Random t-distributed Fields

Jo Røyslien[1]

1998

[1]ADDRESS FOR CORRESPONDENCE: Henning Omre, Department of Mathematical Sciences, The Norwegian University of Science and Technology, 7034 Trondheim, Norway. Email: omre@stat.ntnu.no

Abstract

The t-distribution is a density distribtution with shape similar to the shape of the Gaussian distribution, but with more probability mass in the tails. Because of this, a random t-distributed field is a random field with more variability than a Gaussian random field.

A random field is a function whose values are random variables on a parameter set, for example $\mathbf{R}^2$.

In this report a sequential algorithm for the generation of both conditional and unconditional random t-distributed fields is created. The algorithm uses rejection sampling with the Cauchy distribution as proposal. An EM-algorithm for the estimation of the parameters in a random t-distributed field is also created. ML-estimators are found for all parameters, but to speed up the algorithm under the assumption of a parametrisized covariance matrix, the maximum quasilikelihood estimator is calculated for the parameter of the exponential covariance function.

The algorithm for generation of random t-distributed fields is then tested through computer simulation. From the resulting samples one finds that unconditional random t-distributed fields are very voulnerable to the first couple of sampled values, whereas conditional simulation is quite capable of preserving the wanted behaviour of the original random field, but with more variability than the frequently used Gaussian model. The random t-distributed fields model therefore appears to be an alternative to the Lévy stable model. The simulation of random t-distributed fields with the algorithm presented here is slightly slower than the simulation of Gaussian random fields, as one would expect.

Finally it is shown that in order to generate random t-distributed fields one has to apply an algorithm as the one presented here. Transforming a Gaussian random field appears to be analytically impossible.

A NOTE TO THE READER

This report is a diploma-thesis ($\sim$master-thesis) in statistics written by Jo Røyslien, during the fall 1997, at Department of Mathematical Sciences, The Norwegian University of Science and Technology, Trondheim, Norway. There could be some public interest in this work, so I therefore make it public available as a technical report. I hope the report will be of use for the reader with interest in Random t-distributed Fields.

Please keep in mind that our diploma-thesis consists of 4 months of work with an absolute time-limit and that the candidate had only basic knowledge of random fields and stochastic sampling before he started on this work. Not all topics are covered/cited and the thesis should be read with this in mind.

September 11, 1998, Henning Omre

Contents

Preface

This diploma thesis is written at The Norwegian University of Science & Technology, Department of Mathematical Sciences. The work is partly built on my fourth year project, and the time spent is four months.

The work has mainly been done at The Norwegian Computing Centre, and I would therefore like to thank them for providing me with an office and a computer, and especially Petter Abrahamsen for looking after me. My fellow student Andreas Thorsen has been a valuable consultant throughout the work.

The task was primerally to create an algorithm to generate random t-distributed fields and an algorithm to estimate the parameters of such fields. In order to manage this, however,

I had to gain control over a lot more than random fields,

such as numerical optimization, the EM-algorithm and quasilikelihood. Throughout this process my tutor Henning Omre has been of great help in guiding me to the fields of theory where the answers to the occurring problems were most likely to be found.

Oslo 22.02.98 Jo Røislien.

1 Introduction

The generation of random fields is an important issue in spatial statistics. The most frequently used model has been the Gaussian, see [5], i.e., the generation of Gaussian random fields has been central. The reasons for this are both due to the fact that the Central Limit Theorem ensures that all distributions will converge to the Gaussian, see [10], and that the Gaussian distribution is easy to handle mathematically. Because of this the Gaussian random fields model has been used despite the fact that it in many cases has been known to underestimate variability.

During the last years new models have been suggested, among these a model developed by Scott Painter and Lincoln Paterson, see [1] and [2]. Painter and Paterson claim that their model, which has its base in the Lévy stable distributions, is superior to the existing Gaussian model basically because the Lévy stable distributions contain more probability mass in the tails, thus implying greater variability. The Lévy stable distributions are all symmetric, like the Gaussian, and the class contains the Gaussian distribution as a singularity. One of the problems with this model, though, is the fact that the Lévy stable distributions do not have an analytical closed form, and the algorithms suggested for practical purposes are troublesome and cannot guarantee exact results.

Based on this it is evident that one should look for alternative models, and if the model proposed by Painter and Paterson has its advantage in the heavy tail-behaviour, it is probably worthwhile testing well-known distributions that also possess heavy tails compared to the Gaussian.

This idea is persued in this report by creating a model that better explains the variability in data than the Gaussian does. After having stated that many of the results that make the Gaussian distribution particularly tractable also hold for another and wider class of distributions including the Gaussian, the elliptic symmetric distributions, see [5], the multivariate t-distribution is extracted to put forward a procedure for generating random t-distributed fields. This procedure does not exclude the existing Gaussian model, but includes it as a special case. The esitmation of parameters in a given field that is assumed to be t-distributed is then considered, before the generation and estimation procedures are tested through computer simulation. Finally the transformation of Gaussian random fields to t-distributed fields is discussed.

2 Random fields

First a short overview on random fields is given. The results in section 2.1 are taken from [3].

2.1 Definition of random fields

A formal definition of a random field is:

Definition 1 ***Random fields***
Let a probability space, $(\Omega, \mathcal{F}, P)$, and a parameter set, $\mathcal{X}$, be given. A random field is then a finite or real valued function $Z(\mathbf{x}, \omega)$ which, for every fixed $\mathbf{x} \in \mathcal{X}$, is a measurable function of $\omega \in \Omega$.

Here $\mathcal{X} = \mathbf{R}^2$. For a fixed $\omega \in \Omega$, $Z(\cdot, \omega)$ is a non-random function of $\mathbf{x}$. The parameter $\mathbf{x}$ is called the coordinate. What the definition then says, is that a random field Z on $\mathbf{R}^2$, i.e., $\mathbf{x} \in \mathbf{R}^2$, is a function whose values are random variables for any $\mathbf{x} \in \mathbf{R}^2$.

2.2 Description of random fields

A random field is usually described by the finite-dimensional cumulative distribution

$$F_{\mathbf{x_1},\dots,\mathbf{x_k}}(z_1, \dots, z_k) = Prob\{Z_{\mathbf{x_1}} \leq z_1, \dots, Z_{\mathbf{x_k}} \leq z_k\}$$

Alternatively it can be described by the finite-dimensional density distribution

$$f_{\mathbf{x_1},\dots,\mathbf{x_k}}(z_1, \dots, z_k) = \frac{\partial^k F_{\mathbf{x_1},\dots,\mathbf{x_k}}(z_1, \dots, z_k)}{\partial z_1 \cdots \partial z_k}$$

if it exists, i.e., if the cumulative distribution is differentiable.

This way of writing density functions, as $f_{\mathbf{x}}(\mathbf{z})$ rather than $f_{\mathbf{Z}}(\mathbf{z})$, may be unfamiliar to some, but it is only a way of writing that it is the random field generated from a particular densitiy distribution that is being considered. Throughout this report the notation $f_{\mathbf{Z}}(\mathbf{z})$ will be used when the density distributions are discussed in general, and $f_{\mathbf{x}}(\mathbf{z})$ when they are applied to the theory of random fields.

2.3 Generation of random fields

If one wants to generate a random field in two dimensions on an area $\mathcal{D} \subset \mathbf{R}^2$, one way of doing this is to discretisize the area $\mathcal{D}$ into a lattice $\mathcal{L}_{\mathcal{D}}$ and then generate the field on the finite number of points $\{x_{ij}; i,j \in \mathcal{L}_{\mathcal{D}}\}$ only. From probability theory the relation

$$f_{\mathbf{x}}(\mathbf{z}) = f_{\mathbf{x}}(z_n | z_1, \dots, z_{n-1}) \cdot f_{\mathbf{x}}(z_{n-1} | z_1, \dots, z_{n-2}) \cdots f_{\mathbf{x}}(z_2 | z_1) \cdot f_{\mathbf{x}}(z_1) \tag{1}$$

is known and this can now be used in an algorithm to generate random fields by letting n be the number of points in the lattice $\mathcal{L}_{\mathcal{D}}$ and then sequentially run through all these points $\{x_{ij}; i,j \in \mathcal{L}_{\mathcal{D}}\}$ using expression (1).

Algorithm 1 ***General sequential algorithm***

- *Initiation: Draw a sample from the one-dimensional unconditional distribution $f_{\mathbf{x}}(z_1)$.*

- *Iterate $i = 2 \dots n$: Draw a sample conditioned on all previously sampled values $f_{\mathbf{x}}(z_i|z_1, \dots, z_{i-1})$.*
- *Return the complete multivariate distribution*
  $f_{\mathbf{x}}(\mathbf{z}) = f_{\mathbf{x}}(z_n|z_1, \dots, z_{n-1}) \cdot f_{\mathbf{x}}(z_{n-1}|z_1, \dots, z_{n-2}) \cdots f_{\mathbf{x}}(z_2|z_1) \cdot f_{\mathbf{x}}(z_1)$

The most efficient way of doing this is by computer simulation.

Notice that in order to be able to use the method sketched here, one needs to be able to fully specify conditional distributions. If normalizing constants are computationally hard to find, which often is the case, other methods should be used. For more on this, see Section 6.2.2

3 The Gaussian distribution

Most of the work in multivariate statistics has been with the Gaussian distribution. The reason for this preferance is both mathematical and due to its nature. The Central Limit Theorem, see [10], ensures that all distributions will approach the Gaussian in distribution as the number of observations increase to infinity: Mathematically, the Gaussian distribution is easy to do calculations with, and it holds many simplifying qualities, see [5] for reference.

3.1 Qualities

The Gaussian distribution can be defined through its probability density distribution as follows:

Definition 2 ***Gaussian distribution***
The stochastic vector $\boldsymbol{Z}$ ($m\times$ 1) is said to be Gaussian with parameters $\boldsymbol{\mu}$ ($m\times$ 1) and $\boldsymbol{\Sigma}$ ($m\times$ m) if the distribution is of the form

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}$$

If an m-dimensional variable $\mathbf{Z}$ is Gaussian, this is denoted $\mathbf{Z} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu} \\ Cov(\mathbf{Z}) &= \boldsymbol{\Sigma} \end{aligned}$$

The distribution is fully described for given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For partitioned vectors the following theorem applies:

Theorem 1 ***Partitioned Gaussian vectors***
If $\boldsymbol{Z}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}$$

with $\mathbf{Z}_1$ *and* $\boldsymbol{\mu_1}$ $k \times 1$ *and* $\boldsymbol{\Sigma_{11}}$ $k \times k$*, then the marginal distribution of* $\mathbf{Z_1}$ *is* $\mathbf{Z_1} \sim N_k(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_{11}})$*, and, given that they exist,*

$$\begin{aligned} E(\mathbf{Z}_1|\mathbf{Z}_2) &= \boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu_2}) \\ Cov(\mathbf{Z}_1|\mathbf{Z}_2) &= \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}} \end{aligned} \tag{2}$$

In addition the conditional distribution of $\mathbf{Z}_1|\mathbf{Z}_2$ *is k-variate Gaussian;*
$\mathbf{Z}_1|\mathbf{Z}_2 \sim N_k(E(\mathbf{Z}_1|\mathbf{Z}_2), Cov(\mathbf{Z}_1|\mathbf{Z}_2))$.

In expression (2) one observes that $Cov(\mathbf{Z}_1|\mathbf{Z}_2)$ is independent of $\mathbf{Z}_2$, meaning that the conditional covariance is independent of the data given. This is one of the aspects of the frequently used Gaussian model that has sometimes been criticized.

3.2 Gaussian random fields

A Gaussian random field is defined by, see [3]:

Definition 3 ***Gaussian random fields***
A Gaussian random field is a random field where all the finite dimensional distributions $f_{\mathbf{x}_1,\dots,\mathbf{x}_k}(z_1,\dots,z_k)$ *are Gaussian for any choice of* k *and* $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$.

The generation of Gaussian random fields is particularly easy since the explicit form of the conditional distributions needed in expression (1) are known and Gaussian.

4 Elliptic symmetric distributions

Though the Gaussian distribution has been used extensively, the assumption of normality is often unrealistic due to the fact that when one presumes an underlying Gaussian model, one implicitely assumes low variability in the data. This, again, due to the fact that the Gaussian distribution has very little probability mass in its tails. Actually so little that this is a known problem with the model's capability of explaining observed variability.

In the search for other models the elliptic symmetric distributions have been studied with good result, especially in robustness studies. For reference, see [4]. The elliptic symmetric distributions are symmetric distributions that for equal probabilty possess the same hyperelliptic shape as the Gaussian distribution, see [5]. The elliptic symmetric distributions include both long- and shorttailed distributions compared to the Gaussian, making it possible to construct a model that covers a wider range of possibilities to explain variability in data, and since the Gaussian distribution itself is an elliptic symmetric distribution, one can create a model that includes the Gaussian, thereby not excluding the existing model.

4.1 Qualities

The elliptic symmetric distributions can be defined through their probability density distributions as follows:

Definition 4 ***Elliptic symmetric distributions***
The stochastic vector $\mathbf{Z}$ *(m× 1) is said to be elliptically distributed with parameters* $\boldsymbol{\mu}$ *(m× 1) and* $\mathbf{V}$ *(m× m) if the distribution is of the form*

$$f_{\mathbf{Z}}(\mathbf{z}) = c_m(\det\mathbf{V})^{-1/2}h((\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})) \tag{3}$$

for some function $h(\cdot)$*,* c_m *a constant not dependent on* $\mathbf{z}$*, and* $\mathbf{V}$ *positive definite.*

when $\mathbf{Z}$ has some elliptic distribution this is written $\mathbf{Z} \sim E_m(\boldsymbol{\mu}, \mathbf{V})$, and

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu} \\ Cov(\mathbf{Z}) &= \alpha\mathbf{V} \end{aligned}$$

for a constant α given by the characteristic function. Notice that when writing $\mathbf{Z} \sim E_m(\boldsymbol{\mu}, \mathbf{V})$ this does not mean that $\mathbf{Z}$ has a particular elliptic distribution. In fact for every choice of $\boldsymbol{\mu}$ and $\mathbf{V}$ $E_m(\boldsymbol{\mu}, \mathbf{V})$ describes a whole family of elliptic symmetric distributions.

Notice that the Gaussian distribution is elliptic with $h(z) = e^z$ and $\alpha = 1$, implying $\boldsymbol{\Sigma} = Cov(\mathbf{Z}) = \mathbf{V}$. A reason why the elliptic distributions are worthwhile studying, is the fact that many of the tractable qualities that the Gaussian distribution posesses, also hold for all the elliptic distributions. For elliptic distributions the following theorem applies:

Theorem 2 ***Partitioned elliptic vectors***
If $\mathbf{Z}$*,* $\boldsymbol{\mu}$ *and* $\mathbf{V}$ *are partitioned as*

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \mathbf{V} = \begin{pmatrix} \mathbf{V_{11}} & \mathbf{V_{12}} \\ \mathbf{V_{21}} & \mathbf{V_{22}} \end{pmatrix}$$

with $\mathbf{Z}_1$ *and* $\boldsymbol{\mu_1}$ $k \times 1$ *and* $\mathbf{V}$ $k \times k$*, then the marginal distribution of* $\mathbf{Z}_1$ *is* $\mathbf{Z}_1 \sim E_k(\boldsymbol{\mu_1}, \mathbf{V_{11}})$*, and, given that they exist,*

$$\begin{aligned} E(\mathbf{Z}_1|\mathbf{Z}_2) &= \boldsymbol{\mu_1} + \mathbf{V_{12}}\mathbf{V_{22}}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu_2}) \\ Cov(\mathbf{Z}_1|\mathbf{Z}_2) &= \xi(\mathbf{Z}_2)(\mathbf{V_{11}} - \mathbf{V_{12}}\mathbf{V_{22}}^{-1}\mathbf{V_{21}}) \end{aligned}$$

for some function $\xi(\cdot)$*. Moreover; the conditional distribution of* $\mathbf{Z}_1|\mathbf{Z}_2$ *is* k*-variate elliptic.*

For the Gaussian distribution $\xi(\mathbf{Z}_2) = 1$. Among the elliptic symmetric distributions are both long- and shorttailed distributions compared to the Gaussian.

