On quantile-based asymmetric family of distributions: properties and inference

Peer-reviewed author version

Gijbels, Irène; KARIM, Rezaul & VERHASSELT, Anneleen (2019) On quantile-based asymmetric family of distributions: properties and inference. In: INTERNATIONAL STATISTICAL REVIEW, 87(3), p. 471-504..

DOI: 10.1111/insr.12324
Handle: http://hdl.handle.net/1942/28220

On quantile-based asymmetric family of distributions: properties and inference*

Irène Gijbels[1], Rezaul Karim[1,2] and Anneleen Verhasselt[2]

February 28, 2018

[1] Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Celestijnenlaan 200B, Box 2400, B-3001 Leuven (Heverlee), Belgium.
[2] Center for Statistics, Hasselt University, Agoralaan-building D, B-3590 Diepenbeek, Belgium.

Abstract

In this paper we provide a detailed study of a general family of asymmetric densities. In the general framework we establish expressions for important characteristics of the distributions, and discuss estimation of the parameters via method-of-moments as well as maximum likelihood estimation. Asymptotic normality results for the estimators are provided. The results under the general framework are then applied to some specific examples of asymmetric densities. The use of the asymmetric densities is illustrated in real data analysis.

Keywords and phrases: asymmetric density, asymptotics, location parameter, maximum likelihood estimation, method-of-moments, quantiles.

1 Introduction

Although the normal or bell-shape density is the most standard reference density, together with the Student's-t density when heavier tails seem to be more appropriate, the symmetry of both densities makes them unsuitable for many applications. Investment return data and household income data are just a few examples of data that can only be described appropriately with asymmetric distributions. Simon (1955) realized the importance of asymmetric distributions in sociology, economics, and in many biological phenomena. Apart from the classical asymmetric distributions such as the log-normal, the chi-squared, and the Fisher distributions, there are many classes of non-classical asymmetric distributions that have been proposed in the literature.

* This research was supported by the FWO research project G.0826.15N (Flemish Science Foundation) and the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy). The first author gratefully acknowledges support from the GOA/12/014 project of the Research Fund KU Leuven. The third author acknowledges support from the Flemish Science Foundation (FWO research grant 1518917N), and from the Special Research Fund (Bijzonder Onderzoeksfonds) of Hasselt University.

One can distinguish some major different approaches for constructing asymmetric distributions. We briefly review these. Azzalini (1985, 1986, 2005) proposed a general class of skew distributions with probability density function

$$f_\lambda(y) = 2\Pi(\lambda y) f(y), \tag{1.1}$$

where $f(y)$ is a given density, symmetric around 0, and $\Pi(y)$ is an absolutely continuous distribution function with probability density function $\Pi'(y)$ that is symmetric around 0. The real-valued parameter λ determines together with f and Π the specific element in this class of asymmetric densities.

Another approach towards transforming a symmetric distribution into a skew distribution was proposed by Fernández and Steel (1998). For a given unimodal and symmetric around 0 density f and a scalar index $\gamma \in (0, +\infty)$, a density in this class of skew distributions is defined by

$$f_\gamma(y) = \frac{2}{\gamma + \frac{1}{\gamma}} \begin{cases} f(\gamma y) & \text{if} \quad y \le 0 \\ f(\frac{y}{\gamma}) & \text{if} \quad y > 0. \end{cases} \tag{1.2}$$

Arellano-Valle et al. (2005) proposed a general family of skew distributions that includes (1.2) as a special case. For a given symmetric around 0 density f, a real-valued parameter α, and positive asymmetric functions $a(\cdot)$ and $b(\cdot)$, a density in this family is defined as

$$f_\alpha(y) = \frac{2}{a(\alpha) + b(\alpha)} \begin{cases} f(\frac{y}{b(\alpha)}) & \text{if} \quad y \le 0 \\ f(\frac{y}{a(\alpha)}) & \text{if} \quad y > 0. \end{cases} \tag{1.3}$$

When taking $a(\alpha) = b(\alpha)$, the density $f_\alpha(y)$ reduces to a symmetric density.

For constructing an asymmetric density, Nassiri and Loris (2013) start from a given symmetric around 0 density f, and positive real parameters λ_1 and λ_2, and define

$$f_{\lambda_1,\lambda_2}(y) = \frac{2\lambda_1\lambda_2}{\lambda_1 + \lambda_2} \begin{cases} f(\lambda_1 y) & \text{if} \quad y \le 0 \\ f(\lambda_2 y) & \text{if} \quad y > 0. \end{cases} \tag{1.4}$$

For any $\lambda_1 = \lambda_2$ the density f_{λ_1,λ_2} is symmetric, with a special case $f_{\lambda_1,\lambda_2} = f$ when $\lambda_1 = \lambda_2 = 1$. When λ_1 is larger (respectively smaller) than λ_2 one gets a right-skew (respectively left-skew) density. Note that the family of densities in (1.2) is a special case of the more general family of Nassiri and Loris (2013) by taking $\lambda_1 = \gamma$ and $\lambda_2 = \frac{1}{\gamma}$. The Arellano-Valle et al. (2005) family given in (1.3) is also a special case of the Nassiri and Loris (2013) family with $\lambda_1 = \frac{1}{b(\alpha)}$ and $\lambda_2 = \frac{1}{a(\alpha)}$.

The family of asymmetric densities (1.4) is thus quite broad, and a detailed study of it has not been done yet. A first aim of this paper is to establish general properties for this interesting family of asymmetric densities. A first merit of studying the general family (1.4), is that several known asymmetric densities can be seen as special cases of it, with our study allowing to get probabilistic and statistical results on these. Secondly, and more importantly, this study (i) reveals new insights for existing asymmetric densities and fills in gaps in the literature; and (ii) provides a detailed study of properties of and inference for many new asymmetric densities. When it comes to estimation, we focus on a special setting by taking $\lambda_1 = 1 - \alpha$ and $\lambda_2 = \alpha \in (0, 1)$, which allows for sufficient modelling flexibility and avoids an additional parameter (to be estimated). In addition this parametrization has important specific merits. Firstly, the parameters μ and ϕ are orthogonal (see Remark 3.1). Secondly, symmetry of f_{λ_1,λ_2} is then equivalent to $\alpha = 0.5$, which facilitates for example the development of a test of symmetry. Thirdly, the location parameter of the family corresponds to the αth-quantile of the distribution, which makes the family very-well suited in studies where quantiles are a main focus, as opposed to a single mean parameter.

Apart from studying the probabilistic properties of the general family of asymmetric densities in (1.4), in Section 2.2, we also discuss estimation of the parameters in the reduced family ($\lambda_1 = 1 - \alpha$ and $\lambda_2 = \alpha$), and establish the asymptotic distributional properties of the estimators, in Section 3. In Sections 4 and 5 we apply the general results to the interesting special cases of asymmetric Laplace and normal densities. In Section 6 we illustrate the use of the discussed large class of families in a real data application. Some further discussions are provided in Section 7. The Supplemental Material contains additional results, including a small simulation study, proofs of some of the theoretical results of Sections 2.2 and 3, a further extension of the statistical estimation part, a study of asymmetric Student's-t and asymmetric logistic densities, and an additional real data example.

2 Quantile-based asymmetric family of densities

2.1 Location-scale asymmetric family

In (1.4) the reference symmetric density f is considered to be a standard version of the density in a location-scale family of densities, such as a standard normal density, a standard Laplace density, a standard Cauchy density. By introducing a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\phi > 0$, we obtain

$$f_{\lambda_1,\lambda_2}(y; \mu, \phi) = \frac{2\lambda_1\lambda_2}{\phi(\lambda_1 + \lambda_2)} \begin{cases} f\big(\lambda_1(\frac{\mu - y}{\phi})\big) & \text{if} \quad y \leq \mu \\ f\big(\lambda_2(\frac{y-\mu}{\phi})\big) & \text{if} \quad y > \mu. \end{cases} \tag{2.1}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}^+$. With $\mu = 0$ and $\phi = 1$ this reduces to (1.4).

In a recent review of existing asymmetric densities Jones (2015) classified these into four families. The family given in (2.1) is a member of what he termed as Family 3A asymmetric densities ("Transformation of Scale including Two-piece") which takes the form $f_S(y) = 2f(W^{-1}(y))$, with f a density in a location-scale family of symmetric densities. Indeed, by taking

$$W^{-1}(y) = \frac{\phi(\lambda_1 + \lambda_2)}{\lambda_1\lambda_2}\Big|\frac{y - \mu}{\phi}\Big|\Big\{\lambda_1\mathbb{I}(y \leq \mu) + \lambda_2\mathbb{I}(y > \mu)\Big\},$$

with $\mathbb{I}(A)$ denoting the indicator function (i.e. $\mathbb{I}(A) = 1$ if A holds, and 0 otherwise), we get the density in (2.1).

The following proposition formally states that the above family of densities constitutes a location-scale family, as soon as f is an element of a location-scale family.

Proposition 2.1. Suppose that f belongs to a location-scale family of symmetric densities. If $Y \sim f_{\lambda_1,\lambda_2}(\cdot; \mu, \phi)$, then, for any $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_0 + \beta_1 Y \sim f_{\lambda_1,\lambda_2}(\cdot; \beta_0 + \beta_1\mu, |\beta_1|\,\phi)$.

2.2 Properties of the asymmetric family of densities

Consider a random variable Y with density $f_{\lambda_1,\lambda_2}(\cdot;\mu,\phi)$ in (2.1). In this section we provide explicit expressions for (i) the cumulative distribution function of Y and the βth-quantile of Y (in Theorem 2.1); (ii) the central moments of Y (in Theorem 2.2); and (iii) the characteristic function of Y (in Theorem 2.3). Denote by F and F^{-1} respectively the cumulative distribution and quantile function of the standard symmetric around 0 density f. The symmetry property of f implies that $F(0) = 0.5$ and $F^{-1}(0.5) = 0$. The proofs of Theorems 2.1 and 2.2 are provided in the Supplemental Material.

Theorem 2.1. If Y is a random variable with asymmetric density $f_{\lambda_1,\lambda_2}(\cdot;\mu,\phi)$ as in (2.1), then the cumulative distribution function of Y equals

$$F_{\lambda_1,\lambda_2}(y;\mu,\phi) = \begin{cases} \frac{2\lambda_2}{\lambda_1+\lambda_2} F\big(\lambda_1(\frac{y-\mu}{\phi})\big) & \text{if} \quad y < \mu \\ \frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2} + \frac{2\lambda_1}{\lambda_1+\lambda_2} F\big(\lambda_2(\frac{y-\mu}{\phi})\big) & \text{if} \quad y \geq \mu, \end{cases} \tag{2.2}$$

and for any $\beta \in (0,1)$, the βth-quantile of Y is given by

$$F^{-1}_{\lambda_1,\lambda_2}(\beta) = \begin{cases} \mu + \frac{\phi}{\lambda_1} F^{-1}\big(\frac{\beta(\lambda_1+\lambda_2)}{2\lambda_2}\big) & \text{if} \quad \beta < \frac{\lambda_2}{\lambda_1+\lambda_2} \\ \mu + \frac{\phi}{\lambda_2} F^{-1}\Big(\frac{\beta(\lambda_1+\lambda_2)+(\lambda_1-\lambda_2)}{2\lambda_1}\Big) & \text{if} \quad \beta \geq \frac{\lambda_2}{\lambda_1+\lambda_2}, \end{cases} \tag{2.3}$$

with in particular

$$F^{-1}_{\lambda_1,\lambda_2}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right) = \mu. \tag{2.4}$$

Furthermore, $F^{-1}_{\lambda_1,\lambda_2}(0)$ is the minimum and $F^{-1}_{\lambda_1,\lambda_2}(1)$ is the maximum value of Y.

For the special standard setting that $\mu = 0$ and $\phi = 1$, expressions (2.2) and (2.3) can be found in Nassiri and Loris (2013).

We next investigate central moments of Y about μ. Generally, Y has finite rth order moment ($r \in \mathbb{R}$) if and only if a corresponding moment of the symmetric density f exists. This is formally stated in Theorem 2.2. Of special interest are the expressions for the mean $E(Y)$, the variance $\text{Var}(Y)$, the skewness and the kurtosis of Y, with the latter two defined as respectively

$$\gamma_{\text{sk}} = \frac{E\left[(Y-E(Y))^3\right]}{\{E[(Y-E(Y))^2]\}^{\frac{3}{2}}} \qquad \text{and} \qquad \gamma_{\text{ku}} = \frac{E\left[(Y-E(Y))^4\right]}{\left\{E\left[(Y-E(Y))^2\right]\right\}^2}. \tag{2.5}$$

Theorem 2.2. If Y is a random variable with asymmetric density $f_{\lambda_1,\lambda_2}(\cdot;\mu,\phi)$ as in (2.1), then the rth central moment of Y about μ , with $r \in \mathbb{R}$, is given by

$$E(Y-\mu)^r = \frac{\phi^r}{(\lambda_1+\lambda_2)}\Big[\frac{\lambda_1^{r+1} + (-1)^r \lambda_2^{r+1}}{\lambda_1^r \lambda_2^r}\Big]\mu_r,$$

where

$$\mu_r = 2\int_0^\infty s^r f(s) ds. \tag{2.6}$$

Furthermore, the mean, variance, skewsness and kurtosis are given by, respectively

$$E(Y) = \mu + \frac{\phi(\lambda_1 - \lambda_2)}{\lambda_1\lambda_2}\mu_1 \tag{2.7}$$

$$\mathrm{Var}(Y) \overset{\text{not.}}{=} V(Y) = \frac{\phi^2}{\lambda_1^2\lambda_2^2}[(\lambda_1 - \lambda_2)^2(\mu_2 - \mu_1^2) + \lambda_1\lambda_2\mu_2]$$

$$\gamma_{\text{sk}} = \frac{(\lambda_1 - \lambda_2)[(\lambda_1 - \lambda_2)^2(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3) + \lambda_1\lambda_2(2\mu_3 - 3\mu_1\mu_2)]}{[(\lambda_1 - \lambda_2)^2(\mu_2 - \mu_1^2) + \lambda_1\lambda_2\mu_2]^{\frac{3}{2}}} \tag{2.8}$$

and

$$\gamma_{\text{ku}} = \frac{(\lambda_1^5 + \lambda_2^5)\mu_4 - (\lambda_1 + \lambda_2)(\lambda_1 - \lambda_2)^2[4(\lambda_1^2 + \lambda_2^2)\mu_1\mu_3 - 6(\lambda_1^2 - \lambda_1\lambda_2 + \lambda_2^2)\mu_1^2\mu_2 + 3(\lambda_1 - \lambda_2)^2\mu_1^4]}{(\lambda_1 + \lambda_2)[(\lambda_1 - \lambda_2)^2(\mu_2 - \mu_1^2) + \lambda_1\lambda_2\mu_2]^2}. \tag{2.9}$$

Remark 2.1. Note from (2.8) and (2.9) that the skewness and the kurtosis do not depend on the parameters μ and ϕ, but only on λ_1 and λ_2, and on the moment-type quantities (μ_1, μ_2, μ_3 and μ_4) of the reference symmetric density f.

Theorem 2.3 provides the expression for the characteristic function of the asymmetric density (2.1).

Theorem 2.3. *If Y is a random variable with asymmetric density $f_{\lambda_1,\lambda_2}(\cdot;\mu,\phi)$ as in (2.1), then the characteristic function of Y is given by*

$$\varphi(t) = \frac{2e^{it\mu}}{(\lambda_1 + \lambda_2)}\Big[\lambda_2\varphi^+\big(-\frac{\phi t}{\lambda_1}\big) + \lambda_1\varphi^+\big(\frac{\phi t}{\lambda_2}\big)\Big],$$

where

$$\varphi^+(t) = \int_0^\infty e^{ity} f(y)dy. \tag{2.10}$$

Proof. The calculation is straightforward using changes of variables:

$$\begin{aligned} E[e^{itY}] &= \int_{-\infty}^{\infty} e^{ity} f_{\lambda_1,\lambda_2}(y;\mu,\phi)dy \\ &= \frac{2\lambda_2}{(\lambda_1 + \lambda_2)}e^{it\mu}\int_0^\infty e^{i(-\frac{\phi t}{\lambda_1})x} f(x)dx + \frac{2\lambda_1}{(\lambda_1 + \lambda_2)}e^{it\mu}\int_0^\infty e^{i(\frac{\phi t}{\lambda_2})z} f(z)dz \\ &= \frac{2e^{it\mu}}{(\lambda_1 + \lambda_2)}\Big[\lambda_2\varphi^+\big(-\frac{\phi t}{\lambda_1}\big) + \lambda_1\varphi^+\big(\frac{\phi t}{\lambda_2}\big)\Big]. \end{aligned}$$

□

It is of interest to look at the basic idea underlying the construction of the asymmetric density (2.1). With f a symmetric density, the basic idea is to introduce scale factors λ_1 and λ_2 in the positive and the negative orthants (with respect to μ) such that the resulting density retains the mode μ. From the expression of the cumulative distribution function in Theorem 2.1 it is easy to see that

$$\frac{P(Y > \mu)}{P(Y \le \mu)} = \frac{\lambda_1/(\lambda_1 + \lambda_2)}{\lambda_2/(\lambda_1 + \lambda_2)} = \frac{\lambda_1}{\lambda_2}, \tag{2.11}$$

or equivalently

$$\lambda_2\, P(Y > \mu) = \lambda_1\, P(Y \leq \mu).$$

from which it is clear that λ_1 and λ_2 control the allocation of mass to each side of the mode μ.

