Newcastle University

School of Mathematics & Statistics

A copula approach to spatio-temporal modelling

Student:
Aamir Khan

Supervisor:
Prof. Richard Boys

2013-2014

Abstract

The sparsity of well-known and flexible multivariate distributions can lead to problems when trying to effectively model a set of multivariate data. Hence we look at the use of copulas to describe dependence between univariate random variables, specifically investigating the Gaussian and skew t copulas. We motivate the use of these by emphasising their advantages when using them as dependence structures for rainfall volumes, as an example. By employing Markov chain Monte Carlo methods we perform Bayesian inference on the copula parameters, as well as simulate realisations from these complex models.

Contents

1 Introduction

1.1 Background

When thinking about multivariate statistics, there are only a limited number of well known distributions that come to mind, for instance the multivariate Normal is a favoured distribution, along with the multinomial, multivariate t and Dirichlet distributions to name a few. In the context of real life applications, if we were to model a multivariate set of continuous data, $\boldsymbol{X} = (X_1, \ldots, X_p)^T$, then modelling the univariate marginal distributions, $f(x_1), \ldots, f(x_p)$, would be straightforward enough. However, we would be lucky to find a multivariate distribution $f(\boldsymbol{x})$ with these marginals. We will look into using a type of distribution function known as a copula to describe the dependence between univariate random variables, with a particular emphasis on modelling rainfall volumes.

One of the most significant areas of research, at present, is climate change. There are many factors that contribute to the ever-changing distribution of weather patterns, caused by both natural and human events, and the effects of this are felt in very different ways dependent on location.

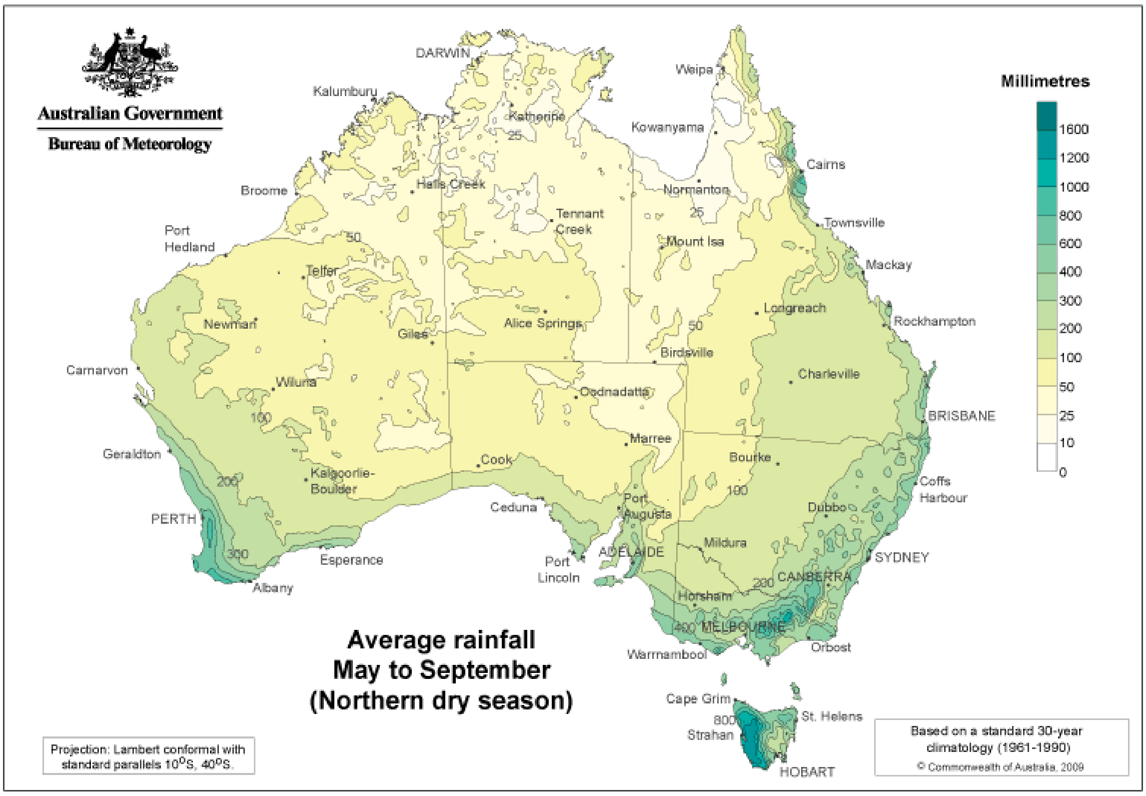

Figure 1.1: Average rainfall (mm) for Northern dry season (May to September) from 1961 to 1990. [Bureau of Meteorology, Australia]

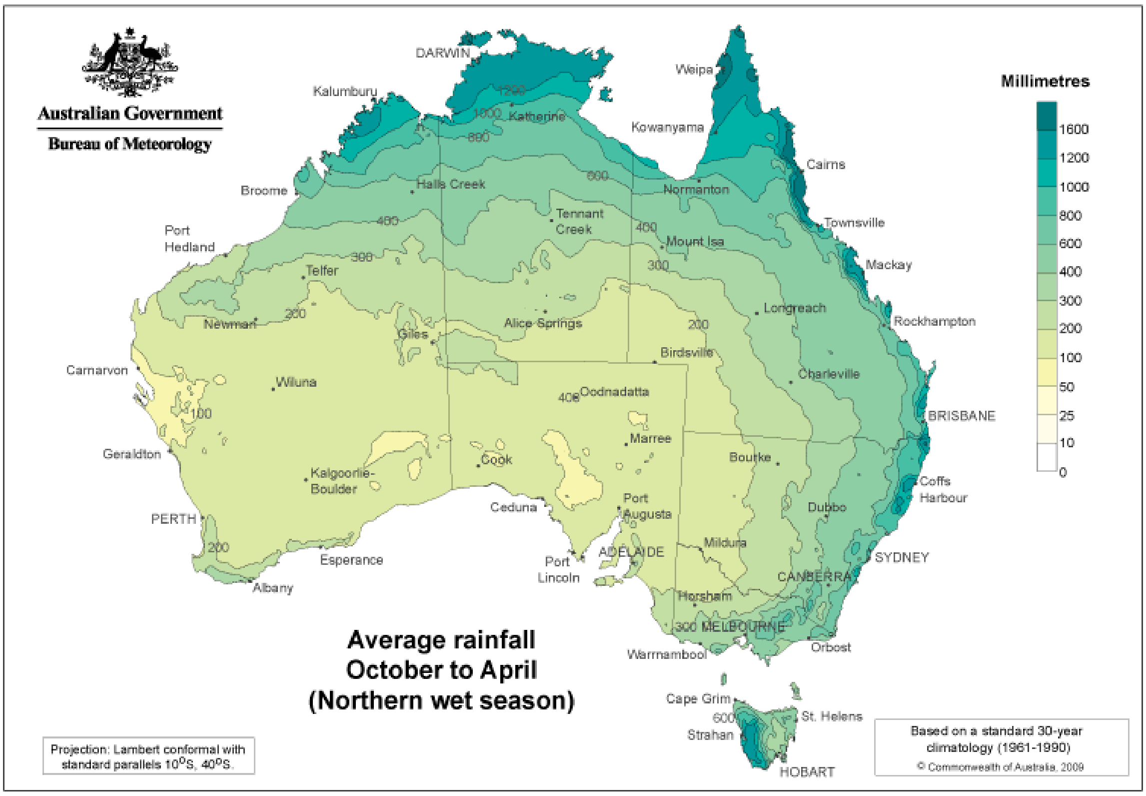

Figure 1.2: Average rainfall (mm) for Northern wet season (October to April) from 1961 to 1990. [Bureau of Meteorology, Australia]

Australia's climate is heavily affected by ocean currents, with a phenomenon known as the El Niño Southern Oscillation. The varying rainfall amounts, due to this, results in periods of drought in certain areas, whilst causing periods of heavy rainfall in others. Whether it be the flooding caused by this, or conversely the drought, the areas in question will need vital aid in order to minimise the amount of both damage and casualties, thus, a way of better understanding the distribution of rainfall, and subsequently forecasting rainfall volumes, is essential.

Figures 1.1 and 1.2, [Bureau of Meteorology, Australia], show us the average rainfall in millimetres over the two seasonal periods in an annual cycle. We see that rainfall is minimal in the centre of the country, regardless of the seasonal period due to the desert-like climate in this area. The rainfall averages are much higher for the coastal regions, dependent on the season, due to the contrasting weather system found here as opposed to the centre of the country. The mountainous regions of the country also affect the rainfall averages, such as northeastern Queensland having high rainfall totals during the Northern wet season, and Tasmania having high totals over both seasons.

In 2013, Australia observed it's warmest year since records began back in 1910, [Bureau of Meteorology, Australia]. There were several drastic bush fires in a number of areas and annual rainfall was below average across a variety of regions including western Queensland and inland New South Wales, yet above average over areas of Tasmania and parts of the south coast of Western Australia.

It is undoubtable that research in to climate change on both regional and international scales is vital, not only for the physical attributes of the landscape, but to the way we adapt as a consequence of this research.

Objective

We set out to find an appropriate flexible multivariate model to describe both the temporal and spatial dependence in rainfall amounts in a specific area of Australia. A copula based model will be used to do this. We will be assuming that the data $\mathcal{D} = \{\boldsymbol{x}_t : t = 1, \ldots, T\}$, with $\boldsymbol{x}_t = (x_{t,1}, \ldots, x_{t,p})$, can be described using a joint model $f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1})$. The T here represents the final time point of the data we will be using, since we essentially have a time series of rainfall volumes, and the p will represent the number of sites used in the model, thus $x_{t,i}$, $i = 1, \ldots, p$, is the rainfall amount for site i at time t. Modelling the site-specific time-series should be straightforward enough and then we must describe the dependence between each of these sites by utilising a multivariate distribution known as a copula.

1.2 Copulas

Before we explain how the data will be modelled in more depth, we first need to give some background about what copulas actually are and how they can be used. Nelsen [2006] gives two descriptions of what copulas are, one stating that they are functions that couple multivariate distributions to their one dimensional marginal distributions. The other statement explains that copulas are continuous multivariate distributions with uniform marginals on the interval $(0, 1)$, which can also be found in Pitt et al. [2006] along with the following, where $C : [0, 1]^p \to [0, 1]$.

$$C(u_1, \ldots, u_p) = Pr(U_1 \leq u_1, \ldots, U_p \leq u_p). \tag{1.1}$$

We can formally define a copula, originally defined by Abe Sklar in 1959, from Bouyé et al. [2000].

Definition 1. *A p-dimensional copula is a function C with the following properties:*

1. *Domain of $C = [0, 1]^p$;*
2. *C is grounded and p-increasing;*
3. *C has univariate margins C_n which satisfy $C_n(u) = C(1, \ldots, 1, u, 1, \ldots, 1) = u$ for all $u \in [0, 1]$. Note: The u in the argument of C is in the n^{th} position.*

With this formal definition Sklar's theorem, [Nelsen, 2006], shows that every multivariate distribution function can be written in terms of a copula.

Theorem 1.2.1 (Sklar's theorem)**.** *Let F be a p-dimensional joint distribution function with margins $F_1, \ldots, F_p$. Then there exists a copula C such that for all $x_1, \ldots, x_p \in \mathbb{R}$,*

$$F(x_1, \ldots, x_p) = C(F_1(x_1), \ldots, F_p(x_p)).$$

The above theorem then forms the basis of how we will construct an appropriate spatio-temporal model, by joining our univariate marginals together by means of a copula. This is better explained through an example. Say we have two random variables, $X_1 \sim Exp(\lambda)$ and $X_2 \sim N(\mu, \sigma)$, with known parameters λ, μ and σ, and cumulative distribution functions F_1 and F_2, respectively. Via the Probability Integral Transform, $F_i(X_i) = U_i \sim U(0, 1)$. Therefore, the joint distribution of $\boldsymbol{X} = (X_1, X_2)^T$ is

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{1.2}$$

Since the marginals are continuous, the joint density function can be found by differentiating (1.2) to get

$$f(x_1, x_2) = f_1(x_1) f_2(x_2)\, c(F_1(x_1), F_2(x_2)), \tag{1.3}$$

where f_1 and f_2 are the density functions for an $Exp(\lambda)$ and $N(\mu, \sigma)$, respectively. This result can then be generalised to p-dimensions to obtain

$$f(x_1, \ldots, x_p) = f_1(x_1) \ldots f_p(x_p)\, c(F_1(x_1), \ldots, F_p(x_p)), \tag{1.4}$$

where $c(u_1, \ldots, u_p) = \frac{\partial^p}{\partial \boldsymbol{u}} C(u_1, \ldots, u_p)$ is the copula density. The next step is to choose our copula C.

Archimedean copulas

A popular class of copulas are the Archimedean copulas. As Nelsen [2006] describes, they are popular due to the large variety of copulas available, the ease of which they can be constructed and the fact that, even in high dimensions, the dependence model can be described by only one parameter. We can define an Archimedean copula as follows, but first we must define what is known as the generator $\varphi(\cdot)$, where $\varphi : [0, 1] \to [0, \infty]$ is a continuous, strictly decreasing convex function with $\varphi(1) = 0, \varphi(0) \leq \infty$ [Fischer et al., 2009]. We then define the pseudo-inverse of $\varphi, \varphi^{[-1]}$ by

$$\varphi^{[-1]}(t) \equiv \begin{cases} \varphi^{-1}(t) & \text{for } 0 \leq t \leq \varphi(0), \\ 0 & \text{for } \varphi(0) \leq t \leq \infty. \end{cases}$$

For the 2-dimensional case, we can define a copula by

$$C(u_1, u_2) = \varphi^{[-1]} \left(\varphi(u_1) + \varphi(u_2) \right). \tag{1.5}$$

This can be extended to the p-dimensional case as follows

$$C(u_1, \ldots, u_p) = \varphi^{[-1]} \left(\varphi(u_1) + \ldots + \varphi(u_p) \right). \tag{1.6}$$

Family	Generator $\varphi(t\|\theta)$	Parameter
Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$\theta \in (0,\infty)$
Frank	$-\log\left(e^{-\theta t}-1/e^{-\theta}-1\right)$	$\theta \in (0,\infty)$
Gumbel	$-\log(t)^{\theta}$	$\theta \in (1,\infty)$
Joe	$-\log\left[1-(1-t)^{\theta}\right]$	$\theta \in (1,\infty)$

Table 1.1: Popular generators for Archimedean copulas

As Fischer et al. [2009] also states, this is a p-dimensional Archimedean copula if and only if φ^{-1} is completely monotonic on $\mathbb{R}^+$. Table 1.1 [adapted from Nelsen [2006]] gives a few popular generators $\varphi(\cdot)$.

Figure 1.3 shows the theoretical contours of the Clayton copula for $\theta = 1$ as well as 50K simulated values for $\theta = 1$ (black) and $\theta = 2$ (red). Evidently we see a weak but positive correlation in the data, and then increasing this one copula parameter will then increase the correlation in the data. Note that here the relation between the copula parameter and the correlation in the data seems quite straightforward, although for most cases outside of the family of elliptical distributions, the relation isn't as simple [Schmidt, 2006].

