A COPULA REPRESENTATION OF THE ASYMMETRIC LAPLACE DISTRIBUTION

Tong Wang

Supervisor: Doctor Clara Grazian

School of Mathematics and Statistics
UNSW Sydney

December 2020

Submitted in partial fulfilment of the requirements of the degree of
Master of Statistics

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: Tong Wang Date: 9/12/20

Acknowledgements

I would like to show my great respect and deep gratitude to my supervisor, Dr. Clara Grazian. In these past two months, I have been suffering from the most difficult time. She gave me a lot of help, not only academically, but also spiritually. During this difficult time and with the remotely project. Dr. Clara Grazain gave me reply with no lag and in a very detailed way every time. She explained everything very clear with her patience. And she kept encouraging me to get over my illness and move forward. Without her help, I cannot make it so far.

Besides, I would like to thank my parents and my wife. They gave me a lot of support during this period on everything.

Tong Wang, 27 November 2020.

Abstract

The scaled-generalised multivariate asymmetric Laplace distribution, which has the abbreviation of SMAL. The SMAL is a continuous probability distribution which is the general form of the multivariate Laplace distribution with the generalised unequal scale parameter and covariance matrix and has the definition on the quantile. In 2003, Kotz, Kozubowski, and Podgorski have found that the asymmetric Laplace distribution is a multivariate extension of the univariate, symmetric Laplace distribution. We will focus on the relationship between the SMAL distribution and quantile regression analysis based on the knowledge of copulas. Thus we can study if the SMAL distribution is a reasonable assumption for a real dataset.

Contents

Chapter 1 Introduction 1

- 1.1 Introduce to asymmetric Laplace distribution 1
- 1.2 Introduction to multivariate asymmetric Laplace distribution 2
- 1.3 Plan 3

Chapter 2 Bayesian Statistics And Quantile Regression 4

- 2.1 Bayes' theorem 4
- 2.2 Posterior distribution and Likelihood function 4
- 2.3 Quantile regression 5
- 2.4 Quantile regression and SMAL distribution 6

Chapter 3 Copula 7

- 3.1 Introduction to Copulas 7
- 3.2 Copulas in this thesis 7
- 3.3 Frequentist and Bayesian statistics with respect to copulas 9

Chapter 4 The dependence structure of the SMAL 10

- 4.1 Spearman's rho 10
- 4.2 Spearman's ρ of the SMAL distribution and specific Copulas 12
- 4.3 Tail dependence coefficients 14

Chapter 5 Simulation Study 18

- 5.1 Simulation study on Spearman's ρ 18
- 5.2 Simulation study on tail dependence coefficients 20

Chapter 6 Real data inspection 23

- 6.1 Introduction to the dataset "body" 23
- 6.2 Non-parametric Bayesian quantile regression 24
- 6.3 Simulate An SMAL distribution for testing 25

Chapter 7 Conclusion 27

References 28

Chapter 1

Introduction

In the past few years, quantile regression has become more and more popular and a widely used technique to describe the distribution of a response variable given a set of explanatory variables.(Geraci M and Bottai M.,2007[1]). According to the seminal work by Koenker and Bassett(1978), we have a new method besides the traditional least squares estimation while the least squares estimates the conditional mean, quantile regression estimates the conditional median or other quantiles. We will discuss the quantile regression in details in Chapter 2. To introduce the Scale-generalized multivariate asymmetric Laplace distribution, we first need to have a good knowledge of the univariate asymmetric Laplace distribution.

1.1 Introduce to asymmetric Laplace distribution

First we need to know the definition of the asymmetric Laplace distribution, as is mentioned in Kotz et.al 2001[2], we consider a random variable Y which has the univariate asymmetric Laplace distribution with location parameter μ with the mean $\mu+\eta$, and variance $\eta^2+\delta^2$, thus $Y \sim \mathcal{AL}(\mu, \eta, \delta)$, where $\mu \in \mathbb{R}, \eta \in \mathbb{R}$ and $\delta \geq 0$. The characteristic function of Y is in the form

$$\psi(t) = \frac{e^{i\mu t}}{1 + \frac{1}{2}\delta^2 t^2 - i\eta t} \tag{1.1.1}$$

The most important aspect of the asymmetric Laplace distribution is that we define it with respect to the quantile or the skewness. See(Kotz et.al 2001[2]), if we let τ be the skewness parameter related to the location μ and the scale δ, we can write

$$\tau = \frac{\sqrt{2}\delta}{\eta + \sqrt{2\delta^2 + \eta^2}} = \frac{\sqrt{2\delta^2 + \eta^2} - \eta}{\sqrt{2}\delta} \tag{1.1.2}$$

with $\tau > 0$. And $\eta = \frac{\delta}{\sqrt{2}}\left(\frac{1}{\tau} - \tau\right)$.

Now we assume a random variable Y is distributed as an asymmetric Laplace distribution with location parameter μ, scale parameter $\delta > 0$, and skewness parameter $\tau \in (0, 1)$, i.e., $Y \sim \mathcal{AL}(\mu, \delta, \tau)$, we have the following density function.(Lea Petrella and Valentina Raponi,2019[3])

$$f(y; \mu, \delta, \tau) = \tau(1-\tau)\delta^{-1} \exp[-\rho_\tau\{(y-\mu)/\delta\}], \tag{1.1.3}$$

where $\rho_\tau(v) = v(\tau - I(v \leqslant 0))$ is a loss function. We can inspect some plots of the densities of the univariate Laplace distribution from the following figure.

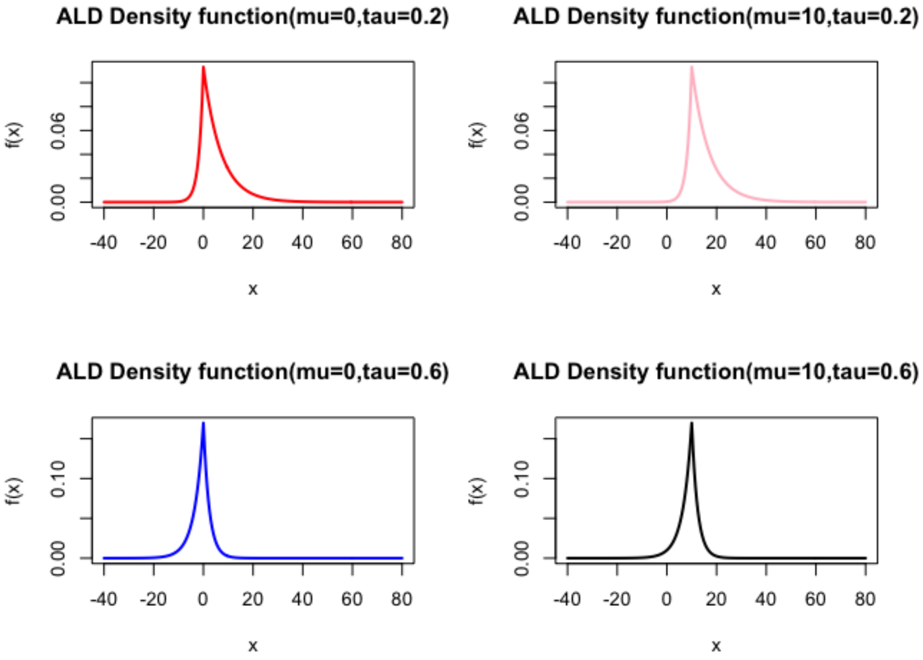


Figure 1.1: ALD density functions with different location parameters and skewness parameters

According to Kotz et.al 2001[2], we could represent the above as a Gaussian mixture form

$$Y = \mu + \xi_\tau U + \theta_\tau \sqrt{\delta U} Z$$

where U follows an exponential distribution with the rate parameter $1/\theta$ and Z is a random variable with a normal distribution with mean 0 and variance being 1. And we have $\xi_\tau = (1-2\tau)/\{\tau(1-\tau)\}, \quad \theta_\tau = 2/\{\tau(1-\tau)\}$.

1.2 Introduction to multivariate asymmetric Laplace distribution

Based on the knowledge of the univariate AL, we can extend the result of asymmetric Laplace distribution to the multivariate Laplace distribution. For the multivariate asymmetric Laplace distribution we let the response variables, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$, to be the p-dimensional response variables with $i \in \{1, ..., n\}$ entries for each response variable. From the details given in Kotz et.al 2001[2], for the $\boldsymbol{Y}_i$ which in the law of the multivariate asymmetric Laplace distribution with covariance matrix $\boldsymbol{\Sigma}$, and $K_v(\cdot)$ is the modified Bessel function of the third kind $v = (2-p)/2$. Thus we will have the following density function.

$$g(\mathbf{y}) = \frac{2}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \left(\frac{\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y}}{2}\right)^{v/2} K_v\left(\sqrt{2\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y}}\right) \tag{1.2.1}$$

Although we have the density function of the p-dimensional multivariate asymmetric Laplace distribution, we need to transform this density function into the alternative form, similar to the form we have discussed in the previous section. So now we can consider the following multivariate asymmetric Laplace distribution with density function from (Lea Petrella and Valentina Raponi,2019[3]),

$$f_Y\left(\boldsymbol{y}_i \mid \boldsymbol{\beta}_\tau \boldsymbol{X}_i, \boldsymbol{D}\xi, \boldsymbol{D}\Sigma\boldsymbol{D}\right) = \frac{2\exp\left\{\left(\boldsymbol{y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i\right)^\top \boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{-1}\xi\right\}}{(2\pi)^{p/2}|\boldsymbol{D}\Sigma\boldsymbol{D}|^{1/2}}\left(\frac{m}{2+d}\right)^{v/2} K_v\{\sqrt{(2+d)m}\} \tag{1.2.2}$$

where $\beta_\tau \boldsymbol{X}_i$ is the location parameter vector, $\boldsymbol{D}\xi \in \mathbb{R}^p$ is the scale parameter, with $\boldsymbol{D} = \begin{bmatrix} \delta_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \delta_n \end{bmatrix}$, $\delta_j > 0$,and $\xi = \left(\xi_1, \ldots, \xi_p\right)^\top$, $\xi_j = \left(1 - 2\tau_j\right) / \left\{\tau_j\left(1 - \tau_j\right)\right\}$. $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Psi\Lambda}$, with $\boldsymbol{\Psi}$ being a correlation matrix and $\boldsymbol{\Lambda} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \end{bmatrix}$, where $\sigma_j^2 = 2/\left\{\tau_j\left(1 - \tau_j\right)\right\}$ for all $j \in \{1, \ldots, p\}$. Also, $m = \left(\boldsymbol{y} - \boldsymbol{\beta}_\tau \boldsymbol{X}_i\right)^\top \left(\boldsymbol{D\Sigma D}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{\beta}_\tau \boldsymbol{X}_i\right)$, $d = \boldsymbol{\xi}^\top \boldsymbol{\Sigma\xi}$. Here we can also turn this into the Gaussian mixture form,

$$\boldsymbol{Y}_i = \boldsymbol{\beta}_\tau \boldsymbol{X}_i + \boldsymbol{D\xi}W + \sqrt{W}\boldsymbol{D\Sigma}^{1/2}\boldsymbol{Z}, \tag{1.2.3}$$

where W is an exponentially distributed random variable with mean 1, independent from $\boldsymbol{Z}$, and $\boldsymbol{Z}$ is a p-dimensional standard normal distribution. We also have the constrains as the univariate case, $\xi_j = \left(1 - 2\tau_j\right) / \left\{\tau_j\left(1 - \tau_j\right)\right\}$ and $\sigma_j^2 = 2/\left\{\tau_j\left(1 - \tau_j\right)\right\}$.

