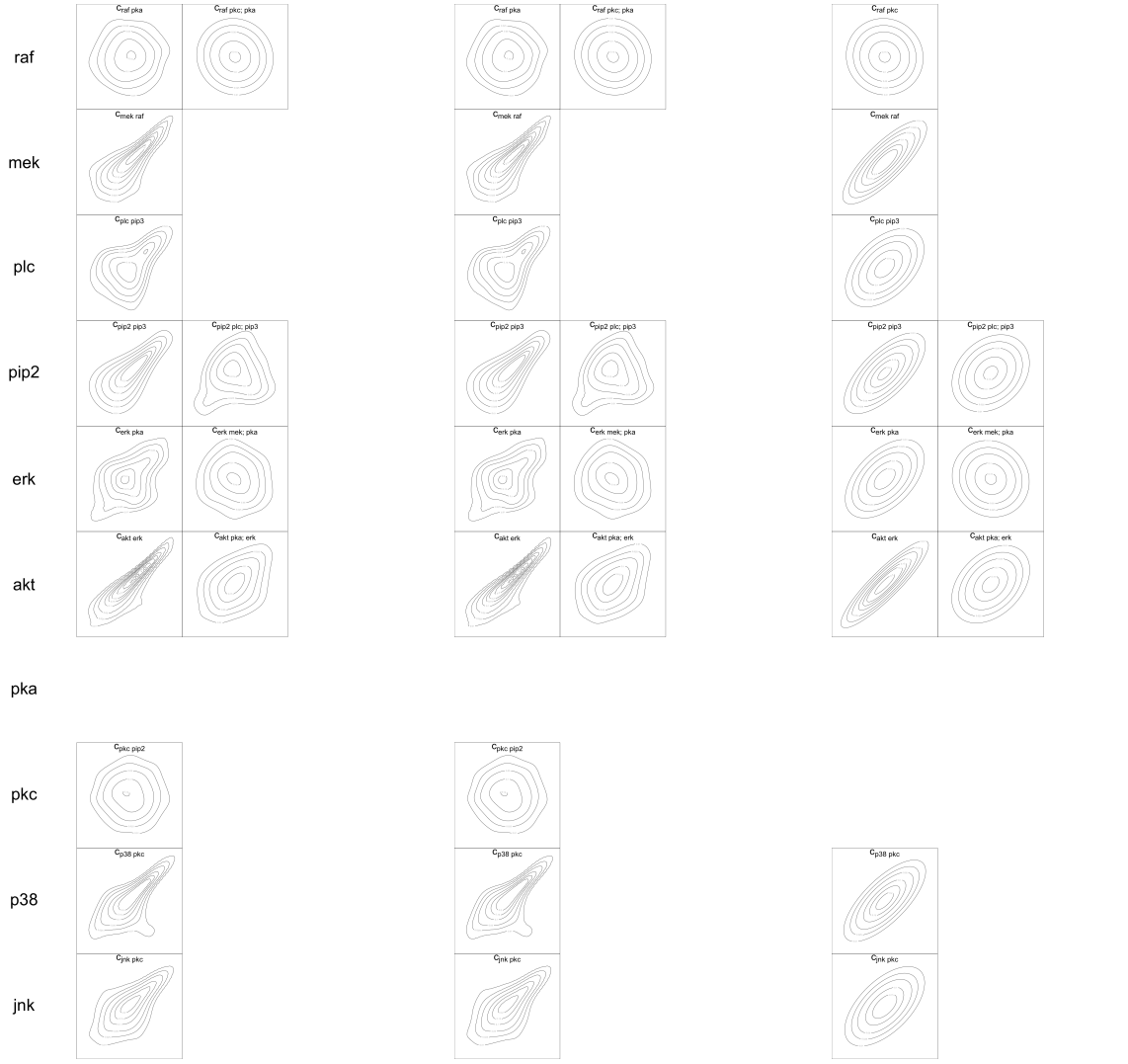
| Node | D-vine order | Copula log-lik. | $AIC_C$ | $BIC_C$ |
|---|---|---|---|---|
| raf | pkc | 1.02 | -0.03 | 5.43 |
| mek | raf | 920.61 | -1839.22 | -1833.77 |
| plc | pip3 | 222.61 | -443.22 | -437.76 |
| pip2 | pip3, plc | 613.28 | -1222.57 | -1211.65 |
| erk | pka, mek | 263.49 | -522.98 | -512.06 |
| akt | erk, pka | 1549.02 | -3094.04 | -3083.12 |
| pka | | | | |
| pkc | | | | |
| p38 | pkc | 647.86 | -1293.71 | -1288.25 |
| jnk | pkc | 373.11 | -744.23 | -738.77 |
| $\sum$: | | 4591.00 | -9160.00 | -9099.95 |

Table 13.9: Log-likelihood of the copula terms, $AIC_C$, $BIC_C$ and order of the D-vine for each node in the $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ model

We observe that only eleven of the twenty possible copulas are modeled. Additional to the ones missing in the $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ model and the $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model also the dependencies $pka \to raf$ and $pip2 \to pkc$ are missing. The contour plots are displayed in Figure 13.5.

| Node | Pair copula | Family | Copula log-lik. | $AIC_C$ | $BIC_C$ | Par. | Est. Ken. $\tau$ |
|---|---|---|---|---|---|---|---|
| raf | raf pkc | gaussian | 1.02 | -0.03 | 5.43 | -0.03 | 0.02 |
| mek | mek raf | gaussian | 920.61 | -1839.22 | -1833.77 | 0.81 | 0.62 |
| plc | plc pip3 | gaussian | 222.61 | -443.22 | -437.76 | 0.48 | 0.27 |
| pip2 | pip2 pip3 | gaussian | 550.89 | -1099.79 | -1094.33 | 0.69 | 0.26 |
| | pip2 plc; pip3 | gaussian | 62.39 | -122.78 | -117.32 | 0.27 | 0.52 |
| erk | erk pka | gaussian | 258.03 | -514.06 | -508.60 | 0.51 | 0.25 |
| | erk mek; pka | gaussian | 5.46 | -8.92 | -3.46 | -0.08 | -0.05 |
| akt | akt erk | gaussian | 1419.65 | -2837.31 | -2831.85 | 0.90 | 0.75 |
| | akt pka; erk | gaussian | 129.37 | -256.73 | -251.27 | 0.37 | 0.30 |
| p38 | p38 pkc | gaussian | 647.86 | -1293.71 | -1288.25 | 0.73 | 0.53 |
| jnk | jnk pkc | gaussian | 373.11 | -744.23 | -738.77 | 0.59 | 0.43 |

Table 13.10: Summary of all copulas fitted in the $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ model




(a) $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ (b) $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ (c) $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$

Figure 13.5: Contour plots of the respective copulas in the models allowing for parametric and non-parametric pair copulas where kernel density margins or Gaussian mixture margins were used for the PIT and the model where Gaussian margins and only Gaussian copulas were used on the $cd3cd28icam2$ and $pma$ data set

### 13.3.4 Comparison of D-vine Regression Models

What stands out first looking at the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model is that compared to all the earlier models with kernel density margins or Gaussian mixture margins used for the PIT and parametric and non-parametric copulas fitted on the different data sets the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model are the only ones where one node is modeled independent of all others. We observe that the only parent dependency of the node *pka*, $pkc \to pka$, is not modeled in both the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model. Additionally, there are many further similarities between these two models. In both of them only thirteen of the twenty possible dependencies are modeled, all nine D-vines have the same order and all copulas have the same family and a similar number of effective parameters, log-likelihood, $\mathrm{AIC}_C$ and $\mathrm{BIC}_C$. Therefore, unsurprisingly, as a consequence also their contour plots in Figure 13.5 are almost indistinguishable.