In Figure 1.4, we have demonstrated how we can use copulas to join univariate marginals. In this case we have two standard Normal marginals fitted with a Gumbel copula with parameter $\theta = 2$. We are accustomed to seeing Normal distributions having regular ellipses as their joint density, whereas here, using a copula, the joint density does not

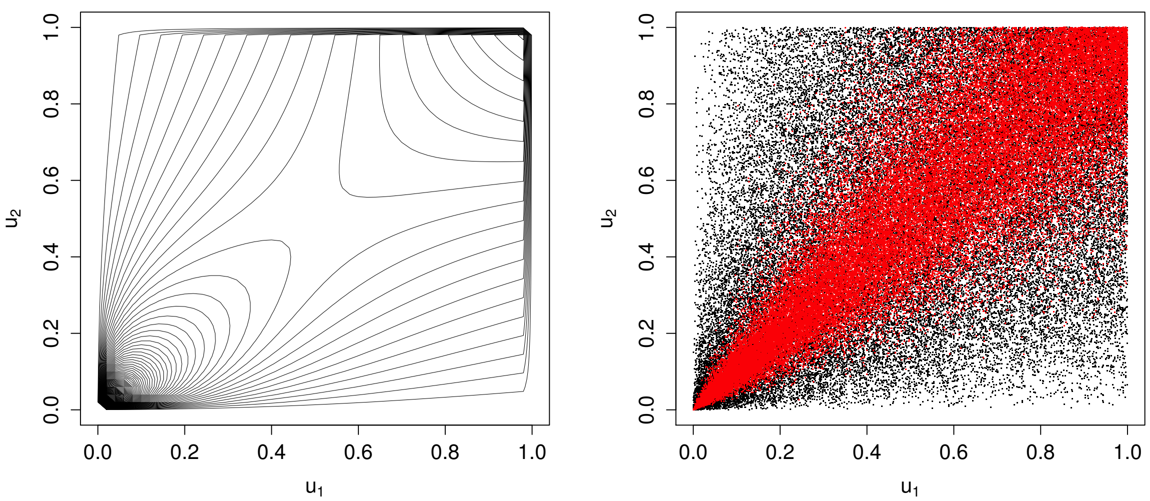

Figure 1.3: Theoretical and simulated values from the Clayton copula.

resemble the bivariate Normal density at all due to the skewed ellipses produced. Some may find these single parameter copulas to be sufficient enough, for instance, Hennessy and Lapan [2002] discuss how the structural form of these copulas can be used in the area of risk and investments, specifically, using them to model portfolio allocations.

For our analysis we need a more flexible model, which is why we will look at elliptical copulas for the remainder of this report, starting with the Gaussian copula. See Nelsen [2006] for further properties of Archimedean copulas.

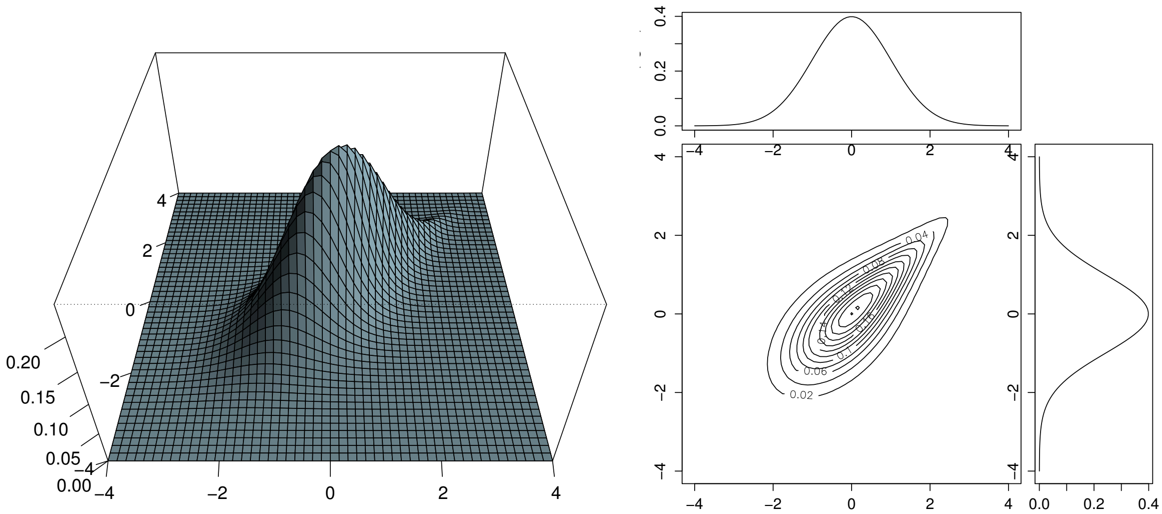

Figure 1.4: Standard Normal marginals fitted with a Gumbel copula with $\theta = 2$

2 | The Gaussian copula

2.1 Definition

The Gaussian copula is derived from the multivariate Normal distribution and is one of the more popular copulas used in the financial sector, and areas of risk management and insurance with more listed in Balakrishnan and Lai [2009]. Some experts believe that the Gaussian copula had a key role in the financial meltdown of 2008-09, when it was formerly being used to assess risk by estimating the dependence between a number of individual debt securities [Lee, 2009]. In comparison to the Archimedean copulas we looked at in §1.2, the Gaussian copula has more than one parameter for dimensions $p \geq 2$; in fact there are $p(p-1)/2$. As seen in Balakrishnan and Lai [2009], the bivariate Gaussian copula is defined explicitly below in (2.2), with Φ_2 denoting the cumulative distribution function (cdf) for the standard bivariate Normal distribution, and Φ representing the cdf of the standard univariate Normal distribution.

Definition 2 (Bivariate Gaussian copula function)**.** *The Gaussian copula function in 2-dimensions is defined as*

$$\begin{aligned} C(u_1, u_2) &= \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)|\rho) \qquad (2.1) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left\{\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right\} \mathrm{d}s\, \mathrm{d}t, \end{aligned}$$

where ρ is the dependence parameter in the copula.

We can generalize this to p dimensions to obtain the following definition.

Definition 3 (Gaussian copula function)**.** *The p-dimensional Gaussian copula function is defined as*

$$C(u_1, \dots, u_p) = \Phi_p(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)|P), \qquad (2.2)$$

where Φ_p denotes the cdf of the p-dimensional Normal distribution.

Here P is the $p \times p$ correlation matrix, thus being a symmetric positive semi-definite matrix with 1s along the diagonal. This means the number of parameters of the Gaussian copula is equal to the number of entries in the upper triangular giving us $p(p-1)/2$.

Using Song [2000], we can define the Gaussian copula density as follows, with the full derivation found in Appendix A.1.1.

Definition 4 (Gaussian copula density)**.** *The Gaussian copula has density*

$$\begin{aligned} c(u_1, \dots, u_p) =& \frac{\phi_p(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)|P)}{\prod_{i=1}^{p} \phi(u_i)} \\ =& (2\pi)^{-p/2}|P|^{-1/2} \\ & \times \exp\left\{-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_p)\end{pmatrix}^T (P^{-1} - I_p)\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_p)\end{pmatrix}\right\}. \qquad (2.3) \end{aligned}$$

Figure 2.1 shows the bivariate Gaussian copula density for values of $\rho = 0.5$ and $\rho = -0.7$, respectively, to demonstrate how altering the value of $\rho \in (-1, 1)$ affects the copula. For large positive values of ρ the contours show a positive correlation, and conversely, for negative values of ρ, the contours show negative correlation between the two uniform variables u_1, u_2.

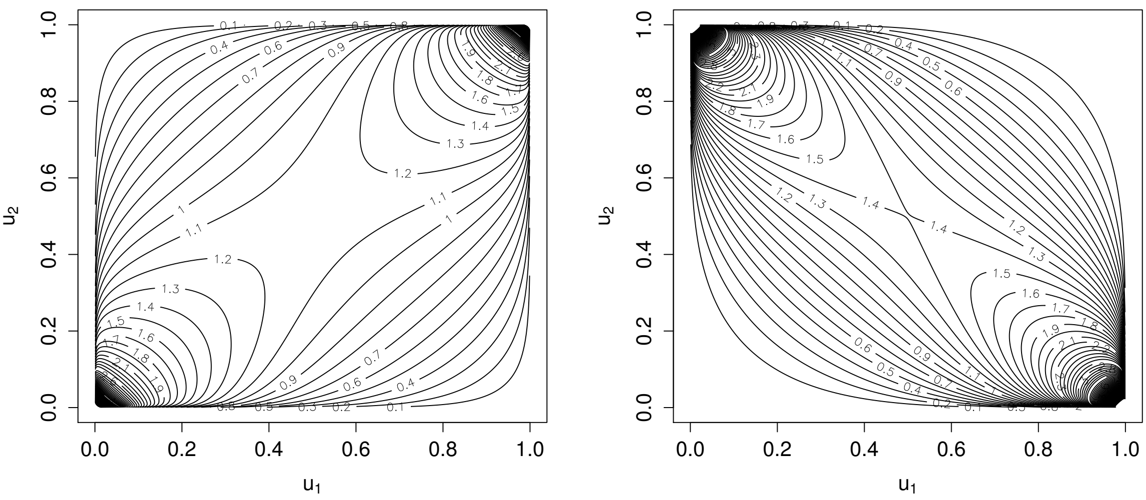

Figure 2.1: Contours of the bivariate Gaussian copula with $\rho = 0.5$ (left) and $\rho = -0.7$ (right).

2.2 Simulation of realisations

By simulating values from the bivariate Normal distribution, $N_2(\mathbf{0}, P)$, where P is the correlation matrix, it is easy to demonstrate how the copula can be used to join univariate distributions. Using the Probability Integral Transform on different marginal distributions we are able to have jointly distributed data using the copula, as previously seen in §1.2 with the Gumbel copula with Normal marginals. Adequately simulating from these copula based models is important since we will be performing inference on simulated data first, before applying our methods to actual data in order to check that our inferences are just.

Exponential and Normal marginals

Returning to our example in §1.2 where we had two random variables, $X_1 \sim Exp(\lambda)$ and $X_2 \sim N(\mu, \sigma)$, with known parameters λ, μ and σ, we will couple these two distributions together using the Gaussian copula. Let's say for X_2 that $\lambda = 0.5$ and that X_1 is standard Normal. Figure 2.2 shows both a 3D density plot of the joint density function and 50K simulated values from this joint distribution with their marginal histograms. We've managed to couple two univariate distributions with the Gaussian copula. In the

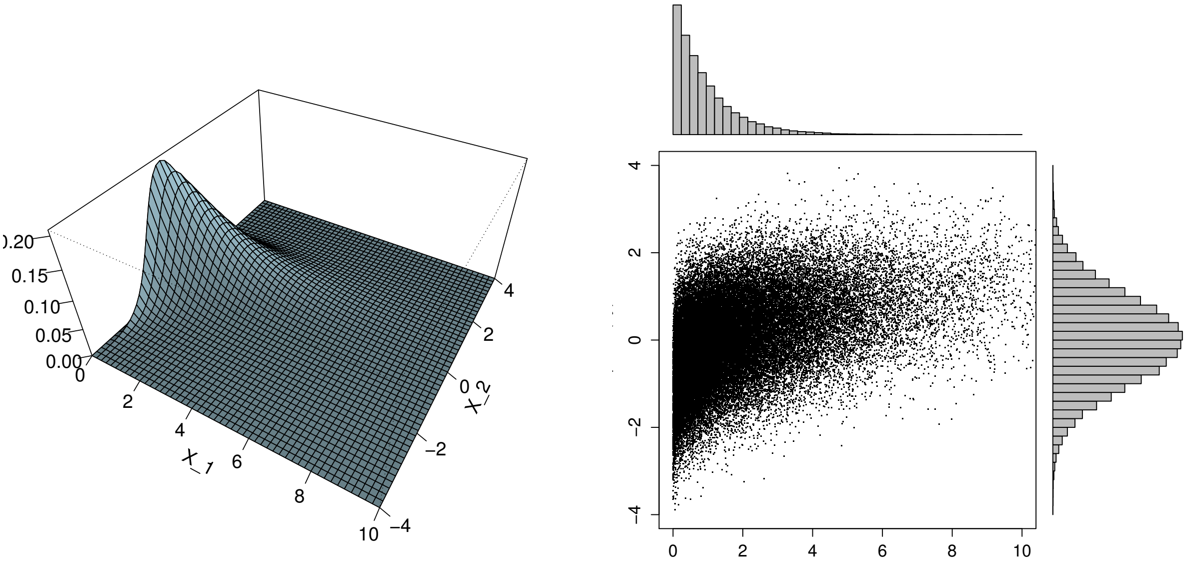

Figure 2.2: Joint density (left) and 50K simulated values (right) from the joint distribution of $X_1 \sim Exp(0.5)$ and $X_2 \sim N(0,1)$ using the Gaussian copula for $\rho = 0.5$, with their marginal histograms.

context of rainfall data, we will be taking the observations at each site and coupling them together using an appropriate copula model. Thinking about the rainfall example, rainfall volumes can never be negative and so a Normal distribution wouldn't be suitable to describe the distributions at each site and so this example should be discarded when considering an appropriate model for rainfall. An interesting note however, although self-evident to see, is that if we had both $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$ and used a Gaussian copula with parameter ρ, then the resulting joint distribution is simply the standard bivariate Normal, $\boldsymbol{X} = (X_1, X_2)^T \sim N_2(\mathbf{0}, P)$. This can also be generalised to p-dimensions.

Gamma marginals

Now we move on to using the Gamma distribution as our univariate marginals, since this would be a more appropriate choice to say model rainfall data over the Normal distribution, given that the support is $(0, \infty)$. Figure 2.3 shows, on the left, the 3D density plot of the joint distribution of the two Gamma distributions with a large correlation parameter of $\rho = 0.8$ in the copula. On the right we see simulated values from this joint distribution with an evidently large empirical linear correlation. Since our data involves sites which are near to one and other in the context of location we would expect large positive correlations in the observed data, understandable since rain at site A will, most likely, mean rain at site B, if they are close.

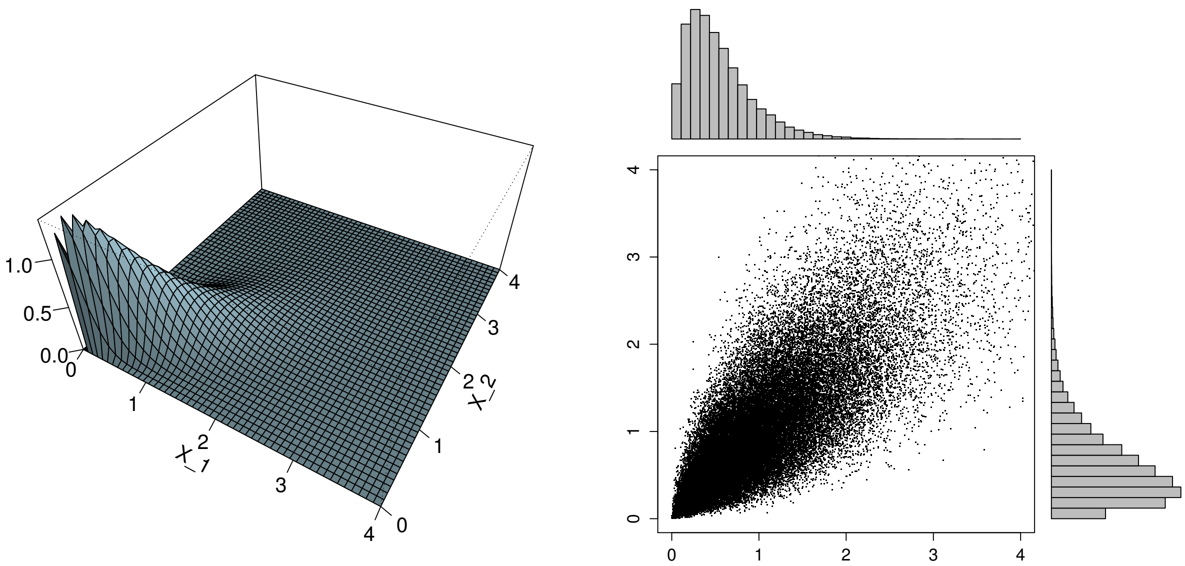

Figure 2.3: Joint density (left) and 50K simulated values (right) from the joint distribution of $X_1 \sim Ga(2,2)$ and $X_2 \sim Ga(2,2)$ using the Gaussian copula for $\rho = 0.8$, with their marginal histograms.