1.3 Plan

A brief review on the Bayesian statistics and the quantile regression will be given in Chapter 2. We will also discuss the relationship between the SMAL and the quantile regression in that Chapter. The dependence structure of the SMAL will be presented in Chapter 3 and the simulation study will be provided in Chapter 4. And Chapter 5 contains the real dataset application and the testing between it and the SMAL. Finally, the conclusion follows in Chapter 6.

Chapter 2

Bayesian Statistics And Quantile Regression

Bayesian statistics, named after Thomas Bayes, is a statistical theory based on Bayesian probability interpretation, in which probability represents the credibility of events. Bayesian statistics use the Bayes' theorem to find the probability or credibility after being given a new data, assuming we already have some knowledge of the given data according to our prior experience or the results from previous experiments.

2.1 Bayes' theorem

First, we need to talk about the Bayes' theorem in a detailed way. Bayes' theorem is about the conditional probability and the marginal probability of random events, A and B. And P(A | B) is the probability that A occurs when B is given. We know that if B is given, the probability of event A will occur should be $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. Similarly, if A is given, the probability that B will occur should be $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$. From these two we have $P(A \cap B) = P(A \mid B)^{*}P(B) = P(B \mid A)^{*}P(A)$. Therefore, we will reach the formula of the Bayes' theorem $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)^{*}P(A)}{P(B)}$.

2.2 Posterior distribution and Likelihood function

In the Bayesian statistics, the posterior probability of a random or uncertain event is the conditional distribution which is based on our prior knowledge or the experience from previous work. From (Christopher M. Bishop (2006)[4]), we have the definition of the posterior distribution. The posterior probability is the probability of the parameters θ given the evidence $X : p(\theta \mid X)$. The prior probability of θ is $p(\theta)$, and the observations $X = (x_1, ..., x_n)$ have the likelihood $p(x \mid \theta)$, then the posterior probability can be calculated as the following:

$$p(\theta \mid x) = \frac{p(x \mid \theta)}{p(x)} p(\theta) \tag{2.2.1}$$

Since we can consider $p(x) = \int_{\Theta} f(\mathbf{X} \mid \theta) p(\theta) d\theta$ as a constant. Therefore, we can write the posterior probability distribution as Posterior probability $\propto$ Likelihood $\times$ Prior probability.

Likelihood function is a joint density function of each observations of X, say, if

$\boldsymbol{X} = (x_1, ..., x_n)$ has a joint density function $f\left(x_1, x_2, \ldots, x_n \mid \theta\right)$, then the likelihood function with respect to θ can be written as

$$L(\theta) = L\left(\theta \mid x_1, x_2, \ldots, x_n\right) = f\left(x_1, x_2, \ldots, x_n \mid \theta\right), \tag{2.2.2}$$

if $X_1, X_2, \ldots, X_n$ are i.i.d., then $L(\theta) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right)$.

2.3 Quantile regression

Before talking about the quantile regression, we need to review the regression analysis. Including the linear regression, polynomial regression, most of regression analysis, we first assume the data from some fixed function and let the function fit the data to find the unknown parameters. We often use the MSE to let the function fit more data, where $MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}\left(x_i\right)\right)^2 = E(y - \hat{f}(x))^2$. Actually, we have got the expectation of Y. Hence, we can conclude from those analysis, in essence, the function obtained by all regression analysis is a conditional expectation function,i.e., in the condition of fixed x to find the expectation of Y.

Now we introduce the quantile regression which focuses on not only the expectation of Y, but also the whole distribution of Y. According to the seminal work by Koenker and Bassett(R. Koenker, G. Bassett Jr,1978[5]), quantile regression has been recognized in recent years as a robust statistical procedure that offers a powerful alternative to ordinary mean regression. While standard regression methods usually focus on modelling the conditional mean of the response variable, quantile regression models the conditional median or other quantiles of the response variable. From(Geraci M and Bottai M.,2007[1]), if we have a data in the form of $\left(\mathbf{x}_i^{\mathrm{T}}, y_i\right)$, for $i = 1, \ldots, N$ where y_i are independent scalar observations of a continuous random variable with common cumulative distribution function F_y, whose shape is not exactly known, and $\mathbf{x}_i^{\mathrm{T}}$ are row p -vectors of a known design matrix X. The linear conditional quantile functions are defined as $G_{y_i}\left(\tau \mid \mathbf{x}_i\right) = \mathbf{x}_i^{\mathrm{T}}\beta, \quad i = 1, \ldots, N$ with $\tau \in (0,1)$ being the quantile, $G_{y_i}(\cdot) \equiv F_{y_i}^{-1}(\cdot)$, and $\beta \in \mathbb{R}^p$ is a column vector of length p with unknown fixed parameters. We can minimize the following equation to find the quantile

$$\hat{Q}_Y(\tau) = \operatorname{argmin}_{\xi_\tau \in R}\left(\sum_{i:Y_i > \xi_\tau} \tau\left|Y_i - \xi_r\right| + \sum_{i:Y_i < \xi_\tau}(1-\tau)\left|Y_i - \xi_r\right|\right)$$

$$\min_{\beta \in \mathbb{R}^p} V_N(\beta), V_N(\beta) = \sum_{i \in \left(i: y_i \geqslant \mathbf{x}_i^{\mathrm{T}}\beta\right)}^{N} \tau\left|y_i - \mathbf{x}_i^{\mathrm{T}}\beta\right| + \sum_{i \in \left(i: y_i < \mathbf{x}_i^{\mathrm{T}}\beta\right)}^{N}(1-\tau)\left|y_i - \mathbf{x}_i^{\mathrm{T}}\beta\right|, \tag{2.3.1}$$

yields

$$\beta_\tau = \operatorname{argmin}_{\beta \in \mathbb{R}^p}\left(\sum_{i \in \left(i: y_i \geqslant \mathbf{x}_i^{\mathrm{T}}\beta\right)}^{N} \tau\left|y_i - \mathbf{x}_i^{\mathrm{T}}\beta\right| + \sum_{i \in \left(i: y_i < \mathbf{x}_i^{\mathrm{T}}\beta\right)}^{N}(1-\tau)\left|y_i - \mathbf{x}_i^{\mathrm{T}}\beta\right|\right), \tag{2.3.2}$$

2.4 Quantile regression and SMAL distribution

The basic notion of the SMAL distribution has been discussed in Chapter 1. Now we will consider what is the relationship between the SMAL and the quantile regression and why does the SMAL is so used in practical.

For the univariate case first. In [3], let $\mathcal{Q}_{y_i}(\tau \mid \boldsymbol{x}_i)$ denote the quantile regression function of y_i given $\boldsymbol{x}_i$ with the quantile parameter $\tau \in (0,1)$. Thus the quantile function can be modified as

$$\mathcal{Q}_{y_i}(\tau \mid \boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \tag{2.4.1}$$

where $\boldsymbol{x}_i$ is a $k \times 1$ vector, and $\boldsymbol{\beta}_\tau$ is a $k \times 1$ vector of regression coefficients. Thus

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau + \epsilon_i \tag{2.4.2}$$

Therefore,the error term gives us a zero τ-quantile,i.e., $\tau = 0$ thus from the equation (2.3.2) we obtain

$$\begin{aligned}
\beta_\tau &= \operatorname{argmin}_{\beta \in \mathbb{R}^k} \left(\sum_{i \in \left(i: y_i \geqslant \mathbf{x}_i^{\mathrm{T}} \beta\right)}^{N} \tau \left| y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \right| + \sum_{i \in \left(i: y_i < \mathbf{x}_i^{\mathrm{T}} \beta\right)}^{N} (1-\tau) \left| y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \right| \right) \\
&= \operatorname{argmin}_{\beta \in \mathbb{R}^k} \left(\sum_{i \in \left(i: y_i < \mathbf{x}_i^{\mathrm{T}} \beta\right)}^{N} (1-0) \left| y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \right| \right) \\
&= \operatorname{argmin}_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau \right)
\end{aligned} \tag{2.4.3}$$

where ρ_τ is a loss function.

Next, we move to the multivariate case from([3]). $\boldsymbol{\beta}_\tau$ has become a $p \times k$ matrix of the regression coefficients.

$$\begin{bmatrix} \mathcal{Q}_{Y_{i1}}(\tau_1 \mid \boldsymbol{X}_i) \\ \vdots \\ \mathcal{Q}_{Y_{ip}}(\tau_p \mid \boldsymbol{X}_i) \end{bmatrix} = \boldsymbol{\beta}_\tau \boldsymbol{X}_i, \tag{2.4.4}$$

where $\mathcal{Q}_{Y_{ij}}(\tau_j \mid \boldsymbol{X}_i)$ denotes the τ_j -level quantile regression function of Y_{ij} given $\boldsymbol{X}_i$. Since the quantile regression can provide us the whole distribution of the dataset without knowing its shape, namely, non-parametric. And the SMAL has a direction definition based on the quantiles which can be applied directly to the quantile regression analysis. And in practice, when we are given a data, without any prior knowledge of the shape of the data, we thus cannot apply the common linear regression or polynomial regression. This makes the scale-generalized multivariate asymmetric Laplace distribution a widely used distribution to fit the real data.

Chapter 3

Copula

The very important tool we are using in the thesis is copula. When given a multivariate distribution, we can transform the density distribution into the copula form which can help us separate the marginal distributions from the dependency structure.(Martin Haugh,2016[6]). Especially, we can talk more about the research on the measurement on the association between variables, including the Spearman's ρ and the tail dependence coefficients, those we will explain detailedly in Chapter 4.

