UNIVERSITEIT GENT

FACULTEIT WETENSCHAPPEN

VAKGROEP TOEGEPASTE WISKUNDE, INFORMATICA EN STATISTIEK

Skew-symmetric distributions and associated inferential problems

Elissa Burghgraeve

Promotor : Prof. Christophe LEY

Masterproef ingediend tot het behalen van de academische graad van master in de wiskunde

Academiejaar 2016-2017

UNIVERSITEIT GENT

FACULTEIT WETENSCHAPPEN

VAKGROEP TOEGEPASTE WISKUNDE, INFORMATICA EN STATISTIEK

Skew-symmetric distributions and associated inferential problems

Elissa Burghgraeve

Promotor : Prof. Christophe LEY

Masterproef ingediend tot het behalen van de academische graad van master in de wiskunde

Academiejaar 2016-2017

Preface

Ever since childhood, I've had a special interest in logical reasoning and analyzing. As I got older, mathematics was what I loved most at school and therefore it wasn't a very hard choice to pursue studying mathematics. It certainly wasn't always easy, but it gave me so much gratification to acquire new insights and to gain a deeper understanding of mathematics. When the bachelor came to a close, it became clear to me that, although I found the pure mathematical subjects interesting, applied mathematics was much better for me. The course Statistical Inference' by prof. Christophe Ley was one of the subjects that really appealed to me. By working on a project for this subject, this interest was further enhanced. This was mainly due to the combination of statistics with techniques from algebra and analysis. So when Prof. Ley proposed to write a thesis following my project, I did not have to think long.

So this is really the last step of my education and that would not have been possible without a number of people.

First of all I would like to thank my promotor, Prof. Christophe Ley, for offering me this topic and the extremely good guidance. I would like to thank him for helping me when I was stuck or when I did not understand something, for every time he reviewed my thesis with me and helped me improve my thesis. Without Prof. Ley, I absolutely would not have been able to complete this thesis.
I would like to thank my parents, Anne and Guido, for their support over the years. There were sometimes setbacks, but they always kept believing in me and helped me reach my final goal.
I also want to thank my sister, Lara, for the positive vibes and for proofreading this thesis. Her English expertise has certainly come in handy.
Finally, I would like to thank my group of friends for countless days in the library, supporting and motivating each other to continue working and to finish this thesis.

Toelating tot bruikleen

Elissa Burghgraeve,
mei 2017

Abstract

Data sets in many practical applications are not symmetric or normal, even though we would like them to be. So the data can not be fitted using the popular normal distribution. In the 20^{th} century a new family of distributions was developed to handle this skewness, the skew-symmetric distributions.

In this thesis, we will explore the skew-symmetric distributions and we will look more closely at the inferential problems they may have. To do this I mainly made use of a few important articles concerning skew-symmetric distributions. I have analyzed these articles and brought together the different ideas explained in them. I have worked out in detail the results given in the articles.

In the first chapter, we give a historical overview on the development of skewed distributions. First attempts were made by modifying the skewed data to fit the normal curve. Mathematicians like Edgeworth (1899) [27] elaborated this method. One of the first to define a new family of distributions was Pearson (1895) [54] with his four-parameter system of continuous distributions. His method to obtain this is given in more detail in this thesis. A very innovative proposal to construct non-normal distributions was given by de Helguero (1909) [23, 24]. We also take a closer look at the construction of his skewed distributions. More recently, the widely known skew-normal distributions were popularized by Azzalini (1985) [7], this family of distributions extends the normal one. Its probability density function (pdf) is given by

$$\phi(z;\delta) = 2\phi(z)\Phi(\delta z), \qquad -\infty < z < \infty,$$

where ϕ is the standard Gaussian pdf and Φ the standard Gaussian cumulative distribution function. To finish this chapter we also give some applications of the skew-symmetric distributions. These are applications from many different fields and they show how widespread the use of skew-symmetric distributions is.

In the second chapter, we will look at the skew-symmetric distributions from a more theoretical perspective. More specifically, we will investigate the skew-normal and skew-t distributions. The pdf of the skew-normal distributions is given above. The pdf of the skew-t distributions can be expressed as follows:

$$t(z;\delta,\nu) = 2t(z;\nu)T\left(\delta z\sqrt{\frac{\nu+1}{\nu+z^2}};\nu+1\right), \qquad -\infty < z < +\infty,$$

where t and T denote the standard Student-t density function and distribution function, respectively, and ν stands for the degrees of freedom. In both cases we start by giving some properties with proof. For the skew-normal family we continue by giving the moment generating function and computing the moments. Lastly for the skew-normal distributions we give the extended skew-normal distribution. For the skew-t family we calculate the moments by stating that we can write a skew-t random variable as a ratio

$$Y = \frac{Z}{\sqrt{\frac{U}{\nu}}}$$

with Z a standard skew-normal variate and U follows the chi-squared distribution with ν degrees of freedom, Z and U are independent.

In the third and final chapter, we introduce the associated inferential problems of the skew-symmetric distributions. This is again applied to the two examples used in the second chapter, the skew-normal and the skew-t distributions. In both examples the score function and the Fisher information matrix are calculated. In case of the skew-normal distributions the Fisher information matrix is singular in the vicinity of symmetry which will lead to slower convergence rates of the estimated skewness parameter, it will in fact drop to a $\sqrt[6]{n}$-rate. To prove this fact, Lemma 3 from Rotnitzky *et al.* (2000) [59] and a Proposition proved by Chiogna (2005) [21] are given. After establishing the problem, two reparametrizations to overcome the problem of singularity of the Fisher information matrix are presented and analyzed. The first is the centred parametrization, first proposed by Azzalini (1985) [7]. The second uses orthogonalization, proposed by Hallin and Ley (2014) [39] which uses the Gram-Schmidt orthogonalization process. The orthogonalization process needs to be applied twice because of a so called double singularity problem of the skew-normal distributions. With both reparametrizations, a new set of parameters is obtained and the Fisher information matrix is calculated with respect to these parameters. In both cases the Fisher information matrix will no longer be singular. For the skew-t family, the Fisher information matrix is not singular and thus there is no singularity problem here unless the degrees of freedom ν go to infinity. But then the skew-t distribution tends to the skew-normal one, for which we already know the solution.

Contents

Preface i

Abstract iv

1 An introduction to the skew-symmetric distributions **1**

1.1 Some history of skewed distributions 1

1.1.1 Early attempts 1

1.1.2 Later developments 13

1.2 Applications 17

2 Skew-symmetric family **21**

2.1 Skew-normal family 22

2.1.1 Properties 23

2.1.2 Moment generating function and moments 25

2.1.3 Extended skew-normal distribution 28

2.2 Skew-t family 30

2.2.1 Properties 31

2.2.2 Moments 32

3 Singularity problem of skew-symmetric distributions **35**

3.1 Skew-normal family 35

3.1.1 Centred parametrizaton 44

3.1.2 Orthogonalization 47

3.2 Skew-t family 54

3.3 Conclusion 60

Appendix A Nederlandstalige samenvatting **61**

Appendix B **63**

Chapter 1

An introduction to the skew-symmetric distributions

Symmetry is a concept that is present in our everyday lives. It is something we try to seek naturally in everything. Symmetry is therefore in many ways seen as a beauty ideal. But not everything in the world is symmetric, in fact most things are not. So the idea of finding symmetry in all things is very unrealistic. The same is true in statistics. Some kind of symmetry is supposed in most classical procedures. However, most datasets are not symmetric (or normal). More so, asymmetry or absence of symmetry are much more common in data then symmetry is. So we either need to test whether or not the data is symmetric or we need procedures that do not need for the data to be symmetric. So there is a necessity for skewed distributions for a few different reasons :

- There will be a better fit to the data.
- They give an alternative for tests in symmetry.
- These distributions form the foundation of new, more general procedures.

1.1 Some history of skewed distributions

1.1.1 Early attempts

During the 19^{th} century statistical methods became more widely used than only in the natural sciences. The normal distribution, developed for describing the variation of errors of measurement was utilized to describe the variation of different characteristics of individuals. However, people came across asymmetric data which instigated the need for non-normal distributions. Then of course, it was natural to adapt the normal distribution.
The first proposals of non-symmetric and non-normal distributions were made in the late 19^{th} century as stated in the article by Ley (2014) [47].

Francis Ysidro Edgeworth

One of the earliest attempts was proposed by Francis Ysidro Edgeworth (1845-1926), an Irish polymath. In the 1880's he was involved in trying to fit non-normal data. In one of his publications he described how distributions such as those of bank reserves and price changes could be examined to see if they satisfied the assumption of normality. He suggested first testing symmetry and then determining whether or not the normal distribution was the best fit of symmetric curves, which were limited in amount. In 1886 [**?**] he tried to find asymmetric distributions to fit asymmetric frequency data and he is usually considered as the first to do so. Over time Edgeworth tried different approaches to model skew data. According to Wallis (2014) [64], the first one was the 'method of translation' which consists of fitting a normal curve to transformed data. Another method was called the 'method of separation' or mixture of normals. These methods were suggested in the first two parts of his five-part article 'On the representation of statistics by mathematical formulae'. In the third part Edgeworth considers the 'method of composition', in which he fitted two half-normal curves to the left and right sides of the distribution to construct a 'composite probability-curve'. The figure below shows the accompanying figure Edgeworth gave in his paper [27] on the method of composition.

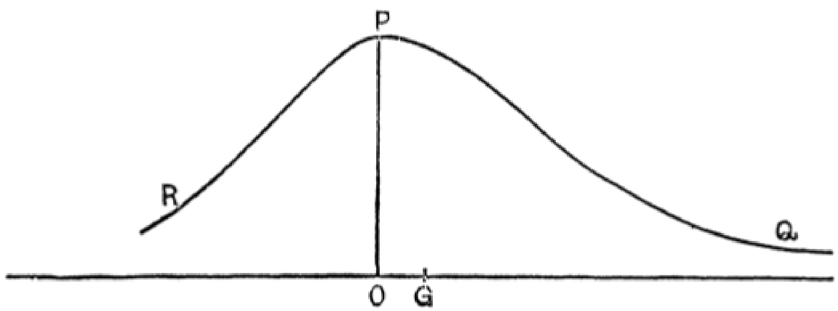

Figure 1.1.1: Edgeworth's composite probability-curve

When a sample mean and second and third sample moments are given, Edgeworth estimates the parameters using their definitions to get a cubic equation in function of the distance between the mean and the mode. Solving this equation gives the required parameters estimates.

Karl Pearson

A few years after the publication of Edgeworth, Karl Pearson (1857-1936) started investigating the fitting of asymmetric data. His interest was sparked by the work of a zoologist Walter Weldon on Plymouth shore crabs. Weldon had found that one of his distributions of data was not symmetrical while all the others were. Even more so all the others showed normal-like behavior. Pearson wanted to find an alternative way to interpret the data instead of trying to normalize it since it did not produce a normal curve, as we can read in Hald (2004) [37]. He wanted to understand the shape of the distribution without having to deform the original shape. Pearson had to construct a new statistical system to interpret Weldon's data since such systems did not exist at the time. He did this by adjusting mathematics of mechanics and using the method of moments. In one of his first attempts he dissected an asymmetric frequency curve into two normal curves. So this resulted in a mixture of two normal distributions. However, he found the model to be too limited and felt it was necessary to find continuous distributions to describe Waldon's data. His breakthrough came with his definition of a generalized form of the normal curve of an asymmetric character. This result started some kind of feud between Edgeworth and Pearson. For instance, Edgeworth stated that the curved line defined by Pearson had already been derived by Erastus de Forest, which Pearson did not deny. Pearson's next attempt was even more innovative. In 1895 he defined several probability distributions in his article [54] as the foundation of Pearson's four-parameter system of continuous distributions, a family of distributions that was studied exhaustively and is still used today. We will now take a brief look at Pearson's derivation of his system of distributions as obtained in his article [54], making use of the elaborations in Hald (2004) [37].

Pearson's system of continuous distributions Pearson defines the moments as

$$\mu'_r = \mathbb{E}(x^r), \qquad r = 0, 1, \ldots \tag{1.1.1}$$

$$\mu_r = \mathbb{E}\left((x-\mu'_r)^r\right), \qquad r = 2, 3, \ldots \tag{1.1.2}$$

and

$$\beta_1 = \frac{\mu_3^2}{\mu_2^2}, \qquad \beta_2 = \frac{\mu_4}{\mu_2^2}, \qquad \beta_3 = \frac{\mu_3\mu_5}{\mu_2^4}.$$

He derives the normal distribution by stating that a polygon formed by plotting the terms of the point-binomial

$$\left(\frac{1}{2}+\frac{1}{2}\right)^n = \sum_{x=0}^{n}\binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x}$$

at distance c from each other coincides very closely with the contour of a normal frequency curve when n is only moderately large, defining the symmetric binomial as

$$p(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n.$$

The height of a random term is given by $p(x)$ and the relative slope by

$$\begin{aligned}
\frac{\text{slope}}{\text{mean ordinate}} &= \frac{p(x+1)-p(x)}{\frac{c}{2}\left(p(x+1)+p(x)\right)} \\
&= \frac{\binom{n}{x+1}\left(\frac{1}{2}\right)^n - \binom{n}{x}\left(\frac{1}{2}\right)^n}{\frac{c}{2}\left(\binom{n}{x+1}\left(\frac{1}{2}\right)^n + \binom{n}{x}\left(\frac{1}{2}\right)^n\right)} \\
&= \frac{n-2x-1}{\frac{c}{2}(n+1)} \\
&= \frac{cn-c(x+x+1)}{c^2(n+1)/2}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{2c(x'+x'+1)/2}{2c^2(n+1)/4} \\
&= -\frac{2 \times \text{mean abcissa}}{2\sigma^2}
\end{aligned}
$$

with $x' = x - \frac{n}{2}$ and $\sigma^2 = c^2(n+1)/4$. We can see that we have found the same expression as for the slope of the normal curve of frequency $y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$. So we can say that this binomial polygon and the normal curve are very similar. Pearson concludes this by differentiation. We thus have

$$
\begin{aligned}
\frac{\text{slope}}{\text{ordinate}} &= -\frac{2 \times \text{abcissa}}{2\sigma^2} \\
\Longleftrightarrow \frac{p(x+1)-p(x)}{cp(x)} &= -\frac{2cx'}{2\sigma^2} \\
\Longleftrightarrow \frac{p'(x)}{p(x)} &= -\frac{2c^2(x-\frac{n}{2})}{2\sigma^2}
\end{aligned}
$$

So we have found that the corresponding continuous distribution satisfies

$$
\frac{d \ln p(x)}{dx} = -\frac{x - \frac{n}{2}}{(n+1)/4}.
$$

Solving this differential equation we get

$$
\begin{aligned}
\int d \ln p(x) &= -\frac{4}{n+1} \int \left(x - \frac{n}{2}\right) dx \\
\Longleftrightarrow \ln p(x) &= -\frac{4}{2(n+1)} \left(x - \frac{n}{2}\right)^2 \\
\Longleftrightarrow p(x) &= \exp\left(-\frac{\left(x-\frac{n}{2}\right)^2}{(n+1)/2}\right).
\end{aligned}
$$

The solution is the normal distribution with mean $n/2$ and variance $(n+1)/2$.
Analogous, Pearson then analyzes the skew binomial and the hypergeometric distribution (point-binomial $(p+q)^n$). For the hypergeometric distribution he finds

$$
p(x) = \binom{n}{x} \frac{(Np)^x (Nq)^{n-x}}{N^n}
$$

which gives

$$
S = -\frac{y}{\beta_1 + \beta_2 x + \beta_3 x^2}
$$

with $y = x + \frac{1}{2} - \mu$ and μ, β_1, β_2 and β_3 constants depending on the parameters of S.

Hence the corresponding continuous density satisfies

$$\frac{d\ln p(x)}{dx} = -\frac{x-\alpha}{\beta_1 + \beta_2 x + \beta_3 x^2}.$$

The solution depends on the sign of the discriminant of the denominator $\beta_2^2 - 4\beta_1\beta_3$. We will derive an expression for the solution of this differential equation. Writing p for $p(x)$, Pearson's system is based on the differential equation

$$\frac{d\ln p(x)}{dx} = \frac{x+a}{b_0 + b_1 x + b_2 x^2}. \tag{1.1.3}$$

It follows that

$$x^r(b_0 + b_1 x + b_2 x^2)p' = x^r(x+a)p.$$

Integrating this equation and using partial integration we get, with μ'_r as in (1.1.2),

$$-rb_0\mu'_{r-1} - (r+1)b_1\mu'_r - (r+2)b_2\mu'_{r+1} = \mu'_{r+1} + a\mu'_r, \qquad r = 0, 1, \ldots$$

assuming that $x^r(b_0 + b_1 x + b_2 x^2)p$ is zero at the endpoints of the support of p. For successive positive integer values of r from zero to 3, we get, as in Lloyd (1983) [49], 4 equations from which we can calculate the constants:

$$a\mu'_0 + b_1\mu'_0 + 2b_1\mu'_1 = -\mu'_1,$$

$$a\mu'_1 + b_1\mu'_0 + 2b_1\mu'_1 + 3b_1\mu'_2 = -\mu'_2,$$

$$a\mu'_2 + 2b_1\mu'_1 + 3b_1\mu'_2 + 4b_1\mu'_3 = -\mu'_3,$$

$$a\mu'_3 + 3b_1\mu'_2 + 4b_1\mu'_3 + 5b_1\mu'_4 = -\mu'_4.$$

Hence there is a one-to-one correspondence between a, b_0, b_2 and b_2 and the first four moments, so p is uniquely determined by the first four moments. Equation (1.1.3) then becomes

$$\frac{d\ln p(x)}{dx} = \frac{x + \frac{M_1}{M_2}}{(M_3 + M_1 x + M_4 x^2)/M_2}$$

with

$$M_1 = \sqrt{\mu'_2\beta_1}(\beta_2 + 3),$$

$$M_2 = 2(5\beta_2 - 6\beta_1 - 9),$$

$$M_3 = \mu'_2(4\beta_2 - 3\beta_1),$$

$$M_4 = 2\beta_2 - 3\beta_1 - 6.$$

The solution depends on the roots of the equation

$$(M_3 + M_1 x + M_4 x^2)/M_2 = 0$$

i.e. on $\frac{M_1^2}{4M_3M_4}$, which expressed in the terms of the moments gives the criterion

$$\kappa = \frac{\beta_1(\beta_2+3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}.$$

Pearson distiguishes different types of distributions depending on the value of κ. This results in the following table.

Table 1.1.1: Table of Pearson's Type I to VII distributions.

Type	Equation	Origin for x	Limits for x	criterion
I	$y = y_0\left(1+\frac{x}{a_1}\right)^m\left(1-\frac{x}{a_2}\right)^m$	Mode	$-a_1 \leq x \leq a_2$	$\kappa<0$
II	$y = y_0\left(1-\frac{x^2}{a^2}\right)^m$	Mean (= mode)	$-a \leq x \leq a$	$\kappa = 0$
III	$y = y_0 e^{-\gamma x}\left(1+\frac{x}{a}\right)^{\gamma a}$	Mode	$-a \leq x < \infty$	$\kappa = \infty$
IV	$y = y_0 e^{-\nu \tan^{-1} x/a}\left(1+\frac{x^2}{a^2}\right)^{-m}$	Mean $+\frac{\nu a}{r}$, $r = 2m-2$	$-\infty < x < -\infty$	$0 < \kappa < 1$
V	$y = y_0 e^{-\gamma/x} x^{-p}$	At start of curve	$0 \leq x < \infty$	$\kappa = 1$
VI	$y = y_0(x-a)^{q_2} x^{-q_1}$	At or before start of curve	$a \leq x < \infty$	$\kappa > 1$
VII	$y = y_0\left(1+\frac{x^2}{a^2}\right)^{-m}$	Mean (= mode)	$-\infty < x < \infty$	$\kappa = 0$

At the end of his paper [54], Pearson gives a lot of examples by fitting his distributions to a variety of data coming from different fields of research. So he did not only give a new set of distributions theoretically, but he also showed that they were able to actually fit data in practice.
We can see that when $\nu = 0$, Pearson IV becomes the Student-t distribution. So Pearson IV is an asymmetric version of the Student-t distribution. Figure 1.1.2 below shows a plot of the Pearson IV probability density function (pdf) while Figure 1.1.3 compares the pdf of Pearson IV when $\nu = 0$ with the pdf of the Sudent-t distribution.
Although Pearson was one of the first to derive this general form of the Student-t distribution, it was named after William Sealy Gosset who worked under the pseudonym 'Student'. Student refers to the distribution as the frequency distribution of standard deviations of samples drawn from a normal population in his 1908 paper [61].
Edgeworth's reaction on Pearson's newly derived distributions was a paper on his 'Method of translation', a concept to transform data to make the resulting transformed data follow the normal distribution, as we discussed in the previous section. This technique to deal with asymmetric or non-normal data was

already used before but was formally developed by Edgeworth by taking a suitable selected function of the observations as normally distributed. He had the support of Kapteyn, a statisticien who also generalized the idea of transforming the data. So besides the rivalry between Pearson and Edgeworth, a discussion started between Pearson and Kapteyn each claiming their own family of skew curves was better then the other.

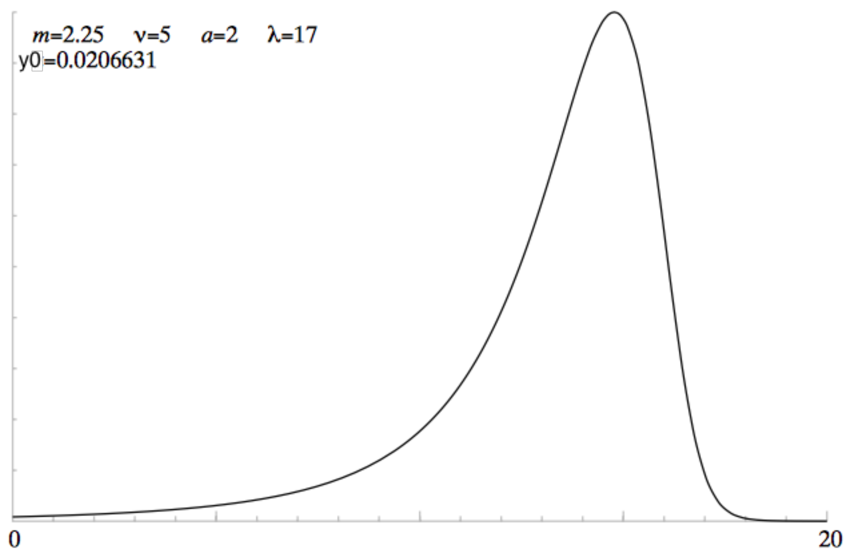

Figure 1.1.2: Pearson IV with $m = 2.25$, $\nu = 5$ and $a = 2$.

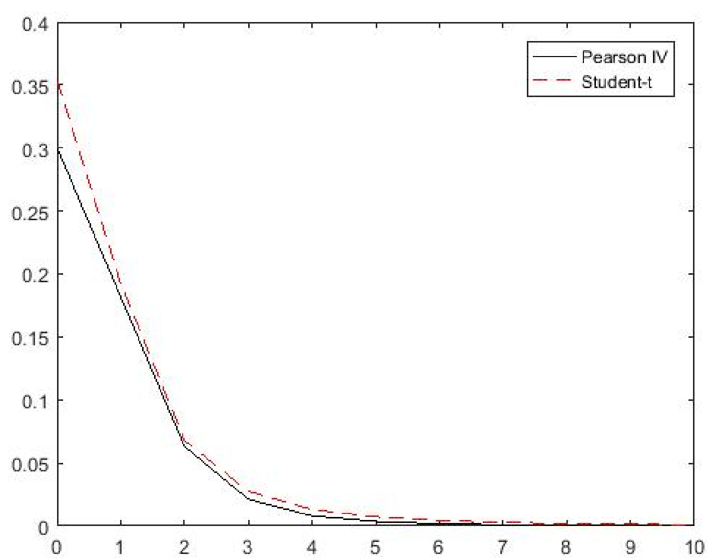

Figure 1.1.3: Pdf of Pearson IV with $m = 2.25$, $y_0 = 0.3$ and $a = 2$ compared to the pdf of the Student-t distribution with $\nu = 2$.

Carl Gustav Fechner

Around the same time, in 1897, a book came out by Carl Gustav Fechner (1801-1887) [30]. His manuscript was completed and published by Gottlob Friedrich Lipps, which explaines the publication after his death. In his book Fechner introduced a skew curve by binding together two halves of normal curves, each having the same mode but different standard deviations. With this he had thus laid the foundation of a model for non-symmetric distributions that is still used today, namely the two-piece distributions.
However, Fechner's idea was heavely disputed by Pearson on both historical and statistical grounds because he saw it as a rival to his own family of curves, as we can see in Ley (2014) [47]. Pearson claimed that Fechner's work was not original, but that a same proposal was already made by De Vries in 1894. From a statistical viewpoint, he argued that Fechner's curves were not general enough, in contrast to his own. Due to this strong opposition by Pearson, Fechner's work dissapeared from statistical literature until it reappeared in Hald's history in 1998 [37]. Meanwhile it was re-discovered on a few different occasions. An early rediscovery was given by Edgeworth [29]. He considers the 'Method of composition', a method in which he constructs 'a composite probability-curve'. This curve consists of two half-probability curves of different types, put together at the mode to get a continuous curve. The second one appeared much later in the physics literature by Gibbons and Mylroie in 1973 [35] under the name 'joined half-Gaussian' distribution. This distribution is fitted by the method of moments. Third was the 'three-parameter two-piece normal' distribution of John in 1982 [42]. This was published in the statistical literature. John compared estimation by the method of moments and maximum likelihood. The fourth rediscovery in the meteorology literature was by Toth and Szentimrey in 1990 [63]. They presented the 'binormal' distribution, which was again fitted by the maximum likelihood. Very recently, in 2016, it has reappeared again. This time in the financial literature in an article 'A Simple Skewed Distribution with Asset Pricing Applications' by Frans de Roon and Paul Karehnke [25].