When instead restricting to Gaussian copulas and using Gaussian margins for the probability integral transform the goodness of fit is notably worsened. One reason is that in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model only eleven dependencies are modeled. Additional to the node *pka* being modeled independently as in the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model also the node *pkc* is modeled independently of all others. Note that four of six dependencies with *pka* as a parent node are not modeled. This is something we have observed similar for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3\&cd3+ly}}$ model.

## 13.4 Goodness of Fit Measures Model Comparison

One last time, we add up the goodness of fit measures of the margins and the copulas of the D-vine models fitted on the $cd3cd28icam2$ and $pma$ data set such that we are able to compare the results with the fitted Linear Gaussian Bayesian Network. The results for the log-likelihood are displayed in Table 13.11, for the $\text{AIC}_F$ in Table 13.12, for the $\text{BIC}_F$ in Table 13.13 and for the number of (effective) parameters used in the models in Table 13.14.

Comparing the two models in which we allow for parametric and non-parametric copulas we can see that they are extremely similar for all three goodness of fit measures. For the log-likelihood and the $\text{AIC}_F$ it seems that the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model produces a slightly better output than the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model. This holds for all nodes when considering the log-likelihood and all but one node, $jnk$, when considering the $\text{AIC}_F$. This leads to the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model being the best model of the four when comparing the log-likelihood or the $\text{AIC}_F$. However, when taking the absolute value into account the differences between the two models can almost be neglected.

On the other hand, comparing the $\text{BIC}_F$ the results for most nodes in the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model are slightly better than in the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model. This is due of the increased amount of effective parameters used in the kernel density margins. Therefore, making the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model the best of the four models when looking at the $\text{BIC}_F$. However, as before when considering the log-likelihood or the $\text{AIC}_F$, taking the absolute value into account no significant differences are present for the $\text{BIC}_F$.

While the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model and the $\text{LGBN}_{\mathbf{icam2\&pma}}$ have a worse fit than the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model they both range on the same level. Comparing the log-likelihood and the $\text{AIC}_F$ they are even almost identical. Only when considering the $\text{BIC}_F$ we derive that the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model fits a little bit better than the $\text{LGBN}_{\mathbf{icam2\&pma}}$.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{LGBN_{icam2\&pma}}$ | -2005 | -1070 | -1482 | -2147 | -2410 | -2038 | -420 | -1773 | -1968 | -1106 | -1781 | -18200 |
| $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ | -1929 | -716 | -1001 | -1549 | -2355 | -1834 | 130 | -1602 | -1825 | -573 | -1421 | -14675 |
| $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ | -1938 | -729 | -1020 | -1558 | -2368 | -1853 | 102 | -1612 | -1835 | -593 | -1428 | -14831 |
| $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ | -2006 | -1070 | -1482 | -2149 | -2410 | -2038 | -420 | -1773 | -1968 | -1106 | -1781 | -18204 |

Table 13.11: Global log-likelihood of the model fitted on the *cd*3*cd*28*icam*2 and *pma* data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{LGBN_{icam2\&pma}}$ | 4019 | 2150 | 2969 | 4303 | 4824 | 4085 | 850 | 3551 | 3944 | 2220 | 3571 | 36485 |
| $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ | 3926 | 1543 | 2084 | 3191 | 4731 | 3805 | -111 | 3232 | 3734 | 1231 | 2931 | 30299 |
| $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ | 3929 | 1546 | 2106 | 3198 | 4757 | 3815 | -62 | 3240 | 3737 | 1259 | 2929 | 30454 |
| $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ | 4017 | 2146 | 2969 | 4307 | 4824 | 4085 | 848 | 3549 | 3941 | 2218 | 3569 | 36473 |

Table 13.12: $AIC_F$ of the models fitted on the *cd*3*cd*28*icam*2 and *pma* data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LGBN}_{\mathbf{icam2\&pma}}$ | 4040 | 2178 | 2986 | 4324 | 4835 | 4106 | 877 | 3573 | 3966 | 2241 | 3593 | 36720 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ | 4115 | 1844 | 2307 | 3444 | 4789 | 4180 | 298 | 3308 | 3965 | 1463 | 3174 | 32889 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ | 4075 | 1786 | 2286 | 3419 | 4817 | 4114 | 326 | 3283 | 3922 | 1457 | 3130 | 32615 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ | 4033 | 2163 | 2986 | 4329 | 4835 | 4106 | 870 | 3560 | 3952 | 2235 | 3585 | 36654 |

Table 13.13: $\text{BIC}_F$ of the models fitted on the *cd*3*cd*28*icam*2 and *pma* data set for each node. Values in green show that it is at least 10% better than the respective value in the Linear Gaussian Bayesian Network.

| Model | raf | mek | plc | pip2 | pip3 | erk | akt | pka | pkc | p38 | jnk | $\sum$: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\text{LGBN}_{\mathbf{icam2\&pma}}$ | 4 | 5 | 3 | 4 | 2 | 4 | 5 | 3 | 4 | 4 | 4 | 42 |
| $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ | 35 | 55 | 41 | 46 | 11 | 69 | 75 | 14 | 42 | 43 | 44 | 474 |
| $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ | 27 | 44 | 33 | 40 | 11 | 55 | 71 | 8 | 34 | 36 | 37 | 396 |
| $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ | 3 | 3 | 3 | 4 | 2 | 4 | 4 | 2 | 2 | 3 | 3 | 33 |

Table 13.14: Number of (effective) parameters of the models fitted on the *cd*3*cd*28*icam*2 and *pma* data set for each node

## 13.5 Simulation Based Comparison

Similar to when we fitted models to the Sachs data set or the data from other chosen experiments, we again simulated from the fitted models as described in Chapter 8. The $cd3cd28icam2$ and $pma$ data set consists of 1736 samples and so we also sampled 1736 times from the models to be able to compare the output with the $cd3cd28icam2$ and $pma$ data set. First, we look at how the kernel density estimates fitted to the data of each node in the four models behave compared to kernel density estimates fitted to each node in the original data set.

This is displayed in Figure 13.6. We already discussed earlier that this mostly gives information about the fit of the marginals of the models. Hence, it does not surprise that the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model, where the margins fit quite well, outperform the other two models. This holds especially for the nodes which do not seem to be Gaussian distributed, i.e., all but $erk$, as the marginals in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model and the $\text{LGBN}_{\mathbf{icam2\&pma}}$ are Gaussian distributed.

