Revista de la Facultad de Ciencias Universidad Nacional de Colombia, Sede Medellín
V 11 N°2 julio-diciembre de 2022 • ISSN-e 2357-5749 • Research paper • Pages 54 a 73
DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99967

FLAT LIKELIHOODS: SKEW NORMAL DISTRIBUTION CASE[a]

VEROSIMILITUDES PLANAS: CASO DE LA DISTRIBUCIÓN NORMAL SESGADA

JOSÉ A. MONTOYA[b*], GUDELIA FIGUEROA PRECIADO[c]

Recibido 06-12-2021, aceptado 10-03-2022, versión final 04-05-2022
Research paper

ABSTRACT: Several references argue in favor of alternative estimation methods, rather than the likelihood one, when the likelihood function exhibits flat regions. However, in the case of the skew normal distribution we present a discussion describing the interpretation of those flat likelihoods. This distribution is widely used in several interesting applications and contains the normal distribution as a nested model and the half-normal as an embedded model. Here, we show that flat likelihoods provide relevant information that should be carefully analyzed before discarding its use and proposing other estimation methods. Two well-known examples, that have been reported as troublesome, are analyzed here, including also an exhaustive computational study. The analysis of different scenarios allows to understand not only the reason of this likelihood function shape, but also to discover the information this behavior provides.
KEYWORDS: Flat likelihood function; parameter relationship; embedded models; likelihood contours; profile likelihood, simulation study.

RESUMEN: Diversas referencias argumentan a favor de métodos de estimación alternativos al de verosimilitud, cuando la función de verosimilitud exhibe regiones planas. Sin embargo, para el caso de la distribución normal sesgada se presenta una discusión donde se describe la interpretación de esas verosimilitudes planas. Esta distribución es ampliamente utilizada en diversas aplicaciones interesantes y tiene a la distribución normal como modelo anidado y a la distribución half-normal como modelo empotrado. Aquí se muestra que las verosimilitudes planas proporcionan información relevante que debe analizarse cuidadosamente, antes de abandonar su uso y proponer otros métodos de estimación. Se analizan dos ejemplos muy conocidos, que han sido reportados como problemáticos y se incluye también un estudio computacional exhaustivo. El análisis de diferentes escenarios permite comprender no sólo la razón de esta forma de la función de verosimilitud, sino también descubrir la información que este comportamiento proporciona.
PALABRAS CLAVE: Función de verosimilitud plana; relación entre parámetros, modelos empotrados, contornos de verosimilitud; verosimilitud perfil, estudio de simulación.

---

[a]Montoya, J. A. & Figueroa-Preciado, G. (2022). Flat likelihoods: Skew normal distribution case. *Rev. Fac. Cienc.*, 11 (**2**), 54–73. DOI: https://doi.org/10.15446/rev.fac.cienc.v11n2.99967
[b]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.
[*]Corresponding author: arturo.montoya@unison.mx
[c]PhD in Probability and Statistics. Statistics Professor. Department of Mathematics. University of Sonora, México.

1. INTRODUCTION

Many data generated in real life processes often exhibit asymmetry, being then poorly described with symmetric distributions, like the popular and easy to use normal distribution. The need for study and analysis of skewed distributions is fundamental, and various proposals can be found in statistical literature, like those presented in Jones (2015). Here, we focus on a particular and appropriate distribution to analyze unimodal and skewed data, that is proposed by Azzalini (1985) in a well-known paper, where skew normal distribution is introduced; being this an extension of the normal distribution family, where a shape parameter is added in order to control skewness. In this way, this three parameter distribution includes the normal distribution, when considering that skewness parameter is equal to zero. Skew normal distribution allows to reflect different degrees of skewness, so it can be used in many interesting applications, like the one related to the analysis of microRNA data that can be found in Hossain & Beyene (2015), reliability studies applied to a strength-stress model, presented in Gupta & Brown (2001), and also addressing geostatistical problems, like the ones exposed in Allard & Naveau (2007), among many applications. On the other hand, some properties and characterizations of the skew normal distribution are discussed in Genton (2004), Gupta *et al.* (2004), Figueiredo & Gomes (2013), as well as in Azzalini (2005).

The skew normal density function is given by

$$f(x;\xi,\omega,\alpha) = \frac{2}{\omega\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\xi}{\omega}\right)^2\right) \Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right), \tag{1}$$

where $x \in \mathbb{R}$, $\alpha, \xi \in \mathbb{R}$, $\omega \in \mathbb{R}^+$, and

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.$$

It is common to use the notation $X \sim SN(\xi,\omega,\alpha)$ when a random variable X has the density function presented in (1). In this parametrization α is the shape parameter, while ξ and ω are location and scale parameters, respectively. When $\alpha = 0$, a regular normal distribution $N(\xi,\omega^2)$ is obtained.

In some statistical literature, such as Azzalini & Arellano-Valle (2013) and Arellano-Valle & Azzalini (2008), two main inferential problems linked to skew normal distribution are exposed. One occurs when parameter $\alpha = 0$, since for any random sample $X_1, X_2, \cdots, X_n$ from (1), the profile log-likelihood function for α has an inflection point, and the expected information matrix becomes singular, even when the distribution is identifiable. Nevertheless, this singularity problem is due to parametrization of model (1). When the original parametrization (ξ,ω,α) is replaced for the centred one (μ,σ,γ_1), proposed by Azzalini (1985),

$$\mu = \xi + \omega\mu_\alpha, \quad \sigma^2 = \omega^2\left(1-\mu_\alpha^2\right), \quad \gamma_1 = \frac{4-\pi}{2}\left(\frac{\mu_\alpha}{\sqrt{1-\mu_\alpha^2}}\right)^3, \tag{2}$$

where $\mu_\alpha = \alpha\sqrt{2/\pi}/\sqrt{1+\alpha^2}$, then all these problems vanish (Azzalini & Arellano-Valle, 2013). It is important to emphasize that Rubio & Genton (2016) identified another problem related to the skew normal distribution, occurring when α is in an interval around zero, and although the skew normal is identifiable, it is hard to distinguish from a normal distribution. It is relevant to notice that parameter α is quite important, since it controls several features like the mode, the mean, the tail-weight and the spread of the distribution; so it is expected to require more data to be properly estimated.