In this paper we focus in particular on the setting when $\lambda_1 = 1-\alpha$ and $\lambda_2 = \alpha \in (0,1)$. In this case the asymmetric density is given by

$$f_\alpha(y;\mu,\phi) = \frac{2\alpha(1-\alpha)}{\phi} \begin{cases} f\big((1-\alpha)(\frac{\mu-y}{\phi})\big) & \text{if} \quad y \leq \mu \\ f\big(\alpha(\frac{y-\mu}{\phi})\big) & \text{if} \quad y > \mu \end{cases} \tag{2.12}$$

and (2.11) reduces to

$$\frac{P(Y>\mu)}{P(Y\leq\mu)} = \frac{1-\alpha}{\alpha} \Longleftrightarrow \alpha\; P(Y>\mu) = (1-\alpha)\, P(Y \leq \mu),$$

with now only α controling the mass allocation in the density to the left and the right of the mode μ. From expression (2.2) in Theorem 2.1 we also know that, with $\lambda_1 = 1-\alpha$ and $\lambda_2 = \alpha$, the quantile of order α is equal to μ, the location parameter. As such, the family of asymmetric densities in (2.12) is tailored towards quantiles, with the αth-quantile its location parameter.

Remark 2.2. For the well known family of densities (1.2) studied by Fernández and Steel (1998), which corresponds to taking $\lambda_1 = \gamma$ and $\lambda_2 = \frac{1}{\gamma}$, the ratio (2.11) becomes γ^2. One may be tempted to think of (2.12) with $\mu = 0$ and $\phi = 1$ as a reparametrization of (1.2), by simply matching the expression for $P(Y > \mu)/P(Y \leq \mu)$, which yields $\alpha = \frac{1}{1+\gamma^2}$, a one-to-one transformation between $\mathbb{R}^+$ to (0,1). This is of course not true. Using the above transformation, the density (2.12) yields the density

$$f_\gamma(y) = \frac{2\gamma^2}{(1+\gamma^2)^2} \begin{cases} f(\frac{\gamma^2 y}{1+\gamma^2}) & \text{if} \quad y \leq 0 \\ f(\frac{y}{1+\gamma^2}) & \text{if} \quad y > 0 \end{cases}$$

which is clearly different from the density in (1.2).

From the previous theorems, we easily deduce the following properties for the family of asymmetric densities in (2.12).

Corollary 2.1. If Y is a random variable with asymmetric density $f_\alpha(\cdot;\mu,\phi)$ in (2.12), then

(i). the cumulative distribution function of Y is given by

$$F_\alpha(y;\mu,\phi) = \begin{cases} 2\alpha F\big((1-\alpha)(\frac{y-\mu}{\phi})\big) & \text{if} \quad y < \mu \\ 2\alpha - 1 + 2(1-\alpha)F\big(\alpha(\frac{y-\mu}{\phi})\big) & \text{if} \quad y \geq \mu. \end{cases}$$

and for any $\beta \in (0,1)$, the βth-quantile of Y is

$$F_\alpha^{-1}(\beta) = \begin{cases} \mu + \frac{\phi}{1-\alpha} F^{-1}\big(\frac{\beta}{2\alpha}\big) & \text{if} \quad \beta < \alpha \\ \mu + \frac{\phi}{\alpha} F^{-1}\Big(\frac{1+\beta-2\alpha}{2(1-\alpha)}\Big) & \text{if} \quad \beta \geq \alpha, \end{cases}$$

with in particular $F_\alpha^{-1}(\alpha) = \mu$.

(ii). Furthermore, the rth central moment of Y about μ, with $r \in \mathbb{R}$, is

$$E(Y-\mu)^r = \phi^r \Big[\frac{(1-\alpha)^{r+1} + (-1)^r \alpha^{r+1}}{\alpha^r (1-\alpha)^r}\Big] \mu_r = \phi^r k_r$$

where we denoted

$$k_r = \Big[\frac{(1-\alpha)^{r+1} + (-1)^r \alpha^{r+1}}{\alpha^r (1-\alpha)^r}\Big] \mu_r. \tag{2.13}$$

In particular, the mean, variance, skewsness and kurtosis of Y are, respectively

$$E(Y) = \mu + \frac{\phi(1-2\alpha)\mu_1}{\alpha(1-\alpha)} \tag{2.14}$$

$$V(Y) = \frac{\phi^2}{\alpha^2(1-\alpha)^2}[(1-2\alpha)^2(\mu_2-\mu_1^2) + \alpha(1-\alpha)\mu_2]. \tag{2.15}$$

$$\gamma_{\text{sk}} = \frac{(1-2\alpha)[(1-2\alpha)^2(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3) + \alpha(1-\alpha)(2\mu_3 - 3\mu_1\mu_2)]}{[(1-2\alpha)^2(\mu_2-\mu_1^2) + \alpha(1-\alpha)\mu_2]^{\frac{3}{2}}}, \tag{2.16}$$

and

$$\gamma_{\text{ku}} = \frac{\big[(1-\alpha)^5 + \alpha^5\big]\mu_4 - (1-2\alpha)^2\big[4(1-2\alpha+2\alpha^2)\mu_1\mu_3 - 6(1-3\alpha+3\alpha^2)\mu_1^2\mu_2 + 3(1-2\alpha)^2\mu_1^4\big]}{[(1-2\alpha)^2(\mu_2-\mu_1^2) + \alpha(1-\alpha)\mu_2]^2}.$$

(iii). The characteristic function of Y is given by

$$\varphi(t) = 2e^{it\mu}\Big[\alpha\varphi^+\Big(-\frac{\phi t}{1-\alpha}\Big) + (1-\alpha)\varphi^+\Big(\frac{\phi t}{\alpha}\Big)\Big].$$

In the sequel we focus on the family of quantile-based asymmetric densities in (2.12).

3 Parameter estimation in the quantile-based asymmetric family of densities

Let $Y_1, \ldots, Y_n$ be an i.i.d. sample from Y with density (2.12). For a given reference symmetric around 0 density f and a given index-parameter $\alpha \in (0,1)$, the asymmetric density (2.12) depends on two parameters μ and ϕ. A first aim in this section is to provide estimators for the parameters μ and ϕ. A secondary aim is also to consider estimation of the index-parameter α. We discuss maximum likelihood estimation as well as estimation via a method-of-moments.

3.1 Method-of-moments estimation

When α is known, we consider the first two moments of Y to obtain method-of-moments estimators for μ and ϕ. Expressions (2.14) and (2.15) lead to

$$\begin{cases} E(Y) = \mu + k_1\phi \\ E(Y^2) = \mu^2 + 2\mu k_1 \phi + k_2\phi^2 \end{cases} \tag{3.1}$$

where k_r is defined in (2.13), and specifically $k_1 = \frac{1-2\alpha}{\alpha(1-\alpha)}\mu_1$ and $k_2 = \frac{(1-\alpha)^3+\alpha^3}{\alpha^2(1-\alpha)^2}\mu_2$.

We need to invert the system of equations in (3.1). Squaring the first equation and subtracting it from the second equation we get

$$E(Y^2) - (E(Y))^2 = (k_2 - k_1^2)\phi^2 \iff \phi^2 = \frac{1}{(k_2 - k_1^2)}\{E(Y^2) - (E(Y))^2\}. \tag{3.2}$$

Substituting this expression for ϕ into the first equation in (3.1) leads to the inverted system:

$$\begin{cases} \mu = E(Y) - \frac{k_1}{\sqrt{k_2 - k_1^2}}\sqrt{E(Y^2) - (E(Y))^2} \\ \phi = \frac{1}{\sqrt{k_2 - k_1^2}}\sqrt{E(Y^2) - (E(Y))^2}. \end{cases} \tag{3.3}$$

Note from (3.2) that $k_2 - k_1^2 = \phi^{-2}V(Y) > 0$.

Based on an i.i.d. sample $Y_1, \ldots, Y_n$ from Y, method-of-moments estimators for μ and ϕ are obtained by replacing the population moment $E(Y^j); j = 1, 2, \ldots$ in the inverted system of equations (3.3) by their empirical counterparts

$$M_j = \frac{1}{n}\sum_{i=1}^{n} Y_i^j \qquad j = 1, 2, \ldots$$

which then leads to the method-of-moments estimators

$$\begin{cases} \widehat{\mu}_n = M_1 - \frac{k_1}{\sqrt{k_2 - k_1^2}}\sqrt{M_2 - M_1^2} \\ \widehat{\phi}_n = \frac{1}{\sqrt{k_2 - k_1^2}}\sqrt{M_2 - M_1^2}. \end{cases} \tag{3.4}$$

When α is unknown, we have to consider a third population moment. Recall from its definition (see (2.5)) that the skewness depends on the three first population moments. From (2.16) we can also see that for a density (2.12), the skewness is a known function of α, say $h(\alpha)$. A possible approach would thus be to estimate the skewness by replacing the three first population moments by their empirical counterparts M_j, $j = 1, 2, 3$, leading to an empirical estimate $\widehat{\gamma}_{\text{sk}}$ for the skewness. Assuming that the function h is invertible (i.e., $h^{-1}(\gamma_{sk})$ exists) then the index-parameter α can be estimated by solving

$$\widehat{\alpha}_n^{(1)} = h^{-1}(\widehat{\gamma}_{\text{sk}}). \tag{3.5}$$

Alternatively, we can exploit the fact that for the density (2.12), it holds that $P\{Y \leq \mu\} = F_\alpha(\mu; \mu, \phi) = \alpha$. This leads to the estimator

$$\widehat{\alpha}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(Y_i \leq \widehat{\mu}_n), \tag{3.6}$$

with $\widehat{\mu}_n$ as in (3.4). Since this estimator depends on $\widehat{\mu}_n$, method-of-moments estimators for (μ, ϕ, α) are obtained via an iterative procedure. The estimate of α from equation (3.5) could be a good initial value of $\widehat{\alpha}_n$ for the iteration procedure.

3.2 Maximum likelihood estimation

The likelihood function for the full vector of parameters (μ, ϕ, α) for density (2.12) is

$$L_n(\mu,\phi,\alpha) = \Big[\frac{2\alpha(1-\alpha)}{\phi}\Big]^n \prod_{i=1}^{n} \Big[f\Big((1-\alpha)\big(\frac{\mu - Y_i}{\phi}\big)\Big)\Big]^{\mathbb{I}(Y_i\le\mu)} \times \Big[f\Big(\alpha\big(\frac{Y_i-\mu}{\phi}\big)\Big)\Big]^{\mathbb{I}(Y_i>\mu)},$$

leading to the log-likelihood function

$$\ln[L_n(\mu,\phi,\alpha)] = n\ln[2\alpha(1-\alpha)] - n\ln(\phi) + \sum_{i=1}^{n}\mathbb{I}(Y_i\le\mu)\ln\Big[f\Big((1-\alpha)\big(\frac{\mu-Y_i}{\phi}\big)\Big)\Big] + \sum_{i=1}^{n}\mathbb{I}(Y_i>\mu)\ln\Big[f\Big(\alpha\big(\frac{Y_i-\mu}{\phi}\big)\Big)\Big]. \tag{3.7}$$

The maximum likelihood estimator (MLE) of $\boldsymbol{\theta} = (\mu,\phi,\alpha)^T$ is obtained as a solution to the problem $\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ln[L_n(\mu,\phi,\alpha)]$ where $\boldsymbol{\Theta} = \mathbb{R}\times\mathbb{R}^+\times(0,1)$ is the parameter space of $\boldsymbol{\theta}$. We assume throughout that the function f is a differentiable function.

If α is known and equals 0.5, then $\mu = F_\alpha^{-1}(\alpha)$, i.e. the median of Y. In that case the log-likelihood function is differentiable, with respect to μ and ϕ, and maximization is straightforward.

If α is known but $\alpha \neq 0.5$, or α is not known, then maximizing the log-likelihood is not straightforward since the log-likelihood function is not differentiable with respect to the parameter μ at the points $\mu = Y_i$. At points $\mu \notin \{Y_1,\dots,Y_n\}$ the log-likelihood function is differentiable with respect to μ. In contrast, the log-likelihood function is differentiable with respect to the parameter ϕ and the index-parameter α, at all points of the domain. There is a package `fmincon` in MATLAB software for solving linear and nonlinear constraints minimization problems with excellent (close to true value) initial values. Ardalan et al. (2012) showed with a numerical example that this package may fail to find the global maximum value, and proposed an algorithm to find the MLE for a two-piece normal-Laplace distribution. We can generalize this algorithm to find the MLE for the considered asymmetric family of densities. Assume throughout this section that the reference symmetric density f is unimodal.

Let $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ denote the order statistics associated to the i.i.d. sample $Y_1, Y_2, ..., Y_n$ from Y. Denoting with $\widehat{\mu}$ the MLE of μ, and assuming that $Y_{(1)} \le \widehat{\mu} \le Y_{(n)}$, we can proceed as follows

(i) given $\phi \in (0,\infty)$ and $\alpha \in (0,1)$, the function $g_1(\mu) \equiv \ln[L_n(\mu,\phi,\alpha)]$ is a concave function of $\mu \in (-\infty,\infty)$;

(ii) given $\mu \in (Y_{(1)}, Y_{(n)})$ and $\alpha \in (0,1)$, let $\eta = \frac{1}{\phi}$, the function $g_2(\eta) \equiv \ln[L_n(\mu,\phi,\alpha)]$ is a concave function of $\eta \in (0,\infty)$;

(iii) given $\mu \in (Y_{(1)}, Y_{(n)})$ and $\phi \in (0,\infty)$, the function $g_3(\alpha) \equiv \ln[L_n(\mu,\phi,\alpha)]$ is a concave function of $\alpha \in (0,1)$.

Denote the left-hand derivative and right-hand derivative of the concave function $g_1(\mu)$ by $g'_{1-}(\mu)$ and $g'_{1+}(\mu)$, respectively. Let m be the first index in the set $\{1,\dots,n\}$ for which $g'_{1+}(Y_{(m)}) \le 0$, i.e $g'_{1+}(Y_{(m-1)}) > 0$. Then the following two cases can happen. (i) If $g'_{1+}(Y_{(m)})) = 0$ and $g'_{1-}(Y_{(m)})) = 0$, then the function $g_1(\mu)$ reaches its maximum at the point $Y_{(m)}$, i.e. $\widehat{\mu} = Y_{(m)}$. (ii) If $g'_{1+}(Y_{(m)}) \leq 0$ and $g'_{1-}(Y_{(m)})) < 0$, it follows that $Y_{(m-1)} < \widehat{\mu} < Y_{(m)}$, since $g'_{1+}(Y_{(m-1)}) > 0$. But in this case $g_1(\mu)$ is differentiable at $\widehat{\mu}$, with

$$g'_1(\widehat{\mu}) = g'_{1-}(\widehat{\mu}) = g'_{1+}(\widehat{\mu}) = \frac{\partial}{\partial\mu}\sum_{i=1}^{m-1}\ln\Big[f\Big((1-\alpha)\big(\frac{\mu - Y_{(i)}}{\phi}\big)\Big)\Big]\Big|_{\mu=\widehat{\mu}}$$
$$+\frac{\partial}{\partial\mu}\sum_{i=m}^{n}\ln\Big[f\Big(\alpha\big(\frac{Y_{(i)}-\mu}{\phi}\big)\Big)\Big]\Big|_{\mu=\widehat{\mu}}. \tag{3.8}$$

Setting $g'_1(\widehat{\mu}) = 0$ based on equation (3.8), then leads to the MLE of μ in that second case. Overall, denote the MLE by $\widehat{\mu} = \widehat{\mu}(\phi, \alpha)$.

Clearly, maximizing $\ln[L_n(\mu, \phi, \alpha)]$ over $\phi \in (0, \infty)$ is equivalent to maximizing the function $g_2(\eta)$ over $\eta \in (0, \infty)$, where $\eta = \frac{1}{\phi}$. For given (μ, α), the MLE of ϕ is $\widehat{\phi}(\mu, \alpha)$ which is obtained by solving the equation $\frac{\partial}{\partial\phi}\ln[L_n(\mu, \phi, \alpha)] = 0$. Similarly, for given (μ, ϕ), the MLE of α is $\widehat{\alpha}(\mu, \phi)$ obtained by solving the equation $\frac{\partial}{\partial\alpha}\ln[L_n(\mu, \phi, \alpha)] = 0$.

From the above it is clear that the maximum likelihood (ML) method results into an iterative procedure, which needs of course some starting values. Since for given index-parameter α, the parameter μ is nothing but the αth-quantile of Y, we consider as plausible starting values for (α, μ) the values $(\alpha_0^{(j)}, \mu_0^{(j)}) = (\frac{j}{n}, Y_{(j)})$, for $j = 1, \ldots, n$. For a given starting value $(\alpha_0^{(j)}, \mu_0^{(j)})$, we then first calculate $\widehat{\phi}(\mu_0^{(j)}, \alpha_0^{(j)})$, leading to starting values for the whole parameter vector (μ, ϕ, α). We then iterate until convergence of the procedure, to get to a first MLE for (μ, ϕ, α), which we denote by $(\widehat{\mu}^{(j)}, \widehat{\phi}^{(j)}, \widehat{\alpha}^{(j)})$. Running this computation for all n starting values, and calculating $\ln\Big[L_n\Big(\widehat{\mu}^{(j)}, \widehat{\phi}^{(j)}, \widehat{\alpha}^{(j)}\Big)\Big]$ for $j = 1, \ldots, n$, we select as the MLE of (μ, ϕ, α), the value $(\widehat{\mu}^{(k)}, \widehat{\phi}^{(k)}, \widehat{\alpha}^{(k)})$, for which $\ln\Big[L_n\Big(\widehat{\mu}^{(k)}, \widehat{\phi}^{(k)}, \widehat{\alpha}^{(k)}\Big)\Big]$ is maximal among the n values.