3.1 Introduction to Copulas

Following(Thorsten Schmidt, 2006[7]), we have a definition of the d- dimensional copula. A d-dimensional copula C: $[0,1]^p :\rightarrow [0,1]$ is a function which is a cumulative distribution function with uniform marginals. Assume we have a uniform distribution $U \sim U[0,1]$, F_X is a continuous CDF of X, thus $P\left(F^{\leftarrow}(U) \leq x\right) = F_X(x)$ and the inverse function should be, $F_X(X) \sim U[0,1]$. Given a p-variate vector $\mathbf{X} = (X_1, \ldots, X_p)$ with CDF F_X and continuous and increasing marginals, thus $F_{X_1}\left(X_1\right), \ldots, F_{X_p}\left(X_p\right)$ is a copula, we can denote it by C_X, and

$$\begin{aligned} C_X\left(u_1, \ldots, u_p\right) &= P\left(F_{X_1}\left(X_1\right) \leq u_1, \ldots, F_{X_p}\left(X_d\right) \leq u_p\right) \\ &= P\left(X_1 \leq F_{X_1}^{-1}\left(u_1\right), \ldots, X_p \leq F_{X_p}^{-1}\left(u_p\right)\right), \\ &= F_{\mathbf{X}}\left(F_{X_1}^{-1}\left(u_1\right), \ldots, F_{X_p}^{-1}\left(u_p\right)\right) \end{aligned} \tag{3.1.1}$$

if we let $u_i := F_{X_i}\left(x_i\right)$ for $i \in \{1, ..., p\}$, then

$$F_{\mathbf{X}}\left(x_1, \ldots, x_d\right) = C_X\left(F_{X_1}\left(x_1\right), \ldots, F_{X_p}\left(x_p\right)\right) \tag{3.1.2}$$

And Equation (3.1.2) gives us the Sklar's Theorem.

3.2 Copulas in this thesis

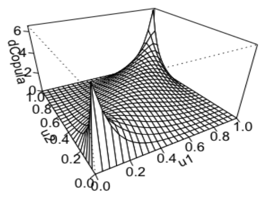
In this thesis, we will use three typical copulas, the Gaussian copula the Clayton copula, and the Gumbel copula. First, we introduce the Gaussian copula. The Gaussian copula is a transformation of the multivariate Gaussian distribution, i.e., the Gaussian copula is a distribution over the unit cube $[0,1]^d$ which is constructed from the d-dimensional multivariate normal distribution. We use the Gaussian copula instead of the bivariate normal because of its generality.

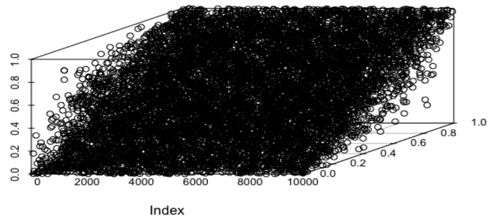
And the Clayton and Gumbel copulas are bivariate copulas, called the Archimedean copula(Nelsen, R. B. (2006)[8]). Archimedean copulas admit an explicit formula, which is, sometimes, not possible for the Gaussian copula. And we have the following explicit expression of the Clayton copula(Clayton, David G.,1978[9]) and the Gumbel copula

Clayton	$\left[\max\left\{u^{-\theta}+v^{-\theta}-1;0\right\}\right]^{-1/\theta}$	$\theta \in [-1,\infty)\backslash\{0\}$
Gumbel	$\exp\left[-\left((-\log(u))^{\theta}+(-\log(v))^{\theta}\right)^{1/\theta}\right]$	$\theta \in [1,\infty)$

(3.2.1)

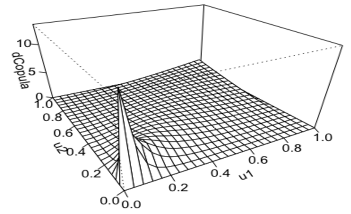
From the defination equation of the Clayton copula, we can tell when $\theta \to 0$ we obtain the independence copula, and when $\theta \to \infty$ the Clayton copula becomes the comonotonicity copula. Also, considering the Gumbel copula, for $\theta = 1$ we reach the independence copula, then when $\theta \to \infty$ the Gumbel copula becomes the comonotonicity copula(Thorsten Schmidt,2006[7]). The following figures gives us the density plots and the scatter plots of the Gaussian copula, Clayton copula, and Gumbel copula.

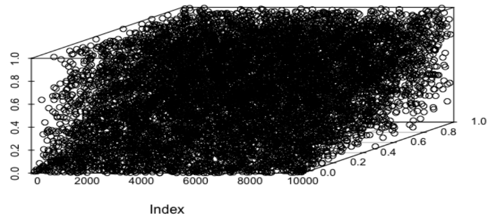




(a) density plot of Gaussian copula with $\rho = 0.8$

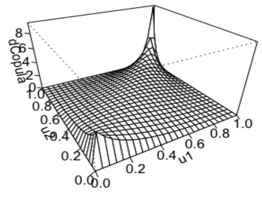
(b) scatter plot of Gaussian copula with $\rho = 0.8$

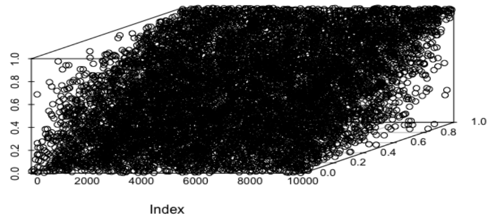




(c) density plot of Clayton copula with $\theta = 2$

(d) scatter plot of Clayton copula with $\theta = 2$





(e) density plot of Gumbel copula with $\theta = 2$

(f) scatter plot of Gumbel copula with $\theta = 2$

3.3 Frequentist and Bayesian statistics with respect to copulas

It is possible to show (Clara Grazian and Brunero Liseo,2017[10]) that the density of a vector $\boldsymbol{X} = (X_1, .., X_p)$ has a copula representation in the continuous case, which is given by

$$f(x \mid \lambda, \theta) = c(u; \theta) \prod_{j=1}^{p} f_j \left(x_j \mid \lambda_j\right). \tag{3.3.1}$$

where $c(u; \theta)$ is the derivative with respect to θ of the p-dimensional function C_X according to the equation (3.1.2). Since we have discussed the posterior distribution of the vector in section 2.2. Thus we can express the posterior as the following equation

$$\begin{aligned}
\pi(\theta, \lambda \mid x) &\propto L(x \mid \lambda, \theta) \times \pi(\theta, \lambda) \\
&\propto \prod_{j=1}^{p} f_j \left(x_j \mid \lambda_j, \theta\right) \times \pi(\theta, \lambda) \\
&\propto \pi(\theta, \lambda) \prod_{i=1}^{n} \left[c\left(u_i; \theta\right) \prod_{j=1}^{p} f\left(x_{ij}; \lambda_j\right) \right]
\end{aligned} \tag{3.3.2}$$

The most difference between frequentist statistics and Bayesian statistics is the way we treat a data. For the frequentist statistics, all the estimations of the parameters are based on the original data itself. The frequentist method interprets the probability as a statistical mean by applying a huge number of independent experiments, while, the Bayesian method would rely on the prior of the parameter. Thus compared with the frequentist methods, the Bayesian method is more flexible since it assumes the parameter has a prior distribution instead of being a fixed value, especially, for the non-parametric case. In the parametric case, the estimations from the frequentist methods are based on maximizing the value of θ and λ from the likelihood estimation, i.e., the MLE of those parameters. (Joe,2015[11]) However, for the Bayesian method, the joint posterior distribution will be evaluated by the Monte Carlo algorithm with θ, and $(\lambda_1, .., \lambda_p)$ generated separately by using a Gibbs sampling(Clara Grazian and Brunero Liseo,2017[10]). Therefore, we have another way to look back at the frequentist statistics and Bayesian statistics. For the most of the time, the frequentist change the problem into an optimization problem while the Bayesian method will change the same problem into the sampling problem. This is the biggest difference between these two methods.

Chapter 4

The dependence structure of the SMAL

Now we are moving to the dependence structure of the SMAL distribution. To talk about the dependence structure we need consider two factors, Spearman's rho and the tail dependence coefficients.

4.1 Spearman's rho

Spearman's rho is also called Spearman's rank correlation coefficient, it is statistical parameter which is widely used in the non-parametric cases for the measurement of dependence between the rankings of two variables. For the bivariate case, the definition of the Spearman's rho is given by

$$\rho_n := \operatorname{Corr}\left(F_1\left(X_1\right), F_2\left(X_2\right)\right) \tag{4.1.1}$$

where F_1, F_2 are the marginals of two random variable X_1, X_2, respectively.(Thorsten Schmidt,2006[7]).

Let X and Y be continuous random variables with the distribution functions F and G, respectively. And in the copula representation, $u = F(x)$ and $v = G(y)$. Therefore, the Spearman's ρ in terms of $C(u,v)$ is given by (Schweizer and Wolff, 1981[12])

$$\begin{aligned}\rho &= 12\int_0^1\int_0^1 (C(u,v) - uv)dudv \\ &= 12\int_0^1\int_0^1 C(u,v)dudv - 12\int_0^1\int_0^1 uvdudv \\ &= 12\int_0^1\int_0^1 C(u,v)dudv - 3.\end{aligned} \tag{4.1.2}$$

Now we will estimate the Spearman's ρ, which means we need to find the posterior distribution of the Spearman's ρ, say, $\hat{\rho}$. (Clara Grazian and Brunero Liseo,2017[10])) gives us an estimation of Spearman's ρ which is given by

$$\hat{\rho}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{12}{n^2-1}R(x_i)R(y_i)\right) - 3\frac{n+1}{n-1}, \tag{4.1.3}$$

where

$$R(x_i) = \operatorname{rank}(x_i) = \sum_{k=1}^{n} \mathbb{I}(x_k \leq x_i), \quad R(y_i) = \operatorname{rank}(y_i) = \sum_{k=1}^{n} \mathbb{I}(y_k \leq y_i), \quad i = 1, \ldots, n. \tag{4.1.4}$$