Next we want to analyze how well the models are able to resemble the dependence structures in the $cd3cd28icam2$ and $pma$ data set. For this we compare the pairs plots of the simulated data sets with the original data set.

Figure 13.7 shows the pairs plots of the simulated data from the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model compared to the pairs plots of the $cd3cd28icam2$ and $pma$ data set. First, to say is that we hardly observe any differences between the pairs plots of the two models. Comparing them to pairs plots of the original data set many similarities are present. This holds for dependencies directly modeled as well as for dependencies not directly modeled. We already derived a similar result when analyzing the models with kernel density margins or Gaussian mixture margins and parametric and non-parametric copulas on the $cd3cd28 + aktinhib$ or $cd3cd28$ and $cd3cd28 + ly$ data set.

The results for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model and the $\text{LGBN}_{\mathbf{icam2\&pma}}$ are displayed in Figure 13.9. As both models are only able to model Gaussian dependence structures it does not surprise that both models have problems recreating more complicated non-normal dependence structures like for example between the nodes $pkc$ and $p38$.

Therefore, compared to the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model the ability of the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model and the $\text{LGBN}_{\mathbf{icam2\&pma}}$ to recreate the pairs plots of the original data set is as expected way worse. Looking for example at the $raf \leftrightarrow mek$ pairs plot it seems that data simulated from the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model is slightly closer to the original data than the simulated data from the $\text{LGBN}_{\mathbf{icam2\&pma}}$.

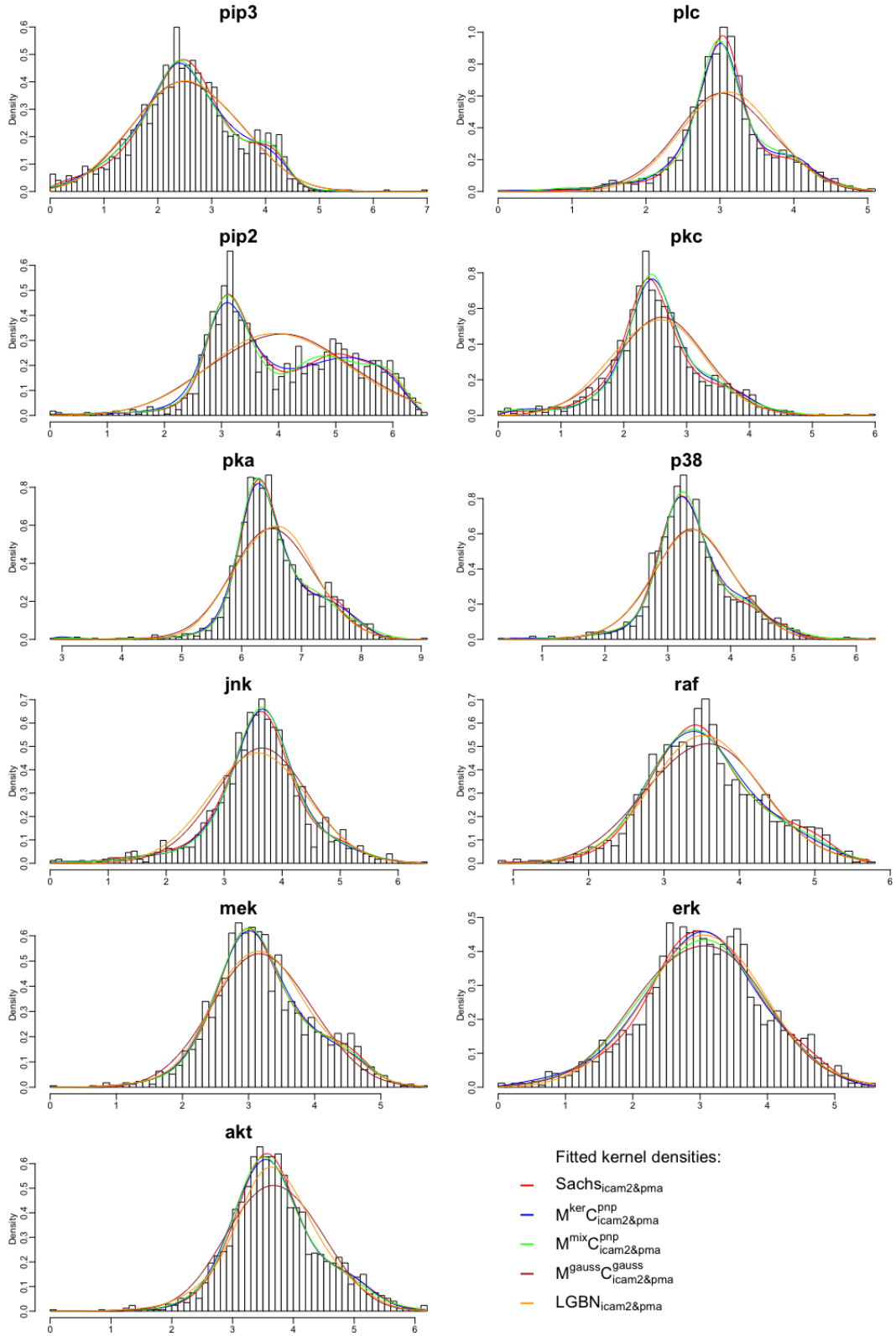
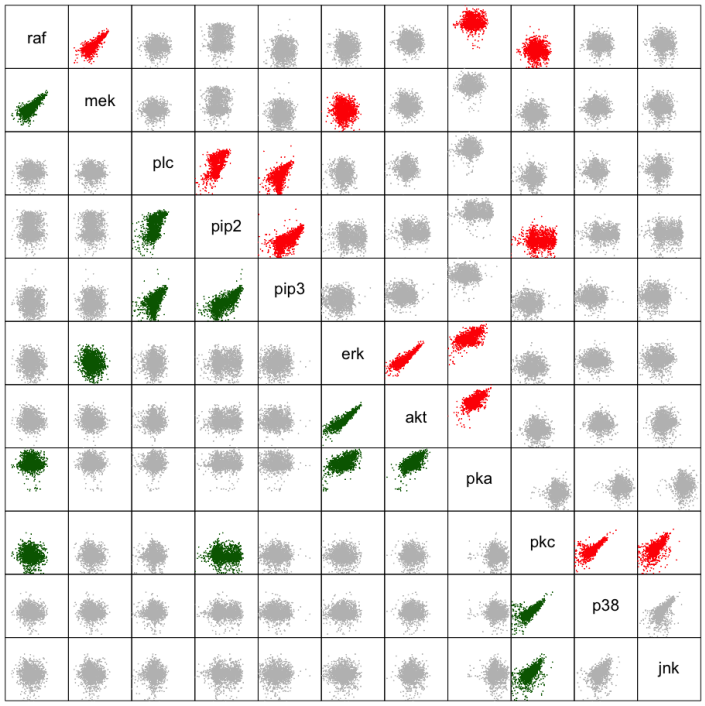
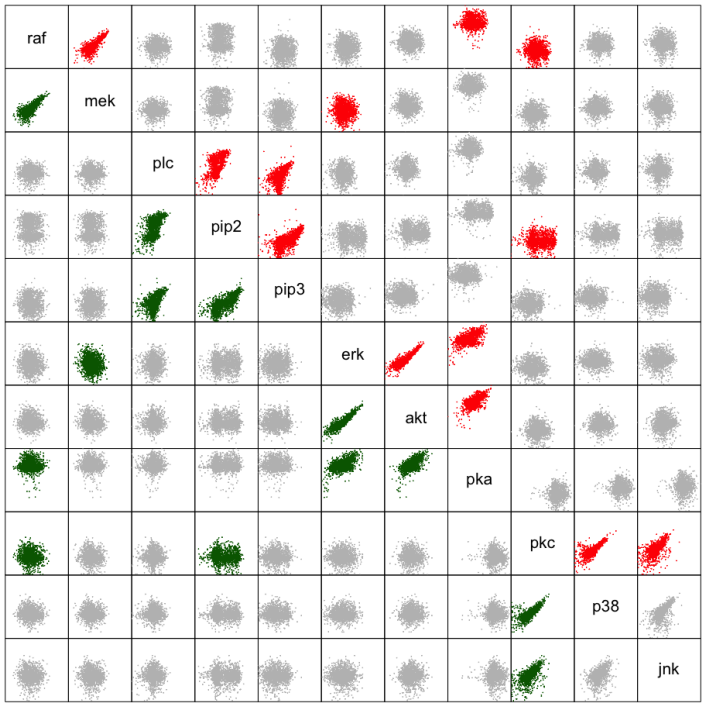