Obtaining the maximum likelihood estimators is computationally more involved than calculating the method-of-moments estimators. See also Section 3.3.3 for a further discussion on and comparison of both estimation methods.

3.3 Asymptotic behaviour of the parameter estimators

We next investigate the asymptotic properties of both, the method-of-moments estimators and the maximum likelihood estimators. We distinguish between the cases that the index-parameter α is given, and the case that it is not given and needs to be estimated. Allowing some mild notational ambiguity, we denote in the former case the unknown parameter vector by $\boldsymbol{\theta} = (\mu, \phi)^T$ and its true value by $\boldsymbol{\theta}_0 = (\mu_0, \phi_0)^T$; whereas in the latter case this is $\boldsymbol{\theta} = (\mu, \phi, \alpha)^T$ and $\boldsymbol{\theta}_0 = (\mu_0, \phi_0, \alpha_0)^T$ respectively.

The proofs of Theorems 3.1 and 3.4 are deferred to the Appendix, whereas the proofs of all other results are provided in the Supplemental Material.

3.3.1 Asymptotic behaviour of the method-of-moments estimator

Denote the true moments of Y by

$$\mu_{j,Y} = \mu_{j,Y}(\boldsymbol{\theta}_0) \equiv E(Y^j), \quad j = 1, 2, 3, 4,$$

and use the notation

$$\mu^*_{3,Y} = \mu^*_{3,Y}(\boldsymbol{\theta}_0) \equiv P\{Y \leq \mu_0\} = E\left[\mathbb{I}\{Y \leq \mu_0\}\right] = \alpha_0.$$

The method-of-moments estimator for (μ, ϕ), when the index-parameter α is known, is given in (3.4). In case α is unknown, the additional estimator in (3.6) is used, and the estimation procedure becomes an iterative one. In both cases, and for notational ease, denote the method-of-moments (MoM) estimator by $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})}$. Theorems 3.1 and 3.2 state the asymptotic normality result for $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})}$, in case the index-parameter is known, respectively not known. Herein N_k stands for a k-variate normal distribution.

Theorem 3.1. (*Known index-parameter* α)
If $\mu_{4,Y} = E(Y^4) < \infty$, then

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_2(\mathbf{0}, \Gamma(\boldsymbol{\theta}_0)) \qquad \text{as} \quad n \to \infty,$$

with

$$\Gamma(\boldsymbol{\theta}) = \frac{\phi^2}{4}\frac{1}{(k_1^2 - k_2)^2} \times \begin{bmatrix} k_1^2 k_4 - k_1^2 k_2^2 - 4\,k_1 k_2 k_3 + 4\,k_2^3 & 2\,k_1^3 k_2 + 2\,k_1^2 k_3 - 5\,k_1 k_2^2 - k_1 k_4 + 2\,k_2 k_3 \\ 2\,k_1^3 k_2 + 2\,k_1^2 k_3 - 5\,k_1 k_2^2 - k_1 k_4 + 2\,k_2 k_3 & k_4 - 4\,k_1^4 + 8\,k_1^2 k_2 - 4\,k_1 k_3 - k_2^2 \end{bmatrix}, \tag{3.9}$$

where k_r, for $r = 1, 2, 3, 4$, is as in (2.13).

Theorem 3.2. (*Unknown index-parameter* α)
Denote by $\boldsymbol{\mu}_Y(\boldsymbol{\theta}) = (\mu_{1,Y}(\boldsymbol{\theta}), \mu_{2,Y}(\boldsymbol{\theta}), \mu^*_{3,Y}(\boldsymbol{\theta}))^T$ the vector of population moments and by $\boldsymbol{M}_n = (M_1, M_2, M_3^*)^T$ with $M_3^* = n^{-1}\sum_{i=1}^n \mathbb{I}\{Y_i \leq \mu\}$. Further, let $\bar{\boldsymbol{m}}_n(\boldsymbol{\theta}) = \boldsymbol{M}_n - \boldsymbol{\mu}_Y(\boldsymbol{\theta}) = \frac{1}{n}\sum\limits_{i=1}^{n} \boldsymbol{m}(Y_i, \boldsymbol{\theta})$ is a vector of moment restrictions which is non-differentiable with respect to the unknown parameter, and denote the method-of-moments estimator by $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})}$. Assume that

(A1) $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})}$ is a consistent estimator of the true parameter $\boldsymbol{\theta}_0 = (\mu_0, \phi_0, \alpha_0)^T$ that satisfies $\bar{\boldsymbol{m}}_n(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})}) = \mathbf{0}$.

(A2) $E[\bar{\boldsymbol{m}}_n(\boldsymbol{\theta})]$ is differentiable at $\boldsymbol{\theta}_0$ with matrix of first order derivatives $\mathbf{M}$ such that $\mathbf{M}$ is nonsingular.

(A3) Denote $v_n(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}[\boldsymbol{m}(Y_i, \boldsymbol{\theta}_0) - E(\boldsymbol{m}(Y_i, \boldsymbol{\theta}_0))]$. Assume that $\{v_n(\cdot) : n \geq 1\}$ is stochastically equicontinuous.

(A4) The matrix $\mathbf{S} = E[\boldsymbol{m}(Y, \boldsymbol{\theta}_0)\boldsymbol{m}(Y, \boldsymbol{\theta}_0)^T]$ exists.

Then we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_3(\mathbf{0}, \mathbf{M}^{-1}\mathbf{S}(\mathbf{M}^{-1})^T) \qquad \text{as} \quad n \to \infty, \tag{3.10}$$

where $\mathbf{S} = E[\boldsymbol{m}(Y, \boldsymbol{\theta}_0)\boldsymbol{m}(Y, \boldsymbol{\theta}_0)^T]$.

Note that we have an explicit expression for the asymptotic variance-covariance matrix $\Gamma(\boldsymbol{\theta}_0)$ in the known index-parameter case.

3.3.2 Asymptotic behaviour of the maximum likelihood estimators

The true value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta} = (\mu, \phi, \alpha)^T$ is estimated by the MLE $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} = (\widehat{\mu}_n^{(\mathrm{MLE})}, \widehat{\phi}_n^{(\mathrm{MLE})}, \widehat{\alpha}_n^{(\mathrm{MLE})})^T$ obtained as a maximizer of (3.7).

One of the requirements for applying asymptotic normality results from standard maximum likelihood theory is that the objective function (the log-likelihood) is twice continuous differentiable. However, the asymmetric density has a non-differentiable peak at the mode μ, and hence the log-likelihood function $\ln[L_n(\boldsymbol{\theta})]$ is non-differentiable at the points $Y_i = \mu$. Therefore, the standard maximum likelihood asymptotic theory is not directly applicable when one of the unknown parameters is μ, and one has to use special asymptotic results which account for such an irregularity in the statistical model. By relying however on fundamental results, such as Theorems 2.1 and 7.1 of Newey and McFadden (1994) and Theorem 3 as well as its corollary in Huber (1967) we can still establish the usual asymptotic properties for our maximum likelihood estimators. The following assumptions are needed.

Assumptions:

(B1) Let $\boldsymbol{\Theta}_R = [-\mu_u, \mu_u] \times [\phi_l, \phi_u] \times [\alpha_l, \alpha_u]$, where $|\mu_u| < \infty$, $0 < \phi_l \leq \phi \leq \phi_u < \infty$, and $0 < \alpha_l \leq \alpha \leq \alpha_u < 1$, be a compact subset of $\boldsymbol{\Theta}$, and assume that $\boldsymbol{\theta}_0 \in \mathring{\boldsymbol{\Theta}}_R$, with $\mathring{\boldsymbol{\Theta}}_R$ the interior of $\boldsymbol{\Theta}_R$.

(B2) $\int_0^\infty \big|\ln f(s)\big| f(s) ds < \infty$; where $f(s)$ is the reference symmetric density.

(B3) $\gamma_r = \int_0^\infty s^{r-1} \cdot \frac{(f'(s))^2}{f(s)} ds < \infty \quad$ for $r = 1, 2, 3$.

(B4) $\lim\limits_{s\to\infty} s f(s) = 0$ or $\int_0^\infty s f'(s) ds = -\frac{1}{2}$.

Theorem 3.3 guarantees the consistency of the maximum likelihood estimator, whereas Theorem 3.4 establishes the asymptotic normality result.

Theorem 3.3. Under Assumptions (B1) and (B2), the MLE $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} = (\widehat{\mu}_n^{(\mathrm{MLE})}, \widehat{\phi}_n^{(\mathrm{MLE})}, \widehat{\alpha}_n^{(\mathrm{MLE})})^T$ of $\boldsymbol{\theta}$ is (weakly) consistent. That is, $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} \xrightarrow{P} \boldsymbol{\theta}_0$, as $n \to \infty$.

Before deriving the asymptotic distribution of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}$, we establish some result on the score vector (in Proposition 3.1) and subsequently derive the Fisher information matrix in Proposition 3.2.

Proposition 3.1. If Assumption (B4) holds, then the expectation with respect to the true underlying distribution of the score vector for Y, denoted by $\frac{\partial \ln f_\alpha(Y;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, is zero, i.e.

$$E\Big[\frac{\partial \ln f_\alpha(Y;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}.$$

Proposition 3.2. Suppose Assumptions (B3) and (B4) hold. Then the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}) = \Big[E\Big\{\frac{\partial}{\partial\theta_i} \log f_\alpha(Y;\boldsymbol{\theta}) \cdot \frac{\partial}{\partial\theta_j} \log f_\alpha(Y;\boldsymbol{\theta})\Big\}\Big]_{i,j=1,2,3}$ for $\boldsymbol{\theta} = (\mu, \phi, \alpha)^T$ is

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{2\alpha(1-\alpha)\gamma_1}{\phi^2} & 0 & -\frac{2\gamma_2}{\phi} \\ 0 & \frac{1}{\phi^2}(2\gamma_3 - 1) & -\frac{(1-2\alpha)(2\gamma_3-1)}{\alpha(1-\alpha)\phi} \\ -\frac{2\gamma_2}{\phi} & -\frac{(1-2\alpha)(2\gamma_3-1)}{\alpha(1-\alpha)\phi} & \frac{[\alpha^3+(1-\alpha)^3]2\gamma_3-(1-2\alpha)^2}{\alpha^2(1-\alpha)^2} \end{bmatrix}. \tag{3.11}$$

Remark 3.1.

1. Note from the Fisher information matrix in (3.11) that the parameters μ and ϕ are always orthogonal (see Cox and Reid, 1987). As consequences we mention: (i) the ML estimates of μ and ϕ are asymptotically independent, (ii) the asymptotic variance for estimating μ is the same whether ϕ is known or unknown. In addition this orthogonality property may lead to simpler numerical determination of the ML estimates for (μ, ϕ).

2. Note that the quantity γ_r, defined in Assumption (B3), is always a positive real number.

Theorem 3.4. Suppose Assumptions (B1)—(B4) hold. Then the MLE $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}$ is asymptotically normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $[\mathcal{I}(\boldsymbol{\theta}_0)]^{-1}$:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_3(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1}) \qquad \text{as} \quad n \to \infty,$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix given in (3.11), with inverse

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{\gamma_3 \phi^2}{2\alpha(1-\alpha)(\gamma_1\gamma_3-\gamma_2^2)} & \frac{(1-2\alpha)\gamma_2\phi^2}{2\alpha(1-\alpha)(\gamma_1\gamma_3-\gamma_2^2)} & \frac{\gamma_2\phi}{2(\gamma_1\gamma_3-\gamma_2^2)} \\ \frac{(1-2\alpha)\gamma_2\phi^2}{2\alpha(1-\alpha)(\gamma_1\gamma_3-\gamma_2^2)} & [\mathcal{I}(\boldsymbol{\theta})^{-1}]_{22} & \frac{(1-2\alpha)\gamma_1\phi}{2(\gamma_1\gamma_3-\gamma_2^2)} \\ \frac{\gamma_2\phi}{2(\gamma_1\gamma_3-\gamma_2^2)} & \frac{(1-2\alpha)\gamma_1\phi}{2(\gamma_1\gamma_3-\gamma_2^2)} & \frac{\alpha(1-\alpha)\gamma_1}{2(\gamma_1\gamma_3-\gamma_2^2)} \end{bmatrix}$$

where

$$[\mathcal{I}(\boldsymbol{\theta})^{-1}]_{22} = \frac{(6\,\alpha^2\gamma_1\gamma_3 + 2\,\gamma_2{}^2\alpha^2 - 4\,\alpha^2\gamma_1 - 6\,\alpha\,\gamma_1\gamma_3 - 2\,\gamma_2{}^2\alpha + 4\,\alpha\,\gamma_1 + 2\,\gamma_1\gamma_3 - \gamma_1)\,\phi^2}{2\alpha(1-\alpha)(2\gamma_3 - 1)(\gamma_1\gamma_3 - \gamma_2^2)}.$$

If α is known, then the asymptotic variance-covariance matrix of the MLE $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n^{(\mathrm{MLE})}, \widehat{\phi}_n^{(\mathrm{MLE})})^T$ of $(\mu, \phi)^T$ is

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{\phi^2}{2\alpha(1-\alpha)\gamma_1} & 0 \\ 0 & \frac{\phi^2}{2\,\gamma_3 - 1} \end{bmatrix}. \tag{3.12}$$

If ϕ and α are both known, then the asymptotic variance (abbreviated AVar) of $\widehat{\mu}_n^{(\mathrm{MLE})}$ is $\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MLE})}) = \frac{\phi^2}{2\alpha(1-\alpha)\gamma_1}$.

Note that for the ML estimation method we also have a closed-form expression for the asymptotic variance-covariance matrix when the index-parameter α is unknown. Since in case α known we have closed-form expressions for the asymptotic variance-covariance for both estimation methods, we can compare the performances of the estimators. See Section 3.3.3.

3.3.3 Comparison between asymptotic properties of MoM and MLE estimators

Suppose that the index-parameter α is given (fixed). From the expressions for the asymptotic variances of the method-of-moments and the maximum likelihood estimators for μ

and ϕ, provided in (3.9) and (3.12), respectively, we obtain that

$$\mathrm{RAVar}(\widehat{\mu}_n) \equiv \frac{\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MoM})})}{\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MLE})})} = \frac{\alpha\,(1-\alpha)\,(k_1^2 k_4 - k_1^2 k_2^2 - 4\,k_1 k_2 k_3 + 4\,k_2^3)\,\gamma_1}{2\,(k_2 - k_1^2)^2},$$

and

$$\mathrm{RAVar}(\widehat{\phi}_n) \equiv \frac{\mathrm{AVar}(\widehat{\phi}_n^{(\mathrm{MoM})})}{\mathrm{AVar}(\widehat{\phi}_n^{(\mathrm{MLE})})} = \frac{(k_4 - 4\,k_1^4 + 8\,k_1^2 k_2 - 4\,k_1 k_3 - k_2^2)(2\gamma_3 - 1)}{4\,(k_2 - k_1^2)^2}.$$

Recall the definition of k_r in (2.13), denote $k_r(\alpha) = (1-\alpha)^{r+1} + (-1)^r \alpha^{r+1}$, and write $k_r = \{k_r(\alpha)/[\alpha(1-\alpha)]^r\}\ \mu_r$. Further noting that $k_r(\alpha) = (1-\alpha)^{r+1} - (-\alpha)^{r+1}$ and using the polynomial identity $a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \ldots + ab^{n-2} + b^{n-1})$, for $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$, reveals that $k_r(\alpha)$, for r an integer, is a polynomial of degree r in α. Specifically, we find

$$\begin{aligned} k_1(\alpha) &= 1 - 2\alpha & k_3(\alpha) &= 1 - 4\alpha + 6\alpha^2 - 4\alpha^3 \\ k_2(\alpha) &= 1 - 3\alpha + 3\alpha^2 & k_4(\alpha) &= 1 - 5\alpha + 10\alpha^2 - 10\alpha^3 + 5\alpha^4. \end{aligned} \tag{3.13}$$

With the above notations we get

$$\mathrm{RAVar}(\widehat{\mu}_n) = \frac{k_1^2(\alpha)k_4(\alpha)\mu_1^2\mu_4 - k_1^2(\alpha)k_2^2(\alpha)\mu_1^2\mu_2^2 - 4k_1(\alpha)k_2(\alpha)k_3(\alpha)\mu_1\mu_2\mu_3 + 4k_2^3(\alpha)\mu_2^3}{2\left(k_2(\alpha)\,\mu_2 - k_1^2(\alpha)\mu_1^2\right)^2 \alpha(1-\alpha)}\,\gamma_1 \tag{3.14}$$

$$\mathrm{RAVar}(\widehat{\phi}_n) = \frac{k_4(\alpha)\mu_4 - 4k_1^4(\alpha)\mu_1^4 + 8k_1^2(\alpha)k_2(\alpha)\mu_1^2\mu_2 - 4k_1(\alpha)k_3(\alpha)\mu_1\mu_3 - k_2^2(\alpha)\mu_2^2}{4\left(k_2(\alpha)\mu_2 - k_1^2(\alpha)\mu_1^2\right)^2}(2\gamma_3 - 1). \tag{3.15}$$

The quantity $\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MoM})})/\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MLE})})$ in (3.14) is thus the ratio of a polynomial of degree six in α and, in the denominator, a polynomial of degree 4 in α multiplied with the factor $\alpha(1-\alpha)$. The quantity $\mathrm{AVar}(\widehat{\phi}_n^{(\mathrm{MoM})})/\mathrm{AVar}(\widehat{\phi}_n^{(\mathrm{MLE})})$ in (3.15) is the ratio of two polynomials of degree four in α.