Here we can try to prove the equation (4.1.3)

$$\begin{aligned}
\rho = \frac{S_{xy}}{S_x S_y} &= \frac{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \mathbb{E}(R(x))\right)\cdot\left(R(y_i) - \mathbb{E}(R(y))\right)}{\sqrt{\operatorname{var}(R(x))}\cdot\sqrt{\operatorname{var}(R(y))}} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)\cdot\left(R(y_i) - \overline{R(y)}\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)^2\right)\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\left(R(y_i) - \overline{R(y)}\right)^2\right)}} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} R(x_i)R(y_i) - \frac{1}{n}\cdot\frac{n+1}{2}\sum_{i=1}^{n} R(x_i) - \frac{1}{n}\cdot\frac{n+1}{2}\sum_{i=1}^{n} R(y_i) + \left(\frac{n+1}{2}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i)\right)^2 - \left(\frac{n+1}{2}\right)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} R(x_i)R(y_i) - \frac{n+1}{n}\sum_{i=1}^{n} R(x_i) + \left(\frac{n+1}{2}\right)^2}{\frac{1}{n}\cdot\frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} R(x_i)R(y_i) - \frac{n+1}{n}\cdot\frac{n(n+1)}{2} + \left(\frac{n+1}{2}\right)^2}{\frac{n+1}{12}\left[2(2n+1) - 3(n+1)\right]} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} R(x_i)R(y_i) - \frac{(n+1)^2}{2} + \frac{(n+1)^2}{4}}{\frac{n+1}{12}(4n+2-3n-3)} \\
&= \frac{1}{n}\sum_{i=1}^{n} R(x_i)R(y_i)\cdot\frac{12}{(n-1)(n+1)} - \frac{12}{(n-1)(n+1)}\frac{(n+1)^2(2-1)}{4} \\
&= \frac{12}{n(n^2-1)}\sum_{i=1}^{n} R(x_i)R(y_i) - 3\cdot\frac{n+1}{n-1} \\
&= \hat{\rho}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{12}{n^2-1}R(x_i)R(y_i)\right) - 3\frac{n+1}{n-1}
\end{aligned} \tag{4.1.5}$$

The frequentist method to find the estimator of the Spearman's ρ is not directly computed, however, it can be found through the asymptotic variance of ρ_n from(Borkowf,2002 [13]), that is

$$\sigma_{\mathrm{a}}^2(\hat{\rho}_{\mathrm{n}}) = 144\left(-9\phi_1^2 + \phi_3 + 2\phi_4 + 2\phi_5 + 2\phi_6\right) \tag{4.1.6}$$

where

$$\begin{aligned}
&\phi_1 = \mathbb{E}\left[F(X_1)G(Y_1)\right] \\
&\phi_2 = \mathbb{E}\left[(1 - F(X_1))^2(1 - G(Y_1))^2\right] \\
&\phi_3 = \mathbb{E}\left[(1 - F(X_1) - G(Y_2) + H(X_1, Y_2))(1 - F(X_2))(1 - G(Y_1))\right] \\
&\phi_4 = \mathbb{E}\left[(1 - F(\max\{X_1, X_2\}))(1 - G(Y_1))(1 - G(Y_2))\right] \\
&\phi_5 = \mathbb{E}\left[(1 - F(X_1))(1 - F(X_2))(1 - G(\max\{Y_1, Y_2\}))\right]
\end{aligned} \tag{4.1.7}$$

with H being the joint distribution of (X_1, Y_1) and (X_2, Y_2), F and G being the marginal distribution of X and Y, respectively. Based on the knowledge above, we can apply the algorithm designed by (Clara Grazian and Brunero Liseo,2017[10]).

Algorithm 0: Spearman's ρ algorithm

```
Change the simulated data into the copula form
Generate a sample of n pseudo-observations from the previous copula
    $$u = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ \cdots & \cdots \\ u_{n1} & u_{n2}, \end{pmatrix} \qquad (4.1.8)$$
(i) for $b = 1 \to B$ do
    draw $\rho^{(b)}$ from the assumed prior distribution, such as
      $\rho^{(b)} \sim Unif(-1, 1)$
    Compute a nonparametric estimate of the Spearman's $\rho$ :
      $\hat{\rho}_n = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{12}{n^2-1} R(x_i) R(y_i) \right) - 3\frac{n+1}{n-1}$
    where $R(x_i) = \text{rank}(x_i)$,$R(y_i) = \text{rank}(y_i)$ , $i = 1, ..., n$
    Compute the Bayesian exponentially tilted empirical likelihood
end
(ii) return $\hat{\rho}_n$.
```

4.2 Spearman's ρ of the SMAL distribution and specific Copulas

After introducing all the background knowledge, we can now start to find the first correlation coefficient, Spearman's ρ, of the SMAL distribution. To do this, we first need to simulate a SMAL distribution. Here we let the number of the Monte Carlo replications be 100, B = 100. The number of observations is 1000. And we use the bivariate asymmetric Laplace distribution instead of the multivariate Laplace distribution to get a conclusion without loss of generality and it will be easier for us to make comparison between the results from SMAL distribution and those three copulas. Hence the dimension parameter, $p = 2$. And we call this simulated SMAL distribution, Y. Next, we need to change Y into the copula form in order to perform the algorithm. Since we obtained this Y with 100 replication, we thus got an array with 100 matrices and each matrix being 1000×2. Then we can make a scatter plot of one entry of the array of the pseudo-observations.

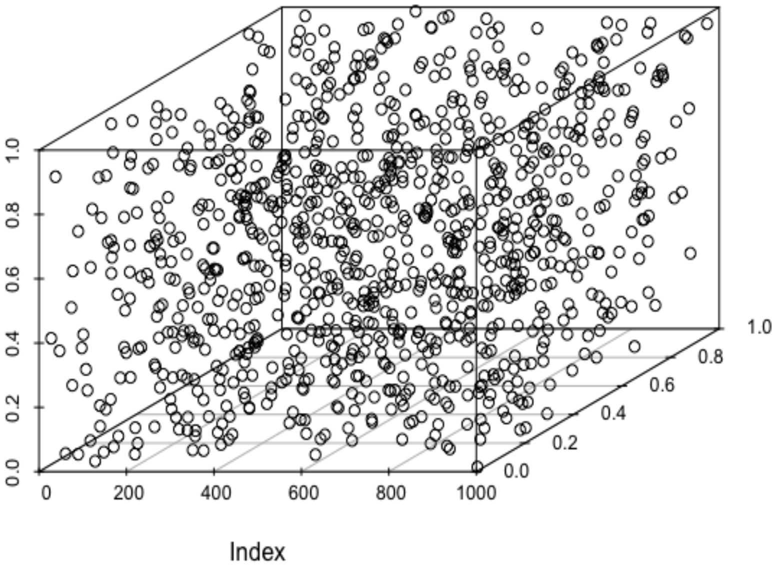
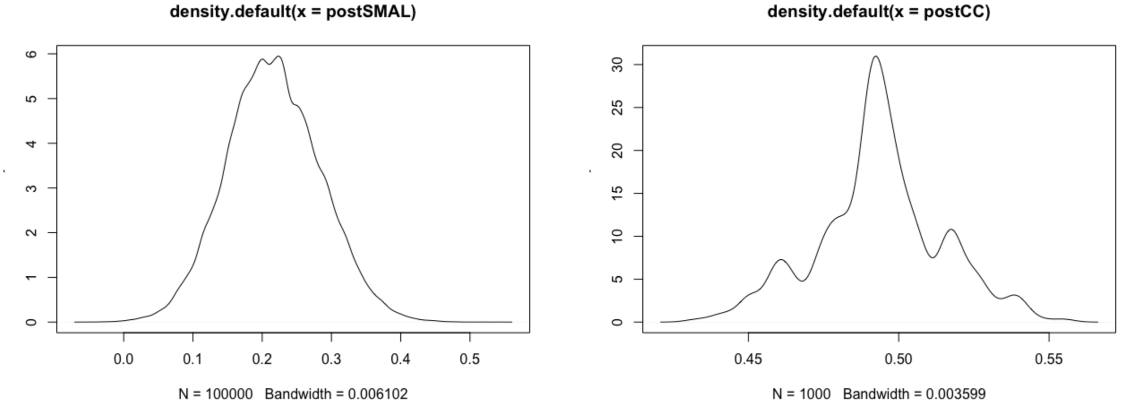


Figure 4.1: Scatter plot of the uu[,,1]

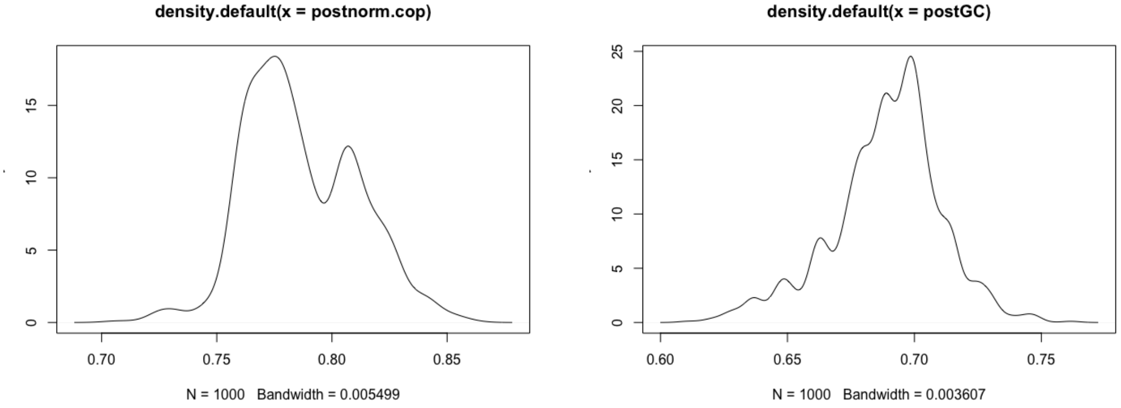
The next step is applying the algorithm to find the posterior distributions of the Spearman's ρ for both the SMAL distribution and the other three copulas. Therefore, we obtain the following density plots of those.

From those plots, we can tell, the concentration for the SMAL is around 0.2, while the concentrations for the other three copulas are around 0.5, 0.76, and 0.7 respectively. Since the closer the absolute value of Spearman's ρ to 1, the better the correlation between two variables, the strength of the association of three copulas is much better than the SMAL distribution. However, the Spearman's ρ is a non-parametric test, without having any prior knowledge or experience of the given data, it is acceptable to obtain such a 0.2 Spearman's ρ.