Fernando de Helguero

In the beginning of the 20^{th} century an innovative way to construct non-normal distributions was given by a young Italian statisticien, Fernando de Helguero (1880-1908). He did this from a entirely different point of view on what he called abnormal curves. He wanted to present an alternative to Pearson's family of curves which at the time was predominant. In two papers [23, 24], both published posthumously in 1909, he presents his own method to handle non-normal data. In his work he also criticized Edgeworth's and Pearson's work by remarking that their proposals are only mathematical constructions and do not show us which mechanism might have generated the data, even though they are better than the normal distribution because they are generalizations. His own idea consists in giving a formulation for modelling non-normal frequency distributions by perturbating the normal density via a uniform distribution function. He does this because he assumes that the normal distribution naturally arises but that some external action might have caused a perturbation leading to the observed asymmetry.
Unfortunately de Helguero died very young, at the age of 28 by an earthquake. We can only guess what important developments he could have made, had he survived. A recent article on his work was written by Azzalini and Regoli [15] where they take a look at the original work of de Helguero and modify some of it. We will now give the elaboration of both, following de Helguero [24] and Azzalini and Regoli [15].

Mathematical development de Helguero derived the equation of his abnormal curve as follows : he starts by giving an equation of what he calls the hypothetical normal variation

$$\frac{c}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2} \tag{1.1.4}$$

which would not have the external perturbation cause.
The probability that an individual in class x is affected by the perturbation cause is a function of x, say $\theta(x)$. In class x there will be $y\theta(x)$ individuals impacted with y the number of individuals in class x. Consiquently, $y - y\theta(x) = y(1-\theta(x))$ individuals will remain in class x. For the curve with equation (1.1.4) we assumed that there was no external perturbation, meaning that all the individuals would remain in the class. So with an external perturbation cause, the individuals remaining in the class will also follow the curve with equation (1.1.4) multiplied by the probability that they remain in the class, namely $1-\theta(x)$. Therefore the pertubated curve will have the equation

$$\frac{c}{\sigma\sqrt{2\pi}}(1-\theta(x))\, e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}.$$

We just need some more information on the function $\theta(x)$. $\theta(x)$ is a probability, so it lies between 0 and 1. In his paper [23], de Helguero states that he assumes a linear selection law but notes that it is also possible to make different assumptions. He just thought is was the simplest and the most important.
He continues by stating that $\theta(x) = A(x-b)+B$ with b the mean of the hypothetical variation. $\theta(x)$ will be 0 when $x = b - \frac{B}{A}$ which must lie outside the range of the variation if we have a simple selection law and $\theta(x)$ is 1 when $x = b + \frac{1-B}{A}$ which represents the bound of the variation because then we have all of the individuals in class x. Using the substitution $y_0 = c(1-B)$ and $\alpha = -\sigma\frac{A}{1-B}$ we then get

$$\frac{y_0}{\sigma\sqrt{2\pi}}\left(1 - \frac{\alpha(x-b)}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}.$$

Since we see that the factor $\left(1-\frac{\alpha(x-b)}{\sigma}\right)$ is proportional to the distribution of a uniform random variable, we find that this equation is of the currently known form

$$f(x) = k(\lambda_0)G_0(\lambda_0 + w(x;\lambda))\, f_0(x)$$

with λ, λ_0 real parameters, $k(\lambda_0)$ a normalizing constant, f_0 a symmetric density about 0, G_0 a distribution function with density symmetric about 0 and $w(x;\lambda)$ an odd function depending on λ .
Next de Helguero tried to find the four coefficients, namely the normalizing constant y_0, the mean b and the standard deviation σ of the hypothetical normal distribution and the coefficient of perturbation α. Here normalization is meant as equalizing the integral of the curve to the number of observations (instead of 1). The process consists of calculating the moments up to order 3, equating the theoretical moments to the observed ones, and working out the equations with respect to the coefficients. To compute these moments, de Helguero however takes only the condition $1-\phi(x) > 0$ into account what makes him work with the distribution

$$y = \begin{cases} 0 & \text{if } x \leq x_1 \\ \frac{y_0}{\sigma\sqrt{2\pi}}\left(1+\frac{\alpha(x-b)}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2} & \text{if } x_1 \leq x \end{cases} \tag{1.1.5}$$

with x_1 the point where $1-\phi(x)=0$ and assuming $\alpha > 0$.
To calculate the moments

$$v_n = \int_{-\infty}^{+\infty} x^n dF_X(x)$$

of (1.1.5) after application of the translation $b = 0$, de Helguero assumes that $\alpha > 0$. Consider the integral

$$I_n = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{\infty} x^n e^{-\frac{x^2}{2\sigma^2}} dx$$

such that

$$\begin{aligned} I_0 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{\sigma}{\alpha}} e^{-\frac{x^2}{2\sigma^2}} dx \end{aligned}$$

wich is the standard normal distribution function evaluated in $\frac{1}{\alpha}$ and

$$\begin{aligned} I_1 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{\infty} x e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2\alpha^2}} \\ &= \sigma z\left(\frac{1}{\alpha}\right) \end{aligned}$$

with $z(.)$ the standard normal density. Using partial integration we get the recursive formulae

$$\begin{aligned} I_n &= -\frac{\sigma}{\sqrt{2\pi}} \left[x^{n-1} e^{-\frac{x^2}{2\sigma^2}} \right]_{-\frac{\sigma}{\alpha}}^{\infty} + (n-1)\sigma^2 \frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{\infty} x^{n-2} e^{-\frac{x^2}{2\sigma^2}} \\ &= \sigma\left(-\frac{\sigma}{\alpha}\right)^{n-1} z\left(\frac{1}{\alpha}\right) + (n-1)\sigma^2 I_{n-2}. \end{aligned}$$

We get the expression

$$\begin{aligned} v_n &= \frac{y_0}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{+\infty} x^n \left(1 + \frac{\alpha x}{\sigma}\right) e^{-\frac{x^2}{2\sigma^2}} dx \\ &= y_0 \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{+\infty} x^n e^{-\frac{x^2}{2\sigma^2}} dx + \frac{\alpha}{\sigma} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{+\infty} x^{n+1} e^{-\frac{x^2}{2\sigma^2}} \right) \\ &= y_0 \left(I_n + \frac{\alpha}{\sigma} I_{n+1} \right). \end{aligned}$$

With this expression we can calculate the moments up to order 3. After re-shifting the distribution back to location b, we can derive an expression for the normalizing factor

$$y_0 = v_0 \frac{1}{I_0 + \alpha z(\alpha^{-1})}$$

where we set $v_0 = 1$ to apply today's convention of setting the first moment equal to 1. Calculating the mean, the variance and the coefficient of skewness, we get

$$\begin{aligned}
&v_1 = \mathbb{E}(X-b) \\
\iff &\sigma y_0 \alpha I_0 = \mu_1 - b \\
\iff &b = \mu_1 - \sigma y_0 \alpha I_0 \\
&= \mu_1 - \sigma H^{-1},
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left((X-b)^2\right) = v_2 - v_1^2 \\
\iff &\mu_2 = \sigma^2 y_0 \left(I_0 + 2\alpha z(\alpha^{-1}) - y_0 \alpha^2 I_0^2\right) \\
\iff &\sigma^2 = \mu_2 \left(y_0 \left(I_0 + 2\alpha z(\alpha^{-1}) - y_0 \alpha^2 I_0^2\right)\right)^{-1} \\
&= \mu_2 \frac{H^2}{2H^2 - \alpha^{-1}H - 1},
\end{aligned}$$

$$\begin{aligned}
\beta_1 = \frac{\mu_3^2}{\mu_2^3} &= \frac{\left(v_3 - 3v_1v_2 + 2v_1^3\right)^2}{\left(v_2 - v_1^2\right)^2} \\
&= \frac{y_0^2\sigma^6}{y_0^3\sigma^6} \frac{\left(-z(\alpha^{-1}) + 3\alpha I_0 - 3\alpha y_0 I_0^2 - 6\alpha^2 y_0 I_0 z(\alpha^{-1}) + 2\alpha^3 y_0^2 I_0^3\right)^2}{\left(I_0 + 2\alpha z(\alpha^{-1}) - \mathbf{y}_0 \alpha^2 I_0^2\right)^3} \\
&= \frac{1}{y_0} \frac{\left(z(\alpha^{-1}) + 3\alpha^2 y_0 I_0 z(\alpha^{-1}) - 2\alpha^3 y_0^2 I_0^3\right)^2}{\left(I_0 + 2\alpha z(\alpha^{-1}) - \mathbf{y}_0 \alpha^2 I_0^2\right)^3} \\
&= \frac{\left(\alpha^{-1} F H^2 + 3FH - 2\right)^2}{\left(2H^2 - \alpha^{-1}H - 1\right)^3}
\end{aligned}$$

with

$$F = \frac{z(\alpha^{-1})}{I_0}, \qquad H = \frac{1}{\alpha} + F,$$

μ_1 the mean of the observed distribution and μ_2 and μ_3 its central moments of order 2 and 3, respectively. To estimate b, σ and α, de Helguero replaces μ_1, μ_2 and μ_3 with their sample counterparts and solves the equations above.
All the further steps after dropping the condition $1 - \theta(x) < 1$ are coherent with this revised model. So (1.1.5) is normalized properly and its moments are correct. Consequently the estimation procedure based on the method of moments gives consistent estimates.

Preserving the original conditions We will now see if there would have been a different outcome if both conditions $0 < \theta(x)$ and $\theta(x) < 1$ were considered as done by Azzalini and Regoli in their paper [15]. Denote that de Helguero demands for the parameters A and B to be such that the intersection points of $\theta(x)$ with 0 and 1 fall outside the range of variation of the data. This suggests $0 < B < 1$. Set

$$y_0 = c(1-B), \qquad \alpha = -\sigma\frac{A}{1-B}, \qquad \beta = -\sigma\frac{A}{B},$$

then we can write x_0 and x_1, the points where $\theta(x)$ takes the values 0 and 1, respectively, as

$$x_0 = b - \frac{B}{A} = b + \frac{\sigma}{\beta}, \qquad x_1 = b + \frac{1-B}{A} = b - \frac{\sigma}{\alpha}.$$

Here β is an additional parameter. This is necessary because θ was originally a function depending on two parameters, hence it cannot be written as a function of α only.
Assuming $\alpha > 0$, we have $x_1 < x_0$ and the density function is

$$y = \begin{cases} 0 & \text{if } x \le x_1 \\ \frac{\beta}{\alpha+\beta}\frac{c}{\sigma\sqrt{2\pi}}\left(1+\frac{\alpha(x-b)}{\sigma}\right)e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2} & \text{if } x_1 \le x \le x_0 \\ \frac{c}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2} & \text{if } x \ge x_0 \end{cases} \tag{1.1.6}$$

where we have taken $\theta(x) = 0$ for $x > x_0$ by continuity and monotonicity. If $\alpha < 0$, then $x_0 < x_1$ and all inequalities in (1.1.6) must be reversed. We define the integral

$$I_n(\xi) = \int_{\xi}^{\infty} x^n \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}dx$$

and writing v_n as the n^{th} order moment of (1.1.6) shifted to $b = 0$, we get

$$\begin{aligned} v_0 &= c\left(\frac{\beta}{\alpha+\beta}\left(I_0(x_1)-I_0(x_0)+\frac{\alpha}{\sigma}\left(I_1(x_1)-I_1(x_0)\right)\right)+I_0(x_0)\right) \\ &= \frac{c}{\alpha+\beta}\left(\alpha\Phi(-\beta^{-1})+\beta\Phi(\alpha^{-1})+\alpha\beta\left(z(\alpha^{-1})-z(\beta^{-1})\right)\right) \end{aligned}$$

and similarly we get

$$v_n = \frac{c}{\alpha+\beta}\left(\alpha I_n(x_0)+\beta I_n(x_1)+\alpha\beta\left(I_{n+1}(x_1)-I_{n+1}(x_0)\right)\right)$$

Nowadays we want a density normalized to 1, so we set $v_0 = 1$. Therefrom we can write c in function of α and β. In the special case $\alpha = \beta$, we obtain

$$\frac{\alpha}{\alpha+\beta} = \frac{1}{2} \qquad \begin{aligned} v_0 &= \frac{c}{2\alpha}\left(\alpha\Phi(-\alpha^{-1})+\alpha\Phi(\alpha^{-1})+\alpha^2\left(z(\alpha^{-1})-z(\alpha^{-1})\right)\right) \\ &= \frac{c}{2\alpha}\left(\alpha(1-\Phi(\alpha^{-1}))+\alpha\Phi(\alpha^{-1})\right) \\ &= \frac{c}{2}. \end{aligned}$$

Hence $c = 2$ when $v_0 = 1$. This leads to a density of the type $f(x) = 2G_0(w(x;\lambda))f_0(x)$ where, up to a b shift, the normal density in (1.1.6) is multiplied by the distribution function of a random variable on the interval $]-\frac{\sigma}{\alpha}, \frac{\sigma}{\alpha}[$.
Figure 1.1.4 shows the curves of (1.1.5) and the symmetric interval case of (1.1.6) with $\alpha = \beta$, with $\sigma = 1$ in both cases. For $\alpha = \beta = 1$ the curves are very similar, while for $\alpha = \beta = 2$ there is a noticeble difference. The curve of (1.1.5) is smooth over the whole support, while the curve of (1.1.6) is spiked on a point at the right end of the interval $]-\frac{\sigma}{\alpha}, \frac{\sigma}{\alpha}[$.

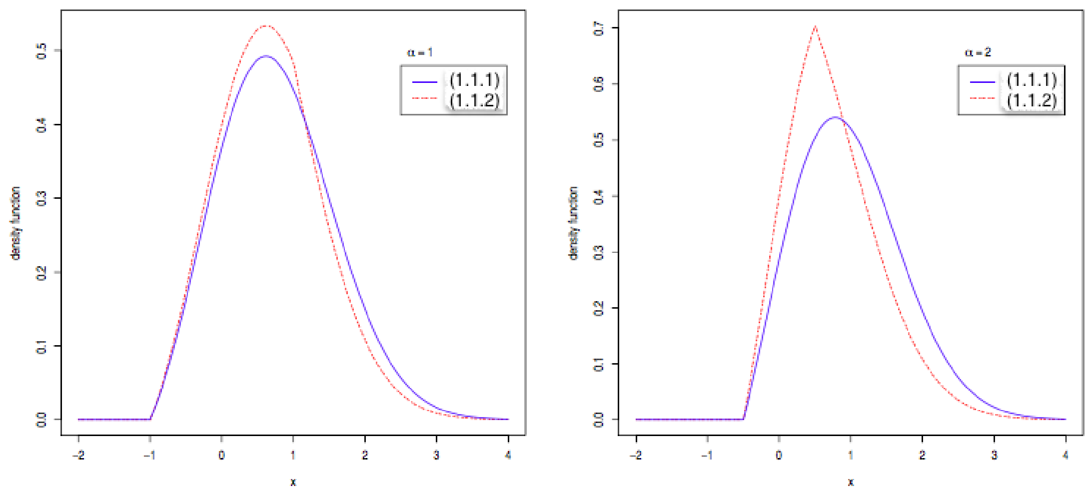

Figure 1.1.4: The de Helgeuro curve (1.1.5) and the density function (1.1.6) in the symmetric interval case with $\alpha = \beta$, with $\sigma = 1$ in both cases.

1.1.2 Later developments

It is clear that, looking to the current literature on skew-symmetric distribution, de Helguero's distribution is the precursor of the renowned skew-normal distribution. It re-appeared in different shapes in the literature as the result of the manipulation of normal variates and involves some of the mechanisms described in the next section, to handle a specific applied problem.

Early reappearences

The idea to construct a family of distributions from the normal distribution by modifying it to model skewness can probably be found in Birnbaum's work of 1950 [18] and independently in the work of O'Hagan and Leonard, published mush later in 1976 [53] as described in Kotz and Vicari (2005) [45]. Weinstein dealt with an analogous problem in 1964 [65] but represented it in a different way. In 1966, Roberts developed his model by selecting the largest or smallest value of normal variables which led to an equivalent proposal [58]. Aigler, Lovell and Schmidt handled the same problem by utilizing the transformation method involving two normal variables in 1977 [1].
We will now take a look at each of the different approaches in more detail as Azzalini (2005) [8] did.

Birnbaum : conditional inspection and selective sampling Birnbaum discussed the following problem when he came across a practical difficulty in educational testing. Let U_1 be the score a given individual received on an educational test, where U_1 can be obtained as a linear combination of several such tests. Let U_0 be the score the same individual received in the admission examination. Suppose that (U_0, U_1) follows the bivariate normal distribution with unit marginals and correlation ρ. Subjects are examined in the subsequent tests given that the admission score exceeds a certain threshold τ', so the distribution will be the one of $Z = (U_1 | U_0 > \tau')$. This will result in what we now know as the extended skew-normal distribution (see Chapter 2)

$$\phi(z)\frac{\Phi(\tau\sqrt{1+\delta^2}+\delta z)}{\Phi(\tau)}$$

with $\delta = \rho/\sqrt{1-\rho^2}$ and $\tau = -\tau'$. This reduces to the skew-normal distribution when $\tau = 0$. We can assume without loss of generality that the marginal distributions of U_0 and U_1 have the same location parameters since a potential difference can be absorbed in τ. When we have the location parameter equal to zero and the scale parameter equal to 1, we can use the transformation $Y = \xi + \omega Z$.

Roberts : selecting maxima Assume (U_0, U_1) as in the previous paragraph and consider the distribution of $\max(U_0, U_1)$ and of $\min(U_0, U_1)$. Roberts has analyzed this problem in the studies of twins, where U_0 and U_1 are the measurements taken on a pair of twins. Because it were twins being measured, assuming an equal distribution of the two components seems reasonable. The joint density of (U_0, U_1) as derived in [17] is

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{y^2-2xy\rho+x^2}{2(1-\rho^2)}\right) \qquad \text{for } -\infty < x < \infty, \quad -\infty < y < \infty$$

with ρ the correlation coefficient of X and Y.
Analogous to the proof of Roberts (1966) [58] for the minimum, we can find the density for $Z = \max(U_0, U_1)$.

Theorem 1.1.1. *The density for* $Z = \max(U_0, U_1)$ *is*

$$h(z) = \frac{2}{\sqrt{2\pi}}\Phi\left(z\sqrt{\frac{1-\rho}{1+\rho}}\right)e^{-\frac{z^2}{2}} \qquad \textit{for } -\infty < z < \infty$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du$.

Proof. Define $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$ and let $H(Z) = P(Z \leq z)$. We have $H(Z) = F(z,z)$. However, using the Leibniz integral rule

$$\begin{aligned}\frac{d}{dz}F(z,z) &= 2\int_{-\infty}^{z} f(z,y)dy \\ &= 2\int_{-\infty}^{z}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{y^2-2zy\rho+z^2}{2(1-\rho^2)}\right)dy\end{aligned}$$

$$= \frac{2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \int_{-\infty}^{z} \exp\left(-\frac{(y-\rho z)^2}{2(1-\rho^2)}\right) dy$$
$$= \frac{2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Phi\left(z\sqrt{\frac{1-\rho}{1+\rho}}\right)$$

Observing that

$$h(z) = \frac{d}{dz} F(z,z),$$

the proof is complete. □

The distribution of $\max(U_0, U_1)$ is thus the skew-normal distribution (see Chapter 2)

$$2\phi(z)\Phi(\delta z)$$

with shape parameter $\delta = \sqrt{1-\rho}/\sqrt{1+\rho}$. To obtain the distribution of $\min(U_0, U_1)$ we have to reverse the sign of the shape parameter or see Roberts(1966) [58] for the proof.

Weinstein : convolution of normal and truncated-normal Weinstein was interested in the cumulative distribution function of the sum of two independent normal variables V_0 and V_1, when V_0 is truncated by limiting it so it would not exceed a certain threshold. Say if V_0 and V_1 are independent, $V_0, V_1 \sim N(0,1)$ and $\alpha \in]1,1[$, then as proved in Kim (2006) [43]

$$Z = \frac{1}{\sqrt{1+\alpha^2}}|V_0| + \frac{\alpha}{\sqrt{1+\alpha^2}} V_1$$

follows the extended skew-normal distribution (see Chapter 2).

O'Hagan & Leonard O'Hagan and Leonard discussed a closely related construction, even though they formulated it differently. Let θ be the mean value of a normal population for which previous considerations suggest that $\theta > 0$ but we are not entirely certain about this. We can deal with this uncertainty by constructing the previous distribution of θ in two stages, assuming that $\theta|\mu \sim N(\mu, \sigma^2)$ and that μ has a distribution of type $N(\mu_0, \sigma_0^2)$ truncated when smaller than 0. The resulting distribution of θ as found by O'Hagan & Leonard (1976) [53] is

$$\pi(\theta) = \phi\left((\sigma^2 + \sigma_0^2)^{\frac{1}{2}}(\theta - \mu_0)\right)\Phi\left((\sigma^{-2} + \sigma_0^{-2})^{-\frac{1}{2}}(\sigma^{-2}\theta + \sigma_0^{-2}\mu_0)\right)$$

where $\phi(.)$ and $\Phi(.)$ respectively denote the standard normal density and distribution function. We get a distribution corresponding to the sum of a normal and a truncated normal variable as the distribution for θ. When the threshold value of the variable V_0 coincides with $\mathbb{E}(V_0)$, the sum will take the form $a|V_0| + bV_1$, for some real values a and b, and $|V_0|$ is a half-normal variable. Without loss of generality we may consider the special case

$$Z = \alpha|V_0| + \sqrt{1-\alpha^2} V_1$$

where V_0 and V_1 are independent $N(0,1)$ variables, and $\alpha \in]-1,1[$. The distribution of Z is the skew-normal distribution with shape parameter $\alpha/\sqrt{1-\alpha^2}$.

Aigler, Lovell and Schmidt : transformation method The Z discussed in the paragraph above has the structure of the random term showing up in the econometric literature dealing with stochastic frontier analysis and thus also in the paper of Aigner *et al.* Here the response variable is provided by the output produced by some economic unit of a given type, and a regression model is constructed to represent the relationship between the response variable and a set of covariates which expresses the input factors used to acquire the corresponding output. This regression model differs from ordinary regression models mainly because here the stochastic component is the sum of two terms: one is a standard error term centred around zero and the other is an essentially negative quantity, which stands for the inefficiency of a production unit, producing an output level below the curve of technical efficiency. Like V_1 in the previous paragraph, the purely random term is normal and the inefficiency is assumed to be of type $\alpha|V_0|$ with $\alpha < 0$. We thus have a regression model with an error term of the skew-normal type.

Adelchi Azzalini

Considering the skew-normal distribution as a distribution of independent interest instead of via certain transformations of normal variates, for its ability to incorporate skewness in the data modelling process is a more recent idea.
This seems to start with Adelchi Azzalini and the skew-normal owes its fame to Azzalini's 1985 paper [7], which is among the most quoted papers in the literature on skewed distributions. It consists of modifying the normal probability density function by multiplication with a skewing function. Azzalini stated that

$$2f(x)G(\delta x)$$

is a pdf where f is the density of a variable symmetric around 0, and G is the cdf of another independent random variable. By combining different symmetric distributions (normal, t, logistic, uniform, double exponential , etc.) numerous families of skewed distributions may be generated. Years later, the original result was extended to the multivariate case by Azzalini and Dalla Valle (1996) [13], which also generated a lot of attention. Further work on the properties of the class of skew-normal densities and on the associated inferential problems has been developed by several authors, including Azzalini himself together Reinaldo Arellano-Valle and Antonella Capitanio.
More on this skew-normal distribution and its properties can be found in the next chapters.

Barry Arnold

An important publication by Arnold *et al.* (1993) [6] provided applications and further elaborations and interpretations. Arnold also considered the extended skew-normal distribution

$$\phi(z)\frac{\Phi(\tau\sqrt{1+\delta^2}+\delta z)}{\Phi(\tau)}$$

extensively, after Azzalini had briefly considered them, see Section 2.1.3. Arnold also developed diverse skewing methods, including hidden truncation.

Marc Genton

Genton is one of the main contributers to the multivariate skewed distributions. He and his coworkers initiated further research in the multivariate case of the skew-normal distribution.
The early years of the 21st century also produced a number of valuable results dealing with generalized skew elliptical distributions which led to the book edited by Genton on skew-elliptical distributions : *Skew-Elliptical Distributions and Their Applications : A Journey Beyond Normality*' [31]. The probability density function of generalized skew-elliptical distributions is as follows

$$\frac{2}{|\Omega|^{\frac{1}{2}}} g\left(\Omega^{-1/2}(z-\xi)\right) \pi\left(\Omega^{-1/2}(z-\xi)\right)$$

with $\xi \in \mathbb{R}^p$ the location vector parameter, $\Omega \in \mathbb{R}^{pxp}$ the scale matrix parameter, g the pdf of a spherical distribution and π a skewing function. $|\Omega|$ signifies the absolute value of the determinant of Ω. Skew-elliptical distributions include skew-normal ones as well as elliptical ones.