On the other hand, Azzalini & Arellano-Valle (2013) and Arrué & Arellano-Valle & Gómez (2016) have pointed out another estimation problem that is also linked to the skew normal model. They found a non negligible probability that the maximum likelihood estimator (MLE) of the skewness parameter α goes to ∞ or $-\infty$, in samples of moderate sizes; that is, the MLE of α can become divergent. Azzalini & Arellano-Valle (2013) stated that if $X \sim SN(0,1,\alpha)$, then skew normal density (1), seen as a function of α, is proportional to the standard normal distribution $\Phi(\cdot)$ when this is evaluated in αx,

$$f(x;0,1,\alpha) \propto \Phi(\alpha x).$$

They also mentioned the fact that $\Phi(\cdot)$ is a monotonically increasing function, so then the maximum of the log-likelihood for an observed sample $x = x_i$ occurs if α goes to ∞ when $x_i > 0$, or if α goes to $-\infty$ when $x_i < 0$. They related this behavior with the probability of a divergent MLE, considering the fact that $P(X < 0) = 0.5 + \pi^{-1}\arctan\alpha$, when $\xi = 0$ and $\omega = 1$; then, the probability of a divergent MLE can be written as

$$p_{n,\alpha} = \left(\frac{1}{2} - \frac{\arctan\alpha}{\pi}\right)^n + \left(\frac{1}{2} + \frac{\arctan\alpha}{\pi}\right)^n. \tag{3}$$

Azzalini & Arellano-Valle (2013) reported that this probability goes rapidly to zero, when sample size n goes to ∞; but in the case of small or moderate sample sizes, there is a non negligible probability, ($p_{25,\alpha=5} \approx 0.197$ and $p_{50,\alpha=5} \approx 0.039$). These authors also mentioned that there is no characterization for those samples coming from a three parameter skew normal distribution $SN(\xi,\omega,\alpha)$, whose MLE diverges. It is important to note that the MLE is just one feature of the likelihood function shape, and what is truly fundamental is to understand the reason of that shape; only such a knowledge provides the basis for legitimate criticisms that could support new estimation methods, alternative computational optimizations, adequate reparametrizations, etcetera; such a knowledge could even provide practical clues for a model selection. Actually, as we previously stated, the skew normal singularity problem can be addressed with an adequate parametrization of the model presented in (1), as is explained in Pewsey (2000), and some references cited therein.

On the other hand, we are aware that there are some issues that deserve to be analyzed in depth, like the one related with a non negligible probability for the MLE of α of being divergent; conducting such a study may help to understand why the likelihood function for parameter α can result flat or even increasing over a wide range of parameter values. Now, although we consider important to quantify the frequency of cases where the MLE of α goes to ∞ or $-\infty$, in a skew normal with two unknown parameters, we consider even more important to understand the flatness of the likelihood function for α and its relationship with a half normal embedded model problem.

In this manuscript we basically analyze the behavior of the shape of the likelihood function for parameter α, since it controls the shape in a skew normal model. In the analyses presented here we consider two cases: the first one concerns with a skew normal model, where only parameter α is considered unknown; this case is included just for the understanding of the simplest case. The second one is related to a skew normal model where α and ξ, shape and location parameters, are both unknown. We believe that these cases provide enough information for the understanding of the shape that takes the likelihood function in a skew normal model, since location parameter ξ is linked to the support of the distribution in the embedded half normal model.

In order to meet the objectives we have set out in this manuscript, we conducted some simulation studies. In the case of unknown parameter α, we analyze the limit of the relative likelihood function for α, when α goes to infinity. Now, when (ξ,α) are both unknown, we study the limit of the profile relative likelihood function for α, when α goes to infinity; in both cases we denote this limit as $R(\infty)$. When only parameter α is unknown, simulation results showed that skew normal samples, with no negative values lead to an increasing likelihood function for α, reaching its maximum when α goes to ∞, $R(\infty)=1$, and there is a large probability that a hypothesis test for an embedded half normal model will not be rejected. On the other hand, when a skew normal sample has at least one negative observation, then a half normal model is not plausible, $R(\infty)=0$, result that any reasonable inferential method should yield, since the probability of assuming a negative value for a half normal variable is zero. Furthermore, mathematically, the likelihood function for α turns out informative and provides reliable confidence intervals.

In the case of unknown α and ξ parameters, our simulation results indicated that those skew normal samples leading to profile likelihood functions for α whose maximum is reached at infinity, $R(\infty)=1$, there is a large probability of not rejecting a hypothesis of an embedded shifted half normal model (shifted ξ units). Conversely, when $0 \leq R(\infty) < 1$, the relationship between parameter α value and sample size n, plays a fundamental role for determining the flatness of the likelihood function. For instance, when $\alpha \leq 3$ and $n \geq 200$, it is possible to construct 95 % likelihood-confidence intervals with a finite upper bound, since most of the times $R(\infty) \approx 0$. Again, for any of these skew normal samples, the shape of the likelihood function for parameter α is informative and provides reasonable inferences.

In Section 2 we introduce the likelihood function for a skew normal model where three parameters are unknown; presenting also the relative and profile likelihood functions for a specific parameter of interest in this model. Section 3 includes two illustrative examples, taken from statistical literature. The first one allows us to detect when centred parametrization can yield a well-behaved likelihood function, and the second one enables us to identify cases where the relative profile likelihood for α grows until reaching its maximum, decreasing later and staying essentially flat; features that are revealed, in a different way, when using centred parametrization. A more in-depth exploration about the flatness of the likelihood function is performed in Section 4, where two simulation studies are included. Finally, some important conclusions are presented in Section 5.

2. SKEW NORMAL LIKELIHOOD

In this section we define the likelihood function and the profile and relative likelihood functions for the general $SN(\xi,\omega,\alpha)$ case, where all three parameters are unknown. The cases where only α is unknown or (ξ,α) are both unknown, discussed in this manuscript, can be easily obtained just fixing the real values of the remaining parameters. Definitions included in this manuscript are useful and valid for the purposes described here, but more details regarding the theoretical likelihood framework can be found in Barndorff-Nielsen (1988) and Pawitan (2001), just to mention some.

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables, skew normal distributed with unknown parameters (ξ,ω,α). Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ be realization of this sample. The likelihood function of the parameters vector (ξ,ω,α), based on the observed sample $\boldsymbol{x}$ and in model (1), can be written as

$$L(\xi,\omega,\alpha) \propto \omega^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\xi}{\omega}\right)^2\right)\prod_{i=1}^{n}\Phi\left(\alpha\left(\frac{x_i-\xi}{\omega}\right)\right), \tag{4}$$

where $\alpha,\xi \in \mathbb{R}$, and $\omega \in \mathbb{R}^+$. For notation simplicity, the dependence of L on the sample is not made explicit.

When considering the centred reparametrization given in (2), then the likelihood function of the parameter vector (μ,σ,γ_1), can be written as:

$$L(\mu,\sigma,\gamma_1) \propto L\left(\xi = \xi_{\mu,\sigma,\gamma_1}, \omega = \omega_{\sigma,\gamma_1}, \alpha = \alpha_{\gamma_1}\right), \tag{5}$$

where

$$\xi_{\mu,\sigma,\gamma_1} = \mu - \sigma c_{\gamma_1}, \quad \omega_{\sigma,\gamma_1} = \sigma\sqrt{1+c_{\gamma_1}^2}, \quad \alpha_{\gamma_1} = \frac{c_{\gamma_1}}{\sqrt{b^2-(1-b^2)\,c_{\gamma_1}^2}},$$

$b = \sqrt{2/\pi}$, and $c_{\gamma_1} = [2\gamma_1/(4-\pi)]^{1/3}$. It is important to note that throughout this manuscript, we use both, direct parametrization (ξ,ω,α) of the skew normal model, as well as centred reparametrization (μ,σ,γ_1), in order to exemplify and analyze the likelihood shape. Nevertheless, it will be sufficient to define the relative and profile likelihood functions for the centred parametrization.