4.2 Elliptic symmetric random fields

Due to the fact that the marginal distributions of elliptic symmetric distributions also are elliptic symmetric distributions, the definition of an elliptic random field is a direct generalization of that for the Gaussian;

Definition 5 ***Elliptic symmetric random fields***
An elliptic symmetric random field is a random field where all the finite dimensional distributions, $f_{\mathbf{x}_1,\dots,\mathbf{x}_k}(z_1,\dots,z_k)$, are elliptic symmetric distributions for any choice of k and $\{\mathbf{x}_1,\dots,\mathbf{x}_k\}$.

5 The multivariate *t*-distribution

In the proceeding section it was stated that many of the tractable qualities of the Gaussian distribution also hold for the elliptic symmetric distributions. One of the reasons for the investigation of these ditributions is the fact that, in addition to these qualities, one knows that this class contains long-tailed distributions compared to the Gaussian, thus implying greater variability.

The class of elliptic symmetric distributions is a rather large one, and it is not too difficult to generate new elliptic symmetric distributions either, see [6], so the question of which one of these distributions to choose is at its place. A good choice would be a distribution that is both in the class and already familiar, and such a distribution exist, namely the multivariate t-distribution. The advatage with this distribution is that it includes the Gaussian as a special case, but generally has heavier tails. The price to pay is that one has to estimate another parameter, ν, the number of degrees of freedom.

The multivariate t-distribution is a generalization of the Student t-distribut-ion in one dimension, and dates back to Cornish, see [7]. There are in fact several possible generalizations, see [8], chapter 37. Here the most common one will be used.

5.1 Qualities

The multivariate t-distribution can be defined through its probability density distribution as follows:

Definition 6 ***Multivariate t-distribution***
The stochastic vector $\boldsymbol{Z}$ ($m\times$ 1) is said to be multivariate t-distributed with parameters $\boldsymbol{\mu}$ ($m\times$ 1), $\boldsymbol{V}$ ($m\times$ m) and ν, the number of degrees of freedom, if the distribution is of the form

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{\Gamma[\frac{1}{2}(\nu+m)]}{\Gamma(\frac{1}{2}\nu)(\nu\pi)^{m/2}}(\det\mathbf{V})^{-1/2}[1+\frac{1}{\nu}(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})]^{-(\nu+m)/2} \tag{4}$$

This expression is valid for all $\nu > 0$. If a variable is multivariate t-distributed one writes $\mathbf{Z} \sim T_m(\boldsymbol{\mu}, \mathbf{V}, \nu)$ and

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu} \quad , \nu \geq 2 \\ Cov(\mathbf{Z}) &= \frac{\nu}{\nu - 2}\mathbf{V} \ , \ \nu \geq 3 \end{aligned} \tag{5}$$

For proofs, see [7]. Notice that the first and second order moments only exist for particular values of ν. For a partitioned vector $\mathbf{Z}$ the following theorem applies:

Theorem 3 ***Partitioned t-distributed vectors***
If $\boldsymbol{Z}$, $\boldsymbol{\mu}$ and $\boldsymbol{V}$ are partitioned as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \mu = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \mathbf{V} = \begin{pmatrix} \mathbf{V_{11}} & \mathbf{V_{12}} \\ \mathbf{V_{21}} & \mathbf{V_{22}} \end{pmatrix}$$

with $\mathbf{Z}_1$ and μ_1 $k \times 1$ and V $k \times k$, then the marginal distribution of $\mathbf{Z}_1$ is $\mathbf{Z}_1 \sim T_k(\boldsymbol{\mu_1}, \mathbf{V_{11}}, \nu)$, and, given that they exist,

$$E(\mathbf{Z}_1|\mathbf{Z}_2) = \mu_1 + \mathbf{V_{12}}\mathbf{V_{22}}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu_2}) \quad , \nu \geq 2 \tag{6}$$

$$Cov(\mathbf{Z}_1|\mathbf{Z}_2) = \frac{\nu + (\mathbf{Z}_2 - \boldsymbol{\mu_2})^T\mathbf{V_{22}}^{-1}(\mathbf{Z}_2 - \boldsymbol{\mu_2})}{\nu + m - (k+2)} \tag{7}$$

$$\cdot\, (\mathbf{V_{11}} - \mathbf{V_{12}}\mathbf{V_{22}}^{-1}\mathbf{V_{21}}) \quad , \nu \geq 3 \tag{8}$$

Moreover; the conditional distribution of $\mathbf{Z}_1|\mathbf{Z}_2$ is k-variate elliptic, but not k-variate t-distributed.

In the limit $\nu \to \infty$ the t-distribution converges to the Gaussian in distribution, and for $\nu = 1$ it coinsides with the Cauchy distribution. For $0 < \nu < 2$ neither first nor second order moments exist. For comparison of the univariate t-distribution with its special cases, see Figure 1.

5.2 Random t-distributed fields

Since the multivariate t-distribution is an elliptic symmetric distribution and all marginals are t-distributed the definition of a random t-distributed field reads:

Definition 7 ***Random t-distributed fields***
A random t-distributed field is a random field where all the finite dimensional distributions, $f_{\mathbf{x}_1,...,\mathbf{x}_k}(z_1, \cdots, z_k)$, are multivariate t-distributed for any choice of k and $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$.

Section 6 is devoted to the generation of random t-distributed fields.

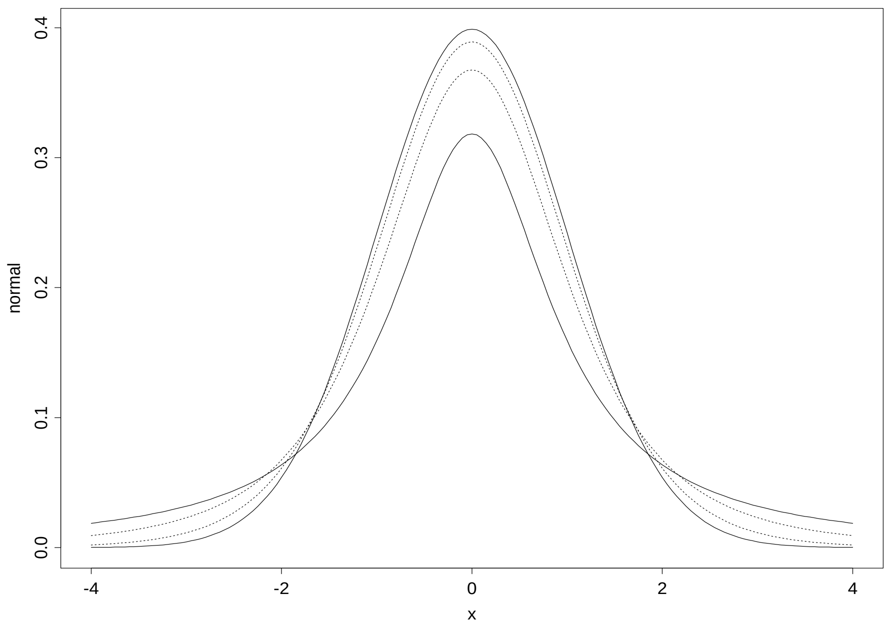


Figure 1: *In descending order; Gaussian distribution, t-distribution with 10 and 3 degrees of freedeom (dotted lines) and Cauchy distribution*

6 Algorithm for the simulation of random t-distributed fields

Algorithm 1 shows how random fields can be generated. This, combined with the results from the previous section, makes it possible to set up an algorithm for the simulation of random t-distributed sfields. In the case of the t-distribution, however, there are a few more obstacles than with the Gaussian distribution. In this section both the general elements of generating a random field by (1) and the specific problems for the t-distribution will be treated. As mentioned in Section 2.3 the usual notation $f_{\mathbf{Z}}(\mathbf{z})$ will here be abandoned for the random fields description $f_{\mathbf{x}}(\mathbf{z})$.

6.1 Walking through the grid

It is of interest to generate realizations of random t-distributed fields $\{Z(\mathbf{x});$ $\mathbf{x} \in \mathcal{D}\}, \mathcal{D} \subset \mathbf{R}^2$. This can be done with the method mentioned in Section 2.3, i.e., by generating values z on a two-dimensional grid $\mathcal{L}_{\mathcal{D}}$. The values $Z(\mathbf{x})$ are not independent, and the interaction between them could for example be expressed as a function of their distances in $\mathbf{R}^2$.

The way in which one walks through the grid $\mathcal{L}_{\mathcal{D}}$, i.e., the order in which the values in the grid $\mathcal{L}_{\mathcal{D}}$ are calculated, is not insignificant. The problems that may arise are not statistical but numerical, and is due to the effect of handling unbalanced matrices. One way to avoid such problems is by spanning out some points far from one another to stretch out the field and then refill in a recursive way. Another alternative is to walk randomly, but add a test to avoid clustering.

6.2 Sampling procedure

In this subsection the general framework for sampling the values in each of the grid nodes of $\mathcal{L}_{\mathcal{D}}$ will be discussed.

6.2.1 Conditional distribution

By applying Agorithm 1 as the simulation algorithm it is evident that the conditional elliptical distribution is needed. This can be found be recognizing that for an m-dimensional vector $\mathbf{Z}$ partitioned as $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)^T$ with $\mathbf{Z}_1$ being $k \times 1$, the conditional distritubion of $\mathbf{Z}_1|\mathbf{Z}_2$ is always given as

$$f_{\mathbf{x}_1,\dots,\mathbf{x}_k|\mathbf{x}_{k+1},\dots,\mathbf{x}_m}(\mathbf{z}_1|\mathbf{z}_2) = \frac{f_{\mathbf{x}_1,\dots,\mathbf{x}_m}(\mathbf{z})}{f_{\mathbf{x}_{k+1},\dots,\mathbf{x}_m}(\mathbf{z_2})}$$

with $f_{\mathbf{x}_1,\dots,\mathbf{x}_m}(\mathbf{z}) = f_{\mathbf{x}_1,\dots,\mathbf{x}_k,\mathbf{x}_{k+1},\dots,\mathbf{x}_m}(\mathbf{z}_1, \mathbf{z}_2)$ and $f_{\mathbf{x}_{k+1},\dots,\mathbf{x}_m}(\mathbf{z_2})$ the margi-nal distribution of the given data. Having decided on using a sequential algorithm means that in every step i a new value from a one-dimensional distribution

$$f_{\mathbf{x}_i|\mathbf{x}_s}(z_i|\mathbf{z}_s) = \frac{f_{\mathbf{x}_i,\mathbf{x_s}}(z_i, \mathbf{z_s})}{f_{\mathbf{x_s}}(\mathbf{z_s})} \tag{9}$$

must be generated. In (9) $\mathbf{z_s} = (z_1, \dots, z_{i-1})^T$ denotes the vector with the already sampled values and $\mathbf{x_s} = (\mathbf{x_1}, \dots, \mathbf{x_{i-1}})^T$ the coordinates associated with these values. All marginals are multivariate t-distributed, so both numinator and denuminator are multivariate t-distributed, but the conditional distribution is not, as stated in Section 5.1. By taking (4), putting it into (9) and canselling terms one gets

$$\begin{aligned} f_{\mathbf{x}_i|\mathbf{x}_s}(z_i|\mathbf{z}_s) &= \frac{\Gamma(\frac{1}{2}(\nu+i))}{\Gamma(\frac{1}{2}(\nu+i-1))} \cdot \frac{1}{(\nu\pi)^{1/2}} \cdot \frac{1}{|v_{ii} - \mathbf{V_{is}}\mathbf{V_{ss}^{-1}}\mathbf{V_{si}}|^{1/2}} && (10) \\ &\cdot \frac{[1 + \frac{1}{\nu}(\mathbf{z_s} - \boldsymbol{\mu_s})^{\mathrm{T}}\mathbf{V_{ss}^{-1}}(\mathbf{z_s} - \boldsymbol{\mu_s})]^{(\nu+m-1)/2}}{[1 + \frac{1}{\nu}(\mathbf{z} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{V^{-1}}(\mathbf{z} - \boldsymbol{\mu})]^{(\nu+m)/2}} && (11) \end{aligned}$$

where $\mathbf{Z}$, μ and $\mathbf{V}$ have been partitioned as

$$\mathbf{Z} = \begin{pmatrix} z_i \\ \mathbf{z}_s \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_i \\ \boldsymbol{\mu_s} \end{pmatrix}, \mathbf{V} = \begin{pmatrix} v_{ii} & \mathbf{V_{is}} \\ \mathbf{V_{si}} & \mathbf{V_{ss}} \end{pmatrix}$$

with subscript s denoting the already sampled values. $\mathbf{Z}$ and $\boldsymbol{\mu}$ are $m \times 1$ and $\mathbf{V}$ is $m \times m$. Expressions (10) and (11) are valid for all $\nu > 0$, and this is the distribution values should be sampled from, except the first value, which is sampled from an unconditional t-distribution.

¿From expressions (10) and (11) it is evident that it is not $\mathbf{V}$ that is needed, but rather $\mathbf{V^{-1}}$. By looking at the results given in Appendix A, one finds that the different elements in

$$\mathbf{V^{-1}} = \begin{pmatrix} \mathbf{V^{11}} & \mathbf{V^{12}} \\ \mathbf{V^{21}} & \mathbf{V^{22}} \end{pmatrix} \tag{12}$$

can be identified as

$$\begin{aligned}
\mathbf{V^{11}} &= \left(v_{ii} - \mathbf{V_{is}}\mathbf{V_{ss}}^{-1}\mathbf{V_{si}}\right)^{-1} \\
\mathbf{V^{22}} &= \mathbf{V_{ss}^{-1}} + \mathbf{V_{ss}^{-1}}\mathbf{V_{si}}\mathbf{V^{11}}\mathbf{V_{is}}\mathbf{V_{ss}^{-1}} \\
\mathbf{V^{12}} &= -\frac{1}{v_{ii}}\mathbf{V_{is}}\mathbf{V^{22}} \\
\mathbf{V^{21}} &= -\mathbf{V_{ss}^{-1}}\mathbf{V_{si}}\mathbf{V^{11}}
\end{aligned}$$

One then observes the fact that in order to spesify $\mathbf{V^{-1}}$ in iteration step i one only has to combine already known quantities and thereby eliminate the inversion process completely, since $\mathbf{V_{ss}^{-1}}$ already is known from the previous iteration, $i-1$.

6.2.2 Sampling from the conditional distribution

There are several possible methods one can apply to sample from a multivariate density distribution $f_{\mathbf{x}}(\mathbf{z})$. If one wants to sample from the complete multivariate distribution, only iterative methods, like Markov chain Monte Carlo simulation, can be used, but if one makes use of Algorithm (1) both iterative and direct approach methods can be applied in each iteration step i. Iterative MCMC algorithms are the most frequently used, and the general principles, see [16], are outlined in Appendix B.1. MCMC algorithms can always be applied, and are quite easy to use. The main argument against these methods, however, is the number of iterations sometimes needed before one can be sure of convergence.

The direct approach methods, like invertion and rejection sampling, see [11], do rarely work for practical purposes, since, as mentioned in Section 2.3, normalizing constants are needed, and these may be computationally hard or analytically impossible to find. With MCMC algorithms one does not have to know normalizing constants. In the case with the multivariate t-distribution, however, rejection sampling has proved to work quite well together with the direct approach of Algorithm 1. The reason for the success is the number of values one has to sample in each iteration step i. Whereas one in an MCMC algorithm needs at least 100 values to be somewhat confident of convergence, the rejection sampling procedure has, in this particular case, been able to manage with 4-5, and for this reason rejection sampling will be used here, see Appendix B.2 for details.

When applying rejection sampling to sample from the one-dimensional distribution in expressions (10) and (11) an obvious choice for the proposal distribution is the Gaussian since it is easy to sample from and it too is elliptic symmetric. The rejection sampling procedure, however, includes the number

$$M \geq \max_x \{\mathcal{M}(x)\} = \max_x \left\{\frac{f(x)}{g(x)}\right\},$$

with $\mathcal{M}(x) = \frac{f(x)}{g(x)}$, $f(x)$ the distribution to sample from and $g(x)$ the proposal distribution. When the proposal density has lighter tails, i.e., approaches zero faster as $x \to \pm\infty$, than the sample distribution, this number will approach infinity. This implies that the Gaussian distribution cannot be used as proposal. But for $\nu = 1$ the t-distribution coinsides with the Cauchy distribution which has heavier tails than a t-distribution with $\nu > 1$ and same scale, so this would be a natural choice for our proposal. For the case $0 < \nu < 1$ there is no wellknown distribution to sample from with the same bell-like shape, so in this case a uniform distribution on the sample space $\Omega = \mathbf{R^1}$ is one of few simple choices available. In order for it to be a valid density distribution, however, the uniform distribution must be truncated, see Section 6.4.3 for more.

