NTNU
Norwegian University of
Science and Technology

Closed-skew Distributions

Simulation, Inversion and Parameter Estimation

Daniel Høyer Iversen

Master of Science in Physics and Mathematics
Submission date: June 2010
Supervisor: Karl Henning Omre, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

To further develope theory based on the closed-skew Gaussian distribution and if possible generalize it to represent heavy-tailed, skew phenomena.

Particular issues:

- improved simulation algorithm
- improved parameter estimators
- definitions of spatially, stationary models
- generalizations

This MSc-thesis is to be carried out at the Department of Mathematical Science under guidance by Professor Henning Omre.

Assignment given: 18. January 2010
Supervisor: Karl Henning Omre, MATH

Preface

This report is the result of my work in the TMA4905 Master Thesis in the final semester of my Master of Science degree in Industrial Mathematics. The study is performed at the Department of Mathematical Sciences, NTNU, during the spring semester 2010,

I would like to thank my supervisor Professor Henning Omre, for his helpful advice and guidance through the entire year. I would also like to thank my "officemates" Eivind Gausland and Morten Valberg.

Daniel Høyer Iversen
Trondheim, June 14, 2010

Abstract

Bayesian closed-skew Gaussian inversion is defined as a generalization of traditional Bayesian Gaussian inversion. Bayesian inversion is often used in seismic inversion, and the closed-skew model is able to capture the skewness in the variable of interest. Different stationary prior models are presented, but the generalization comes at a cost, simulation from high-dimensional pdfs and parameter inference from data is more complicated. An efficient algorithm to generate realizations from the high-dimensional closed-skew Gaussian distribution is presented. A full-likelihood is used for parameter estimation of stationary prior models under exponential dependence structure. The simulation algorithms and estimators are evaluated on synthetic examples. Also a closed-skew T-distribution is presented to include heavy tails in the pdf and the model is presented with some examples. In the last part the simulation algorithm, the different prior models and parameter estimators are demonstrated on real data from a well in the Sleipner Øst field. The full-likelihood estimator seems to be the best estimator for data with exponential dependence structure

Contents

Chapter 1

Introduction

The objectives of inverse problems are to get information about a parameterized physical system from observational data, based on prior information and theoretical relationships between model parameters and data. Inverse problem theory is used in geophysics to get information about the Earth's properties from physical measurements at the surface. An important inverse problem is found in seismology, where recorded seismic waves at the Earth's surface are used to estimate subsurface parameters. These inverse problems are often multidimensional and strongly affected by noise and measurement uncertainty.

One important subject is prediction of the P-wave impedance, z, in reservoirs from seismic stacked data. The impedance is defined as the product of the P-wave velocity, v_p, and the density, ρ. P-wave is a longitudinal wave, which means that the particles in the material moves parallel to the direction of the travel of the wave energy. The inverted properties are used to get information about reservoir characteristics as lithology-fluid and porosity-permeability. Knowledge about these characteristics are important for development and production of petroleum reservoirs. Oil companies rely on seismic analysis before selecting sites for exploratory oil wells. The seismic waves are generated by an explosion and geophones are used to detect the signal. More information about seismology can be found in Sheriff et al. (1995).

From a statistical point of view the solution of an inverse problem is represented by a posterior probability distribution. The aim of the inversion is not only to obtain a best solution, but also to characterize the uncertainty in the solution. A Bayesian setting makes it possible to combine available prior knowledge with the information from the measured data. In Mosegaard and Tarantola (1995) a model that only can be assessed by iterative algorithms is used. This approach is too computer demanding to be applied in large scale 3D studies. The approach in Buland and Omre (2003) relies on an analytical evaluation, hence it is extremely computer efficient.

The approach in Buland and Omre (2003) relies heavily on Gaussian assump-

tions, so it can be termed Bayesian Gaussian inversion. It is assumed that the prior model of the log-transformed elastic material properties is Gaussian. The log-transform is needed to obtain a linear relation between the seismic data and the elastic properties. A convolved linearized Zoeppritz relation with an additive Gaussian error term defines the likelihood model. These assumptions provides a Gaussian posterior model with model parameters analytical assessable from the prior and the likelihood model. Then the posterior model can be determined for high-dimensional 3D problems, but some numerical approximations may be needed to asses the numerical values of the solution.

In Karimi et al. (2009) the Gaussian assumption in Buland and Omre (2003) is replaced by a closed-skew (CS) Gaussian assumption, which introduce skewness in the model for the variables involved. Under the assumptions of a CS-Gaussian prior also the posterior model will be CS-Gaussian, with model parameters analytically assessable.

The generalization to the CS-Gaussian distribution comes at a cost. The number of parameters in the model is larger than for the Gaussian distribution and the inference from data is more complicated. In this study we focus on the inference of the CS-Gaussian distribution, different stationary priors and parameter estimation. Algorithms to generate realizations from the CS-Gaussian distribution are also studied.

The thesis proceeds as follows: In Chapter 2 the CS-Gaussian distribution is introduced and properties of this distribution are discussed. Bayesian CS-Gaussian inversion is also presented. Further are different algorithms to generate realizations from high dimensional CS-Gaussian distribution given. Different stationary prior models are defined in Chapter 4. Parameter estimation for the different priors is discussed in Chapter 5. In Chapter 6 is a closed-skew T-distribution defined with properties and some examples. In the last part are the simulation algorithms, prior models and parameter estimators tested on real seismic data in a Bayesian setting. Finally we draw some conclusions, and outline possible further work in Chapter 8.

Chapter 2

Methodology

2.1 The CS-Gaussian distribution

The multivariate closed-skew (CS) Gaussian distribution was first introduced in González-Farías et al. (2004). It is a generalization of the Gaussian distribution, and the most important extension is that it adds skewness into the Gaussian model. The most convenient way to define the CS-Gaussian distribution is to consider a random vector $\boldsymbol{t} \in \mathbb{R}^n$ and a random vector $\boldsymbol{v} \in \mathbb{R}^q$. Let the joint probability density function (pdf) of $(\boldsymbol{t}, \boldsymbol{v})$ be multivariate Gaussian as:

$$\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{N}_{n+q} \left(\begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \Sigma_t & \Gamma_{tv} \\ \Gamma_{vt} & \Sigma_v \end{bmatrix} \right), \tag{2.1}$$

where $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_v$ are expectation vectors and Σ_t, Σ_v and Γ_{vt} define covariance matrices with proper dimensions. The multivariate Gaussian distribution of a $\boldsymbol{y} \in \mathbb{R}^k$ with dimension k, expectation $\boldsymbol{\mu}_y \in \mathbb{R}^k$ and covariance $\Sigma_y \in \mathbb{R}^{k \times k}$ is defined by the pdf:

$$\boldsymbol{y} \sim \mathrm{N}_k(\boldsymbol{\mu}_y, \Sigma_y) = \phi_k(\boldsymbol{y}; \boldsymbol{\mu}_y, \Sigma_y) = \frac{1}{(2\pi)^{k/2}|\Sigma_y|^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu}_y)' \Sigma_y^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_y) \right\}.$$

Then the CS-Gaussian variable of interest $\boldsymbol{x} \in \mathbb{R}^n$ is defined as

$$\boldsymbol{x} = [\boldsymbol{t} | \boldsymbol{v} \geq \boldsymbol{0}].$$

Where $\boldsymbol{a} \geq \boldsymbol{b}$ for vectors with elements a_i and b_i is defined as $a_i \geq b_i$ for all elements i. Let $\mathrm{p}(\boldsymbol{x})$ be the generic term for a pdf of the argument, and the term $\mathrm{p}(\boldsymbol{x}|\boldsymbol{d})$ represents the conditional pdf of $\boldsymbol{x}$ given $\boldsymbol{d}$. The pdf of the CS-Gaussian distribution is then given as:

$$\begin{aligned} \boldsymbol{x} \sim \mathrm{p}(\boldsymbol{x}) = p(\boldsymbol{t} | \boldsymbol{v} \geq \boldsymbol{0}) &= \frac{\mathrm{p}(\boldsymbol{v} \geq \boldsymbol{0} | \boldsymbol{t}) \mathrm{p}(\boldsymbol{t})}{\mathrm{p}(\boldsymbol{v} \geq \boldsymbol{0})} \\ &= [1 - \Phi_q(\boldsymbol{0}; \boldsymbol{\mu}_v, \Sigma_v)]^{-1} \left[1 - \Phi_q(\boldsymbol{0}; \boldsymbol{\mu}_{v|t}, \Sigma_{v|t}) \right] \phi_n(\boldsymbol{t}; \boldsymbol{\mu}_t, \Sigma_t), \end{aligned} \tag{2.2}$$

with $\Phi_n(\cdot;\boldsymbol{\mu},\Sigma)$ and $\phi_n(\cdot;\boldsymbol{\mu},\Sigma)$ being the cumulative distribution function (cdf) and the pdf of an n-dimensional Gaussian distribution. So the CS-Gaussian distribution can be seen as a Gaussian distribution, which is skewed by a multiplication of a cumulative Gaussian distribution. The pdf will be unimodal, since it is a product of an unimodal function and a monotone function. The usual parameterization of the CS-Gaussian distribution is

$$\begin{aligned}\mathrm{CSN}_{n,q}\left(\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta\right) &= \psi_{n,q}(\boldsymbol{x};\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta) \\ &= \left[\Phi_q(\mathbf{0};\boldsymbol{\nu},\Delta+\Gamma\Sigma\Gamma')\right]^{-1}\Phi_q(\Gamma(\boldsymbol{x}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta)\phi_n(\boldsymbol{x};\boldsymbol{\mu},\Sigma),\end{aligned} \tag{2.3}$$

with $\boldsymbol{\mu}=\boldsymbol{\mu}_t$, $\Sigma=\Sigma_t$, $\Gamma=\Gamma_{vt}\Sigma_t^{-1}$, $\boldsymbol{\nu}=-\boldsymbol{\mu}_v$, $\Delta=\Sigma_v-\Gamma_{vt}\Sigma_t^{-1}\Gamma_{tv}$ and $\Psi_{n,q}(\boldsymbol{x};\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta)$ being the cdf of the CS-Gaussian distribution. Δ and Σ must be positive definite matrices. Γ is called the skewness parameter, and $\Gamma=0$ entails that the CS-Gaussian distribution coincide with the Gaussian distribution with parameters $(\boldsymbol{\mu},\Sigma)$. Hence the CS-Gaussian distribution should be seen as a generalization of the Gaussian distribution, and not an alternative distribution. The q-parameter may be seen as the degree of skewness freedom in the pdf, and $q=0$ also entails that the CS-Gaussian distribution coincide to the Gaussian distribution. Note also that $\Delta=\Sigma_{v|t}$, where $\Sigma_{v|t}$ is the covariance of $\boldsymbol{v}$ given $\boldsymbol{t}$.

There exists no analytical solution of the Gaussian cdf, $\Phi_q(\cdot;\cdot,\cdot)$, and there is also computationally hard to estimate the value of $\Phi_q(\cdot;\cdot,\cdot)$. Hence it is not possible to obtain an analytical expression for the CS-Gaussian pdf. That cause problems in several cases, and is the underlying problem in this thesis.

The transformation between the CS-Gaussian distribution for $\boldsymbol{x}$ and the Gaussian distribution for $(\boldsymbol{t},\boldsymbol{v})$ is one-to-one given by:

$$\boldsymbol{x}=[\boldsymbol{t}|\boldsymbol{v}\geq 0]\sim\mathrm{CSN}_{n,q}(\boldsymbol{\mu},\Sigma,\Gamma,\boldsymbol{\nu},\Delta)\iff\begin{bmatrix}\boldsymbol{t}\\ \boldsymbol{v}\end{bmatrix}\sim\mathrm{N}_{n+q}\left(\begin{bmatrix}\boldsymbol{\mu}_t\\ \boldsymbol{\mu}_v\end{bmatrix},\begin{bmatrix}\Sigma_t & \Gamma_{tv}\\ \Gamma_{vt} & \Sigma_v\end{bmatrix}\right) \tag{2.4}$$

with

$$\begin{array}{ll}\boldsymbol{\mu}=\boldsymbol{\mu}_t\in\mathbb{R}^n & \boldsymbol{\mu}_t=\boldsymbol{\mu}\in\mathbb{R}^n\\ \Sigma=\Sigma_t\in\mathbb{R}^{n\times n} & \Sigma_t=\Sigma\in\mathbb{R}^{n\times n}\\ \Gamma=\Gamma_{vt}\Sigma_t^{-1}\in\mathbb{R}^{q\times n} & \Gamma_{vt}=\Gamma_{tv}'=\Gamma\Sigma\in\mathbb{R}^{q\times n}\\ \boldsymbol{\nu}=-\boldsymbol{\mu}_v\in\mathbb{R}^q & \boldsymbol{\mu}_v=-\boldsymbol{\nu}\in\mathbb{R}^q\\ \Delta=\Sigma_v-\Gamma_{vt}\Sigma_t^{-1}\Gamma_{tv}\in\mathbb{R}^{q\times q} & \Sigma_v=\Delta+\Gamma\Sigma\Gamma'\in\mathbb{R}^{q\times q}\end{array}\quad\iff$$

2.2 Properties of the CS-Gaussian Distribution

Some favorable characteristics of the CS-Gaussian distribution that are relevant for inversion problems are proved in González-Farías et al. (2004):

1. Linear combinations of components CS-Gaussian random variables are also CS-Gaussian random variables. If $\boldsymbol{x} \sim \mathrm{CSN}_{n,q}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$, A is a deterministic $l \times n$ matrix with $(l \leq n)$ and $\boldsymbol{c}$ is a deterministic l-dimensional vector. Then

$$[\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{c}] \sim \mathrm{CSN}_{l,q}(\boldsymbol{\mu}_y, \Sigma_y, \Gamma_y, \boldsymbol{\nu}_y, \Delta_y),$$

   where

$$\begin{aligned}
\boldsymbol{\mu}_y &= A\boldsymbol{\mu} + \boldsymbol{c}, \\
\Sigma_y &= A\Sigma A', \\
\Gamma_y &= \Gamma\Sigma A'\Sigma_y^{-1}, \\
\boldsymbol{\nu}_y &= \boldsymbol{\nu}, \\
\Delta_y &= \Delta + \Gamma\Sigma\Gamma' - \Gamma\Sigma A'\Sigma_y^{-1}A\Sigma\Gamma'.
\end{aligned}$$

   So the CS-Gaussian distribution is closed under linear transformations.

2. The composition of two independent CS-Gaussian variables will also be CS-Gaussian. If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two independent CS-Gaussian variables with distribution $\boldsymbol{x}_1 \sim \mathrm{CSN}_{n_{x_1},q_{x_1}}(\boldsymbol{\mu}_{x_1}, \Sigma_{x_1}, \Gamma_{x_1}\boldsymbol{\nu}_{x_1}, \Delta_{x_1})$ and
   $\boldsymbol{x}_2 \sim \mathrm{CSN}_{n_{x_2},q_{x_2}}(\boldsymbol{\mu}_{x_2}, \Sigma_{x_2}, \Gamma_{x_2}, \boldsymbol{\nu}_{x_2}, \Delta_{x_2})$, then

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} \sim \mathrm{CSN}_{n,q}(\boldsymbol{\mu}_y, \Sigma_y, \Gamma_y, \boldsymbol{\nu}_y, \Delta_y).$$

Where

$$\begin{aligned} n =& n_{x_1} + n_{x_2}, \\ q =& q_{x_1} + q_{x_2}, \\ \boldsymbol{\mu}_y =& \begin{bmatrix} \boldsymbol{\mu}_{x_1} \\ \boldsymbol{\mu}_{x_2} \end{bmatrix}, \\ \Sigma_y =& \begin{bmatrix} \Sigma_{x_1} & 0 \\ 0 & \Sigma_{x_2} \end{bmatrix}, \\ \Gamma_y =& \begin{bmatrix} \Gamma_{x_1} & 0 \\ 0 & \Gamma_{x_2} \end{bmatrix}, \\ \boldsymbol{\nu}_y =& \begin{bmatrix} \boldsymbol{\nu}_{x_1} \\ \boldsymbol{\nu}_{x_2} \end{bmatrix}, \\ \Delta_y =& \begin{bmatrix} \Delta_{x_1} & 0 \\ 0 & \Delta_{x_2} \end{bmatrix}. \end{aligned}$$

3. CS-Gaussian random variables conditional of the components are also CS-Gaussian variables. Let $\boldsymbol{x} \sim \mathrm{CSN}_{n,q}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$, with $\boldsymbol{x}_1 \in \mathbb{R}^k$ and $\boldsymbol{x}_2 \in \mathbb{R}^{n-k}$ being two subvectors of $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2']'$. Further let the parameters associated with the subvectors be:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Gamma = [\Gamma_1, \Gamma_2].$$

Then

$$[\boldsymbol{x}_1 | \boldsymbol{x}_2] \sim \mathrm{CSN}_{k,q}(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2}, \Gamma_{1|2}, \boldsymbol{\nu}_{1|2}, \Delta_{1|2}).$$

Where

$$\begin{aligned} \boldsymbol{\mu}_{1|2} =& \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \\ \Sigma_{1|2} =& \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \\ \Gamma_{1|2} =& \Gamma_1, \\ \boldsymbol{\nu}_{1|2} =& \boldsymbol{\nu} - \left(\Gamma_2 + \Gamma_1\Sigma_{12}\Sigma_{22}^{-1}\right)(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \\ \Delta_{1|2} =& \Delta. \end{aligned}$$

Note that a result of property 1. is that also marginal pdfs will be CS-Gaussian by choosing $A = \boldsymbol{b}_{(i)}'$ with $\boldsymbol{b}_{(i)}$ an $n \times 1$ vector with entries zeros except for element number i, so that $x_i = \boldsymbol{b}_{(i)}'\boldsymbol{x}$. A result of property 1. and 2. is that a sum of independent CS-Gaussian variables also will be CS-Gaussian.

2.3 Statistics for the CS-Gaussian Distribution

We will here discuss location statistics and spread statistics.

2.3.1 Location statistics

The mode, mean and median are three location statistics that are useful in characterizing the distribution. These statistics can be used for prediction and are for $\boldsymbol{x} \sim \mathrm{p}(\boldsymbol{x})$ defined as:

- Mode, $\mathrm{Mod}(\boldsymbol{x}) = \arg\max_{\boldsymbol{x}} \mathrm{p}(\boldsymbol{x})$.
- Mean, $\mathrm{E}(\boldsymbol{x}) = [\mathrm{E}(x_1), \mathrm{E}(x_2), ..., \mathrm{E}(x_n)]'$, where $\mathrm{E}(x_i) = \int_{x_i} x_i \mathrm{p}(x_i) \mathrm{d}x_i$.
- Median, $\mathrm{Med}(\boldsymbol{x}) = [\mathrm{Med}(x_1), \mathrm{Med}(x_2), ..., \mathrm{Med}(x_n)]'$, where $\mathrm{Med}(x_i)$ is defined by $\int_{-\infty}^{\mathrm{Med}(x_i)} \mathrm{p}(x_i) \mathrm{d}x_i = 0.5$.

For the Gaussian distribution these three statistics are identical, but for a CS-Gaussian distribution they are in general different, due to the skewness in the CS-Gaussian model. These statistics are also analytical obtainable for the Gaussian distribution, since they are given as the model parameter $\boldsymbol{\mu}$. It would be useful to be able to calculate the mean, mode and median for the CS-Gaussian distribution.

Mode. Since we know that the CS-Gaussian pdf is unimodal, the mode is found as the $\boldsymbol{x}$ that satisfies

$$\nabla_x \mathrm{p}(\boldsymbol{x}) = \boldsymbol{0}$$

where

$$\nabla_x = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n} \right]'.$$

Let $\mathrm{p}(\boldsymbol{t}|\boldsymbol{v} \geq \boldsymbol{0})$ be as given in Expression (2.2). The derivation gives

$$\begin{aligned}
\nabla_t \mathrm{p}(\boldsymbol{t}|\boldsymbol{v} \geq \boldsymbol{0}) =& c_1 \int_{\boldsymbol{v}}^{\boldsymbol{0}} \exp\left\{ -\frac{1}{2} (\boldsymbol{v} - \boldsymbol{\mu}_{v|t})' \Sigma_{v|t}^{-1} (\boldsymbol{v} - \boldsymbol{\mu}_{v|t}) \right\} (\boldsymbol{v} - \boldsymbol{\mu}_v - \Gamma_{vt} \Sigma_t^{-1} (\boldsymbol{t} - \boldsymbol{\mu}_t))' \\
& \cdot \Sigma_{v|t}^{-1} \Gamma_{vt} \Sigma_t^{-1} \mathrm{d}\boldsymbol{v} \cdot \exp\left\{ -\frac{1}{2} (\boldsymbol{t} - \boldsymbol{\mu}_t)' \Sigma_t^{-1} (\boldsymbol{t} - \boldsymbol{\mu}_t) \right\} \\
& + c_2 (1 - \Phi(\boldsymbol{0}; \boldsymbol{\mu}_{v|t}, \Sigma_{v|t})) \exp\left\{ -\frac{1}{2} (\boldsymbol{t} - \boldsymbol{\mu}_t)' \Sigma_t^{-1} (\boldsymbol{t} - \boldsymbol{\mu}_t) \right\} (\boldsymbol{t} - \boldsymbol{\mu}_t) \Sigma_t^{-1} \\
=& \boldsymbol{0}.
\end{aligned}$$

It is hard to solve this equation with respect to $\boldsymbol{t}$, hence an analytical solution for the mode is complicated to obtain. Numerical methods can be used to solve this problem. For example can stochastic simulation be used, by generating realizations from the CS-Gaussian distribution. Then choose the area with the highest density of points as the mode. More traditionally optimization methods are complicated to use, due to the difficulty of analytically compute the value of the CS-Gaussian pdf. These methods are not investigated any further in this project.

Mean. The mean is given as the first moment. If $\boldsymbol{x} \sim \mathrm{CSN}_{n,q}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$, then the moment generating function of $\boldsymbol{x}$ is given in González-Farías et al. (2004) as

$$\mathrm{M}_x(\boldsymbol{s}) = \frac{\Phi_q(\Gamma\Sigma\boldsymbol{s}; \boldsymbol{v}, \Delta + \Gamma\Sigma\Gamma')}{\Phi_q(\boldsymbol{0}; \boldsymbol{v}, \Delta + \Gamma\Sigma\Gamma')} \cdot \exp\left\{\boldsymbol{s}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{s}'\Sigma\boldsymbol{s}\right\}, \qquad \boldsymbol{s} \in \mathbb{R}^n.$$

The mean of the CS-Gaussian distribution is then given as

$$\mathrm{E}(\boldsymbol{x}) = \nabla_s M_x(\boldsymbol{s})|_{\boldsymbol{s}=\boldsymbol{0}}\,.$$

Then

$$\begin{aligned}\nabla_s \mathrm{M}_x(\boldsymbol{s}) =& \Sigma\Gamma' \frac{[\nabla_s \Phi_q(\boldsymbol{s}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')]'}{\Phi_q(\boldsymbol{0}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')} \cdot \exp\left\{\boldsymbol{s}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{s}'\Sigma\boldsymbol{s}\right\} \\ &+ \frac{\Phi_q(\Gamma\Sigma\boldsymbol{s}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')}{\Phi_q(\boldsymbol{0}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')} \cdot \exp\left\{\boldsymbol{s}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{s}'\Sigma\boldsymbol{s}\right\} \cdot (\boldsymbol{\mu} + \Sigma\boldsymbol{s})\end{aligned}$$

So the mean value of $\boldsymbol{x}$ is then

$$\nabla_s \mathrm{M}_x(\boldsymbol{s})|_{s=0} = \boldsymbol{\mu} + \Sigma\Gamma'\boldsymbol{\eta}, \tag{2.5}$$

where

$$\boldsymbol{\eta} = \left.\frac{[\nabla_s \Phi_q(\boldsymbol{s}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')]'}{\Phi_q(\boldsymbol{0}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')}\right|_{s=0}$$

(Dominguez-Molina et al. 2003). For a high dimensional problem it is complicated to compute the mean, since it does not exits any analytical solution for $\Phi_q(\cdot;\cdot,\cdot)$. But for $q = 1$ the mean will be

$$\mathrm{E}(x) = \mu + \Sigma\Gamma' \frac{\phi_1(0; \nu, \Delta + \Gamma\Sigma\Gamma')}{\Phi_1(0; \nu, \Delta + \Gamma\Sigma\Gamma')}.$$

To obtain the mean for a higher dimensional problem a numerical method can be used to compute Expression (2.5). Alternatively, $\mathrm{E}(x_i) = \int_{x_i} x_i \mathrm{p}_i(x_i)\mathrm{d}x_i$ can be approximated by a numerical method. For example by stochastic integration as

$$\widehat{\mathrm{E}(x_i)} = \frac{1}{N}\sum_{j=1}^{N} x_i^{(j)},$$

where $x_i^{(j)}$ is realization number j of element x_i from simulation of the CS-Gaussian distribution.

Median. The median is often used as location measure instead of the mean. The median is not simple to obtain analytically, since it requires an inversion of the cdf for the CS-Gaussian distribution. Therefore it is necessary to predict the median with numerically methods. For example by

$$\widehat{\mathrm{Med}(x_i)} = \mathrm{Med}\left\{x_i^{(1)}, x_i^{(2)}, ..., x_i^{(j)}, ..., x_i^{(N)}\right\},$$

where $x_i^{(j)}$ is realization number j of element x_i from the CS-Gaussian distribution.

It is possible to predict the mean and the median from each marginal distribution. But since the marginal distribution depends of the full $\boldsymbol{v}$-vector it will be faster to generate $\boldsymbol{x}^{(j)}$ from the joint distribution.

2.3.2 Spread statistics

The covariance and the quantiles are two spread statistics that are suitable for characterizing the variability in the model. These statistics can be used to describe the uncertainty in the predicted values and are defined as:

- Covariance, $\mathrm{Cov}(\boldsymbol{x}) = \mathrm{E}\left((\boldsymbol{x} - \mathrm{E}(\boldsymbol{x})(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))'\right)$.
- Quantiles, the α-quantile, $\mathrm{Q}_\alpha(\boldsymbol{x}) = [\mathrm{Q}_\alpha(x_1), \mathrm{Q}_\alpha(x_2), ..., \mathrm{Q}_\alpha(x_n)]'$, where $\mathrm{Q}_\alpha(x_i)$ is defined by $\int_{-\infty}^{\mathrm{Q}_\alpha(x_i)} \mathrm{p}(x_i)\mathrm{d}x_i = \alpha$.