(a) density plot of the posterior distribution of Spearman's ρ of the SMAL distribution

(b) density plot of the posterior distribution of Spearman's ρ of the Clayton copula



(c) density plot of the posterior distribution of Spearman's ρ of the Gaussian copula

(d) density plot of the posterior distribution of Spearman's ρ of the Gumbel copula

Figure 4.2: Density plots of the posterior distributions of the Spearman's ρ from different distributions

4.3 Tail dependence coefficients

Now we will introduce the tail dependence coefficients. Tail dependence coefficients measures the comovements in the tails of the distribution of two random variables (Hartmann, Philip et al, 2004[14]). We can divide the tail dependence coefficients into two parts, including the first part, the upper tail dependence coefficient, and the second part, lower tail dependence coefficient. When talking about the dependence structure of the multivariate distribution, especially the financial risk assessment it is very complicated to reach a conclusion. The reason is we can not only consider the correlation of the whole distribution or the financial derivative, but also consider the tail dependence of the asset price (Sibuya, 1959[15]). That is, the correlation between assets when a small probability event occurs.(Ané and Kharoubi ,2003[16]). Let us focus back on the definition of the tail dependence coefficients. The notion of the tail dependence is the connection with the extreme value on the data at the tail in the bivariate distribution. It is very convenient to build the connection between the tail dependence coefficients and copulas.

Here we can take the conditional probability,$P\left(X_1 > x_1 \mid X_2 > x_2\right)$, into account again. We can interpret the conception of this conditional probability as: if we change the value of $X_2 > x_2$, does it affect the probability of $X_1 > x_1$? And when the values of x_1, x_2 are large, it comes to the tail dependence of X_1, and X_2.
Let $C(u, v)$ be the copula function of (X_1, X_2), where $u, v \in [0, 1]$, thus we can easily obtain

$$P\left(X_1 > x_1 \mid X_2 > x_2\right) = P\left(U_1 > u_1 \mid U_2 > u_2\right). \tag{4.3.1}$$

Besides, we have the survival function of copula of X_1 and X_2 from (Nelsen,2006[15])

$$\hat{C}\left(u_1, u_2\right) = u_1 + u_2 - 1 + C\left(1 - u_1, 1 - u_2\right) \tag{4.3.2}$$

And the joint survival function $\bar{C}$ for two random variables which are uniformly $(0, 1)$ distributed. Therefore, we have the following equation:

$$\begin{aligned}\bar{C}\left(u_1, u_2\right) &= P\left(U_1 > u_1, U_2 > u_2\right) = 1 - u_1 - u_2 + C\left(u_1, u_2\right) \\ &= (1 - u_1) + (1 - u_2) - 1 + C(u_1, u_2) \\ &= \hat{C}(1 - u_1, 1 - u_2)\end{aligned} \tag{4.3.3}$$

Now assume X and Y be two continuous random variables with the marginals $F(X)$ and $G(Y)$, respectively. And the copula function of these two variable being $C(X, Y)$. Therefore, the upper and lower tail dependence coefficients are written as

$$\begin{aligned}\lambda_U &= \lim_{u \to 1} P\left\{Y > G^{-1}(u) \mid X > F^{-1}(u)\right\} \\ &= \lim_{u \to 1} \frac{\hat{C}(1 - u, 1 - u)}{1 - u} \\ &= \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u},\end{aligned} \tag{4.3.4}$$

and

$$\begin{aligned}\lambda_L &= \lim_{u \to 0} P\left\{Y < G^{-1}(u) \mid X < F^{-1}(u)\right\} \\ &= \lim_{u \to 0} \frac{C(u, u)}{u}.\end{aligned} \tag{4.3.5}$$

We will say (X, Y) has upper and lower tail dependence when λ_U and λ_L are greater than 0, and upper and lower tail independence when $\lambda_U = 0$ and $\lambda_L = 0$. Now if we want to apply the algorithm, we have to find the non-parametric estimator of λ_U and λ_L. In (Frahm et al. 2005[17]), for the non-parametric case, the estimation of the tail dependence coefficients can be obtained from the empirical copula $\tilde{C}_n$. And the empirical copula gives the following relationship

$$\tilde{F}_n(x, y) = \tilde{C}_n\left\{\tilde{G}_n(x), \tilde{H}_n(y)\right\}. \tag{4.3.6}$$

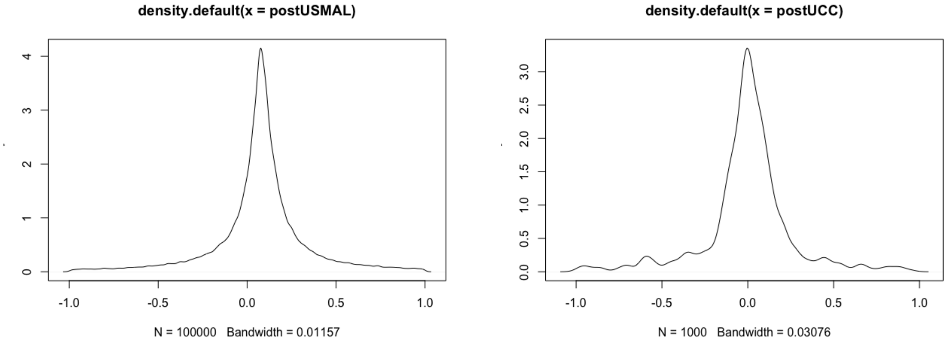
The special case of non-parametric estimation given in (Joe et al. 1992[18]) is

$$
\begin{aligned}
\hat{\lambda}_U &= 2 - \frac{1 - \tilde{C}_n((n-k)/n, (n-k)/n)}{1 - (n-k)/n} \\
\hat{\lambda}_L &= \frac{\tilde{C}_n\left(k/n, k/n\right)}{k/n}, \quad 0 < k \leq n.
\end{aligned}
\tag{4.3.7}
$$

And we have the following table, which gives us the tail dependence coefficients of the multivariate Archimedean copulas see(Salazar and Ng, 2015[19])

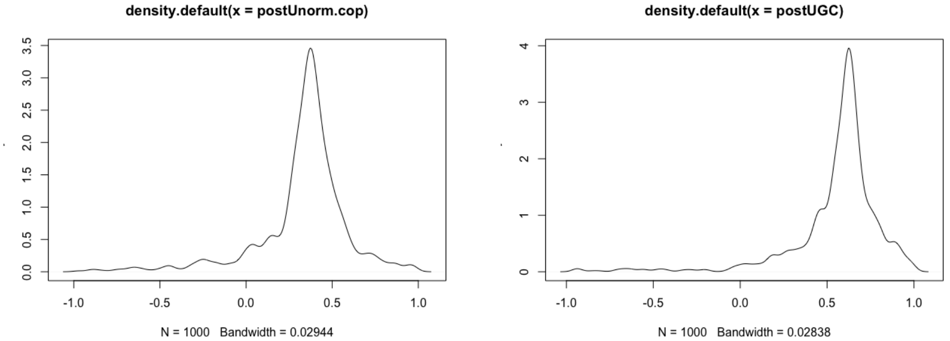
Copula	$C\left(u_1, \cdots, u_d\right)$	λ_L	λ_U
Clayton	$\left(1+\theta\left[\frac{1}{\theta}\sum_{j=1}^{d}\left(u_j^{-\theta}-1\right)\right]\right)^{-1/\theta}$	$d^{-1/\theta}$	0
Frank	$-\frac{1}{\theta}\log\left[1+\frac{\prod_{j=1}^{d}(\exp(-\theta u_j)-1)}{[\exp(-\theta)-1]^{d-1}}\right]$	0	0
Gumbel	$\exp\left[-\left[\sum_{j=1}^{d}\left(-\log u_d\right)^{\theta}\right]^{1/\theta}\right]$	0	$\sum_{r=1}^{d}(-1)^{r+1}\binom{d}{r}r^{1/\theta}$

We now can use the Bayesian estimates to find the posterior distribution of the tail dependence coefficients by presenting the density plots.



(a) density plot of the posterior distribution of upper tail dependence coefficient of the SMAL distribution

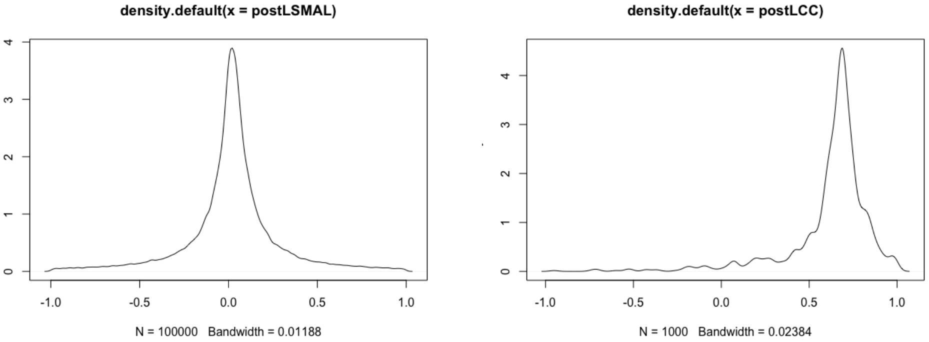
(b) density plot of the posterior distribution of upper tail dependence coefficient of the Clayton copula



(c) density plot of the posterior distribution of upper tail dependence coefficient of the Gaussian copula

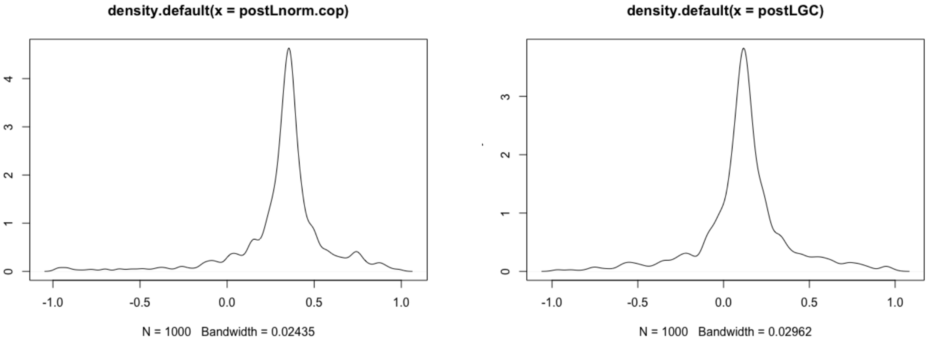
(d) density plot of the posterior distribution of upper tail dependence coefficient of the Gumbel copula

Figure 4.3: Density plots of the posterior distributions of upper tail dependence coefficients from different distributions in two dimensions.



(a) density plot of the posterior distribution of lower tail dependence coefficient of the SMAL distribution

(b) density plot of the posterior distribution of lower tail dependence coefficient of the Clayton copula



(c) density plot of the posterior distribution of lower tail dependence coefficient of the Gaussian copula

(d) density plot of the posterior distribution of lower tail dependence coefficient of the Gumbel copula

Figure 4.4: Density plots of the posterior distributions of lower tail dependence coefficients from different distributions in two dimensions.

