

Article

Generalized Skew-Normal Negentropy and Its Application to Fish Condition Factor Time Series

Reinaldo B. Arellano-Valle ¹, Javier E. Contreras-Reyes ^{2,3,4,*}  and Milan Stehlík ^{4,5,6} 

¹ Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile; reivalle@mat.puc.cl

² Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile

³ Departamento de Evaluación de Recursos, División de Investigación Pesquera, Instituto de Fomento Pesquero (IFOP), Valparaíso 2361827, Chile

⁴ Instituto de Estadística, Universidad de Valparaíso, Valparaíso 2360102, Chile; mlnstehlik@gmail.com

⁵ Institute of Applied Statistics, Johannes Kepler University Linz, Linz 4040, Austria

⁶ Linz Institute of Technology, Johannes Kepler University Linz, Linz 4040, Austria

* Correspondence: jecontrr@uc.cl or javier.contreras@ifop.cl; Tel.: +56-9-7960-8218

Received: 7 September 2017; Accepted: 30 September 2017; Published: 6 October 2017

Abstract: The problem of measuring the disparity of a particular probability density function from a normal one has been addressed in several recent studies. The most used technique to deal with the problem has been exact expressions using information measures over particular distributions. In this paper, we consider a class of asymmetric distributions with a normal kernel, called Generalized Skew-Normal (GSN) distributions. We measure the degrees of disparity of these distributions from the normal distribution by using exact expressions for the GSN negentropy in terms of cumulants. Specifically, we focus on skew-normal and modified skew-normal distributions. Then, we establish the Kullback–Leibler divergences between each GSN distribution and the normal one in terms of their negentropies to develop hypothesis testing for normality. Finally, we apply this result to condition factor time series of anchovies off northern Chile.

Keywords: Shannon entropy; negentropy; skew-normal; modified skew-normal; Kullback–Leibler divergence; condition factor

1. Introduction

Recent studies deal with the problem of measuring the disparity of a particular probability density function (pdf) from the normal one [1]. A typical technique to deal with the problem has been exact expressions using information measures over particular distributions. For example, Vidal et al. [2] measure the sensitivity of the skewness parameter using the L_1 distance between symmetric and asymmetric distributions. Stehlík [3] proved results on the decomposition of Kullback–Leibler (KL) divergences [4] in the gamma and normal family for divergence between the Maximum Likelihood Estimator (MLE) of the canonical parameter and the canonical parameter of the regular exponential family [5]. Contreras-Reyes and Arellano-Valle [6] considered Jeffrey’s (J) divergence [7] to compare the multivariate Skew-Normal (SN) from the normal distribution, and Gómez-Villegas et al. [8] assessed the effect of kurtosis deviations from normality on conditional distributions, such as the multivariate exponential power family. Main et al. [9] evaluated the local effect of asymmetry deviations from normality using the KL divergence measure of the SN distribution and then compared the local sensitivity with Mardia’s and Malkovich–Afifi’s skewness indexes. They also agree on the use of the SN model to regulate the asymmetry of an empirical distribution because it reflects the deviation in a tractable way. Dette et al. [10] characterizes the “disparity” between the skew-symmetric models and their symmetric counterparts in terms of the total variation distance, which is later used to construct

priors. The paper provides additional insights, to those provided in Vidal et al. [2], on the interpretation of this distance and also discusses the usage of the KL divergence among several other distances.

Some recent applications of measuring the disparity of a particular pdf from the normal one using negentropy include those by Gao and Zhang [11] and Wang et al. [12], where the negentropy method has been successfully applied to seismic wavelet estimation. Pires and Ribeiro [13] considered the negentropy to measure the distance of non-Gaussian information from the normal one in independent components, with application to Northern Hemispheric winter monthly variability of a high-dimensional quasi-geostrophic atmospheric model. Furthermore, Pires and Hannachi [14] used a tensorial invariant approximation of the multivariate negentropy in terms of a linear combination of squared coskewness and cokurtosis. Then, the method was applied to global sea surface temperature anomalies, after data anomalies were tested through a non-Gaussian distribution.