2.3 Bayesian inference

Now we begin with Bayesian analysis of such models starting simple, on simulated data from the Gaussian model. We first need to deal with the setup of this problem by correctly defining the model, likelihood, prior distributions and suitable paramaterizations of the copula parameters, the $p(p-1)/2$ elements of P. Since there will be no conjugate analysis here due to the complex nature of the model, we will be implementing appropriate Markov chain Monte Carlo algorithms, commonly abbreviated to MCMC, in order to simulate observations from the posterior distributions of the parameters.

Likelihood

To perform posterior analysis we compute the likelihood of the data, with the full derivation found in Appendix A.1.2. Since the rainfall data between time points will be correlated, we will assume that there is time dependence, i.e. between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$, and so wish to model the distribution of $\boldsymbol{X}_t$ given $\boldsymbol{X}_{t-1}$, that is $f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$. Suppose that the marginal distribution of $X_{t,i}$ depends on the set of parameters $\boldsymbol{\alpha}_i$ and denote the cdf and pdf of $X_{t,i}$ as F_i and f_i, respectively, for $i = 1, \dots, p$. Let $X = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$, $A = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p)$ and $\boldsymbol{\beta}$ be the set of parameters in the copula, $c(\cdot)$, in this case the Gaussian copula. Taking the product of the this over the number of sites, p, and over all time points in the data, T, we get the likelihood, explicitly, (2.4).

$$f(X|A, \boldsymbol{\beta}) = \frac{\prod_{t=2}^{T} f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|A, \boldsymbol{\beta})}{\prod_{t=2}^{T-1} f(\boldsymbol{x}_t|A, \boldsymbol{\beta})}, \tag{2.4}$$

where

$$f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|A, \boldsymbol{\beta}) \left\{ \prod_{i=1}^{p} f_i(x_{t,i}|\boldsymbol{\alpha}_i) f_i(x_{t-1,i}|\boldsymbol{\alpha}_i) \right\} \times c(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|A, \boldsymbol{\beta}).$$

Due to the possibility of performing this on many sites over a long time interval, the log-likelihood function would be more favourable in this instance as we can often run in to numerical issues when taking products of products such as integer overflow in packages such as `R` and so we will use the log-likelihood, simply by taking the log of (2.4).

Posterior distribution

As we know, Bayesian analysis allows us to combine both our beliefs and what the data actually tells us, so our posterior distribution is found since it is proportional to our prior beliefs multiplied by the likelihood. We will assume that our prior beliefs for the parameters of the marginal distributions and the copula parameters are independent. The posterior density is

$$\pi(A, \boldsymbol{\beta}|\mathcal{D}) \propto \pi(A)\pi(\boldsymbol{\beta}) f(X|A, \boldsymbol{\beta}).$$

As stated previously, it is understandable that no conjugate analysis is possible, so MCMC, specifically Metropolis-Hastings algorithms will be used to draw samples from the posterior distribution. The usual approach with copula based models is to use a two stage process to estimate the parameters, see Danaher and Smith [2011] and Joe [2005]. This method involves estimating the parameters of the marginal distributions, and then estimating the copula parameters conditional upon these.

Marginal parameters

Estimating the marginal distributions in the first stage can be done through a variety of ways including maximum likelihood estimation, method of moments estimation and, the technique we will be performing, MCMC algorithms, full details of which can be found in Appendix A.1.5. We essentially perform analysis assuming spatial and temporal independence to obtain realisations from the independent posterior densities, that is,

$$\tilde{\pi}(\boldsymbol{\alpha}_i|\mathcal{D}) \propto \pi(\boldsymbol{\alpha}_i) \prod_{t=1}^{T} f(x_{t,i}|\boldsymbol{\alpha}_i).$$

We will be using Gamma distributions for our marginals, so $X_i \sim Ga(\alpha_i, \lambda_i)$, thus $\boldsymbol{\alpha}_i = (\alpha_i, \lambda_i)$, for each site $i = 1, \ldots, p$. Since we require both α_i and λ_i to be positive, we impose independent Lognormal prior distributions on both parameters. The MCMC algorithm is specifically a Metropolis Hastings algorithm, so we require a suitable proposal distribution for proposed values of the parameters which are then accepted or rejected. Due to both parameters being positive, we use Lognormal distributions again when proposing values α_i^*, λ_i^*, since using a Normal random walk here would generate negative proposals as well and having to reject these every so often will be quite wasteful.

Copula parameters

Since the parameters of the copula are elements of a correlation matrix P, certain constraints exist, since correlation matrices must be positive semi-definite. This will be the difficult part of the MCMC scheme as proposing $\rho_{ij} \in (-1, 1)$ for the symmetric off-diagonal elements would yield low acceptance probabilities. As Barnard et al. [2000] identify, producing a valid correlation matrix is a challenging statistical issue and so this is tackled by parametrising the correlation matrix using a similar approach to Danaher and Smith [2011], but adapted to consider the lag 1 auto-correlation in the data. Essentially we use the following representation

$$P = \text{diag}(\Sigma)^{-1/2}\,\Sigma\,\text{diag}(\Sigma)^{-1/2},$$

where

$$P = \begin{pmatrix} P_1 & P_2 \\ P_2 & P_1 \end{pmatrix} \tag{2.5}$$

is a $2(p \times p)$ matrix where P_1 represents the correlations between $X_1, \ldots, X_p$ at time t and P_2 is the lag 1 autocorrelation matrix, that is, describing the dependence between $X_1, \ldots, X_p$ at times t and $t-1$. Here $\Sigma = \text{Cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-1})$ is constructed from two further matrices R_m and R_v, which are both upper triangular matrices containing our new parameters $\boldsymbol{r}$. See Appendix A.1.4 for full technical details. Below we have the two matrices for the $p = 2$ case:

$$R_m = \begin{pmatrix} r_{m1} & r_{m2} \\ 0 & r_{m3} \end{pmatrix} \qquad \text{and} \qquad R_v = \begin{pmatrix} 1 & r_{v1} \\ 0 & 1 \end{pmatrix}.$$

This is easily generalised for the p-dimensional case. This parameterization of the correlation matrix means that the constraints of the different r's are the thing to be concerned about. The only constraints here are that we require the diagonal elements of R_m to be positive real values and all the upper-triangular elements of both R_m and R_v must be real. This setup yields the correlation matrix in (2.5). This allows us to generate valid positive semi-definite correlation matrices, as long as we adhere to the constraints of the r's. However we now have p^2 number of parameters, as opposed to the $p(p-1)/2$ number of parameters, needed to parameterize a correlation/covariance matrix.

Prior distributions The prior distributions we impose on these r's will be Uniform, with different end-points dependent on the constraints. Let us denote r_i^+ as the diagonal r's that we require to be positive. Since we want to input fairly uninformative prior information, we impose a $U(0, 5)$ prior on these. The way we propose values for these parameters will be through Lognormal proposal distributions. Let the unconstrained r's be denoted by $r_i^{\pm}$. We impose a $U(-5, 5)$ prior on these, again to be fairly uninformative, and use a Normal random walk proposal. We will be updating each parameter separately, and not as a block update, thus allowing us to tune each accordingly.

Inference on simulated data

Performing the MCMC scheme on data we have simulated from the model allows us to confirm whether or not the algorithm is correct. Simulating from the model is slightly more involved this time around due to the lag 1 autocorrelations which we need to account for and the complete algorithm is found in Appendix A.1.3. We will keep things basic by having $p = 2$, although the code used to run this is already generalised to perform for $p \geq 2$.

Inference on marginal parameters

The marginals we have here will be $X_1 \sim Ga(3, 1)$ and $X_2 \sim Ga(4, 1)$, as these have larger densities for smaller values, since we expect to have more weight down the lower end of the scale of rainfall volumes. Using a sample size of $n = 1,000$ and a copula correlation matrix with components

$$P_1 = \begin{pmatrix} 1.000 & 0.768 \\ 0.768 & 1.000 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0.065 & 0.059 \\ 0.059 & 0.124 \end{pmatrix}, \tag{2.6}$$

stage 1 output was obtained and are shown in Figures 2.4 to 2.5 and Table 2.1. This run was carried out after an initial, shorter, run in order to tune each of the parameters, which we discuss when looking at the autocorrelation. Evidently, the algorithm is clearly producing realisations from the independent posterior densities for the marginal parameters as each one of the (marginal) kernel densities has the respective value, used to simulate the data, near the mean of the posterior distribution.

Parameter	Posterior mean	Posterior SD	Effective sample size
α_1	3.032	0.128	5145.3
α_2	4.007	0.170	3290.7
λ_1	1.024	0.047	5349.5
λ_2	1.016	0.046	3470.6

Table 2.1: Posterior summaries from the stage 1 analysis.

Convergence and Autocorrelation To assess the convergence of the algorithm, we can run three simultaneous schemes starting from different initial values as seen in Figures 2.4 and 2.5. Since the stationary distribution of our algorithm is the posterior distribution, the chain will eventually converge to the posterior no matter where we initialise the chain (provided we initialise in the correct parameter space). Here, it's evident that we have convergence since, even after removing as little as 100 iterations, the chain has already 'burnt-in'. A way to see how much autocorrelation we have in the chain is to look at both autocorrelation plots and the *effective sample sizes* of the parameters. The effective sample size is an estimate of how many uncorrelated realisations the chain contains. An issue here is that there is a fair amount of autocorrelation within the chain, as we get very small effective sample sizes from runs of 10K, only around 500 for each

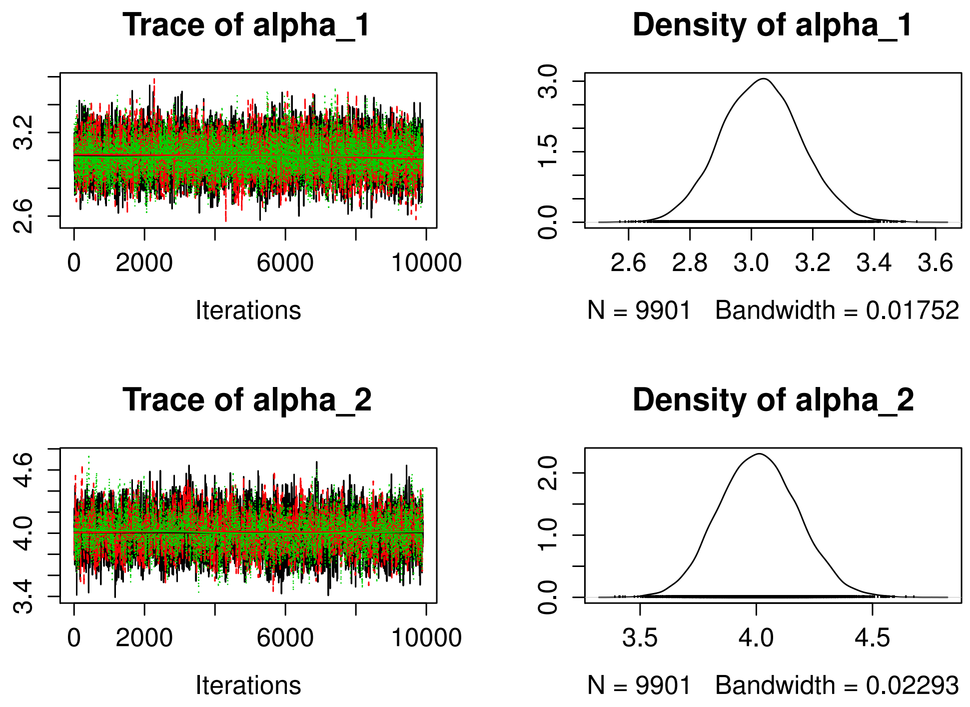

Figure 2.4: Traceplots and densities obtained from 10K posterior realisations of α_1 and α_2.

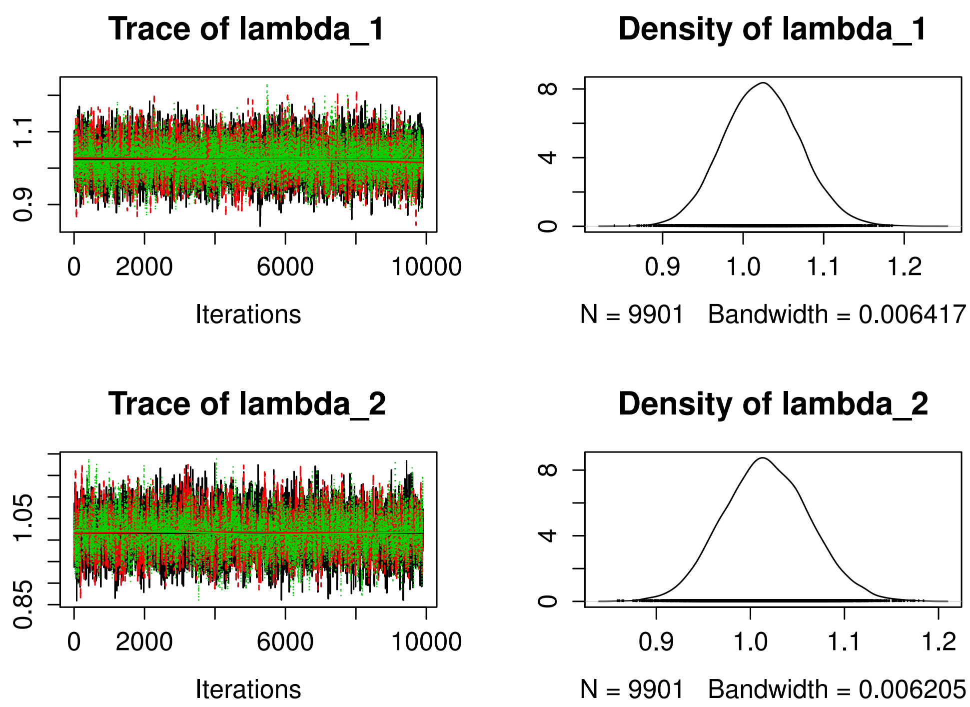

Figure 2.5: Traceplots and densities obtained from 10K posterior realisations of λ_1 and λ_2.

of the parameters, despite having a thinning interval of 10. And so instead, we perform the algorithm again, with appropriate tuning, and a thinning interval of 100, essentially taking every 100^{th} realisation from a run of 100K, to get almost 10K uncorrelated realisations. The effective sample sizes are given in Table 2.1, and we see the autocorrelation plots in Figure 2.6.