For the Gaussian distribution the covariance is given as the model parameter, Σ, while the quantiles must be found by numerical methods. It would be useful to be able to calculate these spread statistics for the CS-Gaussian distribution.

Covariance. The covariance is given as $\mathrm{Cov}(\boldsymbol{x}) = \mathrm{E}\left((\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x})'\right) = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}') - \mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x}')$. So the covariance can be discussed from the first and second moment. The first moment is already found and the second moment can be calculated as

$$\begin{aligned}\mathrm{E}(\boldsymbol{x}\boldsymbol{x}') =& \nabla_s^2 \mathrm{M}_x(\boldsymbol{s})|_{\boldsymbol{s}=\boldsymbol{0}} \\ =& \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\eta}'\Gamma\Sigma + \Sigma\Gamma\boldsymbol{\eta}\boldsymbol{\mu}' + \Sigma\Gamma'\Lambda\Gamma\Sigma,\end{aligned}$$

where

$$\boldsymbol{\Lambda} = \left. \frac{\nabla_s \nabla_s' \Phi_q(\boldsymbol{s}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')}{\Phi_q(\mathbf{0}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')} \right|_{s=0}.$$

So the covariance of $\boldsymbol{x}$ is given as

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{x}) =& \mathrm{E}(\boldsymbol{x}\boldsymbol{x}') - \mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x}') \\ =& \Sigma + \Sigma\Gamma'\Lambda\Gamma\Sigma - \Sigma\Gamma'\boldsymbol{\eta}\boldsymbol{\eta}'\Gamma\Sigma \end{aligned}$$

(Dominguez-Molina et al. 2003). It is not possible to compute the covariance analytically for a high dimensional problem, so we need to use numerical approximations.

Quantiles. First note that $\mathrm{Q}_{0.5}(\boldsymbol{x})$ is identical as the median. So inversion of the CS-Gaussian cdf is required to obtain the quantiles. As for the median it seems that a numerical approximation is best for computing the quantiles. Also here it is natural to use stochastic simulation to calculate the quantiles.

2.4 Examples of CS-Gaussian Distributions

The CS-Gaussian distribution is closely related to the Gaussian distribution and has therefore also many similar properties. To illustrate the characteristics of the CS-Gaussian distribution some examples of the CS-Gaussian distribution will be shown. We will first consider a CS-Gaussian $x \in \mathbb{R}^1$, with a base case as

$$\mu = 5, \quad \Sigma = 9, \quad \Gamma = 1, \quad \nu = 0, \quad \Delta = 0.05. \tag{2.6}$$

Note that Σ, Γ and Δ here are one dimensional matrices.

From Expression (2.3) we know that the CS-Gaussian distribution is proportional to a product of a Gaussian pdf with variance Σ and a Gaussian cdf with variance Δ. A lower variance, Δ, will give more skewness in the model. A higher variance, Σ, for the Gaussian pdf will also lead to more skewness. This can be demonstrated by considering the base case, but with three different values for Σ and Δ:

$$\begin{aligned} \Sigma_1 = 9, \quad \Delta_1 = 0.05 \\ \Sigma_2 = 6, \quad \Delta_2 = 0.5 \\ \Sigma_3 = 3, \quad \Delta_3 = 5. \end{aligned}$$

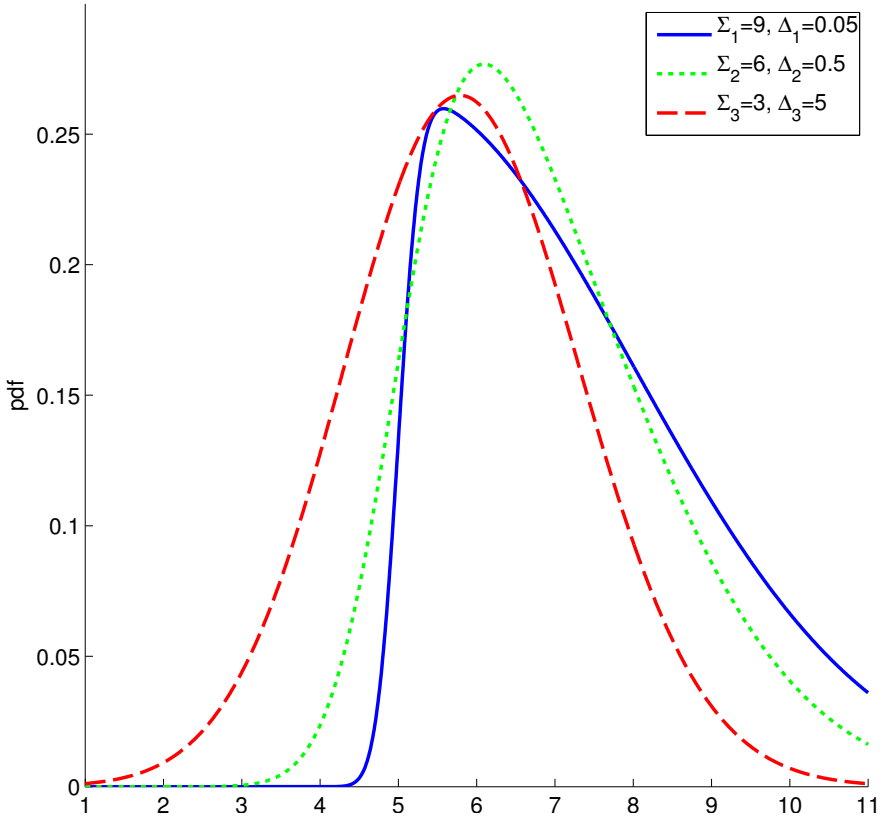


Figure 2.1: Pdf for different Σ and Δ, and $\mu = 5$, $\Gamma = 1$ and $\nu = 0$.

The pdf for these three cases are displayed in Figure 2.1.

The skewness in the model is not only determined by Δ and Σ, it is also influenced by ν. For Δ very small and Σ very large will there still not be any skewness in the model if $\nu = -\infty$. This effect is demonstrated by considering a CS-Gaussian x with the parameters given in Expression (2.6), but with three different values of ν:

$$\begin{aligned} \nu_1 &= 0 \\ \nu_2 &= -3 \\ \nu_3 &= -9. \end{aligned}$$

The three different pdfs are displayed in Figure 2.2. In the first case the pdf is very skewed and in the third case the pdf is almost Gaussian. From these two examples we observe that higher Σ will give longer tail on the right side and a small Δ will give a short left tail. The CSN pdfs can be seen as a Gaussian pdf that here is cut

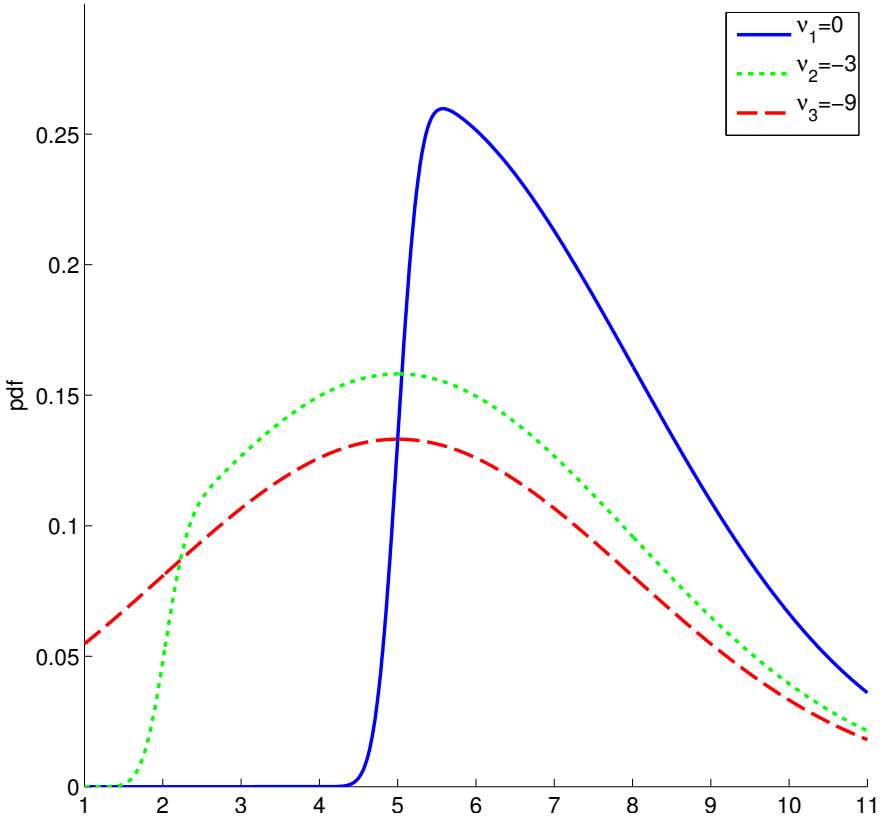


Figure 2.2: Pdf for different ν, and $\mu = 5$, $\Sigma = 9$, $\Gamma = 1$ and $\Delta = 0.05$.

off at the left side. The position of the cut off is decided by ν and the sharpness is decided by Δ.

The parameter μ is the only parameter that not will influence the skewness in the model. In an equivalent way as for the μ in the Gaussian distribution will the μ in the CS-Gaussian distribution only shift the location of the distribution. The translation effect is demonstrated by considering a CS-Gaussian x with the parameters given in Expression (2.6), but with three different values of μ:

$$\mu_1 = 5$$
$$\mu_2 = 7$$
$$\mu_3 = 2.$$

All these three pdfs are identical except for a shift in the density, as displayed in Figure 2.3.

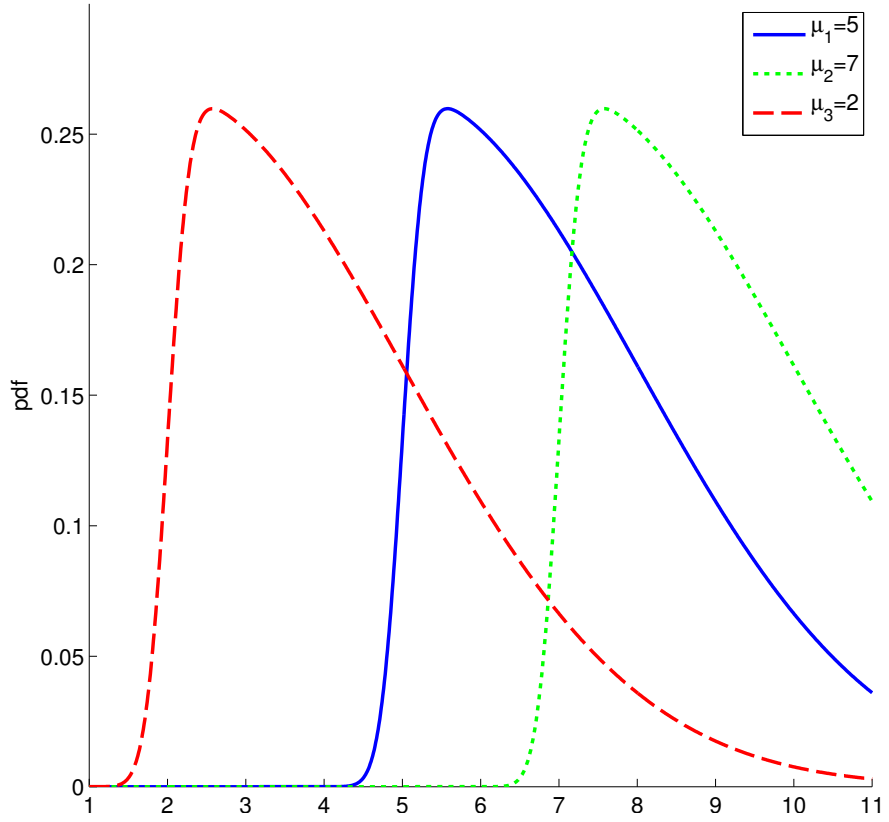


Figure 2.3: Pdf for different μ, and $\Sigma = 9$, $\Gamma = 1$, $\nu = 0$ and $\Delta = 0.05$.

The same situations will occur for higher dimensional problems, and we will here also study a particular CS-Gaussian distributed variable $\boldsymbol{x}$ with $n = q = 2$, and parameters

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 4 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}, \quad \boldsymbol{\nu}_1 = \begin{bmatrix} -2 \\ 6 \end{bmatrix}, \quad \Delta_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The pdf of this two dimensional CS-Gaussian distribution is shown in Figure 2.4. Another example is shown in Figure 2.5 with

$$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1, \quad \Sigma_2 = \Sigma_1, \quad \Gamma_2 = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{\nu}_2 = \boldsymbol{\nu}_1, \quad \Delta_2 = \Delta_1.$$

Note that the only difference between these two examples is one element in Γ_2. It is observed from Figure 2.4 and Figure 2.5 that there is more skewness in the first example. By setting one diagonal element in Γ_2 to zero makes the element

t_2 almost uncorrelated with $\boldsymbol{v}$, hence will the marginal of t_2 be almost Gaussian with mean 7. This shows that the Γ controls the skewness in the model. It is also observed that this small change in the Γ also change the location of the distribution in the x_2-dimension. In the first example the mode is about $(5, 9)$ and in the second example is it about $(5, 7)$.

These examples show that Σ, Γ, $\boldsymbol{\nu}$ and Δ all influence the location, spread and skewness in the model. This makes the inference in the model quite complicated.

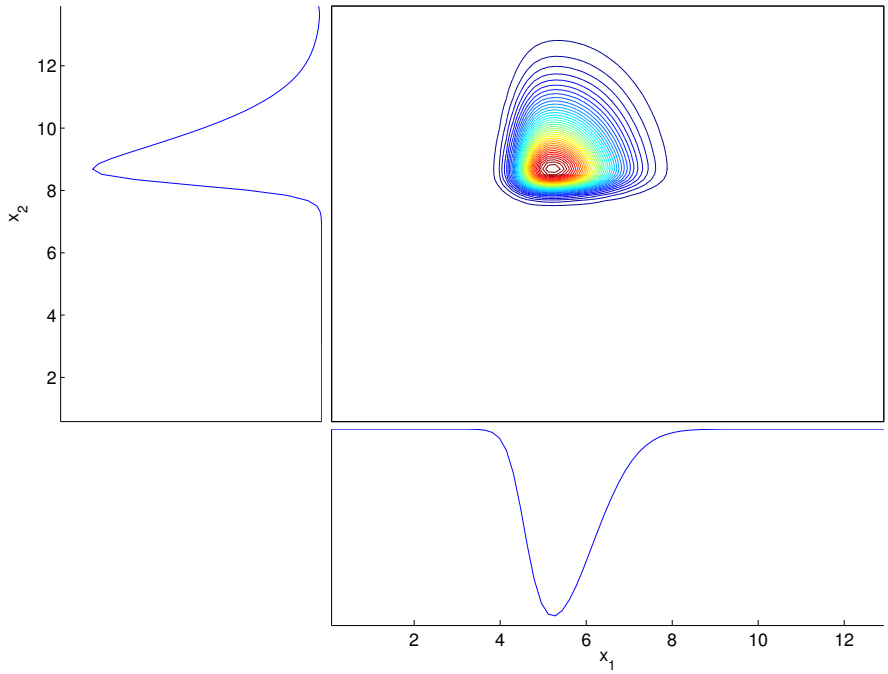


Figure 2.4: First bivariate CS-Gaussian example. Bivariate pdf and marginal pdfs.

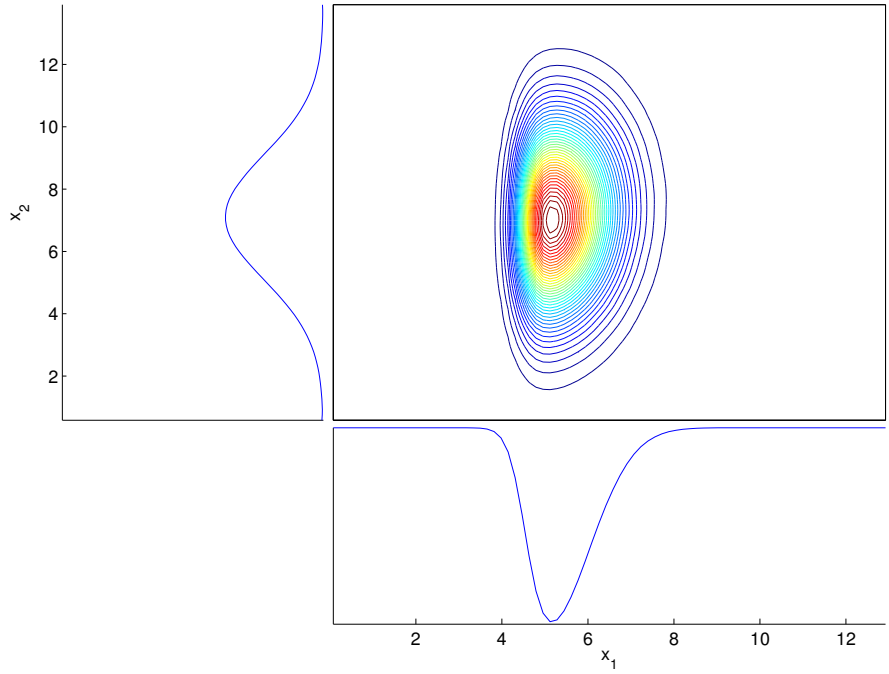


Figure 2.5: Second bivariate CS-Gaussian example. Bivariate pdf and marginal pdfs.

2.5 Bayesian CS-Gaussian Inversion

Let $\boldsymbol{x} \in \mathbb{R}^{n_x}$ be the variable of interest and consider a linear model

$$[\boldsymbol{d}|\boldsymbol{x}] = H\boldsymbol{x} + \boldsymbol{e}, \tag{2.7}$$

where $\boldsymbol{d} \in \mathbb{R}^{n_d}$ is measured data with error $\boldsymbol{e} \in \mathbb{R}^{n_d}$. The object is to assess $\boldsymbol{x}$ given $\boldsymbol{d}$, and there exists several methods to do this. Here we will consider Bayesian inversion, so we assume a probability distribution for $\boldsymbol{x}$ and $\boldsymbol{e}$. In Bayesian inversion we are interested in the probability distribution of an $\boldsymbol{x}$ given some observed data $\boldsymbol{d}$. Bayes' rule gives the posterior model as

$$\mathrm{p}(\boldsymbol{x}|\boldsymbol{d}) = c \cdot \mathrm{p}(\boldsymbol{d}|\boldsymbol{x}) \cdot \mathrm{p}(\boldsymbol{x}),$$

where c is a constant, $\mathrm{p}(\boldsymbol{d}|\boldsymbol{x})$ is the likelihood and $\mathrm{p}(\boldsymbol{x})$ represents the prior knowledge about the variable of interest. The prior model for $\boldsymbol{x}$ is defined as

$$\boldsymbol{x} \sim \mathrm{CSN}_{n_x,q_x}(\boldsymbol{\mu}_x, \Sigma_x, \Gamma_x, \boldsymbol{\nu}_x, \Delta_x).$$

The error, $\boldsymbol{e}$, is assumed to be independent of $\boldsymbol{x}$ and distributed as

$$\boldsymbol{e} \sim \mathrm{p}(\boldsymbol{e}) = \mathrm{CSN}_{n_d,q_d}(\mathbf{0}, \Sigma_e, \Gamma_e, \mathbf{0}, I),$$

where I is the identity matrix of proper dimension. This gives

$$[\boldsymbol{d}|\boldsymbol{x}] \sim \mathrm{p}(\boldsymbol{d}|\boldsymbol{x}) = \mathrm{CSN}_{n_d,q_d}(H\boldsymbol{x}, \Sigma_e, \Gamma_e, \mathbf{0}, I).$$

Then the posterior model $p(\boldsymbol{x}|\boldsymbol{d})$ is given by

$$[\boldsymbol{x}|\boldsymbol{d}] \sim p(\boldsymbol{x}|\boldsymbol{d}) = \mathrm{CSN}_{n_x,q_x+q_d}(\boldsymbol{\mu}_{x|d}, \Sigma_{x|d}, \Gamma_{x|d}, \boldsymbol{\nu}_{x|d}, \Delta_{x|d}).$$

Where the parameters are given in Karimi et al. (2009) as

$$\begin{aligned}
\boldsymbol{\mu}_{x|d} =& \boldsymbol{\mu}_x + \Sigma_x H'[H\Sigma_x H' + \Sigma_e]^{-1}(\boldsymbol{d} - H\boldsymbol{\mu}_x) \\
\Sigma_{x|d} =& \Sigma_x - \Sigma_x H'[H\Sigma_x H' + \Sigma_e]^{-1} H\Sigma_x \\
\Gamma_{x|d} =& \left[\begin{bmatrix} \Gamma_x \Sigma_x \\ 0 \end{bmatrix} - \begin{bmatrix} \Gamma_x \Sigma_x H' \\ \Gamma_e \Sigma_e \end{bmatrix} [H\Sigma_x H' + \Sigma_e]^{-1} H\Sigma_x \right] \Sigma_{x|d}^{-1}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\nu}_{x|d} =& \begin{bmatrix} \boldsymbol{\nu}_x \\ 0 \end{bmatrix} - \begin{bmatrix} \Gamma_x \Sigma_x H' \\ \Gamma_e \Sigma_e \end{bmatrix} [H\Sigma_x H' + \Sigma_e]^{-1} (\boldsymbol{d} - H\boldsymbol{\mu}_x) \\
\Delta_{x|d} =& \begin{bmatrix} \Delta_x + \Gamma_x \Sigma_x \Gamma_x' & 0 \\ 0 & I + \Gamma_e \Sigma_e \Gamma_e' \end{bmatrix} - \begin{bmatrix} \Gamma_x \Sigma_x H' \\ \Gamma_e \Sigma_e \end{bmatrix} [H\Sigma_x H' + \Sigma_e]^{-1} \begin{bmatrix} \Gamma_x \Sigma_x H' \\ \Gamma_e \Sigma_e \end{bmatrix}' \\
& - \Gamma_{x|d} \Sigma_{x|d} \Gamma_{x|d}'.
\end{aligned}$$

This is a generalization of the traditional Bayesian Gaussian inversion, which is presented in Mosegaard and Tarantola (1995) and in Buland and Omre (2003). This can be seen by setting $\Gamma_e = 0$ in the likelihood and $\Gamma_x = 0$ in the prior model. In Bayesian Gaussian inversion the prior is Gaussian and the likelihood is Gauss-linear. Then the posterior model will be Gaussian with parameters $\boldsymbol{\mu}_{x|d}$ and $\Sigma_{x|d}$. The posterior model, $\mathrm{p}(\boldsymbol{x}|\boldsymbol{d})$, can be used to predict $\boldsymbol{x}$ based on the data, $\boldsymbol{d}$.

All the three location statistics presented in Section 2.3 can be used as predictors for $\boldsymbol{x}$ given the data $\boldsymbol{d}$, but we have chosen to use the median. Hence the predictor of $\boldsymbol{x}$ is given as $\widehat{[\boldsymbol{x}|\boldsymbol{d}]} = \mathrm{Med}\{\boldsymbol{x}|\boldsymbol{d}\} = \mathrm{Q}_{0.5}\{\boldsymbol{x}|\boldsymbol{d}\}$. The median is, as argued for in Karimi et al. (2009), a good choice since it is logically related to the natural definition of the $(1-\alpha)$-prediction interval, $[\mathrm{Q}_{\alpha/2}\{\boldsymbol{x}|\boldsymbol{d}\}, \mathrm{Q}_{1-\alpha/2}\{\boldsymbol{x}|\boldsymbol{d}\}]$. The median is also a linear operator with respect to any monotone function $\mathrm{h}(\cdot)$,

i.e. $\mathrm{Q}_{0.5}\{\mathrm{h}(\boldsymbol{x}|\boldsymbol{d})\} = \mathrm{h}(\mathrm{Q}_{0.5}\{\boldsymbol{x}|\boldsymbol{d}\})$. Then for example $[\widehat{\boldsymbol{z}|\boldsymbol{d}}] = \exp\{\widehat{\boldsymbol{x}|\boldsymbol{d}}\}$, where $\boldsymbol{z} = \exp\{\boldsymbol{x}\}$. The median as a predictor is also robust with respect to deviations in the tail of the posterior model $\mathrm{p}(\boldsymbol{x}|\boldsymbol{d})$. Note that the median is frequently used as a predictor for skew distributions instead of the mean, since the mean is less robust towards long tails than the median.

2.5.1 Center of Likelihood Model.

The error term, $\boldsymbol{e}$, of the likelihood model is not centered for all choices of parameters Σ_e and Γ_e. It is normal to assume an error term with location measure zero, and with $\Gamma_e = 0$ the distribution is reduced to a Gaussian distribution with location zero. Another way to make sure that the error term has zero median is to introduce a centered error, $\boldsymbol{e}^c$. The centered error term has the same parameters Σ_e and Γ_e as $\boldsymbol{e}$, but includes an additional $\boldsymbol{\mu}_{e^c}$. The $\boldsymbol{\mu}_{e^c}$ parameter should translate the distribution such that median will be zero, and is therefore defined as

$$\boldsymbol{\mu}_{e^c} = -\mathrm{Med}(\boldsymbol{e}).$$

The median of $\boldsymbol{e}$ must be obtained by a numerical method.

Chapter 3

Simulation

Simulation based inference of the CS-Gaussian distribution will often be required. In order to estimate the median and the quantiles as seen in Section 2.3 it is useful to generate random values, $\boldsymbol{x}$, from the multivariate CS-Gaussian distribution. Even to make density plots as shown in Figure 2.4 and Figure 2.5, it is useful to be able to simulate from the CS-Gaussian distribution, since it is computationally hard to compute the density directly. In this chapter three different algorithms to generate realizations from the multivariate CS-Gaussian distribution will be presented. An empirical study will be done to compare the algorithms in the end of the chapter.