The relative likelihood function of (μ,σ,γ_1) is a standardized version of (5), that takes the value of one at its maximum, the MLE of (μ,σ,γ_1),

$$R(\mu,\sigma,\gamma_1) = \frac{L(\mu,\sigma,\gamma_1)}{\max\limits_{\mu,\sigma,\gamma_1} L(\mu,\sigma,\gamma_1)} = \frac{L(\mu,\sigma,\gamma_1)}{L(\widehat{\mu},\widehat{\sigma},\widehat{\gamma}_1)}, \tag{6}$$

where $\widehat{\mu}$, $\widehat{\sigma}$, and $\widehat{\gamma_1}$ are the MLE's of parameters μ, σ, and γ_1. Note that the likelihood function for (ξ,ω,α) is obtained just replacing in (6) the function $L(\mu,\sigma,\gamma_1)$ by $L(\xi,\omega,\alpha)$ given in (4).

Assume that the full parameter (μ,σ,γ_1) can be partitioned into an interest parameter ψ and a nuisance parameter λ. Then, the profile likelihood and its corresponding relative likelihood function of ψ, standardized to be one at the maximum of the likelihood function, are defined as

$$L_{\max}(\psi) = \max_{\lambda} L(\psi,\lambda),$$
$$R_{\max}(\psi) = \frac{L_{\max}(\psi)}{\max_{\psi,\lambda} L(\psi,\lambda)} = \frac{L_{\max}(\psi)}{L_{\max}(\widehat{\psi})}.$$

For example, the profile likelihood and its corresponding relative likelihood function of $\psi = \gamma_1$ is

$$L_{\max}(\gamma_1) = \max_{\mu,\sigma} L(\mu,\sigma,\gamma_1), \tag{7}$$
$$R_{\max}(\gamma_1) = \frac{L_{\max}(\gamma_1)}{\max_{\mu,\sigma,\gamma_1} L(\alpha,\gamma,\beta)} = \frac{L_{\max}(\gamma_1)}{L_{\max}(\widehat{\gamma_1})}. \tag{8}$$

Now, it is easy to note that profile likelihood of $\psi = \alpha$ (and its corresponding relative likelihood) can be obtained just substituting in (7-8) the function $L(\mu,\sigma,\gamma_1)$ by the function $L(\xi,\omega,\alpha)$ given (4).

In general, the profile or maximized likelihood is a powerful though simple inferential method for estimating a parameter of interest separately from the remaining unknown parameters of the statistical model. More details about the profile likelihood approach are described in Kalbfleisch (1985, Section 10.3), Pawitan (2001, Section 3.4), Sprott (2000, p. 66), Barndorff-Nielsen & Cox (1994, pp. 89-91), Serfling (2002, pp. 155-160), and Murphy & Van Der Vaart (2000). Thus, the profile likelihood function may be used to study and visualize various aspects of a full likelihood function, such as for instance those associated with the flatness of the likelihood surface, as well as the intrinsic relationship between those parameters.

3. EXAMPLES: SKEW NORMAL LIKELIHOOD

In this section we analyze two data sets. In Example 1 we show that centred reparametrization (2) proposed by Azzalini (1985) allows to symmetrize the likelihood function, when data come from a skew normal model where α is closed to zero; so this parameter space transformation favors inferences about model parameters. However, as will be shown in Example 2, when data come from a skew normal model where α is far away from zero, this skew normal model reparametrization generates likelihood functions whose shapes reveal features linked to flat or monotone increasing likelihoods, but those features emerge in a different way.

3.1. Example 1

In this example we construct profile likelihood functions based on a simulated random sample $\boldsymbol{x_1} = (x_1, x_2, \cdots, x_n)$ of size $n = 202$, from a skew normal model $X \sim SN(64.07, 17.68, 0)$, using

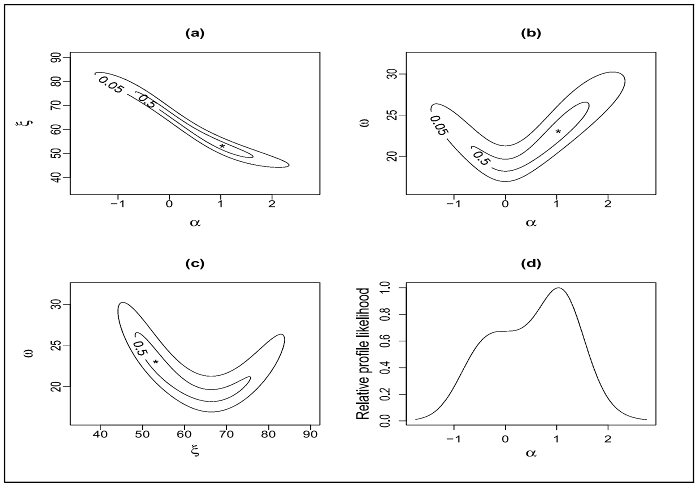

Figure 1: Unknown three parameters ξ, ω, and α: Relative profile likelihood functions for direct parameters corresponding to the simulated sample $\boldsymbol{x_1}$ with size $n = 202$ from a skew normal with parameters $\xi = 64.07$, $\omega = 17.68$, and $\alpha = 0$. Source: Elaborated by the authors.

R software version 4.0.4, and using a seed in order to make it replicable in case of interest: set.seed(as.numeric(as.Date("2021-09-03"))). Sample size selection and parameter values for ξ and ω were selected according to the sample size and the MLE's that were obtained with the AIS, (Australian Institute of Sport) weight data (in kg); see Arellano-Valle & Azzalini (2008). AIS data contain biomedical measurements on 202 Australian athletes and it is available in the library *sn* in R. Shape parameter was fixed in $\alpha = 0$, just for comparing the likelihood shape when direct (ξ, ω, α) or centred (μ, σ, γ_1) reparametrization are used.

Based on simulated $\boldsymbol{x_1}$ data, Figures 1-2 show the profile likelihood contours for parameters in the $SN(\xi, \omega, \alpha)$ and $SN(\mu, \sigma, \gamma_1)$ models, respectively, considering all parameters unknown. As can be observed in Figure 1, for this simulated sample, direct reparametrization (ξ, ω, α) generates elongated likelihood contours, sometimes like a boomerang shape. In this case, profile likelihood function for parameter α is not flat; it is bimodal, with the highest hump corresponding to the MLE location. Everything seems to indicate that this skew normal parametrization brings to light a model parameters relationship. On the other hand, Figure 2 leads to what could be called well-behaved and nearly symmetric likelihoods, favoring simple, efficient and reasonable inferences.