Figure 13.6: Marginal histograms of the *cd*3*cd*28*icam*2 and *pma* data set of each of the eleven nodes in topological order after samples containing a zero value have been deleted and kernel density estimates have been fitted to the simulated data from the different models

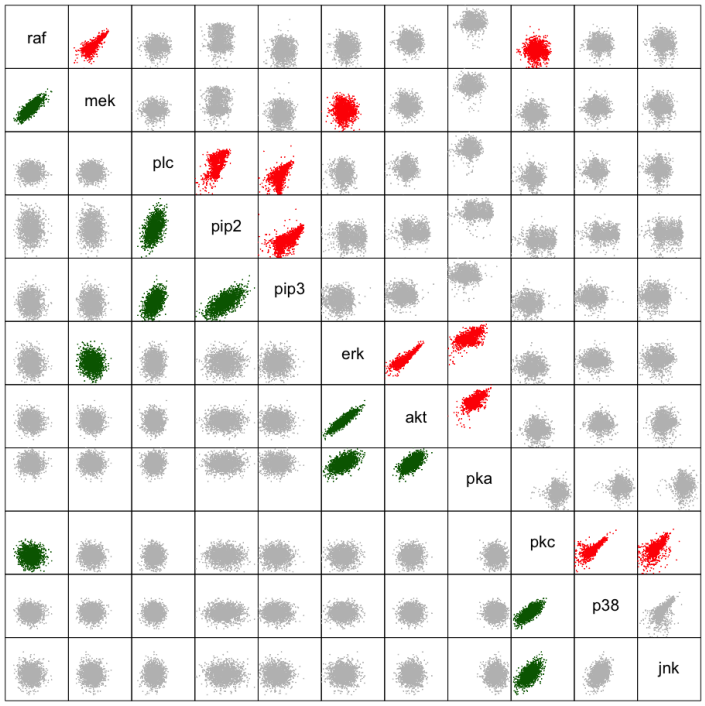



(a) $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$

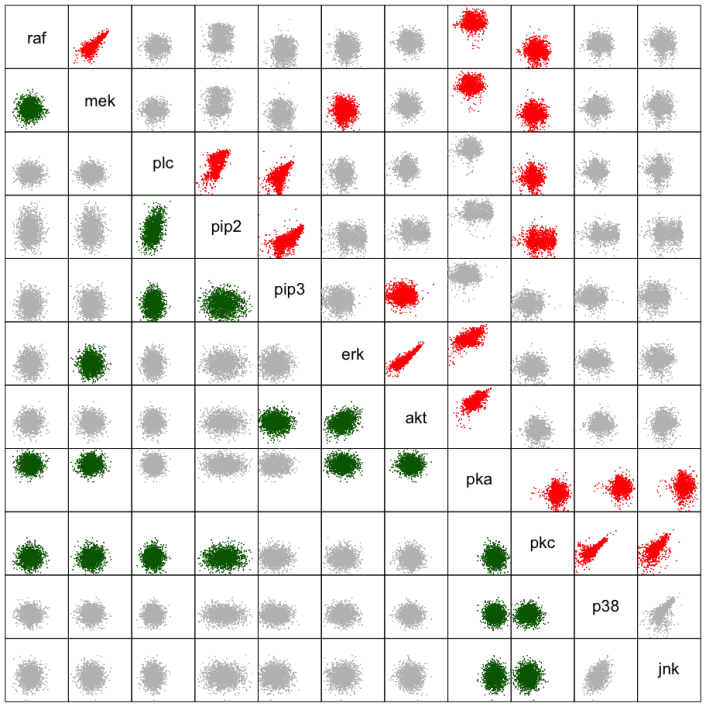



(b) $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$

Figure 13.7: Pairs plots for each pair of nodes of the respective model in the lower panel and of the *cd*3*cd*28*icam*2 and *pma* data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel




(a) $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$




(b) $\text{LGBN}_{\mathbf{icam2\&pma}}$

Figure 13.8: Pairs plots for each pair of nodes of the respective model in the lower panel and of the *cd*3*cd*28*icam*2 and *pma* data set in the upper panel. Green plots in the lower panel indicate that an edge exists between these two nodes in the model. For a better visibility the same pairs are colored in red in the upper panel