3.1 Algorithm A: Brute Force Simulation

The simplest way to simulate, $\boldsymbol{x}$, from a CS-Gaussian distribution is to generate $(\boldsymbol{t}, \boldsymbol{v})$ from the multivariate Gaussian distribution as given in Expression (2.1) with rejection sampling until $\boldsymbol{v} \geq \boldsymbol{0}$.

Algorithm A: Brute Force Simulation

- **Repeat until** $\boldsymbol{v} \geq \boldsymbol{0}$
  Generate $\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{N}_{n+q}\left(\begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \Sigma_t & \Gamma_{tv} \\ \Gamma_{vt} & \Sigma_v \end{bmatrix} \right)$
- $\boldsymbol{x} \leftarrow \boldsymbol{t}$

Then $\boldsymbol{x}$ will be a realization of the CS-Gaussian distribution.

This algorithm is not efficient for $\boldsymbol{\mu}_v \ll \boldsymbol{0}$ and elements in $\boldsymbol{v}$ highly negatively correlated, since it is very unlikely to generate $v_i \geq 0; i = 1, ..., q$ at the same time. When $\boldsymbol{v}$ is high dimensional it will be almost impossible to generate all $v_i \geq 0; i = 1, ..., q$ simultaneously.

3.2 Algorithm B: Two Steps Simulation

A more efficient algorithm is to first generate $\boldsymbol{v}$ until $\boldsymbol{v} \geq \mathbf{0}$, and then generate $\boldsymbol{t}$ conditional on $\boldsymbol{v}$. Consider the joint distribution $\mathrm{p}(\boldsymbol{t}, \boldsymbol{v})$, then

$$\mathrm{p}(\boldsymbol{v}|\boldsymbol{v} \geq \mathbf{0}) = \frac{\mathrm{p}(\boldsymbol{v}, \boldsymbol{v} \geq \mathbf{0})}{\mathrm{p}(\boldsymbol{v} \geq \mathbf{0})} = c \cdot \mathrm{p}(\boldsymbol{v}) \cdot \mathrm{I}(\boldsymbol{v} \geq \mathbf{0}).$$

Where $c = [\mathrm{p}(\boldsymbol{v} \geq \mathbf{0})]^{-1}$ is a constant and $\mathrm{I}(\cdot)$ an indicator function taking value 1 if the argument is true and 0 otherwise. Further we have that

$$\mathrm{p}(\boldsymbol{t}, \boldsymbol{v}|\boldsymbol{v} \geq \mathbf{0}) = \frac{\mathrm{p}(\boldsymbol{t}, \boldsymbol{v}, \boldsymbol{v} \geq \mathbf{0})}{\mathrm{p}(\boldsymbol{v} \geq \mathbf{0})} = c \cdot \mathrm{p}(\boldsymbol{t}, \boldsymbol{v}) \cdot \mathrm{I}(\boldsymbol{v} \geq \mathbf{0}).$$

Then the distribution of $\boldsymbol{x}$ can be written as

$$\begin{aligned}
\boldsymbol{x} = [\boldsymbol{t}|\boldsymbol{v} \geq \mathbf{0}] \sim \mathrm{p}(\boldsymbol{t}|\boldsymbol{v} \geq \mathbf{0}) &= \int_{-\infty}^{\infty} \mathrm{p}(\boldsymbol{t}, \boldsymbol{v}|\boldsymbol{v} \geq \mathbf{0}) \mathrm{d}\boldsymbol{v} \\
&= \int_{-\infty}^{\infty} c \cdot \mathrm{p}(\boldsymbol{t}, \boldsymbol{v}) \cdot \mathrm{I}(\boldsymbol{v} \geq \mathbf{0}) \mathrm{d}\boldsymbol{v} \\
&= \int_{0}^{\infty} c \cdot \mathrm{p}(\boldsymbol{t}, \boldsymbol{v}) \mathrm{d}\boldsymbol{v}.
\end{aligned}$$

We want to generate $\boldsymbol{x}$ from this distribution, and that will be the same as first generate

$$\boldsymbol{v} \sim \mathrm{p}(\boldsymbol{v}|\boldsymbol{v} \geq \mathbf{0}),$$

and then generate $\boldsymbol{x}$ as

$$\boldsymbol{x} = [\boldsymbol{t}|\boldsymbol{v}] \sim \mathrm{p}(\boldsymbol{t}|\boldsymbol{v}).$$

This is valid since $\boldsymbol{x}$ then is generated from

$$\begin{aligned}
\boldsymbol{x} \sim \int_{-\infty}^{\infty} \mathrm{p}(\boldsymbol{t}|\boldsymbol{v}) \cdot \mathrm{p}(\boldsymbol{v}|\boldsymbol{v} \geq \mathbf{0}) \mathrm{d}\boldsymbol{v} &= \int_{-\infty}^{\infty} \mathrm{p}(\boldsymbol{t}|\boldsymbol{v}) \cdot c \cdot \mathrm{p}(\boldsymbol{v}) \cdot \mathrm{I}(\boldsymbol{v} \geq \mathbf{0}) \mathrm{d}\boldsymbol{v} \\
&= \int_{0}^{\infty} c \cdot \mathrm{p}(\boldsymbol{t}, \boldsymbol{v}) \mathrm{d}\boldsymbol{v} \\
&= \mathrm{p}(\boldsymbol{t}|\boldsymbol{v} \geq \mathbf{0}).
\end{aligned}$$

Since $\mathrm{p}(\boldsymbol{t}, \boldsymbol{v})$ is Gaussian will also $\mathrm{p}(\boldsymbol{t}|\boldsymbol{v})$ be Gaussian with mean and variance given as

$$\begin{aligned}
\mathrm{E}(\boldsymbol{t}|\boldsymbol{v}) &= \boldsymbol{\mu} + \Gamma_{tv}\Sigma_v^{-1}(\boldsymbol{v} + \boldsymbol{\nu}) \\
\mathrm{Var}(\boldsymbol{t}|\boldsymbol{v}) &= \Sigma_t - \Gamma_{tv}\Sigma_v^{-1}\Gamma_{tv}'.
\end{aligned}$$

Then we can first generate $\boldsymbol{v}$ until all elements $v_i; i = 1, ..., q$ is greater than zero, and so generate $\boldsymbol{t}$ from the corresponding Gaussian distribution. Hence we reduce the rejection problem to a lower dimension. The algorithm is:

Algorithm B: Two Steps Simulation

- **Repeat until** $\boldsymbol{v} \geq \mathbf{0}$
  Generate $\boldsymbol{v} \sim \mathrm{N}_q(-\boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma)$
- $\boldsymbol{x} \leftarrow \boldsymbol{t}|\boldsymbol{v} \sim \mathrm{N}_n(\boldsymbol{\mu} + \Gamma_{tv}\Sigma_v^{-1}(\boldsymbol{v} + \boldsymbol{\nu}), \Sigma_t - \Gamma_{tv}\Sigma_v^{-1}\Gamma_{tv}')$

Then $\boldsymbol{x}$ will be a realization of the CS-Gaussian distribution.

This is a more efficient algorithm than the first algorithm, since we generate q values instead of $n + q$ values for each step. But also this algorithm requires that $v_i \geq 0$ for all i happen simultaneously.

3.3 Algorithm C: Iterative Simulation

An even more efficient method can be constructed by sequential rejection sampling combined with an MCMC approach. We will use the same principle as in the previous algorithm, by first generate $\boldsymbol{v} \geq \mathbf{0}$ and then generate $\boldsymbol{t}$. But instead of generating all $v_i \geq 0; i = 1, ..., q$ simultaneously will each v_i be generated sequentially by an iterative method.

To generate $v_i \geq 0$ a rejection sampling approach, as described in Robert (1995) is used. First we will present the general rejection algorithm to generate realizations from $\mathrm{h}(w)$. Choose a distribution $\mathrm{g}(w)$ to be similar to $\mathrm{h}(w)$, but such that it is easier to simulate from. Choose c to be a constant such that $\mathrm{h}(w) \leq c \cdot \mathrm{g}(w)$ for every w in the support of $\mathrm{h}(w)$. Then should a w be generated from $\mathrm{g}(w)$ and an u from the uniform distribution between 0 and 1, $\mathrm{U}_1(0, 1)$. The w will be a realization from $\mathrm{h}(w)$, when the w and u satisfies $u \leq \mathrm{h}(w)/(c \cdot \mathrm{g}(w))$.

Let $\mathrm{h}(v; \mu_+, \sigma_+^2, v_0)$ be the truncated univariate Gaussian distribution with truncation point v_0, then the density will be

$$\mathrm{h}(v; \mu_+, \sigma_+^2, v_0) = \mathrm{N}_1^+(\mu_+, \sigma_+^2, v_0) = \begin{cases} \frac{\exp(-(v-\mu_+)^2/2\sigma_+^2)}{\sqrt{2\pi}\sigma_+(1-\Phi((v_0-\mu_+)/\sigma_+))} & , v \geq v_0 \\ 0 & , v < v_0. \end{cases}$$

Where $v_0 \to -\infty$ gives the Gaussian distribution with mean μ_+ and variance σ_+^2. In the rest of the algorithm description we assume with out loss of generality that

$\mu_+ = 0$ and $\sigma_+^2 = 1$. Since v_0 can be updated as $v_0 \leftarrow (v_0 - \mu_+)/\sigma_+$ and then the algorithm can be performed with $\mu_+ = 0$ and $\sigma_+ = 1$, afterwards v is updated as $v \leftarrow \mu_+ + v \cdot \sigma_+$.

A possible choice for the proposal function $\mathrm{g}(\cdot)$ is a translated exponential distribution with density

$$\mathrm{g}(v; \alpha, v_0) = \begin{cases} \alpha \exp(-\alpha(v - v_0)) & , v \geq v_0 \\ 0 & , v < v_0. \end{cases}$$

Translated exponential distributed realizations can be generated by the inversion method by $v = -1/\alpha \cdot \ln(u) + v_0$, where u is generated from the uniform distribution $\mathrm{U}_1(0, 1)$. Where α is a parameter that must be set. For $v \geq v_0$ we have that

$$\begin{aligned} \alpha > v_0\text{:}\ & \exp(\alpha(v - v_0)) \exp(-v^2/2) \leq \exp(\alpha^2/2 - v_0\alpha) = c \\ \alpha \leq v_0\text{:}\ & \exp(\alpha(v - v_0)) \exp(-v^2/2) \leq \exp(-(v_0)^2/2) = c, \end{aligned}$$

then

$$\frac{\mathrm{g}(v)}{c \cdot \mathrm{h}(v)} = \begin{cases} \exp(-v^2/2 + \alpha(v - v_0) - \alpha^2/2 + \alpha v_0) & \text{if } \alpha \geq v_0 \\ \exp(-v^2/2 + \alpha(v - v_0) - \alpha^2/2 + \alpha v_0) & \text{otherwise.} \end{cases}$$

Then from Robert (1995) it is given that

$$\alpha = \frac{v_0 + \sqrt{v_0^2 + 4}}{2}$$

is the optimal factor, which will give the highest probability for acceptance. Hence also the most efficient algorithm.

Then we have an algorithm to generate a value from a truncated univariate Gaussian distribution, but we want to simulate from a truncated multivariate Gaussian distribution. This can be done by an MCMC approach with a Gibbs step, then $\boldsymbol{v}^{(j)}$ in state j is updated as

$$\begin{aligned} v_1^{(j)} &\sim \mathrm{N}_1^+ \left(\mathrm{E}(v_1 | v_2^{(j-1)}, v_3^{(j-1)}, ..., v_q^{(j-1)}), \mathrm{Var}(v_1 | v_2^{(j-1)}, v_3^{(j-1)}, ..., v_q^{(j-1)}), v_{0,1} \right) \\ v_2^{(j)} &\sim \mathrm{N}_1^+ \left(\mathrm{E}(v_2 | v_1^{(j)}, v_3^{(j-1)}, ..., v_q^{(j-1)}), \mathrm{Var}(v_2 | v_1^{(j)}, v_3^{(j-1)}, ..., v_q^{(j-1)}), v_{0,2} \right) \\ &\vdots \\ v_q^{(j)} &\sim \mathrm{N}_1^+ \left(\mathrm{E}(v_q | v_1^{(j)}, v_2^{(j)}, ..., v_{q-1}^{(j)}), \mathrm{Var}(v_q | v_1^{(j)}, v_2^{(j)}, ..., v_{q-1}^{(j)}), v_{0,q} \right), \end{aligned}$$

where $v_{0,i}$ is the truncation point for element v_i. Let $\boldsymbol{v}_{\neg i} = [v_1, .., v_{i-1}, v_{i+1}, ..., v_q]$. Then $\mathrm{E}(v_i | \boldsymbol{v}_{\neg i}) = -\nu_i + \Sigma'_{v,i\neg i} \Sigma^{-1}_{v,\neg i\neg i} (\boldsymbol{v}_{\neg i} + \boldsymbol{\nu}_{\neg i})$ and

$\mathrm{Var}(v_i|\boldsymbol{v}_{\neg i}) = \sigma^2_{v,ii} - \Sigma'_{v,i\neg i}\Sigma^{-1}_{v,\neg i\neg i}\Sigma_{v,i\neg i}$, where $\Sigma_{v,\neg i\neg i}$ is the $(q-1)\times(q-1)$ matrix derived from Σ_v, with elements $\sigma^2_{v,ij}$, by eliminating row i and column i. And $\Sigma_{v,i\neg i}$ is the $(q-1)$ vector derived from the ith column of Σ_v by removing the ith row term. It is also important to note that it is possible to calculate $\Sigma^{-1}_{v,\neg i\neg i}$ without performing a full inversion. Let $V = \Sigma_v^{-1}$ then $\Sigma^{-1}_{v,\neg i\neg i} = V_{\neg i\neg i} - V_{i\neg i}V'_{i\neg i}/v_{ii}$, where v_{ii} is element (i,i) in V. So the inversion of Σ_v in full dimension is the only inversion that needs to be done. An initial state has to be chosen for $\boldsymbol{v}^{(0)}$, and a natural choice is $\boldsymbol{v}^{(0)} = \max\{\boldsymbol{\mu}_v, \boldsymbol{v}_0\} = max\{-\boldsymbol{\nu}, \boldsymbol{v}_0\}$. Where $\max\{\cdot,\cdot\}$ acts component wise on the vectorial argument.

Then we have everything needed to define an efficient algorithm to generate realizations from the CS-Gaussian distribution.

Algorithm C: Iterative Simulation

- $V \leftarrow \Sigma_v^{-1}$
- $\boldsymbol{v} = max\{-\boldsymbol{\nu}, \boldsymbol{v}_0\}$
- Iterate until convergence
- **for** $i = 1, ..., q$
- Calculate $\Sigma^{-1}_{v,\neg i\neg i} \leftarrow V_{\neg i\neg i} - V_{i\neg i}V'_{i\neg i}/v_{ii}$
- Calculate the mean of $v_i|\boldsymbol{v}_{\neg i}$: $\mu_+ \leftarrow -\nu_i + \Sigma'_{v,i\neg i}\Sigma^{-1}_{v,\neg i\neg i}(\boldsymbol{v}_{\neg i} + \boldsymbol{\nu}_{\neg i})$
- Calculate the variance of $v_i|\boldsymbol{v}_{\neg i}$: $\sigma^2_+ \leftarrow \Sigma^2_{v,ii} - \Sigma'_{v,i\neg i}\Sigma^{-1}_{v,\neg i\neg i}\Sigma_{v,i\neg i}$
- Standardize v_0: $v_0 \leftarrow (v_{0,i} - \mu_+)/\sigma_+$
- Calculate the parameter for the proposal distribution: $\alpha \leftarrow v_0 + \left(\sqrt{v_0^2 + 4}\right)/2$
- **Repeat until** $u \leq \exp(-(v-\alpha)^2/2)$ for $v_0 < \alpha$
  and $u \leq \exp(-(v-\alpha)^2/2)\exp(-(v_0-\alpha)^2/2)$ else.
  Generate $v \sim \exp(1/\alpha) + v_0$
  Generate $u \sim \mathrm{U}_1(0,1)$
- $v_i \leftarrow \mu_+ + v\sigma_+$ for $i = 1, ..., q$
- $\boldsymbol{x} \leftarrow \boldsymbol{t}|\boldsymbol{v} \sim \mathrm{N}_n(\boldsymbol{\mu} + \Gamma_{tv}\Sigma_v^{-1}(\boldsymbol{v} + \boldsymbol{\nu}), \Sigma_t - \Gamma_{tv}\Sigma_v^{-1}\Gamma'_{tv})$

Then $\boldsymbol{x}$ will be a realization of the CS-Gaussian distribution.

Contrary to the other algorithms that requires to generate $v_i \geq 0$ for all i simultaneously, this algorithm generates $v_i \geq 0$ one by one. But it require an iterative method for each realization, which can be too computer demanding. It might be possible to generate $\boldsymbol{v}$ simultaneous and without iterations. For example with rejection sampling from a multivariate exponential distribution as proposal distribution, thus avoid using an MCMC method.

The number of iterations necessary for convergence will depend of the problem. But in general it seems that the realizations have a relative low correlation. Since it is only $\boldsymbol{v}^{(j)}$ that depends on the previous state, and the variable of interest

$\boldsymbol{x} = [\boldsymbol{t}|\boldsymbol{v} \geq \boldsymbol{0}]$ is not directly dependent of the previous state, as displayed in Figure 3.1. The truncated multivariate Gaussian distribution is unimodal, hence we are avoiding problems due to convergence and several maxima in the MCMC method. This is further discussed in the next Section and in Section 7.2.

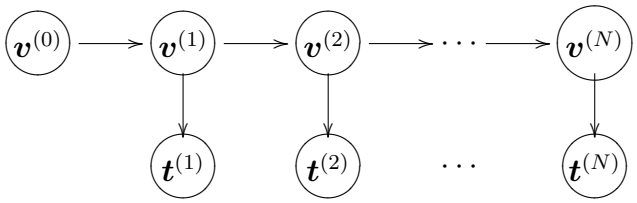


Figure 3.1: Dependency graph.

3.4 Empirical Study

To compare the algorithms an inverse problem, as discussed in Section 2.5, is constructed with an $\boldsymbol{x} \in \mathbb{R}^3$. A stationary model $\mathrm{CSN}_{3,3}(\cdot,\cdot,\cdot,\cdot,\cdot)$ is assumed. The parameters are chosen to $\mu_{x.} = 2$, $\sigma^2_{x.} = 0.3$, $\nu_{x.} = 1$, $\delta_{x.} = 1$, $\Sigma_e = 10^{-2}(HH' + I)$, $\Gamma_e = 0 \cdot I$ and five different values of $\gamma_{x.} = \{-1.5, 1, 0, 1, 1.5\}$. An exponential dependency function is used, $\mathrm{c}(h) = \exp\{-h/5\}$. The H matrix in the likelihood model is chosen to be

$$H = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.20 & 0.60 & 0.20 \\ 0 & 0.25 & 0.75 \end{bmatrix}.$$

First reference values $\boldsymbol{x}$ is generated from the prior, then $\boldsymbol{d} \in \mathbb{R}^3$ is generated as $[\boldsymbol{d}|\boldsymbol{x}] = H\boldsymbol{x} + \boldsymbol{e}$. The posterior model can be found as described in Section 2.5. 50000 realizations are simulated from the posterior model with algorithm A, B and C, and the average time of repeating this for 50 different reference data sets is used to compare the algorithms.

To study the convergence of the *Iterative Simulation* algorithm a traceplot of the L2-norm of each step is presented in Figure 3.2. Three different initial states are used: $\boldsymbol{v^{(0)}} = \boldsymbol{-\nu}$, $\boldsymbol{v^{(0)}} = [-2,-2,-2,-2,-2,-2]'$ and $\boldsymbol{v^{(0)}} = [4,4,4,4,4,4]'$. As seen in Figure 3.2 the algorithm appears to converge almost immediately. In Figure 3.3 the two first components of $\boldsymbol{v}$ (v_1, v_2) are displayed, and also this confirms that the algorithm seems to converge fast for this case. For the *Iterative Simulation* algorithm a burn in period of 50 realizations are used to be sure that the algorithm has converged.

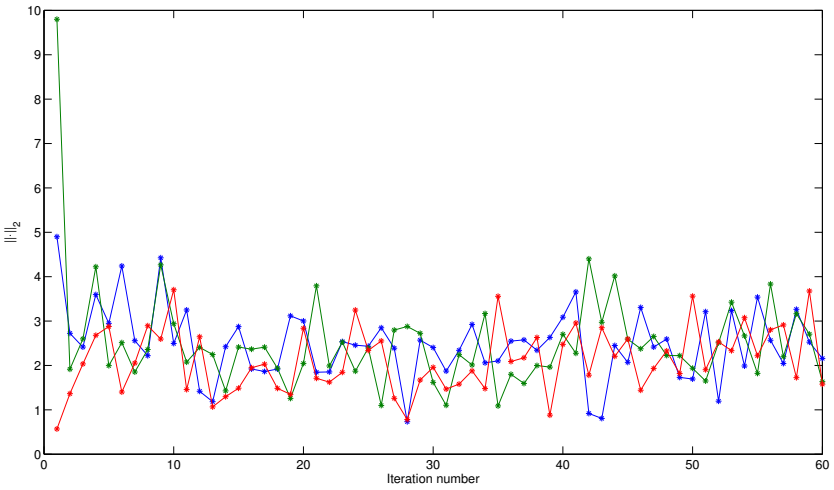


Figure 3.2: Traceplot for the length of $\boldsymbol{v}$.

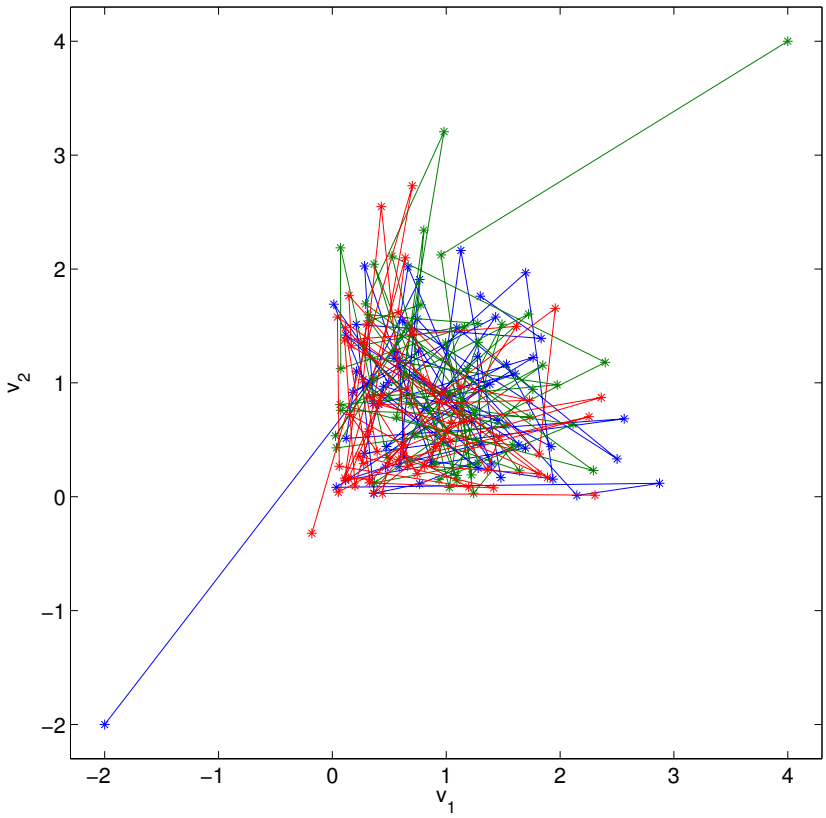


Figure 3.3: Plot of v_1 and v_2 for the three different initial states.

$\gamma_{x_\cdot}$	Brute Force Simulation	Two Steps Simulation	Iterative Simulation
-1.5	42s	30s	6.4s
-1	259s	191s	6.5s
0	679s	494s	6.5s
1	274s	204s	6.5s
1.5	42s	31s	6.5s

Table 3.1: Runtime in seconds (s) for the algorithms.

As seen in Table 3.1 the *Iterative Simulation* algorithm is the most efficient one. As expected is it observed that the *Brute Force Simulation* algorithm has a bit longer runtime than the *Two Steps Simulation* algorithm. The runtime for the *Iterative Simulation* algorithm appear as almost constant for these values of $\gamma_{x_\cdot}$. For a higher dimensional problem the difference in runtime will be even greater.

The runtime of algorithm A and B is particularly influenced by

$$\boldsymbol{\nu}_{x|d} = \begin{bmatrix} \nu_{x_\cdot} \cdot \mathbf{1} - \gamma_{x_\cdot} \sigma_{x_\cdot}^2 C_n H' (H \Sigma_x H' + \Sigma_e)^{-1} (\boldsymbol{d} - H \boldsymbol{\mu}_x) \\ 0 \end{bmatrix},$$

for $\gamma_{e_\cdot} = 0$, since $-\boldsymbol{\nu}_{x|d} \ll \mathbf{0}$ will shift the density such that the probability to generate $\boldsymbol{v} \geq \mathbf{0}$ is low. The sign of the elements in $\boldsymbol{\nu}_{x|d}$ is then given from the sign of $\gamma_{x_\cdot}$, $\boldsymbol{e} = \boldsymbol{d} - H\boldsymbol{\mu}$ and of the value of $\nu_{x_\cdot}$. Some elements e_i will be negative and some will be positive, hence $\boldsymbol{\nu}_{x|d}$ will have both positive and negative values. For $\gamma_{x_\cdot} \ll 0$ will the density shift such that most elements in $\boldsymbol{e}$ is likely to be positive. And for $\gamma_{x_\cdot} \gg 0$ will most elements in $\boldsymbol{e}$ be less than zero. So a low absolute value of $\gamma_{x_\cdot}$ will give that some v_i are very unlikely to be generated greater than zero. Then the *Brute Force Simulation* algorithm and the *Two Steps Simulation* algorithm are expected to propose many times, before a realization is accepted. The *Iterative Simulation* algorithm is not affected by this in the same way, since the rejection sampling use a translated distribution for the proposal. Then the proposal distribution will shift as $\boldsymbol{\nu}_{x|d}$ changes. The algorithms give the same predicted mean and median, but the realizations from algorithm C are not independent as the realizations from algorithm A and B are. The estimated cdfs and pdfs from these three algorithms are almost identical.