3.2. Example 2

This example is based on sample size and parameter values considered in Frontier data, where a set of $n = 50$ values were sampled from $SN(0, 1, 5)$; this dataset is presented by Azzalini & Capitanio (1999). It is well-known that the MLE of α, in this dataset, diverges when a skew normal model with unknown parameters (ξ, ω, α) is considered; see Azzalini & Arellano-Valle (2013).

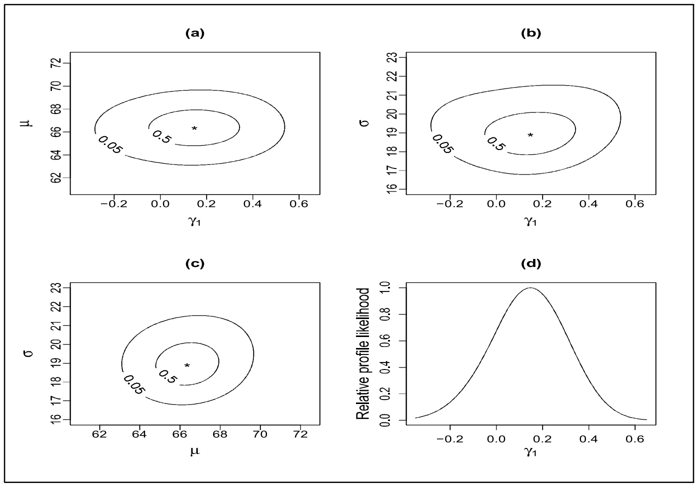

Figure 2: Unknown three parameters μ, σ, and γ_1: Relative profile likelihood functions for centred parameters corresponding to the simulated sample $\boldsymbol{x_1}$ with size $n = 202$ from a skew normal with parameters $\xi = 64.07$, $\omega = 17.68$, and $\alpha = 0$. Source: Elaborated by the authors.

In view of the foregoing, we construct some profile likelihood functions based on a sample $\boldsymbol{x_2} = (x_1, x_2, \cdots, x_n)$ of size $n = 50$, that was simulated from a skew normal random variable, $X \sim SN(0,1,5)$ using the R software. The sample can be easily replicated by setting a seed with the instruction set.seed(as.numeric(as.Date("2021-09-06"))). This sample allows us to show a relative profile likelihood for α, that grows until it reaches its maximum and then decreases, staying essentially flat.

In the case of a $SN(\xi, \omega, \alpha)$ model with all parameters unknown, Figure 3 shows the shape of different relative profile likelihood contours when direct parametrization (ξ, ω, α) is used. It can be seen that the contour plot, associated to the relative profile likelihood of (ξ, ω) based on the observed sample $\boldsymbol{x_2}$, estimates quite well some plausible regions for these parameters. However, contour plots associated to the relative profile likelihood functions of (ξ, α) and (ω, α), lengthen as α goes to infinity. Actually, relative profile likelihood of α seems flat for values greater than 40. It is important to highlight that a similar behavior is observed when considering a $SN(\xi, 1, \alpha)$ model, where just ξ and α, the location and shape parameters are unknown; see Figure 4.

Finally, Figure 5 shows the profile likelihood function of γ_1, based on our 50 simulated observations contained in $\boldsymbol{x_2}$, and also the one that is based on Frontier's data, a set of $n = 50$ values sampled from $SN(0,1,5)$. For these two samples we assume a skew normal model with ξ and α unknown, when using direct parametrization, and μ y γ_1 are treated as unknown, in case of a centred parametrization. In this figure we can observe that the MLE of γ_1, in Frontier data, is reached at the right end of the support of the parameter space for γ_1, given by $\sqrt{(}2)\,(4-\pi)\,/\,(\pi-2)^3\,/2$, and occurring when $\alpha \to \infty$. As we previously stated, this is also the case in a skew normal model with three unknown parameters. On the other hand, for our simulated data, the set of γ_1 values with a relative profile likelihood greater than 0.15, $\{\gamma_1 : R_{\max}(\gamma_1) \geq 0.15\}$, provides a

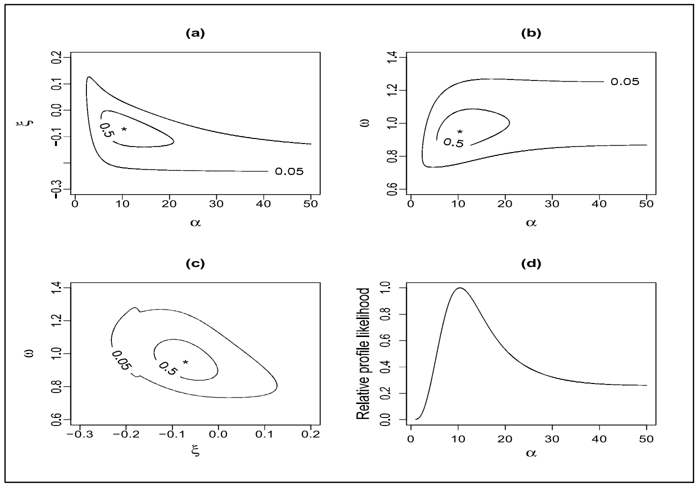

Figure 3: Unknown three parameters ξ, ω, and α: Relative profile likelihood functions for direct parameters corresponding to the simulated sample $\boldsymbol{x_2}$ with size $n = 50$ from a skew normal with $\xi = 0$, $\omega = 1$, and $\alpha = 5$. Source: Elaborated by the authors.

95 % likelihood-confidence interval for γ_1. This includes the right end of the support of the parameter space for γ_1, associated to $\alpha \to \infty$, that is linked to a well-defined model, as this simply corresponds to the half normal distribution. Thus, this does not represent a degenerate case, but a well-defined model.

It is interesting to deeply analyze this example, since it reveals that some features linked to flat or monotone increasing likelihoods for shape parameter α, in the skew normal model, do not vanish when using centred parametrization; actually, those features are inherited and revealed in a different way.

Next section is devoted to analyze, in greater detail, this flat likelihood phenomenon; taking into account the relationship between this likelihood flatness and the closeness between skew normal models and their embedded models.

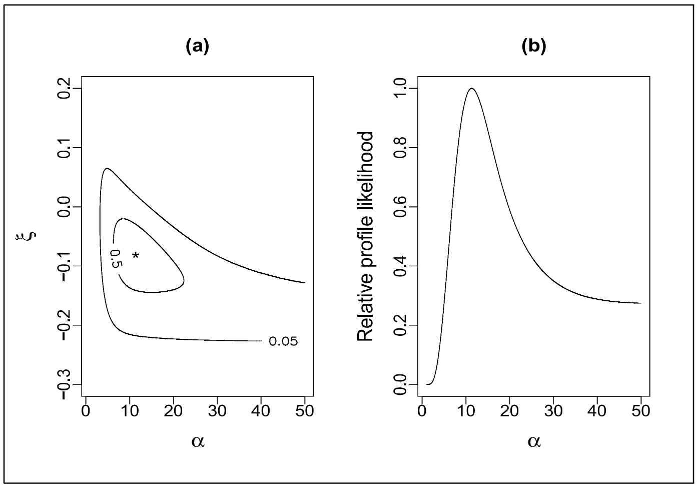

Figure 4: Unknown two parameters ξ and α: Relative profile likelihood functions for direct parameters corresponding to the simulated sample $\boldsymbol{x_2}$ with size $n = 50$ from a skew normal with $\xi = 0$, $\omega = 1$, and $\alpha = 5$. Source: Elaborated by the authors.