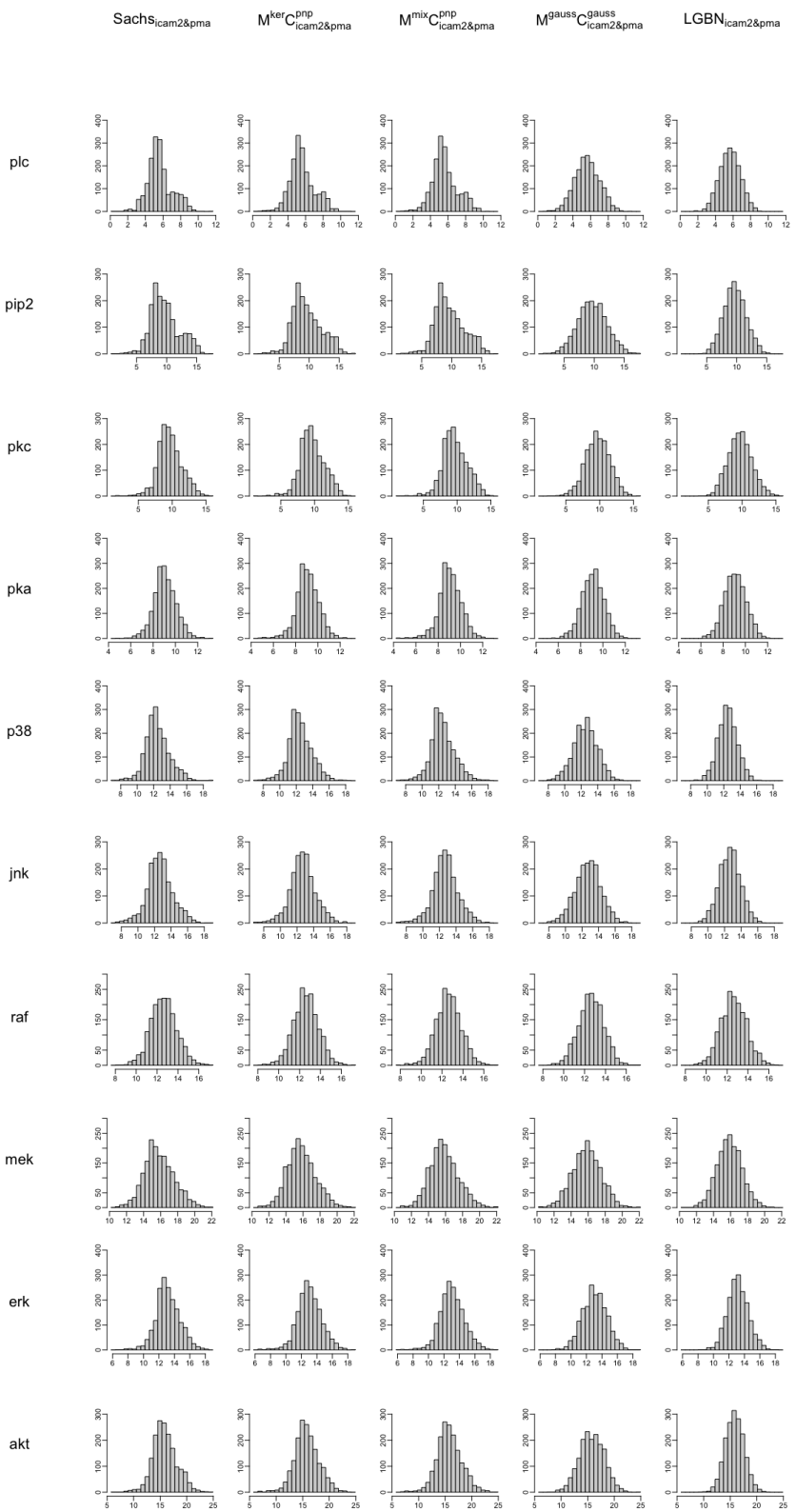



Figure 13.9: Histograms of the sum of each node on the set of its parents according to the consent DAG in the *cd*3*cd*28*icam*2 and *pma* data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ model, $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model, $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ model and the $\mathrm{LGBN}_{\mathbf{icam2\&pma}}$

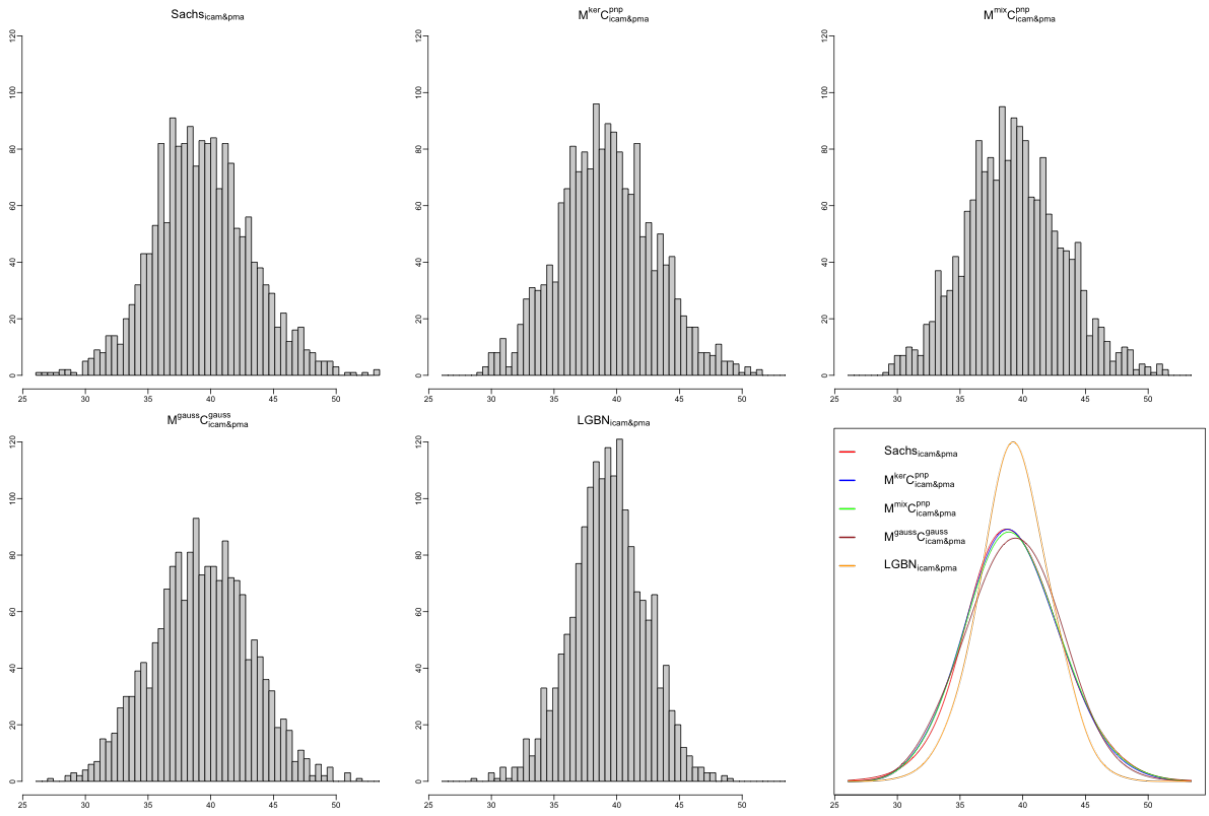



Figure 13.10: Histograms of the sum over all nodes in the *cd*3*cd*28*icam*2 and *pma* data set and in the simulated data sets of the $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ model, $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model, $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ model and the $\mathrm{LGBN}_{\mathbf{icam2\&pma}}$ in the first five plots and fitted kernel density estimates to each of the data sets in the sixth plot

Next, we compare how accumulations of high/low values appear within the samples in the simulated data sets and the *cd*3*cd*28*icam*2 and *pma* data set. As in the fitted models each node is modeled on the set of its parents according to the DAG, we first want to observe how well each of these conditional dependencies is modeled. For this we sum up according to the parents of each node in the data sets and then compare the resulting histograms. In the results in Figure 13.9 we can see that it seems that the histograms of the $\mathbf{M^{ker}C^{pnp}_{icam2\&pma}}$ model and the $\mathbf{M^{mix}C^{pnp}_{icam2\&pma}}$ model are quite close to the histograms of the *cd*3*cd*28*icam*2 and *pma* data set.