The proposal distribution must be centered at the same place as the sample distribution, and for symmetric unimodal distributions like the elliptic, this is the mode. For $f_{\mathbf{x}_i|\mathbf{x}_s}(z_i|\mathbf{z}_s)$ the notation η_i for the mode and τ_i for the scale will be used. If $f(\cdot)$ is elliptic and $g(\cdot)$ is Cauchy, and $f(\cdot)$ and $g(\cdot)$ have the same scale, τ_i, then $\max_x \mathcal{M}(x) = \mathcal{M}(\eta_i)$ In order to sample values from $f_{\mathbf{x}_i|\mathbf{x}_s}(z_i|\mathbf{z}_s)$ one may now in every step i sample values from $Z \sim \text{Cauchy}(\eta_i, \tau_i)$ in a rejection sampling procedure instead.

6.3 Field parameters

Subsection 6.2 showed how to exploit the sequential Algorithm 1 to generate random t-distributed fields. It is then time to look at the parameters in the field, $\boldsymbol{\mu}$, $\mathbf{V}$ and ν, and what model assumptions to make. For estimation of these parameters from a given field $f_{\mathbf{x}}(\mathbf{z})$, see Section 7.

6.3.1 Location

It will here be assumed that the field is located around a fixed level being equal for the whole area $\mathcal{D} \subset \mathbf{R^2}$, i.e., $\boldsymbol{\mu} = \mathbf{1}\mu$ with μ a constant.

6.3.2 Interaction

When dealing with spatial data it is natural to assume some spatial interaction between different coordinates. Points close to one another will be assumed to have greater impact on one another than points further away. The distance between two points p and q will be denoted $\Delta_{pq} = |\mathbf{x_p} - \mathbf{x_q}|$ and interactional effects are modelled as exponential:

$$\rho_{exp}(\Delta_{pq}; R, \nu) = e^{-3(\frac{\Delta_{pq}}{R})^{\beta}} \tag{13}$$

with R describing the range of the interactional effect and β a parameter that can take the values $0 < \beta \leq 2$. One observess that $\rho_{exp}(0) = 1$, so it is desireable to adjust $\rho_{exp}(\Delta_{pq}; R, \nu)$ with a factor $c\sigma^2$ describing the scale of the univariate observations. The different elements of $\mathbf{V}$ can now be identified as

$$\begin{aligned}\mathbf{V}(p,q) &= c\sigma^2 \cdot \rho_{exp}(\Delta_{pq}; R, \beta) \\ &= c\sigma^2 \cdot e^{-3(\frac{\Delta_{pq}}{R})^\beta} \end{aligned} \qquad (14)$$

since the choice of interactional effects, ρ_{exp}, guarantees the demand of positive definiteness of $\mathbf{V}$. Notice that $\mathbf{V}(p,p) = c\sigma^2$.

6.3.3 Degrees of freedom

The last parameter, ν, adjusts the tail behaviour of the t-distribution. Fields with any number of degrees of freedom $\nu > 0$ may be generated, but bare in mind the direct relationship between the existence of moments and the number of degrees of freedom:

1. $3 \leq \nu < \infty$: Both first and second order moments exist.
2. $2 \leq \nu < 3$: First but not second order moment exists.
3. $0 < \nu < 2$: Neither first nor second order moments exist.

In the next subsection it is shown how this affects the sampling procedure.

6.4 Sampling differences for different values of ν

So far in this section the general framework for generating random t-distributed fields has been studied. From Section 5.1, however, it is known that the number of degrees of freedom controls some analytical behaviour of the t-distribution. This was summarized in the previous subsection, and it will now be shown what this means in the actual simulation process.

6.4.1 $3 \leq \nu < \infty$

In Section 6.2 it was stated that the mode η_i and the scale τ_i were needed to be able to use the rejection sampling procedure effectively, and for $\nu \geq 3$ these expressions are given explicitely in (6), (7) and (8) with $k = 1$;

$$\begin{aligned}\eta_i &\equiv E(Z_i|z_1, \ldots, z_{i-1}) = E(Z_i|\mathbf{z}_s) \\ &= \mu + \mathbf{V_{is}}\mathbf{V_{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu}_s) \end{aligned} \qquad (15)$$

$$\begin{aligned}\tau_i &\equiv Cov(Z_i|z_1, \ldots, z_{i-1}) = Cov(Z_i|\mathbf{z}_s) \\ &= \frac{\nu + (\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V_{ss}^{-1}} (\mathbf{z}_s - \boldsymbol{\mu})}{\nu + m - 3} \left(v_{ii} - \mathbf{V_{is}}\mathbf{V_{ss}^{-1}}\mathbf{V_{si}}\right) \end{aligned} \qquad (16)$$

with $\mathbf{V}_{si} = c\sigma^2[\rho_{exp}(\Delta_{i,i-1}), \ldots, \rho_{exp}(\Delta_{i,1})]^T$ and $v_{ii} = c\sigma^2$. Remember from expression (5) that there is a relation between $\mathbf{V}$ and $\mathbf{Z}$, and with the results in section 6.3.2 one finds;

$$\begin{aligned} \mathbf{V} &= \frac{\nu - 2}{\nu} \cdot Cov(\mathbf{Z}) \\ \mathbf{V}(p,q) &= c \cdot \sigma^2 \rho_{exp}(\Delta_{pq}; R, \nu) \end{aligned}$$

with $\Delta_{pq} = |\mathbf{x_p} - \mathbf{x_q}|$. So there is a stationary, isotropic covariance function invoked in $\mathbf{V}$ with $c = \frac{\nu-2}{\nu}$ and σ^2/c being the variance of the marginal univariate t-distributions.

6.4.2 $2 \leq \nu < 3$

With degrees of freedom less than 3, the second order moment seizes to exist, and since the conditional scale depends on the already sampled values, one cannot in general assume that $\tau_i \leq \sigma/\sqrt{c}$, i.e., that the conditional scale is less than or equal to the unconditional, see Appendix C. One may assume, however, that τ_i will not be too far away from $\sigma/\sqrt{c}$. This will be correct unless very extreme values are sampled early in the process. So by setting $\tau_i = \sigma/\sqrt{c}$ for all i, in most cases the scale parameter will be overestimated, which causes no problem since in the cases that the actual τ_i is less than $\sigma/\sqrt{c}$, $\mathcal{M}(x)$ will have its maximum in the mode, and this peak will generally be larger than if one had the exact τ_i. If τ_i is really larger than $\sigma/\sqrt{c}$, $\mathcal{M}(x)$ will not have its maximum value in the mode of $f(x)$, but be bimodal symmetric about it. In this case one can apply a numerical method to locate the maxima of $\mathcal{M}(x)$, and then find the specific value M. For details on a numerical method to find a local maxima of a function, see Appendix D.1. As far as η_i is concerned, this can still be found analytically through expression (15).

6.4.3 $0 < \nu < 2$

When the number of degrees of freedom is below 2, neither first nor second order moments exist. The conditional scale τ_i can be set to $\sigma/\sqrt{c}$ similarly as for $2 \leq \nu < 3$. Finding the location η_i is an easy numerical task, since the conditional t-distribution is centered in the mode, and the problem is therefore reduced to finding the maxima of a symmetric unimodal function $f(x)$. This can be done with the algorithm of Appendix D.1.

Another aspect to consider, is the distinction between $0 < \nu < 1$ and $1 \leq \nu < 2$. In the latter case rejection sampling with a Cauchy distribution as proposal can be applied, and in the first a uniform distribution over the sample space $\Omega = \mathbf{R^1}$ can be used. But in order for the uniform distribution to be a valid denstity distribution it has to be truncated. The question is where. For the Gaussain distribution 99.872% of the sample space is covered by truncating it at $x = \mu \pm 3\sigma$. For comparison let us consider the Cauchy distribution, i.e., a t-distribution with $\nu = 1$. In order to get the same coverage of the sample space, 99.872%, one needs $x = \mu \pm 497.4\sigma$. With $\nu < 1$ this number will grow drastically, and for the uniform distribution even more. Therefore the generation of random t-distributed fields with $0 < \nu < 1$ will not be considered here.

6.5 Algorithm for the generation of random t-distributed fields

It is of interest to generate a random t-distributed field $T_n(\boldsymbol{\mu}, \mathbf{V}, \nu)$ on an area $\mathcal{D} \subset \mathbf{R}^2$ by discretisizing this area $\mathcal{D}$ into a lattice $\mathcal{L}_{\mathcal{D}}$ as suggested in Section 2.3. The n points $\{x_{pq}; p, q \in \mathcal{L}_{\mathcal{D}}\}$ are then visited in some order, in each step applying the sampling procedure of Section 6.2 to generate the new value $Z(\mathbf{x}_i)$ in this point.

The procedure for sampling from the t-distribution $f(\cdot)$ can now schematically be given as a sequential rejection sampling algorithm with the proposal distribution $g(\cdot)$ being the Cauchy distribution. It is assumed that $\nu \geq 1$:

Algorithm 2 ***Sequential algorithm for unconditional sampling of random t-distributed fields***

- *Initiate: Sample a value from the one-dimensional unconditional Student t-distribution with $\eta_1 = \mu$.*
- *Iterate: For i from 2 to n do*
  1. *Calculate the interaction between Z_i and $\mathbf{z_s}$ by (14).*
  2. *Calculate η_i from expression (15) if $\nu \geq 2$, else find it numerically.*
  3. *Calculate τ_i from expression (16) if $\nu \geq 3$, else $\tau_i = \sigma/\sqrt{c}$.*
  4. *Calculate $\mathbf{V}^{-1}$ from expression (12).*
  5. *Calculate $M = \frac{f(\eta_i)}{g(\eta_i)}$ with $f(\cdot)$ given in expressions (10) and (11), and $g(\cdot)$ being Cauchy(η_i, τ_c), if $\nu \geq 3$. If $1 \leq \nu < 3$ find M numerically.*
  6. *Repeat*
     (a) *Sample $y \sim$ Cauchy(η_i, τ_c)*
     (b) *Sample $u \sim$ Uniform[0,1].*
     (c) *If $u \leq \frac{f(y)}{M \cdot g(y)}$ then $z_i = y$*
     until new value z_i is found

6.6 Conditional simulation

It may often be of interest to simulate some conditional random fields, i.e., the generated random fields are to be conditioned on some, say k, observations. These k values will then be known values $z_j = z(\mathbf{x}_j)$, and they can be included in an algorithm by simply putting them first into the table of calculated values. The procedure for sampling from the t-distribution $f(\cdot)$ conditioned on k known values can then schematically be given as a sequential rejection sampling algorithm with the proposal distribution $g(\cdot)$ being the Cauchy distribution. It is assumed that $\nu \geq 1$:

Algorithm 3 ***Sequential algorithm for conditional sampling of random t-distributed fields***

- *Initiate: If there are k values $\mathbf{z}_{known}$ to condition the field on, set $\mathbf{z_s} := \mathbf{z}_{known}$.*
- *Iterate: For i from $k+1$ to n do*
  1. *Calculate the interaction between Z_i and $\mathbf{z_s}$ by (14).*
  2. *Calculate η_i from expression (15) if $\nu \geq 2$, else find it numerically.*
  3. *Calculate τ_i from expressions (16) if $\nu \geq 3$, else $\tau_i = \sigma/\sqrt{c}$.*
  4. *Calculate $\mathbf{V}^{-1}$ from expression (12).*
  5. *Calculate $M = \frac{f(\eta_i)}{g(\eta_i)}$ with $f(\cdot)$ given in expressions (10) and (11), and $g(\cdot)$ being Cauchy(η_i, τ_c), if $\nu \geq 3$. If $1 \leq \nu < 3$ find M numerically.*
  6. *Repeat*
     (a) *Sample $y \sim$ Cauchy(η_i, τ_c)*
     (b) *Sample $u \sim$ Uniform[0,1].*
     (c) *If $u \leq \frac{f(y)}{M \cdot g(y)}$ then $z_i = y$*

     until new value z_i is found

7 Parameter estimation

In addition to creating random t-distributed fields with certain qualities, it is also of interest to estimate what kind of random field a given set of samples is. If this random field is assumed to be t-distributed, the parameters $\boldsymbol{\mu}$, $\mathbf{V}$ and ν need to be estimated. A standard method that can be applied is to calculate the maximum likelihood estimators, or the ML-estimators for short.

For $\boldsymbol{\mu}$ we the model assumption $\boldsymbol{\mu} = \mathbf{1}\mu$ is made, as mentioned in Ssection 6.3.1, for $\mathbf{V}$ interactional effects are assumed to be as described in Section 6.3.2, and ν must be limited to strictly positive values.

7.1 The loglikelihood

The likelihoodfunction is

$$L(\mu, \mathbf{V}, \nu|\mathbf{z}) = \underbrace{\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{n/2}}}_{C(\nu,n)} \frac{1}{|V|^{1/2}} \frac{1}{[1+\frac{1}{\nu}\underbrace{(\mathbf{z}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})}_{D}]^{\frac{\nu+n}{2}}}$$

when assuming only one realization of the field, and $D = (\mathbf{z}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})$ the squared Mahalanobis distance. The loglikelihood, $\ln L(\mu, \mathbf{V}, \nu|\mathbf{z})$, is then

$$l(\boldsymbol{\mu}, \mathbf{V}, \nu|\mathbf{z}) = \ln C(\nu, n) - \frac{1}{2}\ln|\mathbf{V}| - \frac{\nu+n}{2}\ln\left(1+\frac{1}{\nu}D\right) \tag{17}$$

7.2 Location parameter μ

One can optimize the loglikelihood (17) for μ by differentiating:

$$\begin{aligned}
\frac{\partial l}{\partial \mu} &= -\frac{\nu+n}{2} \cdot \frac{\partial}{\partial \mu} \ln \left[1 + \frac{1}{\nu} D\right] \\
&= -\frac{\nu+n}{2} \cdot \frac{\nu}{\nu+D} \cdot \frac{1}{\nu} \cdot \frac{\partial D}{\partial \mu} \\
&= -\frac{\nu+n}{\nu+D} \cdot \frac{1}{2} \cdot \left[2\mu \mathbf{1}^T \mathbf{V}^{-1} \mathbf{1} - (\mathbf{z}^T \mathbf{V}^{-1} \mathbf{1} + \mathbf{1}^T \mathbf{V}^{-1} \mathbf{z})\right] \qquad (18)
\end{aligned}$$

where the last equality is due to the fact that

$$\begin{aligned}
D &= (\mathbf{z} - \mathbf{1}\mu)^T \mathbf{V}^{-1} (\mathbf{z} - \mathbf{1}\mu) \\
&= tr(\mathbf{V}^{-1} \mathbf{z}\mathbf{z}^T) - \mu[\mathbf{z}^T \mathbf{V}^{-1} \mathbf{1} + \mathbf{1}^T \mathbf{V}^{-1} \mathbf{z}] + \mu^2 \mathbf{1}^T \ \mathbf{V}^{-1} \mathbf{1}
\end{aligned}$$

and therefore

$$\frac{\partial D}{\partial \mu} = 2\mu \mathbf{1}^T \ \mathbf{V}^{-1} \mathbf{1} - \mathbf{z}^T \mathbf{V}^{-1} \mathbf{1} - \mathbf{1}^T \mathbf{V}^{-1} \mathbf{z}$$

Set expression (18) equal to zero to find the maxima, and get

$$\begin{aligned}
\hat{\mu} &= \left(\mathbf{z}^T \mathbf{V}^{-1} \mathbf{1} + \mathbf{1}^T \mathbf{V}^{-1} \mathbf{z}\right) \left(2 \mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}\right)^{-1} \\
&= \frac{1}{2} \left(\mathbf{z}^T \mathbf{V}^{-1} \mathbf{1} + (\mathbf{z}^T \mathbf{V}^{-1} \mathbf{1})^T\right) \left(\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}\right)^{-1} \\
&= \mathbf{z}^T \mathbf{V}^{-1} \mathbf{1} \left(\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}\right)^{-1} \\
&= \bar{\mathbf{z}}
\end{aligned}$$

i.e., $\hat{\mu}_{ML} = \bar{\mathbf{z}}$ for all $\nu > 0$.