From the figures above, we can tell both the upper and lower tail dependence coefficients of the SMAL distribution is concentrated around 0. For the Clayton copula, we can see that the upper tail dependence concentrates around 0 and the lower tail dependence concentrates around 0.7 which match the expressions from (Salazar,2015), where $\lambda_U = 0$ and $\lambda_L = d^{-1/\theta} = 2^{-1/2} = 0.707$. From the inspection of the Clayton we can find that the tail dependent coefficients found by the algorithm in this thesis is acceptable.
Overall, according to the Spearman's ρ and the tail dependence coefficients, we can tell that the SMAL distribution is good for fit the data since it has a definition on quantiles first. And second, more events from the SMAL distribution will concentrate on the majority instead of being on the tail along with the acceptable Spearman's ρ.

Chapter 5

Simulation Study

In this chapter, we will move to the simulation study. To do this, I have changed the correlation matrix of the SMAL distribution and making other parameters fixed in order to see how the Spearman's ρ and the tail dependence coefficients change. Here we use the algorithm above with 1000 replications. By fixing a scale on a grid (between -1 and 1), I will repeat the simulation and find the posterior distributions for the Spearman's ρ and the tail dependence coefficients.

As is mentioned in Chapter 1, we know Ψ is the correlation matrix. In our case, Ψ is a 2×2 matrix, $\Psi = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix}$, where we want to change a and b to change the correlation matrix. Here we are considering two cases, $a = b$ and $a = -b$. Since a and b are constrained in $[-1, 1]$, thus we have the following scenarios in which the correlations are equal:

$$\begin{gathered} \Psi_{g1} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \Psi_{g2} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \\ \Psi_{g3} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Psi_{g4} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \Psi_{g5} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \end{gathered} \tag{5.0.1}$$

and in which the correlations are the opposite:

$$\begin{gathered} \Psi_{ng1} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \Psi_{ng2} = \begin{bmatrix} 1 & -0.5 \\ 0.5 & 1 \end{bmatrix} \\ \Psi_{ng3} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Psi_{ng4} = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 1 \end{bmatrix} \Psi_{ng5} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \end{gathered} \tag{5.0.2}$$

Since $\Psi_{g3} = \Psi_{ng3}$, we will repeat four times of the simulations for the case that the correlations are in the opposite sign. The reason why we only change the value out of the diagonal is because the ones on the diagonal are associated with the variance, so there is not much effect on the dependence structure.

5.1 Simulation study on Spearman's ρ

As what we have done before, we plot the density of the posterior distributions of Spearman's ρ of two cases separately.

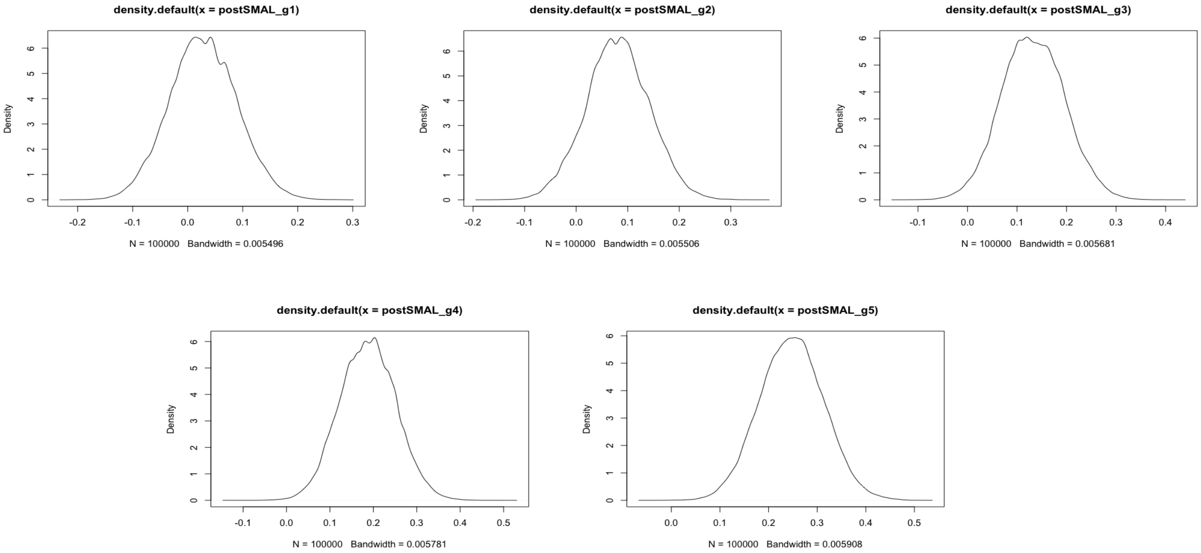


Figure 5.1: Posterior distributions of Spearman's ρ for the equal correlation.

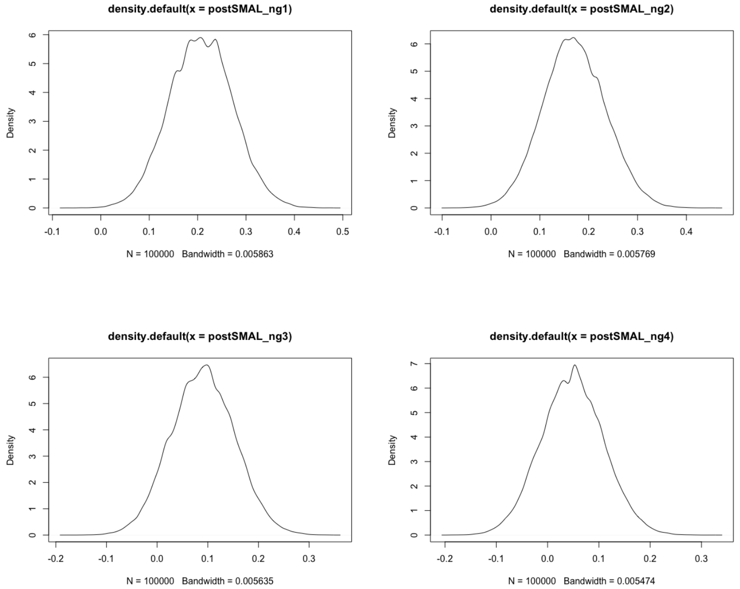


Figure 5.2: Posterior distributions of Spearman's ρ for the opposite correlation.

From the density plot we can tell the curve shifts to the right when adding the correlation from -1 to 1 for the equal correlation case, and from the opposite correlation case, the curve keeps shifting to the right when a is decreasing. To see this in a clearer way, we saved the mean of the posterior distributions of the SMAL distributions from those two cases. The table presents the results as what we mentioned before.

Spearman's rho for equal correlation				
(1,-1,-1,1)	(1,-0.5,-0.5,1)	(1,0,0,1)	(1,0.5,0.5,1)	(1,1,1,1)
0.029021	0.079083	0.1338221	0.186915	0.248423
Spearman's rho for opposite correlation				
(1,-1,1,1)	(1,-0.5,0.5,1)	(1,0,0,1)	(1,0.5,-0.5,1)	(1,1,-1,1)
0.0487604	0.089154	0.1338221	0.1710109	0.2062635

Table 5.1: Simulation study on the Spearman's ρ

5.2 Simulation study on tail dependence coefficients

Now we move to the simulation on the tail dependence coefficients, we give the density plots of the upper and lower tail dependence and the mean of the posterior distributions.

Upper tail dependence for equal correlation				
(1,-1,-1,1)	(1,-0.5,-0.5,1)	(1,0,0,1)	(1,0.5,0.5,1)	(1,1,1,1)
0.06604293	0.06424079	0.06252416	0.09316917	0.1204435
Upper tail dependence for opposite correlation				
(1,-1,1,1)	(1,-0.5,0.5,1)	(1,0,0,1)	(1,0.5,-0.5,1)	(1,1,-1,1)
0.06321448	0.06508844	0.06252416	0.06316775	0.09244834

Table 5.2: Simulation study on the upper tail dependence

Lower tail dependence for equal correlation				
(1,-1,-1,1)	(1,-0.5,-0.5,1)	(1,0,0,1)	(1,0.5,0.5,1)	(1,1,1,1)
-0.0008705791	0.001986895	0.0004194878	0.05808117	0.08717094
Lower tail dependence for opposite correlation				
(1,-1,1,1)	(1,-0.5,0.5,1)	(1,0,0,1)	(1,0.5,-0.5,1)	(1,1,-1,1)
0.000683531	-0.0007985829	0.0004194878	-7.758008e-05	0.08691091

Table 5.3: Simulation study on the lower tail dependence

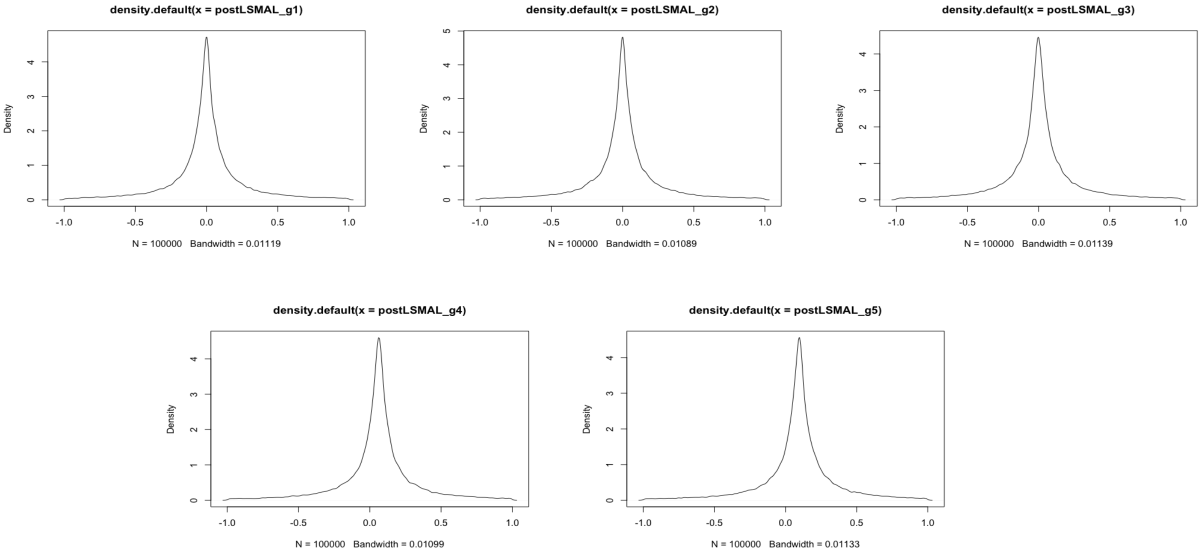


Figure 5.3: Posterior distributions of λ_L for the equal correlation.