1.2 Applications

There are a lot of possible applications of the skew-symmetric distributions. We give a few that can be linked directly to the results described above, as they are described in Azzalini (2005) [8] and Azzalini (2006) [9]. We will also highlight the connection with some areas of work that do not seem related at first sight.

Selective sampling

Assuming normality of the overall population, the goal of this selection is to produce a skew-normal distribution for the observable data. To get a formulation, start from the relationships

$$Y_0 = X_0\beta_0 + U_0, \qquad Y_1 = X_1\beta_1 + U_1,$$

where (U_0, U_1) is a bivariate normal variable, and β_0, β_1 are unknown parameters. The X's and Y's are observable but, because of the method of selection in the sampling process, we observe Y_1 only when $Y_0 > 0$. The construction is then analogous to the genesis by conditioning as noted by Birnbaum, leading to the extended skew-normal distribution.
Selective sampling has been widely studied in quantitative sociology with a model called the 'Heckman model', firstly introduces by Heckman in the 70's. The literature on Heckman model focuses strongly on the normality assumption. This main focus caused a lot of criticism because the normality assumption was often violated in practice which led to the development of a more robust estimation procedure. But both methods were very sensitive to high correlation between the different variables. Many other estimation approaches were proposed over the years. It is possible that they can produce similar but more flexible and realistic methods. One can expect the skew-elliptical distributions, especially the skew-t distribution, as the underlying distribution to be useful. One of the most common deviations from normality in practice is when the distribution of the data has heavier tails than in the normal distribution. This makes it a very natural choice to use the Student-t distribution as proposed by Genton and Marchenko [51].

Observation of the maximal component

In many different situations, observations are set in pairs, specifically in the medical sector. But the main interest is often obtaining the maximal value (or the minimal in other cases). For example, in the ophthalmology, the sharpness of vision in both eyes is often measured, but the maximum of these two values can be considered as the single response value for certain purposes. Assuming joint normality and equal marginal distribution of the two measurements, the distribution of the maximum value is skew-normal, like we obtained in the mechanism of selecting maxima by Roberts (1966) [58].

Financial markets

The presence of long tails in the observed distribution is present almost everywhere in financial applications. It is also required for data modelling that there is a strong formulation for the error term,involving say, a Student-t distribution.
More recently, skewness was taken more and more into consideration for a more accurate data modelling. We can not only motivate this change by support from empirical observations but also by qualitative arguments, since financial markets react inversely but with different amplitude to positive and negative information coming for instance from other markets. The skew-normal distributions seem a good fit, because they also keep the main properties of the economic formulation.

Adaptive designs in clinical trials

The enormous cost of clinical trials carried out for drug development, increases more and more. Therefore people want to limit these costs. To attempt this, adaptive designs are currently of interest in medical statistics. A possible way of working in this context is looking at the combination of the outcome of a phase II study and the outcome of a phase III study. There are two facts we have to take into account when working like this: the first is that the phase III study is only carried out if the phase II was successful, the other is that the two studies often consider a different endpoint. The condition of success of phase II that we need to keep in mind suggests, under normality assumption of the variables, a skew-normal component of the resulting likelihood function can be considered.

Compositional data

We can find compositional data in many different fields, but the regular situation is represented in the geological context. To analyse this kind of data a regularly used method is to transform the $d+1$ original components belonging to the simplex to d components in $\mathbb{R}^d$ using the additive log-ratio transform. This is then followed by an analysis based on methods for normal data. After the additive log-ratio transformation, we can assume skew-normality on the transformed data instead of assuming normality, to improve adequacy in data fitting. This assumption on $\mathbb{R}^d$ brings forth a distribution on the simplex which has some desirable properties, which are due to the properties of closure under marginalisation and affine transformation of the skew-normal distribution, inducing some corresponding properties on the simplex.

Flooding risk

Estimating the flooding risk is a practical application of the skew-elliptical distributions, more precisely the skew-t distributions. This can be constructed by modeling the distribution of the sea levels over a long time and using the skew-t distribution to predict changes in flooding risk associated with rising sea level. The skew-t distribution will prove to be an effective description of the sea level process and can be used to take into account its strong seasonality and other form of nonstationarity.

Chapter 2

Skew-symmetric family

In the historical developments of the skew-symmetric distributions discussed in the previous chapter, we have seen the focus of interest shift from applying certain transformations to making the transformed data follow the normal distribution and then finally to developing an extension to the normal family to incorporate skewness in the data modelling process. In this chapter we will look at these new parametric families from a more theoretical point of view. Some basic properties will be set out along with the moment generating function and the moments based on two examples of families of skew-symmetric distributions.

The skew-symmetric family as defined in Hallin and Ley (2014) [39], is a parametric family of probability density functions of the form

$$x \mapsto f_{\boldsymbol{\vartheta}}^{\Pi}(x) := 2\,\sigma^{-1} f(\sigma^{-1}(x-\mu))\Pi(\sigma^{-1}(x-\mu), \delta), \quad x \in \mathbb{R}, \tag{2.0.1}$$

where

- $\boldsymbol{\vartheta} = (\mu, \sigma, \delta)'$, with $\mu \in \mathbb{R}$ a *location parameter*, $\sigma \in \mathbb{R}_0^+$ a *scale parameter* and $\delta \in \mathbb{R}$ a *skewness parameter*;
- $f : \mathbb{R} \to \mathbb{R}_0^+$, the *symmetric kernel*, is a nonvanishing symmetric pdf (such that, for any $z \in \mathbb{R}$, $0 \neq f(-z) = f(z)$), and
- $\Pi : \mathbb{R} \times \mathbb{R} \to [0, 1]$ is a *skewing function*, that is, it satisfies

  $$\Pi(-z, \delta) + \Pi(z, \delta) = 1, \quad z, \delta \in \mathbb{R}, \text{ and } \Pi(z, 0) = \frac{1}{2}, \quad z \in \mathbb{R}, \tag{2.0.2}$$

  and, in case $(z, \delta) \mapsto \Pi(z, \delta)$ admits a derivative of order s at $\delta = 0$ for all $z \in \mathbb{R}$,

  $$\partial_z^s \Pi(z, \delta)|_{\delta=0} = 0, \ \ z \in \mathbb{R} \quad \text{and, for } s \text{ even,} \quad \partial_\delta^s \Pi(z, \delta)|_{\delta=0} = 0, \ \ z \in \mathbb{R}. \tag{2.0.3}$$

The condition (2.0.3) can be explained by the analogy with skewing functions of the form $\Pi(z,\delta) = \Pi(\delta z)$, which are the most common ones. If Π is s times continuously differentiable, $\partial_z^s \Pi(\delta z) = \delta^s(\partial^s \Pi)(\delta z)$ vanishes at $\delta = 0$, because of multiplication by zero. The fact that $\Pi(-y)+\Pi(y) = 1, y \in \mathbb{R}$, implies that $\partial^s \Pi(\delta z)$ cancels at $\delta = 0$ for even values of s, with $\partial^s \Pi(\delta z)$ the s^{th} derivative of $\Pi(\delta z)$ with respect to δ. This can be shown by deriving s times both sides of the equality $\Pi(-y)+\Pi(y) = 1$. We get

$$(-z)^s \partial^s \Pi(\delta z) + z^s \partial^s \Pi(\delta z) = 0 \qquad (2.0.4)$$
$$\iff \partial^s \Pi(\delta z).((-z)^s + z^s) = 0.$$

So either $(-z)^s + z^s = 0$ or $\partial^s \Pi(\delta z) = 0$. If s is odd, we get $(-z)^s + z^s = -z^s + z^s = 0$, so equation (2.0.4) is always zero no matter what the value for $\partial^s \Pi(\delta z)$ is. If s is even then $(-z)^s + z^s = z^s + z^s = 2z^s \neq 0$. We find for s even that $\partial^s \Pi(\delta z)$ has to be zero for equation (2.0.4) to be true.
We will give more insight in this family by giving a few examples, in particular the skew-normal family and the skew-t family.

2.1 Skew-normal family

A first example of such a skew-symmetric family is the skew-normal family whose probability density function is given by

$$\phi(z;\delta) = 2\phi(z)\Phi(\delta z), \qquad -\infty < z < +\infty, \qquad (2.1.1)$$

as proposed by Azzalini [7], where the symmetric kernel f is the standard Gaussian pdf ϕ and the skewing function $\Pi(z,\delta) = \Phi(\delta z)$ with Φ the standard Gaussian cumulative distribution function. When discussing the skew-normal family we will use the outline of a book by Azzalini (2013) [10].
If Z is a continuous random variable with density function (2.1.1), then the variable $Y = \mu + \sigma Z$ $(\mu \in \mathbb{R}, \sigma \in \mathbb{R}_0^+)$ is a skew-normal variable with density function at $x \in \mathbb{R}$

$$2\sigma^{-1}\phi(\sigma^{-1}(x-\mu))\Phi(\delta\sigma^{-1}(x-\mu)) = \sigma^{-1}\phi(\sigma^{-1}(x-\mu);\delta). \qquad (2.1.2)$$

We will use the notation

$$Y \sim \text{SN}(\mu, \sigma^2, \delta).$$

When $\mu = 0$ and $\sigma = 1$, we have the density (2.1.1) again. We then say that the distribution is normalized. Figure 2.1.1 shows the variation of the pdf with the skewness parameter.

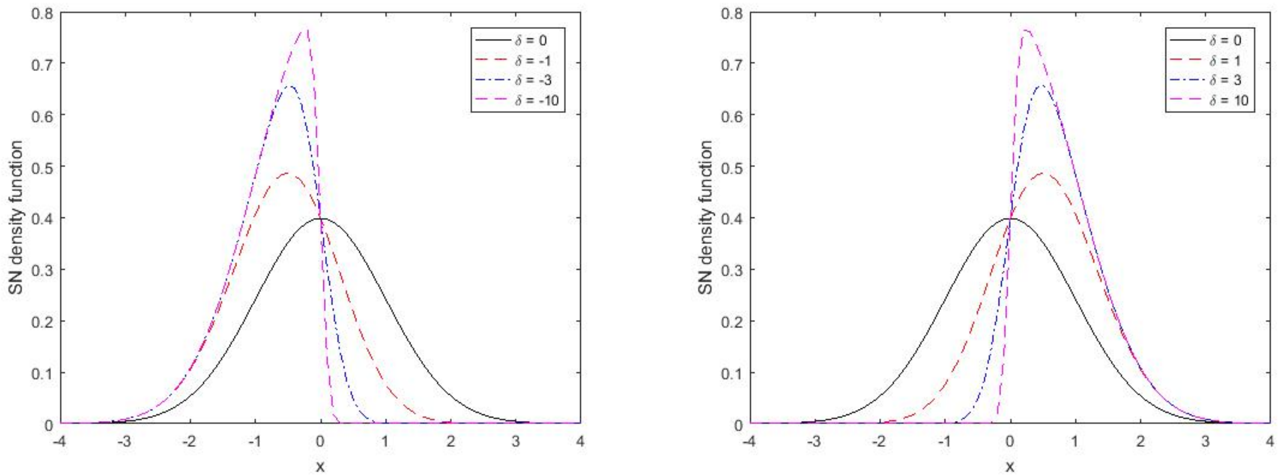

Figure 2.1.1: Skew-normal density functions for varying δ.

2.1.1 Properties

Suppose $Z \sim \text{SN}(0,1,\delta)$. The case $\delta = 0$ corresponds to the standard normal distribution. So the standard normal distribution is an element of the family of skew-normal densities. We will now prove a first property of the skew-normal family, namely the chi-squared property.

Property 2.1.1. *$Z^2 \sim \chi_1^2$, regardless of δ.*

Proof. We will prove this property by showing that $|Z|$ and $|X|$, with $X \sim \text{N}(0,1)$, have identical distributions. It then follows that Z^2 will be identically distributed as X^2, which is χ_1^2.

$$\begin{aligned}
P(|Z| \leq z) &= \int_{-z}^{z} 2\phi(u)\Phi(\delta u)du \\
&= \int_{0}^{z} 2\phi(u)\Phi(\delta u)du + \int_{-z}^{0} 2\phi(u)\Phi(\delta u)du \\
&= \int_{0}^{z} 2\phi(u)\Phi(\delta u)du - \int_{z}^{0} 2\phi(-u)\Phi(-\delta u)du \\
&= \int_{0}^{z} 2\phi(u)\Phi(\delta u)du + \int_{0}^{z} 2\phi(u)\Phi(-\delta u)du \\
&= \int_{0}^{z} 2\phi(u)\big(\Phi(\delta u) + \Phi(-\delta u)\big)du \\
&= \int_{0}^{z} 2\phi(u)du \\
&= P(|X| \leq z).
\end{aligned}$$

This proves the property. □

We will now give some other properties.

Property 2.1.2. *If $Z \sim SN(0,1,\ \delta)$ the following properties are true :*

(a) $\phi(0;\delta) = \phi(0), \forall \delta$;

(b) $-Z \sim SN(0,\ 1,\ -\delta)$, *equivalently* $\phi(-x;\delta) = \phi(x;-\delta), \forall \delta$;

(c) *if* $Z' \sim SN(0,1,\ \delta')$ *with* $\delta' \leq \delta$, *then* $Z' \leq_{st} Z$ *i.e.* $P(Z' > x) \leq P(Z > x), \forall x \in \mathbb{R}$.

Proof. (a) This follows immediately from the definition (2.1.1).

(b) We have

$$\Phi_{-Z}(x;\delta) = P(-Z \leq x) = P(Z \geq -x) = 1 - P(Z \leq -x) = 1 - \Phi_Z(-x;\delta).$$

We derive both sides of the equation to get

$$\phi_{-Z}(x;\delta) = \phi_Z(-x;\delta)$$

where

$$\phi_Z(-x;\delta) = 2\phi(-x)\Phi(-\delta x) = 2\phi(x)\Phi(-\delta x) = \phi_Z(x;-\delta)$$

because of the symmetry of the distribution function ϕ of the normal distribution. We thus find that $-Z \sim$ SN(0,1,$-\delta$).

(c) We consider, for a fixed x and an arbitrary δ, the function $\delta \mapsto h(\delta) = \Phi(x;\delta)$. Because Z is a continuous variable, $\phi(z;\delta)$ is continuous and continuously differentiable. Therefore we can use the Leibniz integral rule, we have

$$\begin{aligned}
h'(\delta) &= 2\int_{-\infty}^{x} \phi(t)\frac{\partial}{\partial\delta}\Phi(\delta t)dt \\
&= 2\int_{-\infty}^{x} t\phi(t)\phi(\delta t)dt \\
&= \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{x} t\phi(t\sqrt{1+\delta^2})dt \\
&= \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{x} -\phi'(t\sqrt{1+\delta^2})dt \\
&= -\frac{2}{\sqrt{2\pi(1+\delta^2)}}\phi(x\sqrt{1+\delta^2})
\end{aligned}$$

where we have used $\sqrt{2\pi}\phi(at)\phi(bt) = \phi(t\sqrt{a^2+b^2})$ and $\phi'(t) = -t\phi(t)$. We have found that $h(\delta)$ is decreasing and that

$$\begin{aligned}
&\Phi(x;\delta') \geq \Phi(x;\delta) \\
\iff\ & P(Z' \leq x) \geq P(Z \leq x) \\
\iff\ & P(Z' > x) \leq P(Z > x).
\end{aligned}$$

We thus find $Z' \leq_{st} Z$.

□

2.1.2 Moment generating function and moments

The result on the normal distibution mentioned below has been stated by numerous authors.

Theorem 2.1.1. *If U ∼ N(0,1) then*

$$\mathbb{E}(\Phi(hU+k)) = \Phi\left(\frac{k}{\sqrt{1+h^2}}\right) \qquad h,k \in \mathbb{R}.$$

Proof. Let Y be a standard normal variable. We define the function $\Psi(h,k), \forall h,k \in \mathbb{R}$ as follows

$$\Psi(h,k) = \int_{-\infty}^{+\infty} \Phi(hy+k)\phi(y)dy.$$

Then $\Psi(h,k) = \mathbb{E}\left(\Phi(hy+k)\right)$. Differentiating $\Psi(hy+k)$ with respect to k and using the Leibniz integral rule because $\Phi(y)$ and $\phi(y)$ are continuous functions, we get

$$\begin{aligned}
\frac{\partial \Psi(hy+k)}{\partial k} &= \int_{-\infty}^{+\infty} \phi(hy+k)\phi(y)dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-\frac{(hy+k)^2+y^2}{2}\right)dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-\frac{(h^2+1)y^2+2hky+k^2}{2}\right)dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-\frac{h^2+1}{2}\left(y^2+\frac{2hky}{1+h^2}+\frac{h^2k^2}{(1+h^2)^2}-\frac{h^2k^2}{(1+h^2)^2}+\frac{k^2}{1+h^2}\right)\right)dy \\
&= \frac{1}{2\pi}\exp\left(-\frac{k^2}{2(1+h^2)}\right)\int_{-\infty}^{+\infty} \exp\left(-\frac{1+h^2}{2}\left(y+\frac{hk}{1+h^2}\right)^2\right)dy
\end{aligned}$$

$$\textbf{subst. : } \boldsymbol{u = \sqrt{1+h^2}\left(y+\frac{hk}{1+h^2}\right)}$$

$$\begin{aligned}
&= \frac{1}{2\pi\sqrt{1+h^2}}e^{-\frac{k^2}{2(1+h^2)}}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}du \\
&= \frac{\sqrt{2\pi}}{2\pi\sqrt{1+h^2}}e^{-\frac{k^2}{2(1+h^2)}} \\
&= \frac{1}{\sqrt{1+h^2}}\phi\left(\frac{k}{\sqrt{1+h^2}}\right).
\end{aligned}$$

Now, integrating with respect to k, we have

$$\Psi(h,k) = \Phi\left(\frac{k}{\sqrt{1+h^2}}\right) + C$$

with C a constant. Letting $k \to \infty$, we find that $C = 0$, which proves the lemma. □

From this result we can find the moment generating function of Y. Y has a skew-normal distribution with expected value μ, standard deviation σ and skewness δ, so if $Z \sim SN(0,1,\delta)$ then $Y = \mu + \sigma Z$.

$$\begin{aligned}
M_Y(t) = \mathbb{E}(e^{Yt}) &= \mathbb{E}(\exp(\mu t + \sigma Z t)) \\
&= 2\int_{-\infty}^{+\infty} \exp(\mu t + \sigma z t)\phi(z)\Phi(\delta z)dz \\
&= 2\exp(\mu t)\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\exp(\sigma z t)\exp\left(\frac{-z^2}{2}\right)\Phi(\delta z)dz \\
&= 2\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-(z-\sigma t)^2}{2}\right)\Phi(\delta z)dz
\end{aligned}$$

$$\textbf{subst. : } \boldsymbol{u = z - \sigma t}$$

$$\begin{aligned}
&= 2\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-u^2}{2}\right)\Phi(\delta u + \sigma\delta t)du \\
&= 2\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\int_{-\infty}^{+\infty} \phi(u)\Phi(\delta u + \sigma\delta t)du \\
&= 2\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\mathbb{E}\left(\Phi(\delta U + \sigma\delta t)\right).
\end{aligned}$$

Using the result of theorem (2.1.1), this becomes

$$M_Y(t) = 2\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\Phi(\sigma t\lambda) \quad \text{where} \quad \lambda = \frac{\delta}{\sqrt{1+\delta^2}}. \tag{2.1.3}$$

We can now compute the moments of $Y \sim \mathrm{SN}(\mu, \sigma^2, \delta)$ via the moment generating function (2.1.3), or equivalently via the cumulant generating function

$$K(t) = \log M_Y(t) = \mu t + \frac{t^2\sigma^2}{2} + \zeta_0(\lambda\sigma t)$$

where

$$\zeta_0(x) = \log(2\Phi(x)).$$

We will also need the derivatives

$$\zeta_r(x) = \frac{d^r}{dx^r}\zeta_0(x) \quad (r = 1, 2, \ldots)$$

whose expressions, for the first few orders, are

$$\zeta_1(x) = \frac{\phi(x)}{\Phi(x)},$$

$$\begin{aligned}\zeta_2(x) &= -\frac{\phi^2(x)}{\Phi^2(x)} - x\frac{\phi(x)}{\Phi(x)}\\ &= -x\zeta_1(x) - \zeta_1^2(x),\end{aligned}$$

$$\begin{aligned}\zeta_3(x) &= -\zeta_1(x) - x\zeta_2(x) - 2\zeta_1(x)\zeta_2(x)\\ &= -\zeta_1(x) - x(-x\zeta_1(x) - \zeta_1^2(x)) - 2\zeta_1(x)(-x\zeta_1(x) - \zeta_1^2(x))\\ &= -\zeta_1(x) + x^2\zeta_1(x) + 3x\zeta_1^2(x) + 2\zeta_1^3(x),\end{aligned}$$

$$\begin{aligned}\zeta_4(x) &= -\zeta_2(x) + 2x\zeta_1(x) + x^2\zeta_2(x) + 3\zeta_1^2(x) + 6x\zeta_1(x)\zeta_2(x) + 6\zeta_1^2(x)\zeta_2(x)\\ &= x\zeta_1(x) + \zeta_1^2(x) + 2x\zeta_1(x) + x^2(-x\zeta_1(x) - \zeta_1^2(x)) + 3\zeta_1^2(x) + 6x\zeta_1(x)(-x\zeta_1(x) - \zeta_1^2(x)) + 6\zeta_1^2(x)(-x\zeta_1(x) - \zeta_1^2(x))\\ &= -6\zeta_1^4(x) - 12x\zeta_1^3(x) - 7x^2\zeta_1^2(x) + 4\zeta_1^2(x) - x^3\zeta_1(x) + 3x\zeta_1(x).\end{aligned}$$

For the expected value and variance of Y we have

$$\mathbb{E}(Y) = \mathbb{E}(\mu + \sigma Z) = \mu + \sigma\mu_Z, \tag{2.1.4}$$

$$\text{var}(Y) = \text{var}(\mu + \sigma Z) = \sigma^2\sigma_Z. \tag{2.1.5}$$

Using the expressions for the first 4 orders of ζ_r, we can derive the derivatives of $K(t)$ up to fourth order immediately. This leads to calculating $\mathbb{E}(Y)$ en $\text{var}(Y)$ in a different way. We get

$$\begin{aligned}\mathbb{E}(Y) &= K'(0)\\ &= \mu + \sigma^2.0 + \lambda\sigma\zeta_1(0)\\ &= \mu + \sigma\lambda b \qquad (2.1.6)\end{aligned}$$

$$\begin{aligned}\text{var}(Y) &= K''(0)\\ &= \sigma^2 + \lambda^2\sigma^2\zeta_2(0)\\ &= \sigma^2(1 - b^2\lambda^2) \qquad (2.1.7)\end{aligned}$$

where

$$b = \zeta_1(0) = \frac{\phi(0)}{\Phi(0)} = \sqrt{\frac{2}{\pi}}.$$

It thus follows that

$$\mu_Z = b\lambda \quad \text{and} \quad \sigma_Z^2 = 1 - b^2\lambda^2.$$

We can also calculate the third and fourth cumulant

$$\begin{aligned}
\mathbb{E}\Big((Y-\mathbb{E}(Y))^3\Big) &= K'''(0) \\
&= \lambda^3\sigma^3\zeta_3(0) \\
&= \lambda^3\sigma^3(2b^3-b) \\
&= \mu_Z^3\sigma^3\frac{4-\pi}{2},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Big((Y-\mathbb{E}(Y))^4\Big) &= K''''(0) \\
&= \sigma^4\lambda^4\zeta_4(0) \\
&= \sigma^4\lambda^4(-6b^4+4b^2) \\
&= 2\sigma^4\mu_Z^4(\pi-3).
\end{aligned}$$

By standardizing this third and fourth cumulant we get the commonly used measures for skewness and kurtosis

$$\gamma_1(Y) = \frac{K'''(0)}{(K''(0))^{\frac{3}{2}}} = \frac{4-\pi}{2}\frac{\mu_Z^3}{\sigma_Z^3},$$

$$\gamma_2(Y) = \frac{K''''(0)}{(K''(0))^2} = 2(\pi-3)\frac{\mu_Z^4}{\sigma_Z^4}.$$

2.1.3 Extended skew-normal distribution

Using Theorem 2.1.1., we can introduce an extension of the skew-normal family of distributions, since

$$\begin{aligned}
&\int_{-\infty}^{\infty}\phi(x)\Phi(\alpha_0+\alpha x)dx = \mathbb{E}\left(\Phi(\alpha_0+\alpha X)\right) \\
\iff &\int_{-\infty}^{\infty}\phi(x)\Phi(\alpha_0+\alpha x)dx = \Phi\left(\frac{\alpha_0}{\sqrt{1+\alpha^2}}\right) \\
\iff &\frac{1}{\Phi\left(\frac{\alpha_0}{\sqrt{1+\alpha^2}}\right)}\int_{-\infty}^{\infty}\phi(x)\Phi(\alpha_0+\alpha x)dx = 1
\end{aligned}$$

for any α_0 and α. It corresponds to adopting a simple modification of the parameters, and to considering the density function

$$\phi(x;\delta,\tau) = \phi(x)\frac{\Phi\left(\tau\sqrt{1+\delta^2}+\delta x\right)}{\Phi(\tau)}, \qquad x\in\mathbb{R}, \tag{2.1.7}$$

with $(\delta,\tau)\in\mathbb{R}\times\mathbb{R}$.