Since the *Iterative Simulation* algorithm is the fastest it is studied in more detail to see how different parameters influence the runtime. It is shown in Figure 3.4 that different values of $\sigma_{e_\cdot}^2$ and $\sigma_{x_\cdot}^2$ do not affect the runtime much. For $\sigma_{e_\cdot}^2$ and $\sigma_{x_\cdot}^2$ between 10^{-5} and 5 the maximum change in runtime is only 1.4 seconds.

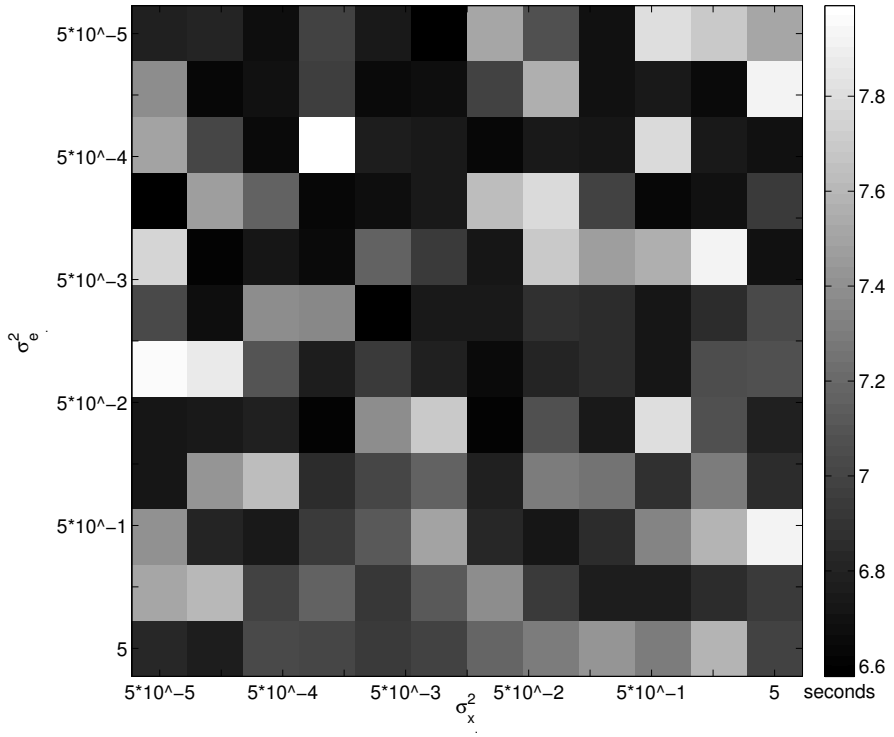


Figure 3.4: Runtime in seconds for algorithm C for different values of $\sigma^2_{x.}$ and $\sigma^2_{e.}$.

The effect of different values of $\gamma_{x.}$ and $\gamma_{e.}$ is displayed in Figure 3.5. The runtime is highest when both $|\gamma_{x.}|$ and $|\gamma_{e.}|$ are large. The covariance of $\boldsymbol{v}|\boldsymbol{d}$ is

$$\begin{aligned}\Sigma_{v|d} =& \Delta + \Gamma\Sigma\Gamma' \\ =& \begin{bmatrix} \delta_{x.} I + \gamma^2_{x.}\Sigma_x - \gamma^2_{x.}\Sigma_x H'AH\Sigma_x & -\gamma_{x.}\gamma_{e.}\Sigma_x H'A\Sigma_e \\ -\gamma_{x.}\gamma_{e.}\Sigma_x H'A\Sigma_e & I + \gamma^2_{e.}\Sigma_e - \gamma^2_{e.}\Sigma_e A\Sigma_e \end{bmatrix},\end{aligned}$$

where $A = [H\Sigma_x H' + \Sigma_e]^{-1}$. So when $\gamma_{x.}$ and $\gamma_{e.}$ have a high absolute value will the correlation between $\boldsymbol{v}$ and $\boldsymbol{v}_e$ be high. This causes higher runtime, but even with very high values of $\gamma_{x.}$ and $\gamma_{e.}$ is the runtime only 1.5 seconds higher than for the lowest runtime.

The runtime for the *Iterative Simulation* algorithm will increase as the dimension grows. Let $T = n_x$ then $q = q_x + q_d = 2T$, so the total dimension is $3T$. Let

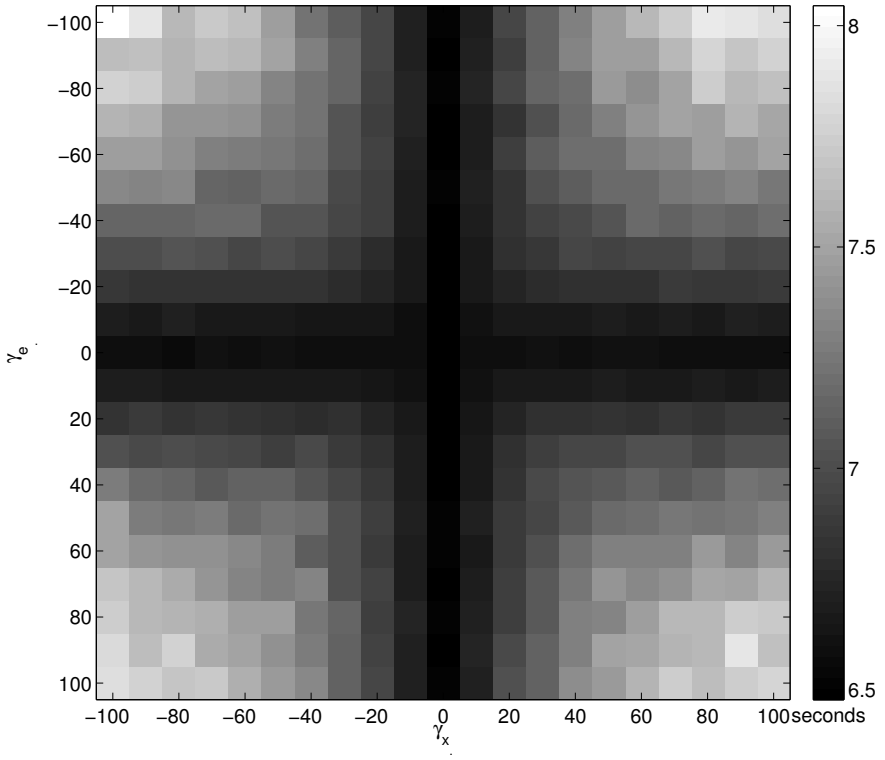


Figure 3.5: Runtime in seconds for algorithm C for different values of $\gamma_{x_\cdot}$ and $\gamma_{e_\cdot}$.

all the parameters be as before with $\gamma_{x_\cdot} = 3$ and the H-matrix is extended as

$$H = \begin{bmatrix} 0.75 & 0.25 & 0 & \cdots & \cdots & 0 \\ 0.20 & 0.60 & 0.20 & 0 & \ddots & 0 \\ 0 & 0.20 & 0.60 & 0.20 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & 0 & 0.20 & 0.60 & 0.20 \\ 0 & \cdots & \cdots & 0 & 0.25 & 0.75 \end{bmatrix} \in \mathbb{R}^{T \times T}.$$

Then we have a posterior model that can be scaled up to any dimension T. To check how the runtime of the algorithm change for different values of T the average time to generate 1000 realizations with 20 different data sets for $T \in \{20, 40, 60, ..., 420\}$ is used. It is observed that the runtime increases a bit faster than linear as a function of T, as displayed in Figure 3.6. This can be a problem for an extremely large trace or in a 3D-setting. The runtime of inverting Σ_v is asymptotically $T \log T$, and will be a problem for large T. So an approximation should be done

such that inverting of the full Σ_v is avoided.

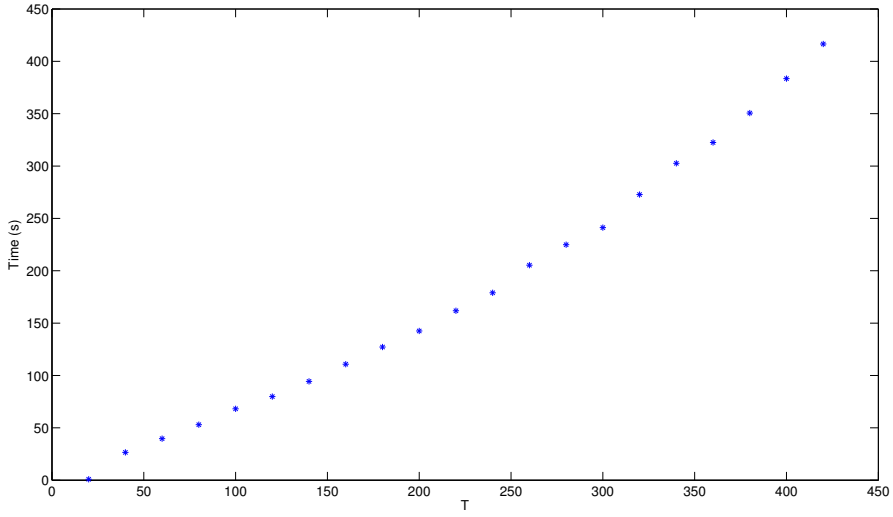


Figure 3.6: Runtime as a function of the dimension, T, of the problem.

Chapter 4

Stationary Prior Models

A stationary prior will not change when shifted in space. So all the marginals, bivariate pdfs and higher dimensional pdfs must be shift invariant. As shown in Expression (2.4) there exists an one-to-one transformation between the CS-Gaussian distribution for $\boldsymbol{x}$ and the Gaussian distribution for $(\boldsymbol{t}, \boldsymbol{v})$. Hence is it possible to specify a prior model in the x-space and one alternative model in the tv-space. So we will here present two different priors, and it is possible to extend them to higher dimensional reference spaces. For both the priors we use a shift invariant dependence matrix $C \in \mathcal{R}^{n_x \times n_x}$ given as

$$C = \begin{bmatrix} 1 & \mathrm{c}(1) & \dots & \dots & \mathrm{c}(n_x - 1) \\ \mathrm{c}(1) & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & \mathrm{c}(1) \\ \mathrm{c}(n_x - 1) & \dots & \dots & \mathrm{c}(1) & 1 \end{bmatrix}.$$

Where $\mathrm{c}(i, j) = c(|i - j|)$ is a spatial positive definite dependence measure with $c(0) = 1$. Here we assume that $c(\cdot)$ is parameterized with given parameters, but in theory it should be possible to also estimate the parameters of the dependence measure.

4.1 x-space Prior

In this prior $\boldsymbol{x} \in \mathcal{R}^{n_x}$ is assumed to be CS-Gaussian, as

$$\boldsymbol{x} \sim \mathrm{CSN}_{n_x,q_x}(\boldsymbol{\mu}_x, \Sigma_x, \Gamma_x, \boldsymbol{\nu}_x, \Delta_x).$$

Let $\mathbf{1}$ be a vector containing only ones, then the parameters can be specified in the x-space as:

$$\begin{array}{ll} \boldsymbol{\mu}_x = \mu_{x_\cdot} \cdot \mathbf{1} \in \mathbb{R}^{n_x} & \boldsymbol{\mu}_t = \mu_{x_\cdot} \cdot \mathbf{1} \in \mathbb{R}^{n_x} \\ \Sigma_x = \sigma^2_{x_\cdot} \cdot C \in \mathbb{R}^{n_x \times n_x} & \Sigma_t = \Sigma_x = \sigma^2_{x_\cdot} \cdot C \in \mathbb{R}^{n_x \times n_x} \\ \Gamma_x = \gamma_{x_\cdot} \cdot I \in \mathbb{R}^{n_x \times n_x} \quad \Leftrightarrow & \Gamma_{vt} = \Gamma_x \Sigma_x = \gamma_{x_\cdot} \sigma^2_{x_\cdot} \cdot C \in \mathbb{R}^{q_x \times n_x} \\ \boldsymbol{\nu}_x = \nu_{x_\cdot} \cdot \mathbf{1} \in \mathbb{R}^{q_x} & \boldsymbol{\mu}_v = -\boldsymbol{\nu} \in \mathbb{R}^{q_x} \\ \Delta_x = \delta_{x_\cdot} \cdot I \in \mathbb{R}^{q_x \times q_x} & \Sigma_v = \Delta_x + \Gamma_x \Sigma_x \Gamma'_x = \delta_{x_\cdot} \cdot I + \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} \cdot C \in \mathbb{R}^{q_x \times q_x}. \end{array}$$

Let $n_x = q_x$ so the free parameters are the scalars $\mu_{x_\cdot}$, $\sigma^2_{x_\cdot}$, $\gamma_{x_\cdot}$, $\nu_{x_\cdot}$ and $\delta_{x_\cdot}$. The parameters of the marginals can be calculated as described in Section 2.2:

$$\begin{aligned} x_i \sim & \mathrm{CSN}_{1,q_x}(\mu_{x_i}, \Sigma_{x_i}, \Gamma_{x_i}, \boldsymbol{\nu}_{x_i}, \Delta_{x_i}) \\ \mu_{x_i} = & \boldsymbol{b}'_{(i)} \boldsymbol{\mu}_x = \mu_{x_\cdot}, \\ \Sigma_{x_i} = & \boldsymbol{b}'_{(i)} \Sigma_x \boldsymbol{b}_{(i)} = \sigma^2_{x_\cdot}, \\ \Gamma_{x_i} = & \Gamma_x \Sigma_x \boldsymbol{b}_{(i)} \frac{1}{\sigma^2_{x_\cdot}} = \gamma_{x_\cdot} \sigma^2_{x_\cdot} C \boldsymbol{b}_{(i)} \frac{1}{\sigma^2_{x_\cdot}} = \gamma_{x_\cdot} C_{(i)}, \\ \boldsymbol{\nu}_{x_i} = & \boldsymbol{\nu}_x, \\ \Delta_{x_i} = & \Delta_x + \Gamma_x \Sigma_x \Gamma'_x - \Gamma_x \Sigma_x \boldsymbol{b}_{(i)} \frac{1}{\sigma^2_{x_\cdot}} \boldsymbol{b}'_{(i)} \Sigma_x \Gamma'_x \\ = & \delta_{x_\cdot} I + \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C - \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C_{(i)} C'_{(i)}, \end{aligned}$$

where $C_{(i)}$ is column i of the dependence matrix C. By transforming the distribution to the corresponding Gaussian distribution of $(\boldsymbol{t}, \boldsymbol{v})$ we obtain parameter values

$$\begin{aligned} \Gamma_{v_i t} = & \Gamma_{x_i} \Sigma_{x_i} = \gamma_{x_\cdot} \sigma^2_{x_\cdot} C_{(i)} \\ \Sigma_{v_i} = & \Delta_{x_i} + \Gamma_{x_i} \Sigma_{x_i} \Gamma'_{x_i} \\ = & \delta_{x_\cdot} I + \gamma^2_{x_\cdot} \sigma^2_x C. \end{aligned}$$

Since $C_{(i)}$ is almost identical for every i except for some deviations caused by the boundary, will the dominating elements be identical for the marginals for a high

dimensional problem. The marginals will be almost identical, except for relatively few variables close to the border. The marginal of x_i is then

$$\mathrm{p}(x_i) = \mathrm{p}(t_i|\boldsymbol{v} \geq \mathbf{0}).$$

So t_i depends on all v_j; $j = 1, ..., q_x$ and not only v_i.

It is possible to constrain the prior by setting $\nu_{x_\cdot} = 0$ and $\delta_{x_\cdot} = 1$. This gives a simpler prior, where the skewness can be controlled by $\gamma_{x_\cdot}$. Hence the free parameters are $\mu_{x_\cdot}$, $\sigma^2_{x_\cdot}$ and $\gamma_{x_\cdot}$. For the model to be valid is it necessary to require that $\sigma^2_{x_\cdot} > 0$. Note that the model is valid for all values of $\gamma_{x_\cdot}$ and $\mu_{x_\cdot}$.

4.2 tv-space Prior

An alternative prior can be specified in the tv-space, where $(\boldsymbol{t}, \boldsymbol{v}) \in \mathcal{R}^{n_x+q_x}$ is assumed to be Gaussian, as

$$\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{N}_{n_x+q_x} \left(\begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \Sigma_t & \Gamma_{tv} \\ \Gamma_{vt} & \Sigma_v \end{bmatrix} \right), \tag{4.1}$$

then the parameters can be specified in the tv-space as:

$$\begin{array}{ll} \boldsymbol{\mu}_t = \mu_{t_\cdot} \mathbf{1} \in \mathbb{R}^{n_x} & \boldsymbol{\mu}_x = \mu_{t_\cdot} \mathbf{1} \in \mathbb{R}^{n_x} \\ \Sigma_t = \sigma^2_{t_\cdot} C \in \mathbb{R}^{n_x \times n_x} & \Sigma_x = \Sigma_t = \sigma^2_{t_\cdot} C \in \mathbb{R}^{n_x \times n_x} \\ \Gamma_{vt} = \gamma_{tv_\cdot} I \in \mathbb{R}^{q_x \times n_x} \quad \Leftrightarrow & \Gamma_x = \frac{\gamma_{tv_\cdot}}{\sigma^2_{t_\cdot}} C^{-1} \in \mathbb{R}^{q_x \times n_x} \\ \boldsymbol{\mu}_v = \mu_{v_\cdot} \mathbf{1} \in \mathbb{R}^{q_x} & \boldsymbol{\nu}_x = -\mu_{v_\cdot} \mathbf{1} \in \mathbb{R}^{q_x} \\ \Sigma_v = \sigma^2_{v_\cdot} I \in \mathbb{R}^{q_x \times q_x} & \Delta_x = \sigma^2_v I - \frac{\gamma^2_{tv_\cdot}}{\sigma^2_{t_\cdot}} C^{-1} \in \mathbb{R}^{q_x \times q_x} \end{array}$$

Let $n_x = q_x$ so the free parameters are the scalars $\mu_{t_\cdot}$, $\sigma^2_{t_\cdot}$, $\gamma_{tv_\cdot}$, $\mu_{v_\cdot}$ and $\sigma^2_{v_\cdot}$. The

parameters of the marginals can be calculated as described in Section 2.2:

$$\begin{aligned}
x_i \sim& \mathrm{CSN}_{1,q_x}(\mu_{x_i}, \Sigma_{x_i}, \Gamma_{x_i}, \boldsymbol{\nu}_{x_i}, \Delta_{x_i}) \\
\mu_{x_i} =& \boldsymbol{b}'_{(i)}\boldsymbol{\mu}_x = \mu_{t_\cdot}, \\
\Sigma_{x_i} =& \boldsymbol{b}'_{(i)}\Sigma_x \boldsymbol{b}_{(i)} = \sigma^2_{t_\cdot}, \\
\Gamma_{x_i} =& \Gamma_x \Sigma_x \boldsymbol{b}_{(i)} \frac{1}{\sigma^2_{t_\cdot}} = \frac{\gamma_{tv_\cdot}}{\sigma^2_{t_\cdot}} C^{-1} \sigma^2_{t_\cdot} C b_{(i)} \frac{1}{\sigma^2_{t_\cdot}} = \frac{\gamma_{tv_\cdot}}{\sigma^2_{t_\cdot}} b_{(i)}, \\
\boldsymbol{\nu}_{x_i} =& -\boldsymbol{\mu}_v, \\
\Delta_{x_i} =& \Delta_x + \Gamma_x \Sigma_x \Gamma'_x - \Gamma_x \Sigma_x \boldsymbol{b}_{(i)} \frac{1}{\sigma^2_{t_\cdot}} \boldsymbol{b}'_{(i)} \Sigma_x \Gamma'_x \\
=& \sigma^2_{v_\cdot} I - \frac{\gamma^2_{tv_\cdot}}{\sigma^2_{t_\cdot}} b_{(i)} b'_{(i)},
\end{aligned}$$

By transforming the distribution to the corresponding Gaussian distribution of $(\boldsymbol{t}, \boldsymbol{v})$ we obtain

$$\begin{aligned}
\Gamma_{vt_i} =& \Gamma_{x_i}\Sigma_{x_i} = \gamma_{tv_\cdot} b_{(i)} \\
\Sigma_{v_i} =& \Delta_{x_i} + \Gamma_{x_i}\Sigma_{x_i}\Gamma'_{x_i} \\
=& \sigma^2_{v_\cdot} I.
\end{aligned}$$

So

$$x_i \sim \mathrm{CSN}_{1,1}\left(\mu_{t_i}, \sigma^2_{t_\cdot}, \frac{\gamma_{tv_\cdot}}{\sigma^2_{t_\cdot}}, -\mu_{v_\cdot}, \sigma^2_{v_\cdot} - \frac{\gamma^2_{tv_\cdot}}{\sigma^2_{t_\cdot}}\right).$$

The marginal of x_i is then

$$\mathrm{p}(x_i) = \mathrm{p}(t_i | \boldsymbol{v} \geq \boldsymbol{0}) = \mathrm{p}(t_i | v_i \geq 0)$$

So each t_i only depends of v_i. Hence this prior model is also stationary, since $\mathrm{p}(x_i) = \mathrm{p}(x_j)$ and independent for all i and j. It is possible to constrain the prior by setting $\mu_{v_\cdot} = 0$ and $\sigma^2_{v_\cdot} = 1$. This gives a more simple prior, where the skewness can be controlled by $\gamma_{tv_\cdot}$. Hence the free parameters are $\mu_{t_\cdot}$, $\sigma^2_{t_\cdot}$ and $\gamma_{tv_\cdot}$. For the model to be valid is it necessary to require that

$$\begin{bmatrix} \sigma^2_{t_\cdot} C & \gamma_{tv_\cdot} I \\ \gamma_{tv_\cdot} I & I \end{bmatrix}$$

is positive definite. This requirement can often be difficult to meet.

4.3 Comparing the Priors

For the x-space prior t_i is connected to all $\boldsymbol{v}$, but for the tv-space prior is t_i only connected to v_i. The latter provides a more logical constructed prior, hence will it also be easier to understand the effects of the different parameters. The marginals for tv-space prior have $q = 1$, but the marginals for the x-space prior have $q = n_x$. So the marginals for the tv-space prior are easier to calculate. But the x-space prior has less complicated requirements for the parameter independence than the tv-space prior. For the tv-space prior it is hard to specify valid parameters, since a determinant has to be calculated to check if a parameter set is valid.

Chapter 5
Parameter Estimation in Prior Models

For an inverse problem in an empirical Bayes spirit will we estimate the parameters in the prior model from observations of a comparable phenomenon. Previous work have considered parameter estimation of the one-dimensional skew-normal distribution, see Pewsey (2000). In this paper "Method of moments" and "Maximum likelihood" methods are used to estimate the parameters. The "Method of moments" is not working well for very skew distributed data sets. The "Maximum likelihood" method must be solved by a numerical approximation. In Pewsey (2000) the problems of finding the global maximum are solved by a grid of different initial values. Here we will present alternative methods to estimate the parameters of the CS-Gaussian distribution for the two different priors presented in the previous section. We will assume that observations $\boldsymbol{x}^o \in \mathcal{R}^{n_x}$ of a comparable one-dimensional phenomenon is available, for example observations along a well trace.

5.1 x-space Prior

For the x-space prior we will present two different methods for estimation, and compare them empirically. The first method, with a full likelihood model, requires an exponential dependency function.

5.1.1 Full Likelihood Model.

For the x-space prior the full likelihood (fL) is defined as:

$$\begin{aligned}\mathrm{L}(\mu_{x_\cdot}, \sigma^2_{x_\cdot}, \gamma_{x_\cdot}, \nu_{x_\cdot}, \delta_{x_\cdot}|\boldsymbol{x^o}) &= \left[\Phi_{q_x}(\mathbf{0}; \nu_{x_\cdot}\mathbf{1}, \delta_{x_\cdot}I + \gamma^2_{x_\cdot}\sigma^2_{x_\cdot}C)\right]^{-1} \\ &\cdot \Phi_{q_x}(\gamma_{x_\cdot}(\boldsymbol{x}^o - \mu_{x_\cdot}\mathbf{1}); \nu_{x_\cdot}\mathbf{1}, \delta_{x_\cdot}I) \\ &\cdot \phi_{n_x}(\boldsymbol{x}^o; \boldsymbol{\mu}_x, \sigma^2_{x_\cdot}C) \\ &= \left[\Phi_{q_x}(\mathbf{0}; \nu_{x_\cdot}\cdot\mathbf{1}, \delta_{x_\cdot}\cdot I + \gamma^2_{x_\cdot}\sigma^2_{x_\cdot}C)\right]^{-1} \\ &\cdot \prod_{i=1}^{q_x}\Phi_1(\gamma_{x_\cdot}(x^o_i - \mu_{x_\cdot}); \nu_{x_\cdot}, \delta_{x_\cdot}) \\ &\cdot \phi_{n_x}(\boldsymbol{x}^o; \mu_{t_\cdot}\mathbf{1}, \sigma^2_{x_\cdot}C),\end{aligned} \tag{5.1}$$

Here we have used that $\Phi_{q_x}(\gamma_{x_\cdot}(\boldsymbol{x}^o - \mu_{x_\cdot}\mathbf{1}); \nu_{x_\cdot}\mathbf{1}, \delta_{x_\cdot}I)$ can be factorized as $\prod_{i=1}^{q_x}\Phi_1(\gamma_{x_\cdot}(x^o_i - \mu_{x_\cdot}); \nu_{x_\cdot}, \delta_{x_\cdot})$, since the covariance $\delta_{x_\cdot}I$ is diagonal. $\phi_{n_x}(\cdot;\cdot,\cdot)$ is analytically obtainable and $\Phi_1(\cdot;\cdot,\cdot)$ is fast to approximate for a one dimensional Gaussian variable. But the normalization constant $\Phi_{q_x}(\cdot;\cdot,\cdot)$ can in general be computationally hard to obtain, so we will here look at the special case where an efficient solution can be found. Then the maximum likelihood, ML, is given as

$$(\hat{\mu}_{x_\cdot}, \hat{\sigma}^2_{x_\cdot}, \hat{\gamma}_{x_\cdot}, \hat{\nu}_{x_\cdot}, \hat{\delta}_{x_\cdot}) = \underset{\mu_{x_\cdot}, \sigma^2_{x_\cdot}, \gamma_{x_\cdot}, \nu_{x_\cdot}, \delta_{x_\cdot}}{\arg\max}\ \mathrm{L}(\mu_{x_\cdot}, \sigma^2_{x_\cdot}, \gamma_{x_\cdot}, \nu_{x_\cdot}, \delta_{x_\cdot}|\boldsymbol{x^o}).$$

Calculation of the normalization constant. The normalization constant is given as

$$\begin{aligned}\Phi_n(\mathbf{0}; \mathbf{0}, \delta_{x_\cdot}\cdot I + \gamma^2_{x_\cdot}\sigma^2_{x_\cdot}C) &= \mathrm{p}(\boldsymbol{x} \le \mathbf{0}) \\ &= \int_{-\infty}^{\mathbf{0}}\phi_n(\boldsymbol{x}; \mathbf{0}, \delta_{x_\cdot}I + \gamma^2_{x_\cdot}\sigma^2_{x_\cdot}C)\mathrm{d}\boldsymbol{x}.\end{aligned}$$

where $\boldsymbol{x}$ is Gaussian distributed as $\mathrm{N}_n(\boldsymbol{x}; \mathbf{0}, \delta_{x_\cdot}I + \gamma^2_{x_\cdot}\sigma^2_{x_\cdot}C)$, so the covariance matrix can be written as $\Sigma = \alpha\cdot I + \beta\cdot C$. Where α and β are constants and C is a correlation matrix. This is in general a very complicated problem, but we will study the special case where C is assumed to be constructed from an exponential correlation function, $c(|i-j|) = \exp\{-|i-j|/r\}$.