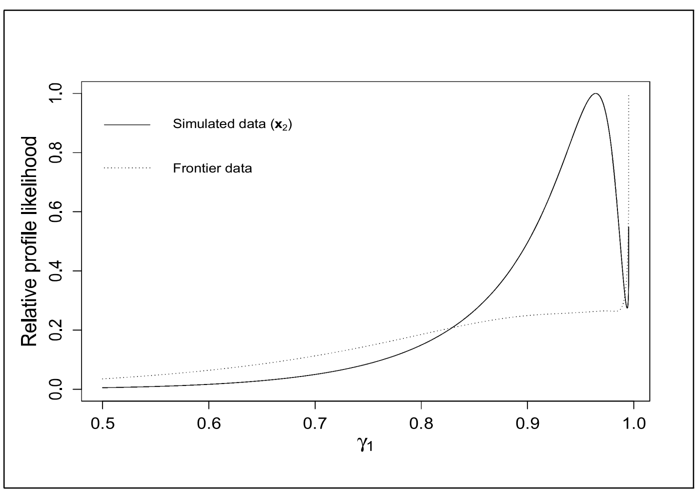

Figure 5: Two unknown parameters: μ and γ_1: Relative profile likelihood functions for γ_1 corresponding to the simulated sample $\boldsymbol{x_2}$, with size $n = 50$ from a skew normal with parameters $\xi = 0$, $\omega = 1$, and $\alpha = 5$, and Frontier data, a set of $n = 50$ values sampled from $SN(0,1,5)$. Source: Elaborated by the authors.

4. FLATNESS OF THE RELATIVE PROFILE LIKELIHOOD OF α

As mentioned by Sprott (2000, p. 11), the MLE and the observed information exhibit two features of the likelihood function; the former is a measure of its location relative to the parameter-axes, while the latter is a measure of its curvature in a neighborhood of the MLE. Thus, when the likelihood function of a scalar parameter is smooth and symmetric with respect to the MLE, then the MLE and the observed information are usually the main features determining the shape of the likelihood function. It should be stressed that flat likelihoods can not be determined by these quantities, without loss of information.

Intuitively, the sensitivity of the MLE depends on the shape of the likelihood function. If the likelihood of α is very curved around MLE, then the MLE results a stable estimate of α. On the other hand, if the likelihood is flat near the MLE, then this becomes highly unstable. However, more important than those facts is the one related with likelihood-confidence intervals of α arising from a flat likelihood, where upper limit can become infinity.

In this section we study, throughout computer simulations, the degree of practical parameter non-identifiability of parameter α, in the sense explained in Cole (2020, p. 38) that "*a parameter is practically non-identifiable if the log-likelihood has a unique maximum, but the parameters's likelihood-based confidence region tends to infinity in either or both directions*". Here, we study the frequency of the occurrence of flat likelihoods of shape parameter α by exploring the limit of the relative likelihood of α, as α approaches infinity. First of all, we analyze the simplest case when only parameter α is unknown, considering later a skew normal model with all parameters unknown. Always having in mind the existence of an embedded half normal model in this flat likelihood problem, our scenarios were designed considering the closeness between the skew normal and half normal models. Total variation distance (TV), as exposed in Huber (1981, p. 34), Jurečková (2006, p. 12) and Rieder (1994), was used to measure the closeness between these distributions and computed through the function TotalVarDist included in the R library *distrEx*. The probability of divergence of the MLE of α, provided in (3), was also measured.

For all simulated samples where an increasing likelihood function pattern ($\widehat{\alpha} \to \infty$) was detected, a p-value associated to a Kolmogorov-Smirnov distribution-free test was obtained; this procedure makes possible to test for general differences in two distributions, as it is explained in Hollander et al. (2014). Resulting p-values allow to quantify the extent to which these samples provide evidence against an embedded half normal model. Note that no p-value computation is needed when the divergence condition for the MLE of α is not met (al least one negative observation), since no negative values are allowed in a half normal model.

4.1. The unknown shape parameter case

To illustrate the degree of practical parameter non-identifiability of the shape parameter in the skew normal model, given a sample size, we will focus on the limit of the relative likelihood function of α as α approaches infinity,

$$R(\infty) = \lim_{\alpha\to\infty} R(\alpha) = \lim_{\alpha\to\infty} \frac{L(\xi=0,\omega=1,\alpha)}{\max\limits_{\xi=0,\omega=1,\alpha} L(\xi,\omega,\alpha)}, \tag{9}$$

where $L(\xi,\omega,\alpha)$ is given in (4). Note that centred parametrization (2) yields

$$\mu = \mu_\alpha, \quad \sigma^2 = \left(1-\mu_\alpha^2\right), \quad \gamma_1 = \frac{4-\pi}{2}\left(\frac{\mu_\alpha}{\sqrt{1-\mu_\alpha^2}}\right)^3,$$

where $\mu_\alpha = \alpha\sqrt{2/\pi}/\sqrt{1+\alpha^2}$. In this way, as $\alpha\to\infty$, $\mu_\alpha \to \sqrt{2/\pi}$, and then

$$\mu = \sqrt{\frac{2}{\pi}}, \quad \sigma^2 = 1-\frac{2}{\pi}, \quad \gamma_1 = \frac{\sqrt{2}\,(4-\pi)}{(\pi-2)^{3/2}}. \tag{10}$$

Moreover, with this reparametrization $\gamma_1 \in [-\gamma_{10},\gamma_{10}]$, where $\gamma_{10} = \sqrt(2)\,(4-\pi)\,/\,(\pi-2)^3\,/2$; this can be interpreted as computationally favorable when searching the MLE of γ_1. For all the foregoing considerations, in our computations we use the following expression for $R(\infty)$, based on (9):

$$R(\infty) = \begin{cases} 1, & \text{if all } x_i > 0, \\ \dfrac{L\left(\mu=\sqrt{2/\pi},\sigma=1-2/\pi,\gamma_1=\gamma_{10}\right)}{\max\limits_{\gamma_1\in[-\gamma_{10},\gamma_{10}]} L\left(\mu=\mu_{\gamma_1},\sigma=\sigma_{\gamma_1},\gamma_1\right)}, & \text{in another case,} \end{cases} \tag{11}$$

where $\mu_{\gamma_1} = c_{\gamma_1}/\sqrt{1+c_{\gamma_1}^2}$ and $\sigma_{\gamma_1} = 1/\sqrt{1+c_{\gamma_1}^2}$.