Note that $k_r(0) = 1$, for all r; $k_r(1) = -1$, for r odd, and $k_r(1) = 1$, for r even; $k_r(0.5) = 0$ for r odd, and $k_r(0.5) = 0.5^r$ for r even. With this knowledge it is easy to find some details of the behaviour of the ratio's in (3.14) and (3.15). The denominator in (3.14) contains the factor $\alpha(1-\alpha)$ and as such equals zero when α either tends to 0 or 1. The numerator in (3.14) however stays finite for both limiting α values. As a consequence the limit of (3.14) is infinite when α either tends to zero or one. Furthermore, the limiting cases for $\alpha = 0$ and $\alpha = 1$ coïncide because of the appearance of $k_1(\alpha)$ and $k_3(\alpha)$ either in a squared version or the product of both quantities, such that the differences in signs in the two limiting cases do not enter. Table 3.1 details the behaviour of the quantities $\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MoM})})/\mathrm{AVar}(\widehat{\mu}_n^{(\mathrm{MLE})})$ and $\mathrm{AVar}(\widehat{\phi}_n^{(\mathrm{MoM})})/\mathrm{AVar}(\widehat{\phi}_n^{(\mathrm{MLE})})$, when α tends to zero or one, and when $\alpha = 0.5$, in which case we are back to the setting of the reference symmetric density.

Table 3.1: *General behaviour of RAVar($\widehat{\mu}_n$) and RAVar($\widehat{\phi}_n$) for* $\alpha = (\lim_{\alpha\searrow 0}, 0.5, \lim_{\alpha\nearrow 1})$.

α	$\alpha = 0.5$	$\lim_{\alpha\searrow 0}$ and $\lim_{\alpha\nearrow 1}$
$\mathrm{RAVar}(\widehat{\mu}_n)$	$2\mu_2\gamma_1$	$+\infty$
$\mathrm{RAVar}(\widehat{\phi}_n)$	$\dfrac{\mu_4 - \mu_2^2}{4\mu_2^2}\,\gamma_1$	$\dfrac{\mu_4 - 4\mu_1^4 + 8\mu_1^2\mu_2 - 4\mu_1\mu_3 - \mu_2^2}{4(\mu_2 - \mu_1^2)^2}\,(2\gamma_3 - 1)$

A possible loss in efficiency should be considered together with the computational advantage of the method-of-methods estimators.

3.4 Examples

Special and appealing choices for the reference symmetric (around 0) density f in (2.12) are: the standard normal density, the Student's-t density with ν degrees of freedom, the standard logistic density and the standard Laplace density. We refer to the resulting asymmetric densities as the asymmetric normal, Student-t, logistic and Laplace densities, abbreviated as AND, ATD, ALD, ALaD. See Sections 5, S.5, S.6 and 4 respectively.

Application of Theorems 2.1, 2.2 and 2.3, and Corollary 2.1 to the setting of specific examples, requires evaluation of the following characteristic quantities of the reference symmetric density f:

- for applying Theorem 2.1 we need to evaluate expressions for the cumulative distribution function and the quantile function;
- for Theorem 2.2 we need the values for μ_r, defined in (2.6), for $r = 1, 2, 3, 4$;
- for Theorem 2.3 we need to calculate the function $\varphi^+(\cdot)$ defined in (2.10).

Table S.1 in the Supplemental Material lists the quantities f, F, F^{-1}, μ_r, for $r = 1, 2, 3, 4$, and φ^+ for the above examples of reference symmetric densities.

For the application of Theorems 3.1, 3.2 and 3.4, we need to verify the assumptions, which also involves calculation of the quantities γ_1, γ_2 and γ_3 (see Assumption (B3)). In Table S.1 in the Supplemental Material we also include the values of these quantities for the special examples of reference symmetric densities.

The information in Tables 3.1 and S.1 allows to study the behaviour of the ratio's $\text{RAVar}(\widehat{\mu}_n) = \text{AVar}(\widehat{\mu}_n^{(\text{MoM})})/\text{AVar}(\widehat{\mu}_n^{(\text{MLE})})$ and $\text{RAVar}(\widehat{\phi}_n) = \text{AVar}(\widehat{\phi}_n^{(\text{MoM})})/\text{AVar}(\widehat{\phi}_n^{(\text{MLE})})$ for these special cases of densities. See Tables S.2 and S.3 (in the Supplemental Material). From Table S.3 it is clearly seen that the ratio's between the asymptotic variance of $\widehat{\boldsymbol{\theta}}_n^{(\text{MoM})}$ and $\widehat{\boldsymbol{\theta}}_n^{(\text{MLE})}$ are always equal for α and $(1-\alpha)$. In Figure 4.1 a plot of $\text{RAVar}(\widehat{\mu}_n)$ and $\text{RAVar}(\widehat{\phi}_n)$ for the asymmetric Laplace density is given. See further Section 4.3.3.

For given α, the asymptotic variance of $\widehat{\mu}_n^{(\text{MLE})}$ for asymmetric normal and Laplace densities equals $\frac{\phi^2}{\alpha(1-\alpha)}$ since $\gamma_1 = 0.5$ for both densities. The asymptotic variance of $\widehat{\mu}_n^{(\text{MoM})}$ is just double that of $\widehat{\mu}_n^{(\text{MLE})}$ for an asymmetric Laplace distribution in case $\alpha = 0.5$. On the other hand, for an asymmetric normal distribution in case of $\alpha = 0.5$ (i.e. for a symmetric normal distribution) the asymptotic variance of $\widehat{\mu}_n^{(\text{MoM})}$ equals that of $\widehat{\mu}_n^{(\text{MLE})}$. For an asymmetric Student's-t distribution with a large value of ν, the value of $\text{RAVar}(\widehat{\boldsymbol{\theta}}_n)$ is very close to that for an asymmetric normal distribution, as can be seen from Tables S.2 and S.3.

It should be mentioned that Table S.2 not only provides insights in a comparison between the asymptotic variances of maximum-likelihood and method-of-moments estimators for the family of asymmetric densities in general, but also reveals new insights about the comparison of the two estimation methods for classical symmetric densities:

- for a normal density, both estimators have the same asymptotic efficiency;
- for a logistic density, the MLE are slightly more efficient (ratio's close to one);

- for a Laplace density, the loss in efficiency when using the method-of-moments estimator is less for the scale estimator than for the location estimator.

In the next sections we look in detail into two special cases of families of asymmetric densities, and apply the general results to these. Two other cases are studied in Sections S.5 and S.6 in the Supplemental Material. In each case we start by discussing the relationship with densities in the literature, highlight differences, and emphasize the new insights gained. We provide the properties of the related densities in (2.1), and discuss the results on MoM and ML estimators for parameters of densities (2.12).

4 Quantile-based asymmetric Laplace densities

In the last several decades, a lot of research has been done on various forms of asymmetric Laplace distributions. For a general discussion on these see for example the book by Kotz et al. (2001). Hinkley and Revankar (1977) proposed an asymmetric double exponential (Laplace) density and discussed ML parameter estimation establishing asymptotic normality results in the context of Pareto law underreported data. Kozubowski and Podgórski (1999) showed that an asymmetric Laplace density is a mixture of two exponential densities with two different rates. In this section, we briefly review contributions to asymmetric Laplace distributions, showing that all can be viewed as special cases of the asymmetric family of densities in (2.1). Important is that the application of our general results in Sections 2.2 and 3 allow us to find existing results for asymmetric Laplace densities as special cases, and moreover to complete important existing gaps in the literature.

4.1 Definition of the asymmetric Laplace densities

With the symmetric density f a standard Laplace density as indicate in Table S.1, the four-parameter asymmetric Laplace density resulting from (2.1) is given by

$$f_{\lambda_1,\lambda_2}(y;\mu,\phi) = \frac{\lambda_1\lambda_2}{\phi(\lambda_1+\lambda_2)} \begin{cases} e^{-\lambda_1\left(\frac{\mu-y}{\phi}\right)} & \text{if} \quad y \le \mu \\ e^{-\lambda_2\left(\frac{y-\mu}{\phi}\right)} & \text{if} \quad y > \mu, \end{cases} \tag{4.1}$$

and the three-parameter asymmetric Laplace density obtained from (2.12) is

$$f_{\alpha}(y;\mu,\phi) = \frac{\alpha(1-\alpha)}{\phi} \begin{cases} e^{-(1-\alpha)\left(\frac{\mu-y}{\phi}\right)} & \text{if} \quad y \le \mu \\ e^{-\alpha\left(\frac{y-\mu}{\phi}\right)} & \text{if} \quad y > \mu. \end{cases} \tag{4.2}$$

If Y has density (4.1) (respectively (4.2)), we denote $Y \sim \text{ALaD}(\mu,\phi,\lambda_1,\lambda_2)$ (respectively $Y \sim \text{ALaD}(\mu,\phi,\alpha)$. The meaning of the location parameter is given by (2.4), revealing the quantile link for the location parameter. Yu and Zhang (2005) also considered the asymmetric Laplace densities in (4.1) and (4.2), and discussed MLE.

There are several reparametrizations of the asymmetric Laplace density in (4.2) in the literature; see Kotz et al. (2001). For example, by reparameterizing the scale parameter $\phi = \frac{\alpha}{\beta}$, the density (4.2) becomes

$$f_{\alpha}(y;\mu,\beta) = \beta(1-\alpha) \begin{cases} e^{-\frac{\beta(1-\alpha)}{\alpha}(\mu-y)} & \text{if} \quad y \le \mu \\ e^{-\beta(y-\mu)} & \text{if} \quad y > \mu. \end{cases}$$

which was proposed by Hinkley and Revankar (1977, eq. (5)). Another popular asymmetric Laplace density was proposed by Kotz et al. (2002), and is defined as

$$f_\kappa(y;\mu,\sigma) = \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2}\begin{cases} e^{-\frac{\sqrt{2}}{\sigma\kappa}(\mu-y)} & \text{if} \quad y \le \mu \\ e^{-\frac{\sqrt{2}\kappa}{\sigma}(y-\mu)} & \text{if} \quad y > \mu, \end{cases} \tag{4.3}$$

which is also a reparameterization of the density (4.2) with $\alpha = \frac{\kappa^2}{1+\kappa^2}$ and $\phi = \frac{\kappa\sigma}{\sqrt{2}(1+\kappa^2)}$. An important advantage of the density in (4.2) is that the location parameter μ equals the αth-quantile of the distribution. Furthermore, the parametrization in (4.2) leads to orthogonality of the parameters μ and ϕ. See Remarks 3.1 and 4.1.

4.2 Properties of the asymmetric Laplace densities

Applying Theorems 2.1, 2.2 and 2.3, and Corollary 2.1 to the setting of Section 4.1, we find the properties of the densities in (4.1) and (4.2). For convenience we collected the findings in Table 4.1. Since most properties are available in the literature we do not provide details on the derivations here. For $\phi = 1$, results for the three and four parameter density are provided in Yu and Zhang (2005); and for the three parameter density they are also in Poiraud-Casanova and Thomas-Agnan (2000).

Table 4.1: *Properties of asymmetric Laplace densities.*

Property	Four parameter density in (4.1)	Three parameter density in (4.2)
cumulative distrib. function	$F_{\lambda_1,\lambda_2}(y;\mu,\phi) = \begin{cases} \frac{\lambda_2}{\lambda_1+\lambda_2}e^{\lambda_1(\frac{y-\mu}{\phi})} & \text{if} \quad y \le \mu \\ 1-\frac{\lambda_1}{\lambda_1+\lambda_2}e^{-\lambda_2(\frac{y-\mu}{\phi})} & \text{if} \quad y > \mu \end{cases}$	$F_\alpha(y;\mu,\phi) = \begin{cases} \alpha e^{(1-\alpha)(\frac{y-\mu}{\phi})} & \text{if} \quad y \le \mu \\ 1-(1-\alpha)e^{-\alpha(\frac{y-\mu}{\phi})} & \text{if} \quad y > \mu \end{cases}$
quantile function	$F^{-1}_{\lambda_1,\lambda_2}(\beta) = \begin{cases} \mu + \frac{\phi}{\lambda_1}\ln\Big(\frac{\beta(\lambda_1+\lambda_2)}{\lambda_2}\Big) & \text{if} \quad \beta \le \frac{\lambda_2}{(\lambda_1+\lambda_2)} \\ \mu - \frac{\phi}{\lambda_2}\ln\Big(\frac{(1-\beta)(\lambda_1+\lambda_2)}{\lambda_1}\Big) & \text{if} \quad \beta > \frac{\lambda_2}{(\lambda_1+\lambda_2)} \end{cases}$	$F^{-1}_\alpha(\beta) = \begin{cases} \mu + \frac{\phi}{1-\alpha}\ln\big(\frac{\beta}{\alpha}\big) & \text{if} \quad \beta \le \alpha \\ \mu - \frac{\phi}{\alpha}\ln\big(\frac{1-\beta}{1-\alpha}\big) & \text{if} \quad \beta > \alpha \end{cases}$
central moment $E(Y-\mu)^r$	$\frac{\phi^r}{(\lambda_1+\lambda_2)}\Big[\frac{\lambda_1^{r+1}+(-1)^r\lambda_2^{r+1}}{\lambda_1^r\lambda_2^r}\Big]\Gamma(r+1)$	$\phi^r\Big[\frac{(1-\alpha)^{r+1}+(-1)^r\alpha^{r+1}}{\alpha^r(1-\alpha)^r}\Big]\Gamma(r+1)$
mean $E(Y)$	$\mu + \frac{\phi(\lambda_1-\lambda_2)}{\lambda_1\lambda_2}$	$\mu + \frac{\phi(1-2\alpha)}{\alpha(1-\alpha)}$
variance $V(Y)$	$\frac{(\lambda_1^2+\lambda_2^2)\phi^2}{\lambda_1^2\lambda_2^2}$	$\frac{\phi^2(1-2\alpha+2\alpha^2)}{(\alpha^2(1-\alpha)^2)}$
mode	μ	μ
skewness γ_{sk}	$\frac{2(\lambda_1^3-\lambda_2^3)}{[\lambda_1^2+\lambda_2^2]^{\frac{3}{2}}}$	$\frac{2(1-2\alpha)(1-\alpha+\alpha^2)}{(1-2\alpha+2\alpha)^{\frac{3}{2}}}$
kurtosis γ_{ku}	$\frac{9\lambda_1^4+6\lambda_1^2\lambda_2^2+9\lambda_2^4}{(\lambda_1^2+\lambda_2^2)^2}$	$\frac{24\alpha^4-48\alpha^3+60\alpha^2-36\alpha+9}{(2\alpha^2-2\alpha+1)^2}$
characteristic function $\varphi(t)$	$\frac{\lambda_1\lambda_2 e^{it\mu}}{(\lambda_1+\lambda_2)}\big((\lambda_1+\phi it)^{-1}+(\lambda_2-\phi it)^{-1}\big)$	$\alpha(1-\alpha)e^{it\mu}\big((1-\alpha+\phi it)^{-1}+(\alpha-\phi it)^{-1}\big)$

4.3 Parameter estimation in asymmetric Laplace densities

Let $Y_1, \ldots, Y_n$ be an i.i.d. sample from Y with density as in (4.2), i.e. $Y \sim \text{ALaD}(\mu, \phi, \alpha)$. We apply results from the general setting in Section 3 to this case.

4.3.1 Method-of-moments estimation

As far as we know there are no results available on method-of-moments estimation for an asymmetric Laplace distribution, and hence we fill in some gap here. In case the index-parameter is known, the method-of-moments estimators for $\boldsymbol{\theta} = (\mu, \phi)^T$ is given in (3.4), with k_1 and k_2 as in (3.13), and μ_1 and μ_2 as listed in Table S.1. Since the fourth moment of a Laplace distribution is finite, a straightforward application of Theorem 3.1, using the quantities tabulated in Table S.1 leads to the asymptotic normality result for the method-of-moments estimator.