This case can be modeled with an $\boldsymbol{x}^0$ and $\boldsymbol{\epsilon}$, as shown in Figure 5.1, where $\boldsymbol{x}^0 \sim \mathrm{N}_n(\mathbf{0}, \beta\cdot C)$, $\boldsymbol{\epsilon} \sim \mathrm{N}_n(\mathbf{0}, \alpha\cdot I)$ and $(\boldsymbol{x}^0, \boldsymbol{\epsilon})$ are independent. Then

$$\boldsymbol{x} = \boldsymbol{x}^0 + \boldsymbol{\epsilon} \sim \mathrm{N}_n(\mathbf{0}, \alpha\cdot I + \beta\cdot C).$$

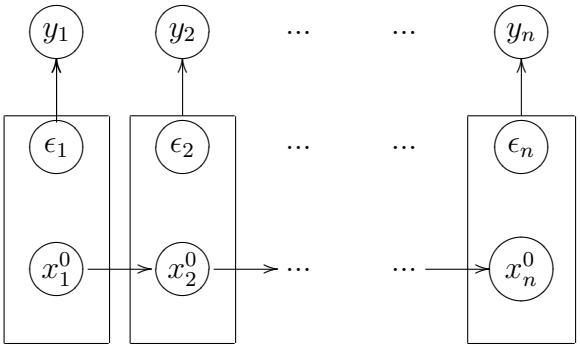


Figure 5.1: Dependency graph.

Define a likelihood model such that

$$[y_i|x_i^0, \epsilon_i] = \begin{cases} 1, & x_i^0 + \epsilon_i \leq 0 \\ 0, & x_i^0 + \epsilon_i > 0. \end{cases} \tag{5.2}$$

Hence will $\Phi_n(\mathbf{0}; \mathbf{0}, \alpha \cdot I + \beta \cdot C) = \mathrm{p}(\boldsymbol{x} \leq \mathbf{0}) = \mathrm{p}(\boldsymbol{y} = \mathbf{1})$.

In Rybicki and Press (1995) it is shown that $\boldsymbol{x}^0$ has Markov property such that $\mathrm{p}(\boldsymbol{x}^0) = \mathrm{p}(x_n^0|x_{n-1}^0) \cdot ... \cdot \mathrm{p}(x_2^0|x_1^0) \cdot \mathrm{p}(x_1^0)$. This require that C is defined from an exponential correlation function. The pdf, $\mathrm{p}(x_i^0|x_{i-1}^0)$, for $i \geq 2$ is the conditional Gaussian distribution given as $\phi_1(x_i^0; \exp(1/r) \cdot x_{i-1}^0, \beta - \beta \cdot \exp(2/r))$ and $\mathrm{p}(x_1^0) = \phi_1(x_1^0; 0, \beta)$.

Then $\mathrm{p}(\boldsymbol{y}) = \mathrm{p}(y_1, y_2, ..., y_n)$ can be written as:

$$\begin{aligned}
\mathrm{p}(y_1, y_2, ..., y_n) =& \mathrm{p}(y_n|y_1, y_2, ..., y_{n-1}) \cdot ... \cdot \mathrm{p}(y_2|y_1) \cdot \mathrm{p}(y_1) \\
=& \int_{\boldsymbol{x}^0} \mathrm{p}(y_n|y_1, y_2, ..., y_{n-1}, \boldsymbol{x}^0) \cdot ... \cdot \mathrm{p}(y_2|y_1, \boldsymbol{x}^0) \cdot \mathrm{p}(y_1|\boldsymbol{x}^0) \cdot \mathrm{p}(\boldsymbol{x}^0)\mathrm{d}\boldsymbol{x}^0 \\
=& \int_{\boldsymbol{x}^0} \mathrm{p}(y_n|x_n^0) \cdot ... \cdot \mathrm{p}(y_2|x_2^0) \cdot \mathrm{p}(y_1|x_1^0) \cdot \mathrm{p}(x_n^0|x_{n-1}^0) \cdot ... \cdot \\
& \mathrm{p}(x_2^0|x_1^0) \cdot \mathrm{p}(x_1^0)\mathrm{d}\boldsymbol{x}^0 \\
=& \int_{\boldsymbol{x}^0} \int_{\boldsymbol{\epsilon}^0} \mathrm{p}(y_n|x_n^0, \epsilon_n) \cdot ... \cdot \mathrm{p}(y_2|x_2^0, \epsilon_2) \cdot \mathrm{p}(y_1|x_1^0, \epsilon_1) \cdot \mathrm{p}(x_n^0|x_{n-1}^0) \\
& \cdot ... \cdot \mathrm{p}(x_2^0|x_1^0) \cdot \mathrm{p}(x_1^0) \cdot \mathrm{p}(\epsilon_n) \cdot ... \cdot \mathrm{p}(\epsilon_2) \cdot \mathrm{p}(\epsilon_1)\mathrm{d}\boldsymbol{\epsilon}\mathrm{d}\boldsymbol{x}^0 \\
=& \int_{x_n^0} \int_{\epsilon_n^0} \mathrm{p}(y_n|x_n^0, \epsilon_n) \cdot \mathrm{p}(\epsilon_n)\mathrm{d}\epsilon_n \\
& \cdot \int_{x_{n-1}^0} \int_{\epsilon_{n-1}^0} \mathrm{p}(y_{n-1}|x_{n-1}^0, \epsilon_{n-1}) \cdot \mathrm{p}(x_n|x_{n-1}) \cdot \mathrm{p}(\epsilon_{n-1})\mathrm{d}\epsilon_{n-1}\mathrm{d}x_n^0 \cdot ... \cdot \\
& \cdot \int_{x_2^0} \int_{\epsilon_2^0} \mathrm{p}(y_2|x_2^0, \epsilon_2) \cdot \mathrm{p}(x_3|x_2) \cdot \mathrm{p}(\epsilon_2)\mathrm{d}\epsilon_2\mathrm{d}x_3^0 \\
& \cdot \int_{x_1^0} \int_{\epsilon_1^0} \mathrm{p}(y_1|x_1^0, \epsilon_1) \cdot \mathrm{p}(x_1) \cdot \mathrm{p}(\epsilon_1)\mathrm{d}\epsilon_1\mathrm{d}x_2^0\mathrm{d}x_1^0
\end{aligned}$$

In our case we want to calculate $\mathrm{p}(\boldsymbol{y} = \mathbf{1})$. Since $\mathrm{p}(y_i|x_i^0, \epsilon_i)$ is zero for $x_i^0 + \epsilon_i > 0$, as given in Expression (5.2), we get the following recursive expression:

$$\begin{aligned}
\mathrm{p}(\boldsymbol{y} = \mathbf{1}) =& \int_{x_n^0} \int_{\epsilon_n < -x_n^0} \mathrm{p}(\epsilon_n)\mathrm{d}\epsilon_n \int_{x_{n-1}^0} \mathrm{p}(x_n^0|x_{n-1}^0) \cdot \int_{\epsilon_{n-1} < -x_{n-1}^0} \mathrm{p}(\epsilon_{n-1})\mathrm{d}\epsilon_{n-1} \cdot ... \cdot \\
& \cdot \int_{x_2^0} \mathrm{p}(x_3^0|x_2^0) \cdot \int_{\epsilon_2 < -x_2^0} \mathrm{p}(\epsilon_2)\mathrm{d}\epsilon_2 \cdot \int_{x_1^0} \mathrm{p}(x_2^0|x_1^0) \cdot \mathrm{p}(x_1^0) \\
& \cdot \int_{\epsilon_1 < -x_1^0} \mathrm{p}(\epsilon_{n-1})\mathrm{d}\epsilon_1\mathrm{d}x_1^0 \cdot ... \cdot \mathrm{d}x_n^0 \\
=& \int_{x_n^0} \Phi_1(x_n^0; 0, 1) \int_{x_{n-1}^0} \phi_1(x_n^0; \exp(1/r) \cdot x_{n-1}^0, \beta - \beta \cdot \exp(2/r)) \\
& \cdot \Phi_1(x_{n-1}^0; 0, 1) \cdot ... \cdot \int_{x_2^0} \phi_1(x_3^0; \exp(1/r) \cdot x_2^0, \beta - \beta \cdot \exp(2/r)) \\
& \cdot \Phi_1(x_2^0; 0, 1) \cdot \int_{x_1^0} \phi_1(x_2^0; \exp(1/r) \cdot x_1^0, \beta - \beta \cdot \exp(2/r)) \\
& \cdot \phi_1(x_1^0; 0, \beta) \cdot \Phi_1(x_1^0; 0, 1)\mathrm{d}x_1^0 \cdot ... \cdot \mathrm{d}x_n^0
\end{aligned}$$

To calculate this a numerical approximation is needed, so let $\tilde{\phi}(\cdot; \cdot, \cdot)$ be the discrete

approximation of the normal distribution $\phi(\cdot;\cdot,\cdot)$. Hence

$$\begin{aligned}\mathrm{p}(\boldsymbol{y}=\mathbf{1}) \approx &\sum_{x_n^0} \Phi_1(x_n^0;0,1) \sum_{x_{n-1}^0} \tilde{\phi}_1(x_n^0;\exp(1/r)\cdot x_{n-1}^0, \beta-\beta\cdot\exp(2/r)) \\ &\cdot \Phi_1(x_{n-1}^0;0,1)\cdot ... \cdot \sum_{x_2^0} \tilde{\phi}_1(x_3^0;\exp(1/r)\cdot x_2^0, \beta-\beta\cdot\exp(2/r))\cdot\Phi_1(x_2^0;0,1) \\ &\cdot \sum_{x_1^0} \tilde{\phi}_1(x_2^0;\exp(1/r)\cdot x_1^0, \beta-\beta\cdot\exp(2/r))\cdot\tilde{\phi}_1(x_1^0;0,\beta))\cdot\Phi_1(x_1^0;0,1)\end{aligned}$$

The sums can be computed recursively and relatively fast, hence is it possible to estimate $\Phi_n(0;0,\alpha\cdot I+\beta C)$ relatively fast. This gives a method to also calculate other Gaussian quantiles, by changing the likelihood model given in Expression (5.2). Hence can the method be very useful in many different situations. The method presented here is much faster than simulate from the corresponding distribution and count the number of points in the desired region.

5.1.2 Localized Pseudo Likelihood Model.

Here we will approximate the likelihood with a pseudo likelihood given as:

$$\begin{aligned}\mathrm{pL}(\mu_{x_\cdot},\sigma_{x_\cdot}^2,\gamma_{x_\cdot},\nu_{x_\cdot},\delta_{x_\cdot}|\boldsymbol{x^o}) = &\left[\Phi_{q_x}(\mathbf{0};\nu_{x_\cdot}\mathbf{1},\delta_{x_\cdot}I+\gamma_{x_\cdot}^2\sigma_{x_\cdot}^2C)\right]^{-n_x} \\ &\cdot\prod_{i=1}^{n_x}\Phi_{q_x}(\gamma_{x_\cdot}C_{(i)}(x_i^o-\mu_{x_\cdot});\nu_{x_\cdot}\mathbf{1},\delta_{x_\cdot}I+\gamma_{x_\cdot}^2\sigma_{x_\cdot}^2C-\gamma_{x_\cdot}^2\sigma_{x_\cdot}^2C_{(i)}C'_{(i)}) \\ &\cdot\prod_{i=1}^{n_x}\phi_1(x_i^o;\mu_{x_\cdot},\sigma_{x_\cdot}^2).\end{aligned}$$

The high dimensional Gaussian quantile, $\Phi_{q_x}(\gamma_{x_\cdot}C_{(i)}(x_i^o-\mu_{x_\cdot});\nu_{x_\cdot}\mathbf{1},\delta_{x_\cdot}I+\gamma_{x_\cdot}^2\sigma_{x_\cdot}^2C-\gamma_{x_\cdot}^2\sigma_{x_\cdot}^2C_{(i)}C'_{(i)})$, is hard to calculate. The method to calculate quantiles for the Gaussian distribution presented in previous section does not work here. It is not possible to express the x-variable as a Markov chain, even for a correlation matrix C constructed from an exponential correlation function. So an approximated marginal need to be used. Consider the equivalent parameterization of $(\boldsymbol{t},\boldsymbol{v})$ for $\boldsymbol{x}$, as given in Expression (2.4). The marginal variable t_i is normally mostly correlated with $v_i, v_{i+1}, v_{i-1}, v_{i-2}, v_{i+2}, ...$, in decreasing order, such that an approximation can be defined by setting the correlation between t_i and v_j for $|j-i| > \lfloor q_x^*/2 \rfloor$ to zero for a given q_x^*. Where $\lfloor\cdot\rfloor$ rounds the element to the nearest integer towards zero. Thus can the marginal distribution be approximated as

$$\mathrm{CSN}_{1,q_x}(\mu_{x_\cdot},\sigma_{x_\cdot}^2,\Gamma_{x_i},\boldsymbol{\nu}_x,\Delta_{x_i}) \approx \mathrm{CSN}_{1,q_x^*}(\mu_{x_\cdot},\sigma_{x_\cdot}^2,\Gamma_{x_i}^*,\boldsymbol{\nu}_{x_i}^*,\Delta_{x_i}^*), \tag{5.3}$$

where

$$\Gamma^*_{x_i} = \gamma_{x_\cdot} \begin{bmatrix} C(\lfloor \frac{q^*_x}{2} \rfloor) \\ \vdots \\ C(1) \\ C(0) \\ C(1) \\ \vdots \\ C(\lfloor \frac{q^*_x}{2} \rfloor) \end{bmatrix} = \gamma_{x_\cdot} C^*_{(i)} \in \mathbb{R}^{q^*_x},$$

$$\boldsymbol{\nu}^*_{x_i} = \begin{bmatrix} \nu \\ \nu \\ \vdots \\ \nu \end{bmatrix} \in \mathbb{R}^{q^*_x},$$

$$\Delta^*_{x_i} = \delta_{x_\cdot} I + \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C^* - \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C^*_{(i)} C'^*_{(i)} \in \mathbb{R}^{q^*_x \times q^*_x}.$$

Hence can the marginals be approximated as

$$\mathrm{p}(x_i) = \mathrm{p}(t_i | \boldsymbol{v} \geq \mathbf{0}) \approx \mathrm{p}(t_i | v_{i-\lfloor \frac{q^*_x}{2} \rfloor} \geq 0, ..., v_{i-1} \geq 0, v_i \geq 0, v_{i+1} \geq 0, ..., v_{i+\lfloor \frac{q^*_x}{2} \rfloor} \geq 0).$$

By using this approximated marginal distribution will the pseudo likelihood be

$$\begin{aligned} \mathrm{pL}^*(\mu_{x_\cdot}, \sigma^2_{x_\cdot}, \gamma_{x_\cdot}, \nu_{x_\cdot}, \delta_{x_\cdot} | \boldsymbol{x^o}) = &\left[\Phi_{q^*_x}(\mathbf{0}; \nu_{x_\cdot} \mathbf{1}, \delta_{x_\cdot} I + \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C^*) \right]^{-n_x} \\ &\cdot \prod_{i=1}^{n_x} \Phi_{q^*_x}(\gamma_{x_\cdot} C^*_{(i)} (x^o_i - \mu_{x_\cdot}); \nu_{x_\cdot} \mathbf{1}, \delta_{x_\cdot} I + \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C^* - \gamma^2_{x_\cdot} \sigma^2_{x_\cdot} C^*_{(i)} C'^*_{(i)}) \\ &\cdot \prod_{i=1}^{n_x} \phi_1(x^o_i; \mu_{x_\cdot}, \sigma^2_{x_\cdot}) \end{aligned}$$

Then the maximum pseudo likelihood, MpL, is given as

$$(\hat{\mu}_{x_\cdot}, \hat{\sigma}^2_{x_\cdot}, \hat{\gamma}_{x_\cdot}, \hat{\nu}_{x_\cdot}, \hat{\delta}_{x_\cdot}) = \underset{\mu_{x_\cdot}, \sigma^2_{x_\cdot}, \gamma_{x_\cdot}, \nu_{x_\cdot}, \delta_{x_\cdot}}{\arg\max} \mathrm{pL}^*(\mu_{x_\cdot}, \sigma^2_{x_\cdot}, \gamma_{x_\cdot}, \nu_{x_\cdot}, \delta_{x_\cdot} | \boldsymbol{x^o})$$

It is observed that this MpL is not defined for all possible choices of parameters, and this problem is avoided by requiring $\hat{\sigma}^2_{x_\cdot} > 0$ and $\hat{\delta}_{x_\cdot} > 0$. Then from the definition of the CS-Gaussian distribution we know that $\hat{\delta}_{x_\cdot} I + \hat{\gamma}^2_{x_\cdot} \hat{\sigma}^2_{x_\cdot} C - \hat{\gamma}^2_{x_\cdot} \hat{\sigma}^2_{x_\cdot} C_{(i)} C_{(i)}$ must be positive definite if C is positive definite. Then a CS-Gaussian distribution will be defined for the estimated parameters. A numerical method can be used to solve the optimization to determine the parameters, and with the constrained prior, $\nu_{x_\cdot} = 0$ and $\delta_{x_\cdot} = 1$, is it possible to find the global maximum. The localized pseudo likelihood will work for every dependency functions, $c(\cdot)$, so a exponential dependency function is not required as for the full likelihood.

5.1.3 Empirical Test for fL and Localized pL

We have constructed a test example with $\mu_{x_\cdot} = 0$, $\sigma^2_{x_\cdot} = 1$ and $\gamma_{x_\cdot} = 3$ to compare the two presented estimators. For different range parameters, r, have we generated 100 data points and the estimated parameters are compared in Table 5.1. The tests are repeated 100 times for each parameter set, and used to estimate an empirical 90% confidence interval.

For the case with $r = 0.00001$ the estimated parameters are identical, which is natural since the data points are almost independent. Hence will the pseudo likelihood be almost identical to the full likelihood. For a higher range parameter will the full likelihood give better results than the pseudo likelihood. That is also expected since the approximation of independency in the pseudo likelihood will be less correct for higher range parameter. The $\mu_{x_\cdot}$ parameter seems to be biased for the pseudo likelihood and highly correlated data. For the case with $r = 30$, the correct value of $\mu_{x_\cdot}$ is not in the empirical confidence interval. The $\sigma^2_{x_\cdot}$ parameter is estimated quite reliably for all range values, but the precision is lower for the higher range parameters. The skewness parameter, $\gamma_{x_\cdot}$, is estimated with a high bias and low precision with the pL estimator. For the pL with range parameter larger than 1 the correct value of $\gamma_{x_\cdot}$ is not in the empirical confidence interval. It is natural that also the full likelihood will be less correct for a higher range parameter, since there are less information in a highly correlated data set than in an uncorrelated data set. Note that when the distance between two points is $3r$, the correlation is still about 5%. So the practical range with $r = 30$ is usually considered to be 90, hence the data is highly correlated.

	Real values					
	$\mu_{x.} = 0$		$\sigma^2_{x.} = 1$		$\gamma_{x.} = 3$	
	$r = 0.00001$					
fL	$\hat{\mu}_{x.} = 0.02$	$[-0.17, 0.19]$	$\hat{\sigma}^2_{x.} = 0.98$	$[0.65, 1.33]$	$\hat{\gamma}_{x.} = 3.18$	$[1.96, 4.50]$
pL*	$\hat{\mu}_{x.} = 0.02$	$[-0.17, 0.19]$	$\hat{\sigma}^2_{x.} = 0.98$	$[0.65, 1.33]$	$\hat{\gamma}_{x.} = 3.18$	$[1.96, 4.50]$
	$r = 1$					
fL	$\hat{\mu}_{x.} = 0.17$	$[-0.10, 0.38]$	$\hat{\sigma}^2_{x.} = 0.85$	$[0.52, 1.17]$	$\hat{\gamma}_{x.} = 2.87$	$[0.83, 5.04]$
pL*	$\hat{\mu}_{x.} = -0.06$	$[-0.21, 0.08]$	$\hat{\sigma}^2_{x.} = 1.18$	$[0.89, 1.50]$	$\hat{\gamma}_{x.} = 5.52$	$[3.07, 9.91]$
	$r = 10$					
fL	$\hat{\mu}_{x.} = -0.12$	$[-0.49, 0.21]$	$\hat{\sigma}^2_{x.} = 0.95$	$[0.77, 1.13]$	$\hat{\gamma}_{x.} = 2.78$	$[0.70, 5.61]$
pL*	$\hat{\mu}_{x.} = 0.23$	$[-0.01, 0.55]$	$\hat{\sigma}^2_{x.} = 1.45$	$[0.85, 2.18]$	$\hat{\gamma}_{x.} = 6.14$	$[4.13, 8.40]$
	$r = 30$					
fL	$\hat{\mu}_{x.} = -0.29$	$[-1.05, 0.19]$	$\hat{\sigma}^2_{x.} = 0.95$	$[0.73, 1.15]$	$\hat{\gamma}_{x.} = 2.49$	$[0.63, 4.49]$
pL*	$\hat{\mu}_{x.} = 0.43$	$[0.18, 0.73]$	$\hat{\sigma}^2_{x.} = 0.99$	$[0.54, 1.57]$	$\hat{\gamma}_{x.} = 6.90$	$[4.26, 11.52]$

Table 5.1: Estimated parameters form 100 data points with exponential correlation, $C(i,j) = \exp\{-|i-j|/r\}$. The estimation is repeated 100 times for each case, and the empirical mean of the parameters are presented with an empirical 90% confidence interval.

5.2 tv-space Prior

For this prior we will also define a full likelihood and a pseudo likelihood for parameter estimation.

5.2.1 Full Likelihood Model.

For the tv-space prior is the full likelihood given as:

$$\begin{aligned} \mathrm{L}(\mu_{t.}, \sigma^2_{t.}, \gamma_{tv.}, \mu_{v.}, \sigma^2_{v.} | \boldsymbol{x}^o) = &\left[\Phi_{q_x}(\mathbf{0}; -\mu_{v.}\mathbf{1}, \sigma^2_{v.} I)\right]^{-1} \\ &\cdot \Phi_{q_x}\left(\frac{\gamma^2_{tv.}}{\sigma^2_{t.}} C^{-1}(\boldsymbol{x}^o - \mu_{t.}\mathbf{1}); -\mu_{v.}\mathbf{1}, \sigma^2_{v.} I - \frac{\gamma^2_{tv.}}{\sigma^2_{t.}} C^{-1}\right) \\ &\cdot \phi_{n_x}(\boldsymbol{x}^o; \mu_t \mathbf{1}, \sigma^2_{t.} C) \end{aligned} \tag{5.4}$$

The problem here is to evaluate $\Phi_{q_x}\left(\frac{\gamma_{tv.}^2}{\sigma_{t.}^2}C^{-1}\cdot(\boldsymbol{x}^o-\mu_{t.}\cdot\mathbf{1});-\mu_{v.}\mathbf{1},\sigma_{v.}^2 I-\frac{\gamma_{tv.}^2}{\sigma_{t.}^2}C^{-1}\right)$. There is no Markov property with this covariance, hence is the earlier presented method for Gaussian quantiles not working here. So we need to use a pseudo likelihood approach.

5.2.2 Pseudo Likelihood Model.

For the prior specified in the tv-space we have to approximate the likelihood with a pseudo likelihood:

$$\begin{aligned}\mathrm{pL}(\mu_{t.},\sigma_{t.}^2,\gamma_{tv.},\mu_{v.},\sigma_{v.}^2|\boldsymbol{x}^o) = &\left[\Phi_1(0;-\mu_{v.},\sigma_{v.}^2)\right]^{-n_x}\\ &\cdot\prod_{i=1}^{n_x}\Phi_1\left(\frac{\gamma_{tv.}}{\sigma_{t.}^2}\cdot(x_i^o-\mu_{t.});-\mu_{v.},\sigma_{v.}^2-\frac{\gamma_{tv.}^2}{\sigma_{t.}^2}\right)\\ &\cdot\prod_{i=1}^{n_x}\phi_1(x_i^o;\mu_{t.},\sigma_{t.}^2).\end{aligned}$$

Note that this pseudo likelihood can be calculated without any approximations, since this pL factorize, as is required with the previous prior. A numerical method is used to solve the optimization to determine the parameters, but it is often difficult to ensure that the global maximum is identified. Therefore a constrained pseudo likelihood approach with $\mu_{v.}=0$ and $\sigma_{v.}^2=1$ can be used as described in Section 4.2. It is observed that this pL is not defined for all possible choices of parameters, and this problem is avoided by requiring $\hat{\sigma}_{t.}^2>0$ and $\hat{\sigma}_{v.}^2-\frac{\hat{\gamma}_{tv.}^2}{\hat{\sigma}_{t.}^2}>0$. With $\sigma_{v.}^2=1$ it is sufficient to require $\hat{\sigma}_{t.}^2>\hat{\gamma}_{tv.}^2>0$. This ensures that the marginals have valid parameters. But to be sure that the simultaneous distribution has valid parameters, is it necessary to require that

$$\begin{bmatrix}\sigma_{t.}^2 C & \gamma_{tv.} I\\ \gamma_{tv.} I & \sigma_{v.}^2 I\end{bmatrix} \tag{5.5}$$

is positive definite.