If $R(\infty)$ is smaller than one, but close to it, then for every real number $\varepsilon > 0$ there exists a real number α_0 such that, for all $\alpha > \alpha_0$, we have $R(\alpha) \in (R(\infty)-\varepsilon, R(\infty)+\varepsilon)$. Thus, there is a set of values of α that makes the observed sample as likely as does the MLE. The frequency of occurrence of this type of behavior represents an index of the degree of practical parameter non-identifiability. In particular, when all parameters are unknown, $R(\infty)=1$ can be interpreted as an extreme case of a practical parameter non-identifiability of the shape parameter in the skew normal model. This situation could suggest that dataset may be fitted with an embedded half normal model, whose parameters are given in (10).

In order to select adequate simulation scenarios that allow us to analyze the behavior of $R(\infty)$, and also its relationship with an embedded half normal model, obtained when α goes to infinity, we focus on the closeness between the skew normal and half normal models. As we previously stated we use the total variation distance, TV, as a measure of the closeness between these distributions. Figure 6a shows the plots for skew normal densities with shape parameter values $\alpha = 0, 5, 50$, and $\alpha\to\infty$; while in Figure 6b we plot TV as a function of α, in order to easily identify TV distances between the densities shown in Figure 6a. Particularly, in Figure 6b we can observe a $TV = 0.5$ between normal density ($\alpha = 0$) and half normal density, difference that is visually evident from Figure 6a, where we can observe that a skew normal with $\alpha = 50$ seems very close to half normal model and $TV = 0.006365677$, quite smaller than $TV = 0.5$. Now, TV distance between

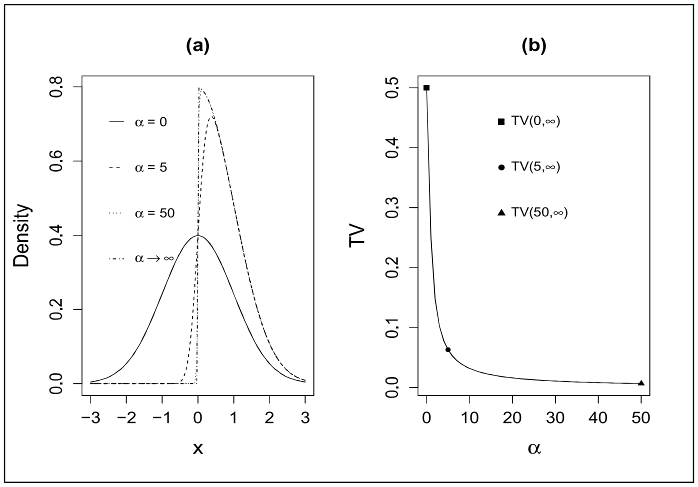

Figure 6: (a) Skew normal densities with shape parameter values $\alpha = 0, 5, 50$ and $\alpha \to \infty$. (b) TV as a function α, measuring closeness between skew normal and an embedded half-normal model. In this plot, TV distances are obtained considering $\alpha = 0, 5, 50$ versus $\alpha = \infty$, according to densities plotted in (a). Source: Elaborated by the authors.

Table 1: Percentages $100p_{n,\alpha}$ of divergence of the MLE of α, for simulation scenarios.

n	TV			
	0.10 ($\alpha = 3$)	0.06 ($\alpha = 5$)	0.03 ($\alpha = 11$)	0.006 ($\alpha = 50$)
25	6.71	19.74	48.09	85.24
50	0.45	3.90	23.13	72.67
100	0	0.15	5.35	52.80
200	0	0	0.29	27.88

a skew normal model where $\alpha = 5$ is closer from a half normal model ($\alpha \to \infty$), than the one obtained when $\alpha = 0$.

Considering the cases previously described, we now simulate independent random samples from a skew normal random variable, $X \sim SN(0, 1, \alpha)$, setting α values in $3, 5, 11$, and 50; these values yield TV distances of $0.1, 0.06, 0.03$, and 0.006, respectively. On the other hand, we select sample sizes of $n = 25$, 50, 100, and 200, in order to have a wide variety for the probability of divergence of the MLE of α. These values are computed throughout $p_{n,\alpha}$ given in (3), and are shown in Table 1.

Table 2 shows the classification of 5000 simulated samples, for each of the selected scenarios, according to whether $0 \leq \widetilde{R}_{\max}(\infty) < 1$ or $\widetilde{R}_{\max}(\infty) = 1$. Note that $\widetilde{R}_{\max}(\infty) = 1$ corresponds to those samples whose MLE of α diverged to ∞. Actually, empirical frequencies corresponding to this column in Table 2, are similar to the percentages shown in Table 1, obtained from the theoretical divergence probabilities. Moreover, all samples considered in this column have positive observations, and at least 95.55% of the p-values yielded by a Kolmogorov-Smirnov test, under the assumption of an embedded half normal model, were greater than 0.05. Thus, at 0.05 significance level, a half normal model will not be rejected approximately 95% of the

Table 2: Classification (in percentage) of 5000 simulated samples $X \sim SN(0,1,\alpha)$.

α	n	$0 \leq \widetilde{R}_{\max}(\infty) < 1$	$\widetilde{R}_{\max}(\infty) = 1$
3	25	93.38	6.62
	50	99.66	0.34
	100	100	0
	200	100	0
5	25	80.28	19.72
	50	96.20	3.80
	100	99.82	0.18
	200	100	0
11	25	51.24	48.76
	50	77.24	22.76
	100	94.74	5.26
	200	99.68	0.32
50	25	15.06	84.94
	50	25.94	74.06
	100	46.28	53.72
	200	72.24	27.76

times, regardless of the α and n values considered in Table 1.

On the other hand, when samples had at least one negative observation, and divergence condition for the MLE of α is not met, for any of the considered scenarios we found that $R(\infty) \in \left[0, 1.691853 \times 10^{-316}\right]$; that is, the right tail of the likelihood functions for α and γ_1, decreases to practically zero. For the particular case when $TV = 0.006$ and $n = 25$, Figure 7 shows the enveloping curves for the relative likelihood functions of γ_1, where simulated samples had at least one negative observation (753 samples, 15.06 %). In this figure we can observe that all relative likelihoods for γ_1 reached the value $R(\infty) = 0$ at $\gamma_1 = \gamma_{10}$. Now, Figure 8 shows the enveloping curves when considering a large TV and also a large sample size ($TV = 0.1$, $n = 200$), in samples with at least one negative observation (5000 samples, 100 %); so likelihood-confidence intervals for γ_1 or α, at typical confidence levels (90 %, 95 %, and 99 %), yield finite upper bounds, since $R(\infty) = 0$.