7.3 Matrix parameter V

If no modelling conditions are imposed on **V**, one may find the ML-estimator as in Appendix E.1. In the case here this will not be the correct approach, since in the generation of new fields, interaction between different values are being modelled as exponential. By this, the different elements of **V** can be identified as

$$\mathbf{V}(p, q) = c\sigma^2 \cdot e^{-3(\frac{\Delta_{pq}}{R})^\beta} \qquad (19)$$

as stated in Section 6.3.2. So when estimating **V**, it is rather σ^2, R and β that are wanted.

For simplicity set $\mathbf{W}(p, q) = \rho_{exp}(\Delta_{pq}; R, \beta) = e^{-3(\frac{\Delta_{pq}}{R})^\beta}$ and from expression (19) $\mathbf{V}(p, q)$ $= c\sigma^2 \mathbf{W}(p, q)$ and $\mathbf{V} = c\sigma^2 \mathbf{W}$. The squared Mahalanobis distance can then be rewritten as

$$\begin{aligned}
D &= (\mathbf{z} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \\
&= \frac{1}{c\sigma^2} \underbrace{(\mathbf{z} - \boldsymbol{\mu})^T \mathbf{W}^{-1} (\mathbf{z} - \boldsymbol{\mu})}_{D_W} \\
&= \frac{1}{c\sigma^2} D_W
\end{aligned}$$

and the loglikelihood is

$$l(\boldsymbol{\mu},\sigma,\mathbf{W},\nu|\mathbf{z}) = \ln C(\nu,n) - \frac{1}{2}\ln|c\sigma^2\mathbf{W}| - \frac{\nu+n}{2}\ln\left(1+\frac{1}{\nu c\sigma^2}D_W\right) \tag{20}$$

7.3.1 Scale parameter σ

Differentiate the loglikelihood in expression (20) with respect to σ;

$$\begin{aligned}\frac{\partial l}{\partial\sigma} &= -\frac{n}{2}\cdot\frac{1}{c\sigma^2}\cdot c2\sigma - \frac{\nu+n}{2}\cdot\frac{1}{1+\frac{D_W}{\nu c\sigma^2}}\cdot\frac{D_W}{\nu c}\cdot(-2)\sigma^{-3}\\ &= -\frac{n}{\sigma}+(\nu+n)\frac{D_W}{\nu c\sigma^3+D_W\sigma}\end{aligned}$$

Setting this equal to zero yields

$$\hat{\sigma}^2_{ML} = \frac{1}{c}\cdot\frac{D_W}{n}$$

As n increases D_W will increase as well, but the fraction $\frac{D_W}{n}$ will be of the same magnitude.

Remember that $c = c(\nu)$ and that D_W is dependent on $\mathbf{W} = \mathbf{W}(R,\beta)$, so that in order to find the ML-estimator of σ one needs ν and $\mathbf{W}$.

7.3.2 Interactional parameters

The next task is to find the the parameters $r = 1/R$ and β of the matrix $\mathbf{W}$ with the form

$$\mathbf{W} = \begin{bmatrix} 1 & e^{-3(\Delta_{12}r)^\beta} & \cdots & e^{-3(\Delta_{1n}r)^\beta}\\ e^{-3(\Delta_{21}r)^\beta} & 1 & \cdots & e^{-3(\Delta_{2n}r)^\beta}\\ \vdots & & \ddots & \vdots\\ e^{-3(\Delta_{n1}r)^\beta} & e^{-3(\Delta_{n2}r)^\beta} & \cdots & 1\end{bmatrix}$$

If one already had an estimat for the unmodelled $\mathbf{W}$, or $\mathbf{V}$, a numerical least squares approach like the one outlined in Appendix E.3 could be used.

The ML-estimators of R and β are calculated in Appendix E.2, but those results cannot be used for practical purposes. An algorithm taking use of the ML-estimators of r and β would run in at least $\mathcal{O}(n^4)$ time, and that is before taking into account that in each of those steps the calculation of a determinant of a $(n-1)\times(n-1)$ matrix is needed.

To get an algorithm for the estimation of $\mathbf{W}$ that works, it is necessary to try other methods than ML-estimation. To make the task a bit easier, it is also wise to reduce the complexity of the interactional form. Therefore replace the matrix $\mathbf{W}$ with

$$\mathbf{W}_\theta = \begin{bmatrix} 1 & e^{-\theta\Delta_{12}} & \cdots & e^{-\theta\Delta_{1n}}\\ e^{-\theta\Delta_{21}} & 1 & \cdots & e^{-\theta\Delta_{2n}}\\ \vdots & & \ddots & \vdots\\ e^{-\theta\Delta_{n1}} & e^{-\theta\Delta_{n2}} & \cdots & 1\end{bmatrix}$$

i.e., set $\beta = 1$ and $\theta = 3/R$ and thereby limiting the number of possible interactional functions slightly.

A method that has prooved to be efficient is quasilikelihood, see [14]. In quasilikelihood one looks at all $N = \frac{n(n-1)}{2}$ pairs of observations and assumes that these are independent. This yields the quasilikelihood

$$ql(\mu, \sigma^2, \nu, \theta) = \prod_{i<j} C(\nu, n) \cdot \frac{1}{|c\sigma^2 \mathbf{W}_\theta|^{1/2}} \cdot [1 + \frac{1}{\nu c\sigma^2}(\mathbf{z} - \boldsymbol{\mu})^T \mathbf{W}_\theta^{-1}(\mathbf{z} - \boldsymbol{\mu})]^{-\frac{\nu+n}{2}}$$

Because of the pairwise consideration $n = 2$. By this $\mathbf{W}$ is 2×2 and so on. Interactional effects arre being modelled as $\sigma^2 e^{-\theta\Delta_{ij}}$, so

$$\begin{aligned}
|c\sigma^2 \mathbf{W}_\theta| &= \left| c\sigma^2 \begin{bmatrix} 1 & e^{-\theta\Delta_{ij}} \\ e^{-\theta\Delta_{ij}} & 1 \end{bmatrix} \right| \\
&= c^2\sigma^4 \underbrace{\left(1 - e^{-2\theta\Delta_{ij}}\right)}_{R(\theta)} \\
&= c^2\sigma^4 \, R(\theta)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{W}_\theta^{-1} &= \begin{bmatrix} 1 & e^{-\theta\Delta_{ij}} \\ e^{-\theta\Delta_{ij}} & 1 \end{bmatrix}^{-1} \\
&= \frac{1}{1 - e^{-2\theta\Delta_{ij}}} \begin{bmatrix} 1 & -e^{-\theta\Delta_{ij}} \\ -e^{-\theta\Delta_{ij}} & 1 \end{bmatrix} \\
&= \frac{1}{R(\theta)} \begin{bmatrix} 1 & -e^{-\theta\Delta_{ij}} \\ -e^{-\theta\Delta_{ij}} & 1 \end{bmatrix}
\end{aligned}$$

so that

$$\begin{aligned}
& (\mathbf{z} - \boldsymbol{\mu})^T \mathbf{W}_\theta^{-1} (\mathbf{z} - \boldsymbol{\mu}) \\
= \; & [z_1 - \mu, z_2 - \mu] \frac{1}{R(\theta)} \begin{bmatrix} 1 & -e^{-\theta\Delta_{ij}} \\ -e^{-\theta\Delta_{ij}} & 1 \end{bmatrix} \begin{bmatrix} z_1 - \mu \\ z_2 - \mu \end{bmatrix} \\
= \; & \frac{1}{R(\theta)} \underbrace{[(z_1 - \mu)^2 + (z_2 - \mu)^2 - 2(z_1 - \mu)(z_2 - \mu)e^{-\theta\Delta_{ij}}]}_{Q(\mu,\theta)} \\
= \; & \frac{Q(\mu, \theta)}{R(\theta)}
\end{aligned}$$

One then, naturally, looks at the logquasilikelihood

$$\begin{aligned}
lql(\mu, \sigma^2, \nu, \theta) &= \ln \prod_{i<j} C(\nu, n) + \ln \prod_{i<j} |c\sigma^2 \mathbf{W}_\theta|^{-\frac{1}{2}} \\
&\quad + \ln \prod_{i<j} [1 + \frac{1}{\nu c\sigma^2}(\mathbf{z} - \boldsymbol{\mu})^T \mathbf{W}_\theta^{-1}(\mathbf{z} - \boldsymbol{\mu})]^{-\frac{\nu+2}{2}}
\end{aligned}$$

Here

$$\begin{aligned}
\ln \prod_{i<j} C(\nu, 2) &= N \ln C(\nu, 2) \\
\ln \prod_{i<j} |c\sigma^2 \mathbf{W}_\theta|^{-\frac{1}{2}} &= \ln \prod_{i<j} \left[c^2\sigma^4(1-e^{-2\theta\Delta_{ij}})\right]^{-1/2} \\
&= -\frac{1}{2}\sum_{i<j} \ln(1-e^{-2\theta\Delta_{ij}}) - Nlnc\sigma^2 \\
\ln \prod_{i<j} [1+\frac{1}{\nu c\sigma^2}(\mathbf{z}-\mu)^T \mathbf{W}_\theta^{-1}(\mathbf{z}-\mu)]^{-\frac{\nu+2}{2}} & \\
&= -\frac{\nu+2}{2}\sum_{i<j} \ln\left(1+\frac{Q(\mu,\theta)}{\nu c\sigma^2 R(\theta)}\right)
\end{aligned}$$

so that

$$\begin{aligned}
lql(\mu,\sigma^2,\nu,\theta) &= N \ln C(\nu,2) - \frac{1}{2}\sum_{i<j} \ln(1-e^{-2\theta\Delta_{ij}}) - N\ln c\sigma^2 \\
&\quad -\frac{\nu+n}{2}\sum_{i<j} \ln\left(1+\frac{Q(\mu,\theta)}{\nu\sigma^2 R(\theta)}\right) \\
&= N\left(\ln C(\nu,2) - \ln c\sigma^2\right) \\
&\quad -\frac{1}{2}\sum_{i<j}\left[\ln R(\theta)\cdot \ln\left(1+\frac{Q(\mu,\theta)}{\nu\sigma^2 R(\theta)}\right)^{\nu+2}\right] \\
&= N \ln \frac{C(\nu,2)}{c\sigma^2} - \frac{1}{2}\sum_{i<j} \ln\left[R(\theta)\left(1+\frac{Q(\mu,\theta)}{\nu c\sigma^2 R(\theta)}\right)^{\nu+2}\right]
\end{aligned}$$

Now, this must be maximied with respect to θ through some numerical method. Either by finding the maxima by a method like the one in Appendix D.1, or by differentiating and finding the zero with a method like the one in Appendix D.2. Differentiating does not reduce the complexity of the problem, so maximizing is preferrable. Therefore

$$\hat{\theta}_{MQL} = \max_\theta \left\{ N \ln \frac{C(\nu,2)}{c\sigma^2} - \frac{1}{2}\sum_{i<j} \ln\left[R(\theta)\left(1+\frac{Q(\mu,\theta)}{\nu c\sigma^2 R(\theta)}\right)^{\nu+2}\right]\right\}$$

Quasilikelihood could of course also be used to find estimators for the other parameters as well, but as long as the maximum likelihood estimators are available, these are better. This is due to the fact that quasilikelihood only consideres pairs of variables and not the complete likelihood. Quasilikelihood does, however, give results that are known to be quite good, and are computationally less time-consuming to find. Be aware of that finding θ is still no easy computational task since there are quite a lot of pairs in a field if one consideres them all. For a 25×25 field, which constitutes of 625 points, one has $\frac{625\,(625-1)}{2} = 195.000$ pairs. Therefore one often includes only the closest neighbours of a point, the local neighbourhood, in the quasilikelihood to reduce computer time. This approach does barely weaken the results, since points far from one another are assumed not to be too dependent on one another, see Section 6.3.2.

The maximum quasilikelihoodestimator of θ will, like the ML-estimator, not be very robust. The quasilikelihoodfunction has a unique maximum, but the function is almost completely flat in that area, making the MQL-estimator of θ very vulnerable to small changes in sampled values.

7.4 Degrees of freedom parameter ν

When finding the ML-estimator of ν, one has to take into account the connection between $\mathbf{V}$ and ν, since for $\nu \geq 3$ $\mathbf{V} = \frac{\nu-2}{\nu}\sigma^2\mathbf{W}$, i.e., the constant c used in the previous subsections is $\frac{\nu-2}{\nu}$, while for $0 < \nu < 3$ there is no direct link, i.e, $c = 1$. This means that these two cases have to be considered separately when differentiating the loglikelihood with respect to ν. Differentiating the loglikelihood with respect to ν with this side condition yields

$$\begin{aligned}
\left.\frac{\partial l}{\partial \nu}\right|_{\nu\geq 3} &= \frac{\partial}{\partial \nu}\left\{\ln\Gamma(\frac{\nu+n}{2}) - \ln\Gamma(\frac{\nu}{2}) - \frac{n}{2}\ln(\nu\pi) - \frac{1}{2}\ln\left[(\frac{\nu-2}{\nu})^n\sigma^{2n}|\mathbf{W}|\right]\right.\\
&\qquad \left.-\frac{\nu+n}{2}\ln\left[1+\frac{1}{\nu-2}\frac{1}{\sigma^2}\underbrace{(\mathbf{z}-\mathbf{1}\mu)^T\mathbf{W}^{-1}(\mathbf{z}-\mathbf{1}\mu)}_{D_W}\right]\right\}\\
&= \frac{\Gamma'(\frac{\nu+n}{2})\cdot\frac{1}{2}}{\Gamma(\frac{\nu+n}{2})} - \frac{\Gamma'(\frac{\nu}{2})\cdot\frac{1}{2}}{\Gamma(\frac{\nu}{2})} - \frac{n}{2}\cdot\frac{1}{\nu} - \frac{1}{2}\frac{n\cdot(\frac{\nu-2}{\nu})^{n-1}}{(\frac{\nu-2}{\nu})^n}\cdot\frac{\nu-(\nu-2)}{\nu^2}\\
&\qquad -\frac{1}{2}\ln\left[1+\frac{1}{\nu-2}\frac{D_W}{\sigma^2}\right] - \frac{\nu+n}{2}\frac{1}{1+\frac{1}{\nu-2}\frac{D_W}{\sigma^2}}\cdot\frac{D_W}{\sigma^2}\frac{-1}{(\nu-2)^2}\\
&= \frac{1}{2}\left\{DG(\frac{\nu+n}{2}) - DG(\frac{\nu}{2}) - \frac{n}{\nu} - \frac{2n}{\nu(\nu-2)}\right. \qquad (21)\\
&\qquad \left.-\ln\left[1+\frac{1}{\nu-2}\frac{D_W}{\sigma^2}\right] + \frac{\nu+n}{(\nu-2)^2}\cdot\frac{\frac{D_W}{\sigma^2}}{1+\frac{1}{\nu-2}\frac{D_W}{\sigma^2}}\right\}\\
&= \frac{1}{2}\cdot h_1(\nu)
\end{aligned}$$

with $DG(\cdot)$ being the digamma function defined by $DG(x) = \frac{\Gamma'(x)}{\Gamma(x)}$. This means that for a given σ and $\mathbf{W}$ one can find the ML-estimator of ν as the solution to the problem

$$\nu \geq 3 \wedge h_1(\nu) = 0 \qquad (22)$$

with $h_1(\nu)$ given in expression (21). Finding a zero is a numerical task which is outlined in Appendix D.2.