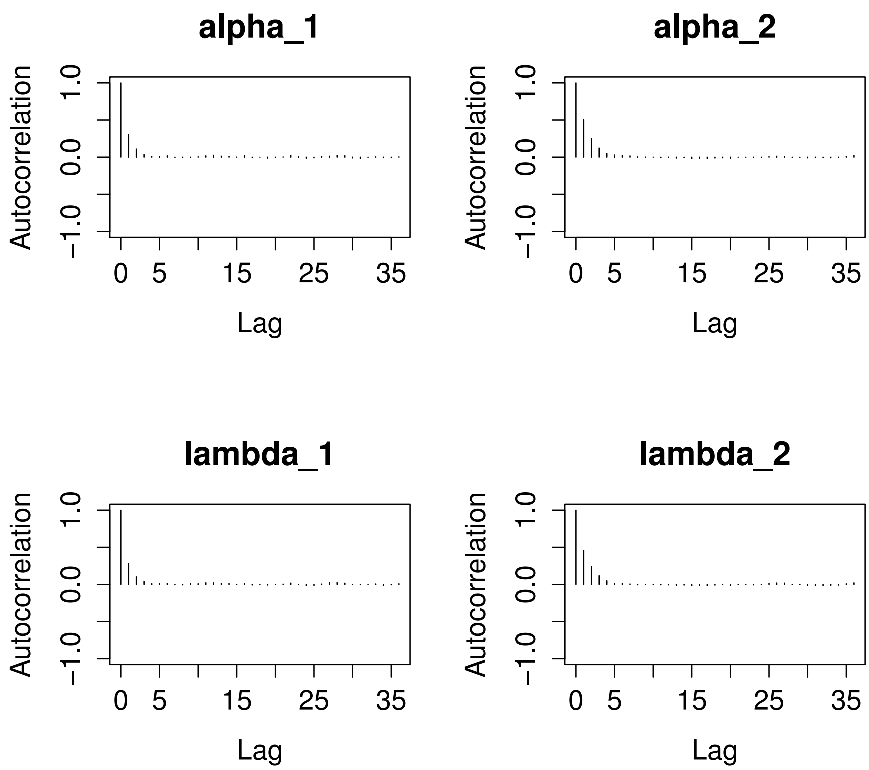

Figure 2.6: Autocorrelation plots for α and λ.

Inference on copula parameters

Using the stage 1 realisations, we use these in the next stage in order to perform inference on the copula parameters, therefore, conditional on the marginal parameters.

$$\pi(\boldsymbol{\beta}|A, \mathcal{D}) \propto \pi(\boldsymbol{\beta}) f(X|A, \boldsymbol{\beta}).$$

The output from this gives realisations from the posterior of the r's, of which we can construct the copula correlation matrix P, so we then have the posterior for the elements of P. Figure 2.7 shows both the trace plots and kernel densities of the (marginal) posterior distributions of the elements of P. The scheme is clearly sampling from the posterior distribution here, since, when comparing the densities of each element with the corresponding element of P used to simulate the data, 2.6, we see that these 'true' values are all captured within the (marginal) posteriors.

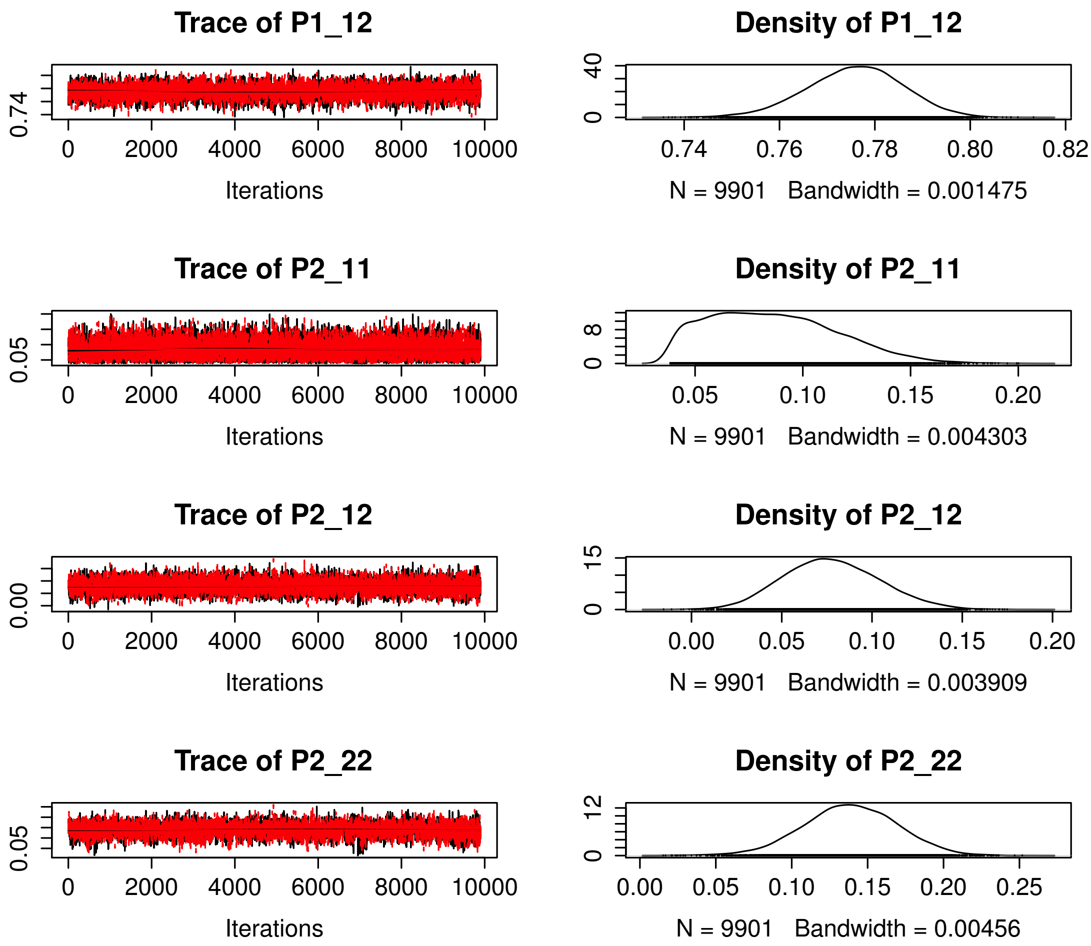

Figure 2.7: Traceplots and densities obtained from 10K posterior realisations of P.

Convergence and Autocorrelation In order to assess that the chain converges for the copula parameters, we can run two chains in parallel with different initial values to check that realisations are in fact from the posterior distribution. It's evident that the chains are converging since both the black and red chains are covering roughly the same parameter space in the trace plots. A thinning interval of 10 was used in this run, meaning we effectively ran the algorithm for 100K iterations and took every 10^{th} realisation, giving us a sample of 10K. We use this with the hopes of removing any autocorrelation from our posterior realisations. As we observe in Figure 2.8, the autocorrelations are quite small past lag 5, and we have effective sample sizes given in Table 2.2.

Since we have quite a large amount of data we have simulated from the model, and used weak priors, it's safe to assume that the MCMC algorithms are in fact producing realisations from the posterior distributions. Although the Gaussian copula works well at describing the dependence for certain applications, we come across issues when describing dependence between extreme events, something which a heavy-tailed distribution would be more suited for. The aim of our analysis is to find a suitable, yet flexible, model for rainfall and so we now move on to look at the skew t copula, which not only allows for heavier tails than the Gaussian, but also allows skewness in each dimension.

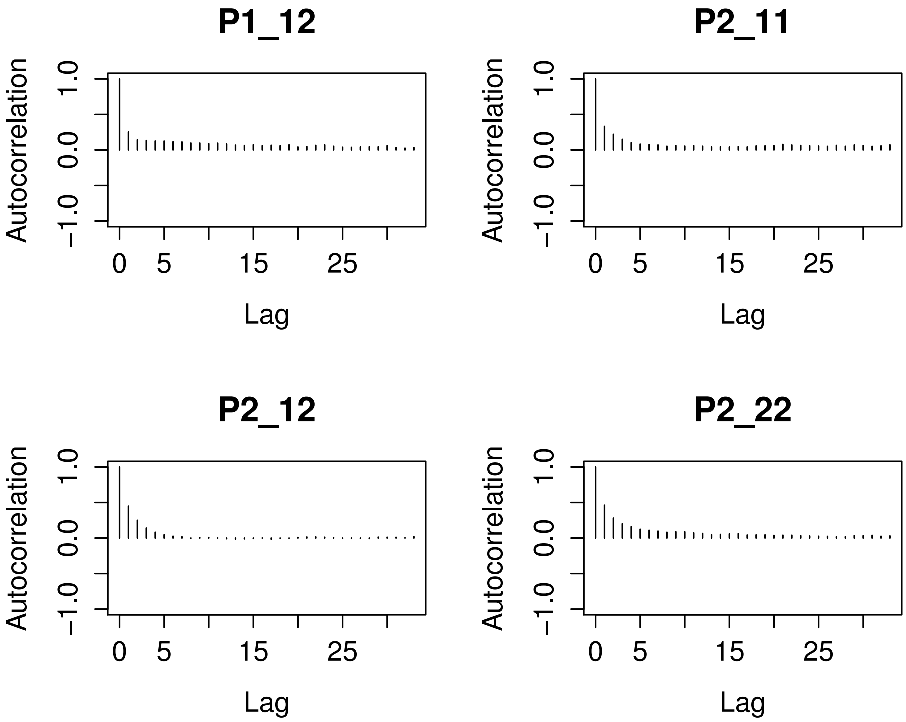

Figure 2.8: Autocorrelation plots for the copula parameters.

Parameter	Posterior mean	Posterior SD	Effective sample size
$P_{\mathbf{1}1,1}$	0.775	0.010	1433.5
$P_{\mathbf{2}1,1}$	0.085	0.029	1043.0
$P_{\mathbf{2}1,2}$	0.077	0.026	3368.9
$P_{\mathbf{2}2,2}$	0.139	0.030	1688.6

Table 2.2: Posterior summaries for the copula parameters.

3 | The skew t copula

3.1 Definition

Although the Gaussian copula has it's applications in certain fields, returning back to our example of rainfall, a dependence structure with increased flexibility would be more suitable. We need one which would not only capture dependence between extreme events more adequately, but also allow for asymmetry, and so we investigate the skew t copula.

The skew t copula is again derived from a multivariate elliptical distribution, namely the skew t distribution. This isn't as widely used as the well-known multivariate Normal distribution that the Gaussian copula is derived from, so we must first define this distribution. Due to the degrees of freedom, ν, the skew t copula has uses in areas such as finance and economics, just like the Gaussian, with the the skew t, and symmetric t copula, being favoured due to the ability to account for extreme events. See Smith et al. [2012] for applications with regional spot prices in the Australian electricity market and dependence of website popularity.

Due to the ability of having different levels of skewness in each dimension, we will be using the distribution as defined by Sahu et al. [2003], as opposed to the one defined by Azzalini and Capitanio [2003].

Definition 5 (Skew t distribution)**.** *Suppose that $\boldsymbol{X}$ and $\boldsymbol{Q}$ are both $p \times 1$ vectors and have a joint $2p$-dimensional t distribution with zero mean/mode, degrees of freedom ν and scale matrix Ω, that is*

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Q} \end{pmatrix} \sim t_\nu \left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \Omega = \begin{pmatrix} P + \Delta^2 & \Delta \\ \Delta & I_p \end{pmatrix} \right),$$

where $\Delta = diag(\delta_1, \dots, \delta_p)$ and $P = (\rho_{ij})$ is a $p \times p$ (positive definite) correlation matrix. Then the skew t distribution is defined to be that of $\boldsymbol{X}|\boldsymbol{Q} > \boldsymbol{0}$, with density

$$f_{St}(\boldsymbol{x}|P, \boldsymbol{\delta}, \nu) = 2^p \left|P + \Delta^2\right|^{-1/2} \psi_\nu \left(\left(P + \Delta^2\right)^{-1/2} \boldsymbol{x} \right) \Pr(\boldsymbol{Q} > \boldsymbol{0}|\boldsymbol{X} = \boldsymbol{x}),$$

where $\psi_\nu(\cdot)$ is the p-dimensional standardised t density and

$$\boldsymbol{Q}|\boldsymbol{X} = \boldsymbol{x} \sim t_{\nu+p} \left(\Delta \left(P + \Delta^2\right)^{-1} \boldsymbol{x}, \frac{\nu + \boldsymbol{x}^T \left(P + \Delta^2\right)^{-1} \boldsymbol{x}}{\nu + p} \left\{ I_p - \Delta \left(P + \Delta^2\right)^{-1} \Delta \right\} \right).$$

We denote $\boldsymbol{X} \sim St_\nu(P, \boldsymbol{\delta})$ if $\boldsymbol{X}$ has density $f_{St}(\boldsymbol{x}|P, \boldsymbol{\delta}, \nu)$ as defined above. As given in Sahu et al. [2003], the first two moments are given below, with

$$\begin{aligned} \mathrm{E}(\boldsymbol{X}) =& \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma\left([\nu - 1]/2\right)}{\Gamma(\nu/2)} \boldsymbol{\delta}, \\ \mathrm{Var}(\boldsymbol{X}) =& (\Gamma + \Delta^2) \frac{\nu}{\nu - 2} \\ & - \left[\left\{ \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma\left([\nu - 1]/2\right)}{\Gamma(\nu/2)} \right\}^2 \boldsymbol{1}\boldsymbol{1}^T - (\boldsymbol{1}\boldsymbol{1}^T - I) \frac{2\nu}{\pi(\nu - 2)} \right] \circ \boldsymbol{\delta}\boldsymbol{\delta}^T, \end{aligned}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^T$, $\mathbf{1}$ is a p-vector of 1's and the operator $\circ$ represents the Hadamard matrix product. The skewness in each dimension is controlled by the skewness parameters $\delta_i \in \mathbb{R}$ for $i = 1, \ldots, p$, with symmetry occurring when $\delta_i = 0$. The degrees of freedom control how heavy the tails are in the distribution and when $\nu \to \infty$ the skew t converges to a skew Gaussian, also known as a skew Normal. When we have $\boldsymbol{\delta} = \mathbf{0}$, the skew t distribution simply becomes the multivariate t distribution, and so it follows that with $\boldsymbol{\delta} = \mathbf{0}$ and $\nu \to \infty$, the skew t converges to the Normal distribution. With the skew t distribution defined, we construct the skew t copula as follows.

Definition 6 (Skew t copula)**.** *The skew t copula is $C_{st}(\boldsymbol{u}|P, \boldsymbol{\delta}, \nu)$, with density*

$$c_{St}(\boldsymbol{u}|P, \boldsymbol{\delta}, \nu) = \frac{f_{St}(\boldsymbol{x}|P, \boldsymbol{\delta}, \nu)}{\prod_{i=1}^{p} f_{St,i}(x_i|\delta_i, \nu)},$$

where $x_i = F_{St,i}^{-1}(u_i|\boldsymbol{\delta}, \nu)$.

Some properties from the multivariate skew t distribution carry to the skew t copula, due to this construction. When $\nu \to \infty$ the skew t copula converges to the skew Gaussian copula, and when $\boldsymbol{\delta} = \mathbf{0}$ also the copula converges to the Gaussian copula defined in the previous chapter.

Data augmentation

If we were to compute the likelihood of the model currently, we would fall into issues due to the difficult calculations required. For example, the need to evaluate the term $\Pr(\boldsymbol{Q} > \mathbf{0}|\boldsymbol{X} = \boldsymbol{x})$ is a rather daunting task. This is a p-dimensional probability for a t distribution, a challenging computation to tackle. We could look in to methods discussed by Genz and Bretz [2002] to calculate this, but we instead look at augmenting the problem. Introducing auxiliary variables will allow us to make the likelihood much simpler. These auxiliary variables aren't observed, but if they were the likelihood would be simplified enough for us to explore the posterior, so we treat these as missing data, where they are dealt with in our MCMC algorithms.