Then the maximum pseudo likelihood, MpL, is given as

$$(\hat{\mu}_{t.},\hat{\sigma}_{t.}^2,\hat{\gamma}_{tv.},\hat{\mu}_{v.},\hat{\sigma}_{v.}^2) = \underset{\mu_{t.},\sigma_{t.}^2,\gamma_{tv.},\mu_{v.},\sigma_{v.}^2}{\arg\max}\ \mathrm{pL}(\mu_{t.},\sigma_{t.}^2,\gamma_{tv.},\mu_{v.},\sigma_{v.}^2|\boldsymbol{x}^o).$$

With $\mu_{t.}$, $\sigma_{t.}^2$ and $\gamma_{tv.}$ as the free parameters, this is almost the same problem as described in Pewsey (2000). The numerical methods to find the maximum of the pL requires initial values for the parameters, but instead of having a grid of initial

values we will here present an alternative algorithm to find the global maximum. It seems that there often will occur two maxima for the log-pseudo likelihood. One gives an almost Gaussian fit ($\gamma_{tv.} \approx 0$) and one with skewness. To understand this we will first look at the log-pseudo likelihood:

$$\begin{aligned}
\log\left(\mathrm{pL}(\mu_{t.}, \sigma_{t.}^2, \gamma_{tv.}, \mu_{v.}, \sigma_{v.}^2 | \boldsymbol{x^o})\right) =& - n_x \log\left(\Phi_1(0; 0, \sigma_{v.}^2)\right) \\
&+ \sum_{i=1}^{n_x} \log\left(\Phi_1\left(\frac{\gamma_{tv.}}{\sigma_{t.}^2}(x_i^o - \mu_{t.}); 0, 1 - \frac{\gamma_{tv.}^2}{\sigma_{t.}^2}\right)\right) \\
&+ \sum_{i=1}^{n_x} \log\left(\phi_1(x_i^o; \mu_{t.}, \sigma_{t.}^2)\right) \\
=& n_x \log(2) \\
&+ \sum_{i=1}^{n_x} \log\left(\Phi_1\left(\frac{\gamma_{tv.}}{\sigma_{t.}^2}(x_i^o - \mu_{t.}); 0, 1 - \frac{\gamma_{tv.}^2}{\sigma_{t.}^2}\right)\right) \\
&- \frac{1}{2}\sum_{i=1}^{n_x}\left(\frac{x_i - \mu_{t.}}{\sigma_{t.}}\right)^2 - n_x \log(\sqrt{2\pi \cdot \sigma_{t.}^2})
\end{aligned}$$

Empirically study shows that $\sum_{i=1}^{n_x} \log\left(\Phi_1\left(\frac{\gamma_{tv.}}{\sigma_{t.}^2}(x_i^o - \mu_{t.}); 0, 1 - \frac{\gamma_{tv.}^2}{\sigma_{t.}^2}\right)\right)$ is unimodal. The last part, $-\frac{1}{2}\sum_{i=1}^{n_x}\left(\frac{x_i - \mu_{t.}}{\sigma_{t.}}\right)^2 - n_x \log(\sqrt{2\pi \cdot \sigma_{t.}^2})$, is recognized from the Gaussian log-likelihood and is unimodal. So the log pseudo likelihood seems to have two modes.

If there exist two modes the parameters can be estimated by initiate $\mu_{t.}$ and $\sigma_{t.}^2$ with the standard estimators for the Gaussian distribution. Since $\gamma_{tv.}$ is limited between $\sigma_{t.}$ and $-\sigma_{t.}$ we will run the numerical method twice with Gaussian estimates for $(\mu_{t.}, \sigma_{t.}^2)$ and the two extreme values of $\gamma_{tv.}$ as initial values.

Algorithm: Maximum pL$(\mu_{t_\cdot}, \sigma^2_{t_\cdot}, \gamma_{tv_\cdot}, 0, 1|\boldsymbol{x}^o)$

- Initiate the parameters:
  - ⋆ $\hat{\mu}_{t_\cdot,0} = \frac{1}{n_x}\sum_{i=1}^{n_x} x_i^o$
  - ⋆ $\hat{\sigma}^2_{t_\cdot,0} = \frac{1}{n_x-1}\sum_{i=1}^{n_x}\left(x_i^o - \hat{\mu}_{t_\cdot,0}\right)^2$
  - ⋆ $\hat{\gamma}_{tv_\cdot,0} = \hat{\sigma}_{t_\cdot,0}$
- Do the optimization to identify $(\hat{\mu}_{t_\cdot}, \hat{\sigma}^2_{t_\cdot}, \hat{\gamma}_{tv_\cdot})^+$. Make sure that the parameters meet the positive definite requirement given in Expression (5.5).
- Initiate the parameters $\hat{\mu}_{t_\cdot,0}$, $\hat{\sigma}^2_{t_\cdot,0}$ and $\hat{\gamma}_{tv_\cdot,0} = -\hat{\sigma}_{t_\cdot,0}$
- Do the optimization to identify $(\hat{\mu}_{t_\cdot}, \hat{\sigma}^2_{t_\cdot}, \hat{\gamma}_{tv_\cdot})^-$. Make sure that the parameters meet the positive definite requirement given in Expression (5.5).
- Choose the parameter set with highest likelihood, $(\hat{\mu}_{t_\cdot}, \hat{\sigma}^2_{t_\cdot}, \hat{\gamma}_{tv_\cdot})$

The two maxima is illustrated by an example in Figure 5.2, where a likelihood for a profile between the two maxima is displayed. The two maxima are found by the given algorithm, and the value of the likelihood on a linear transect between these maxima is calculated. Note that the non-Gaussian maximum has the highest likelihood and also is the solution closest to the correct values of the example which is $\mu_{t_\cdot} = 0$, $\sigma^2_{t_\cdot} = 2$ and $\gamma_{tv_\cdot} = -0.75$.

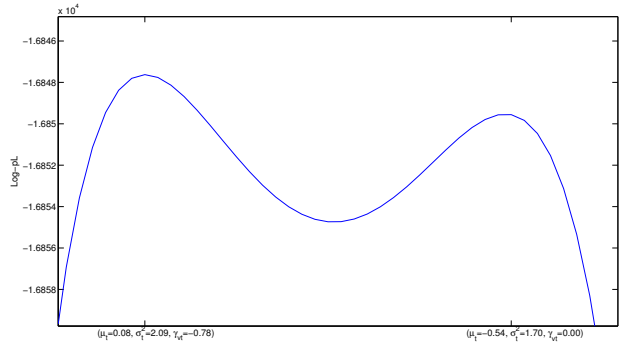


Figure 5.2: Log-pL for an example with real values: $\mu_{t_\cdot} = 0$, $\sigma^2_{t_\cdot} = 2$ and $\gamma_{tv_\cdot} = -0.75$.

5.2.3 Empirical Test for pL

The estimation procedure was tested on the same parameters as in the empirical test for the x-space prior in Section 5.1.3, but it seems to be impossible to a have very skew data with a high correlation, and meet the requirements in Expression (5.5). So we will study another test example than for the x-space prior. We have constructed a test example with 100 data points with $\mu_{t.} = 0$ and $\sigma_{t.}^2 = 4$ and range parameter $r = 3$. We have done parameter estimation with different values of $\gamma_{tv.}$ between -0.8 and 0.8. The test is repeated 100 times for each parameter set. The empirical mean of the parameter estimation is shown in Table 5.2, with an empirical 90% confidence interval. All parameter estimates ensures the constrain in Expression (5.5).

Real values				
$\mu_{t.} = 0$		$\sigma_{t.}^2 = 4$		
				$\gamma_{tv.} = 0.8$ (Real value)
$\hat{\mu}_{t.} = 0.01$	$[-0.76, 0.62]$	$\hat{\sigma}_{t.}^2 = 3.96$	$[2.11, 5.95]$	$\hat{\gamma}_{tv.} = 0.75$ $[0.30, 0.96]$
				$\gamma_{tv.} = 0.5$ (Real value)
$\hat{\mu}_{t.} = 0.21$	$[-0.65, 1.40]$	$\hat{\sigma}_{t.}^2 = 3.93$	$[2.35, 5.53]$	$\hat{\gamma}_{tv.} = 0.22$ $[-0.90, 0.94]$
				$\gamma_{tv.} = 0.1$ (Real value)
$\hat{\mu}_{t.} = -0.03$	$[-1.07, 1.11]$	$\hat{\sigma}_{t.}^2 = 4.24$	$[2.90, 5.93]$	$\hat{\gamma}_{tv.} = 0.18$ $[-0.92, 0.97]$
				$\gamma_{tv.} = -0.1$ (Real value)
$\hat{\mu}_{t.} = 0.02$	$[-1.32, 1.14]$	$\hat{\sigma}_{t.}^2 = 4.16$	$[2.73, 5.53]$	$\hat{\gamma}_{tv.} = -0.12$ $[-0.96, 0.89]$
				$\gamma_{tv.} = -0.5$ (Real value)
$\hat{\mu}_{t.} = -0.19$	$[-1.35, 0.87]$	$\hat{\sigma}_{t.}^2 = 4.10$	$[2.66, 5.93]$	$\hat{\gamma}_{tv.} = -0.16$ $[-0.94, 0.91]$
				$\gamma_{tv.} = -0.8$ (Real value)
$\hat{\mu}_{t.} = 0.01$	$[-0.77, 0.76]$	$\hat{\sigma}_{t.}^2 = 3.81$	$[2.61, 5.22]$	$\hat{\gamma}_{tv.} = -0.72$ $[-0.93, 0.10]$

Table 5.2: Parameter estimation with different values of $\gamma_{tv.}$. 100 data points with $\mu_{t.} = 0$ and $\sigma_{t.}^2 = 4$ and $r = 3$.

It is observed that for data with low skewness the precision in the estimation is very low. Specially the $\gamma_{tv.}$ seems to be hard to estimate. For small values of $|\gamma_{tv.}|$ the value of the pL for the two maxima are often close to each other. Hence will the wrong maxima ($\gamma_{tv.} \approx 0$) be chosen more often for small values of $|\gamma_{tv.}|$. But the answer for small values of $|\gamma_{tv.}|$ will not be affected so much as for higher values of $|\gamma_{tv.}|$. So it seems that the cases with $\gamma_{tv.} = 0.5$ and $\gamma_{tv.} = -0.5$ will be influenced most.

5.3 Closing Remarks

For parameter estimation it seems to be hard to fulfill the requirements in Expression (5.5) for the tv-space prior. Specially to combine high skewness and high correlation seems to be really difficult. So it seems that the x-space prior with the full likelihood estimator is the best choice, since the requirements for the parameters are easier to fulfill. The full likelihood also takes into account the correlation in the data. But the full likelihood requires a exponential correlation measure. For other correlation functions must the pL be used. It will also be possible to let the range parameter be a free parameter, and then optimize the likelihood with respect to $\mu_{x_\cdot}$, $\sigma^2_{x_\cdot}$, $\gamma_{x_\cdot}$ and r. The optimization will be much more computer demanding.

Chapter 6

The CST-distribution

In Karimi et al. (2009) is the inversion done for the elastic material properties: P-wave velocity, S-wave velocity and density. The results shows that the distribution of the P-wave velocity has a bit heavier tails than the CSN-distribution. Therefore it would be useful to have a skew distribution with heavier tails. To meet this requirement the multivariate closed-skew (CS) T-distribution is defined.

The CST-distribution is a generalization of the T-distribution. In a similar way as the CS-Gaussian distribution it adds skewness into the model. First we present some properties of the multivariate T-distribution:

6.1 Multivariate T-distribution

The multivariate T-distribution is a generalization of the Student's t-distribution and a random vector $\boldsymbol{y} \in \mathbb{R}^n$ is multivariate T-distributed

$$\boldsymbol{y} \sim \mathrm{T}_n(\boldsymbol{\mu}, \Omega, \eta)$$

if the pdf is

$$\tau_n(\boldsymbol{y}; \boldsymbol{\mu}, \Omega, \eta) = \frac{\Gamma\left(\frac{\eta+n}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)(\eta \cdot \pi)^{n/2}} |\Omega|^{-\frac{1}{2}} \left(1 + \frac{1}{\eta}(\boldsymbol{y} - \boldsymbol{\mu})' \Omega^{-1} (\boldsymbol{y} - \boldsymbol{\mu})\right)^{-\frac{\eta+n}{2}}$$

where $\Gamma(\cdot)$ is the gamma function, $\boldsymbol{\mu} \in \mathbb{R}^n$ is a centering vector, $\Omega \in \mathbb{R}^{n \times n}$ is a positive definite dependence matrix and $\eta \in \mathbb{R}_+$ is the degree of freedom.

Some relevant characteristics are given in Roislien and Omre (2006):

1. Special case of the T-distribution:

$$\mathrm{T}_n(\mu, \Omega, \eta) \overset{\eta \to \infty}{\to} \mathrm{N}_n(\mu, \Omega)$$

2. Moments of T-distribution: Let $\boldsymbol{x} \sim \mathrm{T}_n(\boldsymbol{\mu}, \Omega, \eta)$ then

$$\mathrm{E}(\boldsymbol{x}) = \boldsymbol{\mu}; \quad \eta \geq 2$$
$$\mathrm{Cov}(\boldsymbol{x}) = \frac{\eta}{\eta - 2} \cdot \Omega; \quad \eta \geq 3$$

3. Linear combinations of T-distributed random variables are also T-distributed. Let $\boldsymbol{x} \sim \mathrm{T}_n(\boldsymbol{\mu}, \Omega, \eta)$ and A be a deterministic $l \times n$ matrix then

$$[A\boldsymbol{x}] \sim \mathrm{T}_l(A\boldsymbol{\mu}, A\Omega A', \eta)$$

4. T-distributed random variables conditional of the components are also T-distributed. Let $\boldsymbol{x} \sim \mathrm{T}_n(\boldsymbol{\mu}, \Omega, \eta)$, with $\boldsymbol{x}_1 \in \mathbb{R}^k$ and $\boldsymbol{x}_2 \in \mathbb{R}^{n-k}$ being two subvectors of $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2']'$. Further let the parameters associated with the subvectors be:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} \Omega_{11} & \Gamma_{12} \\ \Gamma_{21} & \Omega_{22} \end{bmatrix}.$$

   Then

$$[\boldsymbol{x}_1 | \boldsymbol{x}_2] \sim \mathrm{T}_k \left(\boldsymbol{\mu}_1 + \Gamma_{12}\Omega_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}), \kappa(\eta) \cdot (\Omega_{11} - \Gamma_{12}\Omega_{22}^{-1}\Gamma_{21}), \eta + n - k \right),$$

   where

$$\kappa(\eta) = \frac{1}{1 + \frac{n-k}{\eta}} \left(1 + \frac{1}{\eta}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)'\Omega_{11}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) \right).$$

5. Components of T-distributed random variables are non-independent. Let $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2']$ be as above, then

$$\mathrm{T}_n(\boldsymbol{\mu}, \Omega, \eta) \neq \mathrm{T}_n(\boldsymbol{\mu}_1, \Omega_{11}, \eta) \cdot \mathrm{T}_n(\boldsymbol{\mu}_2, \Omega_{22}, \eta); \quad \eta < \infty$$

   for all eligible (μ, Ω). Even for the particular case where Γ_{12} and Γ_{21} being a zero-matrix this holds. The subvectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ will not be independent even if Ω is a block matrix. When $\eta \to \infty$ we get the multivariate Gaussian case and independence can be obtained.

6.2 The CST-distribution

The closed-skew T-distribution adds skewness into the T-distribution. The most convenient way to define the CST-distribution is to consider a random vector

$\boldsymbol{t} \in \mathbb{R}^n$ and a random vector $\boldsymbol{v} \in \mathbb{R}^q$. Let the joint pdf of $(\boldsymbol{t}, \boldsymbol{v})$ be multivariate T-distributed as

$$\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{T}\left(\begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \Omega_t & \Gamma_{tv} \\ \Gamma_{vt} & \Omega_v \end{bmatrix}, \eta \right).$$

Then the CST-distributed variable of interest $\boldsymbol{x} \in \mathbb{R}^n$ is defined as

$$\boldsymbol{x} = [\boldsymbol{t} | \boldsymbol{v} \geq \boldsymbol{0}].$$

Let $\mathrm{p}(\boldsymbol{x})$ be the generic term for a pdf of the argument, and the term $\mathrm{p}(\boldsymbol{x}|\boldsymbol{d})$ represents the conditional pdf of $\boldsymbol{x}$ given $\boldsymbol{d}$. The pdf of $\boldsymbol{x}$ is then given as:

$$\begin{aligned} \boldsymbol{x} \sim \mathrm{p}(\boldsymbol{x}) =& \mathrm{p}(\boldsymbol{t}|\boldsymbol{v} \geq \boldsymbol{0}) = \frac{\mathrm{p}(\boldsymbol{v} \geq \boldsymbol{0}|\boldsymbol{t})\mathrm{p}(\boldsymbol{t})}{\mathrm{p}(\boldsymbol{v} \geq \boldsymbol{0})} \\ =& [1 - \mathcal{T}_q(\boldsymbol{0}; \boldsymbol{\mu}_v, \Omega_v, \eta)]^{-1} \left[1 - \mathcal{T}_q(\boldsymbol{0}; \boldsymbol{\mu}_{v|t}, \Omega_{v|t}, \eta + n)\right] \tau_n(\boldsymbol{t}; \boldsymbol{\mu}_t, \Omega_t, \eta), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{v|t} =& \boldsymbol{\mu}_v + \Gamma_{vt}\Omega_t^{-1}(t - \mu_t) \\ \Omega_{v|t} =& \frac{1}{1 + \frac{n}{\eta}} \left(1 + \frac{1}{\eta}(\boldsymbol{t} - \boldsymbol{\mu}_t)'\Omega^{-1}(\boldsymbol{t} - \boldsymbol{\mu}_t)\right)(\Omega_v - \Gamma_{vt}\Omega_t^{-1}\Gamma_{tv}) \end{aligned}$$

Here $\mathcal{T}_q(\cdot; \boldsymbol{\mu}, \Omega, \eta)$ is the q-dimensional cumulative T-distribution. The closed-skew T-distribution can be parameterized as:

$$\mathrm{CST}_{n,q}(\boldsymbol{\mu}, \Omega, \Gamma, \boldsymbol{\nu}, \Delta, \eta) = [\mathcal{T}_q(\boldsymbol{0}; \boldsymbol{\nu}, \Delta + \Gamma\Omega\Gamma', \eta)]^{-1}\mathcal{T}_q(\Gamma(\boldsymbol{x} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta, \eta + n)\tau_n(\boldsymbol{x}; \boldsymbol{\mu}, \Omega, \eta)$$

with $\boldsymbol{\mu} = \boldsymbol{\mu}_t$, $\Omega = \Omega_t$, $\Gamma = \Gamma_{vt}\Omega_t^{-1}$,$\boldsymbol{\nu} = -\boldsymbol{\mu}_v$ and $\Delta = \Omega_{v|t}$.

The transformation between the CST-distribution for $\boldsymbol{x}$ and the T-distribution for $(\boldsymbol{t}, \boldsymbol{v})$ is one-to-one given by:

$$\boldsymbol{x} = [\boldsymbol{t}|\boldsymbol{v} \geq 0] \sim \mathrm{CST}_{n,q}(\boldsymbol{\mu}, \Omega, \Gamma, \boldsymbol{\nu}, \Delta, \eta) \Leftrightarrow \begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{T}_{n+q}\left(\begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \Omega_t & \Gamma_{tv} \\ \Gamma_{vt} & \Omega_v \end{bmatrix}, \eta \right) \tag{6.1}$$

with

$$\begin{array}{ll} \boldsymbol{\mu} = \boldsymbol{\mu}_t \in \mathbb{R}^n & \boldsymbol{\mu}_t = \boldsymbol{\mu} \in \mathbb{R}^n \\ \Omega = \Omega_t \in \mathbb{R}^{n\times n} & \Omega_t = \Omega \in \mathbb{R}^{n\times n} \\ \Gamma = \Gamma_{vt}\Omega_t^{-1} \in \mathbb{R}^{q\times n} \quad \Longleftrightarrow & \Gamma_{vt} = \Gamma_{tv}' = \Gamma\Omega \in \mathbb{R}^{q\times n} \\ \boldsymbol{\nu} = -\boldsymbol{\mu}_v \in \mathbb{R}^q & \boldsymbol{\mu}_v = -\boldsymbol{\nu} \in \mathbb{R}^q \\ \Delta = \Omega_v - \Gamma_{vt}\Omega_t^{-1}\Gamma_{tv} \in \mathbb{R}^{q\times q} & \Omega_v = \Delta + \Gamma\Omega\Gamma' \in \mathbb{R}^{q\times q} \end{array}$$

6.3 Properties of the CST-distribution

Some favorable characteristics of the CST-distribution that are relevant for inversion problems are given here and proven in Appendix A:

1. Linear combinations of components CST-distributed random variables are also CST-distributed random variables. If $\boldsymbol{x} \sim \mathrm{CST}_{n,q}(\boldsymbol{\mu}, \Omega, \Gamma, \boldsymbol{\nu}, \Delta, \eta)$ and A is a deterministic $l \times n$ matrix with $(l \leq n)$. Then

$$[\boldsymbol{y} = A\boldsymbol{x}] \sim \mathrm{CSN}_{l,q}(\boldsymbol{\mu}_y, \Omega_y, \Gamma_y, \boldsymbol{\nu}_y, \Delta_y, \eta_y),$$

   where

$$\begin{aligned} \boldsymbol{\mu}_y &= A\boldsymbol{\mu}, \\ \Omega_y &= A\Omega A', \\ \Gamma_y &= \Gamma\Omega A'\Omega_y^{-1}, \\ \boldsymbol{\nu}_y &= \boldsymbol{\nu}, \\ \Delta_y &= \Delta + \Gamma\Omega\Gamma' - \Gamma\Omega A'\Omega_y^{-1}A\Omega\Gamma' \\ \eta_y &= \eta. \end{aligned}$$

   So the CST-distribution is closed under linear transformations. This is proven in Appendix A.1. Note that a result is that also marginal pdfs will be CST by choosing $A = \boldsymbol{b}_{(i)}'$ with $\boldsymbol{b}_{(i)}$ an $n \times 1$ vector with entries zeros except for element number i, so that $x_i = \boldsymbol{b}_{(i)}'\boldsymbol{x}$.

2. Components of CST-distributed random variables are presumable non-independent. From property 5 in Section 6.1 we know that components of T-distributed random variables are non-independent, hence is unlikely that there is possible to construct a CST-distributed random variable with independent components. We have not been able to prove this.

3. CST-distributed random variables conditional of the components are also CST-distributed variables. Let $\boldsymbol{x} \sim \mathrm{CST}_{n,q}(\boldsymbol{\mu}, \Omega, \Gamma, \boldsymbol{\nu}, \Delta, \eta)$, with $\boldsymbol{x}_1 \in \mathbb{R}^k$ and $\boldsymbol{x}_2 \in \mathbb{R}^{n-k}$ being two subvectors of $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2']'$. Further let the parameters associated with the subvectors be:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \quad \Gamma = [\Gamma_1, \Gamma_2].$$

Then

$$[\boldsymbol{x}_1 | \boldsymbol{x}_2] \sim \mathrm{CST}_{k,q}(\boldsymbol{\mu}_{1|2}, \Omega_{1|2}, \Gamma_{1|2}, \boldsymbol{\nu}_{1|2}, \Delta_{1|2}, \eta_{1|2}).$$

Where

$$\begin{aligned}
\boldsymbol{\mu}_{1|2} =& \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \\
\Omega_{1|2} =& \kappa(\eta)(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}), \\
\Gamma_{1|2} =& \Gamma_1, \\
\boldsymbol{\nu}_{1|2} =& \boldsymbol{\nu} - \left(\Gamma_2 + \Gamma_1\Omega_{12}\Omega_{22}^{-1}\right)(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \\
\Delta_{1|2} =& \kappa(\eta)\Delta \\
\eta_{1|2} =& \eta.
\end{aligned}$$

This is proven in Appendix A.2. So the CST-distribution is closed under linear transformation and conditioning.

6.4 Examples of the CST-distribution

The CST-distribution is closely related to the CSN-distribution and the T-distribution. To illustrate the characteristics of the CST-distribution some examples of the distribution will be shown. We will first consider a CST-distributed $x \in \mathcal{R}^1$, with the similar base case as in Section 2.4 with

$$\mu = 5, \quad \Omega = 9, \quad \Gamma = 1, \quad \nu = 0, \quad \Delta = 0.05. \tag{6.2}$$

Note that Ω, Γ and Δ here are one dimensional matrices. In Figure 6.1 are the density displayed for four different degrees of freedom, η. For $\eta = \infty$ we have a CSN distribution and the same density as displayed in Figure 2.1. It is observed that lower degrees of freedom gives heavier tails.