If $0 < \nu < 3$ one can perform calculations similar to the ones above, now with $\mathbf{V} = 1\cdot\sigma^2\mathbf{W}$, and find

$$\left.\frac{\partial l}{\partial \nu}\right|_{0<\nu<3} = \frac{1}{2}\cdot h_2(\nu)$$

with

$$h_2(\nu) = \frac{1}{2}\left\{DG(\frac{\nu+n}{2}) - DG(\frac{\nu}{2}) - \frac{n}{\nu} - \ln\left[1+\frac{D_W}{\nu\sigma^2}\right] + \frac{\nu+n}{\nu^2}\cdot\frac{D_W/\sigma^2}{1+\frac{\frac{D_W}{\sigma^2}}{\nu\sigma^2}}\right\}$$

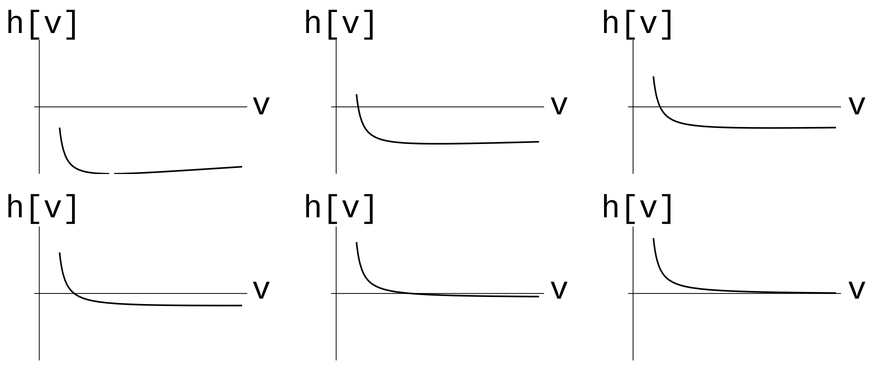


Figure 2: $h(\nu)$ *for increasing values of the Mahalanobis distance (horizontally to the right) and dimension* $625 = 25 \times 25$. $h(\nu) \in [-1.5, 1.5]$ *and* $\nu \in [3, 30]$

Now the system

$$0 < \nu < 3 \wedge h_2(\nu) = 0 \tag{23}$$

must be solved. This can be done in a manner similar to that of expression (22).

The ML-estimator of ν will, like the MQL-estimator of θ, unfortunately not be very robust. By looking at a typical figure of $h(\nu)$ for given dimension n and squared Mahalanobis distance D, Figure 2, one sees that in the cases of a zero, the function has a single crossing, but it is almost flat in that area, making the estimator very vulnerable to differences in sampled values. The behaviour of $h(\nu)$ in Figure 2 is typical for all D. In addition one finds that in order for the systems (22) and (23) to have a solution at all, D and n need to be of the same magnitude.

A way of visualizing the unstability of the ML-estimate would then be to draw the graph of $h(\nu)$ for that given value of n and D. One can also slightly change a few observations and observe what happens to ν and $h(\nu)$. Including graphs of the loglikelihood itself could also be useful, but be aware that one then would need the determinant of $\mathbf{W}$, a calculation that could be rather tedious.

7.5 Algorithm for the estimation of parameters

When estimating the parameters of an assumed random t-distributed field from one single sample, one observes that the ML estimator of μ is only dependent on the observation vector $\mathbf{z}$ and can be calculated independently of the other parameters σ, ν and θ, i.e. $\mathbf{W}$. These parameters, however, depend on one another and must therefore be calculated simultanously. One way of doing this is by an EM type numerical iteration algorithm. For references on the EM algorithm see [9]. If one assumes the number of degrees of freedom to be greater than 3 one would proceed as follows:

Algorithm 4 ***Parameter estimation***

- *Estimate* $\hat{\mu} = \bar{\mathbf{z}}$.
- *Initiate: Choose initial values* $\sigma^{2(0)}$, $\nu^{(0)}$, $\theta^{(0)}$ *and* $D_W(\theta^{(0)})^{(0)}$.
- *Iterate until a satisfactory level of closeness is obtained:*
  1. $c^{(s+1)} = \frac{\nu^{(s)}-2}{\nu^{(s)}}$
  2. $\hat{\sigma}^{2(s+1)} = \frac{D_W^{(s)}}{n \cdot c^{(s+1)}}$
  3. $$\begin{aligned}\hat{\theta}^{(s+1)} &= \max_{\theta}\Bigg\{ N \ln \frac{C(\nu^{(s)}, 2)}{c^{(s+1)}\sigma^{2(s+1)}} \\ &\quad -\frac{1}{2}\sum_{i<j} \ln \left[R(\theta) \left(1 + \frac{Q(\hat{\mu}, \theta)}{\nu^{(s)} c^{(s+1)} \sigma^{2(s+1)} R(\theta)} \right)^{\nu^{(s)}+2} \right] \Bigg\}\end{aligned}$$
  4. $D_W^{(s+1)} = (\mathbf{z} - \mathbf{1}\mu)^T (\mathbf{W}^{-1})^{(s+1)} (\mathbf{z} - \mathbf{1}\mu)$ *with* $(\mathbf{W}^{-1})^{(s+1)}$ *found through* $\mathbf{W}(\theta^{(s+1)})$
  5. $\nu^{(s+1)}$ *is the solution to the system of equations*
  $$\nu \geq 3 \wedge h_1(\nu, \hat{\sigma}^{2(s+1)}, D_W^{(s+1)}) = 0$$

This will find the estimators under the assumption that $\nu \geq 3$. But one must also check whether there is a solution for $1 \leq \nu < 3$. To do this, remove step 1, since $c = 1$, and substitute step 5 with $\nu^{(s+1)}$ as the solution to the system of equations $1 \leq \nu < 3 \wedge h_2(\nu, \hat{\sigma}^{2(s+1)}, D_W^{(s+1)}) = 0$. Most likely one will find only one solution for either $1 \leq \nu < 3$ or $3 \leq \nu$, but due to the unstability of the estimation of ν and the problematic case when $\nu \approx 3$, this is not necessarily the case, but one will find maximum one solution in each interval.

The reason behind the choice of differentiating and finding a zero when searching for ν, and finding a maxima when looking for θ, is that the loglikelihood includes the determinant of $\mathbf{W}$ and this will vanish when differentiating with respect to ν but not otherwise.

Since every repetition of steps 1-5 is rather tedious a lot can be done in speeding up the algorithm by choosing good starting values.

8 Simulation of random t-distributed fields

In Section 6 an algorithm for simulating random t-distributed fields was presented. In this section this method will be tested through computer simulation.

8.1 Unconditional sampling

First some unconditional random t-distributed fields will be generated and commented on.

8.1.1 Results of unconditional sampling

Algorithm 2 has been implemented in the statistical programming language S-Plus to generate some random t-distributed fields with various degrees of freedom. All fields have dimension 10×10 and for all fields $\mu = 0$, $\sigma = 1$ and $\theta = 0.3$ has been used. The value $\theta = 0.3$ implies $R = 10$ and $\beta = 1$ for the parameters of Section 6.3.2. The results are shown in Figures 3 and 4.

8.1.2 Comments on results of unconditional sampling

If one looks at Figure 3 one notices that the surfaces for $\nu \geq 3$ in general appear to be smoother than the ones for $1 \leq \nu < 3$. Be aware of the fact, however, that it doesn't necessarily have to be this way. By taking a look at Figure 4, which is a collection of nine random t-distributed fields with $\nu = 3$, one observes that there are quite large differences between some of these surfaces. An explanation is that since the conditional variance is dependent on the values already sampled, an extreme value early in the sample process, which is attainable due to the heavy tails of the t-distribution, the conditional variance will be large in the next sample step, making it quite possible to draw a new extreme observation. Through the rest of the simulation process the field will keep this high variance, though the interactional effects may lower it some, resulting in a rather ragged field. Likewise, if a couple of values drawn early in the process are very close to the location μ the field will be regulated and flattened out by these and it becomes highly unlikely to draw an extreme observation. As one can see, unconditional simulation of random t-distributed fields is quite vulnerable to the first sampled values.

It is also appearant that random t-distributed fields are quite homogeneous on the whole area. The occational large jump from a rather smooth surface does not happen. This can too be explained with the conditional variance of the distribution. In addition will the interactional effects make such behaviour very unlikely.

8.2 Conditional sampling

When performing conditional sampling, one way to keep control of the result is to use a field that is known in all points, pick k points at random from this field and then sample new fields conditioned on these k points. The resulting fields can then be compared to the true realization.

8.2.1 Results of conditional sampling

For conditional sampling the field in Figure 5 was chosen as the "true" field and from this field five random points were picked to be the five observations to condition the new fields on, see Figure 6. Then Algorithm 3 was implemented in the statistical programming language S-Plus and applied to generate four conditional random t-distributed fields with the same parameters as the field in Figures 5 and 6. The result is shown in Figures 7 and 8.

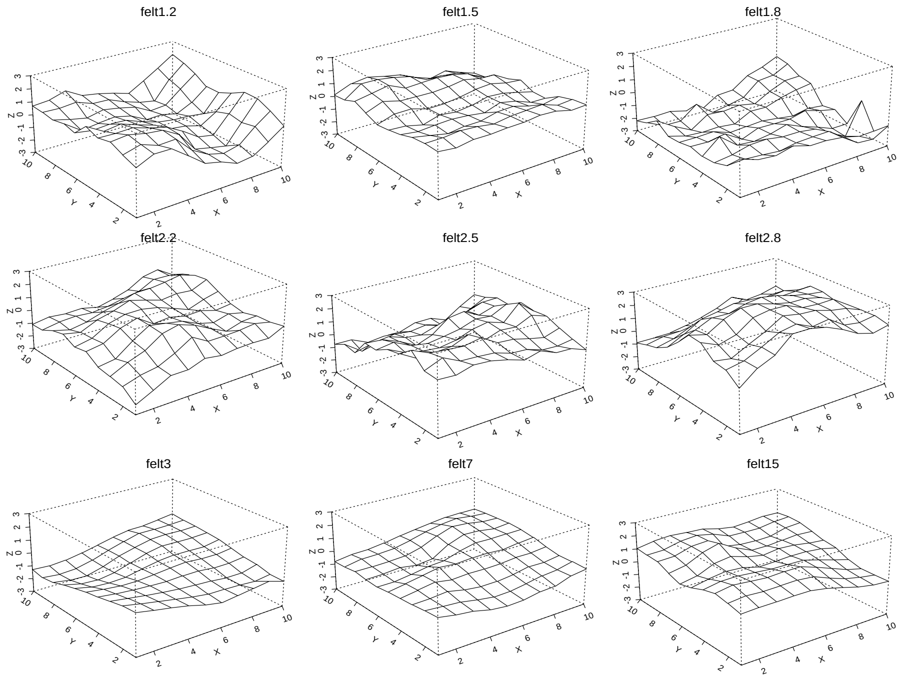


Figure 3: *Random t-distributed fields with* $\nu =$ *1.2, 1.5, 1.8, 2.2, 2.5, 2.8, 3, 7 and 15 degrees of freedom*

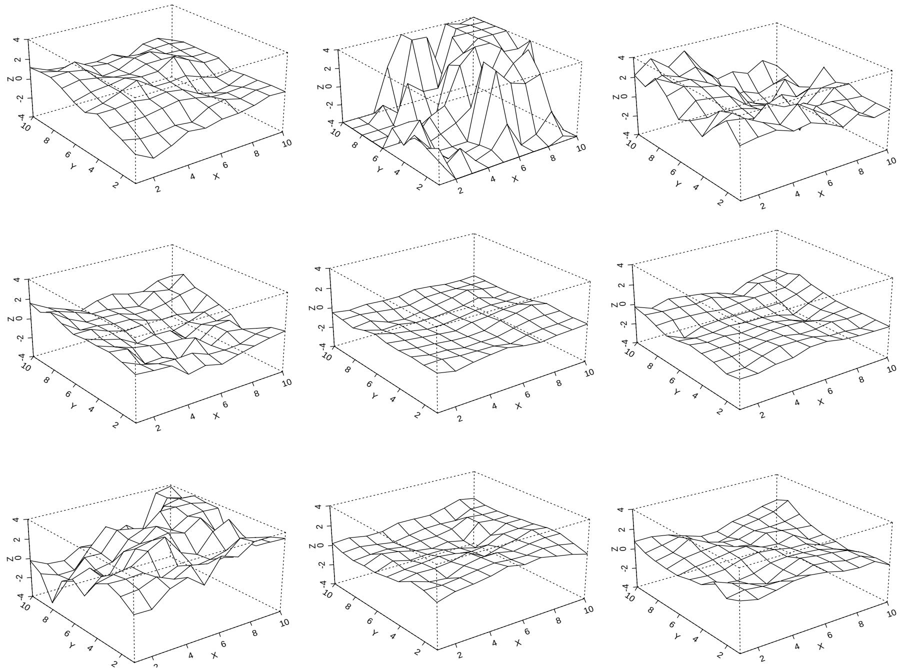


Figure 4: *Random t-distributed fields with* $\nu = 3$ *degrees of freedom*

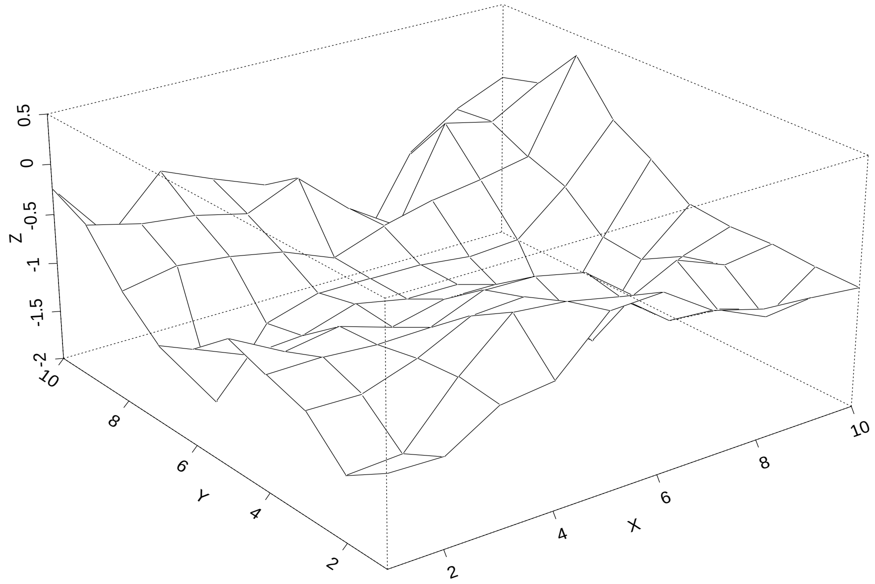


Figure 5: *Random t-distributed field with $\nu = 3$ degrees of freedom.*

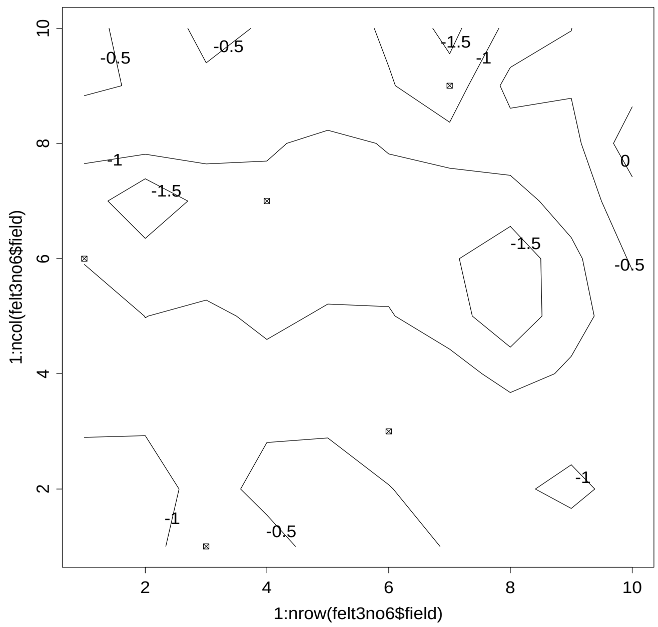


Figure 6: *Contourplot of random t-distributed field with $\nu = 3$ degrees of freedom. The squares are the points the fields are to be conditioned on.*