Conditional Gaussian representation

Due to this augmentation, we represent the skew t as a scale mixture of Normals [Smith et al., 2012]. We get $\boldsymbol{X} \sim St_\nu(P, \boldsymbol{\delta})$, marginally, using auxiliary variables $w, \boldsymbol{q}$, where

$$\boldsymbol{X}|\boldsymbol{q}, w \sim N_p\left(\Delta\boldsymbol{q}, \frac{1}{w}P\right) \qquad \text{and} \qquad \boldsymbol{Q}|w \sim N_p\left(\mathbf{0}, \frac{1}{w}I_p\right),$$

and $W \sim Ga(\nu/2, \nu/2)$; see Appendix A.2.1.

We can then use this representation to now simplify the skew t copula density as

$$c_{St}(\boldsymbol{u}|P, \boldsymbol{\delta}, \nu, \boldsymbol{q}, w) = \frac{\phi_p(\boldsymbol{x}|\Delta\boldsymbol{q}, w^{-1}P)}{\prod_{i=1}^{p} f_{St,i}(x_i|\delta_i, \nu)},$$

where $x_i = F_{St,i}^{-1}(u_i|\boldsymbol{\delta}, \nu)$.

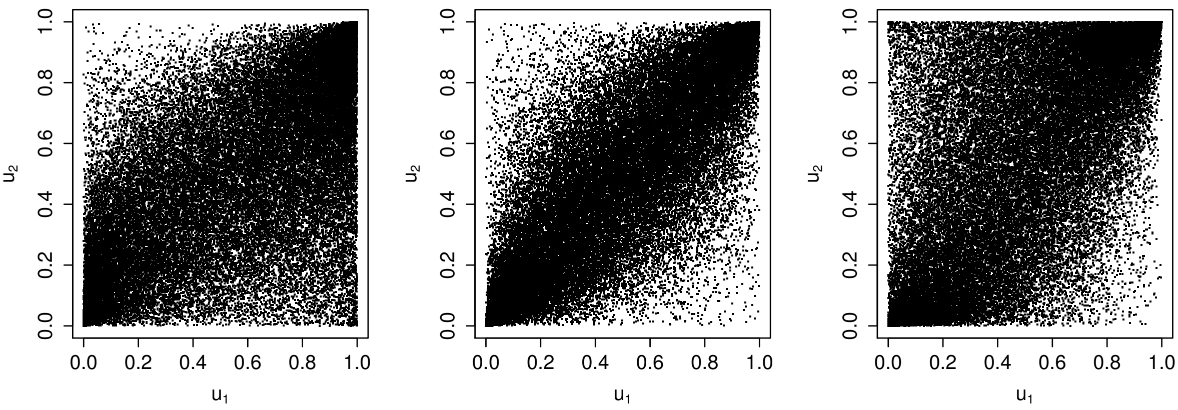

Figure 3.1: Varying skewness parameters, from left to right, $\boldsymbol{\delta} = (1, 0), \boldsymbol{\delta} = (0, 0), \boldsymbol{\delta} = (0, 1)$.

3.2 Simulation of realisations

We can again simulate from this copula model using the algorithm given in Appendix A.2.2. Since we've already discussed how copulas are used to describe the joint distribution of marginal distributions, we'll take a brief look at how uniform variables within the copula are dependent on each other with varying copula parameters. In Figure 3.1, we see three scatterplots of realisations from the joint distribution of $\boldsymbol{U} = (U_1, U_2)$ which are described with a bivariate skew t copula with $\nu = 5$, $\rho = 0.8$ in the correlation matrix P, and varying skewness parameters. These three scatterplots demonstrate the usefulness of having the flexibility of skewness parameters for each dimension.

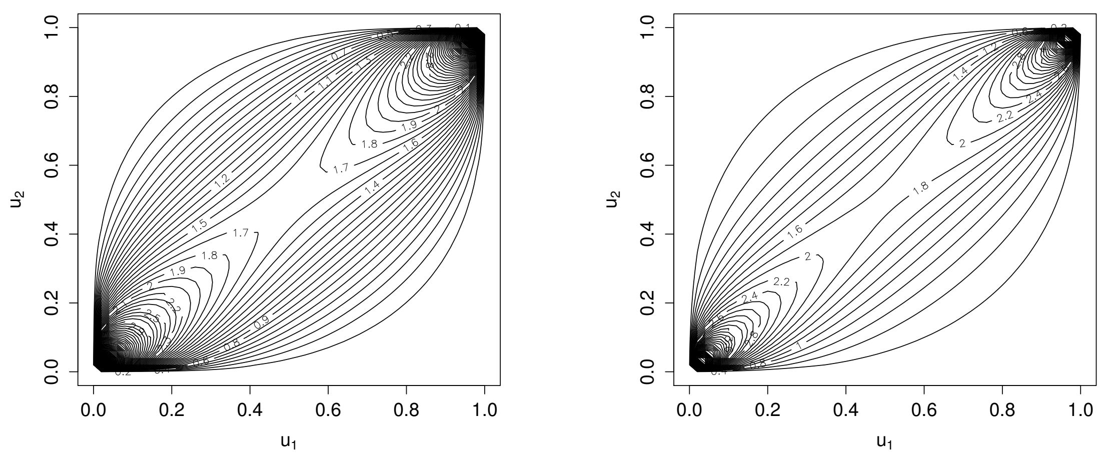

Figure 3.2: Contours of the bivariate Gaussian (left) and t (right) copulas with $\rho = 0.8$, and $\nu = 5$ for the t copula.

An interesting note here is that the centre plot is equivalent to a symmetric t copula, since we have chosen there to be no asymmetry. Next in Figure 3.2, we can compare the contours of the Gaussian and symmetric t copula in order to demonstrate the difference due to the degrees of freedom. We note the closeness of the contours for the Gaussian, due to the steepness in the density. Comparing these to the more dispersed contours of the t copula, it is understandable how the heavy tails of the t distribution affect the copula, allowing us to capture dependence between extreme events effectively.

3.3 Bayesian inference

We now look in to MCMC methods in order to estimate the posterior distributions of the copula parameters. The number of parameters in this case is $p^2 + p + 1$, due to the r's of the P matrix, the p-vector of δ parameters and the degrees of freedom ν.

Likelihood

We are simply changing the copula being used in this problem, and so we still model $f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, just changing the copula function $c(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|A, \boldsymbol{\beta})$. Full details are given in Appendix A.2.3. Let $z_{t,i} = F^{-1}_{St,i}(u_{t,i}|\delta_i, \nu)$, then the skew t copula of $(\boldsymbol{u}_t, \boldsymbol{u}_{t-1})$ is

$$c_{St}(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu) = \frac{f_{St}(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu)}{\prod_{i=1}^{p} f_{St,i}(z_{t,i}|\delta_i, \nu)\, f_{St,i}(z_{t-1,i}|\delta_i, \nu)}.$$

Now we write this in terms of the augmented variables $\boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1}$,

$$c_{St}(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1}) = \frac{f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1})}{\prod_{i=1}^{p} f_{St,i}(z_{t,i}|\delta_i, \nu)\, f_{St,i}(z_{t-1,i}|\delta_i, \nu)},$$

where the distribution of $(\boldsymbol{z}_t, \boldsymbol{z}_{t-1})$ is a multivariate normal. Through some cancellation and matrix manipulation, we get the following representation of our likelihood:

$$\frac{\prod_{t=2}^{T} f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W)}{\prod_{t=2}^{T-1} f(\boldsymbol{z}_t|\boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W)} \propto \frac{|P|^{(T-1)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}P^{-1}\right)\right\}}{|P_1|^{T/2-1} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}^* P_1^{-1}\right)\right\}}, \tag{3.1}$$

where $Q = (\boldsymbol{q}_t)$ and $W = (w_t)$, with $\mathcal{S}$ defined in Appendix A.2.3. This is much easier to handle than if we were to not augment the problem.

Posterior Distribution

We will assume independent prior distributions for the marginal and copula parameters again, and so we have

$$\pi(A, \boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W|\mathcal{D}) \propto \pi(A)\pi(\boldsymbol{r})\pi(\boldsymbol{\delta})\pi(\nu)f(U|A, \boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W),$$

with

$$f(U|A,\boldsymbol{r},\boldsymbol{\delta},\nu,Q,W) \propto \frac{|P|^{(T-1)/2}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}P^{-1}\right)\right\}}{|P_1|^{T/2-1}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}^{*}P_1^{-1}\right)\right\}}.$$

MCMC algorithms will be used here to draw realisations from the posterior distributions, where we will adopt the same approach as in the Gaussian copula with stage 1 analysis on the parameters of the margins and stage 2 analysis on the copula parameters.

Inference on simulated data

Here we look at performing analysis on simulated data from the model, specifically with Normal margins. Due to the large amounts of matrix multiplications, and generally much more complex calculations needed in comparison to the Gaussian, the statistical package `R` can be quite cumbersome, and so computations can be a lot slower, thus we shall perform inference with only $n = 100$ points with $p = 2$. Complete steps are given in Appendix A.2.4. We use margins X_1 and X_2 both distributed $N(1,1)$ and have $\nu = 5, \delta_1 = 0.4, \delta_2 = 0.2$ and the same correlation matrix P found in the Gaussian case, with

$$P_1 = \begin{pmatrix} 1.000 & 0.768 \\ 0.768 & 1.000 \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} 0.065 & 0.059 \\ 0.059 & 0.124 \end{pmatrix}. \tag{3.2}$$

In the analysis that follows, we fix all other parameters to their ‘true’ values, in particular the parameters of the correlation matrix of the copula P. We are also mixing over the augmented variables w and $\boldsymbol{q}$, so a fair amount of uncertainty in the posterior distribution is to be expected.

Prior distributions

The prior distributions of the copula parameters will be taken from Smith et al. [2012], except for the r’s where we shall adopt the same priors as we did in the Gaussian, that is, $U(0,5)$ for the constrained r’s and $U(-5,5)$ otherwise. For ν we induce a uniform prior on the interval [2,50], since this would place non-zero weight on either heavy tailed or near Gaussian distributions on the dependence structure. For the skewness parameters δ_i, for $i = 1,\ldots,p$, we assume a $N(0,5^2)$ prior, thus having a large variance and giving weak prior information.

Posterior analysis for $\boldsymbol{\nu}$

We begin with the degrees of freedom, ν, which affects the tail dependence within the copula. Small values of ν, say under 5, represent heavy tails in the skew t copula. With our prior, we are assuming that any value between 2 and 50 is equally likely, and so our posterior beliefs for ν will built upon the data. Looking at Figure 3.3, we see the posterior for ν is weighted heavily towards smaller values and so our prior beliefs, that is all values of ν between 2 and 50 are equally likely, have changed dramatically in light of the data, since, we know that the data was simulated using $\nu = 5$.

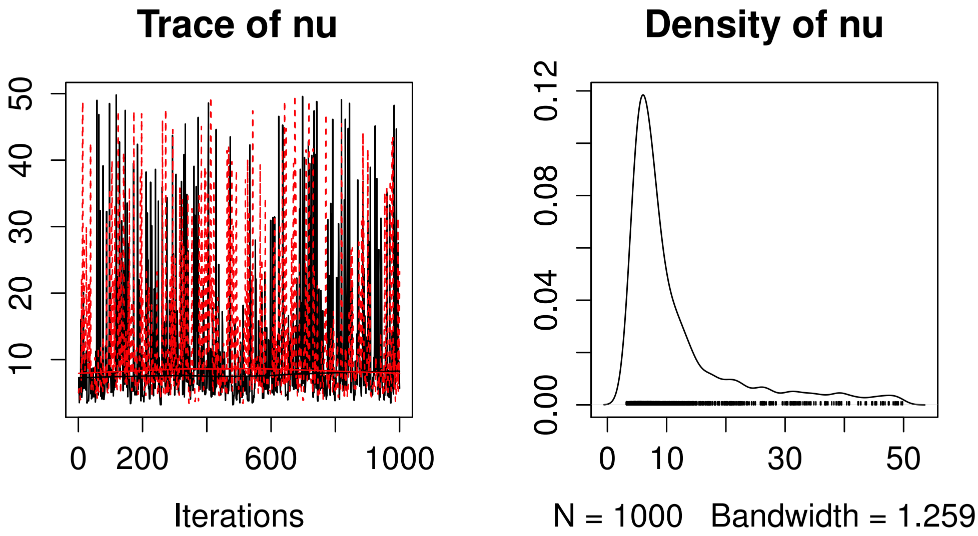

Figure 3.3: 1K iterations from the posterior distribution of ν.

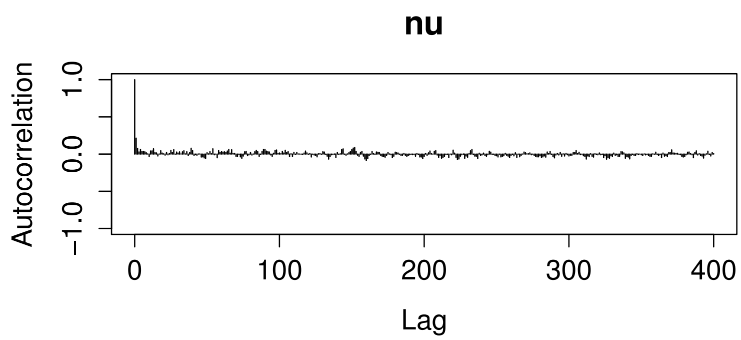

Figure 3.4: Autocorrelation plot for ν.

Parameter	Posterior mean	Posterior SD	Effective sample size
ν	11.417	9.862	646.7

Table 3.1: Posterior summaries for ν.

Convergence and Autocorrelation Figure 3.4 shows the autocorrelation plot for ν, where we see reasonably low autocorrelations and insignificant ones past lag 5. This was after a thinning interval of 20, from a run of 20K, leaving us with 1K iterations, but now we will be left with a higher chance of uncorrelated realisations as a result of this thinning. Table 3.2 shows the posterior summaries and the effective sample size obtained, where we see a large amount of independent realisations from the chain, 647 from 1000. The two chains were ran in parallel in order to asses convergence as well, and both chains seem to be covering the same parameter space.

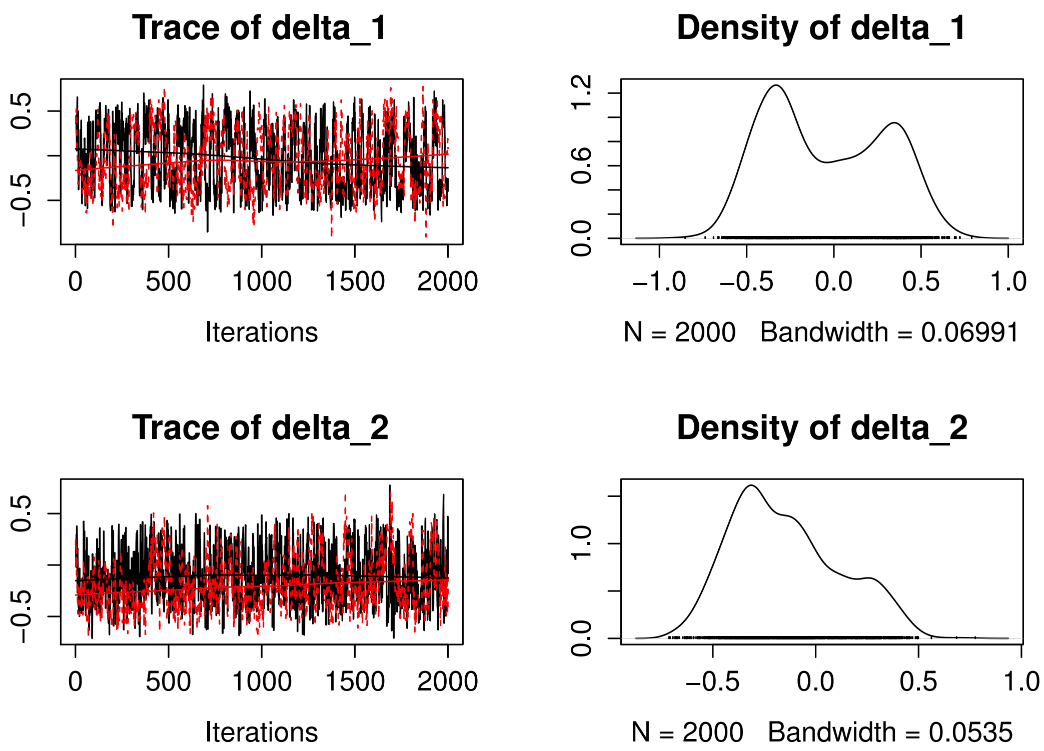

Figure 3.5: 2K iterations from the posterior distributions for δ_1 and δ_2.