If the sum of a node and its parents is not normally distributed or skewed, it seems that both the $\mathbf{M^{gauss}C^{gauss}_{icam2\&pma}}$ model and the $\mathrm{LGBN}_{\mathbf{icam2\&pma}}$ are not able to recreate the histogram. As this is the case for many nodes, like *plc*, *pip*2 or *erk*, not too many similarities can be observed between these two models and the *cd*3*cd*28*icam*2 and *pma* data set.

As we are also interested how the accumulation of high/low values within each sample over all nodes is, we sum up over all nodes and compare the results for the four models to the $cd3cd28icam2$ and $pma$ data set. The results are given in Figure 13.10. It seems that while the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model are able recreate the shape of the histogram of the original data set really good, the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model has problems recreating the height of the peak and the $\mathrm{LGBN}_{\mathbf{icam2\&pma}}$ creates a too high peak. This observation gets confirmed looking at the fitted kernel density estimates fitted to the sums of all nodes in the models. Especially the one from the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model are almost identical with the ones from the $cd3cd28icam2$ and $pma$ data set.

## 13.6 Conditional Simulation Based Comparison

The last analysis to compare the models fitted on the $cd3cd28icam2$ and $pma$ data set is to compare how conditional densities behave if they are conditioned on certain values. This is no indication on which model fits best on the $cd3cd28icam2$ and $pma$ data set but on how the models change if they are fitted on different data sets. Therefore, we condition on the same set of points as in in Chapter 9 where the values are displayed in Table 9.1.

Considering the results in Figure 13.11 we observe that, knowing how similar the results for the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model have been so far, as expected the kernel density estimates fitted to the data generated by the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$ model are very similar. Only when conditioning on points from the tails we can see slight differences for the nodes $pip2$, $p38$ and $jnk$.

More differences can be seen if we compare these two to the kernel density estimates fitted to the data set simulated from the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model. Doing this only similarities for the node $raf$ can be derived. For all other nodes at least one of the plots strongly differs.

Finally looking at the data simulated from the $\text{LGBN}_{\mathbf{icam2\&pma}}$ we observe that there are even less similarities compared to the other models. For no single node all the three, respectively five, fitted kernel density estimates are similar to any other model. Again, the missing similarities to the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model are due to the many dependencies not modeled in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$ model.

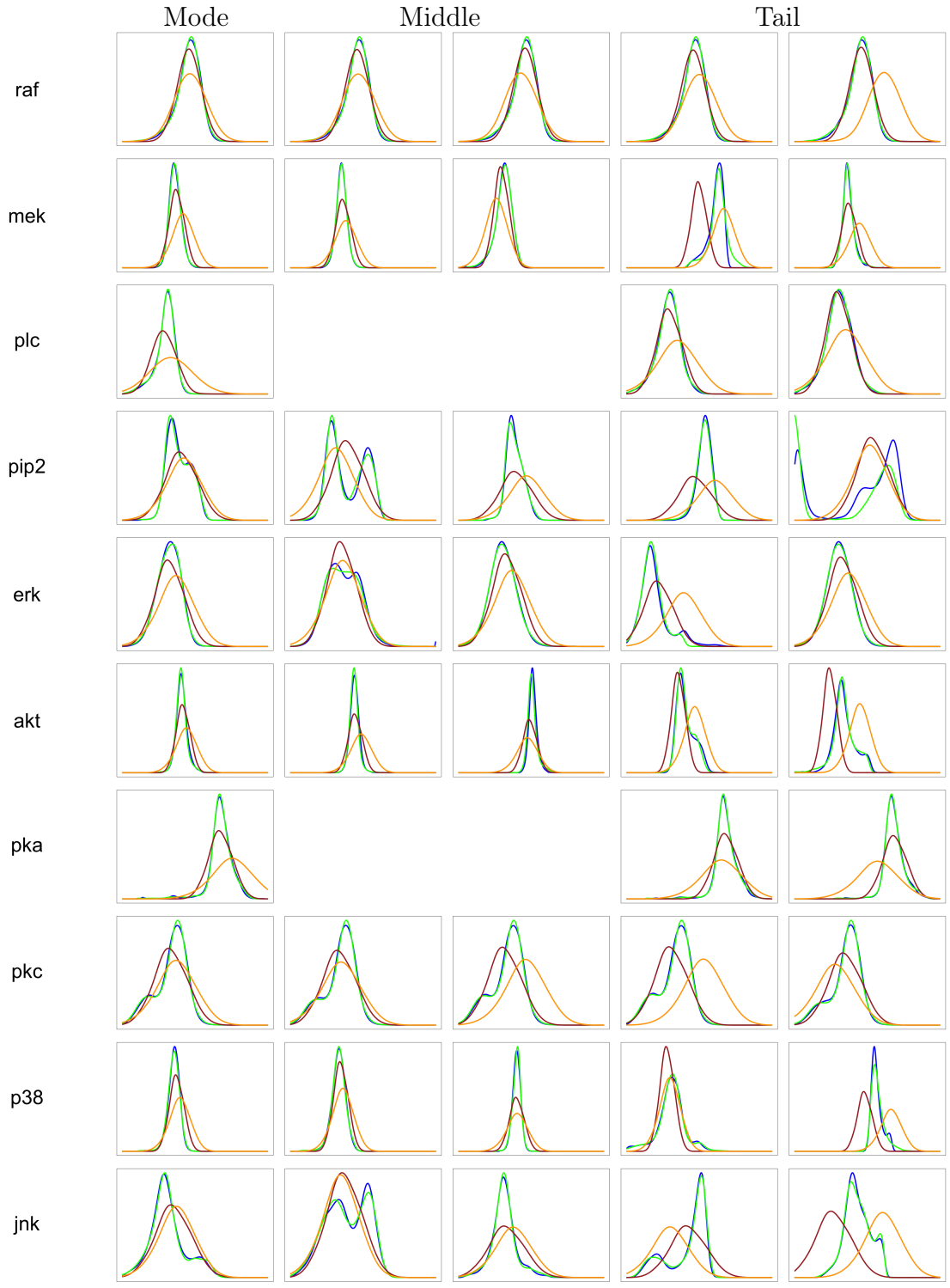



Figure 13.11: Density plots of the kernel density estimates fitted to the simulated data from the different models conditioned on the nodes in Table 9.1 in the same order. Blue: $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$. Green: $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}_{\mathbf{icam2\&pma}}$. Brown: $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{icam2\&pma}}$. Orange: $\mathrm{LGBN}_{\mathbf{icam2\&pma}}$

# Chapter 14

# Summary of the Model Fitting on the Partitions of the Sachs Dataset

In Chapter 10 we have identified three partitions of the Sachs data set to which we fitted a selection of models that we have previously already fitted to the full Sachs data set. We have already observed that especially the models with parametric and non-parametric copulas and either kernel density estimates or Gaussian mixture margins fit quite well on all three partitions. The question now is if we can derive similarities throughout the fitted models on the different partitions and on the complete Sachs data set.