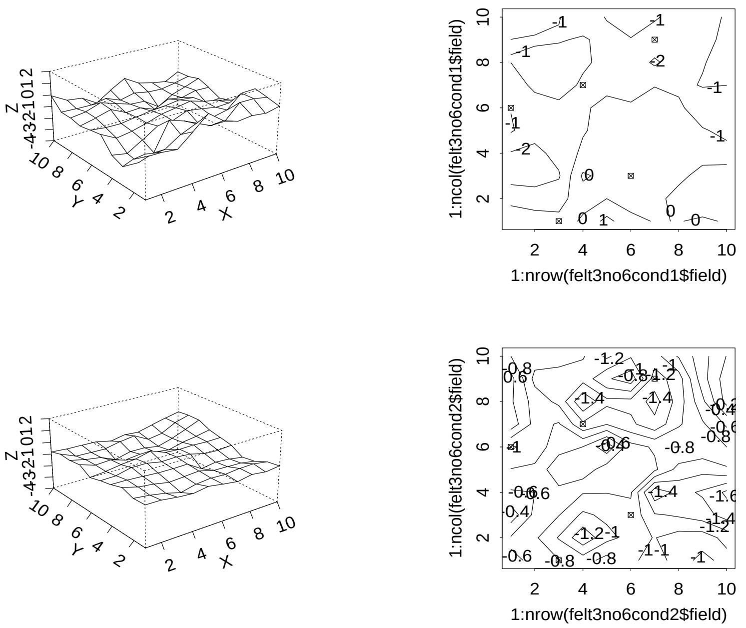


Figure 7: *Conditional random t-distributed fields with $\nu = 3$ degrees of freedom*

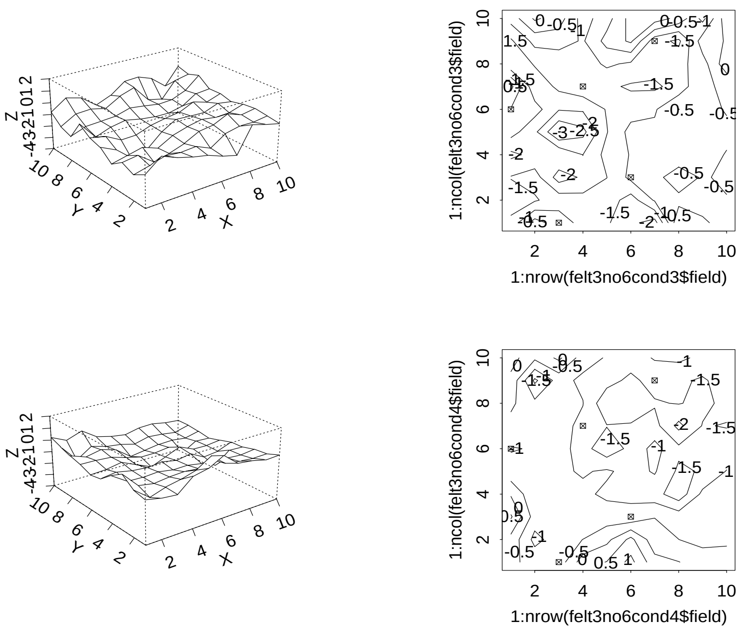


Figure 8: *Conditional random t-distributed fields with $\nu = 3$ degrees of freedom*

8.2.2 Comments on results of conditional sampling

For the conditional fields one observes that they vary in overall variability from field to field, but not as much as the unconditionally sampled fields. The conditional fields preserve the qualities of the original field, but introduce more variability than Gaussian random fields do. Gaussian random fields are also more homogenous in appearance than random t-distributed fields. So, given that one did not know how the original field looked, but only knew the k observations, the random st-distributed fields do therefore give a more varied view on how the original field might be.

9 Estimation of parameters in random t-distributed fields

In Section 7 an algorithm for the estimation of the parameters in a random t-distributed field was presented. In this section this method will be tested through computer simulation and commented on.

9.1 Results of estimation

Algorithm 4 has been implemented in the statistical programming language S-Plus and tested on the three unconditional fields in the first column of Figure 4. For all these 10×10 fields the parameters $\mu = 0$, $\sigma = 1$, $\theta = 0.3$ and $\nu = 3$ were used in the generation process. The results of the estimation is summarized in the following table:

Estimation of parameters in 10×10 fields				
Field	μ	σ	θ	ν
1	0.297	9.21	0.017	11.54
2	-0.127	11.21	0.020	11.54
3	1.791	23.29	0.018	11.55

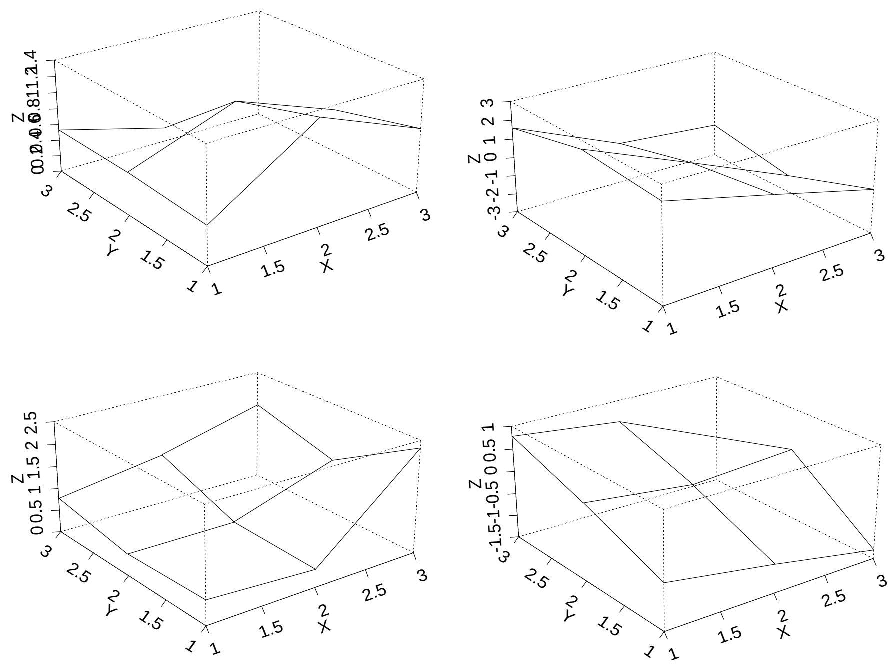
9.2 Comments on results of estimation

When looking at the estimation results of Section 9.1 there is particularly

one column of the table that draws attention, namely the one for ν.

It seems highly unlikely that these three fields will generate almost identical results for the estimation of ν. However, when practically the same ν and θ has been estimated for all fields, it is natural that σ of field three is much larger than the two others.

Estimation of the parameters of other fields in Figure 4 gave similar results, with almost identical ν and θ and varying σ.



$\nu = 3$ degrees of freedom

Another aspect of the numbers in the ν-column, is the particular value 11.5. For 10×10 fields, i.e., fields with dimension $n = 10$, this particular value for ν is very close to $n + 1$. To check whether there was anything in the claim that the implementation estimated ν as $\hat{\nu} \approx n + 1$, Algorithm 4 was used on tiny 3×3 fields with the same field parameters, see Figure 9 for the

realizations, and the following estimation results were obtained:

	Estimation of parameters in 3×3 fields			
Field	μ	σ	θ	ν
1	0.61	7.55	0.01	4.06
2	0.30	3.32	1.13	4.04
3	1.01	12.62	0.01	4.04
4	-0.17	10.82	0.01	4.07

Figure 9: $\nu = 3$ degrees of freedom*Random t-distributed fields with* $\nu = 3$ *degrees of freedom*

Here, too, one observes that the estimate for ν is almost identical for all fields, and very close to $n+1$. And, as for the 10×10 fields, one observes that nearly all fields have almost identical ν and θ, but varying σ. This time, however, the value estimated for θ is worthwhile taking a closer look at too.

The alarming feature with this particular value for θ, i.e., $\theta = 0.1$, is that this value was chosen as the minimum value for θ in the implementation. The value $\theta = 0$ implies a fully deterministic model so it is a sensibel claim that $\theta > 0$. Therefore, by getting the estimate $\theta = 0.1$, this seems to imply that the quasilikelihoodfunction of Section 7.3.2 has a maximum in 0, which is highly unlikely. If so is the case, however, then the function would probably have another maxima for $\theta > 0$. Unfortunately no plots indicate such a thing. It is more likely that the implementation itself causes these errors.

10 Transformation of Gaussian random fields

Section 6 shows a method for generating random t-distributed fields. As observed, some of the work is clearly more troublesome than when generating Gaussian random fields. It would therefore be be of interest to find out whether it is possible to generate a Gaussian random field and then transform this into a random t-distributed field, instead of applying all the work of Section 6.

It is known that for a Gaussian random field, $\mathbf{Z} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, all marginals are Gaussian, so that each single observation, Z_i, is distributed as $Z_i \sim N_1(\mu, \sigma^2)$. One may now transform every single observation into a t-distribut-ed variable, Y_i, through the transformation

$$\begin{aligned} Y_i &= g(Z_i) \\ &= G_T^{-1}\left(G_N(Z_i)\right), \end{aligned} \tag{24}$$

and obtain a situation where all univariate marginals are t-distributed. Knowing that all marginals are t-distributed does, however, not guarantee that the field is t-distributed. The field will be elliptic with location μ, but what kind of elliptic distribution is not necessarily known, nor what will happen to the covariance structure by performing this transformation. Here two alternative approaches to this question will be considered.

10.1 Multivariate transformation

The most preferrable alternative is to perform the complete multivariate transformation of all variables simultaneously and see if this would generate a random t-distributed field. Generally the following applies for multivariate transformation, see [10]: Let $(Z_1, \ldots, Z_n)$ be a random vector with probability density function $f_{\mathbf{X}}(z_1, \ldots, z_n)$. Consider a new random vector, $(U_1, \ldots, U_n)$, defined by $U_1 = g(Z_1, \ldots, Z_n), \ldots, U_n = g(Z_1, \ldots, Z_n)$. When the transformation $(U_1, \ldots, U_n) = (g_1(\mathbf{Z}), \ldots g_n(\mathbf{Z}))$ is one-to-one the inverse can be found, and this is denoted $z_1 = h_1(u_1, \ldots, u_n), \ldots, z_n = h_n(u_1, \ldots, u_n)$. The joint probability density function of $(U_1, \ldots, U_n)$ will now be given by

$$f_{\mathbf{U}}(u_1, \ldots, u_n) = f_{\mathbf{X}}(h_1(u_1, \ldots, u_n), \ldots, h_n(u_1, \ldots, u_n))|J|$$

with J being the Jacobian

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_1(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial h_1(\mathbf{u})}{\partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial h_n(\mathbf{u})}{\partial u_n} \end{vmatrix}$$

By choosing the particular transformation of expression (24) one obtains

$$\begin{aligned}
U_1 &= G_T\left[G_N^{-1}(Z_1)\right] = g_1(Z_1) \\
&\vdots \\
U_n &= G_T\left[G_N^{-1}(Z_n)\right] = g_n(Z_n)
\end{aligned}$$

and therefore

$$\begin{aligned}
Z_1 &= G_N\left[G_T^{-1}(U_1)\right] = h_1(U_1) \\
&\vdots \\
Z_n &= G_N\left[G_T^{-1}(U_n)\right] = h_n(U_n)
\end{aligned}$$

In the Jacobian $\frac{\partial Z_i}{\partial U_j}$ is needed, and this is

$$\frac{\partial Z_i}{\partial U_j} = \frac{\partial}{\partial U_j} G_N\left[G_T^{-1}(U_i)\right] = \begin{cases} 0 & i \neq j \\ \frac{\partial G_N}{\partial G_T^{-1}} \cdot \frac{\partial G_T^{-1}}{\partial U_j} & i = j \end{cases}$$

and the Jacobian will then be

$$J = \begin{vmatrix} \frac{\partial Z_1}{\partial U_1} & \cdots & \frac{\partial Z_1}{\partial U_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z_n}{\partial U_1} & \cdots & \frac{\partial Z_n}{\partial U_n} \end{vmatrix} = \prod_{p=1}^{n} \frac{\partial Z_p}{\partial U_p}$$

A very neat and handy mathematical expression, indeed. For the particular transformationss in question are

$$G_N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

and

$$G_T(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \int_{-\infty}^{x} \frac{1}{\left[1 + \frac{1}{\nu}(\frac{t-\mu}{\sigma})^2\right]^{\frac{\nu+1}{2}}} dt$$

In the heart of these expressions are the integrals

$$\begin{aligned}
\int e^{-t^2} dt &= \frac{\sqrt{\pi}}{2} erf(t) + c_1 \\
\int (1+t^2)^{-\frac{\nu+1}{2}} dt &= t \cdot {}_2F_1(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}, -t^2) + c_2
\end{aligned}$$

The function $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function and is central to many statistical calculations. It has, however, no analytical solution. The second special function, ${}_2F_1(a, b; c; z) = \int_0^1 s^{b-1}(1-s)^{c-b-1}(1-sz)^{-a} ds$, is the solution to a kind of hypergeometric differential equations. Written as a series expansion it is known as the Gauss series. Neither this has an analytical solution. This implies that it is not possible to perform an analytical multivariate transformation of a Gaussian random field. Even if one reduces the problem to a univariate transformation, one cannot find an analytical solution. Whether this problem may be solved numerically is beyond the scope of this report.

10.2 Pairwise consideration

For every pair of normally distributed variables Z_i and Z_j one derives the following

$$\begin{aligned}
Cov(Y_i, Y_j) &= Cov(g(Z_i), g(Z_j)) \\
&= E\left\{[g(Z_i) - E(g(Z_i))]\,[g(Z_j) - E(g(Z_j))]\right\} \\
&= E\left[g(Z_i)g(Z_j)\right] - E\left[g(Z_i)E(g(Z_j))\right] \\
&\quad - E\left[E(g(Z_i))g(Z_j)\right] + E\left[E(g(Z_i))E(g(Z_j))\right] \\
&= E\left[g(Z_i)g(Z_j)\right] - E(g(Z_i))E(g(Z_j))
\end{aligned}$$

With $g(\cdot)$ as in expression (24). From Jensen's unequality

$$\begin{aligned}
E(g(Z_i)) &\geq g(E(Z_i)) \\
&= g(\mu_i)
\end{aligned}$$

if g is convex, and opposite if g is concave. See [10] for proof. This implies that one cannot obtain an analytical expression for the covariance between variables with this, or basically any, transformation. Obtaining bonds is also difficult, since $g(\cdot)$ is not necessarily convex or concave for both variables Z_i and Z_j at the same time. It is not possible to find the expectations explixitely through direct calculations either, since it is impossible to find an analytical expression for the transformation in question.

10.3 Summary transformation

Sections 10.1 and 10.2 show the fact that even if the transformation needed to transform Gaussian random fields to random t-distributed fields is known, it impossible to apply this to get random t-distributed fields. Other, more troublesome approaches, like the method of Section 6, must be used.

11 Closing remarks

The t-distribution is a density distribtution with shape similar to the shape of the Gaussian distribution, but with more probability mass in the tails. Because of this, a random t-distributed field is a random field with more variability than a Gaussian random field. A random field is a function whose values are random variables on a parameter set, for example $\mathbf{R}^2$.

In this report a sequential algorithm for sampling random t-distributed fields is created. The algorithm uses rejection sampling in each iteration step, with the Cauchy distribution as proposal, and testing of the algorithm shows that it is just slightly slower than algorithms for sampling Gaussian random fields.

The algorithm has been used to simulate both unconditional and conditional random t-distributed fields. The generated fields show that unconditional random t-distributed fields are very vulnerable to the first sampled values, and this can be explained with the fact that the conditional variance of the t-distribution is dependent on the already sampled values. Conditional random t-distributed fields vary quite a lot in overall variability from field to field too, but not to the same extent as unconditional random t-distributed fields. The conditional random t-distributed fields therefore appear to be an alternative to the Lévy stable model.

The random t-distributed fields model has, in the lack of such, not been tested on real data. This is one of next steps that should be taken to check whether the random t-distributed fields better explain variability in real data than the Gaussian random fields do. Since conditional random t-distributed fields vary more in appearance from field to field, even with identical parameters, than Gaussian random fields, the random t-distributed fields will in most cases probably give a more realistic picture of the uncertainty of the simulated conditional fields.

An algorithm for the estimation of parameters in a random t-distributed field has also been created. In this EM-algorithm maximum likelihood estimators are found for all parameters, except for the interactional parameters of the exponential covariance function. Here the maximum quasilikelihood estimator is chosen in order to speed up the algorithm.