Posterior analysis for δ

The prior we place upon each individual skewness parameter δ_i places most weight around 0, but even values of ± 5 are highly likely, with the large 95% interval of (-9.799, 9.799). In our simulations, we have $\delta_1 = 0.4$, and $\delta_2 = 0.2$, and so we hope to see the posterior distribution to be more concentrated around these values than the $N(0, 5^2)$. As we can see in Figure 3.5, the posterior distribution for both of our skewness parameters have a much smaller variance than the prior. Even though the 'true' values used to simulate the data are well within the kernel density plots, there is still a fair degree of uncertainty due to the small sample size in our analysis and the augmentation of the problem. The kernel densities aren't typical bell-shaped densities we would hope for, and instead seem to look bi-modal.

Parameter	Posterior mean	Posterior SD	Effective sample size
δ_1	0.099	0.3533	166.8
δ_2	-0.086	0.2720	469.4

Table 3.2: Posterior summaries for skewness parameters δ_1 and δ_2.

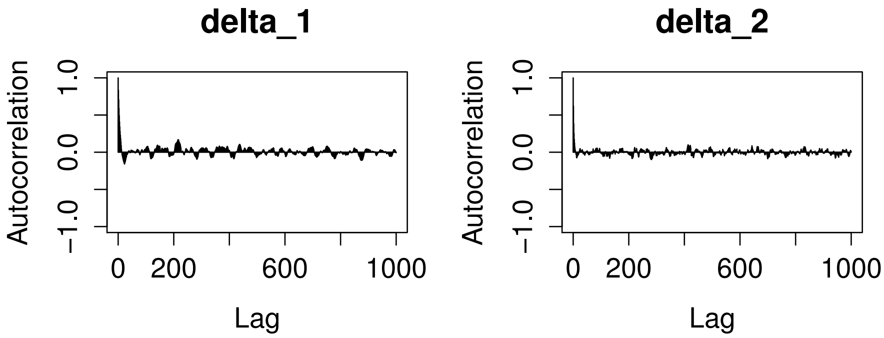

Figure 3.6: Autocorrelation plots for δ_1 and δ_2.

Convergence and Autocorrelation Again, we start two chains in parallel from different initial values in order to assess convergence. Both chains seem to be covering the same parameter space for both parameters. A thinning interval of 15 was used here due to high autocorrelations in the chain, and so we are left with a much smaller chain of only 2K iterations, from a full chain of 30K. The autocorrelations have reduced significantly, as seen in Figure 3.6, meaning that the chain we have here is more likely to contain independent realisations from the posterior. The small number of iterations isn't ideal here, especially since we have small effective sample sizes, although running the algorithm for longer with the same thinning interval would prolong the system time of the algorithm significantly. Remember, we have data augmentation and vast amounts of matrix arithmetic in our scheme, so it would be naive to expect a quick run time in this instance.

Posterior analysis for the joint distribution $(\nu,\boldsymbol{\delta})$

Now we finally look at the joint posterior for $(\nu,\boldsymbol{\delta})$ conditional on the parameters of the correlations matrix of the copula. For this we only ran two chains for 5K iterations due to the very slow algorithm run time. Table 3.3 gives posterior summaries for the parameters and Figure 3.7 gives both trace plots and densities of the marginal densities. Again, all of the 'true' values used to simulate the data are all well within the densities, but with a lot more uncertainty.

Parameter	Posterior mean	Posterior SD	Effective sample size
ν	11.956	9.9610	20.7
δ_1	0.001	0.299	59.7
δ_2	-0.051	0.272	78.6

Table 3.3: Posterior summaries from the joint posterior distributions of $(\nu, \boldsymbol{\delta})$.

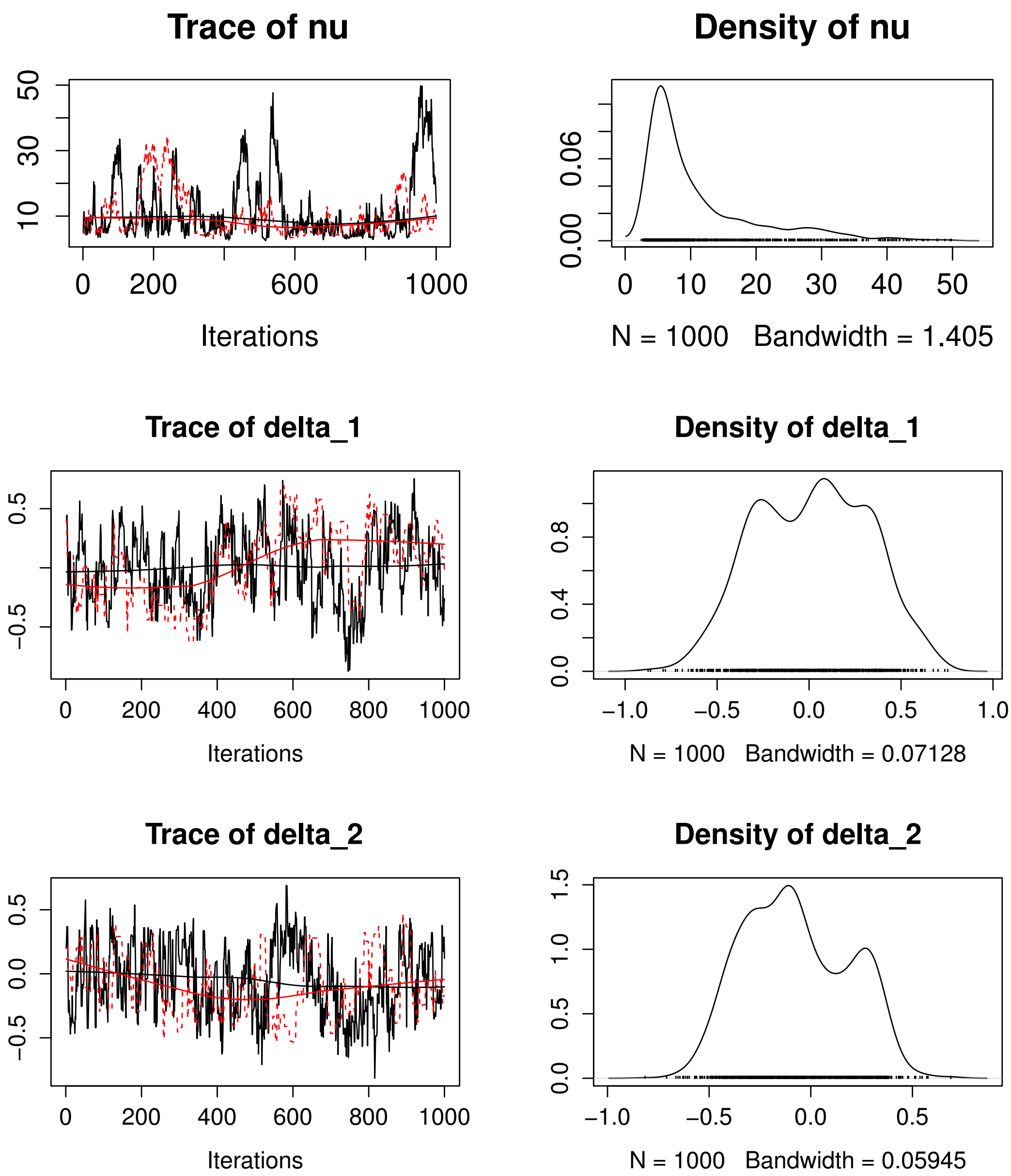

Figure 3.7: 1K realisations from the joint posterior distributions of $(\nu, \boldsymbol{\delta})$.

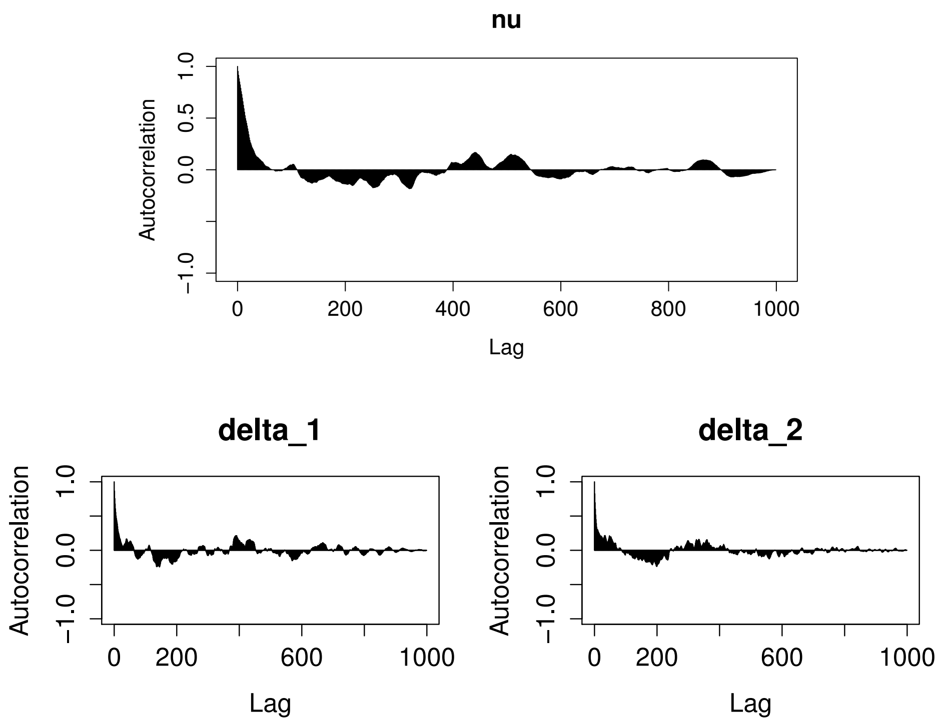

Figure 3.8: Autocorrelation plots for the joint posterior distributions of $(\nu, \boldsymbol{\delta})$.

Convergence and Autocorrelation Thinning the output by 5 does little to help the autocorrelations, as we can see in Figure 3.8. Again we ran two chains to assess convergence and all parameters do seem to be exploring the parameter space, although not at all efficiently as we can see very poor mixing. Looking at Table 3.3, we see very small effective sample sizes and so we have the same issue as previously seen, when we have ν and $\boldsymbol{\delta}$ separately, where we would need to run the algorithm for many more iterations, say 100K, and increase the thinning interval further to obtain a reasonable amount of independent posterior realisations. The increasing uncertainty is to be expected as we allow both ν and $\boldsymbol{\delta}$ to vary with such a small sample size.

Full posterior analysis

Due to mixing over augmented variables, matrix multiplications and hence very slow system run times of these algorithms in `R`, obtaining realisations from the full posterior density (that is, not conditioning upon any parameters) would not be feasible within the time period allowed for this project. The data used here was only 2 dimensional with 100 observations. Imagine using this on actual data with a larger number of dimensions (sites) and a great deal of time points, the inference methods employed here would take far too long to be a viable technique. Thus, our study of the skew t copula ends here.

4 | Conclusions

Modelling with the use of only well-known distributions, analysis falls short when trying to appropriately model dependence in certain real world applications of multivariate statistics. The study was set out to construct and investigate a spatio-temporal model through a different means, specifically with copulas. In Chapter 1 we demonstrated the fundamentals of copulas and how they can be utilised for joint distributions.

Estimating the marginal distributions in our stage 1 analysis was straightforward using independent spatial and temporal assumptions. Exploring the posterior distribution of our model with the Gaussian copula in Chapter 2 wasn't overly challenging either. Although modelling with this is practical in certain areas, we need a model which is appropriate for our example of rainfall volumes, where extreme events need to be taken in to account, an area which the Gaussian fails to capture.

Investigating the posterior distribution of our model with the skew t copula in Chapter 3 was a lot more involved. The complex MCMC schemes developed and mixing over the augmented variables resulted in very long run times, meaning exploring the entire posterior distributions wasn't attainable. We found high autocorrelations in the output, meaning that large thinning intervals were needed and consequently adding to the run times further. An efficient programming language such as Java or `C++` would be more suitable here, and so further work would be aided by ample use of more efficient coding. Once the full posterior could be analysed for simulated data, performing inference on actual rainfall data would be the next logical step.

Modelling real life data such as rainfall is necessary if we want to adequately forecast future outcomes. Hence finding an appropriate joint distribution is crucial, something which we have only begun investigating here due to the need for much more in depth analysis and time.

References

A. Azzalini. *The R `sn` package : The skew-normal and skew-t distributions (version 1.0-0)*. Università di Padova, Italia, 2014. URL http://azzalini.stat.unipd.it/SN.

A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389, 2003.

N. Balakrishnan and C.-D. Lai. *Continuous Bivariate Distributions*. Springer Science+Business Media, New York, NY, second edition, 2009. ISBN 978-0-387-09613-1 (hardback).

J. Barnard, R. McCulloch, and X.-L. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1312, 2000.

E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli. Copulas for finance-a reading guide and some applications. *Available at SSRN 1032533*, 2000.

Bureau of Meteorology, Australia. Average annual, seasonal and monthly rainfall maps, 2014a. URL http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp.

Bureau of Meteorology, Australia. Annual climate report 2013, 2014b. URL http://www.bom.gov.au/climate/annual_sum/2013/AnClimSum2013_LR1.0.pdf.

P. J. Danaher and M. S. Smith. Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science*, 30:4–21, 2011.

M. Fischer, C. Köck, S. Schlüter, and F. Weigert. An empirical analysis of multivariate copula models. *Quantitative Finance*, 9(7):839–854, 2009.

A. Genz and F. Bretz. Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971, 2002.

A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2013. URL http://CRAN.R-project.org/package=mvtnorm. R package version 0.9-9996.

D. A. Hennessy and H. E. Lapan. The use of archimedean copulas to model portfolio allocations. *Mathematical Finance*, 12(2):143–154, 2002.

C. H. Jackson. Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8):1–29, 2011. URL http://www.jstatsoft.org/v38/i08/.

H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401 – 419, 2005.

S. Lee. Formula from hell, Aug. 2009. URL http://www.forbes.com/2009/05/07/gaussian-copula-david-x-li-opinions-columnists-risk-debt.html.

R. B. Nelsen. *An Introduction to Copulas*. Springer Science+Business Media, New York, NY, 2006. ISBN 978-0-387-28678-5 (online).

M. Pitt, D. Chan, and R. Kohn. Bayesian inference for gaussian copula regression models. *Biometrika*, 93:537–554, 2006.

M. Plummer, N. Best, K. Cowles, and K. Vines. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11, 2006. URL http://CRAN.R-project.org/doc/Rnews/.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.