We call this the extended skew-normal distribution since (2.1.7) reduces to (2.1.1) when $\tau = 0$, and more generally for any variable $Y = \mu + \sigma Z$, if Z has density function (2.1.7).
We will use the notation

$$Y \sim SN(\mu, \sigma^2, \delta, \tau)$$

where the occurance of the parameter τ indicates that we are referring to an extended skew-normal distribution. Notice that the value of τ becomes irrelevant when $\delta = 0$.
Figure (2.1.2) shows us the shape of the density for $\delta = 3$ and $\delta = 10$ with different choices for τ. It is clear that the effect of the new parameter τ is dependent on the value of δ. For $\alpha = 3$, the effect of letting τ vary, is much the same as could be achieved by setting $\tau = 0$ and selecting a suitable value of α. For $\alpha = 10$, with the variation of τ, the density function changes in a more elaborate way.

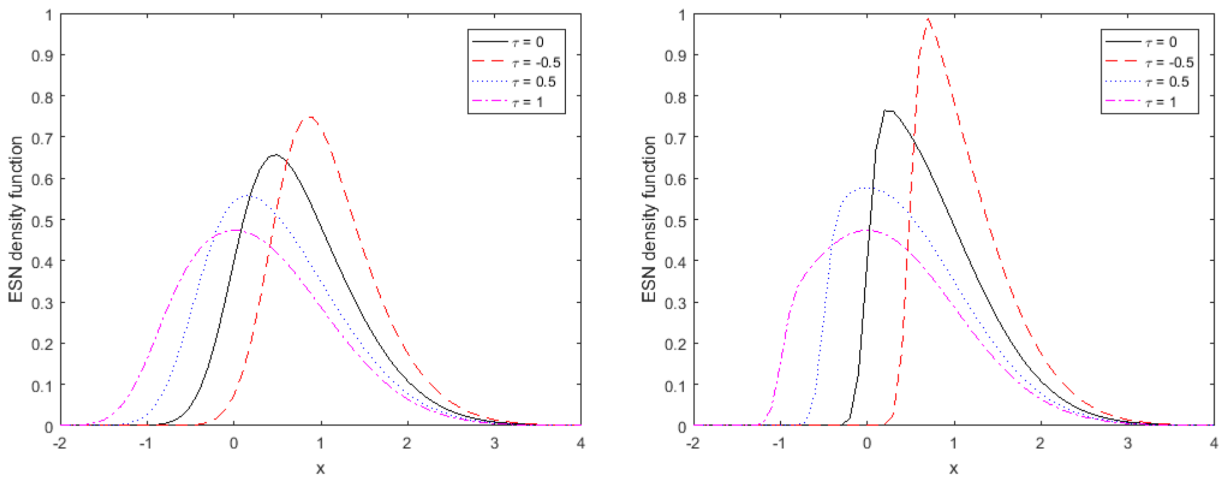

Figure 2.1.2: Extended skew-normal density functions when $\alpha = 3$ and $\alpha = 10$ with varying values of τ.

We can compute the moment generating function of $Y = \mu + \sigma Z$ where $Z \sim SN(0, 1, \delta, \tau)$ the same way we did for the skew-normal case. Making use of Theorem 2.1.1 again, we get

$$\begin{aligned}
M_Y(t) &= \mathbb{E}(\exp(\mu t + \sigma t Z)) \\
&= \int_{-\infty}^{\infty} \exp(\mu t + \sigma t z)\phi(z)\frac{\Phi\left(\tau\sqrt{1+\delta^2} + \delta z\right)}{\Phi(\tau)} dz \\
&= \frac{\exp(\mu t)}{\sqrt{2\pi}\Phi(\tau)} \int_{-\infty}^{\infty} e^{\sigma t z} e^{-\frac{z^2}{2}} \Phi\left(\tau\sqrt{1+\delta^2} + \delta z\right) dz \\
&= \frac{\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)}{\sqrt{2\pi}\Phi(\tau)} \int_{-\infty}^{+\infty} \exp\left(\frac{-(z - \sigma t)^2}{2}\right) \Phi\left(\tau\sqrt{1+\delta^2} + \delta z\right) dz
\end{aligned}$$

$$\textbf{subst. : } \boldsymbol{u = z - \sigma t}$$

$$
\begin{aligned}
&= \frac{\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)}{\sqrt{2\pi}\Phi(\tau)} \int_{-\infty}^{+\infty} \exp\left(\frac{-u^2}{2}\right)\Phi(\tau\sqrt{1+\delta^2} + \delta u + \sigma\delta t)du \\
&= \frac{\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)}{\Phi(\tau)} \int_{-\infty}^{+\infty} \phi(u)\Phi(\tau\sqrt{1+\delta^2} + \delta u + \sigma\delta t)du \\
&= \frac{\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)}{\Phi(\tau)} \mathbb{E}\left(\Phi(\tau\sqrt{1+\delta^2} + \delta U + \sigma\delta t)\right). \\
&= \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)\frac{\Phi(\sigma\lambda t + \tau)}{\Phi(\tau)}
\end{aligned}
$$

with $\lambda = \frac{\delta}{\sqrt{1+\delta^2}}$.
The similarity of the extended skew-normal and the skew-normal moment generating functions implies that many other properties proceed in a similar manner for the two families.

2.2 Skew-t family

A second example of a skew-symmetric family is the skew-t family introduces by Azzalini and Capitanio (2003) [12]. The density function takes the form

$$
t(z;\delta,\nu) = 2t(z;\nu)T\left(\delta z\sqrt{\frac{\nu+1}{\nu+z^2}};\nu+1\right), \qquad -\infty < z < +\infty, \tag{2.2.1}
$$

where t and T denote the standard Student-t density function and distribution function, respectively, and ν stands for the degrees of freedom.
Just as in Section 2.1 of this chapter, we can consider a continuous random variable Z with density function (2.2.1). Again we have that the variable $Y = \mu + \sigma Z$ with $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_0^+$, is a skew-t variable with density function at $x \in \mathbb{R}$

$$
2\sigma^{-1}t(\sigma^{-1}(x-\mu);\nu)T\left(\delta\sigma^{-1}(x-\mu)\sqrt{\frac{\nu+1}{\nu+\sigma^{-2}(x-\mu)^2}};\nu+1\right). \tag{2.2.2}
$$

The skew-t distribution is denoted by

$$
Y \sim ST(\mu,\sigma,\delta,\nu).
$$

When $\delta = 0$, (2.2.2) is reduced to the standard Student t-distribution with ν degrees of freedom. A special case of the skew-t distribution is the skew-normal distribution, obtained as $\nu \to \infty$. Figure 2.2.1 shows some graphs of the skew-t density functions for several values of δ.

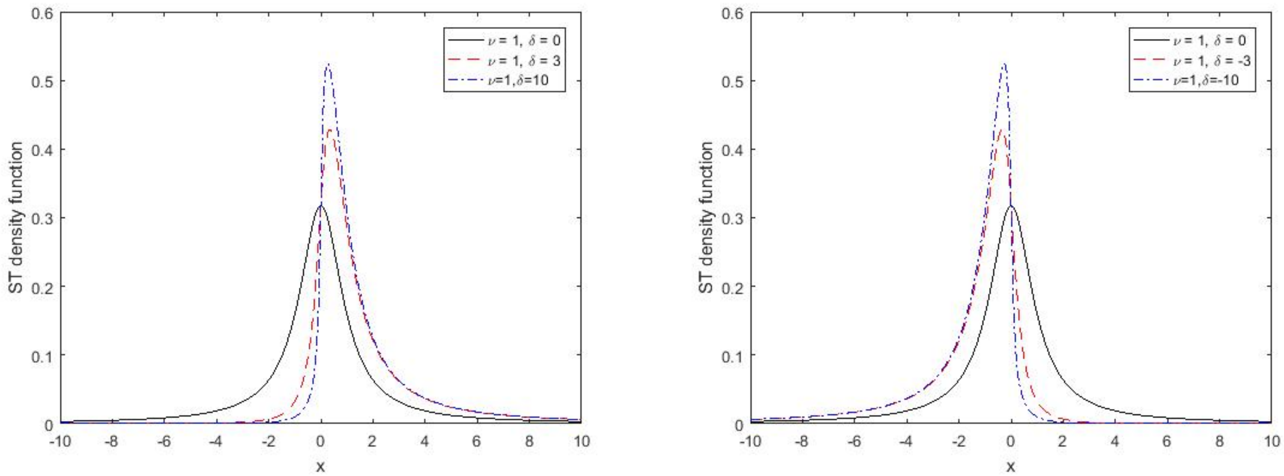

Figure 2.2.1: Skew-t density functions for varying δ and $\nu = 1$.

2.2.1 Properties

Suppose $Z \sim ST(0, 1, \delta, \nu)$. We can find a property for the skew-t family similar to property 2.1.1 for the skew-normal family.

Property 2.2.1. *$Z^2 \sim F_{1,\nu}$, with $F_{1,\nu}$ the F-distribution with parameters 1 and ν.*

Proof. The proof is analogous to the proof of property 2.1.1.

$$\begin{aligned}
P(|Z| \leq z) &= \int_{-z}^{z} 2t(u;\nu)T(\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du \\
&= \int_{0}^{z} 2t(u;\nu)T(\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du + \int_{-z}^{0} 2t(u;\nu)T(\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du \\
&= \int_{0}^{z} 2t(u;\nu)T(\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du - \int_{z}^{0} 2t(-u;\nu)T(-\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du \\
&= \int_{0}^{z} 2t(u;\nu)T(\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du + \int_{0}^{z} 2t(u;\nu)T(-\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)du \\
&= \int_{0}^{z} 2t(u;\nu)\Big(T(\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1) + T(-\delta u\sqrt{\frac{\nu+1}{\nu+u^2}};\nu+1)\Big)du \\
&= \int_{0}^{z} 2t(u;\nu)du \\
&= P(|X| \leq z)
\end{aligned}$$

with $X \sim T(0, 1, \nu)$.
We find that $|Z|$ and $|X|$ are identically distributed. So Z^2 and X^2 will be identically distributed, they both follow the distribution of $F_{1,\nu}$. □

2.2.2 Moments

Let $Z_1 \sim N(0,1)$ and $U \sim \chi^2_\nu$. If Z_1 and U are independent we can construct the t distribution via

$$\frac{Z_1}{\sqrt{\frac{U}{\nu}}}$$

with the degrees of freedom equal to ν.
For the skew-t distribution we can replace the normal variate above by a skew-normal one, Z. Thus we can define the skew-t random variable as follows

$$Y = \frac{Z}{\sqrt{\frac{U}{\nu}}}$$

with $Z \sim SN(0,1,\delta)$ and $U \sim \chi^2_\nu$, Z and U independent. We write $Y \sim ST(0,1,\delta,\nu)$.
The n^{th} moment of Y is given by

$$\mu_n = \mathbb{E}(Y^n) = \nu^{\frac{n}{2}}\mathbb{E}(Z^n)\mathbb{E}(U^{-\frac{n}{2}}). \tag{2.2.3}$$

as noted in Azzalini and Capitanio (2003) [12]. This follows from the fact that the expected value of a product of independent random variables is the product of their expected values. We already know how to calculate the moments of the skew-normal variable Z from Section 1.1.2, so we just need an expression for the n^{th} moment of $U^{-\frac{1}{2}}$.

Lemma 2.2.1. *Let $U \sim \chi^2_\nu$. The n^{th} moment of $U^{-\frac{n}{2}}$ is given by*

$$\mathbb{E}(U^{-\frac{n}{2}}) = \frac{\Gamma(\frac{\nu-n}{2})}{\Gamma(\frac{\nu}{2}).2^{\frac{n}{2}}}, \qquad \textit{where } \nu > n.$$

Proof. The probability density function of the χ^2_ν-distribution is

$$f(x,\nu) = \begin{cases} \frac{x^{(\frac{\nu}{2}-1)}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

We have

$$\begin{aligned} \mathbb{E}(U^{-\frac{n}{2}}) &= \int_0^{+\infty} y^{-\frac{n}{2}} \frac{y^{(\frac{\nu}{2}-1)}e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} dy \\ &= \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^{+\infty} y^{(\frac{\nu-n}{2}-1)}e^{-\frac{y}{2}} dy \\ &= \frac{\Gamma(\frac{\nu-n}{2})2^{\frac{\nu-n}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^{+\infty} \frac{y^{(\frac{\nu-n}{2}-1)}e^{-\frac{y}{2}}}{\Gamma(\frac{\nu-n}{2})2^{\frac{\nu-n}{2}}} dy \\ &= \frac{\Gamma(\frac{\nu-n}{2})}{\Gamma(\frac{\nu}{2}).2^{\frac{n}{2}}}. \end{aligned}$$

□

We can now calculate the moments of $Y \sim ST(\mu, \sigma, \delta, \nu)$ using (2.2.3) with $Z \sim SN(\mu, \sigma^2, \delta)$:

$$\begin{aligned}\mu_1 = \mathbb{E}(Y) &= \sqrt{\nu}(\mu + b\lambda\sigma)\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2}).\sqrt{2}} \\ &= (\mu + b\lambda\sigma)M.\end{aligned}$$

with $M = \sqrt{\frac{\nu}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$. We can see that this first moment depends on all four parameters and exists if and only if $\nu > 1$. From (1.1.6) and (1.1.7) we get

$$\begin{aligned}\mu_2 &= \nu\big(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\big)\frac{\Gamma(\frac{\nu-2}{2})}{\Gamma(\frac{\nu}{2}).2} \\ &= \nu\big(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\big)\frac{\Gamma(\frac{\nu}{2}-1)}{(\frac{\nu}{2}-1)\Gamma(\frac{\nu}{2}-1).2} \\ &= \frac{\nu}{\nu-2}\big(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\big).\end{aligned}$$

From the expressions for μ_1 and μ_2 we can now compute the variance

$$\begin{aligned}Var(Y) &= \mu_2 - \mu_1^2 \\ &= \frac{\nu}{\nu-2}\big(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda\big) - (\mu + b\lambda\sigma)M.\end{aligned}$$

We can also calculate the third and the fourth moment of Y

$$\begin{aligned}\mathbb{E}(Y^3) &= \nu^{\frac{3}{2}}\mathbb{E}(Z^3)\frac{\Gamma(\frac{\nu-3}{2})}{\Gamma(\frac{\nu}{2})2\sqrt{2}} \\ &= \nu^{\frac{3}{2}}(\mu^3 + 3\sigma\mu^2\lambda b + 3\mu\sigma^2 + \sigma^3 b(3\lambda - \lambda^3))\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{\Gamma(\frac{\nu-3}{2})}{\Gamma(\frac{\nu-1}{2})2\sqrt{2}} \\ &= \nu^{\frac{3}{2}}(\mu^3 + 3\sigma\mu^2\lambda b + 3\mu\sigma^2 + \sigma^3 b(3\lambda - \lambda^3))\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{\Gamma(\frac{\nu-3}{2})}{(\frac{\nu-3}{2})\Gamma(\frac{\nu-3}{2})2\sqrt{2}} \\ &= \frac{\nu}{\nu-3}\sqrt{\frac{\nu}{2}}(\mu^3 + 3\sigma\mu^2\lambda b + 3\mu\sigma^2 + \sigma^3 b(3\lambda - \lambda^3))\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \\ &= \frac{\nu}{\nu-3}(\mu^3 + 3\sigma\mu^2\lambda b + 3\mu\sigma^2 + \sigma^3 b(3\lambda - \lambda^3))M,\end{aligned}$$

$$\begin{aligned}\mathbb{E}(Y^4) &= \nu^2\mathbb{E}(Z^4)\frac{\Gamma(\frac{\nu-4}{2})}{\Gamma(\frac{\nu}{2})4} \\ &= \nu^2(\mu^4 + 4\mu^3\sigma b\lambda + 6\mu^2\sigma^2 + 4\mu\sigma^3 b(3\lambda - \lambda^3) + \sigma^4)\frac{\Gamma(\frac{\nu}{2}-2)}{(\frac{\nu}{2}-1)(\frac{\nu}{2}-2)\Gamma(\frac{\nu}{2}-2)4} \\ &= \frac{\nu^2}{(\nu-2)(\nu-4)}(\mu^4 + 4\mu^3\sigma b\lambda + 6\mu^2\sigma^2 + 4\mu\sigma^3 b(3\lambda - \lambda^3) + \sigma^4).\end{aligned}$$

We can now find the expressions for skewness and kurtosis.

$$\gamma_1(Y) = \frac{\mathbb{E}(Y^3)}{(\mathbb{E}(Y^2))^{\frac{3}{2}}} = \frac{\frac{\nu}{(\nu-3)}(\mu^3 + 3\sigma\mu^2\lambda b + 3\mu\sigma^2 + \sigma^3 b(3\lambda - \lambda^3))M}{\frac{\nu^{\frac{3}{2}}}{(\nu-2)^{\frac{3}{2}}}\left(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\right)^{\frac{3}{2}}}$$

$$= \frac{(\nu-2)^{\frac{3}{2}}}{\sqrt{\nu}(\nu-3)} \frac{(\mu^3 + 3\sigma\mu^2\lambda b + 3\mu\sigma^2 + \sigma^3 b(3\lambda - \lambda^3))M}{\left(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\right)^{\frac{3}{2}}},$$

$$\gamma_2(Y) = \frac{\mathbb{E}(Y^4)}{(\mathbb{E}(Y^2))^2} = \frac{\frac{\nu^2}{(\nu-2)(\nu-4)}(\mu^4 + 4\mu^3\sigma b\lambda + 6\mu^2\sigma^2 + 4\mu\sigma^3 b(3\lambda - \lambda^3) + \sigma^4)}{\frac{\nu^2}{(\nu-2)^2}\left(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\right)^2}$$

$$= \frac{(\nu-2)}{(\nu-4)} \frac{(\mu^4 + 4\mu^3\sigma b\lambda + 6\mu^2\sigma^2 + 4\mu\sigma^3 b(3\lambda - \lambda^3) + \sigma^4)}{\left(\mu^2 + \sigma^2 + 2b\mu\sigma\lambda)\right)^4}.$$

Chapter 3

Singularity problem of skew-symmetric distributions

It has been known for some time , since Azzalini (1985) [7] that many skew-symmetric distributions suffer from a Fisher information singularity problem at $\delta = 0$. More specifically, the Fisher information matrix associated with (2.0.1) is singular when coming close to attaining symmetry, i.e. at $\delta = 0$.
It has been shown that this singularity comes from an incompatibility between f and Π, which will be explained in more detail later on in this chapter.
As a result of a singular Fisher information matrix, the consistency rates in the estimation of the skewness parameter (at $\delta = 0$) will be slower than the usual $\sqrt{n}$. Comparably, tests of the null hypothesis of symmetry ($\delta = 0$) will also have slower rates. Therefore, the standard assumptions for root-n asymptotic inference are not met. The rate for "simple singularity" would typically be $\sqrt[4]{n}$. But for example with the skew-normal distributions, this rate drops to $\sqrt[6]{n}$ as we will see in Section 3.1. This is explained by a characteristic of the skew-normal distribution called a "double singularity". This will be discussed further in the Section 2.1.2. In case of "triple singularity" this $\sqrt[6]{n}$-rate can go down to a $\sqrt[8]{n}$ rate. It has been proven by Hallin and Ley (2014) [39] that this is the lowest rate possible.
This singularity problem has been discussed in a lot of papers. In this chapter, we will review the examples of the skew-normal distributions and the skew-t distributions who suffer from the Fisher information singularity problem. We will look at the origin of this singularity in the different skew-symmetric distributions and how this singularity can be overcome using a number of different parametrizations.

3.1 Skew-normal family

We will start again by looking at the skew-normal family and once more using the outline of Azzalini [7]. The log-likelihood function is given by

$$\begin{aligned}\mathscr{L}(\theta^{\mathrm{DP}}; x) &= \log(\sigma^{-1}\phi(\sigma^{-1}(x-\mu); \delta)) \\ &= -\log(\sigma) + \log(\phi(\sigma^{-1}(x-\mu))) + \log(2\Phi(\delta\sigma^{-1}(x-\mu)))\end{aligned}$$

$$= -\log(\sigma) + \log(e^{-\sigma^{-2}\frac{(x-\mu)^2}{2}}) + \log(2\Phi(\delta\sigma^{-1}(x-\mu)))$$

$$= -\log(\sigma) - \sigma^{-2}\frac{(x-\mu)^2}{2} + \zeta_0(\delta\sigma^{-1}(x-\mu)) \tag{1.6}$$

with $\theta^{\text{DP}} = (\mu, \sigma, \delta)'$ and $\zeta_0(x) = \log(2\Phi(x))$. The superscript 'DP' stands for direct parameters because we can read these parameters directly from the expression of the density function. The components of the score vector are

$$l^1_{\theta^{\text{DP}}} = \frac{\partial \mathscr{L}}{\partial \mu} = \sigma^{-2}(x-\mu) - \sigma^{-1}\delta\zeta_0'(\delta\sigma^{-1}(x-\mu))$$

$$= \sigma^{-1}z - \sigma^{-1}\delta\zeta_1(\delta z);$$

$$l^2_{\theta^{\text{DP}}} = \frac{\partial \mathscr{L}}{\partial \sigma} = -\sigma^{-1} + \sigma^{-3}(x-\mu)^2 - \sigma^{-2}(x-\mu)\delta\zeta_0'(\delta\sigma^{-1}(x-\mu))$$

$$= -\sigma^{-1} + \sigma^{-1}z^2 - \sigma^{-1}\delta\zeta_1(\delta z)z; \tag{1.7}$$

$$l^3_{\theta^{\text{DP}}} = \frac{\partial \mathscr{L}}{\partial \delta} = \sigma^{-1}(x-\mu)\zeta_0'(\delta\sigma^{-1}(x-\mu))$$

$$= z\zeta_1(\delta z)$$

with $z = \sigma^{-1}(x-\mu)$ and $\zeta_r(x) = \frac{d^r}{dx^r}\zeta_0(x) \quad (r = 1, 2, \ldots)$. In order to derive the Fisher information matrix, we differentiate the score vector. This leads to

$$\frac{\partial^2 \mathscr{L}}{\partial \mu^2} = \frac{\partial}{\partial \mu}\Big(\sigma^{-1}z - \sigma^{-1}\delta\zeta_1(\delta z)\Big)$$

$$= -\sigma^{-2} + \sigma^{-2}\delta^2\zeta_2(\delta z),$$

$$\frac{\partial^2 \mathscr{L}}{\partial \sigma \partial \mu} = \frac{\partial}{\partial \sigma}\Big(\sigma^{-1}z - \sigma^{-1}\delta\zeta_1(\delta z)\Big)$$

$$= -\sigma^{-2}z - \sigma^{-3}(x-\mu) + \sigma^{-2}\delta\zeta_1(\delta z) + \delta^2\sigma^{-3}(x-\mu)\zeta_2(\delta z)$$

$$= -2\sigma^{-2}z + \sigma^{-2}\delta\zeta_1(\delta z) + \delta^2\sigma^{-2}z\zeta_2(\delta z),$$

$$\frac{\partial^2 \mathscr{L}}{\partial \delta \partial \mu} = \frac{\partial}{\partial \delta}\Big(\sigma^{-1} z - \sigma^{-1}\delta\zeta_1(\delta z)\Big)$$

$$= -\sigma^{-1}\zeta_1(\delta z) - \sigma^{-1}\delta z \zeta_2(\delta z),$$

$$\frac{\partial^2 \mathscr{L}}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\Big(-\sigma^{-1} + \sigma^{-1} z^2 - \sigma^{-1}\delta\zeta_1(\delta z) z\Big)$$

$$= \sigma^{-2} - \sigma^{-2} z^2 - 2\sigma^{-4}(x-\mu)^2 + \sigma^{-2}\delta\zeta_1(\delta z) z + \sigma^{-3}\delta(x-\mu)\zeta_1(\delta z) + \sigma^{-3}\delta^2(x-\mu) z \zeta_2(\delta z)$$

$$= \sigma^{-2} - 3\sigma^{-2} z^2 + 2\sigma^{-2}\delta z \zeta_1(\delta z) + \sigma^{-2}\delta^2 z^2 \zeta_2(\delta z),$$

$$\frac{\partial^2 \mathscr{L}}{\partial \delta \partial \sigma} = \frac{\partial}{\partial \delta}\Big(-\sigma^{-1} + \sigma^{-1} z^2 - \sigma^{-1}\delta\zeta_1(\delta z) z\Big)$$

$$= -\sigma^{-1}\zeta_1(\delta z) z - \sigma^{-1}\delta\zeta_2(\delta z) z^2$$

and

$$\frac{\partial^2 \mathscr{L}}{\partial \delta^2} = \frac{\partial}{\partial \delta}\Big(z\zeta_1(\delta z)\Big)$$

$$= z^2\zeta_2(\delta z).$$

We can now compute the elements of the Fisher informaton matrix. Calculating the mean value of the second derivatives above requires expectations of some expressions in Z. Some of these terms are easy to work out :

$$\mathbb{E}\big(Z^k\zeta_1(\delta Z)\big) = \int_{-\infty}^{+\infty} z^k \frac{\phi(\delta z)}{\Phi(\delta z)} 2\phi(z)\Phi(\delta z) dz$$
$$= \frac{2}{2\pi}\int_{-\infty}^{+\infty} z^k e^{-\frac{z^2(\delta^2+1)}{2}} dz$$

subst. : $\boldsymbol{u = z\sqrt{\delta^2+1}}$

$$= \frac{2}{2\pi\sqrt{\delta^2+1}}\int_{-\infty}^{+\infty} \frac{u^k}{(\delta^2+1)^{\frac{k}{2}}} e^{\frac{-u^2}{2}} du$$
$$= \frac{b}{(\delta^2+1)^{\frac{k+1}{2}}}\mathbb{E}(U^k).$$