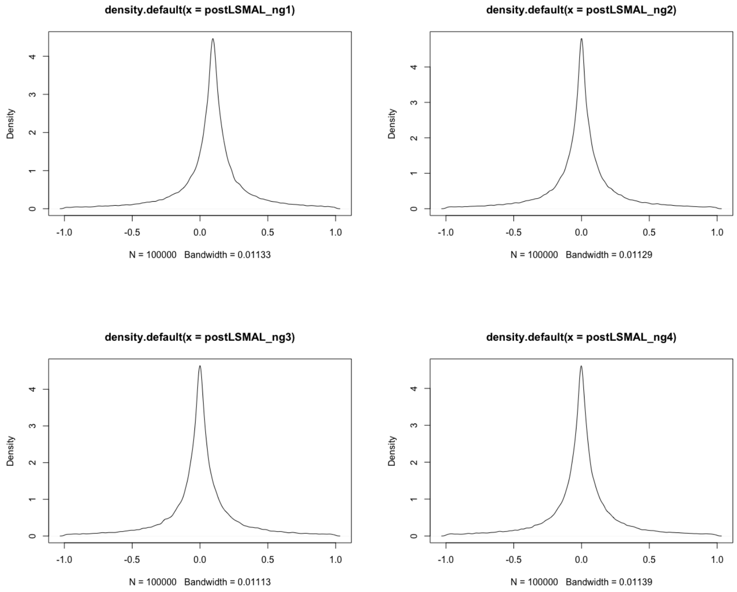


Figure 5.4: Posterior distributions of λ_L for the opposite correlation.

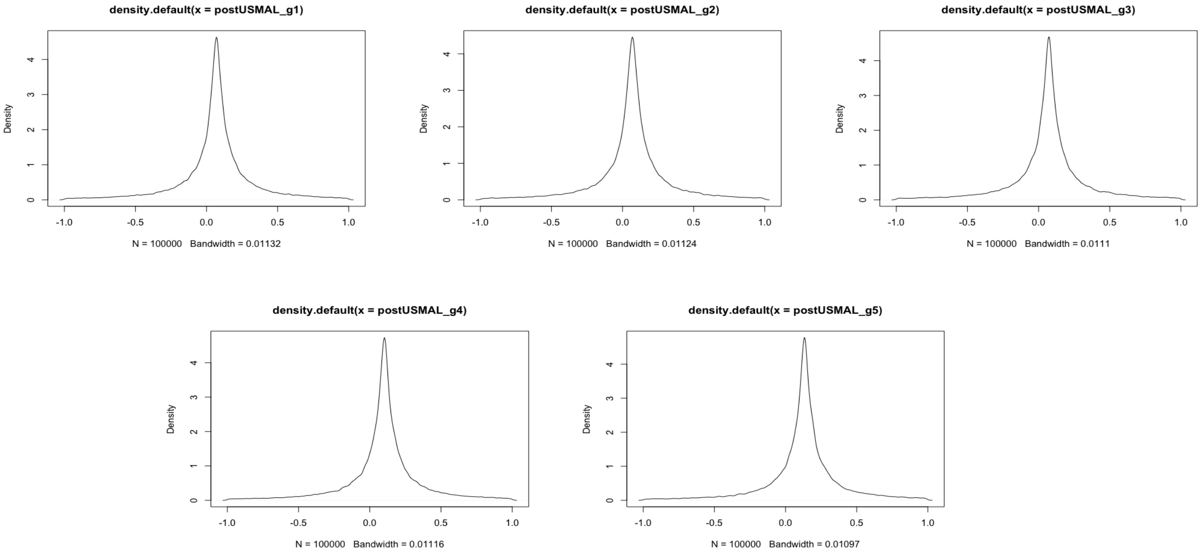


Figure 5.5: Posterior distributions of λ_U for the equal correlation.

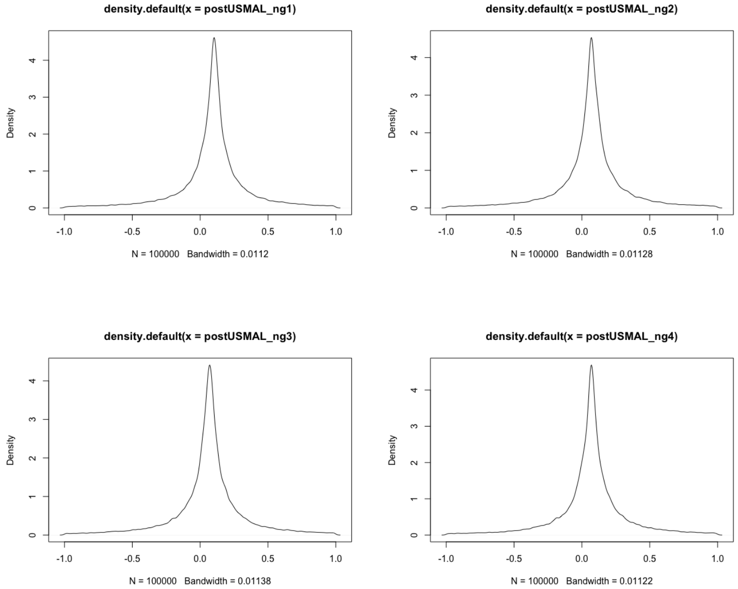


Figure 5.6: Posterior distributions of λ_U for the opposite correlation.

By inspecting the means of the posterior distributions and the density plots, we can reach a conclusion that for the upper and lower tail dependence coefficients, the upper and lower tail dependence coefficients barely vary and concentrate around with the change of the correlation matrix.

Chapter 6

Real data inspection

Now we want to see if the SMAL distribution fits the real data even if it is very used in practice.

6.1 Introduction to the dataset "body"

The real dataset we will use in our thesis to perform the test is "body" from the R package "gclus". This dataset contains 25 variables, including 21 body dimension measurements along with 4 factors, age, weight, height, and gender. The dataset contains the measurements on 507 individuals. The physical measurements include the chest diameter at nipple level(ChestD), wrist diameter(WristD), chest girth(ChestG), calf maximum girth(CalfG), etc in the unit of centimeter. At first we can take a look at the association between some of the random variables in the dataset.

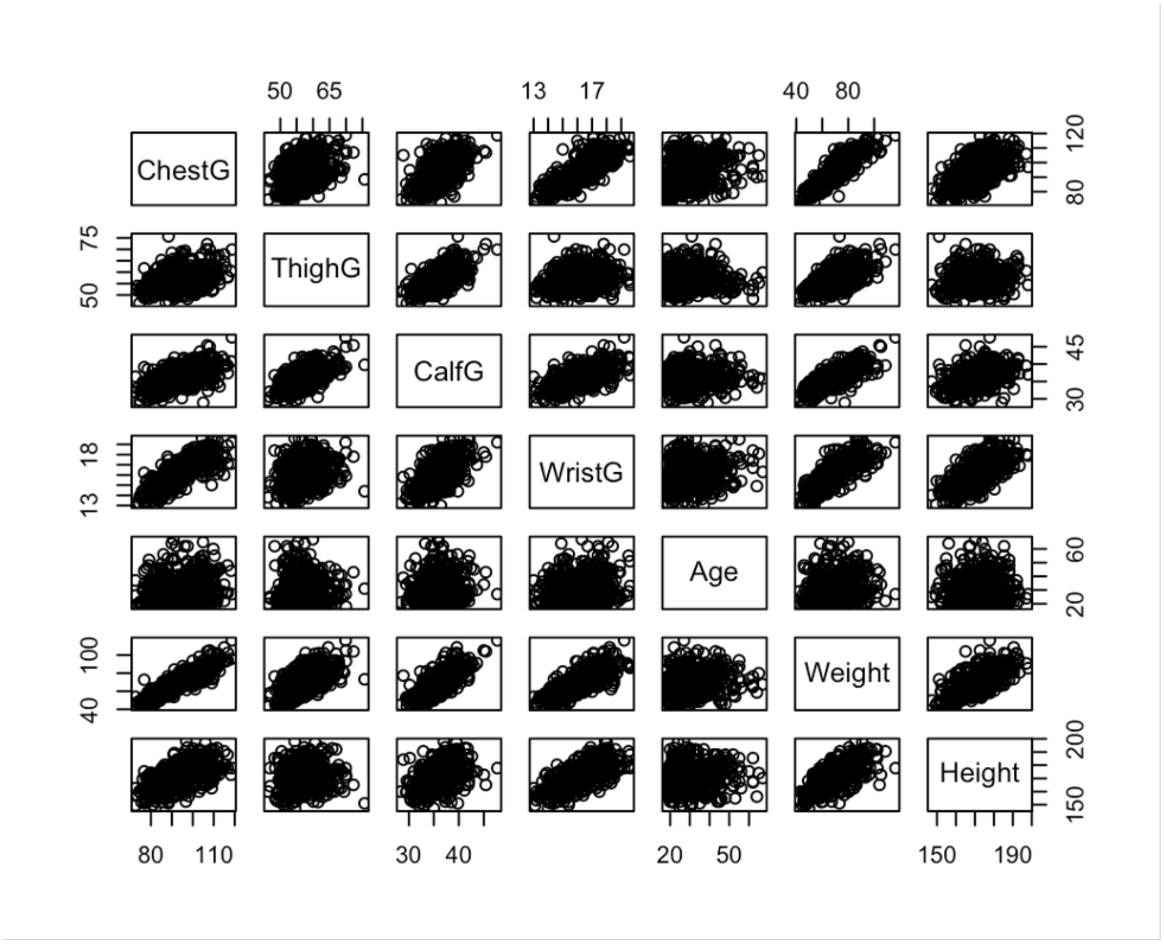


Figure 6.1: Pairs plot of some variables from "body"

From the pairs plot, we can conclude that by looking at the first row of the plots, this single plot gives us an idea of the relationship between each pair of variables in our dataset. Taking the chest girth(ChestG) as an example, we can tell that this measurement seems to be positive correlated to the weight, and less correlated with age.

6.2 Non-parametric Bayesian quantile regression

We determine to use the calf maximum girth and the thigh maximum girth as two response variables, i.e., $Y =$ (Calf maximum girth, Thigh maximum girth) and $X =$(weight, age, BMI index, height), where BMI index $= weight/height^2$ to find the posterior distribution of the Spearman's ρ and the tail dependence coefficients by applying the bivariate analysis by using the Bayesian quantile regression. After applying the algorithm, we can find the results of the non-parametric Bayesian quantile regression on the true dataset. We made the density plots of the Spearman's ρ and the tail dependence coefficients of the true dataset.