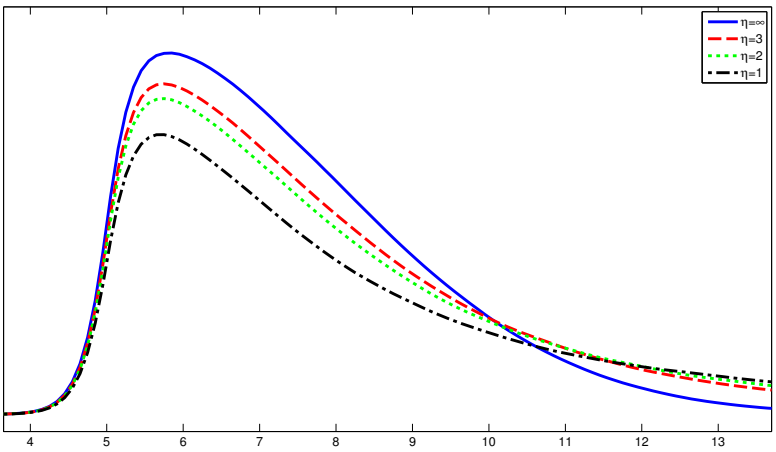
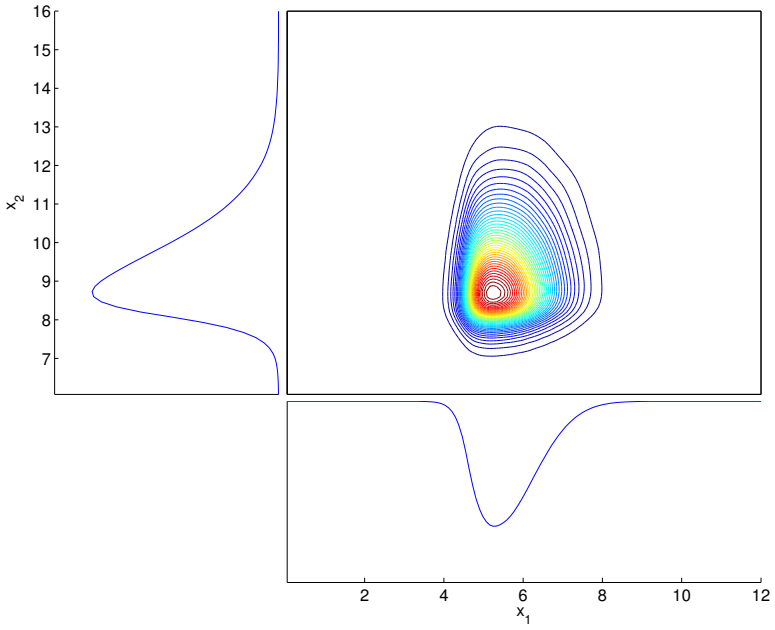


Figure 6.1: Pdf for different degrees of freedom. $\mu = 5$, $\Omega = 9$, $\Gamma = 1$, $\nu = 0$ and $\Delta = 0.05$.

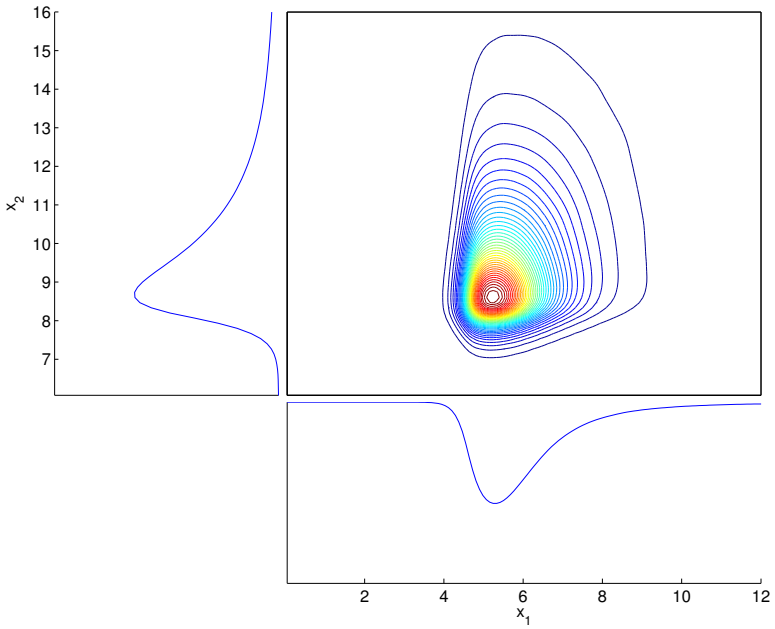
The same situations will occur for higher dimensional problems, and we will here also study a particular CST-distributed variable $\boldsymbol{x}$ with $n = q = 2$, and parameters

$$\boldsymbol{\mu} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 4 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}, \quad \boldsymbol{\nu} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}, \quad \Delta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The pdf of this two dimensional CST distribution is shown in Figure 6.2. The CSN-distribution with $\eta = \infty$ is displayed in Figure 6.2a and the CST-distribution with $\eta = 2$ is displayed in Figure 6.2b. It is observed that the distribution with $\eta = 2$ has heavier tails than the other.



(a) $\eta = \infty$



(b) $\eta = 2$

Figure 6.2: Bivariate pdf and marginal pdfs for a CST-distributed $\boldsymbol{x}$.

6.5 Bayesian CST Inversion

Let $\boldsymbol{x} \in \mathbb{R}^{n_x}$ be the variable of interest and consider a linear model

$$\boldsymbol{d} = H\boldsymbol{x}$$

where $\boldsymbol{d} \in \mathbb{R}^{n_d}$ is measured data, $n_d < n_x$. Note that there can not be an error term, $\boldsymbol{e}$. From property 2 in Section 6.3 we know that it probably not is possible to specify a multivariate CST-distribution of independent vectors $H\boldsymbol{x}$ and $\boldsymbol{e}$. Hence is it not possible to have an independent error term in the model. The prior model for $\boldsymbol{x}$ is defined as

$$\boldsymbol{x} \sim \mathrm{CST}_{n_x,q_x}(\boldsymbol{\mu}_x, \Omega_x, \boldsymbol{\nu}_x, \Delta_x, \eta).$$

Let this define a T-distributed vector as:

$$\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{r} = H\boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{T}_{n_x+n_d+q_x} \left(\begin{bmatrix} \boldsymbol{\mu}_x \\ H\mu_x \\ -\boldsymbol{\nu}_x \end{bmatrix}, \begin{bmatrix} \Omega_x & \Omega_x H' & \Omega_x \Gamma_x \\ H\Omega_x & H\Omega_x H' & H\Omega_x \Gamma_x' \\ \Gamma_x \Omega_x & \Gamma_x \Omega_x H' & \Delta_x + \Gamma_x \Omega_x \Gamma_x' \end{bmatrix}, \eta \right)$$

where

$$\begin{aligned} \boldsymbol{x} =& [\boldsymbol{t} | \boldsymbol{v} \geq \boldsymbol{0}], \quad \boldsymbol{t} \in \mathcal{R}^{n_x}, \boldsymbol{u} \in \mathcal{R}^{q_x} \\ \boldsymbol{d} =& [\boldsymbol{r} | \boldsymbol{v} \geq \boldsymbol{0}] \quad \boldsymbol{r} \in \mathcal{R}^{n_d}, \boldsymbol{v} \in \mathcal{R}^{q_d}. \end{aligned}$$

Then

$$\begin{bmatrix} \boldsymbol{t}|\boldsymbol{r} \\ \boldsymbol{v}|\boldsymbol{r} \end{bmatrix} \sim \mathrm{T}_{n_x+q_x} \left(\begin{bmatrix} \boldsymbol{\mu}_{t|r} \\ \boldsymbol{\mu}_{v|r} \end{bmatrix}, \begin{bmatrix} \Omega_{t|r} & \Gamma_{tv|r} \\ \Gamma_{vt|r} & \Omega_{t|r} \end{bmatrix}, \eta + n_d \right),$$

where

$$\begin{aligned} \boldsymbol{\mu}_{t|r} =& \mu_x + \Omega_x H'(H\Omega_x H')^{-1}(\boldsymbol{r} - H\mu_x) \\ \boldsymbol{\mu}_{v|r} =& -\boldsymbol{\nu}_x + \Gamma_x \Omega_x H'(H\Omega_x H')^{-1}(\boldsymbol{r} - H\boldsymbol{\mu}_x) \\ \Omega_{t|r} =& \kappa(\eta + n_d)(\Omega_x - \Omega H'(H\Omega_x H')^{-1} H\Omega_x) \\ \Omega_{v|r} =& \kappa(\eta + n_d)\left(\Delta_x + \Gamma_x \Omega_x \Gamma_x' - \Gamma_x \Omega_x H(H\Omega_x H')^{-1}(\Gamma_x \Omega_x H')'\right) \\ \Gamma_{v|r} =& \kappa(\eta + n_d)\left(\Gamma_x \Omega_x - \Gamma_x \Omega_x H'(H\Omega_x H')^{-1}\right). \end{aligned}$$

Then from the definition of the CST-distribution we have:

$$[\boldsymbol{x}|\boldsymbol{d}] = [\boldsymbol{t}|\boldsymbol{r}, \boldsymbol{v} \geq \boldsymbol{0}] \sim \mathrm{CST}_{n_x,q_x}(\boldsymbol{\mu}_{x|d}, \Omega_{x|d}, \Gamma_{x|d}, \boldsymbol{\nu}_{x|d}, \Delta_{x|d}, \eta + n_d + q_x)$$

where

$$
\begin{aligned}
\boldsymbol{\mu}_{x|d} =& \mu_x + \Omega_x H'(H\Omega_x H')^{-1}(\boldsymbol{r} - H\mu_x) \\
\Omega_{x|d} =& \kappa(\eta + n_d + q_x)(\Omega_x - \Omega H'(H\Omega_x H')^{-1} H\Omega_x) \\
\Gamma_{d|x} =& \kappa(\eta + n_d + q_x)\left(\Gamma_x\Omega_x - \Gamma_x\Omega_x H'(H\Omega_x H')^{-1}\right)\Omega_{x|d}^{-1} \\
\boldsymbol{\nu}_{x|d} =& \boldsymbol{\nu}_x - \Gamma_x\Omega_x H'(H\Omega_x H')^{-1}(\boldsymbol{r} - H\boldsymbol{\mu}_x) \\
\Delta_{x|d} =& \kappa(\eta + n_d + q_x)\left(\Delta_x + \Gamma_x\Omega_x\Gamma_x' - \Gamma_x\Omega_x H'(H\Omega_x H')^{-1}(\Gamma_x\Omega_x H')'\right) \\
& - \Gamma_{x|d}\Omega_{x|d}\Gamma_{x|d}'.
\end{aligned}
$$

In the limit $\eta \to \infty$ we get $\kappa(\eta) \to 1$ and the parameters are identical as presented in Section 2.5 for the CS Gaussian distribution. Note that there is only one parameter for the degrees of freedom, such that there is not possible to have different degrees of freedom for different marginals. It might be possible to make a CST-distribution with a vector of degrees of freedom, but we have not had time to study this here.

Chapter 7

Inversion Case Study

In Buland and Omre (2003) Bayesian Gaussian inversion is used on data from the Sleipner Øst field in the North Sea. The Sleipner Øst field is a gas condensate field in the southern part of the North Sea. The depth of the reservoir is in the range of 2270 to 2500 meter sub-sea. The prior model is inferred from elastic properties in a well, and the assumptions for a Gaussian model are justified by a probability plot. The data do not fit a Gaussian model very well, but the authors states that assuming a Gaussian prior model is necessary to make the inversion analytically tractable. In Karimi et al. (2009) it is shown that a CS-Gaussian approach makes it possible to capture the skewness in the data and still obtain analytical tractability.

The same data set that is used in Karimi et al. (2009) is used here to generate realizations from the posterior model and to test the efficiency of the simulation algorithms. Then the algorithms can be tested with real data and real parameters. The two different prior models and the different parameter estimators are tested with real data.

We have limited our study to consider the CS-Gaussian distribution. The CST-distribution is more general and could possibly give even better results.

7.1 Seismic Inversion Model

The data is collected from a seismic vessel, generating waves that propagates through the water and the seabed. In the layers in the seabed the properties of the medium change, and portions of the waves are reflected to the surface where hydrophones on the seismic streamers behind the vessel registers the amplitudes of the reflected waves. This results in a set of stacked seismic data. This gives a convolved linear relation between the data and the reflection coefficient, $\boldsymbol{r}$,

$$\boldsymbol{d} = W\boldsymbol{r} + \boldsymbol{e},$$

where W is a wavelet matrix shown in Figure 7.1. The seismic observation error is defined as $\boldsymbol{e} = W\boldsymbol{e}_1 + \boldsymbol{e}_2$, where $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are independent Gaussian distributed with mean zero and variance σ_e^2. Hence $\boldsymbol{e}$ is wavelet colored with mean zero and variance $\Sigma_e = \sigma_e^2(WW'+I)$. Since the CS-Gaussian distribution is a generalization of the Gaussian distribution $\boldsymbol{e}$ is also CS-Gaussian distributed as $\mathrm{CSN}(\mathbf{0}, \Sigma_e, \cdot, \cdot, \cdot)$, where the three latter parameters are unspecified since $q = 0$. The reflection

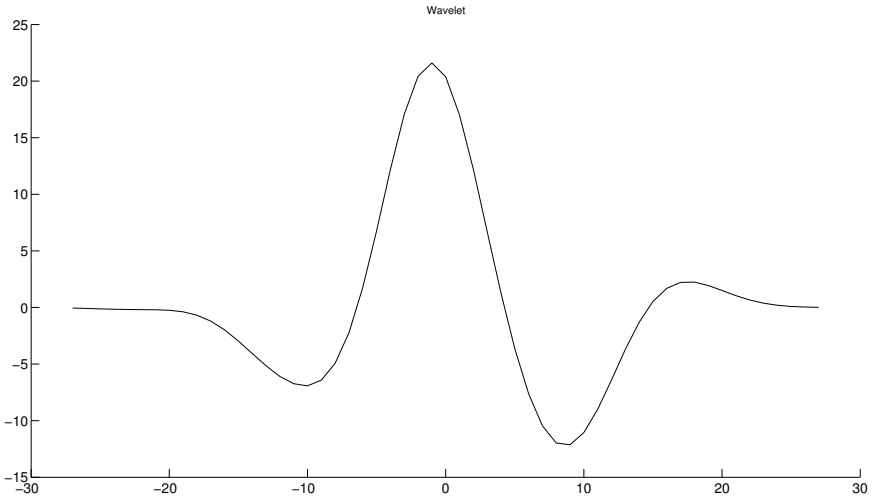


Figure 7.1: Wavelet

coefficient for layer j can be written as $r_j = a_j^{(r)}/a_j^{(i)}$, where $a_j^{(r)}$ is the reflected amplitude of the wave and $a_j^{(i)}$ is the amplitude of the incident wave between layer j and $j+1$. Let v_j be the velocity of the P-wave and let ρ_j be the density in layer j, as shown in Figure 7.2. Then will the impedance, z_j, in layer j be $v_j\rho_j$. The parameters v and ρ from a well at the Sleipner Øst field is used to calculate the impedance, z, and the resulting data are displayed in Figure 7.3. The depth interval used is $[2050, 2378]$ milliseconds with two milliseconds sampling interval making the number of data, $T = 165$.

For normal incidence Zoeppritz' equations reduce to a simple form as described in Sheriff et al. (1995). This gives

$$\begin{aligned} r_j = \frac{a_j^{(r)}}{a_j^{(i)}} &= \frac{v_{j+1}\rho_{j+1} - v_j\rho_j}{v_{j+1}\rho_{j+1} + v_j\rho_j} = \frac{z_{j+1} - z_j}{z_{j+1} + z_j} \\ &\approx \frac{\Delta z_j}{2z_j} \approx \frac{1}{2}\Delta \ln(z_j). \end{aligned}$$

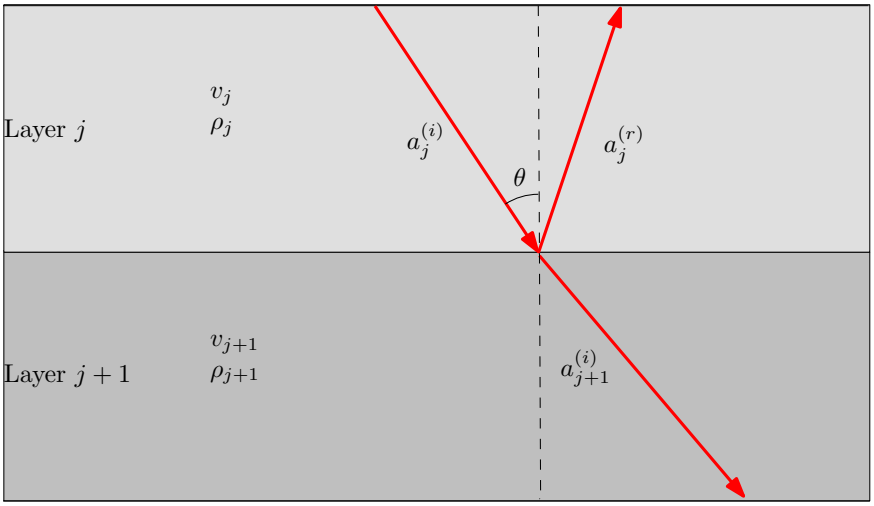


Figure 7.2: Layer j and $j+1$.

Where $\ln(\cdot)$ is the natural logarithm and $\Delta \ln z_j = \ln z_{j+1} - \ln z_j$. Then the linear relation between $\boldsymbol{d}$ and $\boldsymbol{r}$ becomes

$$\begin{aligned} \boldsymbol{d} &= W\boldsymbol{r} + \boldsymbol{e} = WD\ln \boldsymbol{z} + \boldsymbol{e} \\ &= H\boldsymbol{x} + \boldsymbol{e}, \end{aligned} \tag{7.1}$$

where $\boldsymbol{x} = \ln \boldsymbol{z}$, $H = WD$ and D is given as

$$D = \frac{1}{2}\begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \ddots & 0 & -1 & 1 \\ 0 & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix}.$$

Then the model is on the same form as in Expression (2.7). From the model described in Expression (7.1) and the observed impedances, synthetic seismic data, $\boldsymbol{d}$, can be created. The error variance is set to $\sigma_{e_\cdot}^2 = 5 \cdot 10^{-5}$ as used in Karimi et al. (2009). The observed impedance, $\boldsymbol{z}^o$ is displayed in Figure 7.3 and the synthetic seismic data is displayed in Figure 7.4.

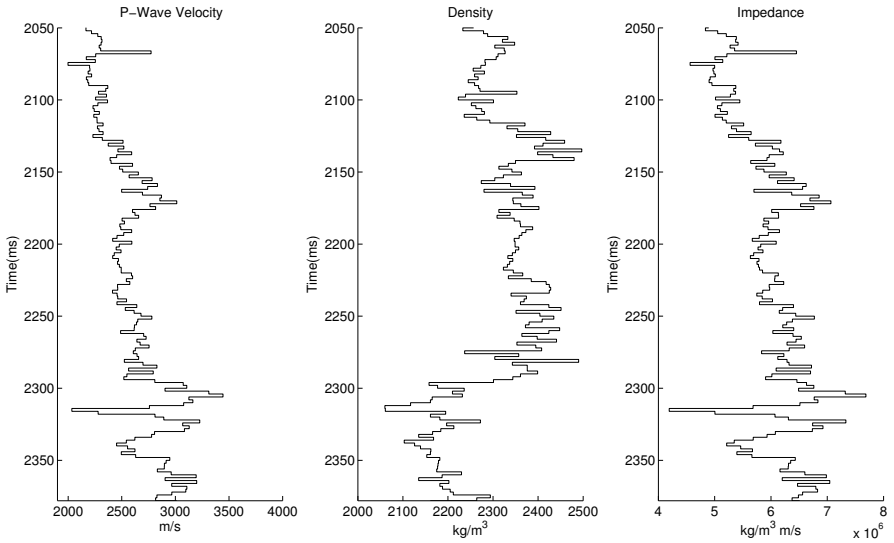


Figure 7.3: Material property observations from a well.

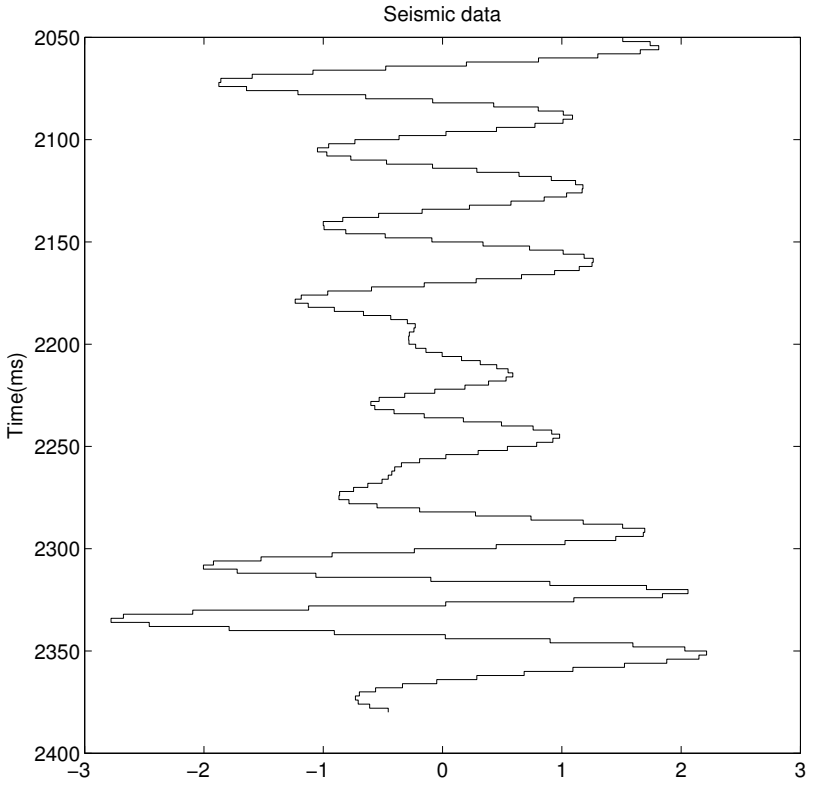


Figure 7.4: Synthetic seismic data.

In order to solve the inverse problem a prior model for $\boldsymbol{x}$ must be determined. It is natural to assume a stationary prior as described in Section 4.1. The used dependency measure is $c(h) = \exp\{-h/15\}$. The empirical correlation function and the used dependency measure are displayed in Figure 7.5. As described in

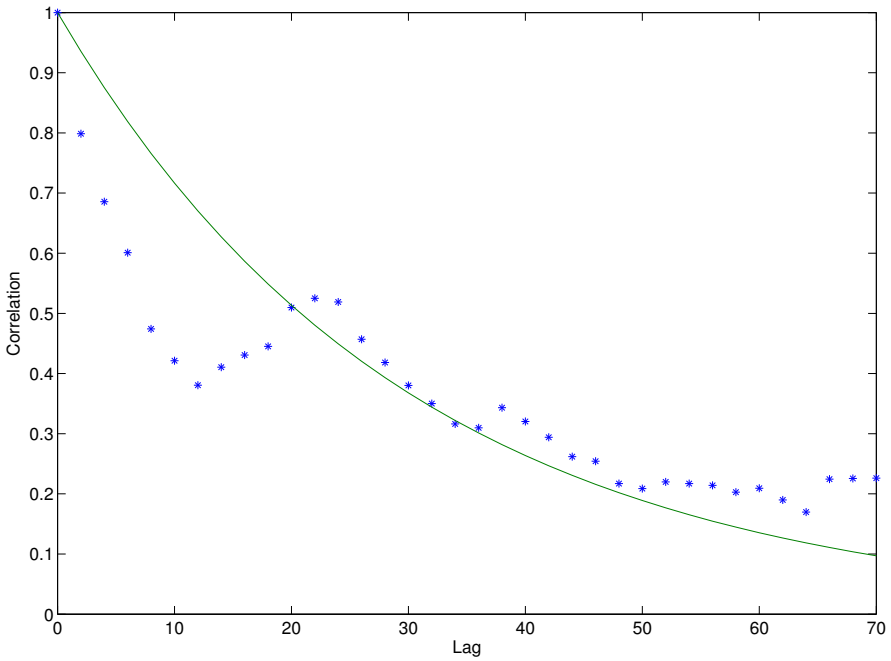


Figure 7.5: Fitted correlation function of the log-impedance $\ln(z)$.

Chapter 4 are there three different priors that can be used. For the x-space prior with the pseudo likelihood estimator is q_x^* set to 3. The results from the different priors are:

x-space prior:

$$\text{fL: } \hat{\mu}_{x_\cdot} = 15.85, \quad \hat{\sigma}^2_{x_\cdot} = 0.0351, \quad \hat{\gamma}_{x_\cdot} = -15.32$$
$$\text{pL}^*\text{: } \hat{\mu}_{x_\cdot} = 15.78, \quad \hat{\sigma}^2_{x_\cdot} = 0.0398, \quad \hat{\gamma}_{x_\cdot} = -32.73$$

tv-space prior:

$$\text{pL: } \hat{\mu}_{t_\cdot} = 15.61, \quad \hat{\sigma}^2_{t_\cdot} = 0.0110, \quad \hat{\gamma}_{tv_\cdot} = -0.019,$$

Gaussian prior:

$$\text{pL: } \hat{\mu}_{t_\cdot} = 15.59, \quad \hat{\sigma}^2_{t_\cdot} = 0.0109.$$

The different prior models are displayed together with the observations in Figure 7.6. Because of the high range parameter and the positive definite restrictions in

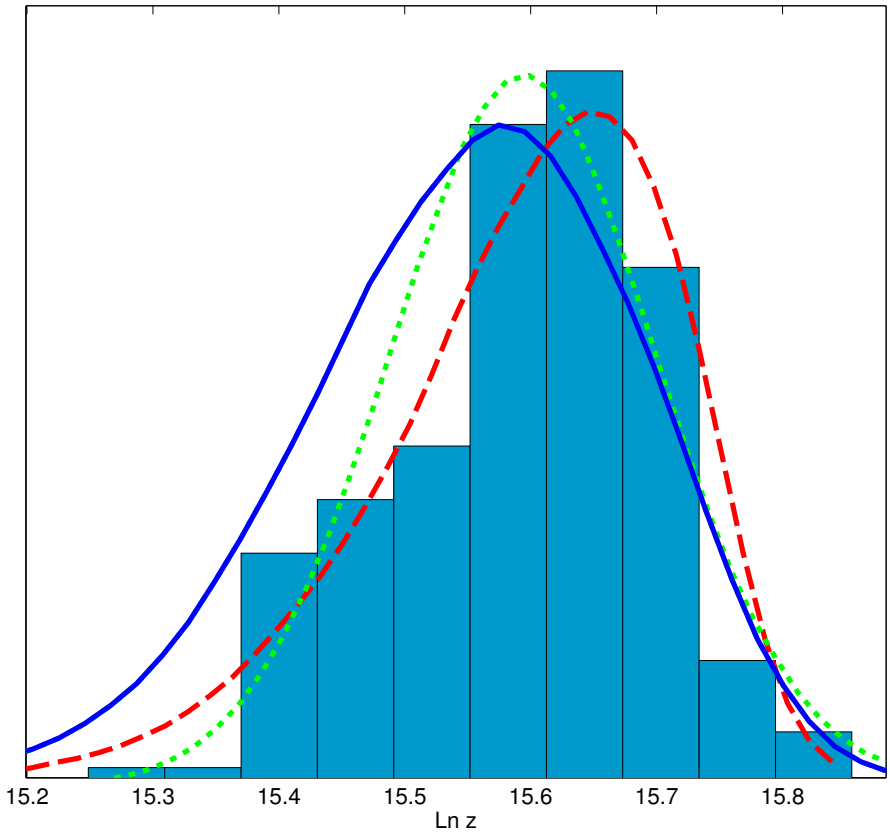


Figure 7.6: Estimated CS-Gaussian model and histogram of the log-impedance $\ln(\boldsymbol{z})$. Solid line (——) is with the fL and the x-space prior. Dashed line $(- - -)$ is the pL* with x-space prior. Doted line, $(\cdots)$, is pL with the tv-space prior.