Note that on all three partitions the fitted Gaussian mixture margins and kernel density margins resulted in a very similar fit which outperformed the fitted Gaussian margins. This is something we have already observed during the model fitting on the complete Sachs data set. Further, using the fitted Gaussian margins for the PIT did not result in uniformly distributed data in any case. Therefore, Gaussian margins do not fulfill the requirements to fit copulas on the data transformed with them them. Recall that the main features that we used to partition the Sachs data set were the fitted kernel densities to each node. Therefore, the margins of the models fitted to different data sets will surely be different and it does not seem reasonable to compare them in more detail.

What we can instead do is to compare how the contour plots of the data look like after applying the the PIT using the fitted margins. We observed that the general shape of the contour plots are the same for all types of margins on a certain data set. Therefore, it is enough to compare the contour plots for the different data sets. We can see that the contours of the same ten pairs of nodes have a non-elliptical form on every partition of the Sachs data. However, for some pairs of nodes, for example for the pair $p38 \leftrightarrow jnk$, the shape changes depending on the data set. Comparing them with the contour plots on the complete Sachs data set we derive that if a pair of nodes has a non-elliptical shape

| | | | Group 1 | | Group 2 | | Group 3 | |
|---|---|---|---|---|---|---|---|---|
| | Sachs | | $cd3cd28 + aktinhib$ | | $cd3cd28$ and $cd3cd28 + ly$ | | $cd3cd28icam2$ and $pma$ | |
| | $\mathbf{MC}^{\mathbf{pnp}}$ | $\mathbf{MC}^{\mathbf{gauss}}$ | $\mathbf{MC}^{\mathbf{pnp}}$ | $\mathbf{MC}^{\mathbf{gauss}}$ | $\mathbf{MC}^{\mathbf{pnp}}$ | $\mathbf{MC}^{\mathbf{gauss}}$ | $\mathbf{MC}^{\mathbf{pnp}}$ | $\mathbf{MC}^{\mathbf{gauss}}$ |
| raf-pka | ✓ | ✓ | ✓ | ✓ | X | X | ✓ | X |
| raf-pkc | ✓ | ✓ | ✓ | ✓ | ✓ | X | ✓ | ✓ |
| mek-raf | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| mek-pkc | ✓ | ✓ | ✓ | ✓ | X | X | X | X |
| mek-pka | ✓ | ✓ | ✓ | X | ✓ | X | X | X |
| plc-pip3 | ✓ | X | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| pip2-plc | ✓ | ✓ | ✓ | X | ✓ | ✓ | ✓ | ✓ |
| pip2-pip3 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| erk-pka | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| erk-mek | ✓ | X | X | X | ✓ | X | ✓ | ✓ |
| akt-erk | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| akt-pka | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| akt-pip3 | ✓ | ✓ | X | X | X | X | X | X |
| pka-pkc | ✓ | ✓ | ✓ | X | ✓ | ✓ | X | X |
| pkc-plc | ✓ | ✓ | X | X | ✓ | ✓ | X | X |
| pkc-pip2 | ✓ | X | ✓ | ✓ | ✓ | X | ✓ | X |
| p38-pka | ✓ | ✓ | ✓ | X | X | X | X | X |
| p38-pkc | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| jnk-pka | ✓ | ✓ | ✓ | X | X | X | X | X |
| jnk-pkc | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Table 14.1: Dependencies modeled in the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model, $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model and $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model on the different data sets. We indicate that a dependency is modeled in a specific model by ✓ and that it is not modeled by X. As the modeled dependencies agree for the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ model and the $\mathbf{M}^{\mathbf{mix}}\mathbf{C}^{\mathbf{pnp}}$ model, if they are modeled on the same data set, we summarize them under $\mathbf{MC}^{\mathbf{pnp}}$. Furthermore we abbreviate $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ by $\mathbf{MC}^{\mathbf{gauss}}$. Pairs of nodes whose normalized contour plots have a highly non-elliptical shape on all three partitions are colored in blue

on the partitions it also has a non-elliptical shape on the whole data set. Nevertheless, the reverse does not hold as there exist many more non-elliptical shaped contour plots on the Sachs data set.

Note that the dependency $p38 \leftrightarrow jnk$ cannot be directly model as there does not exist an edge in the DAG. Therefore, we only see nine edges colored in blue in Table 14.1. Apart from that, dependencies which have a non-elliptical shape are modeled in almost all fitted models. Looking at Table 14.1 we observe that only the $pip2 \leftrightarrow plc$ dependency is not modeled in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}_{\mathbf{cd3+akt}}$ model and the $plc \leftrightarrow pip3$ dependency in the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ model fitted on the Sachs data set as a whole. Recalling Table 10.1 we observe that the pairs where the contour plots have a non-elliptical shape are also the dependencies between which a high dependency, speaking in terms of Kendall's $\tau$, exists in the data of almost all experiments. Apart from that it seems that especially in the models fitted to the data sets of the $cd3cd28$ and $cd3cd28 + ly$ experiments or the $cd3cd28icam2$ and $pma$ experiments many dependencies of the node $pka$ are not modeled.

In general significantly less dependencies are modeled if we only consider parts of the Sachs data instead of the whole set. This indicates that some dependencies are only present in

the data of certain experiments. Note that again independent of which data set we use to fit a **MC** model allowing for parametric and non-parametric copulas and either using Gaussian mixture margins or kernel density margins for the PIT the results are almost identical. Essentially, the same dependencies are modeled in both.

Comparing the $\text{AIC}_F$ of the nodes in the **MC** models with parametric and non-parametric copulas we can see that there the same five nodes are more than 10% better than in the Linear Gaussian Bayesian Networks on all three data sets. Namely these are the nodes *mek*, *pip*2, *akt*, *p*38 and *jnk*. For each of the nodes exists at least one other node with which it has a highly non-elliptical pairs plot. Apart from these nodes the nodes *plc*, *erk* and *pkc* are also significantly better modeled in some of the $\mathbf{MC}^{\mathbf{pnp}}$ models. For all of these nodes exists another node with which they form a highly non-elliptical pairs plot as well.

Comparing the results of the simulation from the different models to the data sets the models are fitted on, we deduce the same results for all three partitions. The $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ models and the $\mathbf{M}^{\mathbf{ker}}\mathbf{C}^{\mathbf{pnp}}$ models create very similar data which is very close to the data the models are fitted on. This does not hold for the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ models and the Linear Gaussian Bayesian Networks where the $\mathbf{M}^{\mathbf{gauss}}\mathbf{C}^{\mathbf{gauss}}$ models seem to recreate the data they are fitted on slightly better.