This algorithm has been implemented and tested, but the results did not give room for conclusions. Since I have reasons to believe that the theory and the algorithm is right, the problem is most likely numerical unstability, since the estimation of both ν and θ is highly unstable. One can, however, not rule out bugs. Finding a way to solve this problem is therefore another step that should be taken in the direction of completing the testing of the behaviour of random t-distributed fields.

References

[1] S. Painter and L. Paterson (1994) *Lévy Stochastic Model for the Variations in the Properties of Sedimentary Rock*, Geophysical Research Letters, 21, 2857-2860

[2] S. Painter and L. Paterson (1995) *Improved Technique for Stochastic Interpolation of Reservoir Properties*, SPE 30599, presented at the SPE Annual Technical Conference & Exhibition in Dalla, USA, 22-25 October, 1995.

[3] Petter Abrahamsen (1997) *A Review of Gaussian Random Fields and Correlation Functions*, Norwegian Computing Centre.

[4] K.L Lange, R.J.A. Little and J.M.G. Taylor (1989) *Robust statistical modelling using the t-distribution*, Journal of the American Statistical Association, 84, 881-896.

[5] Robb J. Muirhead (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.

[6] D.K. McGraw and J.F. Wagner (1967) *Elliptically symmetric distributions*, IEEE Transactions on Information Theory, Vol. IT-14, No.1, January.

[7] E.A. Cornish (1954) *The multivariate t-distribution associated with a set of normal sample eluminates*, Aust. Jour. Phys, 7, 531-542.

[8] N.L. Johnson and S. Kotz (1972) *Distributions in Statistics: Continuous Multivariate distributions*, Wiley, New York.

[9] M. Langaas (1995) *Mixture Codebook Classification, Part 2:Model Selection and Method Evaluation*, Technical Report Statistics 6/1995, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim.

[10] G. Casella and R.L. Berger (1990) *Statistical Inference*, Duxbury Press, Wadsworth.

[11] B.D Ripley (1994) *Stochastic Simulation*, John Wiley & Sons, New York

[12] J.T. Kent, D.E. Tyler and Y. Vardi (1994) *A curious likelihood identity for the multivariate t-distribution*, Communications in Statistics, Simulation and Computation, 23, 441-453.

[13] A.P. Dempster, N.M. Laird and D.B. Rubin (1977) *Maximum likelihood from incomplete data via the EM-algorithm (with discussion)*, Journal of the Royal Statistical Society, Series B, 39, 1-38.

[14] N.L. Hjort and H. Omre (1994) *Topics in spatial statistics*, Scandinavian Journal of Statistics, Vol 21, 289-357.

[15] S.Geman and D. Geman (1984) *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transaction on Pattern Analysis and Machine Intelligence, **6**, 721-41.

[16] W.K. Hastings (1970) *Monte Carlo sampling methods using Markov chains and their applications*, Biomtrika, **57**, 1, 97-109.

[17] S. Chib and E. Greenberg (1995) *Understanding the Metropolis-Hastings Algorithm*, The American Statistician, Vol 49, no 4, november 1995.

Appendices

A Matrix inversion

A.1

For a $k \times k$ nonsingular matrix $\mathbf{A}$, with inverse $\mathbf{A}^{-1}$, partitioned as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{pmatrix}, \mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A^{11}} & \mathbf{A^{12}} \\ \mathbf{A^{21}} & \mathbf{A^{22}} \end{pmatrix}$$

the inverse is found as

$$\begin{aligned} \mathbf{A^{11}} &= \left(\mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{21}}\right)^{-1} \\ \mathbf{A^{22}} &= \left(\mathbf{A_{22}} - \mathbf{A_{21}}\mathbf{A_{11}}^{-1}\mathbf{A_{12}}\right)^{-1} \\ \mathbf{A^{12}} &= -\mathbf{A_{11}}^{-1}\mathbf{A_{12}}\mathbf{A^{22}} \\ \mathbf{A^{21}} &= -\mathbf{A_{22}}^{-1}\mathbf{A_{21}}\mathbf{A^{11}} \end{aligned} \tag{25}$$

For proof, see [5].

A.2

For nonsingular matrices $\mathbf{E}$ and $\mathbf{F}$, $p \times p$ and $q \times q$ respectively, and $\mathbf{C}$ $p \times q$ and $\mathbf{D}$ $q \times p$ the inverse of the matrix $\mathbf{P} = \mathbf{E} + \mathbf{CFD}$ is found as:

$$\mathbf{P}^{-1} = \mathbf{E}^{-1} - \mathbf{E}^{-1}\mathbf{CF}\left(\mathbf{F} + \mathbf{FDE}^{-1}\mathbf{CF}\right)\mathbf{FDE}^{-1} \tag{26}$$

For proof, see [5]

A.3

If the $k \times k$ matrix $\mathbf{A}$ has the particular form

$$\mathbf{A} = \begin{pmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{B} \end{pmatrix}$$

with a a number, $\mathbf{B}$ $(k-1) \times (k-1)$ and $\mathbf{b}$ $(k-1) \times 1$, then, by equations (25)

$$\begin{aligned} \mathbf{A^{11}} &= (a - \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b})^{-1} \\ \mathbf{A^{22}} &= (\mathbf{B} - \frac{1}{a}\mathbf{b}\mathbf{b}^T)^{-1} \\ \mathbf{A^{12}} &= -\frac{1}{a}\mathbf{b}^T\mathbf{A^{22}} \\ \mathbf{A^{21}} &= -\mathbf{B}^{-1}\mathbf{b}\mathbf{A^{11}} \end{aligned}$$

The inversion in $\mathbf{A^{11}}$ is nothing but the inversion of a number, and

$$\begin{aligned}
(\mathbf{B} - \frac{1}{a}\mathbf{b}\mathbf{b}^T)^{-1} &= \left(\mathbf{B} + \mathbf{b}(-\frac{1}{a})\mathbf{b}^T\right)^{-1} \\
&= \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{b}(-\frac{1}{a})\left((-\frac{1}{a}) + (-\frac{1}{a})\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}(-\frac{1}{a})\right)^{-1} \\
&\quad \cdot(-\frac{1}{a})\mathbf{b}^T\mathbf{B}^{-1} \\
&= \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{b}\left(a - \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}\right)^{-1}\mathbf{b}^T\mathbf{B}^{-1} \\
&= \mathbf{B}^{-1} + \mathbf{B}^{-1}b\,\mathbf{A^{11}}\,\mathbf{b}^T\mathbf{B}^{-1}
\end{aligned}$$

by equation (26).

B Stochastic simulation

B.1 Markov Chain Monte Carlo Simulation

As the solution to the problem of sampling a value x from a one-dimensional distribution $f(x)$, written $x \sim f(x)$, MCMC algorithms take use of a reversible Markov chain $\{x(t)\}_{t=0}^{\infty}$ with stationary probability density $f(x)$. For this a transactional cernel P_{xy} is needed, with x being the current state and y being the new one. One then applies the trick of writing $P_{xy} = q_{xy}\alpha_{xy}$ with q_{xy} being a known, chosen distribution to sample the new state y from, and α_{xy} being the probability of accepting this new value. One finds that

$$\alpha_{xy} = \min\left(1, \frac{f(y)q_{yx}}{f(x)q_{xy}}\right)$$

The algorithm then reads

Algorithm 5 ***MCMC simulation***

- *Initiate:* $x \in \Omega_x$, *for example* $x(0) = 0$.
- *Iterate*
  1. $y \sim q_{xy}$
  2. $\alpha_{xy} = \min\left(1, \frac{f(y)q_{yx}}{f(x)q_{xy}}\right)$
  3. $u \sim Uniform(0, 1)$
  4. *If* $u < \alpha_{xy}$ *then* $x_{t+1} = y$ *else* $x_{t+1} = x_t$

The main problem here is the question of convergence, i.e., how many iterations are needed before one is close enough to the stationary distribution. Unfortunately this question cannot be answered. What one usually does, is to try with a lot of iterations and then plot the values as a function of iterations, and see when the values appear to be somewhat stabilized around some level. By doing this, one finds that one in most cases needs at least 100 iterations to be somewhat sure of convergence. Often even more iterations are necessary. The method has, however, proved to work very well and it is very easy to use.

B.2 Rejection sampling

As the solution to the problem of sampling a value x from a distribution $f(x)$, written $x \sim f(x)$, the rejection sampling procedure takes use of another distribution, $g(x)$, which sampling from is quite easy. Loosely the algorithm can be sketched as follows: Sample $y \sim g(y)$ and then return $x = y$ with probability $h(y)$, else sample a new value y. The algorithm thereby keeps on sampling values $y \sim g(y)$ and rejectioning them until one value finally is accepted. This can be rather efficient if $g(y)$ is sufficiently easy to sample from.

In order for this idea to work, one needs to find $h(y)$. The cumulative probability density is

$$\begin{aligned} P(X \le x) &= P(y \le x | y \text{ is accepted}) \\ &= \frac{P(y \le x \cap y \text{ is accepted})}{P(y \text{ is accepted})} \\ &= \frac{\int_{-\infty}^{x} g(y)h(y)dy}{\int_{-\infty}^{\infty} g(y)h(y)dy} \end{aligned}$$

so the result has distribution

$$\frac{g(x)h(x)}{\int_{-\infty}^{\infty} g(y)h(y)dy}$$

It is therefore wanted that

$$\begin{aligned} f(x) &\propto g(x)h(x) \\ &= g(x) \cdot \frac{f(x)}{g(x)} \cdot \frac{1}{M} \end{aligned}$$

and, for this to be a valid dentsity;

$$
\begin{aligned}
\frac{f(x)}{g(x)\cdot M} &\leq 1 \ , \ \forall x \\
M &\geq \frac{f(x)}{g(x)} \ , \ \forall x
\end{aligned}
$$

The rejection sampling algorithm to sample a value from $f(x)$ then reads

Algorithm 6 ***Rejection sampling***

- *Inititate: For a given density $g(x)$, find $M \geq \max \frac{f(x)}{g(x)}$*
- *Iterate:*
  1. *$y \sim g(y)$*
  2. *$u \sim Uniform(0,1)$*
  3. *If $u \leq \frac{f(y)}{g(y)M}$ then return $x = y$*

For this algorithm to work, the proposal density $g(x)$ must be of the same shape as the density $f(x)$, i.e., if $f(x)$ is symmetric about a location μ, then so must $g(x)$ bes. One does not necessarily need to know normalizing constants when applying this algorithm, since they may be invoked in M, but M should, however, be kept as small as possible since M is the expected number of y's to sample before acception.

C Conditional covariance of a t-distribution

The conditional variance of a new variable conditioned on already known values is for the t-distribution given by

$$
\begin{aligned}
Cov(Z_i|\mathbf{z}_s) &= \frac{\nu + (\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu})}{\nu + m - 3} \left(v_{ii} - \mathbf{V}_{\mathbf{is}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{V}_{\mathbf{si}}\right) \\
&< \frac{\nu + (\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu})}{\nu - 2} \left(v_{ii} - \mathbf{V}_{\mathbf{is}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{V}_{\mathbf{si}}\right) \\
&= \frac{\nu + (\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu})}{\nu - 2} \cdot k \cdot v_{ii}
\end{aligned}
$$

with $0 \leq k \leq 1$ since $0 \leq \mathbf{V}_{\mathbf{is}}\mathbf{V}_{\mathbf{ss}}^{-1}\mathbf{V}_{\mathbf{si}} \leq v_{ii}$. It is evident that $Cov(Z_i|\mathbf{z}_s) < Cov(Z_i)$ if

$$
\frac{\nu + (\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu})}{\nu - 2} \cdot k \cdot v_{ii} \leq \frac{\nu}{\nu - 2} v_{ii}
$$

a condition that is met if

$$
(\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu}) \leq \nu \left(\frac{1}{k} - 1\right)
$$

with $0 \leq \frac{1}{k} - 1 < \infty$. Since extreme values have a great impact on the result and $(\mathbf{z}_s - \boldsymbol{\mu})^T \mathbf{V}_{\mathbf{ss}}^{-1}(\mathbf{z}_s - \boldsymbol{\mu})$ therefore may vary a lot, $Cov(Z_i|\mathbf{z}_s)$ may take on any value if the already sampled values are extreme enough. This means that one cannot generally assume that $Cov(Z_i|\mathbf{z}_s) < Cov(Z_i)$.

D Numerical methods

D.1 Finding a local maxima of *f(x)*

One way of solving this numerical problem is by applying the following algorithm

Algorithm 7 ***Finding a local maxima***

1. *Initiaite: Span out three points with similar distance h along the x-axis.*
2. *Iterate until a satistfactory level of closeness is obtained:*
   - *If $f(x_1) < f(x_2) < f(x_3)$; Eliminate x_1, add $x_3 + h$ and calculate $f(x_3 + h)$*
   - *If $f(x_1) > f(x_2) > f(x_3)$; Eliminate x_3, add $x_1 - h$ and calculate $f(x_1 - h)$*
   - *If $f(x_1) < f(x_3) < f(x_2)$; Eliminate x_1, add $x_2 + \frac{h}{2}$ and calculate $f(x_2 + \frac{h}{2})$*
   - *If $f(x_3) < f(x_1) < f(x_2)$; Eliminate x_3, add $x_2 - \frac{h}{2}$ and calculate $f(x_2 - \frac{h}{2})$*

Obvious question are how to choose the first three points, how to choose the first h and how to measure "closeness". Unfortunately no exact or correct answer can be given to the first two questions. "Closeness", however, should be measured through $f(x)$ if the distribution is very narrow, i.e., σ is small, since small differences in x may give large differences in $f(x)$. If the distribution is very flat close to the maxima, like many likelihood functions are, "closeness" should be measured through x, since small differences in $f(x)$ may give large differences in x.

In the particular problem of Section 6.4.2 $x_2 = \mu$ has been chosen as the original value, and it is then assumed that the other values do not drift too far away from the original location. In the same manner $h = \sigma$, i.e., the scale of every unconditional distribution.

D.2 Finding a zero of *f(x)*

It is known that if x_1 and x_2 lie on each side of a zero $f(x_1) \cdot f(x_2) < 0$, and if x_1 and x_2 lie on the same side of a zero, $f(x_1) \cdot f(x_2) > 0$. This can be used in an algorithm for finding a zero:

Algorithm 8 ***Finding a zero***

1. *Initiate: Pick two points x_1 and $x_2 = x_1 + h$ and decide whether the zero lies to the right, to the left or between these two points.*
2. *Iterate until a crossing with the x-axis is located:*
   While $f(x_1) \cdot f(x_2) > 0$;
   - *Eliminate x_1 by setting $x_1 := x_2$ and $x_2 := x_2 + h$ if the zero lies to the right of x_2*
   - *Eliminate x_2 by setting $x_2 := x_1$ and $x_1 := x_1 - h$ if the zero lies to the left of x_1*
3. *Iterate until a satisfactory level of closeness between x_1 and x_2 on either side of the x-axis is obtained:*
   Let $h := h/2$ and set $x_3 = x_1 + h$
   - *If $f(x_1) \cdot f(x_3) < 0$; Eliminate x_2 by setting $x_2 := x_3$*
   - *If $f(x_2) \cdot f(x_3) < 0$; Eliminate x_1 by setting $x_1 := x_3$*

Obvious questions are of course how to pick x_1 and how to pick h. If h is made too large one may step over a double crossing, and if h is too small the routine will be slow. Neither here no exact or correct answer can be given.

E ML-estimator of V in a t-distribution

When finding the ML-estimator of $\mathbf{V}$ one must take into account what form is wanted on the matrix elements. One can either find the general ML-estimator of $\mathbf{V}$, or one can find it when doing model assumptions. Both cases will be considered here.