In summary, simulation results show that when skew normal samples have no negative observations, likelihood function grows until it reaches its maximum at $\alpha = \infty$, being highly probable that an embedded half normal model will not be rejected under a hypothesis test. Now, in case of at least one negative observation, an embedded half normal model results not plausible ($R(\infty) = 0$). Furthermore, mathematically, likelihood yields reliable confidence intervals, sometimes with endpoints $\left(A_{\gamma_1}, \gamma_{10}\right)$, or equivalently (A_α, ∞). Thus, for any of these skew normal samples, likelihood shape resulted informative, providing reasonable inferences.

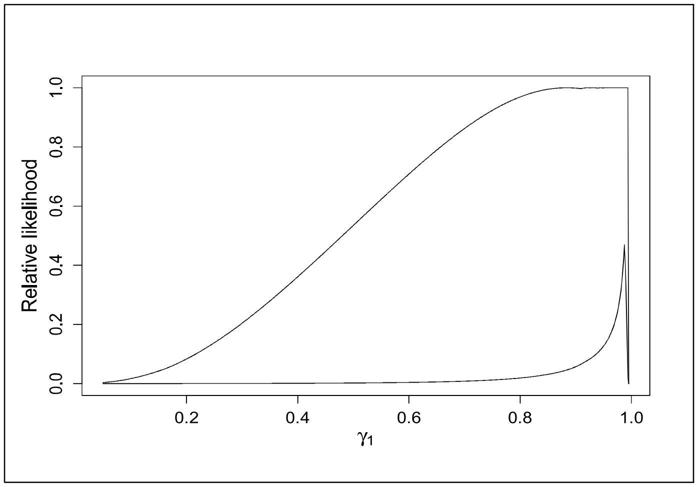

Figure 7: Enveloping curves for relative likelihood functions of γ_1 for simulated skew normal samples where: $TV = 0.006$, $\alpha = 50$ and $n = 25$. Source: Elaborated by the authors.

4.2. The case of unknown shape and location parameters

In the case that parameters (ξ, α) are unknown, we will focus on the limit of the relative profile likelihood function of α as α approaches infinity,

$$R(\infty) = \lim_{\alpha\to\infty} R_{\max}(\alpha) = \lim_{\alpha\to\infty} \frac{L_{\max}(\xi, \omega = 1, \alpha)}{\max\limits_{\xi,\omega=1,\alpha} L(\xi, \omega, \alpha)}, \tag{12}$$

where $L(\xi, \omega, \alpha)$ is given in (4). Note that, in this case, centred parametrization results

$$\mu = \xi + \mu_\alpha, \quad \sigma^2 = \left(1 - \mu_\alpha^2\right), \quad \gamma_1 = \frac{4-\pi}{2}\left(\frac{\mu_\alpha}{\sqrt{1-\mu_\alpha^2}}\right)^3.$$

Then, as $\alpha \to \infty$, $R(\infty)$ can be written as:

$$R(\infty) = \frac{\max\limits_{\xi\in\mathbb{R}} L\left(\mu = \xi + \sqrt{2/\pi}, \sigma = 1 - 2/\pi, \gamma_1 = \gamma_{10}\right)}{\max\limits_{\xi\in\mathbb{R},\gamma_1\in[-\gamma_{10},\gamma_{10}]} L\left(\mu = \mu_{\xi,\gamma_1}, \sigma = \sigma_{\gamma_1}, \gamma_1\right)}, \tag{13}$$

where $\mu_{\xi,\gamma_1} = \xi + c_{\gamma_1}/\sqrt{1 + c_{\gamma_1}^2}$ and $\sigma_{\gamma_1} = 1/\sqrt{1 + c_{\gamma_1}^2}$.

In order to analyze the case where shape and location parameters are unknown, and also for being able of comparing new results with those obtained in previous section, same samples and scenarios will be considered in this section. This makes possible to computationally evaluate the profile likelihood of α when only parameter α is unknown, and also the case when parameters (ξ, α) are both unknown. We again compute the p-values that a Kolmogorov-Smirnov test yields, when an embedded shifted half normal model is assumed,

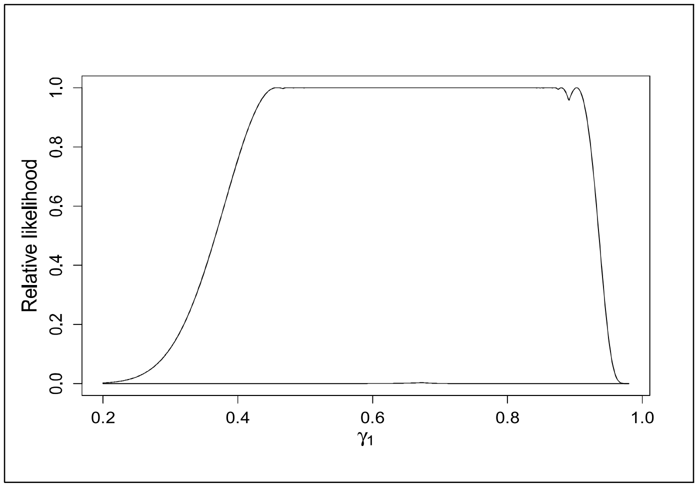

Figure 8: Enveloping curves for relative likelihood functions of γ_1, for simulated skew normal samples where: $TV = 0.1$, $\alpha = 3$ and $n = 200$. Source: Elaborated by the authors.

in order to quantify the occurrence of flat relative profile likelihoods for parameter α.

Table 3 shows a classification of the simulated samples, according to whether $0 \leq \widetilde{R}_{\max}(\infty) < 1$ or $\widetilde{R}_{\max}(\infty) = 1$. Note that $\widetilde{R}_{\max}(\infty) = 1$ corresponds to samples where the MLE of α diverged to ∞. It is important to mention that those scenarios where at least $10\,\%$ of their samples fell in category $\widetilde{R}_{\max}(\infty) = 1$, (five hundred or more of their samples), at least $95.14\,\%$ of them yielded a p-value greater than 0.05 in a Kolmogorov-Smirnov test, providing no evidence against a shifted half normal model. Thus, at 0.05 significance level, a shifted half normal model is not rejected in a Kolmogorov-Smirnov test, $95\,\%$ of the times, independently of sample size n and the α value considered in the simulation scenario; see Tabla 1. Those scenarios with few samples in category $\widetilde{R}_{\max}(\infty) = 1$ were excluded from this analysis, in order to obtain a reliable estimate for the frequency of p-values smaller than 0.05.

On the other hand, when divergence of the MLE of α is not met, the limit of the relative profile likelihood of α, denoted here as $R(\infty)$, has an interesting behavior. For small sample sizes it is highly probable to observe flat likelihoods, but when sample sizes are large, frequency distribution is located and concentrated close to zero. All these can be observed in Figure 9. Scenarios where $TV = 0.006$ ($\alpha = 50$) are not included in this figure, since the limited number of $R(\infty)$ values could yield to non-reliable comparisons. Nevertheless, similar behavior of $R(\infty)$ distribution has been observed in many performed simulations, but due to space limitation are not included here.