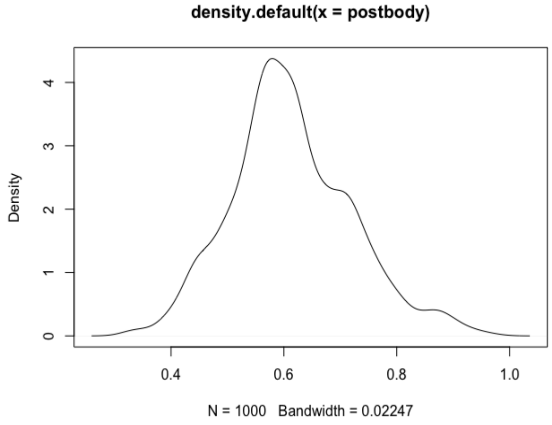
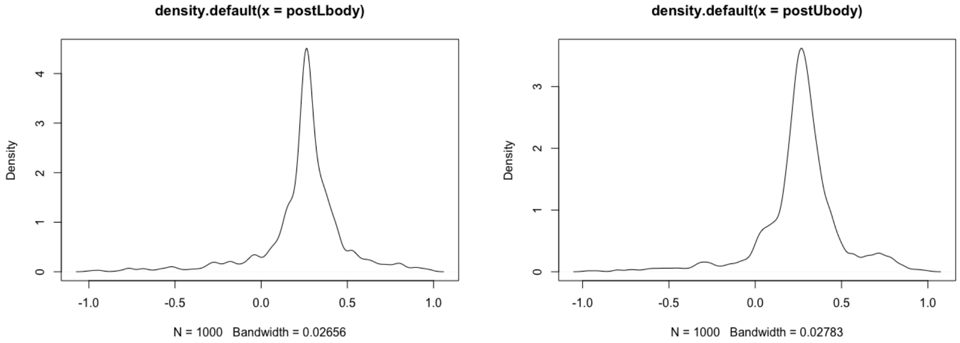


Figure 6.2: Density plot of the posterior distribution of the Spearman's ρ of the real dataset



(a) Density plot of the posterior distribution of lower tail dependence coefficient of the real dataset

(b) Density plot of the posterior distribution of upper tail dependence coefficient of the real dataset

Figure 6.3: density plots of the posterior distributions of the tail dependence coefficients of the real data

The above figures gives us the Spearman's ρ concentrates around 0.6 and both the upper and lower tail dependence coefficients for the real data are around 0.3. Since we have a Spearman's ρ at 0.6 which gives a decent correlation between X and Y, this means Y can almost be interpreted as a monotone function of X. And from the tail dependence coefficients we can tell that the real data sometimes contains the events with the extreme value.

6.3 Simulate An SMAL distribution for testing

In order to test if the SMAL distribution fits the real dataset, we will simulate a new SMAL for testing with the number of dimensions $p = 2$, and the number of regressors $k = 4$. we set the other parameter fixed as before and then we obtain the following density plots

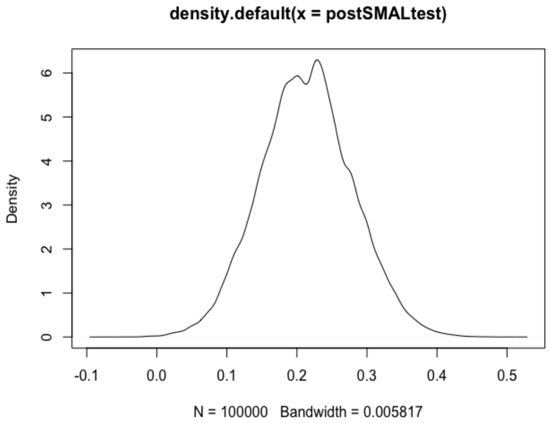
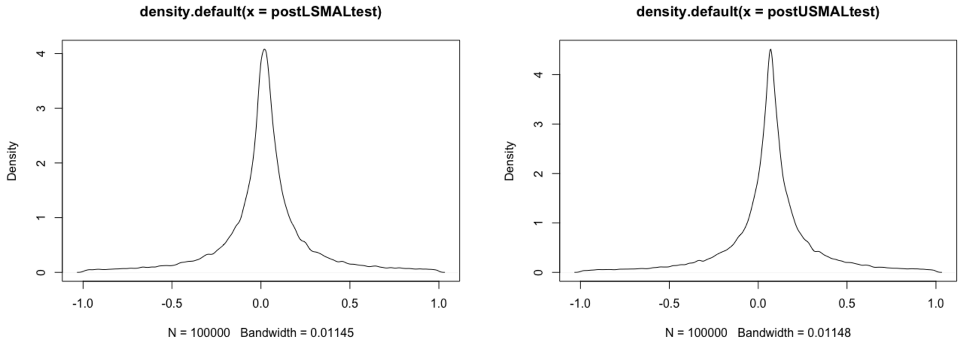


Figure 6.4: Density plot of the posterior distribution of the Spearman's ρ of the testing SMAL



(a) Density plot of the posterior distribution of lower tail dependence coefficient of the testing SMAL

(b) Density plot of the posterior distribution of upper tail dependence coefficient of the testing SMAL

Figure 6.5: Density plots of the posterior distributions of the tail dependence coefficients of the testing SMAL

From the figures above, we will reach a conclusion that both the Spearman's ρ and the tail dependence coefficients are not the same, the Spearman's ρ is about 0.6 in the real dataset instead of 0.2 from the SMAL distribution, also, both of the upper and lower tail dependence coefficients for the real data are around 0.3 instead of 0 from the SMAL.

CHAPTER 7

Conclusion

In this work, we first introduce the asymmetric Laplace distribution and the multivariate asymmetric Laplace distribution. And we proposed the main point of the thesis : is the dependence structure imposed by the SMAL reasonable for the dataset at hand? To obtain the final result, we introduce the non-parametric Bayesian quantile regression and the most important tool we are using in thesis, copulas. According to the definitions and properties of those commonly used copulas, we implement the Bayesian quantile regression algorithm to test the association between two random variables. And the coefficients we focus on are the Spearman's ρ and the tail dependence coefficients. Spearman's ρ can be used to estimate the correlation between two random variables X and Y and the correlation between these two random variables can be expressed by an monotonic function, when one of the variables can be expressed as a good monotone function of the other variable (i.e., the two variables have the same change trend), the Spearman's ρ between the two variables can reach $+1$ or -1. And tail dependence coefficient can be used to estimate the dataset with the extreme value. Also we have conducted the simulation study to find the relationship between the correlation matrix and the dependence structure. Then we apply the test on the real data and make a comparison between the real dataset and the simulating SMAL in the same dimensions and with equal regressors. At last we reach a conclusion that although the SMAL is used for practical reasons, it does not fit our true dataset well. The reason why the SMAL is widely used is just because it has a definition directly on the quantiles, and the likelihood based on the SMAL can provide an automatic choice of the optimal level of penalization.

Of course, there are still some problems to be solved in this thesis and we have the future research prospects, mainly in the following:

First of all, we mainly focused on the two-dimensional case. Although bivariate is the basis of multivariate and has the advantages of intuitive and easy to understand, however, how to promote duality to pluralism is still a problem. Especially, when the dimension increases, the complexity of the model and the difficulty of estimation will increase exponentially. There may be other simpler methods to solve these problems. In addition, extreme value theory is also a method developed in recent years to study the extreme value and tail dependence. The combination of copula and tail dependence coefficients can be dig deeper.

References

[1] Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics. 2007 Jan;8(1):140-54. doi: 10.1093/biostatistics/kxj039. Epub 2006 Apr 24. PMID: 16636139.

[2] Kotz, S., Kozubowski, T. J., amp; Podgorski, K. (2001). The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance. New York: Springer.

[3] Lea Petrella, Valentina Raponi, 2019 Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress, Journal of Multivariate Analysis, Volume 173, Pages 70-84, ISSN 0047-259X

[4] Christopher M. Bishop (2006). Pattern Recognition and Machine Learning. Springer. pp. 21–24. ISBN 978-0-387-31073-2.

[5] Koenker, R., Bassett, G. (1978). Regression Quantiles. Econometrica, 46(1), 33-50. doi:10.2307/1913643

[6] Martin Haugh (2016). An Introduction to Copulas

[7] Schmidt, T. (2006). Coping with Copulas.

[8] Nelsen, R.B. 2006. An Introduction To Copulas. 2nd ed. Portland.

[9] Clayton, David G. (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence". Biometrika. 65 (1): 141–151. doi:10.1093/biomet/65.1.141. JSTOR 2335289.

[10] Grazian, Clara; Liseo, Brunero. Approximate Bayesian Inference in Semiparametric Copula Models. Bayesian Anal. 12 (2017), no. 4, 991–1016. doi:10.1214/17-BA1080.

[11] Joe, H.(2015). Dependence modeling with copulas, volume 134 of Monographs on Statistics and Applied Probability. CRC Press, Boca Raton, FL.

[12] Schweizer, B.; Wolff, E. F. On Nonparametric Measures of Dependence for Random Variables. Ann. Statist. 9 (1981), no. 4, 879–885. doi:10.1214/aos/1176345528.

[13] Borkowf, C. B. (2002). “Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman’s rank correlation.” Comput. Statist. Data Anal., 39(3): 271–286.

[14] Hartmann, Philip; Straetmans, Stefan T.M.; De Vries, Casper G. (2004). "Asset Market Linkages in Crisis Periods". Review of Economics and Statistics. 86 (1): 313–326. doi:10.1162/003465304323023831.

[15] Sibuya, M. Bivariate extreme statistics, I. Ann Inst Stat Math 11, 195–210 (1960).

[16] Ane, Thierry;Kharoubi, Cécile. (2003). Dependence Structure and Risk Measure. The Journal of Business. 76. 411-438. 10.1086/375253.

[17] Frahm, G., Junker, M., and Schmidt, R. (2005). “Estimating the tail-dependence coefficient: properties and pitfalls.” Insurance Math. Econom., 37(1): 80–100.

[18] Joe, H., Smith, R. L., and Weissman, I. (1992). “Bivariate threshold methods for extremes.” J. Roy. Statist. Soc. Ser. B, 54(1): 171–183.

[19] Salazar, Y. and Ng, W. L. (2015). “Nonparametric estimation of general multivariate tail dependence and applications to financial time series.” Statistical Methods & Applications, 24(1): 121–158.