B. Renard and M. Lang. Use of a gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources*, 30(4):897–912, 2007.

S. K. Sahu, D. K. Dey, and M. D. Branco. A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics*, 31 (2):129–150, 2003.

T. Schmidt. Coping with copulas. Technical report, Department of Mathematics, University of Leipzig, Dec. 2006.

M. S. Smith, Q. Gan, and R. J. Kohn. Modelling dependence using skew t copulas: Bayesian inference and applications. *Journal of Applied Econometrics*, 27(3):500–522, 2012.

P. X.-K. Song. Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27:305–320, 2000.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. URL http://www.stats.ox.ac.uk/pub/MASS4. ISBN 0-387-95457-0.

G. R. Warnes, B. Bolker, G. Gorjanc, G. Grothendieck, A. Korosec, T. Lumley, D. MacQueen, A. Magnusson, J. Rogers, and others. *gdata: Various R programming tools for data manipulation*, 2013. URL http://CRAN.R-project.org/package=gdata. R package version 2.13.2.

S. Wilhelm and M. B. G. *tmvtnorm: Truncated Multivariate Normal and Student t Distribution*, 2014. URL http://CRAN.R-project.org/package=tmvtnorm. R package version 1.4-9.

A | Appendix

A.1 Gaussian copula

A.1.1 p-dimensional Gaussian copula density

The Gaussian copula is, adapted from Renard and Lang [2007],

$$C(u_1, \ldots, u_p) = \Phi_p(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)|P).$$

The density is then

$$\begin{aligned}
c(u_1, \ldots, u_p) =& \phi_p(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)|P) \times \left|\textstyle\prod_{i=1}^p \frac{\mathrm{d}}{\mathrm{d}u_i}\Phi^{-1}(u_i)\right| \\
=& \frac{\phi_p(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)|P)}{\prod_{i=1}^p \{\Phi^{-1}(u_i)} \\
=& (2\pi)^{-p/2}|P|^{-1/2} \\
& \times \exp\left\{-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_p)\end{pmatrix}^T (P^{-1} - I_p)\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_p)\end{pmatrix}\right\}.
\end{aligned}$$

A.1.2 Likelihood

We wish to model the distribution of $\boldsymbol{x}_t|\boldsymbol{x}_{t-1}$ due to the temporal dependence. The joint distribution of $\boldsymbol{x}_t, \boldsymbol{x}_{t-1}$ is

$$\begin{aligned}
f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|A, \boldsymbol{\beta}) =& \left\{\prod_{i=1}^p f_i(x_{t,i}|\boldsymbol{\alpha}_i) f_i(x_{t-1,i}|\boldsymbol{\alpha}_i)\right\} \\
& \times c(F_1(x_{t,1}|\boldsymbol{\alpha}_1), \ldots, F_p(x_{t,p}|\boldsymbol{\alpha}_p), \\
& \quad F_1(x_{t-1,1}|\boldsymbol{\alpha}_1), \ldots, F_p(x_{t-1,p}|\boldsymbol{\alpha}_p), |\boldsymbol{\beta}) \\
=& \left\{\prod_{i=1}^p f_i(x_{t,i}|\boldsymbol{\alpha}_i) f_i(x_{t-1,i}|\boldsymbol{\alpha}_i)\right\} \times c(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|A, \boldsymbol{\beta}),
\end{aligned}$$

where $\boldsymbol{u}_t(A) = (F_1(x_{t,1}|\boldsymbol{\alpha}_1), \ldots, F_p(x_{t,p}|\boldsymbol{\alpha}_p))$, that is, each $u_{t,i}(A) = F_i(x_{t,i}|\boldsymbol{\alpha}_i)$. Let $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$, $t = 1, \ldots, T$, then the likelihood is

$$\begin{aligned}
f(X|A, \boldsymbol{\beta}) =& f(\boldsymbol{x}_1|A, \boldsymbol{\beta}) \prod_{t=2}^T f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1} A, \boldsymbol{\beta}) \\
=& f(\boldsymbol{x}_1|A, \boldsymbol{\beta}) \prod_{t=2}^T \frac{f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|A, \boldsymbol{\beta})}{f(\boldsymbol{x}_{t-1}|A, \boldsymbol{\beta})} \\
=& \frac{\prod_{t=2}^T f(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}|A, \boldsymbol{\beta})}{\prod_{t=2}^{T-1} f(\boldsymbol{x}_t|A, \boldsymbol{\beta})}.
\end{aligned}$$

Since the variables in the copula are marginally uniform, we could express this joint density in a second way as follows by letting $U = (\boldsymbol{u}_1(A), \dots, \boldsymbol{u}_T(A))$:

$$f(U|A,\boldsymbol{\beta}) = \frac{\prod_{t=2}^{T} c(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|A,\boldsymbol{\beta})}{\prod_{t=2}^{T-1} c(\boldsymbol{u}_t|A,\boldsymbol{\beta})}.$$

Returning back to $(\boldsymbol{x}_t, \boldsymbol{x}_{t-1})$, we can explicitly show the likelihood as

$$\begin{aligned} f(X|A,\boldsymbol{\beta}) &= \frac{\prod_{t=2}^{T} \left\{\prod_{i=1}^{p} f_i(x_{t,i}|\boldsymbol{\alpha}_i) f_i(x_{t-1,i}|\boldsymbol{\alpha}_i)\right\} c(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|A,\boldsymbol{\beta})}{\prod_{t=2}^{T-1} \left\{\prod_{i=1}^{p} f_i(x_{t,i}|\boldsymbol{\alpha}_i)\right\} c(\boldsymbol{u}_t|A,\boldsymbol{\beta})} \\ &= \frac{\left\{\prod_{i=1}^{p} \prod_{t=2}^{T} f_i(x_{t-1,i}|\boldsymbol{\alpha}_i) c(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|A,\boldsymbol{\beta})\right\} \prod_{i=1}^{p} f_i(x_{T,i}|\boldsymbol{\alpha}_i)}{\prod_{t=2}^{T-1} c(\boldsymbol{u}_t|A,\boldsymbol{\beta})}. \end{aligned}$$

A.1.3 Simulating from the Gaussian copula with time dependence

In order to test that the MCMC algorithm is working, we first consider

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \\ Z_{1,t-1} \\ Z_{2,t-1} \end{pmatrix} \sim N_4(\mathbf{0}, P) \tag{A.1}$$

for the 2 dimensional case (although this is easily generalised for the p-dimensional case), where P is given in (2.5). We form the following conditional distribution, conditional upon $Z_{1,t-1} = z_{1,t-1}, Z_{2,t-1} = z_{2,t-1}$. This allows us to simulate data from the copula model with temporal dependence by simulating values from

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} \Bigg| \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} \sim N_2(\boldsymbol{\mu}, \Sigma) \tag{A.2}$$

where

$$\boldsymbol{\mu} = P_2 P_1^{-1} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} \qquad \text{and} \qquad \Sigma = P_1 - P_2 P_1^{-1} P_2.$$

Algorithm

1. Generate $\boldsymbol{Z}_1 \sim N_2(\mathbf{0}, P_1)$
2. For $t = 2, \dots, T$,
    - Generate subsequent values $\boldsymbol{Z}_t|\boldsymbol{z}_{t-1} \sim N_p(\boldsymbol{\mu}, \Sigma)$.
    - Set $\boldsymbol{U}_t = (\Phi(Z_{t,1}), \dots, \Phi(Z_{t,p}))$.
    - Set $\boldsymbol{X}_t = (F_1^{-1}(U_{t,1}), \dots, F_p^{-1}(U_{t,p}))$.

A.1.4 Paramaterizing the correlation matrix $\boldsymbol{P}$

An effective way of obtaining an appropriate Σ with the correct structure is to take $\boldsymbol{X}_t = \boldsymbol{M} + \boldsymbol{R}_t$, that is, $\boldsymbol{X}_t$ is composed of a random variable $\boldsymbol{M}$ with $\text{Var}(\boldsymbol{M})=V$ and time-dependent noise, $\boldsymbol{R}_t$ with $\text{Var}(\boldsymbol{R}_t) = W$, with $\boldsymbol{M}$ and $\boldsymbol{R}_t$ independent. This choice produces a covariance matrix

$$\Sigma = Cov(X_t, X_{t-1}) = \begin{pmatrix} V+W & V \\ V & V+W \end{pmatrix}. \tag{A.3}$$

The resulting correlation matrix constructed from this then has the correct form as expressed in (2.5).

A.1.5 MCMC algorithm

Stage 1 First we perform analysis assuming spatial and temporal independence to get realisations from the independent posterior densities

$$\tilde{\pi}(\boldsymbol{\alpha}_i|\mathcal{D}) \propto \pi(\boldsymbol{\alpha}_i)\prod_{t=1}^{T} f_i(x_{t,i}|\boldsymbol{\alpha}_i).$$

Stage 2

1. Initialise
   - Simulate the constrained r's from $r_i^{+(0)} \sim U(0,5)$ indep.
   - Simulate the unconstrained r's from $r_i^{\pm(0)} \sim U(-5,5)$ indep.
   - Construct the correlation matrix $P^{(0)}$ using the setup in A.1.4
   - Set $u_{t,i} = F_i(x_{t,i}|\boldsymbol{\alpha}_i)$
2. Each parameter takes their current value in the chain. For iterations $1,\dots,N$
3. For $i = 1,\dots,p$ (constrained r's)
   - simulate a proposal value $r_i^{+c} \sim LN(\log r_i^+, \tau_{r_i}^2)$;
   - construct $P(r_i^{+c})$ and hence $P_1(r_i^{+c})$ and then, assuming uniform priors,
   $$\log B = \log f(U|P(r_i^{+c})) - \log f(U|P(r_i^{+}))$$
   - accept proposal with probability $\min(1,B)$
4. For $i = 1,\dots,p(p-1)$ (unconstrained r's)
   - simulate a proposal value $r_i^{\pm c} \sim N(r_i^{\pm}, \tau_{r_i}^2)$;
   - construct $P(r_i^{\pm c})$ and hence $P_1(r_i^{\pm c})$ and then, assuming uniform priors,
   $$\log B = \log f(U|P(r_i^{\pm c})) - \log f(U|P(r_i^{\pm}))$$
   - accept proposal with probability $\min(1,B)$

A.2 Skew t copula

A.2.1 Conditional Gaussian representation

By augmenting the skew t distribution as follows, we can make our inference on the skew t copula slightly less challenging. Using the full $(\boldsymbol{X}, \boldsymbol{Q})$ representation, we have

$$f(\boldsymbol{x}, \boldsymbol{q}|\boldsymbol{Q} > \boldsymbol{0}) \propto f(\boldsymbol{x}|\boldsymbol{q})\, f(\boldsymbol{q})\, I(\boldsymbol{q} > \boldsymbol{0})$$

for which $\boldsymbol{X} \sim St_\nu(P, \boldsymbol{\delta})$ marginally, where

$$\boldsymbol{X}|\boldsymbol{q} \sim t_{\nu+p}\left(\Delta\boldsymbol{q},\, \frac{\nu + \boldsymbol{q}^T\boldsymbol{q}}{\nu + p}\, P\right).$$

We can further simplify this problem by introducing $W \sim Ga(\nu/2, \nu/2)$, then

$$f(\boldsymbol{x}, \boldsymbol{q}, w|\boldsymbol{Q} > \boldsymbol{0}) \propto f(\boldsymbol{x}|\boldsymbol{q}, w)\, f(\boldsymbol{q}|w)\, I(\boldsymbol{q} > \boldsymbol{0})\, f(w)$$

where

$$\boldsymbol{X}|\boldsymbol{q}, w \sim N_p\left(\Delta\boldsymbol{q},\, \frac{1}{w}\, P\right) \qquad \text{and} \qquad \boldsymbol{Q}|w \sim N_p\left(\boldsymbol{0},\, \frac{1}{w}\, I_p\right).$$

We still maintain $\boldsymbol{X} \sim St_\nu(P, \boldsymbol{\delta})$ marginally. Using this representation, we now simplify the skew t copula as

$$C_{St}(\boldsymbol{u}|P, \boldsymbol{\delta}, \nu) \propto \int \Phi_p\left(\sqrt{w}P^{-1/2}(\boldsymbol{x} - \Delta\boldsymbol{q})\right)\, f(\boldsymbol{q}|w)\, I(\boldsymbol{q} > \boldsymbol{0})\, f(w)\, dw\, d\boldsymbol{q}$$

where $\Phi_p(\boldsymbol{x}) = \prod_{i=1}^p \Phi(x_i)$ is the p-dimensional standard normal distribution function.

Thus

$$c_{St}(\boldsymbol{u}|P, \boldsymbol{\delta}, \nu, \boldsymbol{q}, w) = \frac{f(\boldsymbol{x}|P, \boldsymbol{\delta}, \nu, \boldsymbol{q}, w)}{\prod_{i=1}^p f_{St,i}(x_i|\delta_i, \nu)} = \frac{\phi_p(\boldsymbol{x}|\Delta\boldsymbol{q}, w^{-1}P)}{\prod_{i=1}^p f_{St,i}(x_i|\delta_i, \nu)},$$

where $x_i = F_{St,i}^{-1}(u_i|\delta_i, \nu)$ and $\phi_p(\cdot|\boldsymbol{\mu}, \Sigma)$ is the p-dimensional $N_p(\boldsymbol{\mu}, \Sigma)$ density function.