In summary, simulation results showed that in a skew normal sample whose relative profile likelihood of α reaches its maximum at infinity, there is a large probability that an embedded shifted half normal model will not be rejected in a hypothesis test. On the other hand, when the relative profile likelihood of α is smaller

Table 3: Classification (in percentage) of 5000 simulated samples $X \sim SN(0,1,\alpha)$.

α	n	$0 \le \widetilde{R}_{\max}(\infty) < 1$	$\widetilde{R}_{\max}(\infty) = 1$
3	25	64.18	35.82
	50	91.66	8.34
	100	99.56	0.44
	200	100	0
5	25	37.76	66.24
	50	72.86	27.14
	100	95.04	4.96
	200	99.86	0.14
11	25	10.48	89.52
	50	32.30	67.70
	100	65.48	34.52
	200	92.28	7.72
50	25	2.72	97.28
	50	3.94	96.06
	100	9.24	90.76
	200	26.26	73.74

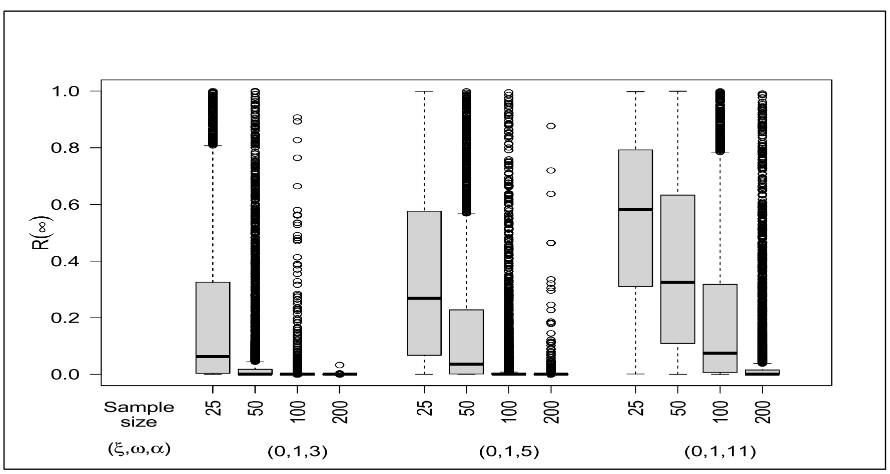

Figure 9: Box plots for three of the skew normal simulation scenarios presented in the Table 1. Source: Elaborated by the authors.

than 1, the relationship between α and n will determine the degree of flatness of the likelihood function. For instance, in Figure 9, for the case where $\alpha \leq 3$ y $n \geq 200$, it is possible to construct 95 % likelihood-confidence intervals with a finite upper bound, since most of the times $R(\infty) \approx 0$; thus, for any of these skew normal samples, likelihood shape is informative and provides reasonable inferences.

5. CONCLUSIONS

Possible conditions of the divergence of the MLE for shape parameter in a skew normal model were studied here, throughout the analysis of its likelihood and profile likelihood function. The two particular cases included in this manuscript and the exhaustive simulation study that was carried out, allowed us to analyze the behavior of the shape of the likelihood function for this parameter. The occurrence of flat likelihoods for shape parameter, where $R(\infty) = 1$, suggested a practical parameter non-identifiability of a skew normal model, observing than an embedded half normal model, or a shifted one, are usually not rejected under a hypothesis test. Some other results presented here, show the fundamental role between shape parameter values and sample size, for determining the flatness of the likelihood function.

As we have shown in this work, flatness of a likelihood function is informative and should not be ignored, much less neglected. The cases discussed in this manuscript reveal, for certain kind of reparametrization, practical non-identifiability problems that can be caused by embedded models not rejected by the data. Flat likelihood regions can also occur when data is unable to provide enough information, either to distinguish nested models or to adequately estimate parameters involved in the distribution under study. The analysis of the shape of the likelihood function and particularly its flatness yields valuable information that usually uncover aspects that must be carefully analyzed. This issue is particularly important when inferential methods are carried out in family models involving nested or embedded models and containing many parameters and covariables.

References

Allard, D. & Naveau, P. (2008). A new spatial skew-normal random field model. *Communications in Statistics - Theory and Methods*, 36(**9**), 1821-1834.

Arellano-Valle, R. B. & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of multivariate analysis*, 99(**7**), 1362-1382.

Arrué, J., Arellano-Valle, R. B. & Gómez, H. W. (2016). Bias reduction of maximum likelihood estimates for a modified skew-normal distribution, *Journal of Statistical Computation and Simulation*, 86(**15**), 2967-2984.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.

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

Azzalini, A. & Arellano-Valle, R. B. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143(**2**), 419-433.

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

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). Inference and asymptotics. Chapman & Hall/CRC. Boca Raton.

Barndorff-Nielsen, O. E. (1988). Parametric Statistical Models and Likelihood. Lecture Notes in Statistics. Springer, New York.

Cole, D. J. (2020). Parameter Redundancy and Identifiability. Chapman and Hall/CRC. New York.

Figueiredo, F. & Gomes, M. I. (2013). The skew-normal distribution in SPC. *REVSTAT - Statistical Journal*, 11(**1**), 83-104.

Gupta, A. K., Nguyen, T. T. & Sanqui, J. A. (2004). Characterization of the skew-normal distribution. *Annals of the Institute of Statistical Mathematics*, 56, 351-360.

Genton, M. G. (2004). Skew-elliptical Distributions and Their Applications: a Journey Beyond Normality. Chapman & Hall/CRC. Boca Raton.

Gupta, R. C. & Brown, N. (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Communications in Statistics-Theory and Methods*, 30, 2427-2445.

Hollander, M., Wolfe, Douglas A. & Chicken, E. (2014). Nonparametric Statistical Methods. Wiley. New Jersey.

Hossain, A. & Beyene, J. (2015). Application of skew-normal distribution for detecting differential expression to microRNA data, *Journal of Applied Statistics*, 42(**3**), 477-491.

Huber, P. J. (1981). Robust Statistics. Wiley. New York

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

Jurečková J. & Picek, J. (2006). Robust Statistical Methods with R. Chapman & Hall/CRC. New York.

Kalbfleisch, J. G. (1985). Probability and Statistical Inference, Vol. 2. Springer-Verlag. New York.

Murphy, S. A. & Van Der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(**450**), 449-465.

Pawitan, Y. (2001). In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press. New York.

Pewsey, A. (2000). Problems of inference for Azzalini's skew normal distribution, *Journal of Applied Statistics*, 27(**2**), 859-870.

Rieder, H. (1994). Robust Asymptotic Statistics. Springer-Verlag. New York.

Rubio, F. J. & Genton, M. G. (2016). Bayesian linear regression with skew-symmetric error distributions with applications to survival analysis. *Statistics in Medicine*, 35(**14**), 2441-2454.

Serfling, R. J. (2002). Approximation Theorems of Mathematical Statistics. John Wiley & Sons. New York.

Sprott, D. A. (2000). Statistical inference in science. Springer-Verlag. New York.