So we need the k^{th} moment of a standard normal variable U. If k is odd then $\mathbb{E}\left(U^k\right) = 0$. When k is even we can obtain an expression for the k^{th} moment of U by applying partial integration.

$$\begin{aligned}
\mathbb{E}(U^k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^k e^{\frac{-u^2}{2}} du \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^{k-1}(u e^{\frac{-u^2}{2}}) du \\
&= \frac{1}{\sqrt{2\pi}} \left(\left[-u^{k-1} e^{\frac{-u^2}{2}} \right]_{-\infty}^{+\infty} + (k-1) \int_{-\infty}^{+\infty} u^{k-2} e^{\frac{-u^2}{2}} du \right) \\
&= \frac{k-1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^{k-2} e^{\frac{-u^2}{2}} du \\
&= (k-1)\mathbb{E}\left(U^{k-2}\right).
\end{aligned}$$

Since $\mathbb{E}(U^0) = 1$, we get the following recursive expression

$$\mathbb{E}(U^k) = (k-1).(k-3)...3.1.$$

In conclusion, we obtain

$$\mathbb{E}\left(Z^k \zeta_1(\delta Z)\right) = \frac{b}{(\delta^2+1)^{\frac{k+1}{2}}} \mathbb{E}(U^k) = \begin{cases} \frac{b}{(\delta^2+1)^{\frac{k+1}{2}}} ((k-1).(k-3)...3.1) & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}. \tag{3.1.1}$$

Other terms are not so manageable such as

$$a_k = a_k(\delta) = \mathbb{E}\left(Z^k \zeta_1^2(\delta Z)\right).$$

Using these results we now calculate the elements of the Fisher information matrix.

$$\begin{aligned}
I_{1,1} &= -\mathbb{E}\Big(\frac{\partial^2 \mathcal{L}}{\partial \mu^2}\Big) \\
&= -\mathbb{E}(-\sigma^{-2} + \sigma^{-2}\delta^2\zeta_2(\delta z)) \\
&= \sigma^{-2} - \sigma^{-2}\delta^2\mathbb{E}(-\zeta_1^2(\delta z) - z\delta\zeta_1(\delta z)) \\
&= \sigma^{-2} + \sigma^{-2}\delta^2 a_0,
\end{aligned}$$

$$\begin{aligned}
I_{1,2} = I_{2,1} &= -\mathbb{E}\Big(\frac{\partial^2 \mathcal{L}}{\partial \sigma \partial \mu}\Big) \\
&= -\mathbb{E}(-2\sigma^{-2}z + \sigma^{-2}\delta\zeta_1(\delta z) + \delta^2\sigma^{-2}z\zeta_2(\delta z))
\end{aligned}$$

$$= 2\sigma^{-2}\mathbb{E}(z) - \sigma^{-2}\delta\mathbb{E}(\zeta_1(\delta z)) - \delta^2\sigma^{-2}\mathbb{E}(-z\zeta_1^2(\delta z) - z^2\delta\zeta_1(\delta z))$$

$$= \frac{2\delta b}{\sigma^2\sqrt{1+\delta^2}} - \frac{\delta b}{\sigma^2\sqrt{1+\delta^2}} + \delta^2\sigma^{-2}a_1 + \frac{\delta^3 b}{\sigma^2(1+\delta^2)^{\frac{3}{2}}}$$

$$= \frac{\delta b(1+2\delta^2)}{\sigma^2(1+\delta^2)^{\frac{3}{2}}} + \delta^2\sigma^{-2}a_1,$$

$$I_{1,3} = I_{3,1} = -\mathbb{E}\Big(\frac{\partial^2\mathscr{L}}{\partial\delta\partial\mu}\Big)$$

$$= -\mathbb{E}(-\sigma^{-1}\zeta_1(\delta z) - \sigma^{-1}\delta z\zeta_2(\delta z))$$

$$= \frac{b}{\sigma\sqrt{1+\delta^2}} + \sigma^{-1}\delta\mathbb{E}(-z\zeta_1^2(\delta z) - z^2\delta\zeta_1(\delta z))$$

$$= \frac{b}{\sigma\sqrt{1+\delta^2}} - \sigma^{-1}\delta a_1 - \frac{\delta^2 b}{\sigma(1+\delta^2)^{\frac{3}{2}}}$$

$$= \frac{b}{\sigma(1+\delta^2)^{\frac{3}{2}}} - \sigma^{-1}\delta a_1,$$

$$I_{2,2} = -\mathbb{E}\Big(\frac{\partial^2\mathscr{L}}{\partial\sigma^2}\Big)$$

$$= -\mathbb{E}(\sigma^{-2} - 3\sigma^{-2}z^2 + 2\sigma^{-2}\delta z\zeta_1(\delta z) + \sigma^{-2}\delta^2 z^2\zeta_2(\delta z))$$

$$= -\sigma^{-2} + 3\sigma^{-2}\mathbb{E}(z^2) - 2\sigma^{-2}\delta\mathbb{E}(z\zeta_1(\delta z)) - \sigma^{-2}\delta^2\mathbb{E}(-z^2\zeta_1^2(\delta z) - z^3\delta\zeta_1(\delta z))$$

$$= 2\sigma^{-2} + \sigma^{-2}\delta^2 a_2,$$

$$I_{2,3} = I_{3,2} = -\mathbb{E}\Big(\frac{\partial^2\mathscr{L}}{\partial\delta\partial\sigma}\Big)$$

$$= -\mathbb{E}(-\sigma^{-1}\zeta_1(\delta z)z - \sigma^{-1}\delta\zeta_2(\delta z)z^2)$$

$$= \sigma^{-1}\mathbb{E}(z\zeta_1(\delta z)) + \sigma^{-1}\delta\mathbb{E}(-z^2\zeta_1^2(\delta z) - z^3\delta\zeta_1(\delta z))$$

$$= -\sigma^{-1}\delta a_2,$$

$$
\begin{aligned}
I_{3,3} &= -\mathbb{E}\Big(\frac{\partial^2 \mathscr{L}}{\partial \delta^2}\Big) \\
&= -\mathbb{E}(z^2 \zeta_2(\delta z)) \\
&= -\mathbb{E}(-z^2 \zeta_1^2(\delta z) - z^3 \delta \zeta_1(\delta z)) \\
&= a_2.
\end{aligned}
$$

The resulting Fisher information matrix takes the form

$$
I^{\mathrm{DP}}(\theta^{\mathrm{DP}}) = \begin{pmatrix} \sigma^{-2} + \sigma^{-2}\delta^2 a_0 & * & * \\ \frac{\delta b(1+2\delta^2)}{\sigma^2(1+\delta^2)^{\frac{3}{2}}} + \delta^2 \sigma^{-2} a_1 & 2\sigma^{-2} + \sigma^{-2}\delta^2 a_2 & * \\ \frac{b}{\sigma(1+\delta^2)^{\frac{3}{2}}} - \sigma^{-1}\delta a_1 & -\sigma^{-1}\delta a_2 & a_2 \end{pmatrix}
$$

where the upper triangle can be uptained by symmetry. At $(\mu, \sigma, 0)' = \theta_0$, the Fisher information matrix becomes

$$
I^{\mathrm{DP}}(\theta_0) = \begin{pmatrix} \sigma^{-2} & 0 & \frac{b}{\sigma} \\ 0 & 2\sigma^{-2} & 0 \\ \frac{b}{\sigma} & 0 & b^2 \end{pmatrix}
$$

where $I_{3,3}^{\mathrm{DP}}(\theta_0)$ comes from

$$
a_2|_{\theta_0} = \mathbb{E}(z^2 \zeta_1^2(0)) = \mathbb{E}(z^2 b^2) = b^2.
$$

We calculate the determinant of $I^{\mathrm{DP}}(\theta_0)$ as follows :

$$
\begin{aligned}
det(I^{\mathrm{DP}}(\theta_0)) &= \begin{vmatrix} \sigma^{-2} & 0 & \frac{b}{\sigma} \\ 0 & 2\sigma^{-2} & 0 \\ \frac{b}{\sigma} & 0 & b^2 \end{vmatrix} \\
&= 2\sigma^{-4} b^2 - \frac{b^2}{\sigma^2} 2\sigma^{-2} \\
&= 0.
\end{aligned}
$$

The skew-normal distribution thus suffers from a Fisher information singularity problem at $\delta = 0$. We can see that this Fisher singularity is caused by the collinearity of l^1 and l^3 at $\delta = 0$. In particular, we get $l^1_{\theta_0} = \frac{z}{\sigma}$ and $l^3_{\theta_0} = \delta z$, from which it then follows $\delta \sigma l^1_{\theta_0} = l^3_{\theta_0}$, so the first and the third components of the score vector are in fact proportional to each other.

We will now look at the estimates of the parameters to get an idea about the slower convergence rates. So we will now estimate the parameters using the method of moments.

The moments of the skew-normal distribution as we have obtained in Section 2.1.2, are given by

$$\mathbb{E}(Y) = \mu + b\lambda\sigma,$$

$$\operatorname{Var}(Y) = \sigma^2(1 - b^2\lambda^2),$$

$$\begin{aligned}\gamma_1 &= \frac{\lambda^3}{(1-b^2\lambda^2)^{\frac{3}{2}}}\left(2b^3 - b\right)\\ &= \frac{\delta^3}{(1+(1-b^2)\delta^2)^{3/2}}\left(2b^3 - b\right)\end{aligned}$$

with $\lambda = \frac{\delta}{\sqrt{1+\delta^2}}$.
Replacing γ_1 by $\frac{\hat{m}_3}{s^3}$, with s^2 the sample variance, we can obtain the estimates for the different parameters. The moment estimators are given by

$$\hat{\mu} = \bar{y} - b\left(\frac{\hat{m}_3}{2b^3 - b}\right)^{\frac{1}{3}},$$

$$\hat{\sigma}^2 = s^2 + b^2\left(\frac{\hat{m}_3}{2b^3 - b}\right)^{\frac{2}{3}},$$

$$\begin{aligned}\hat{\lambda} &= \left(\frac{\hat{m}_3}{\hat{\sigma}^3(2b^3 - b)}\right)^{\frac{1}{3}}\\ &= \left(\frac{\hat{m}_3}{(2b^3-b)}\right)^{\frac{1}{3}}\left(s^2 + b^2\left(\frac{\hat{m}_3}{2b^3 - b}\right)^{\frac{2}{3}}\right)^{-\frac{1}{2}}\\ &= \left(b + s^2\left(\frac{2b^3 - b}{\hat{m}_3}\right)^{\frac{2}{3}}\right)^{-\frac{1}{2}},\end{aligned}$$

$$\begin{aligned}\hat{\delta} &= \frac{\hat{\lambda}}{\sqrt{1-\hat{\lambda}^2}}\\ &= \left(b + s^2\left(\frac{2b^3 - b}{\hat{m}_3}\right)^{\frac{2}{3}} - 1\right)^{-\frac{1}{2}}\end{aligned}$$

where $\bar{y}$ is the sample mean, s^2 is the sample variance, and $m_3 = \frac{1}{n}\sum(y_i - \bar{y})^3$. Therefore, in the neighbourhood of zero, δ is proportional to the cubic root of the third standardized cumulant, i.e. the skewness index γ_1, so that $\hat{\delta} = O_p\left(n^{-\frac{1}{6}}\right)$ because $\gamma_1 = O_p\left(n^{-\frac{1}{2}}\right)$.

This conjecture is confirmed by the result obtained by Rotnitzky *et al.* (2000) [59]. Theorem 3 of Rotnizky *et al.* presumes numerous assumptions for which we will first give some notations used by Rotnitzky *et al*. We consider a $p \times 1$ parameter vector $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. $S_j(\theta)$ denotes the score with respect to θ_j and S_j denotes $S_j(\theta^*)$ with θ^* a point where the information matrix is singular. We asssume that $Y_1, Y_2, \dots, Y_n$ are n independent copies of a random variable Y with density $f(y;\theta^*)$. Let $l(y;\theta)$ denote $\log f(y;\theta)$ and let $l^{(r)}(y;\theta)$ denote $\partial^r \log f(y;\theta)/\partial^{r_1}\theta_1\partial^{r_2}\theta_2\dots\partial^{r_p}\theta_p$. Write $L_n(\theta)$ for $\sum l(Y_i;\theta)$. Define $||\theta||^2$ as $\sum_{k=1}^p \theta_k^2$. And lastly let $S_j^{(s+j)}$ denote $\partial^{s+j} l(Y;\theta)/\partial\theta_1^{s+j}|_{\theta^*}$. Rotnitzky *et al* then assume the following regularity conditions :

1. $\theta^* = (\mu^*, \sigma^*, \delta^*)$ takes its value in a compact subset Θ of $\mathbb{R}^p$ that contains an open neighbourhood $\mathcal{N}$ of θ^*.
2. Distinct values of θ in Θ correspond to distinct probability distributions.
3. $\mathbb{E}\left(\sup_{\theta\in\Theta}|l(Y;\theta)|\right) < \infty$.
4. With probability 1, the derivative $l^{(r)}(Y;\theta)$ exists for all θ in $\mathcal{N}$ and $r \leq 2s+1$ and satifies $\mathbb{E}\left(\sup_{\theta\in\Theta}|l^{(r)}(Y;\theta)|\right) < \infty$. Furthermore, with probability 1 under θ^*, $f(Y;\theta) > 0$ for all θ in $\mathcal{N}$.
5. For $s \leq r \leq 2s+1$, $\mathbb{E}\left(\left(l^{(r)}(Y;\theta^*)\right)^2\right) < \infty$.
6. When $r = 2s+1$ there exists $\epsilon > 0$ and some function $g(Y)$ satisfying $\mathbb{E}\left(g(Y)^2\right) < \infty$ such that for θ and θ' in $\mathcal{N}$, with probability 1,
$$||L_n^{(r)}(\theta) - L_n^{(r)}(\theta')|| \leq ||\theta - \theta'||^\epsilon \sum g(Y_i). \tag{3.1.2}$$
7. The conditions '$S_2, \dots, S_p$ are linearly independent' and '$S_1 = K(S_2, \dots, S_p)^{\mathrm{T}}$' hold with probability 1 for some $1 \times (p-1)$ constant vector K.
8. With probability 1, $\left.\frac{\partial^j l(Y;\theta)}{\partial\theta_1^j}\right|_{\theta^*} = 0, 1 \leq j \leq s-1$.
9. For all $1 \times (p-1)$ vectors K, $S_1^{(s)} \neq K(S_2, \dots, S_p)^{\mathrm{T}}$ with positive probability.
10. If s is even, then for all $1 \times p$ vectors K', $S_1^{(s+1)} \neq K'(S_1^{(s)}, S_2, \dots, S_p)^{\mathrm{T}}$ with positive probability.

The theorem itself[1] then goes as follows

Theorem. *Under these assumptions, when s is odd*

(a) the MLE $\hat{\delta}$ of δ exists when $\delta = \delta^*$, it is unique with a probability tending to 1, and it is a consistent estimator when $\delta = \delta^*$;

(b)
$$\begin{bmatrix} n^{1/(2s)}(\hat{\delta}_1 - \delta_1^*) \\ n^{1/2}(\hat{\delta}_2 - \delta_2^*) \\ \vdots \\ n^{1/2}(\hat{\delta}_p - \delta_p^*) \end{bmatrix} \rightsquigarrow \begin{bmatrix} Z_1^{1/s} \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix},$$

[1] for the proof we refer to Rotnitzky *et al.* (2000) [59]

where $Z = (Z_1, Z_2, \ldots, Z_p)^T$ denotes a mean-zero normal random vector with variance equal to I^{-1}, the inverse of the covariance matrix of $(S_1^{(s)}/s!, S_2, \ldots, S_p)$.

We will use their Theorem 3 to prove Proposition 1, given by Chiogna (2005) [21]. This proof uses the iterative reparametrization used by Rotnitzky *et al.* (2000) [59] until the conditions 9 and 10 are satisfied. This iterative reparametrization is based on orthogonalization of parameters like in Cox and Reid (1987) [22]. Before we give the proposition, we will give some notations used.
We shall indicate the parameter component $(\mu, \sigma)^{\mathrm{T}}$ with χ. Moreover, let $u(\chi, \delta) = (u_\chi(\chi, \delta)^{\mathrm{T}}, u_\delta(\chi, \delta)^{\mathrm{T}})$ denote the score vector for $\theta = (\mu, \sigma, \delta)'$. The expected information matrix will be indicated by $i(\chi, \delta)$ and the observed information matrix by $j(\chi, \delta)$.

Proposition 1. *The random vector*

$$\left(n^{1/2}(\hat{\mu} - \mu^* + b\hat{\sigma}\hat{\delta}), n^{1/2}(\hat{\sigma} - \sigma^* + \frac{1}{2}b^2\hat{\sigma}\hat{\delta}^2), n^{1/6}\hat{\delta}\right)$$

converges under $(\mu, \sigma, \delta)' = (\mu^*, \sigma^*, 0)'$ to $(Z_1, Z_2, Z_3^{1/3})$, with (Z_1, Z_2, Z_3) as in the Theorem of Rotnitzky et al.

Proof. As the first and higher order partial derivatives of the log-likelihood with respect to δ are not zero in $\delta = 0$, we will need to apply the iterative reparametrization procedure of Rotnizky *et al.* to satisfy conditions 9 and 10 so we can apply Theorem 3 of Rotnizky *et al.* (2000) [59]. By looking at the score vector $u(\chi^*, \delta^*)$ for one observation z:

$$u(\chi^*, \delta^*) = \left(\frac{z}{\sigma^*}, \frac{z^2 - 1}{\sigma^*}, bz\right)',$$

with $b = \sqrt{\frac{2}{\pi}}$, we note that $u_\delta(\chi^*, \delta^*) = K u_\chi(\chi^*, \delta^*)$, with $K = (b\sigma^*, 0)$. Therefore, the following reparametrization applies:

$$\theta_{\mathrm{I}} = \theta + (K, 0)'\delta = (\chi_{\mathrm{I}}^{\mathrm{T}}, \delta_{\mathrm{I}})$$

so that $\chi_{\mathrm{I}} = (\mu + \sigma^* b\delta, \sigma)'$ and $\delta_{\mathrm{I}} = \delta$. We will now check the second derivative with respect to δ in the log-likelihood parameterized by θ_{I}. We observe for one individual that

$$\begin{aligned}
j_{\delta\delta}^{\theta_{\mathrm{I}}}(\chi^*, \delta^*) &= \frac{\partial^2}{\partial \delta^2}\left(-\log(\sigma) - \sigma^{-2}\frac{(x - \mu_{\mathrm{I}} + \sigma^* b\delta)^2}{2} + \zeta_0(\delta\sigma^{-1}(x - \mu_{\mathrm{I}} + \sigma^* b\delta)\right)\Big|_{(\chi^*,\delta^*)} \\
&= \frac{\partial}{\partial \delta}\left(-\sigma^{-2}\sigma^* b(x - \mu_{\mathrm{I}} + \sigma^* b\delta) + (\sigma^{-1}(x - \mu_{\mathrm{I}} + 2\sigma^* b\delta)\zeta_1(\delta\sigma^{-1}(x - \mu_{\mathrm{I}} + \sigma^* b\delta)\right)\Big|_{(\chi^*,\delta^*)} \\
&= \left(-\sigma^{-2}\sigma^{*2}b^2 + 2\sigma^{-1}\sigma^* b\zeta_1(\delta\sigma^{-1}(x - \mu_{\mathrm{I}} + \sigma^* b\delta) + (\sigma^{-1}(x - \mu_{\mathrm{I}} + 2\sigma^* b\delta)^2\zeta_2(\delta\sigma^{-1}(x - \mu_{\mathrm{I}} + \sigma^* b\delta)\right)\Big|_{(\chi^*,\delta^*)} \\
&= -b^2 + 2b^2 - z^2 b^2 \\
&= K_1 u_\chi(\chi^*, \delta^*)
\end{aligned}$$

with $K_1 = (0, -\sigma^* b^2)'$. Therefore we carry out the second step in the iterative reparametrization, i.e.

$$\theta_{\text{II}} = \theta + (K,0)'\delta + (1/2K_1,0)'\delta^2,$$

so that $\chi_{\text{II}} = (\mu + \sigma^* b\delta, \sigma - \frac{1}{2}\sigma^* b^2\delta^2)$. The third partial derivative with respect to δ in the log-likelihood newly parameterized by θ_{II} is now neither zero nor a linear combination of the components of $u_\chi(\chi^*,\delta^*)$, the derivative for one individual being when setting $y = \left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-1}(x - \mu_{\text{II}} + \sigma^* b\delta)$ and $y' = \frac{\partial y}{\partial \delta}$

$$\begin{aligned}
\frac{\partial}{\partial\delta} j^{\theta_{\text{II}}}_{\delta\delta}(\chi^*,\delta^*) &= \frac{\partial^3}{\partial\delta^3}\left(-\log\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right) - \frac{y^2}{2} + \zeta_0(\delta y)\right)\Big|_{(\chi^*,\delta^*)} \\
&= \frac{\partial^2}{\partial\delta^2}\left(-\frac{\sigma^* b^2\delta}{\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2} - yy' + (y + \delta y')\zeta_1(\delta y)\right)\Big|_{(\chi^*,\delta^*)} \\
&= \frac{\partial}{\partial\delta}\left(\frac{2\sigma^* b^2(b^2\sigma^*\delta - 2\sigma)}{\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^2} - y'^2 + -yy'' + (2y' + \delta y'')\zeta_1(\delta y) + (y + \delta y')^2\zeta_2(\delta y)\right)\Big|_{(\chi^*,\delta^*)} \\
&= \Bigg(-\frac{4\sigma^{*2} b^4\delta(b^2\sigma^*\delta - 6\sigma)}{\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^3} - 3y'y'' - yy''' + (3y'' + \delta y''')\zeta_1(\delta y) \\
&\quad + 3(2y' + \delta y'')(y + \delta y')\zeta_2(\delta y) + (y + \delta y')^3\zeta_3(\delta y)\Bigg)\Big|_{(\chi^*,\delta^*)} \\
&= z^3(2b^3 - b) - 3b^3 z
\end{aligned}$$

Therefore, the iterative process stops and making use of Theorem 3 of Rotnitzky *et al.* (2000) [59] with $s = 3$, we can complete the proof. The expressions for y and its derivatives with respect to δ along with a more detailed elaboration can be found in the Appendix B. □

We will now look at some other reparametrizations to overcome the problem of singularity of the Fisher information matrix.