Expression (5.5) is it not possible to capture the skewness in the data with this tv-space model. Hence is the parameters with the tv-space prior almost Gaussian.

7.2 Inversion Results

To predict the impedance, $\boldsymbol{z}$, from the seismic data the posterior distribution of $\boldsymbol{x}$ is found as described in Section 2.5. As displayed in Figure 7.7 the *Iterative Simulation* algorithm appears to converge after 15 iterations, because of the fact that the distribution is unimodal as discussed before. After the burn in period are 1000 realizations generated from the posterior model. Realization number 100, 400, 700 and 1000 are displayed in Figure 7.8 with parameters from the x-space prior and fL. The median of $\boldsymbol{z}$ is calculated from these 1000 realizations. Then the impedance is calculated as $[\widehat{\boldsymbol{z}|\boldsymbol{d}}] = \exp([\widehat{\boldsymbol{x}|\boldsymbol{d}}]) = \exp(\mathrm{Q}_{0.5}(\boldsymbol{x}|\boldsymbol{d}))$. In Karimi et al. (2009) is the prediction from the posterior done by an approximation, since they are not able to generate from the full posterior distribution. But with the *Iterative Simulation* algorithm presented in this thesis it is possible to generate realizations from the full posterior distribution. In Figure 7.9 the true impedance is compared with the predicted impedance for the x-space prior with fL, and it is observed that the predicted values fit the true data quite good. The generated realizations are also used to make 80%-prediction interval as displayed in Figure 7.10 for the x-space prior with the fL estimator. The results seems to fit the real values very good. In Figure 7.11 the results for the x-space prior with the pL^* estimator is displayed, hence is it observed that the result is heavy shifted.

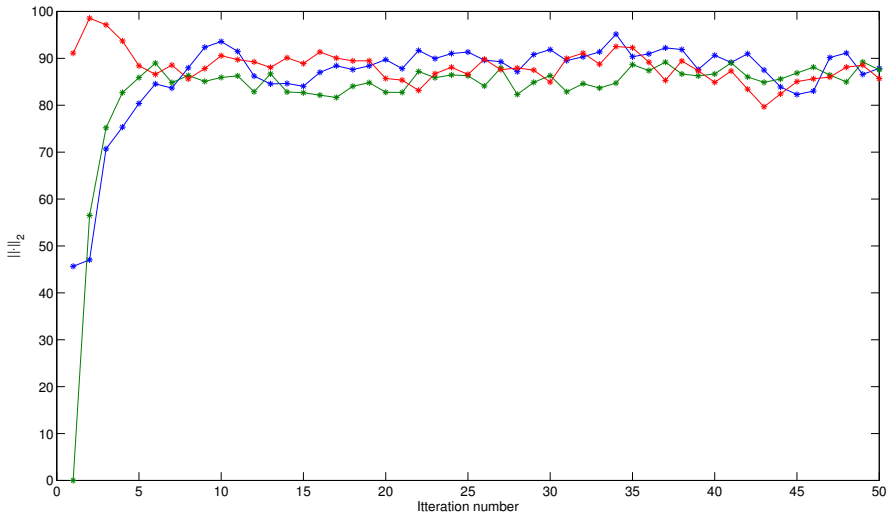
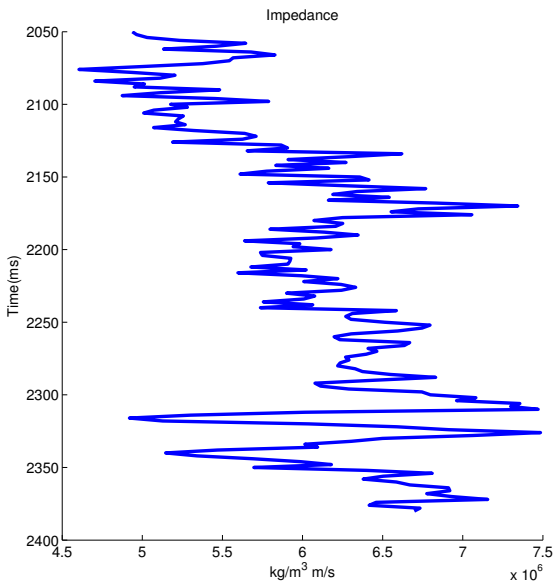
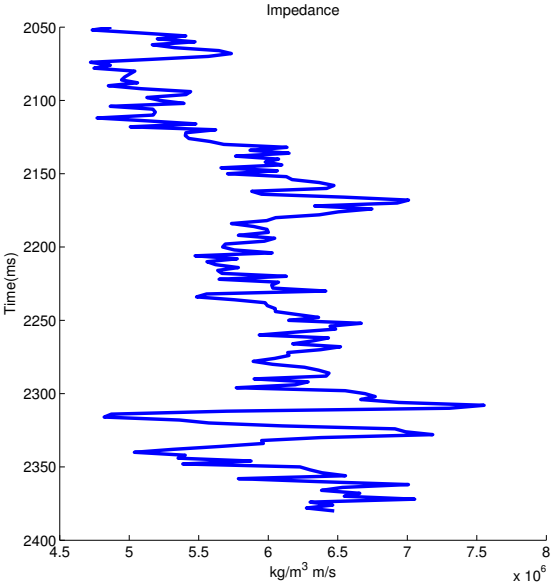


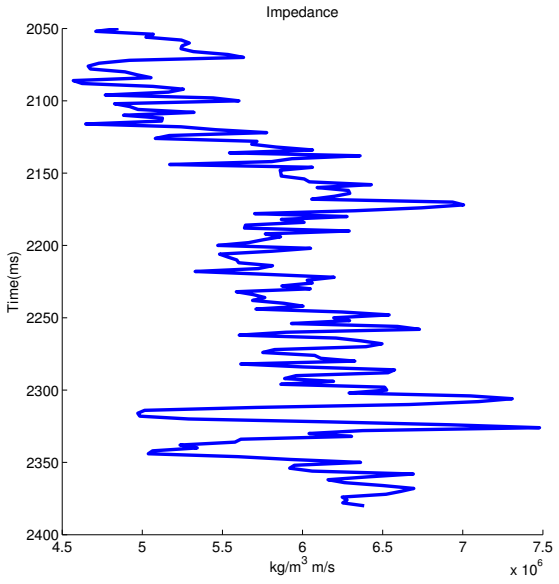
Figure 7.7: Traceplot for convergence of $\boldsymbol{v}$.



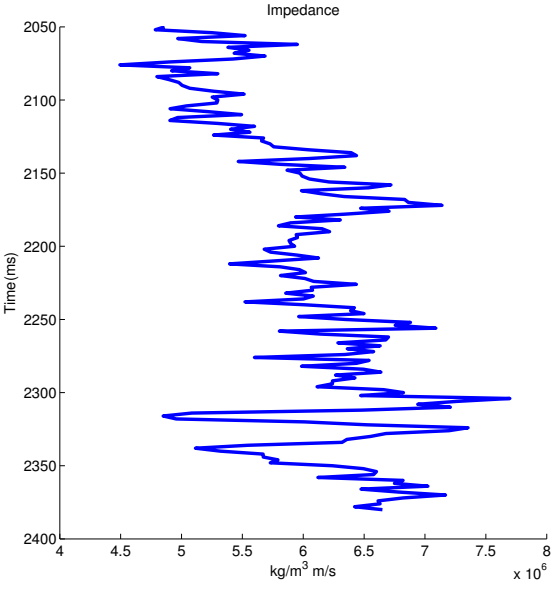
(a) Realization number 100.



(b) Realization number 400.



(c) Realization number 700.



(d) Realization number 1000.

Figure 7.8: Realizations from the posterior model. The x-space prior and the full Likelihood estimator is used.

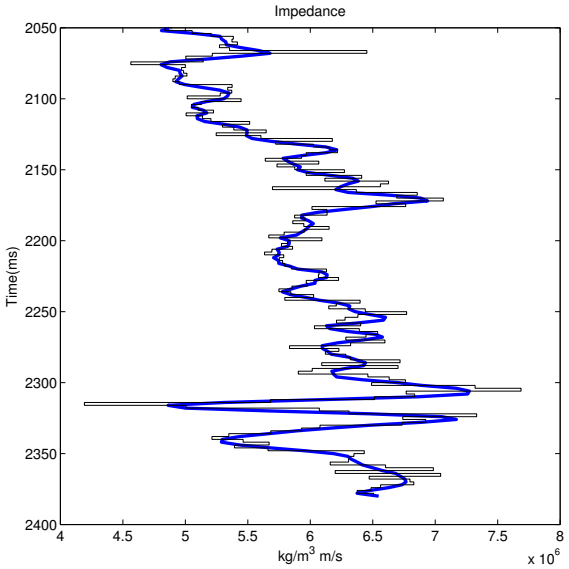


Figure 7.9: Inversion results from Bayesian CS-Gaussian inversion. The true impedance and the predicted impedance. The x-space prior and the full Likelihood estimator is used.

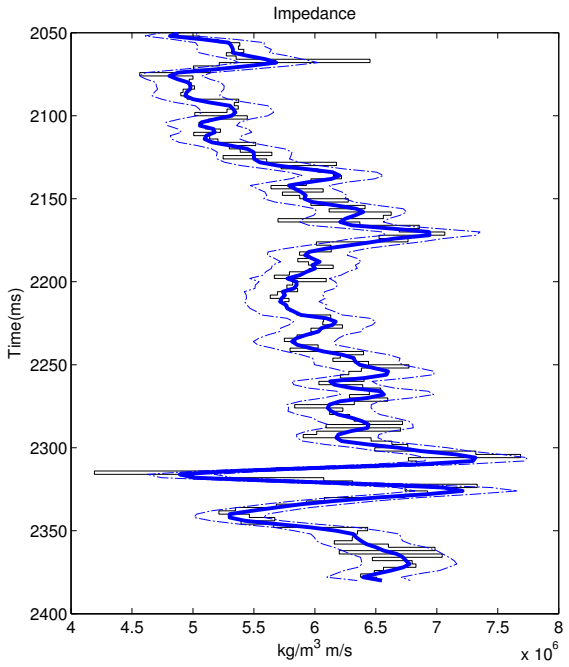


Figure 7.10: 80%-prediction intervals (thin dotted) and predictions (thick solid). The x-space prior and the full Likelihood estimator is used.

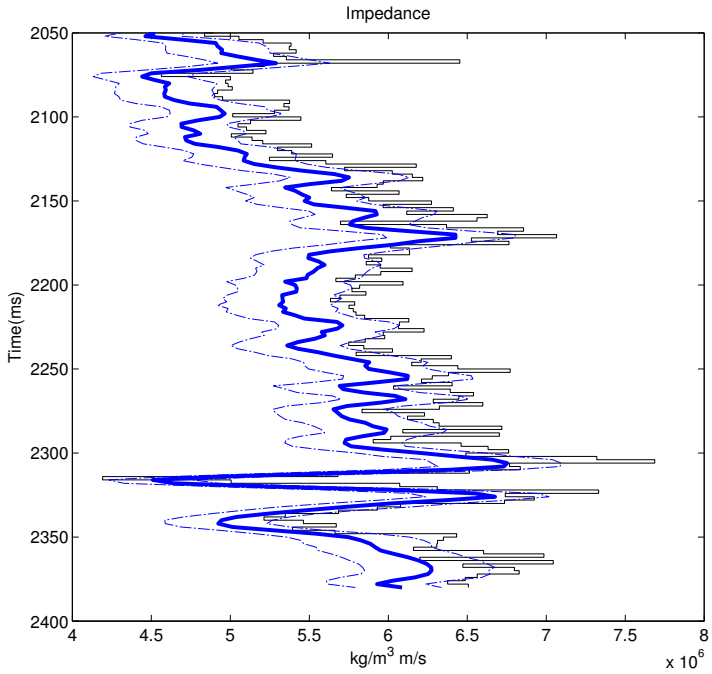


Figure 7.11: 80%-prediction intervals (thin dotted) and predictions (thick solid). The x-space prior and the pseudo Likelihood estimator is used.

In Table 7.1 is the fraction of correct values that lies below the corresponding quantiles listed. The x-space prior with fL estimator gives the best results, but the predicted interval is a bit too wide. The predicted impedance is not centered for the x-space prior with pL*. That can be a result of the approximation for the parameter estimation and a too low q_x^*. To use a higher q_x^* requires that a q_x^*-dimensional Gaussian cdf is evaluated for every step in the numerical optimization. The tv-space prior with pL estimator gives almost the same results as the Gaussian prior. The bias is a bit higher than for the x-space fL estimator, and the predicted interval is wider.

Theoretical:	0.1	0.5	0.9
Estimated, x-space prior fL:	0.048	0.482	0.956
Estimated, x-space prior pL*:	0.006	0.030	0.446
Estimated, tv-space prior pL:	0.036	0.619	0.960
Estimated, Gaussian prior pL:	0.036	0.590	0.970

Table 7.1: Fraction of observations below predicted quantiles.

The *Iterative Simulation* algorithm use only about 35 seconds to generate 1015 realizations and estimate the median, but the *Brute Force Simulation* algorithm and the *Two Steps Simulation* algorithm were not able to generate one single realization during three weeks.

Chapter 8
Closing Remarks and Further Work

In this thesis the properties of the CS-Gaussian distribution have been studied, with a special focus on Bayesian inversion. Bayesian inversion and parameter inference from data is more complicated than for the Gaussian distribution, therefore it is necessary to generate realizations from the CS-Gaussian distribution. An efficient simulation algorithm to generate realizations from a high dimensional problem has been obtained. In contrast to Karimi et al. (2009) is it here possible to generate realizations of the posterior model without any approximations. Different stationary prior models and different methods to do parameter estimation have been developed. For an exponential dependence structure is it possible to use a full-likelihood estimator, hence also capture the dependency in the data.

The simulation algorithm for the CS-Gaussian distribution were implemented and shown to work well on synthetic seismic data inspired by the Sleipner Øst field. Also the different prior models and estimators for the CS-Gaussian distribution are tested on the data set. The full-likelihood estimator seems to be the best estimator for data with exponential dependence structure.

A skew distribution with heavy tails, CST-distribution, is presented with some properties and examples. But it is still much to study about this distribution. A sampling algorithm should be developed, and it would also be useful to have different degrees of freedom for each dimension of the CST-distribution. The Bayesian CS-Gaussian inversion could also be extended to handle 3D problems.

Bibliography

Buland, A. and Omre, H. (2003). Bayesian linearized AVO inversion, *Geophysics* **68**: 185.

Dominguez-Molina, J., González-Farías, G., and Gupta, A. (2003). The multivariate closed skew normal distribution, *Technical report*, Technical Report 03.

González-Farías, G., Domínguez-Molina, J., and Gupta, A. (2004). The closed skew-normal distribution, *Skew-elliptical distributions and their applications: a journey beyond normality*, Chapman & Hall/CRC, Boca Raton, FL, pp. 25–42.

Karimi, O., Omre, H., and Mohammadzadeh, M. (2009). Bayesian closed-skew Gaussian inversion of seismic AVO data into elastic material properties.

Mosegaard, K. and Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res* **100**(12): 431–47.

Pewsey, A. (2000). Problems of inference for Azzalinis skewnormal distribution, *Journal of Applied Statistics* **27**(7): 859–870.

Robert, C. (1995). Simulation of truncated normal variables, *Statistics and computing* **5**(2): 121–125.

Roislien, J. and Omre, H. (2006). T-distributed Random Fields: A Parametric Model for Heavy-tailedWell-log Data1, *Mathematical Geology* **38**(7): 821–849.

Rybicki, G. B. and Press, W. H. (1995). Class of fast methods for processing irregularly sampled or otherwise inhomogeneous one-dimensional data, *Phys. Rev. Lett.* **74**(7): 1060–1063.

Sheriff, R., Geldart, L., and Lees, J. (1995). *Exploration seismology*, Cambridge University Press Cambridge.

Appendix A

Proof of Properties for the CST-distribution

A.1 Closed Under Linear Transformations

Let $\boldsymbol{x} = [\boldsymbol{t}|\boldsymbol{v} \geq \boldsymbol{0}] \sim \mathrm{CST}_{n,q}(\boldsymbol{\mu}, \Omega, \Gamma, \boldsymbol{\nu}, \Delta, \eta)$, and let A be a deterministic $l \times n$ matrix with $(l \leq n)$. If we do the transformation of $\boldsymbol{x}$ to the tv-space as given in Expression (6.1) we obtain

$$\begin{bmatrix} \boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{T}_{n+q}\left(\begin{bmatrix} \boldsymbol{\mu} \\ -\boldsymbol{\nu} \end{bmatrix}, \begin{bmatrix} \Omega & \Gamma\Omega \\ \Omega\Gamma & \Delta + \Gamma\Omega\Gamma' \end{bmatrix}, \eta \right)$$

Hence will the distribution of $(A\boldsymbol{t}, \boldsymbol{v})$ be as given in property 3 by Section 6.1:

$$\begin{bmatrix} A\boldsymbol{t} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{T}_{l+q}\left(\begin{bmatrix} A\boldsymbol{\mu} \\ -\boldsymbol{\nu} \end{bmatrix}, \begin{bmatrix} A\Omega A' & A\Gamma\Omega \\ \Omega\Gamma A' & \Delta + \Gamma\Omega\Gamma' \end{bmatrix}, \eta \right).$$

Then we can do the transformation back to the x-space and we obtain:

$$\boldsymbol{y} = A\boldsymbol{x} = [A\boldsymbol{t}|\boldsymbol{v} \geq \boldsymbol{0}] = \sim \mathrm{CST}_{l,q}(\boldsymbol{\mu}_y, \Omega_y, \Gamma_y, \boldsymbol{\nu}_y, \Delta_y, \eta_y),$$

where the parameters are given as

$$\begin{aligned} \boldsymbol{\mu}_y &= A\boldsymbol{\mu}, \\ \Omega_y &= A\Omega A', \\ \Gamma_y &= \Gamma\Omega A'\Omega_y^{-1}, \\ \boldsymbol{\nu}_y &= \boldsymbol{\nu}, \\ \Delta_y &= \Delta + \Gamma\Omega\Gamma' - \Gamma\Omega A'\Omega_y^{-1}A\Omega\Gamma' \\ \eta_y &= \eta. \end{aligned}$$

A.2 Closed Under Conditioning

Let $\boldsymbol{x} \sim \mathrm{CST}_{n,q}(\boldsymbol{\mu}, \Omega, \Gamma, \boldsymbol{\nu}, \Delta, \eta)$, with $\boldsymbol{x}_1 \in \mathbb{R}^k$ and $\boldsymbol{x}_2 \in \mathbb{R}^{n-k}$ being two subvectors of $\boldsymbol{x} = [\boldsymbol{x}_1', \boldsymbol{x}_2']'$. Further let the parameters associated with the subvectors be:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Omega} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \quad \Gamma = [\Gamma_1, \Gamma_2].$$

If we do the transformation of $\boldsymbol{x}$ to the tv-space as given in Expression (6.1) we obtain

$$\begin{bmatrix} \boldsymbol{t}_1 \\ \boldsymbol{t}_2 \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{T}_{n+q} \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ -\boldsymbol{\nu} \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{11}\Gamma_1' + \Omega_{12}\Gamma_2' \\ \Omega_{21} & \Omega_{22} & \Omega_{21}\Gamma_1' + \Omega_{22}\Gamma_2' \\ \Gamma_1\Omega_{11} + \Gamma_2'\Omega_{21} & \Gamma_1\Omega_{12} + \Gamma_2\Omega_{22} & \Delta + \Gamma\Omega\Gamma' \end{bmatrix}, \eta \right)$$

Hence will the distribution of $[(\boldsymbol{t}_1, \boldsymbol{v})|\boldsymbol{t}_2]$ be as given by property 4 in Section 6.1:

$$\begin{aligned} \boldsymbol{\mu}_{t_1 v|t_2} &= \begin{bmatrix} \boldsymbol{\mu}_{t_1|t_2} \\ \boldsymbol{\mu}_{v|t_2} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\mu}_1 \\ -\boldsymbol{\nu}_2 \end{bmatrix} + \begin{bmatrix} \Omega_{12} \\ \Gamma_2\Omega_{22} + \Gamma_1\Omega_{12} \end{bmatrix} \Omega_{22}^{-1}(\boldsymbol{t}_2 - \boldsymbol{\mu}_2) \\ &= \begin{bmatrix} \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{t}_2 - \boldsymbol{\mu}_2) \\ -\boldsymbol{\nu}_1 + (\Gamma_2 + \Gamma_1\Omega_{12}\Omega_{22}^{-1})(\boldsymbol{t}_2 - \boldsymbol{\mu}_2) \end{bmatrix} \end{aligned}$$

and the correlation matrix of $(\boldsymbol{t}_1, \boldsymbol{v})$ given $\boldsymbol{t}_2$ is

$$\begin{aligned} \Omega_{t_1 v|t_2} &= \begin{bmatrix} \Omega_{t_1|t_2} & \Gamma_{t_1 v|t_2} \\ \Gamma_{v t_1|t_2} & \Omega_{v|t_2} \end{bmatrix} \\ &= \kappa(\eta) \left(\begin{bmatrix} \Omega_{11} & \Omega_{11}\Gamma_1' + \Omega_{12}\Gamma_2' \\ \Gamma_1\Omega_{11} + \Gamma_2\Omega_{21} & \Delta + \Gamma\Omega\Gamma' \end{bmatrix} \right. \\ &\quad \left. - \begin{bmatrix} \Omega_{12} \\ \Gamma_2\Omega_{22} + \Gamma_1\Omega_{12} \end{bmatrix} \Omega_{22}^{-1} \begin{bmatrix} \Omega_{12} \\ \Gamma_2\Omega_{22} + \Gamma_1\Omega_{12} \end{bmatrix}' \right) \\ &= \kappa(\eta) \left(\begin{bmatrix} \Omega_{11} & \Omega_{11}\Gamma_1' + \Omega_{12}\Gamma_2' \\ \Gamma_1\Omega_{11} + \Gamma_2\Omega_{21} & \Delta + \Gamma\Omega\Gamma' \end{bmatrix} \right. \\ &\quad \left. - \begin{bmatrix} \Omega_{12}\Omega_{22}^{-1}\Omega_{21} & \Omega_{12}\Gamma_2' + \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\Gamma_1' \\ \Gamma_2\Omega_{21} + \Gamma_1\Omega_{12}\Omega_{22}^{-1}\Omega_{21} & \Gamma_2\Omega_{22}\Gamma_2' + \Gamma_2\Omega_{21}\Gamma_1' + \Gamma_1\Omega_{12}\Gamma_2' + \Gamma_1\Omega_{12}\Omega_{22}^{-1}\Omega_{21}\Gamma_1' \end{bmatrix} \right) \end{aligned}$$

Then we can do the transformation back to the x-space and we obtain

$$[\boldsymbol{x}_1|\boldsymbol{x}_2] \sim \mathrm{CST}_{k,q}(\boldsymbol{\mu}_{1|2}, \Omega_{1|2}, \Gamma_{1|2}, \boldsymbol{\nu}_{1|2}, \Delta_{1|2}, \eta_{1|2}).$$

Where

$$
\begin{aligned}
\boldsymbol{\mu}_{1|2} =& \boldsymbol{\mu}_{t_1|t_2} = \boldsymbol{\mu}_1 + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),\\
\Omega_{1|2} =& \Omega_{t_1|t_2} = \kappa(\eta)(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}),\\
\Gamma_{1|2} =& \Gamma_{vt_1|t_2}\Omega_{t_1|t_2}^{-1} = \frac{\kappa(\eta)}{\kappa(\eta)}(\Gamma_1\Omega_{11} + \Gamma_2\Omega_{21} - \Gamma_2\Omega_{21} - \Gamma_1\Omega_{12}\Omega_{22}^{-1}\Omega_{21})(\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21})^{-1}\\
=& \Gamma_1\\
\boldsymbol{\nu}_{1|2} =& -\boldsymbol{\mu}_{v|t_2} = \boldsymbol{\nu} - \left(\Gamma_2 + \Gamma_1\Omega_{12}\Omega_{22}^{-1}\right)(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),\\
\Delta_{1|2} =& \Omega_{v|t_2} - \Gamma_{vt_1|t_2}\Omega_{t_1|t_2}\Gamma_{t_1v|t_2}\\
=& \kappa(\eta)\left(\Delta + \Gamma_1\Omega_{11}\Gamma_1' + \Gamma_1\Omega_{12}\Gamma_2' + \Gamma_2\Omega_{21}\Gamma_1' + \Gamma_2\Omega_{22}\Gamma_2'\right.\\
&- \Gamma_2\Omega_{22}\Gamma_2' - \Gamma_2\Omega_{21}\Gamma_1' - \Gamma_1\Omega_{12}\Gamma_2' - \Gamma_1\Omega_{12}\Omega_{22}^{-1}\Omega_{21}\Gamma_1'\\
&\left.-\Gamma_1(\Omega_{11}\Gamma_1' - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}\Gamma_1')\right)\\
=& \kappa(\eta)\Delta\\
\eta_{1|2} =& \eta.
\end{aligned}
$$