A.2.2 Simulating variates from the skew t copula

Algorithm

1. Simulate $w \sim Ga(\nu/2, \nu/2)$
2. Simulate $\boldsymbol{q} \sim N_p(\boldsymbol{0}, w^{-1}I_p)$, constrained to $\boldsymbol{q} > \boldsymbol{0}$
3. Simulate $\boldsymbol{Z} \sim N_p(\Delta\boldsymbol{q}, w^{-1}P)$
4. For $i = 1, \ldots, p$,
   - Set $U_i = F_{St,i}(Z_i|\delta_i, \nu)$
   - Set $X_i = F_i^{-1}(U_i)$

A.2.3 Likelihood

In our model, we have a stationary skew t copula for $(\boldsymbol{u}_t, \boldsymbol{u}_{t-1})$, with

$$c_{St}(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu) = \frac{f_{St}(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu)}{\prod_{i=1}^p f_{St,i}(z_{t,i}|\delta_i, \nu)\, f_{St,i}(z_{t-1,i}|\delta_i, \nu)},$$

where $z_{t,i} = F_{St,i}^{-1}(u_{t,i}|\delta_i, \nu)$. Introducing the augmented variables, $(\boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1})$, we then have

$$c_{St}(\boldsymbol{u}_t, \boldsymbol{u}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1}) = \frac{f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1})}{\prod_{i=1}^p f_{St,i}(z_{t,i}|\delta_i, \nu)\, f_{St,i}(z_{t-1,i}|\delta_i, \nu)},$$

where

$$\left.\begin{pmatrix} \boldsymbol{z}_t \\ \boldsymbol{z}_{t-1} \end{pmatrix}\right| \boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1} \sim N_{2p}\left(\begin{pmatrix} \Delta\boldsymbol{q}_t \\ \Delta\boldsymbol{q}_{t-1} \end{pmatrix}, \operatorname{diag}(w_t^{-1}\, I_p, w_{t-1}^{-1}\, I_p)\, P \right).$$

We have the likelihood, in terms of the uniform variables, with $Q = (\boldsymbol{q}_t)$ and $W = (w_t)$, as

$$\begin{aligned}
f(U|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}, w) &= c_{St}(\boldsymbol{u}_1|\boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W) \prod_{t=2}^T c_{St}(\boldsymbol{u}_t|\boldsymbol{u}_{t-1}, \boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W) \\
&= \frac{f(\boldsymbol{z}_1|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_1, w_1)}{\prod_{i=1}^p f_{St,i}(z_{1,i}|\delta_i, \nu)} \prod_{t=2}^T \frac{f(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1})}{\prod_{i=1}^p f_{St,i}(z_{t,i}|\delta_i, \nu)} \\
&= \frac{f(\boldsymbol{z}_1|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_1, w_1)}{\prod_{t=1}^T \prod_{i=1}^p f_{St,i}(z_{t,i}|\delta_i, \nu)} \prod_{t=2}^T \frac{f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}, w)}{f(\boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}, w)} \\
&= \frac{1}{\prod_{t=1}^T \prod_{i=1}^p f_{St,i}(z_{t,i}|\delta_i, \nu)} \times \frac{\prod_{t=2}^T f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1})}{\prod_{t=2}^{T-1} f(\boldsymbol{z}_t|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, w_t)}.
\end{aligned}$$

We now simplify the numerator and denominator of the second term above:

$$\begin{aligned}
&\prod_{t=2}^T f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}|\boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, \boldsymbol{q}_{t-1}, w_t, w_{t-1}) \\
&\propto \prod_{t=2}^T |P|^{-1/2} \\
&\quad \times \exp\left\{ -\frac{1}{2} \left((\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T (\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})^T \right) \operatorname{diag}(w_t\, I_p, w_{t-1}\, I_p) P^{-1} \begin{pmatrix} \boldsymbol{z}_t - \Delta\boldsymbol{q}_t \\ \boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1} \end{pmatrix} \right\} \\
&\propto |P|^{(T-1)/2} \\
&\quad \times \exp\left\{ \sum_{t=2}^T \operatorname{tr}\left[\left((\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T (\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})^T \right) \operatorname{diag}(w_t\, I_p, w_{t-1}\, I_p) P^{-1} \begin{pmatrix} \boldsymbol{z}_t - \Delta\boldsymbol{q}_t \\ \boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1} \end{pmatrix} \right] \right\} \\
&\propto |P|^{(T-1)/2} \\
&\quad \times \exp\left\{ \operatorname{tr}\left[\sum_{t=2}^T \begin{pmatrix} \boldsymbol{z}_t - \Delta\boldsymbol{q}_t \\ \boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1} \end{pmatrix} \left((\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T (\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})^T \right) \operatorname{diag}(w_t\, I_p, w_{t-1}\, I_p) P^{-1} \right] \right\}.
\end{aligned}$$

If we introduce a new $2(p \times p)$ matrix

$$\begin{aligned}\mathcal{S} &= \begin{pmatrix} \sum_{t=2}^{T} w_t(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T & \sum_{t=2}^{T} w_{t-1}(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)(\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})^T \\ \sum_{t=2}^{T} w_t(\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T & \sum_{t=2}^{T} w_{t-1}(\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})(\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})^T \end{pmatrix} \\ &= \begin{pmatrix} \sum_{t=2}^{T} w_t(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T & \sum_{t=2}^{T} w_{t-1}(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)(\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})^T \\ \sum_{t=2}^{T} w_t(\boldsymbol{z}_{t-1} - \Delta\boldsymbol{q}_{t-1})(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T & \sum_{t=1}^{T-1} w_t(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T \end{pmatrix}\end{aligned}$$

then

$$\prod_{t=2}^{T} f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1} | \boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W) \propto |P|^{(T-1)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}P^{-1}\right) \right\}.$$

Using the same idea,

$$\begin{aligned}\prod_{t=2}^{T-1} f(\boldsymbol{z}_t | \boldsymbol{r}, \boldsymbol{\delta}, \nu, \boldsymbol{q}_t, w_t) &\propto \prod_{t=2}^{T-1} |P_1|^{-1/2} \exp\left\{ -\frac{w_t}{2} \left(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t\right)^T P_1^{-1} \left(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t\right) \right\} \\ &\propto |P_1|^{T/2-1} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1^{-1}\right) \right\}\end{aligned}$$

where $\mathcal{S}^* = \sum_{t=2}^{T-1} w_t(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T$. Thus we have our likelihood in (3.1) as

$$\frac{\prod_{t=2}^{T} f(\boldsymbol{z}_t, \boldsymbol{z}_{t-1} | \boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W)}{\prod_{t=2}^{T-1} f(\boldsymbol{z}_t | \boldsymbol{r}, \boldsymbol{\delta}, \nu, Q, W)} \propto \frac{|P|^{(T-1)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}P^{-1}\right) \right\}}{|P_1|^{T/2-1} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1^{-1}\right) \right\}}.$$

Note that $\mathcal{S}$ is a function of $\boldsymbol{\delta}$ and ν as it is a function of $z_{t,i} = F_{St,i}^{-1}(u_{t,i} | \delta_i, \nu)$, and P is a function of the r's.

We will be performing Metropolis Hastings on each of these parameters, with Normal random walks for all except for the contained r's which will have Lognormal random walks. In the following acceptance probabilities, the proposed value is denoted with a superscript c, as in r_{ij}^c.

A Normal or Lognormal random walk move for the individual r_{ij}, the Metropolis Hastings acceptance probability is $\min(1, B)$, where

$$\begin{aligned}B &= \frac{\pi(r_{ij}^c) f(U | r_{ij}^c, \{\boldsymbol{r} \backslash r_{ij}\}, \cdot)}{\pi(r_{ij}) f(U | \boldsymbol{r}, \cdot)} \\ &= \frac{\pi(r_{ij}^c)}{\pi(r_{ij})} \times \frac{|P(r_{ij}^c)|^{(T-1)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}P(r_{ij}^c)^{-1}\right) \right\}}{|P_1(r_{ij}^c)|^{T/2-1} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1(r_{ij}^c)^{-1}\right) \right\}} \\ &\quad \times \frac{|P_1(r_{ij})|^{T/2-1} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1(r_{ij})\right)^{-1} \right\}}{|P(r_{ij})|^{(T-1)/2} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left(\mathcal{S}P(r_{ij})\right)^{-1} \right\}}.\end{aligned}$$

A Normal random walk move for the individual δ_i has Metropolis Hastings acceptance probability is $\min(1, B)$, where

$$
\begin{aligned}
B &= \frac{\pi(\delta_i^c)\, f(U|\delta_i^c, \{\boldsymbol{\delta}\backslash\delta_i\}, \cdot)}{\pi(\delta_i)\, f(U|\boldsymbol{\delta}, \cdot)} \\
&= \frac{\pi(\delta_i^c)}{\pi(\delta_i)} \times \frac{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\delta_i^c)P^{-1}\right)\right\}}{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\delta_i^c)P_1^{-1}\right)\right\}} \times \frac{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\delta_i)P_1^{-1}\right)\right\}}{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\delta_i)P^{-1}\right)\right\}} \\
&\quad \times \frac{\prod_{t=1}^T f_{St,i}(z_{t,i}|\delta_i, \nu)}{\prod_{t=1}^T f_{St,i}(z_{t,i}^c|\delta_i^c, \nu)}.
\end{aligned}
$$

A Normal random walk move for ν has Metropolis Hastings acceptance probability is $\min(1, B)$, where

$$
\begin{aligned}
B &= \frac{\pi(\nu^c)\, f(U|\nu^c, \cdot)}{\pi(\nu)\, f(U|\nu, \cdot)} \\
&= \frac{\pi(\nu^c)}{\pi(\nu)} \times \frac{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\nu^c)P^{-1}\right)\right\}}{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\nu^c)P_1^{-1}\right)\right\}} \times \frac{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\nu)P_1^{-1}\right)\right\}}{\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\nu)P^{-1}\right)\right\}} \\
&\quad \times \frac{\prod_{t=1}^T \prod_{i=1}^p f_{St,i}(z_{t,i}|\delta_i, \nu)}{\prod_{t=1}^T \prod_{i=1}^p f_{St,i}(z_{t,i}^c|\delta_i, \nu^c)}.
\end{aligned}
$$

For the augmented variables, we have Gibbs steps. The moves for the w_t are Gibbs steps with

$$w_t|\cdot \sim Ga(A, B)$$

where

$$A = \frac{\nu}{2} + m + 1, \qquad B = \frac{1}{2}\left\{(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T P_1^{-1}(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t) + \boldsymbol{q}_t^T\boldsymbol{q}_t + \nu\right\}.$$

The moves for the elements of the $\boldsymbol{q}_t$ are done one at a time from the conditional distribution of $q_{t,i}|\boldsymbol{q}_t\backslash q_{t,i}, \cdot$, constrained so that $q_{t,i} > 0$, where

$$\boldsymbol{q}_t|\cdot \sim N_p\left(w_t(I + w_t\Delta P_1^{-1}\Delta)^{-1}\Delta P_1^{-1}\boldsymbol{z}_t,\, (I + w_t\Delta P_1^{-1}\Delta)^{-1}\right).$$

A.2.4 MCMC algorithm

Since the stage 1 algorithm is the same as the stage 1 for the Gaussian copula, we will look at the stage 2 analysis where we perform inference on the copula parameters. Below is the full algorithm for all copula parameters.

1. Initialise
   - Simulate the skewness parameters $\delta_i^{(0)}$, $i = 1, \ldots, p$ from its prior: $\delta_i \sim N(0, 5^2)$, indep.
   - Simulate the degrees of freedom $\nu^{(0)}$ from its prior: $\nu \sim U(2, 50)$
   - Simulate the constrained r's from $r_i^{+(0)} \sim U(0, 5)$ indep.

- Simulate the unconstrained r's from $r_i^{\pm(0)} \sim U(-5,5)$ indep.
- Construct the correlation matrix $P^{(0)}$ using the setup in A.1.4
- Set $u_{t,i} = F_i(x_{t,i}|\boldsymbol{\alpha}_i)$

2. Each parameter takes their current value in the chain. For iterations $1,\dots,N$

- Set $z_{t,i} = F^{-1}_{St,i}\big(u_{t,i}\big|\delta_i,\nu\big)$
- For $t=1,\dots,T$, simulate $w_t|\cdot \sim Ga(A,B)$ where

$$A = \frac{\nu}{2} + m + 1, \qquad B = \frac{1}{2}\left\{(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t)^T P_1^{-1}(\boldsymbol{z}_t - \Delta\boldsymbol{q}_t) + \boldsymbol{q}_t^T\boldsymbol{q}_t + \nu\right\}.$$

- For $t=1,\dots,T$,
  - for $i=0,\dots,m$, simulate $q_{t,i}|\boldsymbol{q}_t\backslash q_{t,i},\cdot$, constrained so that $q_{t,i}>0$, where

$$\boldsymbol{q}_t|\cdot \sim N_p\left(w_t(I + w_t\Delta P_1^{-1}\Delta)^{-1}\Delta P_1^{-1}\boldsymbol{z}_t,\ (I + w_t\Delta P_1^{-1}\Delta)^{-1}\right).$$

- For $i=1,\dots,p$
  - simulate a proposal value $\delta_i^c \sim N(\delta_i,\tau^2_{\delta_i})$
  - calculate $z^c_{t,i} = F^{-1}_{St,i}(u_{t,i}|\delta_i^c,\nu)$, $t=1,\dots,T$ and then $\mathcal{S}(\delta_i^c)$ and $\mathcal{S}^*(\delta_i^c)$.
  - calculate

$$\begin{aligned}\log B =& \log\pi(\delta_i^c) - \log\pi(\delta_i)\\ &- \frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\delta_i^c)P^{-1}\right) + \frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\delta_i^c)P_1^{-1}\right)\\ &- \frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\delta_i)P_1^{-1}\right) + \frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\delta_i)P^{-1}\right)\\ &+ \sum_{t=1}^T \log f_{St,i}(z_{t,i}|\delta_i,\nu) - \sum_{t=1}^T \log f_{St,i}(z^c_{t,i}|\delta_i^c,\nu)\end{aligned}$$

  - accept proposal with probability $\min(1,B)$.
  - If proposal is accepted then recalculate $z_{t,i} = F^{-1}_{St,i}(u_{t,i}|\delta_i,\nu)$, $t=1,\dots,T$.
- Simulate a proposal value $\nu^c \sim N(\nu,\tau^2_\nu)$. Reject proposal if $\nu^c \notin (2,50)$ as this will have zero prior density. If proposal is not rejected then
  - calculate $z^c_{t,i} = F^{-1}_{St,i}(u_{t,i}|\delta_i,\nu^c)$, $t=1,\dots,T$
  - calculate

$$\begin{aligned}\log B =& \log\pi(\nu^c) - \log\pi(\nu)\\ &- \frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\nu^c)P^{-1}\right) - \frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\nu)P_1^{-1}\right)\\ &+ \frac{1}{2}\operatorname{tr}\left(\mathcal{S}^*(\nu^c)P_1^{-1}\right) + \frac{1}{2}\operatorname{tr}\left(\mathcal{S}(\nu)P^{-1}\right)\\ &+ \sum_{t=1}^T\sum_{i=1}^p \log f_{St,i}(z_{t,i}|\delta_i,\nu) - \sum_{t=1}^T\sum_{i=1}^p \log f_{St,i}(z^c_{t,i}|\delta_i,\nu^c)\end{aligned}$$

- accept proposal with probability $\min(1, B)$

- For $i = 1, \ldots, p$ (constrained r's)
  - simulate a proposal value $r_i^{+c} \sim LN(\log r_i^+, \tau_{r_i}^2)$;
  - construct $P(r_i^{+c})$ and hence $P_1(r_i^{+c})$ and then, assuming uniform priors,

$$\begin{aligned}\log B =& \frac{T-1}{2} \log |P(r_i^{+c})| - \frac{1}{2} \operatorname{tr}\left(\mathcal{S} P(r_i^{+c})^{-1}\right) - \left(\frac{T}{2} - 1\right) \log |P_1(r_i^{+c})| \\ &+ \frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1(r_i^{+c})^{-1}\right) + \left(\frac{T}{2} - 1\right) \log |P_1(r_i^{+})| - \frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1(r_i^{+})^{-1}\right) \\ &- \frac{T-1}{2} \log |P(r_i^{+})| + \frac{1}{2} \operatorname{tr}\left(\mathcal{S} P(r_i^{+})^{-1}\right).\end{aligned}$$

  - accept proposal with probability $\min(1, B)$

- For $i = 1, \ldots, p(p-1)$ (unconstrained r's)
  - simulate a proposal value $r_i^{\pm c} \sim N(r_i^{\pm}, \tau_{r_i}^2)$;
  - construct $P(r_i^{\pm c})$ and hence $P_1(r_i^{\pm c})$ and then, assuming uniform priors,

$$\begin{aligned}\log B =& \frac{T-1}{2} \log |P(r_i^{\pm c})| - \frac{1}{2} \operatorname{tr}\left(\mathcal{S} P(r_i^{\pm c})^{-1}\right) - \left(\frac{T}{2} - 1\right) \log |P_1(r_i^{\pm c})| \\ &+ \frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1(r_i^{\pm c})^{-1}\right) + \left(\frac{T}{2} - 1\right) \log |P_1(r_i^{\pm})| - \frac{1}{2} \operatorname{tr}\left(\mathcal{S}^* P_1(r_i^{\pm})^{-1}\right) \\ &- \frac{T-1}{2} \log |P(r_i^{\pm})| + \frac{1}{2} \operatorname{tr}\left(\mathcal{S} P(r_i^{\pm})^{-1}\right).\end{aligned}$$

  - accept proposal with probability $\min(1, B)$