Theorem 4.1. The methods-of-moment estimator $\widehat{\boldsymbol{\theta}}_n^{(\text{MoM})} = (\widehat{\mu}_n^{(\text{MoM})}, \widehat{\sigma}_n^{(\text{MoM})})^T$ is asymptotically normal distributed:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\text{MoM})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_2(\mathbf{0}, \Gamma(\boldsymbol{\theta}_0)) \qquad \text{as} \quad n \to \infty,$$

where

$$\Gamma(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\left(12\,\alpha^6 - 36\,\alpha^5 + 57\,\alpha^4 - 54\,\alpha^3 + 29\,\alpha^2 - 8\,\alpha + 1\right)\phi^2}{\alpha^2 (1-\alpha)^2 (2\,\alpha^2 - 2\,\alpha + 1)^2} & \frac{\left(6\,\alpha^5 - 15\,\alpha^4 + 22\,\alpha^3 - 18\,\alpha^2 + 7\,\alpha - 1\right)\phi^2}{\alpha(1-\alpha)\,(2\,\alpha^2 - 2\,\alpha + 1)^2} \\ \frac{\left(6\,\alpha^5 - 15\,\alpha^4 + 22\,\alpha^3 - 18\,\alpha^2 + 7\,\alpha - 1\right)\phi^2}{\alpha(1-\alpha)\,(2\,\alpha^2 - 2\,\alpha + 1)^2} & \frac{\left(5\,\alpha^4 - 10\,\alpha^3 + 13\,\alpha^2 - 8\,\alpha + 2\right)\phi^2}{(2\,\alpha^2 - 2\,\alpha + 1)^2} \end{bmatrix}.$$

4.3.2 Maximum likelihood estimation

From the general expression for the log-likelihood function in (3.7), the log-likelihood function of $\boldsymbol{\theta} = (\mu, \phi, \alpha)^T$ for the setting of Section 4.1 equals

$$\begin{aligned} \ln[L_n(\alpha, \mu, \phi)] &= n \ln[\alpha(1-\alpha)] - n\ln(\phi) - \frac{1}{\phi}\sum_{i=1}^n |Y_i - \mu|\big[(1-\alpha)\mathbb{I}(Y_i \leq \mu) + \alpha\mathbb{I}(Y_i > \mu)\big] \\ &= n \ln[\alpha(1-\alpha)] - n\ln(\phi) - \frac{1}{\phi}\sum_{i=1}^n \rho_\alpha(Y_i - \mu), \end{aligned}$$

where $\rho_\alpha(u)$ is the so-called check (or tick) loss function defined by $\rho_\alpha(u) = u(\alpha - \mathbb{I}(u < 0))$ following Koenker and Bassett (1978). The MLE of $\boldsymbol{\theta}$ is a solution to the problem $\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ln[L_n(\alpha, \mu, \phi)]$. Asymptotic results for the maximum likelihood estimation under various forms of asymmetric Laplace densities are discussed by several authors included by Hinkley and Revankar (1977), Kotz et al. (2002), and Yu and Zhang (2005).

The expression for the Fisher information matrix of $\boldsymbol{\theta}$ stated in (3.11) in Proposition 3.2, using the values for the quantities γ_r, for $r = 1, 2, 3$ tabulated in Table S.1, leads to the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ for an $\text{ALaD}(\mu, \phi, \alpha)$ density:

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\alpha\,(1-\alpha)}{\phi^2} & 0 & -\frac{1}{\phi} \\ 0 & \frac{1}{\phi^2} & -\frac{1-2\,\alpha}{\alpha\,(1-\alpha)\phi} \\ -\frac{1}{\phi} & -\frac{1-2\,\alpha}{\alpha\,(1-\alpha)\phi} & \frac{2\,\alpha^2 - 2\,\alpha + 1}{\alpha^2 (1-\alpha)^2} \end{bmatrix}. \tag{4.4}$$

Remark 4.1. For the density with parametrization as in (4.3) Kotz et al. (2002), on page 818 in their paper, provide an expression for the Fisher information matrix. Exploiting the connections with the reparametrization in (4.3), we can show that the expression of the Fisher information matrix in (4.4) coïncides with this in Kotz et al. (2002).

Direct application of Theorem 3.4 leads to the asymptotic normality results for the MLE for the parameters in density (4.2). One only needs to check the conditions of the theorem. Assumptions (B2) is satisfied since for a symmetric Laplace density $\int_0^\infty \left|\ln f(s)\right| f(s) ds \leq 1/2 + \ln(2)$ which is finite. For checking Assumption (B3), we first mention that $f'(s) = -\frac{1}{2}\text{sign}(s)e^{-|s|}$, for $s \neq 0$. From this it is easily seen that

$$\gamma_1 = \int_0^\infty \frac{(f'(s))^2}{f(s)} ds = \frac{1}{2}, \quad \gamma_2 = \int_0^\infty s \frac{(f'(s))^2}{f(s)} ds = \frac{1}{2} \quad \text{and} \quad \gamma_3 = \int_0^\infty s^2 \frac{(f'(s))^2}{f(s)} ds = 1.$$

These values are included in Table S.1. Further, Assumption (B4) is satisfied since for a standard Laplace density it is easy to check that $\lim_{s\to\infty} sf(s) = 0$.

Theorem 4.2. If Assumption (B1) holds, then the MLE $\widehat{\boldsymbol{\theta}}_n^{(\text{MLE})} = (\widehat{\mu}_n^{(\text{MLE})}, \widehat{\phi}_n^{(\text{MLE})}, \widehat{\alpha}_n^{(\text{MLE})})^T$ of $\boldsymbol{\theta}_0$ is consistent and asymptotically normally distributed:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\text{MLE})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_3(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1}) \qquad \text{as} \quad n \to \infty,$$

where

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{2\phi^2}{\alpha(1-\alpha)} & \frac{(1-2\alpha)\phi^2}{\alpha(1-\alpha)} & \phi \\ \frac{(1-2\alpha)\phi^2}{\alpha(1-\alpha)} & \frac{(3\alpha^2-3\alpha+1)\phi^2}{\alpha(1-\alpha)} & (1-2\alpha)\phi \\ \phi & (1-2\alpha)\phi & \alpha(1-\alpha) \end{bmatrix}.$$

If α is known, then the asymptotic variance-covariance matrix is

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{\phi^2}{\alpha(1-\alpha)} & 0 \\ 0 & \phi^2 \end{bmatrix}.$$

Remark 4.2. This asymptotic normality result (case α known) is also available in Hinkley and Revankar (1977) and Kotz et al. (2002). In addition, when focusing only on μ, we also mention the following. For any continuous probability density function $g_Y(y); y \in \mathbb{R}$ and an αth-sample quantile $\widehat{\mu} = \min_{\mu\in\mathbb{R}} \sum_{i=1}^n \rho_\alpha(y_i - \mu)$ of Y, Koenker and Bassett (1978) found $\sqrt{n}(\widehat{\mu}_n - \mu) \approx N\left(0, \frac{\alpha(1-\alpha)}{g_Y^2(\mu)}\right)$, for n large. For an asymmetric Laplace density, we have $g_Y(\mu) = f_\alpha(\mu; \mu, \phi) = \frac{\alpha(1-\alpha)}{\phi}$. Applying this result leads to the asymptotic distribution $\sqrt{n}(\widehat{\mu}_n - \mu) \approx N\left(0, \frac{\phi^2}{\alpha(1-\alpha)}\right)$ which coïncides with the result indicated in Theorem 4.2. However, Yu and Zhang (2005) state that $\sqrt{n}(\widehat{\mu}_n - \mu) \approx N(0, \phi^2)$, $\sqrt{n}(\widehat{\phi}_n - \phi) \approx N(0, \phi^2)$ and $\sqrt{n}(\widehat{\alpha}_n - \alpha) \sim N(0, \frac{\alpha^2(1-\alpha)^2}{(1-2\alpha)^2})$ for $\alpha \neq \frac{1}{2}$, which appears to be not correct.

4.3.3 Comparison of asymptotic variances of MoM and MLE estimators

Using the general considerations in Section 3.3.3 we can get an idea about the possible loss in efficiency when using the computationally very easy method-of-moments estimator, instead of the MLE. The graphical presentation of the quantities $\text{RAVar}(\widehat{\mu}_n) = \text{AVar}(\widehat{\mu}_n^{(\text{MoM})})/\text{AVar}(\widehat{\mu}_n^{(\text{MLE})})$ and $\text{RAVar}(\widehat{\phi}_n) = \text{AVar}(\widehat{\phi}_n^{(\text{MoM})})/\text{AVar}(\widehat{\phi}_n^{(\text{MLE})})$, with (limiting) behaviour as can be read from Tables S.2 and S.3, are plotted in Figure 4.1. The value of $\text{RAVar}(\widehat{\mu}_n)$ decreases rather rapidly for increasing values of α, for $\alpha < 0.30$; and, due to the symmetry, there is a similar rapid increase for $\alpha > 0.70$.

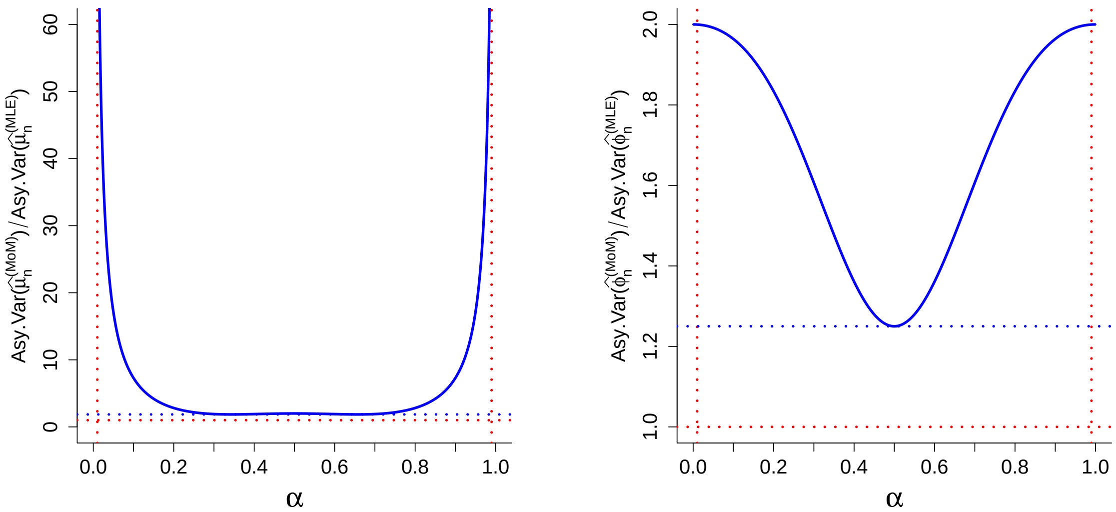

Figure 4.1: *Asymmetric Laplace density. The ratio of asymptotic variances of $\widehat{\mu}_n$ (left) and $\widehat{\phi}_n$ (right) obtained by MoM and MLE as a function of the index-parameter α.*

5 Quantile-based asymmetric normal densities

We first briefly review asymmetric normal densities that can be found in the literature. O'Hagan and Leonard (1976) proposed the following skew normal density

$$f_\lambda(y) = 2\phi(y)\Phi(\lambda y) \qquad -\infty < y < \infty \tag{5.1}$$

where $\lambda \in \mathbb{R}$. This density is a special case of the family in (1.1), by taking $f(y) = \phi(y)$ and $\Pi(y) = \Phi(y)$, respectively the standard normal density and cumulative distribution function. This is the most popular skew normal density and variations of it have been discussed by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), Arnold et al. (2002), Branco and Dey (2001), Chiogna (2004), Arellano-Valle and Genton (2005), among others.

Arellano-Valle et al. (2004) proposed a new class of skew normal densities, so called skew-generalized normal densities, defined as

$$f_{\lambda,\delta}(y) = 2\phi(y)\Phi\Big(\frac{\lambda y}{\sqrt{1+\delta y^2}}\Big) \qquad -\infty < y < \infty$$

where $\lambda \in \mathbb{R}$ and $\delta \geq 0$. Note that (5.1) is a special case of this family, obtained by taking $\delta = 0$. For $\delta = \lambda^2$, the resulting density is called a skew-curved normal density.

Recently, another class of skew normal densities was proposed by Elal-Olivero (2010), called alpha-skew normal densities, defined by

$$f_\alpha(y) = \frac{(1-\alpha y)^2+1}{2+\alpha^2}\phi(y) \qquad -\infty < y < \infty,$$

where $\alpha \in \mathbb{R}$. Taking $\alpha = 0$ gives the standard normal symmetric density.

Another popular approach is to consider so-called split-normal densities, which are obtained by joining at the mode the halves of two normal densities with the same mode but different variances. For example Gibbons and Mylroie (1973) presented the continuous two-piece normal distribution and applied a split-normal model to fit impurity profiles data. Fechner (1897) also studied the continuous two-piece normal distribution. The split-normal density with parameters μ, $\sigma_1 > 0$ and $\sigma_2 > 0$ is as follows (see, Johnson et al., 2002):

$$f(y;\mu,\sigma_1^2,\sigma_2^2) = \frac{2}{\sqrt{2\pi}(\sigma_1+\sigma_2)}\begin{cases} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma_1}\right)^2} & \text{if} \quad y \le \mu \\ e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma_2}\right)^2} & \text{if} \quad y > \mu, \end{cases} \tag{5.2}$$

where $\mu \in \mathbb{R}$ is the location parameter and $\sigma_1 > 0$ and $\sigma_2 > 0$ are the scale parameters. Several authors (see for example, Runnenburg, 1978) applied and studied the split-normal density. For example, split-normal models were used in the estimation of production frontiers in Aigner et al. (1976), and Leffrancois (1989) relied on split-normal models in forecasting processes in econometric phenomena, argueing that a split-normal model provides better forecasting value.

Mudholkar and Hutson (2000) proposed the epsilon-skew normal density that is a special case of the split-normal density, in which $\sigma_1 = (1+\epsilon)\sigma$ and $\sigma_2 = (1-\epsilon)\sigma$, with $-1 < \epsilon < 1$, and henceforth the difference between the two standard deviations of the normal densities (i.e. $\sigma_1 - \sigma_2$) equals $2\epsilon\sigma$. The epsilon-skew normal density is

$$f_\epsilon(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\begin{cases} e^{-\frac{1}{2}\left(\frac{y-\mu}{(1+\epsilon)\sigma}\right)^2} & \text{if} \quad y \le \mu \\ e^{-\frac{1}{2}\left(\frac{y-\mu}{(1-\epsilon)\sigma}\right)^2} & \text{if} \quad y > \mu, \end{cases} \tag{5.3}$$

where $-1 < \epsilon < 1$. The limiting cases of (5.3) as $\epsilon \to \pm 1$ are well known half-normal densities.

Kim (2005) presented a two-piece skew normal density with index-parameter $\lambda \in \mathbb{R}$:

$$f_\lambda(y) = \frac{2\pi}{\pi + 2\tan^{-1}(\lambda)}\phi(y)\Phi(\lambda|y|) \qquad -\infty < y < \infty. \tag{5.4}$$

If $\lambda = 0$, this reduces to the standard normal density. The density in (5.4) is uni/bimodal and a mixture of two truncated skew normal distributions.

A generalized two-piece skew normal distribution was introduced by Jamalizadeh et al. (2011), through a standard bivariate normal distribution with correlation ρ. The density in this case is given by

$$f_{\delta,\lambda,\rho}(y) = \frac{c^*(\delta,\lambda,\rho)}{\sigma}\phi\Big(\frac{y-\mu}{\sigma}\Big)\Phi_2\Big(\frac{\delta(y-\mu)}{\sigma}, \frac{\lambda|y-\mu|}{\sigma}, \rho\Big), \qquad \delta,\lambda,y \in \mathbb{R}, \tag{5.5}$$

where $c^*(\delta,\lambda,\rho) = (4\pi)\left\{\cos^{-1}\left(\frac{-(\rho+\delta\lambda)}{\sqrt{1+\delta^2}\sqrt{1+\lambda^2}}\right) + \cos^{-1}\left(\frac{-(\rho-\delta\lambda)}{\sqrt{1+\delta^2}\sqrt{1+\lambda^2}}\right) + 2\tan^{-1}(\lambda)\right\}^{-1}$ and $\Phi_2(.,.;\rho)$ denotes the cumulative distribution function of $\mathrm{N}_2(0,0,1,1,\rho)$. In the special case when $\rho = 0$, a two-piece skew normal density is obtained:

$$f_{\delta,\lambda}(y) = \frac{4\pi}{\pi + 2\tan^{-1}(\lambda)}\phi(y)\Phi(\delta y)\Phi(\lambda|y|), \qquad \delta,\lambda,y \in \mathbb{R}. \tag{5.6}$$

Note that the density in (5.4) is a special case of this extended class, and is obtained by taking $\delta = 0$ in (5.6).

The main drawback of all these existing skew normal densities, with exception for the split-normal one in (5.2), is that there is no explicit form for their quantile functions. For the asymmetric family of densities considered in Section 2.1 we have the important advantage that explicit expressions are available, as well as estimators with well-studied properties. See the next sections.