E.1 No modelling conditions on V

Let $\mathbf{z}(p)$ denote the p'th element of $\mathbf{z}$ and $\mathbf{IV}(p,q)$ element (p,q) of $\mathbf{V^{-1}}$. Then let $\mathbf{V_{pq}}$ denote the matrix $\mathbf{V}$ with the p'th row and q'th column deleted. In this way $\mathbf{V}(p,q) \notin \mathbf{V_{pq}}$. From the loglikelihood given in expression (17), calculate

$$\frac{\partial l(\boldsymbol{\mu}, \mathbf{V}, \nu|\mathbf{z})}{\partial \mathbf{IV}(p,q)} = \frac{1}{2} \underbrace{\frac{\partial}{\partial \mathbf{IV}(p,q)} \ln|\mathbf{IV}|}_{I} - \frac{\nu+m}{2} \underbrace{\frac{\partial}{\partial \mathbf{IV}(p,q)} \ln[1 + \frac{1}{\nu}D]}_{II} \tag{27}$$

For I

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{IV}(p,q)} \ln|\mathbf{IV}| &= \frac{1}{|\mathbf{IV}|} \frac{\partial}{\partial \mathbf{IV}(p,q)} |\mathbf{IV}| \\
&= \frac{1}{|\mathbf{IV}|} |\mathbf{IV}_{pq}| (-1)^{p+q} \\
&= (\mathbf{IV})^{-1}(p,q) \\
&= \mathbf{V}(p,q)
\end{aligned}$$

and for II

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{IV}(p,q)} \ln[1+\frac{1}{\nu}D] &= \frac{1}{1+\frac{1}{\nu}D}\cdot\frac{1}{\nu}\cdot\frac{\partial}{\partial \mathbf{IV}(p,q)}\left[(\mathbf{z}-\mu)^T\mathbf{V}^{-\mathbf{1}}(\mathbf{z}-\mu)\right] \\
&= \frac{1}{1+\frac{1}{\nu}D}\cdot\frac{1}{\nu}\cdot(\mathbf{z}(p)-\boldsymbol{\mu}(p))(\mathbf{z}(q)-\boldsymbol{\mu}(q)) \\
&= \frac{1}{\nu+D}(\mathbf{z}(p)-\mu)(\mathbf{z}(q)-\mu)
\end{aligned}
$$

Substituting these results back into (27) yields

$$
\begin{aligned}
\frac{\partial l(\mu,\mathbf{V},\nu|\mathbf{z})}{\partial \mathbf{IV}(p,q)} &= \frac{1}{2}\mathbf{V}(p,q)-\frac{\nu+m}{2}\frac{1}{\nu+D}(\mathbf{z}(p)-\mu)(\mathbf{z}(q)-\mu) \\
&= \frac{1}{2}\left[\mathbf{V}(p,q)-\frac{\nu+m}{\nu+D}(\mathbf{z}(p)-\mu)(\mathbf{z}(q)-\mu)\right] \qquad (28)
\end{aligned}
$$

Setting this equal to zero to obtain the maxima yields

$$
\mathbf{V}=\frac{\nu+m}{\nu+D}(\mathbf{z}-\mathbf{1}\mu)(\mathbf{z}-\mathbf{1}\mu)^T
$$

with $D=(\mathbf{z}-\mathbf{1}\mu)^T\mathbf{V}^{-\mathbf{1}}(\mathbf{z}-\mathbf{1}\mu)$. But with $\mathbf{V}$ expressed in this manner, it will be singular. In fact, in order for $\mathbf{V}$ to have a unique solution one needs $n \geq m+1$ with n being the number of realizations of the field, and m the dimension of $\mathbf{V}$. For reference, see [12]. With n realizations the maximum likelihood estimator for $\mathbf{V}$ would take the form

$$
\hat{\mathbf{V}}_{ML}=\frac{1}{n}\sum_{j=1}^{n}\frac{\nu+m}{\nu+D_j}(\mathbf{z}_j-\mathbf{1}\mu)(\mathbf{z}_j-\mathbf{1}\mu)^T
$$

and given that one had $n \geq m+1$, one could use the EM-algorithm to find $\mathbf{V}$ and ν simoultanously. For details, see [13] and [9]. The ML-estimator for ν is calculated in section 7.4.

E.2 Modelling conditions on V

If there are more observations available than there are dimensions on $\mathbf{V}$, one may, as mentioned in the previous section, find the ML-estimator of $\mathbf{V}$. If one thereafter wants to impose some modelling conditions, the already found $\mathbf{V}$ may be used as a starting point to an approximate solution, see Appendix E.3.

If there exists only one observation of the field, which most likely is the case, one may try to find the ML-estimators of β, r and σ directly. The ML-estimator of σ is found in Section 7.3.1, and the remaining parameters are $\alpha=(\beta,r)$ of $\mathbf{W}$. It would then be desireable to perform

$$
\frac{\partial l}{\partial \alpha}\sim\frac{\partial l}{\partial \mathbf{IV}(k,l)}\cdot\frac{\partial \mathbf{IV}(k,l)}{\partial \mathbf{V}(i,j)}\cdot\frac{\partial \mathbf{V}(i,j)}{\partial \alpha}
$$

for the two elements of α. For partial differentiation these differentiations must be performed for all combinations of (i,j) and (k,l), and the full expression for the differentiated loglikelihood is

$$\frac{\partial l}{\partial \alpha} = \sum_{k=1}^{m}\sum_{l=1}^{m}\left[\frac{\partial l}{\partial \mathbf{IV}(k,l)} \cdot \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial \mathbf{IV}(k,l)}{\partial \mathbf{V}(i,j)} \cdot \frac{\partial \mathbf{V}(i,j)}{\partial \alpha}\right]. \tag{29}$$

$\frac{\partial l}{\partial \mathbf{IV}(k,l)}$ is known from expression (28) and $\frac{\partial \mathbf{V}(i,j)}{\partial \alpha}$ is found from differentiating expression (14);

$$\begin{aligned}
\frac{\partial \mathbf{V}(i,j)}{\partial \beta} &= -\frac{3c\sigma^2}{\beta} e^{-3(\Delta_{ij}r)^\beta} (\Delta_{ij}r)^\beta \ln(\Delta_{ij}r)^\beta \\
\frac{\partial \mathbf{V}(i,j)}{\partial r} &= -\frac{3c\sigma^2}{r} e^{-3(\Delta_{ij}r)^\beta} (\Delta_{ij}r)^\beta
\end{aligned}$$

Finally

$$\begin{aligned}
\frac{\partial \mathbf{IV}(k,l)}{\partial \mathbf{V}(i,j)} &= \frac{\partial}{\partial \mathbf{V}(i,j)}\left[\frac{|\mathbf{V_{kl}}|}{|\mathbf{V}|}(-1)^{k+l}\right] \\
&= (-1)^{k+l}\frac{1}{|\mathbf{V}|}\underbrace{\frac{\partial}{\partial \mathbf{V}(i,j)}|\mathbf{V_{kl}}|}_{I} + (-1)^{k+l}|\mathbf{V_{kl}}|\underbrace{\frac{\partial}{\partial \mathbf{V}(i,j)}\frac{1}{|\mathbf{V}|}}_{II}
\end{aligned} \tag{30}$$

If $i = k$ or $j = l$ then $\frac{\partial}{\partial \mathbf{V}(i,j)}|\mathbf{V_{kl}}| = 0$. If $(i,j) \neq (k,l)$

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{V}(i,j)}|\mathbf{V_{kl}}| &= \frac{\partial}{\partial \mathbf{V}(i,j)} \sum_{s=1,s\neq l}^{m} \mathbf{V}(i,s)|\mathbf{V_{ki,ls}}| \\
&= |\mathbf{V_{ki,lj}}|
\end{aligned}$$

with $|\mathbf{V_{ki,lj}}|$ denoting matrix $\mathbf{V}$ with row i and k and column l and j removed. II is calculated as

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{V}(i,j)}\frac{1}{|\mathbf{V}|} &= -|\mathbf{V}|^{-2}\frac{\partial}{\partial \mathbf{V}(i,j)}\sum_{s=1}^{m}\mathbf{V}(i,s)|\mathbf{V_{is}}|(-1)^{i+s} \\
&= \frac{|\mathbf{V_{ij}}|}{|\mathbf{V}|}(-1)^{i+j+1}
\end{aligned}$$

Substituting these results back into expression (30) yields

$$\begin{aligned}
\frac{\partial \mathbf{IV}(k,l)}{\partial \mathbf{V}(i,j)} &= \begin{cases} (-1)^{k+l}\frac{|\mathbf{V_{ki,lj}}|}{|\mathbf{V}|} + (-1)^{k+l+i+j+1}\frac{|\mathbf{V_{kl}}|\mathbf{V_{ij}}}{|\mathbf{V}|^2} & i \neq k, j \neq l \\ (-1)^{k+l+i+j+1}\frac{|\mathbf{V_{kl}}|\mathbf{V_{ij}}}{|\mathbf{V}|^2} & \text{else} \end{cases} \\
&= \begin{cases} \frac{(-1)^{k+l}}{(c\sigma^2)^2}\left[\frac{|\mathbf{W_{ik,jl}}|}{|\mathbf{W}|} + (-1)^{i+j+1}\frac{|\mathbf{W_{kl}}||\mathbf{W_{ij}}|}{|\mathbf{W}|^2}\right] & i \neq k, j \neq l \\ \frac{(-1)^{k+l+i+j+1}}{(c\sigma^2)^2}\frac{|\mathbf{W_{kl}}||\mathbf{W_{ij}}|}{|\mathbf{W}|^2} & \text{else} \end{cases}
\end{aligned}$$

Now that all the factors of expression (29) are found, one may find the theoretical expressions for the ML-estimators of β and r. Differentiating with respect to β and setting equal to zero yields

$$
\begin{aligned}
&\sum_{k=1}^{m}\sum_{l=1}^{m}\Bigg\{\Bigg(\mathbf{V}(k,l)-\frac{\nu+m}{\nu+D}(\mathbf{z}(k)-\mu)(\mathbf{z}(l)-\mu)\Bigg)\\
&\cdot\Bigg[\sum_{i=1,i\neq k}^{m}\ \sum_{j=1,j\neq l}^{m} e^{-3(\Delta_{ij}r)^\beta}(\Delta_{ij}r)^\beta\underbrace{\ln(\Delta_{ij}r)^\beta}_{\delta}\\
&\qquad\qquad\cdot\Bigg((-1)^{k+l}\frac{|\mathbf{W_{ik,jl}}|}{|\mathbf{W}|}+(-1)^{i+j+1}\frac{|\mathbf{W_{kl}}||\mathbf{W_{ij}}|}{|\mathbf{W}|^2}\Bigg)\\
&\quad+\sum_{i=1,i\neq l}^{m}(-1)^{k+i+1}\frac{|\mathbf{W_{kl}}||\mathbf{W_{il}}|}{|\mathbf{W}|^2}e^{-3(\Delta_{il}r)^\beta}(\Delta_{il}r)^\beta\underbrace{\ln(\Delta_{il}r)^\beta}_{\delta}\\
&\quad+\sum_{j=1}^{m}(-1)^{l+j+1}\frac{|\mathbf{W_{kl}}||\mathbf{W_{kj}}|}{|\mathbf{W}|^2}e^{-3(\Delta_{kj}r)^\beta}(\Delta_{kj}r)^\beta\underbrace{\ln(\Delta_{kj}r)^\beta}_{\delta}\Bigg]\Bigg\}=0
\end{aligned}
$$

Doing the same for r yields a similar expression, but without the logarithmic terms δ. It is evident that equations like this is not very useful for generating estimates, both due to the fact that quadruple sums generate an enormous amount of factors even for small matrices, and that the calculation of determinants is rather tedious. If it was possible to isolate β or r on one side, this expression could be of some help, but when it is to be included in a numerical iteration algorithm to solve for σ, β, r and ν simultanously, which one would have to apply, this is not of any practical use.

E.3 Modelling conditions on V. Approximate estimation

In subsection E.1 $\mathbf{V}$ was found with aid of the EM algorithm when no particular form on $\mathbf{V}$ was assumed. In cases where interactional effects are modelled like in, for example, expression (14) when creating new fields, however, this will not be the optimal approach. So when estimating parameters from a given set of data, it would be natural to assume the same interactional effects, and thereby assuming $\mathbf{V}$ to have the form

$$
\mathbf{V}=k\cdot\begin{bmatrix}
1 & e^{-3(\Delta_{12}R^{-1})^\beta} & \cdots & e^{-3(\Delta_{1n}R^{-1})^\beta}\\
e^{-3(\Delta_{21}R^{-1})^\beta} & 1 & \cdots & e^{-3(\Delta_{2n}R^{-1})^\beta}\\
\vdots & & \ddots & \vdots\\
e^{-3(\Delta_{n1}R^{-1})^\beta} & e^{-3(\Delta_{n2}R^{-1})^\beta} & \cdots & 1
\end{bmatrix}
$$

with $\Delta_{pq}=|\mathbf{x_p}-\mathbf{x_q}|$ and $k=c\sigma^2$ being the scale of the univariate distributions. This means that finding the best $\mathbf{V}$ implies finding the optimal R and β from the given data. One knows is that $\mathbf{V}$ is symmetric, since $\Delta_{ij}=\Delta_{ji}$, and that the estimated $\mathbf{V}$ in Appendix E.1, $\hat{\mathbf{V}}_{ML}$, also is symmetric, since $\big((\mathbf{z}-\mathbf{1}\mu)^T(\mathbf{z}-\mathbf{1}\mu)\big)^T=(\mathbf{z}-\mathbf{1}\mu)^T(\mathbf{z}-\mathbf{1}\mu)$, so one may then consider the optimization problem

$$
\mathbf{V}(k,r,\beta)=\hat{\mathbf{V}}_{ML},
$$

with $r = 1/R$, written out as

$$k \cdot \begin{bmatrix} 1 & e^{-3(\Delta_{12}r)^{\beta}} & \cdots & e^{-3(\Delta_{1n}r)^{\beta}} \\ e^{-3(\Delta_{21}r)^{\beta}} & 1 & \cdots & e^{-3(\Delta_{2n}r)^{\beta}} \\ \vdots & & \ddots & \vdots \\ e^{-3(\Delta_{n1}r)^{\beta}} & e^{-3(\Delta_{n2}r)^{\beta}} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix}$$

This is to be optimized with respect to k, r and β. All diagonal elements should be equal, since all variables are identically distributed, and it would therefor be natural to estimate k by the mean value of the diagonal elements, i.e.,

$$k = \frac{1}{n}\sum_{i=1}^{n} v_{ii} = \frac{trace(\hat{\mathbf{V}})}{n}$$

Then introduce the notaion $\frac{v_{ij}}{k} = u_{ij}$ and consider the system

$$\begin{bmatrix} 1 & e^{-3(\Delta_{12}r)^{\beta}} & \cdots & e^{-3(\Delta_{1n}r)^{\beta}} \\ e^{-3(\Delta_{12}r)^{\beta}} & 1 & \cdots & e^{-3(\Delta_{2n}r)^{\beta}} \\ \vdots & & \ddots & \vdots \\ e^{-3(\Delta_{1n}r)^{\beta}} & e^{-3(\Delta_{2n}r)^{\beta}} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & u_{12} & \cdots & u_{1n} \\ u_{12} & 1 & \cdots & u_{2n} \\ \vdots & & \ddots & \vdots \\ u_{1n} & u_{2n} & \cdots & 1 \end{bmatrix}$$

in order to find r and β. The set of data is

$$e^{-3(\Delta_{ij}r)^{\beta}} = v_{ij} \quad , \quad i = 1, \ldots, n \quad , \quad j = i+1, \ldots, n$$

in all $N = \frac{n(n-1)}{2}$ data. Take logarithms on both sides and rewrite it as

$$\begin{aligned} e^{-3(\Delta_{ij}r)^{\beta}} &= v_{ij} && ; i, j \\ \beta \ln(\Delta_{ij}r) &= \ln(\ln v_{ij}^{-1/3}) && ; i, j \\ &= \omega_{ij} && ; i, j \end{aligned}$$

with $\omega_{ij} = \ln(\ln v_{ij}^{-1/3})$. This can be interpreted as if there are multiple observations of systems consisting of one equation in two unknowns. By choosing a value for β one can find $\hat{r}$ as

$$\begin{aligned} \hat{r} &= \bar{r} \\ &= \sum_{l=1}^{N} r_l \\ &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{1}{\Delta_{ij}} e^{\frac{\omega_{ij}}{\beta}} \end{aligned} \tag{31}$$

To check the choice for β one can calculate the square error

$$I_{\beta} = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \left(e^{-3(\Delta_{ij}\hat{r})^{\beta}} - v_{ij}\right)^2$$

The optimal choice for β and r is the combination that minimizes I_β, so a solution to the problem is finding β according to

$$\min_{\beta} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(e^{-3(\Delta_{ij}\hat{r})^{\beta}} - v_{ij} \right)^2 \right\}$$

with $\hat{r}$ as in expression (31).