3.1.1 Centred parametrizaton

Due to this singularity problem, we are unable to use the direct parameters, which we can read directly from the expression from the density function, for making inferences. We introduce a reparametrization, suggested by Azzalini (1985) [7], intended to solve the singularity problem at $\delta = 0$. We rewrite Y as

$$Y = \xi + \omega Z_0, \qquad Z_0 = \frac{Z - \mu_Z}{\sigma_Z} \sim SN\left(-\frac{\mu_Z}{\sigma_Z}, \frac{1}{\sigma_Z^2}, \delta\right)$$

where $\xi = \mathbb{E}(Y)$ and $\omega^2 = \text{Var}(Y)$ are given by (2.1.4) and (2.1.5), respectively. Consider the centred parameters $\theta^{\text{CP}} = (\xi, \omega, \gamma_1)'$ instead of the DP parameters. These parameters are called centered because the reparametrization involves Z_0, which is centred around 0. Here γ_1 is the measure of skewness. We get the correspondance between DP and CP

$$\xi = \mu + b\sigma\frac{\delta}{\sqrt{1+\delta^2}} = \mu + b\sigma\mu_Z,$$

$$\omega = \sigma\left(1 - b^2\frac{\delta^2}{1+\delta^2}\right) = \sigma\sigma_Z,$$

$$\gamma_1 = \frac{4-\pi}{2}\frac{b^3\delta^3}{(1+(1-b^2)\delta^2)^{\frac{3}{2}}} = \frac{4-\pi}{2}\frac{\mu_Z^3}{\sigma_Z^3}$$

and the inverse mapping is given by

$$\mu = \xi - b\sigma\mu_Z = \xi - \omega\frac{\mu_Z}{\sigma_Z},$$

$$\sigma = \frac{\omega}{\sigma_Z},$$

$$\delta = \frac{R}{\sqrt{\frac{2}{\pi} - (1-\frac{2}{\pi})R^2}}$$

with $R = \frac{\mu_Z}{\sigma_Z} = \sqrt[3]{\frac{2\gamma_1}{4-\pi}}$. We now want to compute the Fisher information matrix for θ^{CP}. This can be obtained from the Fisher information matrix for θ^{DP}. Utilizing the chain rule we get

$$\begin{aligned} I^{\mathrm{CP}}(\theta^{\mathrm{CP}}) &= -E\left(\frac{\partial^2\mathscr{L}(\theta^{\mathrm{CP}};x)}{\partial\theta^{\mathrm{CP}\,2}}\right) \\ &= -E\left(\frac{\partial^2\mathscr{L}(\theta^{\mathrm{DP}};x)}{\partial\theta^{\mathrm{DP}\,2}}\left|\frac{\partial\theta^{\mathrm{DP}}}{\partial\theta^{\mathrm{CP}}}\right|^2\right). \end{aligned}$$

We get the formulae

$$I^{\mathrm{CP}}\left(\theta^{\mathrm{CP}}\right) = D^{\mathrm{T}}I^{\mathrm{DP}}(\theta^{\mathrm{DP}})D$$

where D is the Jacobian matrix

$$D = \left(\frac{\partial\theta^{\mathrm{DP}}}{\partial\theta^{\mathrm{CP}}}\right) = \begin{pmatrix} 1 & -\frac{\mu_Z}{\sigma_Z} & \frac{\partial\mu}{\partial\gamma_1} \\ 0 & \frac{1}{\sigma_Z} & \frac{\partial\sigma}{\partial\gamma_1} \\ 0 & 0 & \frac{\partial\delta}{\partial\gamma_1} \end{pmatrix}.$$

We calculate the elements of the last column of D. We can rewrite μ as a function of γ_1. We get

$$\mu = \xi - \omega\sqrt[3]{\frac{2\gamma_1}{4-\pi}}.$$

By deriving μ with respect to γ_1 we get

$$
\begin{aligned}
\frac{\partial \mu}{\partial \gamma_1} &= \frac{\partial}{\partial \gamma_1}\left(\xi - \omega \sqrt[3]{\frac{2\gamma_1}{4-\pi}}\right) \\
&= -\frac{\omega}{3}\left(\frac{2\gamma_1}{4-\pi}\right)^{-\frac{2}{3}} \frac{2}{4-\pi} \\
&= -\frac{\omega}{3}\frac{2\sigma_Z^2}{(4-\pi)\mu_Z^2} \\
&= -\frac{\omega}{3}\frac{2\sigma_Z^3}{(4-\pi)\mu_Z^3}\frac{\mu_Z}{\sigma_Z} \\
&= -\frac{\omega}{3\gamma_1}\frac{\mu_Z}{\sigma_Z}.
\end{aligned}
$$

We can do the same for σ and δ

$$
\begin{aligned}
\frac{\partial \sigma}{\partial \gamma_1} &= \frac{\partial}{\partial \gamma_1}\left(\frac{\omega}{\sigma_Z}\right) \\
&= -\frac{\omega}{\sigma_Z^2}\frac{\partial \sigma_Z}{\partial \gamma_1} \\
&= -\frac{\omega}{\sigma_Z^2}\frac{\partial \sigma_Z}{\partial \delta}\frac{\partial \delta}{\partial \gamma_1}
\end{aligned}
$$

$$
\begin{aligned}
\text{with } \frac{\partial \sigma_Z}{\partial \delta} &= \frac{\partial}{\partial \delta}\left(\sqrt{1-b^2\frac{\delta^2}{1+\delta^2}}\right) \\
&= -\frac{b^2}{2\sqrt{1-b^2\frac{\delta^2}{1+\delta^2}}}\frac{2\delta(1+\delta^2)-2\delta^3}{(1+\delta^2)^2} \\
&= -\frac{b^2}{\sigma_Z}\frac{\delta}{(1+\delta^2)^2} \\
&= -\frac{\mu_Z}{\sigma_Z}\frac{b}{(1+\delta^2)^{\frac{3}{2}}},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \delta}{\partial \gamma_1} &= \frac{\partial}{\partial \gamma_1}\left(\frac{R}{\sqrt{\frac{2}{\pi}-(1-\frac{2}{\pi})R^2}}\right) \\
&= \frac{\frac{\partial R}{\partial \gamma_1}T - \frac{R}{2}T^{-1}(-(1-\frac{2}{\pi})2R)\frac{\partial R}{\partial \gamma_1}}{T^2} \\
&= \frac{2}{3(4-\pi)}\frac{TR^{-2}+(1-\frac{2}{\pi})T^{-1}}{T^2} \\
&= \frac{2}{3(4-\pi)}\left(\frac{1}{R^2T}+\frac{1-\frac{2}{\pi}}{T^3}\right)
\end{aligned}
$$

$$
\begin{aligned}
\text{with } T &= \sqrt{\frac{2}{\pi}-\left(1-\frac{2}{\pi}\right)R^2} \\
\text{and } \frac{\partial R}{\partial \gamma_1} &= \frac{\partial}{\partial \gamma_1}\left(\left(\frac{2\gamma_1}{4-\pi}\right)^{\frac{1}{3}}\right) \\
&= \frac{1}{3}\left(\frac{2\gamma_1}{4-\pi}\right)^{-\frac{2}{3}}\frac{2}{4-\pi} \\
&= \frac{2}{3(4-\pi)}R^{-2}.
\end{aligned}
$$

We can now calculate $I^{\text{CP}}(\theta^{\text{CP}})$ numerically. This computation shows that $I^{\text{CP}}(\theta^{\text{CP}})$ approaches $\text{diag}\left(\frac{1}{\sigma^2}, \frac{2}{\sigma^2}, \frac{1}{6}\right)$ when γ_1 approaches 0.

Now using Proposition 1, proven by Chiogna (2005) [21], we have in the neighbourhood of zero, $(\mu, \sigma) = \chi_{\text{II}}, \gamma_1 = (2b^3 - b)\delta^3$, as:

$$
\xi = \mu + \sigma b \delta,
$$

$$\omega = \sigma - \frac{1}{2}\sigma b^2\delta^2,$$

$$\gamma_1 = (2b^3 - b)\delta^3.$$

Therefore, $\gamma_1 = O(\delta^3)$. As the sampling fluctuations in $\hat{\delta}$ are $O_p\left(n^{-1/6}\right)$, this parametrization brings the order of the convergence of the MLE estimator of the skewness parameter $\hat{\gamma}_1$ back to the usual $O_p\left(n^{1/2}\right)$.

3.1.2 Orthogonalization

We will now look at a different reparametrization, first proposed by Hallin and Ley (2014) [39]. The collinearity between the first and the third score vector evaluated in θ_0, $l^1_{\theta_0}$ and $l^3_{\theta_0}$ respectively, is solved by a Gram-Schmidt orthogonalisation process applied to the components of the score vector. This process orthonormalizes a set of vectors, in this case the components of the score vector, by determining the component of $l^3_{\theta_0}$ orthogonal to $l^1_{\theta_0}$ and $l^2_{\theta_0}$. This corresponds to the score for skewness $l^3_{\theta_0}$ becoming orthogonal to the score for location $l^1_{\theta_0}$, since $l^3_{\theta_0}$ and $l^2_{\theta_0}$ are already independent ($\mathrm{Cov}(l^2_{\theta_0}, l^3_{\theta_0}) = I^{DP}_{2,3}(\theta_0) = 0$).
The general Gram-Schmidt orthogonalization process is as follows : the projection operator is defined by

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{<\mathbf{u},\mathbf{v}>}{<\mathbf{u},\mathbf{u}>}\mathbf{u}$$

with $<\mathbf{u},\mathbf{v}>$ the inner product of the vectors $\mathbf{u}$ and $\mathbf{v}$. This operator projects $\mathbf{v}$ orthogonally on to $\mathbf{u}$. The process itself then works as follows

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 \\
\mathbf{u}_2 &= \mathbf{v}_2 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_2) \\
\mathbf{u}_3 &= \mathbf{v}_3 - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \mathrm{proj}_{\mathbf{u}_2}(\mathbf{v}_3) \\
&\vdots \\
\mathbf{u}_k &= \mathbf{v}_k - \sum_{j=1}^{k-1}\mathrm{proj}_{\mathbf{u}_j}(\mathbf{v}_k)
\end{aligned}$$

We will now apply this process to $l^1_{\theta_0}$, $l^2_{\theta_0}$ and $l^3_{\theta_0}$.

$$l^{1(1)}_{\theta_0} = l^1_{\theta_0},$$

$$\begin{aligned}
l^{2(1)}_{\theta_0} &= l^2_{\theta_0} - l^1_{\theta_0}\frac{\mathrm{Cov}(l^1_{\theta_0}, l^2_{\theta_0})}{\mathrm{Var}(l^1_{\theta_0})} \\
&= l^2_{\theta_0},
\end{aligned}$$

$$
\begin{aligned}
l_{\theta_0}^{3(1)} &= l_{\theta_0}^{3} - l_{\theta_0}^{1}\frac{\mathrm{Cov}(l_{\theta_0}^{1}, l_{\theta_0}^{3})}{\mathrm{Var}(l_{\theta_0}^{1})} - l_{\theta_0}^{2}\frac{\mathrm{Cov}(l_{\theta_0}^{2}, l_{\theta_0}^{3})}{\mathrm{Var}(l_{\theta_0}^{2})} \\
&= l_{\theta_0}^{3} - l_{\theta_0}^{1}\frac{\mathrm{Cov}(l_{\theta_0}^{1}, l_{\theta_0}^{3})}{\mathrm{Var}(l_{\theta_0}^{1})}
\end{aligned}
$$

with $\mathrm{Cov}(l_{\theta_0}^{1}, l_{\theta_0}^{2}) = \mathrm{Cov}(l_{\theta_0}^{2}, l_{\theta_0}^{3}) = 0$ because of the independence. We can now substitute the values for $\mathrm{Cov}(l_{\theta_0}^{1}, l_{\theta_0}^{3})$ and $\mathrm{Var}(l_{\theta_0}^{1})$ in the last equation. We get

$$
l_{\theta_0}^{3(1)} = zb - z\sigma^{-1}\frac{b\sigma^{-1}}{\sigma^{-2}} = 0.
$$

This orthogonal system of scores corresponds with the reparametrization $\theta = (\mu^{(1)}, \sigma^{(1)}, \delta)'$, with

$$
\mu^{(1)} = \mu + \delta\frac{\mathrm{Cov}(l_{\theta_0}^{1}, l_{\theta_0}^{3})}{\mathrm{Var}(l_{\theta_0}^{1})} = \mu + \delta b\sigma,
$$

$$
\sigma^{(1)} = \sigma.
$$

We find the expression for $\mu^{(1)}$ by using the same reparametrization.
The density function at $x \in \mathbb{R}$ becomes

$$
f_{\mu^{(1)},\sigma^{(1)},\delta}(x) = 2\left(\sigma^{(1)}\right)^{-1}\phi\left(\left(\sigma^{(1)}\right)^{-1}\left(x - \mu^{(1)} + \sqrt{\frac{2}{\pi}}\delta\sigma^{(1)}\right)\right)\Phi\left(\delta\left(\sigma^{(1)}\right)^{-1}\left(x - \mu^{(1)} + \sqrt{\frac{2}{\pi}}\delta\sigma^{(1)}\right)\right). \tag{3.1.3}
$$

At $\delta = 0$ this reparametrization becomes $(\mu^{(1)}, \sigma^{(1)}, 0)' = (\mu, \sigma, 0)' = \theta_0$.
The score for skewness is canceled by this reparametrization at $\delta = 0$ and therefore so is the linear term in the Taylor expansion of the log-likelihood. Thus we have to look at the second derivatives with respect to δ. Taylor expansion of the log-likelihood about θ_0 gives us

$$
\begin{aligned}
\mathscr{L}(\theta_0; x) &= \log f_{\theta_0}(x) + (\delta - 0)\partial_\delta \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0} + \frac{(\delta-0)^2}{2}\partial_\delta^2 \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0} + \ldots \\
&= \log f_{\theta_0}(x) + \delta\partial_\delta \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0} + \frac{\delta^2}{2}\partial_\delta^2 \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0} + \ldots \\
&= \log f_{\theta_0}(x) + \frac{\delta^2}{2}\partial_\delta^2 \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0} + \ldots
\end{aligned}
$$

where $\delta\partial_\delta \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0}$ is zero because $\partial_\delta \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0} = l_{\theta_0}^{3(1)} = 0$. So the first local approximation is given by the quadratic term $\frac{\delta^2}{2}\partial_\delta^2 \log f_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\theta_0}$. Consequently, if the impact on the log-likelihood of δ is of the central-limit magnitude $n^{-\frac{1}{2}}$, then $\delta = O\left(n^{-\frac{1}{4}}\right)$. Since we only have a factor δ^2 in the expression for the Taylor expansion, information about its sign is lost.
The existence of second-order derivatives recommends reparametrizing skewness in terms of $\delta^{(1)} = \mathrm{sign}(\delta)\delta^2$ instead of δ. Consider the reparametrization $\theta^{(1)} = (\mu^{(1)}, \sigma^{(1)}, \delta^{(1)})'$.

We will now differentiate $\log f_{\mu^{(1)},\sigma^{(1)},\delta^{(1)}}$ with respect to $\delta^{(1)}$.

$$
\begin{aligned}
\partial_{\delta^{(1)}} \log f_{\theta^{(1)}} &= \partial_{\delta^{(1)}}(\delta)\partial_\delta \log f_{\theta^{(1)}} \\
&= \partial_{\delta^{(1)}}(\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2})\partial_\delta \log f_{\theta^{(1)}} \\
&= \frac{1}{2\sqrt{|\delta^{(1)}|}}\partial_\delta \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}} \qquad \text{if } \delta^{(1)} \neq 0.
\end{aligned}
$$

At $\delta^{(1)} = 0$ we apply l'Hospital's rule once to get

$$
\begin{aligned}
\partial_{\delta^{(1)}} \log f_{\theta^{(1)}} &= \lim_{\delta^{(1)}\to 0} \frac{1}{2\sqrt{|\delta^{(1)}|}}\partial_\delta \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}} \\
&\overset{\text{H}}{=} \lim_{\delta^{(1)}\to 0} \frac{\partial_{\delta^{(1)}}\partial_\delta \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}}}{\partial_{\delta^{(1)}} 2\sqrt{|\delta^{(1)}|}} \\
&= \lim_{\delta^{(1)}\to 0} \frac{1}{2\frac{1}{2\sqrt{|\delta^{(1)}|}}}\partial_{\delta^{(1)}}(\delta)\partial_\delta^2 \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}} \\
&= \lim_{\delta^{(1)}\to 0} \sqrt{|\delta^{(1)}|}\frac{1}{2\sqrt{|\delta^{(1)}|}}\partial_\delta^2 \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}} \\
&= \lim_{\delta^{(1)}\to 0} \pm\frac{1}{2}\partial_\delta^2 \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}} \\
&= \pm\frac{1}{2}\partial_\delta^2 \log f_{\theta^{(1)}}\Big|_{\delta=0}.
\end{aligned}
$$

The plus minus sign is necessary because $\delta = \operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}$.
Combining these results we get

$$
\partial_{\delta^{(1)}} \log f_{\theta^{(1)}} = \begin{cases} \frac{1}{2\sqrt{|\delta^{(1)}|}}\partial_\delta \log f_{\theta^{(1)}}\Big|_{\delta=\operatorname{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}} & \text{if } \delta^{(1)} \neq 0 \\ \pm\frac{1}{2}\partial_\delta^2 \log f_{\theta^{(1)}}\Big|_{\delta=0} & \text{if } \delta^{(1)} = 0 \end{cases}. \tag{3.1.4}
$$

The sign at $\delta = 0$ can not be defined because the left derivative and the right derivative are not the same. Set $y = (\sigma^{(1)})^{-1}(x - \mu^{(1)} + \sqrt{\frac{2}{\pi}}\delta\sigma^{(1)})$. The log-likelihood function of (2.1.1) is

$$
\begin{aligned}
\log f_{\theta^{(1)}} &= -\log(\sigma^{(1)}) + \log\phi(y) + \log 2\Phi(y\delta) \\
&= -\log(\sigma^{(1)}) - \frac{(x - \mu^{(1)} + \sqrt{\frac{2}{\pi}}\delta\sigma^{(1)})^2}{2(\sigma^{(1)})^2} + \zeta_0(y\delta).
\end{aligned}
$$

Therefrom, together with (2.1.2), it follows that

$$
\begin{aligned}
\partial_{\delta^{(1)}} \log f_{\theta^{(1)}} &= \pm\frac{1}{2}\partial_\delta^2 \log f_{\theta^{(1)}} \\
&= \pm\frac{1}{2}\partial_\delta^2\Big(-\log(\sigma^{(1)}) - \frac{(x - \mu^{(1)} + \sqrt{\frac{2}{\pi}}\delta\sigma^{(1)})^2}{2(\sigma^{(1)})^2} + \zeta_0(y\delta)\Big)
\end{aligned}
$$

$$= \pm\frac{1}{2}\partial_\delta\Bigg(-\frac{(x-\mu^{(1)}+\sqrt{\frac{2}{\pi}}\delta\sigma^{(1)})}{\sigma^{(1)}}\sqrt{\frac{2}{\pi}}+\Big(\sigma^{(1)})^{-1}(x-\mu^{(1)})+2\sqrt{\frac{2}{\pi}}\delta\Big)\zeta_1(y\delta)\Bigg)$$
$$= \pm\frac{1}{2}\Bigg(-\frac{2}{\pi}+2\sqrt{\frac{2}{\pi}}\zeta_1(y\delta)+\Big(\sigma^{(1)})^{-1}(x-\mu^{(1)})+2\sqrt{\frac{2}{\pi}}\delta\Big)^2\zeta_2(y\delta)\Bigg).$$

In θ_0 this becomes

$$\begin{aligned}
\partial_{\delta^{(1)}}\log f_{\theta^{(1)}}\Big|_{\theta_0} &= \pm\frac{1}{2}\Bigg(-\frac{2}{\pi}+2\frac{2}{\pi}-\frac{2}{\pi}\Big(\sigma^{-1}(x-\mu)\Big)^2\Bigg)\\
&= \pm\frac{1}{2}\Bigg(\frac{2}{\pi}-\frac{2}{\pi}\sigma^{-2}(x-\mu)^2\Bigg)\\
&= \pm\frac{1}{\pi}\Bigg(1-\sigma^{-2}(x-\mu)^2\Bigg)
\end{aligned}$$

hence

$$\begin{aligned}
l_{\theta_0^{(1)}}(x) &= \Bigg(l^1_{\theta_0^{(1)}}(x), l^2_{\theta_0^{(1)}}(x), l^3_{\theta_0^{(1)}}(x)\Bigg)'\\
&= \begin{pmatrix} \partial_{\mu^{(1)}}\log f_{\theta^{(1)}}\Big|_{\theta_0} \\ \partial_{\sigma^{(1)}}\log f_{\theta^{(1)}}\Big|_{\theta_0} \\ \partial_{\delta^{(1)}}\log f_{\theta^{(1)}}\Big|_{\theta_0} \end{pmatrix}\\
&= \begin{pmatrix} \sigma^{-2}(x-\mu) \\ -\sigma^{-1}+\sigma^{-3}(x-\mu)^2 \\ \pm\frac{1}{\pi}\Bigg(1-\sigma^{-2}(x-\mu)^2\Bigg) \end{pmatrix}.
\end{aligned}$$

We now want to calculate the covariance. Because $l^1_{\theta_0}$ and $l^2_{\theta_0}$ stay unaltered, we already have

$$I(\theta_0^{(1)}) = \begin{pmatrix} \sigma^{-2} & 0 & I^{13}(\theta_0^{(1)}) \\ 0 & 2\sigma^{-2} & I^{23}(\theta_0^{(1)}) \\ I^{13}(\theta_0^{(1)}) & I^{23}(\theta_0^{(1)}) & I^{33}(\theta_0^{(1)}) \end{pmatrix}.$$

We compute the remaining elements by calculating $I^{ij}(\theta_0^{(1)}) = \mathbb{E}\Bigg(l^i_{\theta_0^{(1)}}(x)l^j_{\theta_0^{(1)}}(x)\Bigg)$ using (2.1.1).

$$\begin{aligned}
I^{13}(\theta_0^{(1)}) = I^{31}(\theta_0^{(1)}) &= \mathbb{E}\Bigg(l^1_{\theta_0^{(1)}}(z)l^3_{\theta_0^{(1)}}(z)\Bigg)\\
&= \pm\frac{1}{\pi\sigma}\mathbb{E}\Big(z\Big(1-z^2\Big)\Big) = 0,
\end{aligned}$$

$$\begin{aligned}
I^{23}(\theta_0^{(1)}) = I^{32}(\theta_0^{(1)}) &= \mathbb{E}\Bigg(l^2_{\theta_0^{(1)}}(z)l^3_{\theta_0^{(1)}}(z)\Bigg)\\
&= \mp\frac{1}{\pi\sigma}\mathbb{E}\Big((1-z^2)^2\Big) = \mp\frac{2}{\pi\sigma},
\end{aligned}$$

$$
\begin{aligned}
I^{33}(\theta_0^{(1)}) &= \mathbb{E}\left(\left(l^3_{\theta_0^{(1)}}(z)\right)^2\right) \\
&= \frac{1}{\pi^2}\mathbb{E}\left(\left(1-z^2\right)^2\right) = \frac{2}{\pi^2}.
\end{aligned}
$$

Combining all these results, we get

$$
I(\theta_0^{(1)}) = \begin{pmatrix} \sigma^{-2} & 0 & 0 \\ 0 & 2\sigma^{-2} & \pm\frac{2}{\pi\sigma} \\ 0 & \pm\frac{2}{\pi\sigma} & \frac{2}{\pi^2} \end{pmatrix}.
$$

We can easily see that the determinant of this matrix will be zero because of the collinearity of $l^2_{\theta_0^{(1)}}$ and $l^3_{\theta_0^{(1)}}$. We thus find a double singularity for the skew-normal family. We will need to do a second reparametrization the way we did with the first one. Applying the Gram-Schmidt orthogonalisation process again, but now with the score for scale instead of the score for location, we determine the component of $l^3_{\theta_0^{(1)}}$ orthogonal to $l^1_{\theta_0^{(1)}}$ and $l^2_{\theta_0^{(1)}}$. The resulting score of skewness will be zero at $\theta_0^{(1)}$:

$$
\begin{aligned}
l^3_{\theta_0^{(1)}} - l^2_{\theta_0^{(1)}}\frac{\mathrm{Cov}(l^2_{\theta_0^{(1)}}, l^3_{\theta_0^{(1)}})}{\mathrm{Var}(l^2_{\theta_0^{(1)}})} &= \pm\frac{1}{\pi}\left(1-\sigma^{-2}(x-\mu)^2\right) - (-\sigma^{-1}+\sigma^{-3}(x-\mu)^2)\frac{\mp\frac{2}{\pi\sigma}}{2\sigma^{-2}} \\
&= \pm\frac{1}{\pi}\left(1-\sigma^{-2}(x-\mu)^2\right) + (1-\sigma^{-2}(x-\mu)^2)\left(\mp\frac{1}{\pi}\right) = 0.
\end{aligned}
$$

This projection leads to a reparametrization of the form $(\mu^{(2)}, \sigma^{(2)}, \delta)'$, with

$$
\mu^{(2)} = \mu^{(1)} = \mu + \delta\sigma b,
$$

$$
\sigma^{(2)} = \sigma^{(1)} + \delta^{(1)}\frac{\mathrm{Cov}(l^2_{\theta_0^{(1)}}, l^3_{\theta_0^{(1)}})}{\mathrm{Var}(l^2_{\theta_0^{(1)}})} = \sigma^{(1)}\left(1-\frac{\delta^2}{\pi}\right)
$$

applying the orthogonalization process to find the expression for $\sigma^{(2)}$.
The density function at $x \in \mathbb{R}$ becomes

$$
\begin{aligned}
f_{\mu^{(2)},\sigma^{(2)},\delta}(x) = 2(\sigma^{(2)})^{-1}\left(1-\frac{\delta^2}{\pi}\right)\phi\left((\sigma^{(2)})^{-1}\left(1-\frac{\delta^2}{\pi}\right)\left(x-\mu^{(2)}+\frac{b\pi\delta\sigma^{(2)}}{\pi-\delta^2}\right)\right) \\
\times\Phi\left(\delta(\sigma^{(2)})^{-1}\left(1-\frac{\delta^2}{\pi}\right)\left(x-\mu^{(2)}+\frac{b\pi\delta\sigma^{(2)}}{\pi-\delta^2}\right)\right).
\end{aligned}
\tag{3.1.5}
$$

Analogous to the first time we applied the orthogonalization process we can see that keeping δ as the skewness parameter gives a $n^{1/6}$ consistency rate. This is because the first two derivatives with respect to δ become zero at $\delta = 0$, so that the derivatives of order three will become dominant in local approximations of log-likelihoods. This appearance of third derivatives suggests reparametrizing skewness in terms of $\delta^{(2)} = \delta^3$, giving the reparametrization $\theta^{(2)} = (\mu^{(2)}, \sigma^{(2)}, \delta^{(2)})'$, with $\theta_0^{(2)} = (\mu, \sigma, 0)' = \theta_0$.