5.1 Definition of the asymmetric normal densities

Using for the reference symmetric density, a standard normal density, we obtain from (2.1) the four parameter asymmetric normal density

$$f_{\lambda_1,\lambda_2}(y;\mu,\sigma) = \frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_2)}\sqrt{\frac{2}{\pi\sigma^2}}\begin{cases} e^{-\frac{\lambda_2^2}{2}\left(\frac{y-\mu}{\sigma}\right)^2} & \text{if} \quad y > \mu \\ e^{-\frac{\lambda_1^2}{2}\left(\frac{\mu-y}{\sigma}\right)^2} & \text{if} \quad y \le \mu, \end{cases} \tag{5.7}$$

and from (2.12) the three parameter asymmetric normal density

$$f_{\alpha}(y;\mu,\sigma) = \alpha(1-\alpha)\sqrt{\frac{2}{\pi\sigma^2}}\begin{cases} e^{-\frac{\alpha^2}{2}\left(\frac{y-\mu}{\sigma}\right)^2} & \text{if} \quad y > \mu \\ e^{-\frac{(1-\alpha)^2}{2}\left(\frac{\mu-y}{\sigma}\right)^2} & \text{if} \quad y \le \mu. \end{cases} \tag{5.8}$$

For a random variable Y having density (5.7), respectively density (5.8), we denote $Y \sim \mathrm{AND}(\mu,\sigma,\lambda_1,\lambda_2)$, respectively $Y \sim \mathrm{AND}(\mu,\sigma,\alpha)$. Note that we denote $\phi = \sigma$ in this example.

Remark 5.1. The continuous two-piece normal (split-normal) normal density in (5.2) is a special case of the asymmetric normal density in (5.7), obtained by taking $\lambda_1 = \frac{\sigma}{\sigma_1}$ and $\lambda_2 = \frac{\sigma}{\sigma_2}$. Also the epsilon-skew normal density provided in (5.3) is a special case of the above family, for which $\lambda_1 = \frac{1}{1+\epsilon}$ and $\lambda_2 = \frac{1}{1-\epsilon}$.

Our focus will be on the family of asymmetric normal densities given in (5.8). Some densities from this family are depicted in Figure S.1 in the Supplemental Material. From these plots the impact and meaning of the different parameters are clearly visible.

The asymmetric normal densities given in (5.8) are quite different in construction from the most popular existing skew normal density given in (5.1). To illustrate this we consider the latter density with some extra location parameter μ and scale parameter σ^2 in the reference normal density and cumulative distribution function, leading to the skew normal density

$$f_{\lambda}(y;\mu,\sigma) = \frac{1}{\sigma\pi}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\int_{-\infty}^{\lambda\left(\frac{y-\mu}{\sigma}\right)} e^{-\frac{t^2}{2}}dt \qquad -\infty < y < \infty, \tag{5.9}$$

which is symmetric in case $\lambda = 0$, and with μ the location parameter. This is in contrast to the asymmetric normal density in (5.8) for which μ is the αth-quantile of the distribution. Figure 5.1 depicts several densities from respectively the asymmetric normal density in (5.8) (left panel), and the skew normal density in (5.9) (right panel) for various values of the index-parameters (α and λ). From these plots it is clear that density (5.8) retains the same mode μ, which equals the αth-quantile of $f_\alpha(\cdot)$. In contrast, the modes of the skew normal densities (5.9) are different for different values of the index-parameter λ.

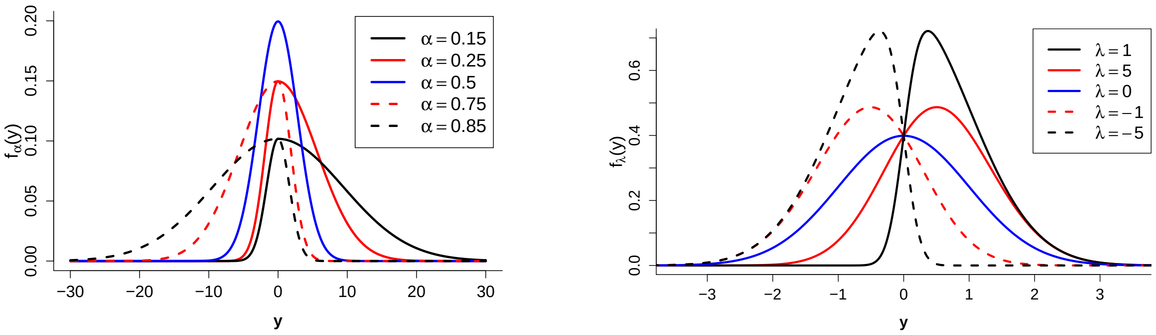

Figure 5.1: *Left: Asymmetric normal densities from* (5.8) *for different index values* α. *Right: Skew normal densities as in* (5.9) *for different index values* λ. *Location and scale parameters are* $\mu = 0$ *and* $\phi = 1$.

Further differences between the skew normal density and the asymmetric normal density are transparant from Table 5.1 (taking $\lambda_1 = 1 - \alpha$ and $\lambda_2 = \alpha$) in which we provide properties for the asymmetric normal density (5.7), as established by the results in Section 2.2. There is no explicit form of a quantile function for the existing skew normal distribution. On the other hand, there is an explicit form of the quantile function for the asymmetric normal density, and the quantile function is a linear function of μ.

5.2 Properties of the asymmetric normal densities

When appyling Theorems 2.1, 2.2 and 2.3, and Corollary 2.1 to the setting of Section 5.1, we essentially need to evaluate the characteristic quantities of the reference symmetric density f, as explained in Section 3.4.

We first look into the cumulative distribution function and the quantile function. If $Y \sim \mathrm{N}(0,1)$, then $F(y) = \frac{1}{2}[1 + \mathrm{erf}(\frac{y}{\sqrt{2}})]; y \geq 0$, where $\mathrm{erf}(x)$ is the so-called Gauss error function, defined as $\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}dt = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ for non-negative values of x, and for which $\mathrm{erf}(-x) = -\mathrm{erf}(x)$. Using Theorem 2.1, we obtain

$$F_{\lambda_1,\lambda_2}(y;\mu,\phi) = \begin{cases} \frac{\lambda_2}{\lambda_1+\lambda_2}\left[1 - \mathrm{erf}\left(\frac{\lambda_1}{\sqrt{2}}\left(\frac{\mu-y}{\sigma}\right)\right)\right] & \text{if} \quad y < \mu \\ \frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2} + \frac{\lambda_1}{\lambda_1+\lambda_2}\left[1 + \mathrm{erf}\left(\frac{\lambda_2}{\sqrt{2}}\left(\frac{y-\mu}{\sigma}\right)\right)\right]) & \text{if} \quad y \geq \mu. \end{cases}$$

We next use that $\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}}\gamma(\frac{1}{2}, x^2) = 1 - \frac{1}{\sqrt{\pi}}\Gamma(\frac{1}{2}, x^2)$ for non-negative values of x, where $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$ denotes the upper incomplete gamma function and $\gamma(s,x)$ is the lower incomplete gamma function, $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$, for $x \geq 0$. This all together then leads to the expression for the cumulative distribution function in Table 5.1.

In the derivation of the quantile function we need $\Gamma^{-1}(\frac{1}{2}, y)$ the inverse of the incomplete gamma function, i.e. $x = \Gamma^{-1}(s,y)$, which is equivalent to $y = \Gamma(s,x)$.

Table 5.1: *Properties of the four parameter asymmetric normal densities in* (5.7).

Property	
cumulative distrib. function	$F_{\lambda_1,\lambda_2}(y;\mu,\sigma^2) = \begin{cases} \frac{\lambda_2}{(\lambda_1+\lambda_2)}\frac{1}{\sqrt{\pi}}\Gamma\big(\frac{1}{2},\frac{\lambda_1^2}{2}(\frac{\mu-y}{\sigma})^2\big) & \text{if} \quad y \leq \mu \\ \frac{\lambda_2}{(\lambda_1+\lambda_2)} + \frac{\lambda_1}{(\lambda_1+\lambda_2)}\frac{1}{\sqrt{\pi}}\gamma\big(\frac{1}{2},\frac{\lambda_2^2}{2}(\frac{\mu-y}{\sigma})^2\big) & \text{if} \quad y > \mu \end{cases}$
quantile function	$F^{-1}_{\lambda_1,\lambda_2}(\beta) = \begin{cases} \mu - \frac{\sigma}{\lambda_1}\sqrt{2\Gamma^{-1}\big(\frac{1}{2},\frac{\sqrt{\pi}\beta(\lambda_1+\lambda_2)}{\lambda_2}\big)} & \text{if} \quad \beta \leq \frac{\lambda_2}{(\lambda_1+\lambda_2)} \\ \mu + \frac{\sigma}{\lambda_2}\sqrt{2\gamma^{-1}\big(\frac{1}{2},\frac{\sqrt{\pi}\beta(\lambda_1+\lambda_2)-\sqrt{\pi}\lambda_2}{\lambda_1}\big)} & \text{if} \quad \beta > \frac{\lambda_2}{(\lambda_1+\lambda_2)} \end{cases}$
central moment $E(Y-\mu)^r$	$\frac{(\sigma\sqrt{2})^r}{\sqrt{\pi}(\lambda_1+\lambda_2)}\Big[\frac{\lambda_1^{r+1}+(-1)^r\lambda_2^{r+1}}{\lambda_1^r\lambda_2^r}\Big]\Gamma\Big(\frac{r+1}{2}\Big)$
mean $E(Y)$	$\mu + \frac{\sqrt{2\sigma^2}(\lambda_1-\lambda_2)}{\sqrt{\pi}\lambda_1\lambda_2}$
variance $V(Y)$	$\frac{\sigma^2}{\pi\lambda_1^2\lambda_2^2}\Big[(\lambda_1-\lambda_2)^2(\pi-2)+\pi\lambda_1\lambda_2\Big]$
skewness γ_{sk}	$\gamma_{sk} = \frac{\sqrt{2}(\lambda_1-\lambda_2)[(\lambda_1-\lambda_2)^2(4-\pi)+\pi\lambda_1\lambda_2]}{\Big[(\lambda_1-\lambda_2)^2(\pi-2)+\pi\lambda_1\lambda_2\Big]^{\frac{3}{2}}}$
kurtosis γ_{ku}	$\frac{3\pi^2(\lambda_1^5+\lambda_2^5)-16\pi(\lambda_1-\lambda_2)(\lambda_1^4-\lambda_2^4)+12\pi(\lambda_1-\lambda_2)^2(\lambda_1^3+\lambda_2^3)-12(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)^4}{(\lambda_1+\lambda_2)\Big[(\lambda_1-\lambda_2)^2(\pi-2)+\pi\lambda_1\lambda_2\Big]^2}$
characteristic function $\varphi(t)$	$\frac{\lambda_2}{(\lambda_1+\lambda_2)}e^{it\mu-\frac{t^2\sigma^2}{2\lambda_1^2}}\Big[1-\text{erf}\Big(\frac{it\sigma}{\sqrt{2}\lambda_1}\Big)\Big] + \frac{\lambda_1}{(\lambda_1+\lambda_2)}e^{it\mu-\frac{t^2\sigma^2}{2\lambda_2^2}}\Big[1+\text{erf}\Big(\frac{it\sigma}{\sqrt{2}\lambda_2}\Big)\Big]$

Calculation of the quantity μ_r is also simple. We find

$$\mu_r = 2\int_0^\infty s^r f(s)ds = \frac{2}{\sqrt{2\pi}}\int_0^\infty s^r e^{-\frac{1}{2}s^2}ds = \frac{(\sqrt{2})^r}{\sqrt{\pi}}\Gamma\Big(\frac{r+1}{2}\Big),$$

leading to the expression for μ_r, and subsequently for $E(Y-\mu)^r$ in Table 5.1, and the specific values for μ_r, for $r = 1,2,3,4$, for the standard normal density in Table S.1.

For the function $\varphi^+(\cdot)$ we find, putting $u = y - it$ and next $u = \sqrt{2}s$,

$$\begin{aligned} \varphi^+(t) = \int_0^\infty e^{ity}f(y)dy = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{ity}e^{-\frac{1}{2}y^2}dy &= \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int_0^\infty e^{-\frac{1}{2}(y-it)^2}dy \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\int_{-it}^\infty e^{-\frac{1}{2}u^2}du \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\Big[\frac{\sqrt{2\pi}}{2} - \int_0^{-\frac{it}{\sqrt{2}}} e^{-s^2}ds\Big] \\ &= \frac{1}{2}e^{-\frac{t^2}{2}}[1+\text{erf}(\frac{it}{\sqrt{2}})] \end{aligned}$$

where we used that $\text{erf}(-x) = -\text{erf}(x)$.

From this we find the characteristic function of the asymmetric normal density (5.7)

$$\begin{aligned} \varphi(t) &= \frac{2e^{it\mu}}{(\lambda_1+\lambda_2)}\Big[\lambda_2\varphi^+\big(-\frac{\sigma t}{\lambda_1}\big) + \lambda_1\varphi^+\big(\frac{\sigma t}{\lambda_2}\big)\Big] \\ &= \frac{\lambda_2}{(\lambda_1+\lambda_2)}e^{it\mu-\frac{t^2\sigma^2}{2\lambda_1^2}}\Big[1-\text{erf}\Big(\frac{it\sigma}{\sqrt{2}\lambda_1}\Big)\Big] + \frac{\lambda_1}{(\lambda_1+\lambda_2)}e^{it\mu-\frac{t^2\sigma^2}{2\lambda_2^2}}\Big[1+\text{erf}\Big(\frac{it\sigma}{\sqrt{2}\lambda_2}\Big)\Big]. \end{aligned}$$

5.3 Parameter estimation in asymmetric normal densities

Let $Y_1, \ldots, Y_n$ be an i.i.d. sample from Y with density (5.8), i.e. $Y \sim \text{AND}(\mu, \sigma, \alpha)$. We now apply the theoretical results from Section 3 to this setting, and discuss method-of-moments and maximum likelihood estimation and the asymptotic properties for the associated estimators.

5.3.1 Method-of-moments estimation

Since the fourth moment of a standard normal distribution is finite, the following result is a straightforward application of Theorem 3.1, using k_1 and k_2 as in (3.13), with μ_1 and μ_2 as listed in Table S.1.

Theorem 5.1. The methods-of-moment estimator $\widehat{\boldsymbol{\theta}}_n^{(\text{MoM})} = (\widehat{\mu}_n^{(\text{MoM})}, \widehat{\sigma}_n^{(\text{MoM})})^T$ is asymptotically normal distributed:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\text{MoM})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_2(\mathbf{0}, \Gamma(\boldsymbol{\theta}_0)) \qquad \text{as} \quad n \to \infty,$$

where

$$\Gamma(\boldsymbol{\theta}) = \begin{bmatrix} \Gamma(\boldsymbol{\theta})_{1,1} & \Gamma(\boldsymbol{\theta})_{1,2} \\ \Gamma(\boldsymbol{\theta})_{2,1} & \Gamma(\boldsymbol{\theta})_{2,2} \end{bmatrix},$$

$$\Gamma(\boldsymbol{\theta})_{1,1} = \frac{(2.5855\,\alpha^6 - 7.7566\,\alpha^5 + 13.4837\,\alpha^4 - 14.0398\,\alpha^3 + 8.1598\,\alpha^2 - 2.4326\,\alpha + 0.4448)\,\sigma^2}{\alpha^2\,(1-\alpha)^2\,(1.4248\,\alpha^2 - 1.4248\,\alpha + 1.1416)^2}$$

$$\Gamma(\boldsymbol{\theta})_{1,2} = \Gamma(\boldsymbol{\theta})_{2,1} = -\frac{0.62668\,(1-2\alpha)^3\,(0.274\,\alpha^2 - 0.274\,\alpha + 0.566)\,\sigma^2}{\alpha\,(1-\alpha)\,(1.4248\,\alpha^2 - 1.4248\,\alpha + 1.1416)^2}$$

$$\Gamma(\boldsymbol{\theta})_{2,2} = \frac{(1.0699\,\alpha^4 - 2.1398\,\alpha^3 + 3.8429\,\alpha^2 - 2.7730\,\alpha + 0.9349)\,\sigma^2}{(1.4248\,\alpha^2 - 1.4248\,\alpha + 1.1416)^2}.$$

5.3.2 Maximum likelihood estimation

The general expression for the log-likelihood function of $\boldsymbol{\theta} = (\mu, \phi, \alpha)^T$ in (3.7), deduces to the following expression for the setting of Section 5.1, when f is the standard symmetric normal density, and we denote $\phi = \sigma$:

$$\ln[L_n(\mu, \sigma, \alpha)] = n \ln[\sqrt{2}\alpha(1-\alpha)] - \frac{n}{2}\ln(\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2\Big[(1-\alpha)^2\mathbb{I}(Y_i \leq \mu) + \alpha^2\mathbb{I}(Y_i > \mu)\Big]. \quad (5.10)$$

The log-likelihood function (5.10) is differentiable with respect to σ and α, but non-differentiable with respect to μ at the points $\mu = Y_i$, and we need to use the algorithm presented in Section 3.2 for finding the maximum likelihood estimators. We therefore verify here that all working conditions of the algorithm are fulfilled in this setting. Note first of all that the function $t(\mu) = -(Y_i - \mu)^2$ is concave and hence $-\sum_{i=1}^{n}(Y_i - \mu)^2$ is also concave. Therefore, given σ and α, the function $g_1(\mu) = \ln[L_n(\mu; \sigma, \alpha)]$ is a concave function of $\mu \in (-\infty, \infty)$. The left-hand and right-hand derivatives of $g(\mu)$ are:

$$g'_{1-}(\mu) = \frac{(1-\alpha)^2}{\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)\mathbb{I}(Y_i < \mu) + \frac{\alpha^2}{\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)\mathbb{I}(Y_i \geq \mu)$$

$$g'_{1+}(\mu) = \frac{(1-\alpha)^2}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)\mathbb{I}(Y_i \le \mu) + \frac{\alpha^2}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)\mathbb{I}(Y_i > \mu).$$

It is easily verified that $g'_{1+}(Y_{(1)}) > 0$ and $g'_{1-}(Y_{(n)}) < 0$. Hence, there is a point, $\widehat{\mu}$, between $Y_{(1)}$ and $Y_{(n)}$ which maximizes $g_1(\mu)$ and, therefore, satisfies $g'_{1-}(\widehat{\mu}) = 0$ and $g'_{1+}(\widehat{\mu}) = 0$. It is also easily shown that, given σ and α, the log-likelihood function $\ln[L_n(\mu;\sigma,\alpha)]$ is maximized at $\widehat{\mu} = Y_{(1)}$ if $\mu < Y_{(1)}$ and at $\widehat{\mu} = Y_{(n)}$ if $\mu > Y_{(n)}$. Therefore, $Y_{(1)} \le \widehat{\mu} \le Y_{(n)}$. Given $\mu \in (Y_{(1)}, Y_{(n)})$ and $\alpha \in (0,1)$, let $g_2(\eta) = \ln[L_n(\eta;\mu,\alpha)]$ with $\eta = \frac{1}{\sigma}$. The second derivative of $g_2(\eta)$ is non-positive and hence $g_2(\eta)$ is a concave function of $\eta \in (0,\infty)$. We also see that, given $\mu \in (Y_{(1)}, Y_{(n)})$ and $\sigma \in (0,\infty)$, the second derivative of the function $g_3(\alpha) = \ln[L_n(\alpha;\mu,\sigma)]$ is non-positive and hence $g_3(\alpha)$ is also a concave function of $\alpha \in (0,1)$. Hence all working conditions described in Section 3.2 are satisfied. In Section S.4.1 we present further details of the ML estimation for $\boldsymbol{\theta} = (\mu,\phi,\alpha)^T$ in (3.7).