Finally, we can also compare the behavior of the conditional simulations when we conditioned on the values from Table 9.1. We will here only compare the models with parametric and non-parametric copulas as the other two models are only able to model conditional Gaussian distributions which therefore will always have a similar shape. What is interesting is that we can already see notable differences in the conditional densities of some nodes for which we said that the data is similarly distributed for all five experiments when we partitioned the Sachs data set in Chapter 10. These are namely the nodes *erk*, *akt*, *pka* and *p*38. While for the node *akt* the conditional densities seem to be mostly similar this does not hold for *erk* and *p*38 especially when looking at the plots where we condition on points from the tails. Note that we cannot compare the conditional densities for the node *pka* as it is modeled independent of all other nodes in the models fitted on the data of the *cd*3*cd*28 and *cd*3*cd*28 + *ly* experiments. Hence, it does not surprise that in the nodes for which we assumed that the distribution is different when we partitioned the Sachs data set, e.g., *pip*2 and *jnk*, more differences can be observed. These are again especially visible considering the plots where we condition on points from the tails of the distributions. Overall it seems reasonable that many differences in the conditional simulations are present as we choose the data sets in a way that the distributions of the nodes are different on the different data sets.

# Chapter 15

# Conclusion

In this thesis we combined the D-vine regression from Kraus and Czado (2017) with a graphical approach, i.e., beforehand modeling the dependence structure as a Bayesian Network following Pearl (1988) and applied this to the Sachs data set. By first fitting a Bayesian Network to the data set we reduced the number of dependencies we need to consider modeling. This helps to reduce to computation time of the model fitting if in the Bayesian Network the parent sets of the nodes are of reasonable size, which is the case here. We then modeled each conditional density given by the Bayesian Network as a D-vine.

For our investigations we varied between Gaussian margins, Gaussian mixture margins or kernel density margins and allowed for three different sets of copulas, Gaussian copulas, parametric copulas or parametric and non-parametric copulas. The results were then compared to a Linear Gaussian Bayesian Network using the full log-likelihood, $\text{AIC}_F$ and $\text{BIC}_F$. There, we have observed that especially when we allowed for parametric and non-parametric copulas and used Gaussian mixtures margins or kernel density margins for the probability integral transform the results outperformed the LGBN. For a further analysis we chose these two models, the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model, together with the $\mathbf{M^{gauss}C^{gauss}}$ model and the Linear Gaussian Bayesian Network and analyzed how they are able to recreate the Sachs data set.

Here, we have again seen that the results of the $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model outperformed the other two. However, some of the results were not as good as expected. As a reason for that we identified the violation of the requirement of identically distributed data throughout the Sachs data set which seems reasonable as the Sachs data set is composed of data from nine different experiments.

Therefore, we considered the individual data sets and identified two pairs of experiments for which the assumption of identical distributed data within them seemed reasonable. Together with these two pairs we chose the data of a third experiment on which we have repeated our analysis. This time, we only fitted four models on each of the data sets, three D-vine regressions, two of them with parametric and non-parametric copulas and either Gaussian mixture margins or kernel density margins, as well as one with Gaussian margins and Gaussian copulas and again a Linear Gaussian Bayesian Network.

As expected on all three data sets the respective $\mathbf{M^{ker}C^{pnp}}$ model and the $\mathbf{M^{mix}C^{pnp}}$ model outperformed the other two. Their results leave almost no room for improvement in recreating the underlying data sets in the performed analysis. We have observed that this does not hold for the $\mathbf{M^{gauss}C^{gauss}}$ models and the Linear Gaussian Bayesian Networks where it seems that the results for the $\mathbf{M^{gauss}C^{gauss}}$ models were slightly better. This might indicated the correctness of the assumption that the data in the Sachs data set is not identically distributed.

Overall, we have seen that on all four data sets the fitted models of a D-vine regression based Bayesian Network approach outperformed the Linear Gaussian Bayesian Network especially when we used either Gaussian mixture margins or kernel density margins for the PIT and parametric and non-parametric copulas. Additionally it seems that the models are in general a good approach to describe the Sachs data set or only the data from certain experiments.

# List of Figures

# Bibliography


Aragam, B., Gu, J., and Zhou, Q. (2019). Learning large-scale bayesian networks with the sparsebn package. *Journal of Statistical Software*, 91(11).

Bedford, T. and Cooke, R. M. (2002). Vines–a new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068.

Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6).

Bevacqua, E., Maraun, D., Haff, I., Widmann, M., and Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in ravenna. *Hydrology and Earth System Sciences Discussions*, 21(6).

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29(2):610–611.

Cheng, Y., Du, J., and Ji, H. (2020). Multivariate joint probability function of earthquake ground motion prediction equations based on vine copula approach. *Mathematical Problems in Engineering*, pages 1–12.

Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas*. Lecture Notes in Statistics. Springer.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4).

Härdle, W. K. and Simar, L. (2015). *Applied Multivariate Statistical Analysis*. Springer Berlin Heidelberg.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference: Likelihood and Bayes*. Springer Berlin Heidelberg.

Joe, H. (1996). *Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters*, volume 28 of *Lecture Notes–Monograph Series*, pages 120–141. Institute of Mathematical Statistics.

Kauermann, G., Schellhase, C., and Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical b-splines. *Scandinavian Journal of Statistics*, 40(4):685–705.

Kendall, M. and Stuart, A. (1961). *The Advanced Theory of Statistics*. Number Bd. 2 in The Advanced Theory of Statistics. Hafner Publishing Company.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press.

Kraus, D. and Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics and Data Analysis*, 110:1–18.

Loader, C. (1999). *Local regression and likelihood*. Statistics and Computing. Springer New York.

Maechler, M. (2019). *nor1mix: Normal aka Gaussian (1-d) Mixture Models (S3 Classes and Methods)*. R package version 1.3-0.

Mathias Drton, Bernd Sturmfels, S. S. (2009). *Lectures on Algebraic Statistics*. Birkhäuser Basel.

Morales, O., Kurowicka, D., and Roelen, A. (2008). Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety*, 93(5):699 – 710.

Nagler, T. (2019). *vinereg: D-Vine Quantile Regression*. R package version 1.0.3.

Nagler, T., Schellhase, C., and Czado, C. (2017). Nonparametric estimation of simplified vine copula models: comparison of methods. *Dependence Modeling*, 5(1):99 – 120.

Nagler, T. and Vatter, T. (2019). *kde1d: Univariate Kernel Density Estimation*. R package version 1.0.2.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Series in Representation and Reasoning. Morgan Kaufmann Publishers Inc.

Pickands, J. (1981). Multivariate extreme value distributions. *Proceedings of the 43rd Session of the International Statistical Institute*, 2:859–878.

Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence. Pearson Education.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., and Nolan, G. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308:523–9.

Sancetta, A. and Satchell, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20:535–562.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231.

Sun, L.-H., Huang, X.-W., Alqawba, M. S., Kim, J.-M., and Emura, T. (2020). *Copula-Based Markov Models for Time Series: Parametric Inference and Process Control*. JSS Research Series in Statistics. Springer Singapore.

Team, R. C. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York.

Weglarczyk, S. (2018). Kernel density estimation and its application. *ITM Web of Conferences*, 23.

Yang, J., Wang, F., and Xie, Z. (2020). *Bernstein Copulas and Composite Bernstein Copulas*, pages 183–217. Springer Singapore.