We will now determine the new score for skewness by differentiating $\log f_{\mu^{(2)},\sigma^{(2)},\delta^{(2)}}$ with respect to $\delta^{(2)}$.

$$\begin{aligned}
\partial_{\delta^{(2)}} \log f_{\theta^{(2)}} &= \partial_{\delta^{(2)}}(\delta)\partial_\delta \log f_{\theta^{(2)}} \\
&= \partial_{\delta^{(2)}}\left((\delta^{(2)})^{1/3}\right)\partial_\delta \log f_{\theta^{(2)}} \\
&= \frac{1}{3(\delta^{(2)})^{2/3}}\partial_\delta \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}} \quad \text{if } \delta^{(2)} \neq 0.
\end{aligned}$$

At $\delta^{(2)} = 0$ we apply l'Hospital's rule twice to get

$$\begin{aligned}
\partial_{\delta^{(2)}} \log f_{\theta^{(2)}} &= \lim_{\delta^{(2)}\to 0} \frac{1}{3(\delta^{(2)})^{2/3}}\partial_\delta \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}} \\
&\overset{H}{=} \lim_{\delta^{(2)}\to 0} \frac{\partial_{\delta^{(2)}}\partial_\delta \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}}}{\partial_{\delta^{(2)}} 3(\delta^{(2)})^{2/3}} \\
&= \lim_{\delta^{(2)}\to 0} \frac{\partial_{\delta^{(2)}}(\delta)\partial_\delta^2 \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}}}{2(\delta^{(2)})^{-1/3}} \\
&= \lim_{\delta^{(2)}\to 0} \frac{\partial_\delta^2 \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}}}{6(\delta^{(2)})^{1/3}} \\
&\overset{H}{=} \lim_{\delta^{(2)}\to 0} \frac{\partial_{\delta^{(2)}}\partial_\delta^2 \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}}}{\partial_{\delta^{(2)}} 6(\delta^{(2)})^{1/3}} \\
&= \lim_{\delta^{(2)}\to 0} \frac{\partial_{\delta^{(2)}}(\delta)\partial_\delta^3 \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}}}{2(\delta^{(2)})^{-2/3}} \\
&= \lim_{\delta^{(2)}\to 0} \frac{1}{6}\partial_\delta^3 \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}} \\
&= \frac{1}{6}\partial_\delta^3 \log f_{\theta^{(2)}}\Big|_{\delta=0}.
\end{aligned}$$

Combining these results we have

$$\partial_{\delta^{(2)}} \log f_{\theta^{(2)}} = \begin{cases} \frac{1}{3(\delta^{(2)})^{2/3}}\partial_\delta \log f_{\theta^{(2)}}\Big|_{\delta=(\delta^{(2)})^{1/3}} & \text{if } \delta^{(1)} \neq 0 \\ \frac{1}{6}\partial_\delta^3 \log f_{\theta^{(2)}}\Big|_{\delta=0} & \text{if } \delta^{(1)} = 0 \end{cases}. \tag{3.1.6}$$

Set $y = (\sigma^{(2)})^{-1}\left(1-\frac{\delta^2}{\pi}\right)\left(x-\mu^{(2)}+\frac{b\pi\delta\sigma^{(2)}}{\pi-\delta^2}\right)$. The log-likelihood of (3.1.5) is

$$\begin{aligned}
\log f_{\theta^{(2)}} &= -\log(\sigma^{(2)}) + \log\left(1-\frac{\delta^2}{\pi}\right) + \log\phi(y) + \log 2\Phi(\delta y) \\
&= -\log(\sigma^{(2)}) + \log\left(1-\frac{\delta^2}{\pi}\right) - \frac{\left(1-\frac{\delta^2}{\pi}\right)^2\left(x-\mu^{(2)}+\frac{b\pi\delta\sigma^{(2)}}{\pi-\delta^2}\right)^2}{2(\sigma^{(2)})^2} + \zeta_0(\delta y).
\end{aligned}$$

Therefrom together with (3.1.6), it follows that

$$\begin{aligned}
\partial_{\delta^{(2)}} \log f_{\theta^{(2)}} &= \frac{1}{6}\partial_\delta^3 \log f_{\theta^{(2)}} \\
&= \frac{1}{6}\partial_\delta^3 \left(-\log(\sigma^{(2)}) + \log\left(1-\frac{\delta^2}{\pi}\right) - \frac{\left(1-\frac{\delta^2}{\pi}\right)^2\left(x-\mu^{(2)}+\frac{b\pi\delta\sigma^{(2)}}{\pi-\delta^2}\right)^2}{2(\sigma^{(2)})^2} + \zeta_0(\delta y)\right) \\
&= \frac{1}{6}\partial_\delta^2 \left(\frac{-2\delta}{\pi-\delta^2} + (\sigma^{(2)})^{-1} y \left(2\frac{\delta}{\pi}(x-\mu^{(2)}) + b\sigma^{(2)}\right) + (\sigma^{(2)})^{-1}\left(\left(1-\frac{3\delta^2}{\pi}\right)(x-\mu^{(2)}) + 2\delta b\sigma^{(2)}\right)\zeta_1(\delta y)\right) \\
&= \frac{1}{6}\partial_\delta \left(-2\frac{\pi+\delta^2}{(\pi-\delta^2)^2} + (\sigma^{(2)})^{-2}\left(2\frac{\delta}{\pi}(x-\mu^{(2)}) + b\sigma^{(2)}\right)^2 + (\sigma^{(2)})^{-1} y \left(\frac{2}{\pi}(x-\mu^{(2)})\right)\right. \\
&\quad \left. +(\sigma^{(2)})^{-1}\left(-\frac{6\delta^2}{\pi}(x-\mu^{(2)}) + 2b\sigma^{(2)}\right)\zeta_1(\delta y) + (\sigma^{(2)})^{-2}\left(\left(1-\frac{3\delta^2}{\pi}\right)(x-\mu^{(2)}) + 2\delta b\sigma^{(2)}\right)^2 \zeta_2(\delta y)\right) \\
&= \frac{1}{6}\left(-4\delta\frac{3\pi-2\delta\pi-\delta^4}{(\pi-\delta^2)^4} + \frac{6}{\pi}(x-\mu^{(2)})(\sigma^{(2)})^{-2}\left(2\frac{\delta}{\pi}(x-\mu^{(2)}) + b\sigma^{(2)}\right) - (\sigma^{(2)})^{-1}\frac{12\delta}{\pi}(x-\mu^{(2)})\zeta_1(\delta y)\right. \\
&\quad +3\left(-\frac{6\delta}{\pi}(x-\mu^{(2)}) + 2b\sigma^{(2)}\right)(\sigma^{(2)})^{-2}\left(\left(1-\frac{3\delta^2}{\pi}\right)(x-\mu^{(2)}) + 2\delta b\sigma^{(2)}\right)\zeta_2(\delta y) \\
&\quad \left. +(\sigma^{(2)})^{-3}\left(\left(1-\frac{3\delta^2}{\pi}\right)(x-\mu^{(2)}) + 2\delta b\sigma^{(2)}\right)^3 \zeta_3(\delta y)\right).
\end{aligned}$$

In θ_0 this becomes

$$\begin{aligned}
\partial_{\delta^{(2)}} \log f_{\theta^{(2)}}\Big|_{\theta_0} &= \frac{1}{6}\left(\frac{6b}{\pi}(x-\mu^{(2)})(\sigma^{(2)})^{-1} - \frac{12b}{\pi}(\sigma^{(2)})^{-1}(x-\mu^{(2)}) + (\sigma^{(2)})^{-3}(x-\mu^{(2)})^3\left(-\sqrt{\frac{2}{\pi}} + \frac{4}{\pi}\sqrt{\frac{2}{\pi}}\right)\right) \\
&= -\frac{b}{\pi}z + \frac{z^3}{6}\left(-b + \frac{4}{\pi}b\right)
\end{aligned}$$

hence

$$\begin{aligned}
l_{\theta_0^{(2)}}(z) &= \left(l^1_{\theta_0^{(2)}}(z), l^2_{\theta_0^{(2)}}(z), l^3_{\theta_0^{(2)}}(z)\right)' \\
&= \begin{pmatrix} \partial_{\mu^{(2)}} \log f_{\theta^{(2)}}\Big|_{\theta_0} \\ \partial_{\sigma^{(2)}} \log f_{\theta^{(2)}}\Big|_{\theta_0} \\ \partial_{\delta^{(2)}} \log f_{\theta^{(2)}}\Big|_{\theta_0} \end{pmatrix} \\
&= \begin{pmatrix} \sigma^{-1}z \\ -\sigma^{-1} + \sigma^{-1}z^2 \\ -\frac{b}{\pi}z + \frac{z^3}{6}\left(-b + \frac{4}{\pi}b\right) \end{pmatrix}.
\end{aligned}$$

By the symmetry of the distribution of Z we have that $\mathbb{E}\left(l^1_{\theta_0^{(2)}}(x), l^2_{\theta_0^{(2)}}(x)\right) = \mathbb{E}\left(l^3_{\theta_0^{(2)}}(x), l^2_{\theta_0^{(2)}}(x)\right) = 0$. The elements $I^{11}(\theta_0^{(2)})$ and $I^{22}(\theta_0^{(2)})$ of the Fisher information matrix stay the same.

The remaining elements are

$$I^{13}(\theta_0^{(2)}) = I^{31}(\theta_0^{(2)}) = \mathbb{E}\left(l^1_{\theta_0^{(2)}}(z), l^3_{\theta_0^{(2)}}(z)\right) = -\frac{b}{\pi}\sigma^{-1}\mathbb{E}(z^2) + \frac{1}{6}\sigma^{-1}\left(-b + \frac{4}{\pi}b\right)\mathbb{E}(z^4)$$

$$= -\frac{1}{\pi}\sqrt{\frac{2}{\pi}}\sigma^{-1} + \frac{1}{2}\sigma^{-1}\left(-\sqrt{\frac{2}{\pi}} + \frac{4}{\pi}\sqrt{\frac{2}{\pi}}\right)$$

$$= \sigma^{-1}\frac{2-\pi}{\pi\sqrt{2\pi}},$$

$$I^{33}(\theta_0^{(2)}) = \mathbb{E}\left(\left(l^3_{\theta_0^{(2)}}(z)\right)^2\right) = \frac{b^2}{\pi^2}\mathbb{E}\left(z^2\right) - \frac{b}{3\pi}\left(-b + \frac{4}{\pi}b\right)\mathbb{E}\left(z^4\right) + \frac{1}{36}\left(-b + \frac{4}{\pi}b\right)^2\mathbb{E}\left(z^6\right)$$

$$= \frac{4}{\pi^3} - \frac{\sqrt{2}}{\pi\sqrt{\pi}}\left(-\sqrt{\frac{2}{\pi}} + \frac{4}{\pi}\sqrt{\frac{2}{\pi}}\right) + \frac{15}{36}\left(-\sqrt{\frac{2}{\pi}} + \frac{4}{\pi}\sqrt{\frac{2}{\pi}}\right)^2$$

$$= -\frac{4}{\pi^3} + \frac{2}{\pi^2} + \frac{15}{36}\left(\frac{2}{\pi} - \frac{16}{\pi^2} + \frac{32}{\pi^3}\right)$$

$$= \frac{5}{6\pi} - \frac{14}{3\pi^2} + \frac{40}{3\pi^3}.$$

The Fisher information matrix is the following

$$I(\theta_0^{(2)}) = \begin{pmatrix} \sigma^{-2} & 0 & \sigma^{-1}\frac{2-\pi}{\pi\sqrt{2\pi}} \\ 0 & 2\sigma^{-2} & 0 \\ -\frac{1}{2}\sqrt{\frac{2}{\pi}}\sigma^{-1} & 0 & \frac{80-28\pi+10\pi^2}{6\pi^3} \end{pmatrix}.$$

The determinant of this matrix is not equal to zero. So we have found a singularity-free reparametrization. We know that $I(\theta_0^{(2)})$ has full rank, so the root-n consistency rates are achieved for $\delta^{(2)} = \delta^3$. This means that at any $\delta \neq 0$ the same root-n rates imply. However, at $\delta = 0$ an $n^{1/2}$ rate for $\delta^{(2)}$ means an $n^{1/6}$ rate for $\delta = (\delta^{(2)})^{1/3}$. This is the same $n^{1/6}$ rate established by Chiogna (2005) [21] as we have seen in the previous sections.

3.2 Skew-t family

We will now retake the example of the skew-t family and take a look at its inferential aspects by making use of Di Ciccio and Monti (2011) [26]. The log-likelihood function is given by

$$\mathscr{L}(\theta^{DP}; x) = \log(\sigma^{-1}t(\sigma^{-1}(x-\mu); \delta, \nu))$$

$$= -\log(\sigma) + \log(t(\sigma^{-1}(x-\mu); \nu)) + \log\left(2T(\delta\sigma^{-1}(x-\mu)\sqrt{\frac{\nu+1}{\nu+\sigma^{-2}(x-\mu)^2}}; \nu+1)\right)$$

$$= -\log(\sigma) + \log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\right) - \frac{\nu+1}{2}\log\left(1 + \frac{\sigma^{-2}(x-\mu)^2}{\nu}\right) + \eta_0\left(\delta\sigma^{-1}(x-\mu)\sqrt{\frac{\nu+1}{\nu+\sigma^{-2}(x-\mu)^2}}; \nu+1\right)$$

with $\theta^{DP} = (\mu, \sigma, \delta, \nu)'$ and $\eta_0(x; \nu) = \log(2T(x; \nu))$.

The components of the score vector are

$$
\begin{aligned}
l^1_{\theta^{DP}} = \frac{\partial \mathscr{L}}{\partial \mu} &= \frac{2}{\nu}\sigma^{-2}(x-\mu)\frac{\nu+1}{2}\frac{\nu}{\nu+\sigma^{-2}(x-\mu)^2} \\
&+\delta\sigma^{-1}\eta_1\left(\delta\sigma^{-1}(x-\mu)\sqrt{\frac{\nu+1}{\nu+z^2}};\nu+1\right)\sqrt{\nu+1}\Big(\sigma^{-2}(x-\mu)^2(\nu+\sigma^{-2}(x-\mu)^2)^{-\frac{3}{2}}-(\nu+\sigma^{-2}(x-\mu)^2)^{-\frac{1}{2}}\Big) \\
&=\sigma^{-1}z\frac{\nu+1}{\nu+z^2}+\delta\sigma^{-1}\sqrt{\frac{\nu+1}{\nu+z^2}}\eta_1\left(\delta z\sqrt{\frac{\nu+1}{\nu+z^2}};\nu\right)\Big(z^2(\nu+z^2)^{-1}-1\Big) \\
&=\sigma^{-1}z\tau^2-\frac{\delta\sigma^{-1}\tau\nu}{\nu+z^2}\eta_1(\delta z\tau;\nu+1),
\end{aligned}
$$

$$
\begin{aligned}
l^2_{\theta^{DP}} = \frac{\partial \mathscr{L}}{\partial \sigma} &= -\sigma^{-1}+\frac{\nu+1}{2}\frac{\nu}{\nu+\sigma^{-2}(x-\mu)^2}\frac{2(x-\mu)^2\sigma^{-3}}{\nu} \\
&+\delta(x-\mu)\eta_1(\delta\sigma^{-1}(x-\mu)\tau;\nu+1)\sqrt{\nu+1}\Big(-\sigma^{-2}(\nu+\sigma^{-2}(x-\mu)^2)^{-\frac{1}{2}}+\sigma^{-4}(x-\mu)^2(\nu+\sigma^{-2}(x-\mu)^2)^{-\frac{3}{2}}\Big) \\
&=-\sigma^{-1}+\sigma^{-1}z^2\frac{\nu+1}{\nu+z^2}+\delta z\sigma^{-1}\eta_1(\delta z\tau;\nu+1)\sqrt{\frac{\nu+1}{\nu+z^2}}\Big(-1+z^2(\nu+z^2)^{-1}\Big) \\
&=-\sigma^{-1}+\sigma^{-1}z^2\tau^2-\frac{\delta z\nu\tau\sigma^{-1}}{\nu+z^2}\eta_1(\delta z\tau;\nu+1),
\end{aligned}
$$

$$
\begin{aligned}
l^3_{\theta^{DP}} = \frac{\partial \mathscr{L}}{\partial \delta} &= \sigma^{-1}(x-\mu)\sqrt{\frac{\nu+1}{\nu+\sigma^{-2}(x-\mu)^2}}\eta_1\left(\delta\sigma^{-1}(x-\mu)\sqrt{\frac{\nu+1}{\nu+\sigma^{-2}(x-\mu)^2}};\nu+1\right) \\
&= z\tau\eta_1(\delta z\tau;\nu+1),
\end{aligned}
$$

$$
l^4_{\theta^{DP}} = \frac{\partial \mathscr{L}}{\partial \nu} = c_\nu - \frac{1}{2}\log(1+\frac{z^2}{\nu})+\frac{\nu+1}{2}\frac{\nu}{\nu+z^2}\frac{z^2}{\nu^2}+p_{\nu+1}(z\delta\tau)
$$

with

$$
\begin{aligned}
z &= \sigma^{-1}(x-\mu), \\
\tau &= \sqrt{\frac{\nu+1}{\nu+z^2}}, \\
\eta_r(x) &= \frac{d^r}{dx^r}\eta_0 \quad (r=1,2,\ldots).
\end{aligned}
$$

$$
\begin{aligned}
c_\nu &= \frac{\partial}{\partial \nu}\log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\right)=\frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)-\frac{1}{\nu}\right), \\
p_\nu(x) &= \frac{\partial}{\partial \nu}(\eta_0(x;\nu)),
\end{aligned}
$$

First we will evaluate $\eta_1(\delta\tau z;\nu)$ in $\delta=0$, because we will need this to evaluate the components of the score vector in $\delta=0$.

$$
\eta_1(0;\nu)=\frac{t(0;\nu)}{T(0;\nu)}=\frac{2\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}
$$

and by applying the Leibniz integral rule

$$
\begin{aligned}
p_{\nu+1}(\delta\tau z) &= \frac{1}{T(\delta\tau z,\nu+1)}\left(t(\delta\tau z,\nu+1)\frac{\delta z}{2\tau}\frac{z^2-1}{(\nu+z^2)^2}+\int_{-\infty}^{\delta\tau z}\frac{\partial}{\partial\nu}t(u,\nu+1)du\right)\\
&= \frac{t(\delta\tau z,\nu+1)}{T(\delta\tau z,\nu+1)}\frac{\delta z}{2\tau}\frac{z^2-1}{(\nu+z^2)^2}+\frac{1}{T(\delta\tau z,\nu+1)}\left(\left(\int_{-\infty}^{\delta\tau z}\frac{\partial}{\partial\nu}\left(\frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}\right)\left(1+\frac{u^2}{\nu+1}\right)^{-\frac{\nu+2}{2}}du\right.\right.\\
&\left.+\int_{-\infty}^{\delta\tau z}t(u,\nu+1)\left(-\frac{1}{2}\log\left(1+\frac{u^2}{\nu+1}\right)+\frac{u^2}{2(\nu+1+u^2)}\right)du\right).
\end{aligned}
$$

Calculating the derivative in the second term in this equation we get

$$
\begin{aligned}
\frac{\partial}{\partial\nu}\left(\frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}\right) &= \frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}\left(-\frac{1}{2(\nu+1)}-\frac{1}{2}\psi\left(\frac{\nu+1}{2}\right)+\frac{1}{2}\psi\left(\frac{\nu+2}{2}\right)\right)\\
&= c_{\nu+1}\frac{\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}.
\end{aligned}
$$

Substituting this result in the expression for $p_{\nu+1}(\delta\tau z)$ gives us

$$
\begin{aligned}
p_{\nu+1}(\delta\tau z) &= \frac{t(\delta\tau z,\nu+1)}{T(\delta\tau z,\nu+1)}\frac{\delta z}{2\tau}\frac{z^2-1}{(\nu+z^2)^2}+\frac{c_{\nu+1}}{T(\delta\tau z,\nu+1)}\int_{-\infty}^{\delta\tau z}t(u;\nu+1)du\\
&+\frac{1}{T(\delta\tau z,\nu+1)}\int_{-\infty}^{\delta\tau z}t(u,\nu+1)\left(-\frac{1}{2}\log\left(1+\frac{u^2}{\nu+1}\right)+\frac{u^2}{2(\nu+1+u^2)}\right)du\\
&= \frac{t(\delta\tau z,\nu+1)}{T(\delta\tau z,\nu+1)}\frac{\delta z}{2\tau}\frac{z^2-1}{(\nu+z^2)^2}+c_{\nu+1}+\frac{\gamma}{T(\delta\tau z,\nu+1)}.
\end{aligned}
$$

In $\delta=0$ this becomes

$$
\begin{aligned}
p_{\nu+1}(0) &= \frac{1}{2}\left(\psi\left(\frac{\nu+2}{2}\right)-\psi\left(\frac{\nu+1}{2}\right)-\frac{1}{\nu+1}\right)+2\gamma_0\\
&= \frac{1}{2}\left(\psi\left(\frac{\nu+2}{2}\right)-\psi\left(\frac{\nu+1}{2}\right)-\frac{1}{\nu+1}\right)+\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}+1\right)+\frac{1}{\nu+1}\right)\\
&= \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}+1\right)+\frac{1}{\nu+1}\right)
\end{aligned}
$$

because, using the result of Di Ciccio and Monti (2011) [26],

$$
\gamma_0=\frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}+1\right)+\frac{1}{\nu+1}\right).
$$

Evaluating these components of the score vector in $\delta = 0$ we get

$$\begin{pmatrix} \partial_\mu \log f_{\theta^{DP}}\Big|_{\delta=0} \\ \partial_\sigma \log f_{\theta^{DP}}\Big|_{\delta=0} \\ \partial_\delta \log f_{\theta^{DP}}\Big|_{\delta=0} \\ \partial_\nu \log f_{\theta^{DP}}\Big|_{\delta=0} \end{pmatrix} = \begin{pmatrix} \sigma^{-1} z\tau^2 \\ -\sigma^{-1} + \sigma^{-1} z^2\tau^2 \\ z\tau \frac{2\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})} \\ \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \log(1+\frac{z^2}{\nu}) + \frac{z^2-1}{\nu+z^2}\right) \end{pmatrix}.$$

We can now calculate the elements of the Fisher information matrix. We have by the symmetry of the distribution of Z that $\mathbb{E}\left(l^1, l^2\right) = \mathbb{E}\left(l^1, l^4\right) = \mathbb{E}\left(l^2, l^3\right) = \mathbb{E}\left(l^3, l^4\right) = 0$. We compute the non-zero elements of the Fisher information matrix by using the change of the variable $u = (1+\frac{z^2}{\nu})^{-1}$, elaborated by Arellano-Valle and Genton (2010) [5] .

$$\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)^k \left(1+\frac{z^2}{\nu}\right)^{-m/2}\right) = \frac{B\left(\frac{\nu+m-2k}{2}, \frac{1+2k}{2}\right)}{B\left(\frac{\nu}{2}, \frac{1}{2}\right)},$$

$$\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)^k \left(1+\frac{z^2}{\nu}\right)^{-m/2} \log\left(1+\frac{z^2}{\nu}\right)\right) = -\frac{B\left(\frac{\nu+m-2k}{2}, \frac{1+2k}{2}\right)}{B(\frac{\nu}{2}, \frac{1}{2})}\left(\psi\left(\frac{\nu+m-2k}{2}\right) - \psi\left(\frac{\nu+m+1}{2}\right)\right),$$

$$\begin{aligned}\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)^k \left(1+\frac{z^2}{\nu}\right)^{-m/2} \left(\log\left(1+\frac{z^2}{\nu}\right)\right)^2\right) = \frac{B\left(\frac{\nu+m-2k}{2}, \frac{1+2k}{2}\right)}{B(\frac{\nu}{2}, \frac{1}{2})}&\left(\left(\psi\left(\frac{\nu+m-2k}{2}\right) - \psi\left(\frac{\nu+m+1}{2}\right)\right)^2\right. \\ &\left. + \psi'\left(\frac{\nu+m-2k}{2}\right) - \psi'\left(\frac{\nu+m+1}{2}\right)\right).\end{aligned}$$