Applying Theorem 3.4 leads to the asymptotic normality result for the MLE of the parameters in the asymmetric normal density (5.8), stated in Theorem 5.2 below. In order to apply Theorem 3.4 we need to check Assumptions (B2)—(B4). Assumption (B2) is satisfied because

$$\begin{aligned}\int_0^\infty \big|\ln f(s)\big| f(s)ds &= \int_0^\infty \Big|-\frac{1}{2}\ln(2\pi) - \frac{s^2}{2}\Big| f(s)ds \\ &\le \Big|\frac{1}{2}\ln(2\pi)\Big|\int_0^\infty f(s)ds + \int_0^\infty \Big|-\frac{s^2}{2}\Big| f(s)ds \\ &= \frac{1}{4}(\ln(2\pi)+1) < \infty.\end{aligned}$$

Regarding Assumption (B3), for f the standard normal density, for which $f'(s) = -\left(\sqrt{2\pi}\right)^{-1}e^{-s^2/2}$, we get

$$\begin{aligned}\gamma_1 &= \int_0^\infty \frac{(f'(s))^2}{f(s)}ds = \frac{1}{\sqrt{2\pi}}\int_0^\infty s^2 e^{-\frac{1}{2}s^2}ds = \frac{1}{2} \\ \gamma_2 &= \int_0^\infty \frac{s(f'(s))^2}{f(s)}ds = \frac{1}{\sqrt{2\pi}}\int_0^\infty s^3 e^{-\frac{1}{2}s^2}ds = \frac{\sqrt{2}}{\sqrt{\pi}} \\ \gamma_3 &= \int_0^\infty s^2\cdot\frac{(f'(s))^2}{f(s)}ds = \frac{1}{\sqrt{2\pi}}\int_0^\infty s^4 e^{-\frac{1}{2}s^2}ds = \frac{3}{2}.\end{aligned}$$

Finally, the first part of Assumption (B4) is obviously satisfied for the standard normal density.

Theorem 5.2. If Assumption (B1) holds, then the ML estimator $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} = (\widehat{\mu}_n^{(\mathrm{MLE})}, \widehat{\sigma}_n^{(\mathrm{MLE})}, \widehat{\alpha}_n^{(\mathrm{MLE})})^T$ of $\boldsymbol{\theta}_0$ is consistent and asymptotically normally distributed:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_3(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1}) \qquad \text{as} \quad n\to\infty,$$

where

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \frac{1}{3\,\pi-8}\begin{bmatrix} \frac{3\pi\,\sigma^2}{\alpha\,(1-\alpha)} & \frac{2\sqrt{2\pi}(1-2\,\alpha)\sigma^2}{\alpha\,(1-\alpha)} & 2\sqrt{2\pi}\sigma \\ \frac{2\sqrt{2\pi}(1-2\,\alpha)\sigma^2}{\alpha\,(1-\alpha)} & \frac{\left(5\,\alpha^2\pi+8\,\alpha^2-5\,\pi\,\alpha-8\,\alpha+2\,\pi\right)\sigma^2}{2\alpha\,(1-\alpha)} & (1-2\,\alpha)\,\sigma\pi \\ 2\sqrt{2\pi}\sigma & (1-2\,\alpha)\,\sigma\pi & \alpha\,(1-\alpha)\,\pi \end{bmatrix}.$$

If α is known, then the variance-covariance matrix is

$$\mathcal{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \frac{\phi^2}{\alpha(1-\alpha)} & 0 \\ 0 & \frac{\phi^2}{2} \end{bmatrix}.$$

As before we can access the possible loss in asymptotic efficiency when using the computationally very easy method-of-moments estimator. A plot similar to Figure 4.1 is provided in Figure S.2 in the Supplemental Material. All this is of course based on the study of the asymptotic behaviour of the MoM and ML estimators in case α is known. In order to evaluate the finite-sample performance of both type of estimators in case α is known, as well as in case α is unknown (and hence needs to be estimated), we present a small simulation study in Section S.4.3 in the Supplemental Material. That small simulation study revealed that for small sample sizes the MoM estimator is sometimes better, and for estimation of the index-parameter α both methods perform very comparable.

6 Real data example

In this section we illustrate the use of the asymmetric densities discussed in previous sections. We consider a dataset concerning the heights (in centimeters) of 100 Australian female athletes. These data were collected by the AIS (Australian Institute of Sport). These data have been used extensively in the literature, e.g. in Cook and Weisberg (2009). The data are available in the R-software package *sn* and can be downloaded from `http://azzalini.stat.unipd.it/SN/index.html`. Jamalizadeh et al. (2011) used a generalized two-piece skew normal density (5.5) to model the distribution of these data.

The main question here is which density of a set of (a)symmetric densities provides the best distributional model for these data. We consider four normal densities, three Student's-t densities and three logistic densities. All the densities have location and scale parameters μ and σ. For the normal densities, we have: (1) a symmetric $\mathrm{N}(\mu, \sigma^2)$ density; (2) an asymmetric normal density (5.1) with location and scale parameters and additional index-parameter λ; (3) an asymmetric normal density (5.5) with location and scale parameters and additional parameters δ, λ and ρ; (4) an asymmetric normal density (5.8) with μ, σ, and index-parameter α. All Student's-t densities have degrees of freedom ν and scale and location parameters μ and ϕ, and further; (5) is a symmetric Student's-t density with parameters μ, σ and ν; (6) is an asymmetric Student's-t density (S.5) with parameters μ, σ and ν and index-parameter λ; and (7) is an asymmetric Student's-t density (S.10) with parameters μ, σ, ν and index-parameter α. The three logistic densities are: (8) a symmetric logistic density with location and scale parameters μ and σ; (9) an asymmetric logistic density (S.22) with parameters μ and σ, and index-parameter λ; and (10) an asymmetric logistic density (S.25) with parameters μ, σ, and index-parameter α.

The parameters in all models are estimated by using the maximum likelihood estimation method. In order to judge about the appropriateness of a density model, we calculate the Akaike's information number

$$\mathrm{AIC} = -2\ln\left(L_n\left(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}\right)\right) + 2k,$$

where k is the number of estimated parameters in the model, and $L_n\big(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}\big)$ is the realized maximal likelihood value. The AIC-value should be as small as possible.

We also conduct for each model a Kolmogorov-Smirnov (KS) goodness-of-fit test, for testing

H_0 : Sample data come from the stated distribution

H_1 : Sample data do not come from the stated distribution.

The Kolmogorov-Smirnov test statistic is defined as :

$$D_n = \sup_y |F_0(y) - F_n(y)|, \tag{6.1}$$

where $F_0(y)$ is the cumulative function of the stated distribution (with ML estimated parameters) and $F_n(y)$ is the empirical distribution function.

The sample mean, variance, skewness and kurtosis are 174.5940, 67.9339, −0.5598 and 4.1967 respectively. Table 6.1 presents the ML estimates (MLE) of the parameters in all normal models, whereas the results for the Student's-t and logistic densities are in Table 6.2. For parameters that are not involved in a model we indicate: NAP = Not Applicable. For each model we also list the maximum log-likelihood value $\ln\big(L_n\big(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}\big)\big)$, the AIC-value, the value of the KS test statistic in (6.1), as well as the associated P-value.

Table 6.1: *Normal densities. MLE (with NAP = Not Applicable), maximal log-likelihood and corresponding AIC-value, value of the Kolmogorov-Smirnov test statistic and corresponding P-value.*

Density	symmetric normal (1)	asymmetric normal densities (2): in (5.1)	(3): in (5.5)	(4) in (5.8)
$\widehat{\mu}$	174.594	174.392	165.921	177.02
$\widehat{\sigma}$	8.201	8.199	9.131	3.879
$\widehat{\delta}$	NAP	NAP	0.498	NAP
$\widehat{\lambda}$	NAP	0.0314	0.539	NAP
$\widehat{\alpha}$	NAP	NAP	NAP	0.60
$\widehat{\rho}$	NAP	NAP	−0.965	NAP
Log-Lik.	−352.318	−352.318	**−347.088**	−350.845
AIC	708.636	710.636	704.176	707.689
KS	0.0894	0.0714	**0.0389**	0.0739
P-value	0.4011	0.6878	**0.9981**	0.6457

The maximal log-likelihood is almost the same for all skew Student's-t and logistic densities. Among all candidate models, the asymmetric logistic density (S.25) has minimum AIC-value. For all considered models, the P-values are larger than 0.05 which indicates that there is no strong evidence against H_0, for none of the models. The largest log-likelihood value is obtained with the asymmetric normal density (5.5). For this model the KS-statistic has also the smallest value with the largest associated P-value. Consequently, this asymmetric normal density constitutes the best model, but it is an overfitted model (the AIC-value is not the smallest). The smallest AIC-value is achieved for the asymmetric logistic density (S.10), which is a more parsimonious, and hence more appropriate model for these data.

Figure 6.1 depicts a histogram of the data, together with the three fitted symmetric densities (with ML estimated parameters), in dashed lines, as well as the three fitted

Table 6.2: *Student-t and logistic densities. MLE (with NAP = Not Applicable), maximal log-likelihood and corresponding AIC-value, value of the Kolmogorov-Smirnov test statistic and corresponding P-value.*

Densities	Student-t densities			logistic densities		
	symmetric	asymmetric densities		symmetric	asymmetric densities	
	(5) Student-t	(6): in (S.5)	(7): in (S.10)	(8) logistic	(9): in (S.22)	(10): in (S.25)
$\widehat{\mu}$	175.142	177.267	177.02	174.998	179.419	177.021
$\widehat{\phi}$	6.217	6.446	3.07	4.409	5.15	2.098
$\widehat{\lambda}$	NAP	−0.364	NAP	NAP	−0.862	NAP
$\widehat{\alpha}$	NAP	NAP	0.60	NAP	NAP	0.6
$\widehat{v}$	4.240	4.20	5.00	NAP	NAP	NAP
LogLik.	−349.364	−348.744	−348.465	−349.594	−348.731	−348.488
AIC	704.728	705.488	704.929	703.188	703.462	**702.976**
KS	0.0841	0.0462	0.0485	0.0440	0.0433	0.0429
P-value	0.4787	0.9727	0.9903	0.9920	0.9920	**0.9928**

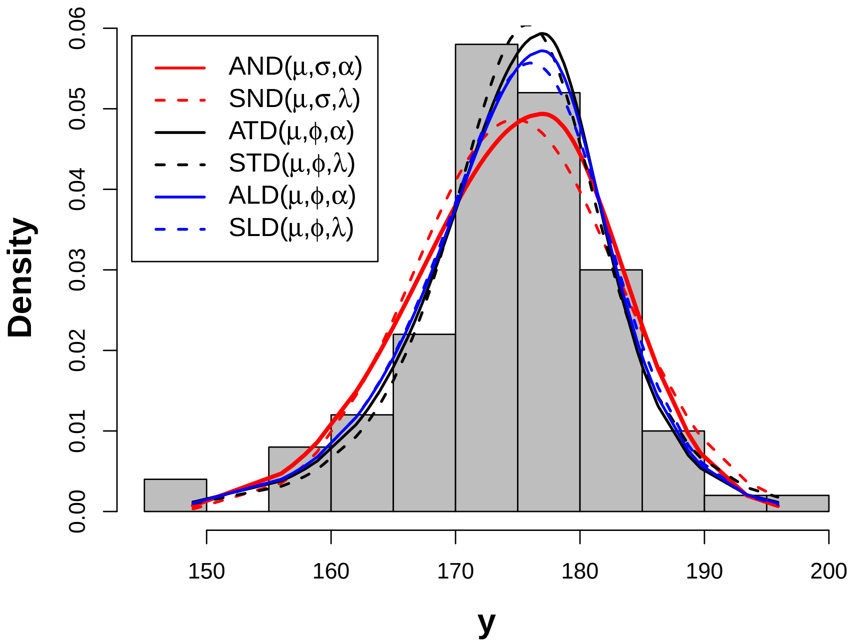

Figure 6.1: *Fitted symmetric densities (1), (5) and (8), and asymmetric densities AND, ATD and ALD, and the histogram of the data set.*

asymmetric densities that are elements of the family (1.4) (with $\lambda_1 = 1 - \alpha$ and $\lambda_2 = \alpha$) studied in this paper, in solid lines. The three asymmetric densities are clearly a better fit than the symmetric densities, which already appeared from Tables 6.1 and 6.2. Based on the MLE of the parameters, and exploiting the established expressions, we plot the quantile functions of the three asymmetric densities, densities (4), (7) and (10), in the left panel of Figure 6.2.

Using the estimated quantiles for the selected asymmetric logistic density, as well as empirical quantiles, we provide a QQ-plot for this density. For convenience a 45-degree reference line is also plotted. Note that most Q-Q values are reasonably close to the reference line.

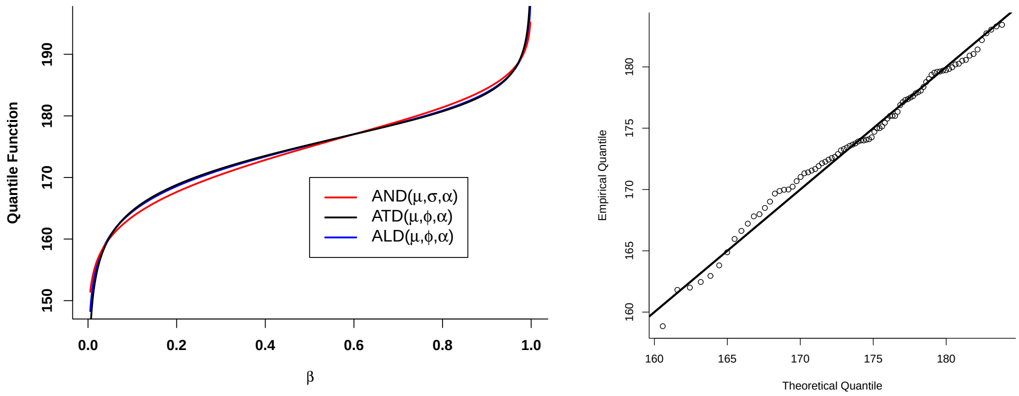

Figure 6.2: *Left: Estimated quantile functions of* $AND(\widehat{\mu}, \widehat{\sigma}, \widehat{\alpha})$*,* $ATD(\widehat{\mu}, \widehat{\phi}, \widehat{\alpha})$ *and* $ALD(\widehat{\mu}, \widehat{\phi}, \widehat{\alpha})$*; Right: Q-Q plot for the asymmetric logistic distribution ALD.*

A second real data application is provided in Section S.7 in the Supplemental Material.

7 Discussion and Conclusion

In this paper we studied a general family of asymmetric densities, established its probabilistic properties, discussed estimation of the parameters, and obtained asymptotic normality results for the estimators. The specific merit of studying the general family is that the results can readily be applied to the many examples of asymmetric densities that follow from it. As such our general results contribute on the one hand to lacking results for existing asymmetric densities in the literature, but on the other hand provide a full study of several new interesting classes of asymmetric densities.

For readability we restricted ourselves in Section 3 to the setting that the reference standard symmetric density f involves no extra parameters. In case this density comes with an extra parameter vector, say $\boldsymbol{\kappa}$, then it is rather straightforward to extend Theorems 3.2 and 3.4 for estimation of the extended parameter vector $\boldsymbol{\theta} = (\mu, \phi, \alpha, \boldsymbol{\kappa}^T)^T$. See Theorems S.3.1 and S.3.2 in the Supplemental Material, as well as Section S.5.