Using these expressions and $z^2\tau^2 = \frac{\nu+1}{\nu+z^2} = (\nu+1)\left(1+\frac{z^2}{\nu}\right)^{-1}\left(\frac{z^2}{\nu}\right)$ we get

$$\begin{aligned} I^{11}(\theta^{DP}) = \mathbb{E}\left((l^1)^2\right) &= \sigma^{-2}\mathbb{E}\left(z^2\tau^4\right) \\ &= \sigma^{-2}\frac{(\nu+1)^2}{\nu}\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-2}\right) \\ &= \sigma^{-2}\frac{(\nu+1)^2}{\nu}\frac{B\left(\frac{\nu+2}{2}, \frac{3}{2}\right)}{B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \\ &= \sigma^{-2}\frac{\nu+1}{\nu+3}, \end{aligned}$$

$$
\begin{aligned}
I^{13}(\theta^{DP}) = I^{31}(\theta^{DP}) = \mathbb{E}\left(l^1, l^3\right) &= \sigma^{-1}\frac{2\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}\mathbb{E}\left(z^2\tau^3\right) \\
&= \sigma^{-1}\frac{(\nu+1)^{3/2}}{\sqrt{\nu}}\frac{2\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-3/2}\right) \\
&= \sigma^{-1}\frac{(\nu+1)^{3/2}}{\sqrt{\nu}}\frac{2\Gamma(\frac{\nu+2}{2})}{\sqrt{(\nu+1)\pi}\Gamma(\frac{\nu+1}{2})}\frac{B\left(\frac{\nu+1}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)} \\
&= \sigma^{-1}(\nu+1)\sqrt{\nu}\frac{\Gamma(\frac{\nu+1}{2})}{2\sqrt{\pi}\Gamma(\frac{\nu+4}{2})},
\end{aligned}
$$

$$
\begin{aligned}
I^{22}(\theta^{DP}) = \mathbb{E}\left((l^2)^2\right) &= \sigma^{-2}\mathbb{E}\left((1-z^2\tau^2)^2\right) \\
&= \sigma^{-2}\mathbb{E}\left(1-2z^2\tau^2+z^4\tau^4\right) \\
&= \sigma^{-2}\left(1-2(\nu+1)\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-1}\right)+(\nu+1)^2\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)^2\left(1+\frac{z^2}{\nu}\right)^{-2}\right)\right) \\
&= \sigma^{-2}\left(1-2(\nu+1)\frac{B\left(\frac{\nu}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}+(\nu+1)^2\frac{B\left(\frac{\nu}{2},\frac{5}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}\right) \\
&= \sigma^{-2}\left(-1+3\frac{\nu+1}{\nu+3}\right),
\end{aligned}
$$

$$
\begin{aligned}
I^{24}(\theta^{DP}) = I^{42}(\theta^{DP}) = \mathbb{E}\left(l^2, l^4\right) &= -\frac{\sigma^{-1}}{2}\left(\psi\left(\frac{\nu+1}{2}\right)-\psi\left(\frac{\nu}{2}\right)\right)\left(1-\mathbb{E}\left(z^2\tau^2\right)\right)-\frac{\sigma^{-1}}{2}\left(\mathbb{E}\left(\log\left(1+\frac{z^2}{\nu}\right)\right)\right. \\
&\left.-\mathbb{E}\left(z^2\tau^2\log\left(1+\frac{z^2}{\nu}\right)\right)\right)-\frac{\sigma^{-1}}{2}\left(\mathbb{E}\left(\frac{z^2-1}{\nu+z^2}\right)-\mathbb{E}\left(\frac{(z^2-1)z^2\tau^2}{\nu+z^2}\right)\right) \\
&= -\frac{\sigma^{-1}}{2}\left(\frac{1}{\nu}\mathbb{E}\left(\frac{z^2}{\nu}\left(1+\frac{z^2}{\nu}\right)^{-1}\right)-\frac{1}{\nu}\mathbb{E}\left(\left(1+\frac{z^2}{\nu}\right)^{-1}\right)-\frac{(\nu+1)^2}{\nu}\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)^2\left(1+\frac{z^2}{\nu}\right)^{-3}\right)\right. \\
&\left.+\frac{(\nu+1)}{\nu}\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-2}\right)\right) \\
&= -\frac{\sigma^{-1}}{2\nu}\left(\frac{B\left(\frac{\nu}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}-\frac{B\left(\frac{\nu+1}{2},\frac{1}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}-(\nu+1)^2\frac{B\left(\frac{\nu+2}{2},\frac{5}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}+(\nu+1)\frac{B\left(\frac{\nu+2}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}\right) \\
&= -\frac{\sigma^{-1}}{2\nu}\left(\frac{1}{\nu+1}-\frac{\nu}{2}\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2-\frac{3\nu(\nu+1)}{(\nu+5)(\nu+3)}+\frac{\nu}{\nu+3}\right) \\
&= -\frac{\sigma^{-1}}{2}\left(\frac{1}{\nu(\nu+1)}-\frac{1}{2}\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^2-\frac{2(\nu-1)}{(\nu+5)(\nu+3)}\right),
\end{aligned}
$$

$$
\begin{aligned}
I^{33}(\theta^{DP}) = \mathbb{E}\left((l^3)^2\right) &= \frac{4\Gamma^2(\frac{\nu+2}{2})}{(\nu+1)\pi\Gamma^2(\frac{\nu+1}{2})}\mathbb{E}\left(z^2\tau^2\right) \\
&= (\nu+1)\frac{4\Gamma^2(\frac{\nu+2}{2})}{(\nu+1)\pi\Gamma^2(\frac{\nu+1}{2})}\mathbb{E}\left(\left(\frac{z^2}{\nu}\right)\left(1+\frac{z^2}{\nu}\right)^{-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{4\Gamma^2(\frac{\nu+2}{2})}{\pi\Gamma^2(\frac{\nu+1}{2})} \frac{B\left(\frac{\nu}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)} \\
&= \frac{4\Gamma^2(\frac{\nu+2}{2})}{(\nu+1)\pi\Gamma^2(\frac{\nu+1}{2})},
\end{aligned}
$$

$$
\begin{aligned}
I^{44}(\theta^{DP}) = \mathbb{E}\left((l^4)^2\right) &= \mathbb{E}\left(\frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \log(1+\frac{z^2}{\nu}) + \frac{z^2-1}{\nu+z^2}\right)^2\right). \\
&= \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)^2 - \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)\mathbb{E}\left(\log(1+\frac{z^2}{\nu}) - \frac{z^2-1}{\nu+z^2}\right) \\
&+ \frac{1}{2}\mathbb{E}\left(\left(\log(1+\frac{z^2}{\nu})\right)^2\right) - \mathbb{E}\left(\log(1+\frac{z^2}{\nu})\frac{z^2-1}{\nu+z^2}\right) + \frac{1}{2}\mathbb{E}\left(\frac{(z^2-1)^2}{(\nu+z^2)^2}\right) \\
&= \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)^2 \left(2 - \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)\right) \\
&- \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)\frac{B\left(\frac{\nu}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)} + \frac{1}{2}\left(\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right)\right) \\
&+ \frac{1}{2}\left(\frac{B\left(\frac{\nu}{2},\frac{5}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)} - \frac{2}{\nu}\frac{B\left(\frac{\nu+2}{2},\frac{3}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)} + \frac{1}{\nu}\frac{B\left(\frac{\nu+2}{2},\frac{1}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)}\right) \\
&= \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)^2 \left(2 - \frac{1}{2}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right)\right) \\
&- \frac{1}{2(\nu+1)}\left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right) + \frac{1}{2}\left(\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right)\right) + \frac{\nu+4}{2(\nu+1)(\nu+3)}.
\end{aligned}
$$

We get

$$
I(\theta^{DP}) = \begin{pmatrix}
\sigma^{-2}\frac{\nu+1}{\nu+3} & 0 & \sigma^{-1}(\nu+1)\sqrt{\nu}\frac{2\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu+4}{2})} & 0 \\
0 & \sigma^{-2}\left(-1+3\frac{\nu+1}{\nu+3}\right) & 0 & I^{24}(\theta^{DP}) \\
\sigma^{-1}(\nu+1)\sqrt{\nu}\frac{2\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu+4}{2})} & 0 & \frac{4\Gamma^2(\frac{\nu+2}{2})}{(\nu+1)\pi\Gamma^2(\frac{\nu+1}{2})} & 0 \\
0 & I^{42}(\theta^{DP}) & 0 & I^{44}(\theta^{DP})
\end{pmatrix}.
$$

We find that for a finite ν, the information matrix $I(\theta^{DP})$ is invertible, in contrast to the information matrix of the skew-normal family.
However, as $\nu \to \infty$, the skew-t distribution tends to the skew-normal one. The components of the score function in $\delta = 0$ become

$$S_\mu = \sigma^{-1}z,$$

$$S_\sigma = -\sigma^{-1} + \sigma^{-1}z^2,$$

$$S_\delta = zb,$$

$$S_\nu = 0.$$

We can now compute the Fisher information matrix easily

$$I(\theta^{DP}) = \begin{pmatrix} \sigma^{-2} & 0 & b\sigma^{-1} & 0 \\ 0 & 2\sigma^{-2} & 0 & 0 \\ b\sigma^{-1} & 0 & b^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This matrix is clearly singular with rank 2, thus when omitting the zero column and zero row the obtained 3 × 3-matrix

$$\begin{pmatrix} \sigma^{-2} & 0 & b\sigma^{-1} \\ 0 & 2\sigma^{-2} & 0 \\ b\sigma^{-1} & 0 & b^2 \end{pmatrix}.$$

is still singular.We again found a singularity problem. The skew-t distribution suffers from a Fisher information singularity problem at $\delta = 0$ if $\nu \to \infty$.
We can overcome this problem by using the centred parametrization like we did in Section 2.1.1. We consider the centred parameters $(\xi, \omega, \gamma_1, \gamma_2)'$ instead of the direct parameters. Here γ_1 and γ_2 are the measures for skewness and kurtosis, respectively. The elaboration is completely analogue, see also Di Ciccio and Monti (2011) [26].

3.3 Conclusion

We have now discussed two existing solutions to the inferential problems that arise when the Fisher information matrix is singular. When we find ourselves in this case, there is thus not one unique way to work. One can choose between the two methods mentioned above, namely centred parametrization or orthogonalization. The parameters both obtained by the centred parametrization as by orthogonalization do not suffer from the singularity problem and thus there is no longer a problem when carrying out inference as we normally would.

So we can compute the score functions and thus the maximum likelihood estimator by evaluating the log-likelihood in the new parameters and deriving with respect to these parameters. Normally we would also use traditional tests of the null hypothesis of symmetry like the Score Test. For the expression of the test statistic consider $Y_1, \ldots, Y_n$. The Y_i's are independent and identically distributed with density $f(y|\theta)$, where θ is $b \times 1$. Consider the null hypothesis $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$. The formula for the test statistic is

$$T_S = S(\theta_0)^T \, (I(\theta_0))^{-1} \, S(\theta_0).$$

Because of the singularity, the factor $(I(\theta_0))^{-1}$ can not be determined in the original parametrization. By using the new parameters, we can calculate this test statistic.

Appendix A

Nederlandstalige samenvatting

In veel praktische toepassingen zijn datasets niet symmetrisch en niet normaal, ook al zouden we dat misschien graag zo hebben. De data zullen dus niet de populaire normale distributie volgen. In de 20^{ste} eeuw werd er een nieuwe familie van verdelingen ontwikkeld om met deze scheefheid om te gaan, de scheef-symmetrische verdelingen.

In deze thesis zullen we de scheef-symmetrische verdelingen onderzoeken en zullen we de mogelijke inferentiële problemen bekijken. Om dit te doen, heb ik vooral gebruik gemaakt van enkele belangrijke artikelen omtrent scheef-symmetrische verdelingen. Ik heb deze artikels geanalyseerd en heb de verschillende ideeën hieruit samengebracht. Ik heb ook de gegeven resultaten uitgewerkt om tot gelijkaardige uitkomsten te komen.

In het eerste hoofdstuk wordt er een historisch overzicht gegeven van de ontwikkeling van scheve verdelingen. Als eerste poging probeerde men de scheve data aan te passen zodat het de normale curve zou volgen. Wiskundigen zoals Edgeworth (1899) [27] werkten zo'n methode uit. Eén van de eersten die een nieuwe familie van distributies definieerde was Pearson (1895) [54] met zijn systeem van continue distributies bestaand uit vier parameters. Zijn methode om dit te bekomen wordt in detail uitgewerkt.Een zeer innovatief voorstel om niet-normale verdelingen te construeren werd gegeven door de Helguero (1909) [23, 24]. Ook hier zullen we wat beter kijken naar de constructie van zijn scheve verdelingen. Recentelijk stelde Azzalini (1985) [7] zijn algemeen bekend sheef-normale verdelingen voor, deze familie van distributies breidt die van de normale uit. Zijn waarschijnlijkheidsdichtheid is gegeven door

$$\phi(z;\delta) = 2\phi(z)\Phi(\delta z), \qquad -\infty < z < \infty,$$

waar ϕ de standaard Gaussische waarschijnlijkheidsdichtheid is en Φ de standaard Gaussische verdelingsfunctie. Om dit hoofdstuk te beëindigen worden nog enkele toepassingen van scheef-symmetrische verdelingen gegeven. Deze toepassingen komen uit verschillende velden en tonen aan hoe wijdverspreid het gebruik van scheefsymmetrische verdelingen is.

In het tweede hoofdstuk, kijken we naar de scheefsymetrische verdelingen vanuit een theoretisch standpunt. Meer bepaald, zullen we de scheef-normale en de scheef-t distributies als voorbeelden onderzoeken. De waarschijnlijkheidsdichtheid is hierboven al gegeven. De waarschijnlijkheidsdichtheid van de scheef-t verdelingen kunnen we op de volgende manier uitdrukken:

$$t(z;\delta,\nu) = 2t(z;\nu)T\left(\delta z\sqrt{\frac{\nu+1}{\nu+z^2}};\nu+1\right), \qquad -\infty < z < +\infty,$$

met t en T de standaard Student-t waarschijnlijkheidsdichtheid and verdelingsfunctie, respectievelijk, and ν staat voor het aantal vrijheidsgraden. In beide gevallen starten we met het geven van enkele eigenschappen met bewijs. Voor de scheef-normale familie gaan we verder met het geven van de momentgenererende functie en met het berekenen van de momenten. Tot slot wordt voor de scheef-normale verdelingen nog de uitgebreidde scheef-normale verdeling gegeven.Voor de scheef-t familie bepalen we de momenten door te stellen dat we een willekeurige scheef-t variable kunnen schrijven als de ratio

$$Y = \frac{Z}{\sqrt{\frac{U}{\nu}}}$$

met Z een standaard scheef-normale variabele en U volgt de Chi-kwadraatverdeling, Z en U zijn onafhankelijk.

In het derde en laatste hoofdstuk introduceren we de geassocieerde inferentiële problemen van de scheef-symmetrische verdelingen. Dit wordt opnieuw toegepast op de voorbeelden van de scheef-normale en de scheef-t distributies. In beide voorbeelden berekenen we de score functie and de Fisher information matrix. In het geval van de scheef-normale verdelingen is deze matrix singulier in de nabijheid van symmetrie wat leidt tot tragere convergentie snelheden, het zal meer bepaald zakken tot een $\sqrt[6]{n}$-rate. Om dit feit te bewijzen worden Lemma 3 van Rotnitzky *et al.* (2000) [59] en een Propositie bewezen door Chiogna (2005) [21] gegeven. Nadat het probleem tot stand is gebracht, worden er twee reparametrizaties gegeven om dit singulariteitsprobleem to overkomen. De eerste is de gecentreerde parametrisatie, als eerste voorgesteld door Azzalini (1985) [7]. De tweede is orthogonalisatie, voorgesteld door Hallin en Ley (2014) [39] wat gebruik maakt van het Gram-Schmidt orthogonalisatie proces. Het orthogonalisatie proces moet twee keer worden toegepast. De scheef-normale verdelingen hebben namelijk het zogenaamde dubbele singulariteitsprobleem. Bij beide reparametrisaties worden nieuwe parameters bekomen en de Fisher information matrix bepaald ten opzichte van deze parameters. In beide gevallen zal de Fisher information matrix niet langer singulier zijn. Voor de scheef-t familie, is de Fisher information matrix niet singulier en is er dus geen singulariteitsprobleem tenzij het aantal vrijheidsgraden ν naar oneindig gaat. Maar dan gaat de scheef-t distributie naar scheef-normale en daarvoor kennen we de oplossing al.

Appendix B

Set $y = \left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-1}(x - \mu_{\text{II}} + \sigma^* b\delta)$ and $y' = \frac{\partial y}{\partial \delta}$. We have

$$y' = -\sigma^* b^2\delta\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-2}(x - \mu_{\text{II}} + \sigma^* b\delta) + b\sigma^*\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-1}$$

$$\begin{aligned}y'' = &-\sigma^* b^2\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-2}(x - \mu_{\text{II}} + \sigma^* b\delta) + 2\sigma^{*2} b^4\delta^2\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-3}(x - \mu_{\text{II}} + \sigma^* b\delta)\\ &-\sigma^{*2} b^3\delta\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-2} - b^3\sigma^{*2}\delta\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-2}\end{aligned}$$

$$\begin{aligned}y''' = &\,2\sigma^{*2} b^4\delta\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-3}(x - \mu_{\text{II}} + \sigma^* b\delta) - \sigma^{*2} b^3\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-2}\\ &+ 4\sigma^{*2} b^4\delta\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-3}(x - \mu_{\text{II}} + \sigma^* b\delta) - 6\sigma^{*3} b^6\delta^2\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-4}(x - \mu_{\text{II}} + \sigma^* b\delta)\\ &+ 2\sigma^{*3} b^5\delta\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-3} - 2\sigma^{*2} b^3\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-2} + 4\sigma^{*3} b^5\delta^2\left(\sigma_{\text{II}} + \frac{1}{2}\sigma^* b^2\delta^2\right)^{-3}\end{aligned}$$

In (χ^*, δ^*) this becomes

$$\begin{aligned}y\big|_{(\chi^*,\delta^*)} &= \sigma^{*-1}(x - \mu^*) = z\\ y'\big|_{(\chi^*,\delta^*)} &= b\\ y''\big|_{(\chi^*,\delta^*)} &= -b^2 z\\ y'''\big|_{(\chi^*,\delta^*)} &= -3b^3\end{aligned}$$

Replacing these expressions in the equation at the end of the proof of Proposition 1, we get

$$\begin{aligned}\frac{\partial}{\partial\delta} j^{\theta_{\text{II}}}_{\delta\delta}(\chi^*, \delta^*) &= 3b^3 z + 3b^3 z + (-3b^2 z)b + 3(2bz)(-b^2) + z^3(2b^3 - b)\\ &= z^3(2b^3 - b) - 3b^3 z\end{aligned}$$

Bibliography

[1] Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model. *J. Econometrics*, 12:21–37.

[2] Arellano-Valle, R. B. (2010). On the information matrix of the multivariate skew-t model. *International Journal of Statistics*, 68(3):106–129.

[3] Arellano-Valle, R. B. and Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99:1362–1382.

[4] Arellano-Valle, R. B. and Azzalini, A. (2013). The centred parameterization and related quantities of the skew-t distribution. *Journal of Multivariate Analysis*, 113:73–90.

[5] Arellano-Valle, R. B. and Genton, M. G. (2010). Multivariate extended skew-t distributions and related families. *International Journal of Statistics*, 67(3):201–234.

[6] Arnold, B. C., Beaver, R. J., A.Groeneveld, R., and Meeker, W. Q. The non-truncated marginal of a truncated bivariate normal distribution. *Psychometrika*.

[7] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

[8] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32(2):159–188.

[9] Azzalini, A. (2006). Some recent developments in the theory of distributions and their applications. Atti Della XLIII Riunione della Società Italiana di Statistica, volume Riunione plenarie e specializzati. :51-64.

[10] Azzalini, A. (2013). *Skew-Normal and Related Families*. Cambridge University Press.

[11] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society : Series B*, 61(3):579–602.

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

[13] Azzalini, A. and Dalla Valle, A. The multivariate skew-normal distribution. *Biometrika*.

[14] Azzalini, A. and Genton, M. G. (2008). Robust likelihood methods based on the skew-"t" and related distributions. *International Statistical Review*, 76(1):106–129.

[15] Azzalini, A. and Regoli, G. (2012). The work of Fernando de Helguero on non-normality arising from selection. *Chilean Journal of Statistics*, 3(2):113–129.

[16] Barbé, L. (2010). *Francis Ysidro Edgeworth: A Portrait with Family and Friends*. Edward Elgar Publishin.

[17] Bertsekas, D. P. and Tsitsiklis, J. N. (2008). *Introduction to Probability*. Athena Scientific.

[18] Birnbaum, Z. W. (1950). Effect of linear truncation on a multinormal population. *Ann. Math. Statist.*, 21:272–279.

[19] Blasi, F. and Scarlatti, S. (2012). From normal vs skew-normal portfolios: Fsd and ssd rules. *Journal of Mathematical Finance*, 2:90–95.

[20] Brown, N. D. (2001). Reliability studies of the skew normal distribution. *Electronic Theses and Dissertations*.

[21] Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods and Applications*, 14(3):331–341.

[22] Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society*, 49:1–39.

[23] de Helguero, F. (1899). Sulla rappresentazione analitica delle curve statistiche. *Giornale degli Economisti*, 38:241–265.

[24] de Helguero, F. (1909). Sulla rappresentazione analitica delle curve abnormali. *In Castel- nuovo, G. (ed.)*, 3:288–299.

[25] De Roon, F. and Karehnke, P. (2016). A simple skewed distribution with asset pricing applications. *Review of Finance, Forthcoming*.

[26] Di Ciccio, T. J. and Monti, A. C. (2011). Inferential aspects of the skew t-distribution. *Quaderni di Statistica*, 13:1–21.

[27] Edgeworth, E. Y. (1899a). On the representation of statistics by mathematical formulÃę. *Journal of the Royal Statistical Society*, 62(2):373–385.

[28] Edgeworth, F. Y. (1886). The law of error and the elimination of chance. *Philosophical Magazine*, 21:308–324.

[29] Edgeworth, F. Y. (1899b). On the representation of statistics by mathematical formulae (part iii). *Journal of the Royal Statistical Society*, 62:373–385.

[30] Fechner, C. G. (1897). *Kollektivmasslehre*. Gottlieb Friedrich Lipps.

[31] Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. CRC Press.

[32] Genton, M. G., He, L., and Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics & Probability Letters*, 51:319–325.

[33] Genton, M. G. and Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, 57(2):389–401.

[34] Genton, M. G. and Thompson, K. R. (2004). Skew-elliptical time series with application to flooding risk. *The IMA Volumes in Mathematics and its Applications*, 45:169–185.

[35] Gibbons, J. F. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-gaussian distributions. *Applied Physics Letters*, 22:568–569.

[36] Haas, M. (2012). A note on the moments of the skew-normal distribution. *Economics Bulletin*, 32(4):3306–3312.

[37] Hald, A. (2004). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713 to 1935*. John Wiley & Sons, Inc.

[38] Hallin, M. and Ley, C. (2012). Skew-symmetric distributions and Fisher information - a tale of two densities. *Bernoulli*, 18(3):747–763.

[39] Hallin, M. and Ley, C. (2014). Skew-symmetric distributions and Fisher information: The double sin of the skew-normal. *Bernoulli*, 20(3):1432–1453.

[40] Hasan, A. M. (2013). A study of non-central skew t distributions and their applications in data analysis and change point detection. Thesis at Bowling Green State University.

[41] Heinrich, J. (2004). A guide to the Pearson type iv distribution. Thesis at University of Pennsylvania.

[42] John, S. (1982). The three-parameter two-piece normal family of distributions and its fitting. *Communications in Statistics Theory and Methods*, 11:879–885.

[43] Kim, H. (2006). On the distribution and its properties of the sum of a normal and a doubly truncated normal. *The Korean Communications in Statistics*, 13(2):225–266.

[44] Kim, H. and Mallick, B. (2003). Moments of random vectors with skew t distribution and their quadratic forms. *Statistics & Probability Letters*, 63(4):417–423.

[45] Kotz, S. and Vicari, D. (2005). Survey of developments in the theory of continuous skewed distributions. *International Journal of Statistics*, 68(2):225–261.

[46] Lee, L. (1993). Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. *Econometric Theory*, 9(3):413–430.

[47] Ley, C. (2014). Flexible modelling in statistics: past, present and future. *Journal de la Société Française de Statistique*, 156(1):76–96.

[48] Ley, C. and Paindaveine, D. On the singularity of multivariate skew-symmetric models. *Journal of Multivariate Analysis*, 101.

[49] Lloyd, G. E. (1983). Strain analysis using the shape of expected and observed continuous freqency distributions. *Journal of Structural Geology*, 5(3):225–231.

[50] Marchenko, Y. V. (2010). Multivariate skew-t distributions in econometrics and environmetrics (texas a&m university).

[51] Marchenko, Y. V. and Genton, M. G. (2012). A Heckman selection-t model. *Journal of the American Statistical Association*, 107:304–317.

[52] Ngunkeng, G. Statistical analysis of skew normal distribution and its applications.

[53] O'Hagan, A. and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63(1):201–203.

[54] Pearson, K. (1895a). Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, 186:343–414.

[55] Pearson, K. (1895b). Mathematical contributions to the theory of evolution. x. supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London. Series A*, 186:443–459.

[56] Pewsey, A. (2000). Problems of inference for azzalini's skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870.

[57] Pourahmadi, M. (2007). Construction of skew-normal random variables: Are they linear combinations of normal and half-normal? Thesis at Northern Illinois University.

[58] Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61(316):1184–1190.

[59] Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6(3):243–284.

[60] Stigler, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press.

[61] Student (1908). The probable error of a mean. *Biometrika*, 6(1):1–25.

[62] Tjetjep, A. and Seneta, E. (2006). Skewed normal variance-mean models for asset pricing and the method of moments. *International Statistical*, 74(1):109–126.

[63] Toth, Z. and Szentimrey, T. (1990). The binormal distribution: a distribution for representing asymmetrical but normal-like weather elements. *Journal of Climate*, 3:128–136.

[64] Wallis, K. F. (2014). The two-piece normal, binormal, or double gaussian distribution: its origin and rediscoveries. *Statistical Science*, 29(1):106–112.

[65] Weinstein, M. A. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6:104–105.