Although the general family of asymmetric densities involves index-parameters λ_1 and λ_2, the reduced family with only one index-parameter α appears to be flexible enough for statistical modeling. Often a first question to answer is whether the underlying density is symmetric or not. Since symmetry under this general family of asymmetric densities is equivalent to having $\alpha = 0.5$ or not, developing testing procedures for testing for symmetry would be quite feasible in this framework.

A specific feature of the general family of asymmetric densities is that its location parameter μ equals the αth-quantile of the density, where α is the index-parameter of the family. This family of densities thus has an implicit focus on estimation of quantiles. Of particular interest in future research would be to look into a regression setting in which one would extend the studied family to a setting of conditional densities, where among others estimation of the regression quantiles would be of particular interest.

References

Aigner, D. J., Amemiya, T., and Poirier, D. J. (1976). On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, pages 377–396.

Ardalan, A., Sadooghi-Alvandi, S., and Nematollahi, A. (2012). The two-piece normal-laplace distribution. *Communications in Statistics–Theory and Methods*, 41(20):3759–3785.

Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, 96(1):93–116.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2004). A new class of skew-normal distributions. *Communications in Statistics-Theory and Methods*, 33(7):1465–1480.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128(2):427–443.

Arnold, B. C., Beaver, R. J., Azzalini, A., Balakrishnan, N., Bhaumik, A., Dey, D., Cuadras, C., and Sarabia, J. M. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, 11(1):7–54.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, pages 171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46(2):199–208.

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

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.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79(1):99–113.

Chiogna, A. A.-M. (2004). Some results on the stress–strength model for skew-normal variates. *Metron*, 62(3):315–326.

Cook, R. D. and Weisberg, S. (2009). *An introduction to regression graphics*, volume 405. John Wiley & Sons.

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, Series B*, 49:1–39.

Elal-Olivero, D. (2010). Alpha-skew-normal distribution. *Proyecciones (Antofagasta)*, 29(3):224–240.

Fechner, G. T. (1897). *Kollektivmasslehre*. Engelmann.

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Gibbons, J. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-gaussian distributions. *Applied Physics Letters*, 22(11):568–569.

Hinkley, D. V. and Revankar, N. S. (1977). Estimation of the Pareto law from underreported data: a further analysis. *Journal of Econometrics*, 5(1):1–11.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233.

Jamalizadeh, A., Arabpour, A., and Balakrishnan, N. (2011). A generalized skew two-piece skew-normal distribution. *Statistical Papers*, 52(2):431–446.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (2002). *Continuous Multivariate Distributions, Volume 1, Models and Applications*, volume 59. New York: John Wiley & Sons.

Jones, M. (2015). On families of distributions with shape parameters. *International Statistical Review*, 83(2):175–192.

Kim, H.-J. (2005). On a class of two-piece skew-normal distributions. *Statistics*, 39(6):537–553.

Koenker, R. and Bassett, Jr, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). Asymmetric laplace distributions. In *The Laplace Distribution and Generalizations*, pages 133–178. Springer.

Kotz, S., Kozubowski, T. J., and Podgórski, K. (2002). Maximum likelihood estimation of asymmetric laplace parameters. *Annals of the Institute of Statistical Mathematics*, 54(4):816–826.

Kozubowski, T. J. and Podgórski, K. (1999). A class of asymmetric distributions. *Actuarial Research Clearing House*, 1:113–134.

Leffrancois, P. (1989). Allowing for asymmetry in forecast errors: Results from a montecarlo study. *International Journal of Forecasting*, 5(1):99–110.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon–skew–normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, 83(2):291–309.

Nassiri, V. and Loris, I. (2013). A generalized quantile regression model. *Journal of Applied Statistics*, 40(5):1090–1105.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

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

Poiraud-Casanova, S. and Thomas-Agnan, C. (2000). About monotone regression quantiles. *Statistics & Probability Letters*, 48(1):101–104.

Runnenburg, J. T. (1978). Mean, median, mode. *Statistica Neerlandica*, 32(2):73–79.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.

Yu, K. and Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics–Theory and Methods*, 34(9-10):1867–1879.

Appendix: Proofs of Theorems 3.1 and 3.4

A.1 Proof of Theorem 3.1

To prove this theorem, we use results from standard large sample theory (e.g., see, Serfling, 1980). Let $\boldsymbol{M}_n = (M_1, M_2)^T$ be the vector of the first two sample moments, and denote by $\boldsymbol{\mu}_Y = (\mu_{1,Y}(\boldsymbol{\theta}_0), \mu_{2,Y}(\boldsymbol{\theta_0}))^T$ the corresponding vector of population moments. From (3.1) we know that the latter vector is given by $\left(\mu_0 + k_1\phi_0, \mu_0^2 + 2k_1\mu_0\phi_0 + k_2\phi_0^2\right)^T$.

Applying Theorem 2.2.1B in Serfling (1980, pp. 68) we obtain

$$\sqrt{n}(\boldsymbol{M}_n - \boldsymbol{\mu}_Y) \xrightarrow{d} N_2(\boldsymbol{0}, \Sigma(\boldsymbol{\theta}_0)), \quad \text{where} \quad \Sigma(\boldsymbol{\theta}_0) = \Big((\mu_{i+j,Y}(\boldsymbol{\theta}_0) - \mu_{i,Y}(\boldsymbol{\theta}_0)\mu_{j,Y}(\boldsymbol{\theta}_0))_{i,j}\Big)_{1\le i,j\le 2}.$$

Since the method-of-moments estimator $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})} = (\widehat{\mu}_n^{(\mathrm{MoM})}, \widehat{\phi}_n^{(\mathrm{MoM})})^T$, with components as in (3.4), takes the form $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})} = (g_1(M_1, M_2), g_2(M_1, M_2))^T$ we can use Theorem 3.3A and its corollary in (pp. 122-126 Serfling, 1980). This then leads to

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MoM})} - \boldsymbol{\theta}_0) \xrightarrow{d} N_2(\boldsymbol{0}, \Gamma(\boldsymbol{\theta}_0)) \qquad \text{as} \quad n \to \infty,$$

where $\Gamma(\boldsymbol{\theta}) = D(\boldsymbol{\theta})\Sigma(\boldsymbol{\theta})D(\boldsymbol{\theta})^T$, with the matrix $\Sigma(\boldsymbol{\theta})$ as above and $D(\boldsymbol{\theta})$ is a 2×2 matrix with as (i,j)th-element $\frac{\partial g_i(M_1,M_2)}{\partial M_j}\Big|_{\boldsymbol{M}_n=\boldsymbol{\mu}_Y}$ $(i,j \in \{1,2\})$. Using (3.1) and Theorem 2.2, we find that

$$\Sigma(\boldsymbol{\theta}) = \begin{bmatrix} \phi^2\left(k_2 - {k_1}^2\right) & -\phi^2\left(2\,\mu\,{k_1}^2 + \phi\,k_1k_2 - 2\,k_2\mu - \phi\,k_3\right) \\ -\phi^2\left(2\,\mu\,{k_1}^2 + \phi\,k_1k_2 - 2\,k_2\mu - \phi\,k_3\right) & \Sigma_{22} \end{bmatrix}, \tag{A.1}$$

with $\Sigma_{22} = -4\,\mu^2\phi^2{k_1}^2 - 4\,\mu\,\phi^3k_1k_2 - \phi^4{k_2}^2 + 4\,\mu^2\phi^2k_2 + 4\,\mu\,\phi^3k_3 + k_4\phi^4$.

For finding the elements of $D(\boldsymbol{\theta})$ we get that

$$\frac{\partial g_1(M_1, M_2)}{\partial M_1} = 1 + \frac{k_1}{\sqrt{k_2 - k_1^2}} \frac{M_1}{\sqrt{M_2 - M_1^2}},$$

and hence

$$\frac{\partial g_1(M_1, M_2)}{\partial M_1}\Big|_{\boldsymbol{M}_n=\boldsymbol{\mu}_Y} = 1 + \frac{k_1(\mu_0 + k_1\phi_0)}{(k_2 - k_1^2)\phi_0} = \frac{\phi_0 k_2 + \mu_0 k_1}{\phi_0(k_2 - k_1^2)}. \tag{A.2}$$

Likewise, we obtain

$$\frac{\partial g_2(M_1,M_2)}{\partial M_1}\Big|_{\boldsymbol{M}_n=\boldsymbol{\mu}_Y} = -\frac{\mu_0+k_1\phi_0}{\phi_0(k_2-k_1^2)} \qquad \begin{aligned} \frac{\partial g_1(M_1,M_2)}{\partial M_2}\Big|_{\boldsymbol{M}_n=\boldsymbol{\mu}_Y} &= -\frac{k_1}{2\phi_0(k_2-k_1^2)} \\ \frac{\partial g_2(M_1,M_2)}{\partial M_2}\Big|_{\boldsymbol{M}_n=\boldsymbol{\mu}_Y} &= \frac{1}{2\phi_0(k_2-k_1^2)}. \end{aligned} \tag{A.3}$$

Combining (A.1), (A.2) and (A.3) leads to the expression for $\Gamma(\boldsymbol{\theta}) = D(\boldsymbol{\theta})\Sigma(\boldsymbol{\theta})D(\boldsymbol{\theta})^T$ as in (3.9).

A.2 Proof of Theorem 3.4

Huber (1967) studied asymptotic normality of MLE under nonstandard conditions like a non-differentiable likelihood function. Since the likelihood function $L_n(\boldsymbol{\theta})$ is non-differentiable at the points $Y_i = \mu$, we will use Theorem 3 and its corollary of Huber (1967), and start by checking the conditions.

Denote by

$$\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta}) = \begin{bmatrix} \psi_1(Y_i, \boldsymbol{\theta}) \\ \psi_2(Y_i, \boldsymbol{\theta}) \\ \psi_3(Y_i, \boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\Big(\frac{\partial^-}{\partial\mu}[\ln f_\alpha(Y_i;\mu,\phi)] + \frac{\partial^+}{\partial\mu}[\ln f_\alpha(Y_i;\mu,\phi)]\Big) \\ \frac{\partial}{\partial\phi}[\ln f_\alpha(Y_i;\mu,\phi)] \\ \frac{\partial}{\partial\alpha}[\ln f_\alpha(Y_i;\mu,\phi)] \end{bmatrix}$$

where $\frac{\partial^-}{\partial\mu}[\ln f_\alpha(Y_i,\mu,\phi)]$ and $\frac{\partial^+}{\partial\mu}[\ln f_\alpha(Y_i,\mu,\phi)]$ are respectively the left-hand and right-hand derivatives of $\ln f_\alpha(Y_i;\mu,\phi)$ with respect to μ. Since

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = E[\boldsymbol{\Psi}(Y, \boldsymbol{\theta})]$$

exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, where expectations are always taken with respect to the true underlying distribution $f_{\alpha_0}(Y_i;\mu_0,\phi_0)$ with parameter vector $\boldsymbol{\theta}_0 = (\mu_0,\phi_0,\alpha_0)^T, \boldsymbol{\theta}_0 \in \mathring{\boldsymbol{\Theta}}_R$.

The first condition **(N-1)** of Theorem 3 of Huber (1967) states that for each fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, $\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta})$ is Ω-measurable, and $\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta})$ is separable (see **(A-1)** of Huber (pp. 222, 1967)). This assumption ensures measurability of the supremum and limits which is irrelevant to us. However, since $\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta})$ is continuous and under assumption (B1), it can be easily shown that for each fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}_R$, the function $\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta})$ is Ω-measurable and separable (Billingsley, 1995). The second condition **(N-2)** is $\lambda(\boldsymbol{\theta_0}) = \mathbf{0}$. The function $\boldsymbol{\Psi}(Y, \boldsymbol{\theta})$ and its expectation with respect to the underlying true distribution, $\boldsymbol{\lambda}(\boldsymbol{\theta}_0)$, are presented in Proposition 3.1. Using this proposition, we see $\boldsymbol{\lambda}(\boldsymbol{\theta_0}) = \mathbf{0}$ which satisfies the condition **(N-2)**. The condition **(N-4)** holds, since from Proposition 3.2, we have

$$E[|\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta}_0)|^2] = E\Big\{[\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta}_0)][\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta}_0)]^T\Big\} = \text{Trace}\mathcal{I}(\boldsymbol{\theta}_0) < \infty.$$

Since the MLE $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{R}}} L_n(\boldsymbol{\theta})$ of $\boldsymbol{\theta}_0$ is consistent, we have $\sum_{i=1}^n \psi(Y_i, \widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}) = \mathbf{0}$, evidently and equation (27) of Huber (1967) holds. That is, for every sequence $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}$ satisfying $\widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})} \to \boldsymbol{\theta}_0$, in probability, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Psi}(Y_i, \widehat{\boldsymbol{\theta}}_n^{(\mathrm{MLE})}) \to \mathbf{0} \quad \text{in probability.}$$

The remaining condition **(N-3)** of Theorem 3 of Huber (1967) reads as: there are strictly positive number a, b, c, d_0 such that

(i) $\|\boldsymbol{\lambda}(\boldsymbol{\theta})\| \geq a\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|, \quad \text{for} \quad \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d_0,$

(ii) $E[u(Y_i, \boldsymbol{\theta}, d)] \leq bd, \quad \text{for} \quad \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + d \leq d_0, d \geq 0,$

(iii) $E[u(Y_i, \boldsymbol{\theta}, d)^2] \leq cd, \quad \text{for} \quad \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + d \leq d_0, d \geq 0,$

where $u(Y_i, \boldsymbol{\theta}, d) = \sup_{\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq d} \|\boldsymbol{\Psi}(Y_i, \boldsymbol{\theta}^*) - \boldsymbol{\Psi}(Y_i, \boldsymbol{\theta})\|$; $\boldsymbol{\theta}^* = (\mu^*, \phi^*, \alpha^*)^T \in \boldsymbol{\Theta}_R$ satisfies $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq d$ and $\|\boldsymbol{\theta}\|$ denotes any norm equivalent to the Euclidean norm.

Regarding part (i) above: since $\boldsymbol{\lambda}(\boldsymbol{\theta})$ is a continuous in the neighborhood of $\boldsymbol{\theta}_0$, we have the Taylor expansion of $\boldsymbol{\lambda}(\boldsymbol{\theta})$ at the point $\boldsymbol{\theta}_0$:

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) - \mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|),$$

since $\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\lambda}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\mathcal{I}(\boldsymbol{\theta}_0)$, and hence

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) + \mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|), \tag{A.4}$$

since $\boldsymbol{\lambda}(\boldsymbol{\theta}_0) = 0$. Now, using spectral decomposition of $\mathcal{I}(\boldsymbol{\theta}_0)$, we have

$$\mathcal{I}(\boldsymbol{\theta}_0) = \sum_{i=1}^3 \tau_i^2 \mathbf{e}_i\mathbf{e}_i^T,$$

where τ_1, τ_2, τ_3 are the eigenvalues of $\mathcal{I}(\boldsymbol{\theta}_0)$ and $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the corresponding (orthonormal) eigenvectors. Note that $\tau_i > 0$, since $\mathcal{I}(\boldsymbol{\theta}_0)$ is positive-definite. Then, putting $a = 0.5 \times \min\{\tau_1, \tau_2, \tau_3\}$,

$$\|\mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|^2 = \sum_{i=1}^3 \tau_i^2[\mathbf{e}_i^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]^2 \geq 4a^2\sum_{i=1}^3[\mathbf{e}_i^T(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]^2 = 4a^2\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2.$$

Hence,

$$2a\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \|\mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| \leq \|\boldsymbol{\lambda}(\boldsymbol{\theta})\| + \|\boldsymbol{\lambda}(\boldsymbol{\theta}) + \mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\|.$$

But from (A.4), it follows that there exists d_0 such that, for any $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < d_0$, we have

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) + \mathcal{I}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq a\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|.$$

Therefore,

$$\|\boldsymbol{\lambda}(\boldsymbol{\theta})\| \geq a\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \qquad \text{for } \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < d_0,$$

and part (i) of condition **(N-3)** holds.

We now check (ii) and (iii) of condition **(N-3)**. The function $u(Y_i, \boldsymbol{\theta}, d)$ is continuous on the compact set $\boldsymbol{\Theta}_R$. Therefore $u(Y_i, \boldsymbol{\theta}, d)$ is compact and bounded on $\boldsymbol{\Theta}_R$. Hence, parts (ii) and (iii) of condition **(N-3)** must hold for strictly positive numbers b and c.

Applying Theorem 3 and its corollary of Huber (1967) which state that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $[-\mathcal{I}(\boldsymbol{\theta}_0)]^{-1}[\mathcal{I}(\boldsymbol{\theta}_0)][-\mathcal{I}(\boldsymbol{\theta}_0)]^{-1} = [\mathcal{I}(\boldsymbol{\theta}_0)]^{-1}$. The elements of $[\mathcal{I}(\boldsymbol{\theta})]^{-1}$ can be calculated using $[\mathcal{I}(\boldsymbol{\theta})]^{-1} = \frac{1}{\det(\mathcal{I}(\boldsymbol{\theta}))}\mathrm{adj}(\mathcal{I}(\boldsymbol{\theta}))$.