In this paper, we develop a procedure, based on KL divergences, to test the significance of the skewness parameter in the Generalized Skew-Normal (GSN) distributions, a flexible class of distributions that includes the SN and normal ones as particular cases. We consider asymptotic expansions of moments and cumulants for the negentropy of two particular cases: the SN and Modified Skew-Normal (MSN) distributions. Given that SN distributions do not accomplish the regularity condition of Fisher Information Matrix (FIM) at $\eta = 0$, normality is tested based on the MSN distribution [15]. This allows one to implement an asymptotic normality test for testing the significance of the skewness parameter. Numerical results are studied by: (a) comparing numerical integration methods with proposed asymptotic expansions; (b) comparing the asymptotic test with the likelihood ratio test and the asymptotic normality test given by Arrué et al. [15]; and (c) applying the proposed test to condition factor time series of anchovy (*Engraulis ringens*).

This paper is organized as follows: information theoretic measures are described in Section 2. In Section 3, we provide an asymptotic expansion in terms of the corresponding cumulants for the GSN, SN and MSN negentropies. We also express the KL and J divergences among each GSN distribution and the normal one in terms of negentropies (as cumulants' expansion series) to develop the hypothesis test about the significance of the skewness parameter together with a simulation study (Section 4). A simulation study is given in Section 5. In Section 6, the real data of the condition factor time series of anchovies off northern Chile illustrate the usefulness of the developed methodology. The discussion concludes the paper.

2. Shannon Entropy and Related Measures

The Shannon Entropy (SE) of a random variable Z with pdf f is given by:

$$H(Z) = -E\{\log f(Z)\}. \quad (1)$$

The SE of a localization-scale random variable $X = \mu + \sigma Z$ does not depend on μ and is such that $H(X) = \log \sigma + H(Z)$ (see, e.g., [16]). The SE could serve to define a measure of disparity from normality, the so-called negentropy [17], which is zero for a Gaussian variable and positive for any distribution. It is defined by:

$$N(Z) = H(Z'_0) - H(Z), \quad (2)$$

where Z'_0 is a normal random variable with the same mean and variance as those of Z . Equation (2) expresses the negentropy in terms of the standardized version of Z , say Z^* , as $N(Z) = H(Z_0) - H(Z^*) = N(Z^*)$; here, Z^* has zero mean and unit variance. Thus, negentropy measures essentially the amount of information that departs from the normal entropy. Furthermore, clearly, the negentropy becomes the KL divergence (see Equation (3) below) between Z^* and Z_0 .

Given that the calculus of negentropy presents a computational challenge, where the integral involves the pdf of Z [16,18], different approximations of negentropy are used, such as cumulants' expansion series [17,19]. Withers and Nadarajah [19] provided exact and explicit series expansions

for the SE and negentropy of a standardized pdf f on $\mathbb{R}$, in terms of cumulants. Yet, they did not perform numerical studies that allow evaluation and comparison with other procedures in some specific families of distributions.

Other measures related to the SE are KL and J divergences. They measure the degree of divergence between the distributions of two random variables Z_1 and Z_2 with pdfs f_1 and f_2 , respectively. The KL divergence of the pdf for Z_1 from the pdf for Z_2 is defined as:

$$K(Z_1, Z_2) = E \left[\log \left\{ \frac{f_1(Z_1)}{f_2(Z_1)} \right\} \right]. \quad (3)$$

as indicated in the notation, the expectation is defined with respect to the pdf f_1 for Z_1 . Since in general $K(Z_1, Z_2)$ differs from $K(Z_2, Z_1)$, the J divergence is considered as a symmetric version of the KL divergence, which is defined by:

$$J(Z_1, Z_2) = K(Z_1, Z_2) + K(Z_2, Z_1). \quad (4)$$

3. Generalized Skew-Normal Distributions

An attractive class of Skew-Symmetric (SS) distributions defined in terms of the pdf appears in Azzalini [20], Azzalini and Capitanio [21] and Gupta et al. [22]:

$$f(z; \eta) = 2f(z)G\{w(z; \eta)\}, \quad z \in \mathbb{R}, \quad (5)$$

where $\eta \in \mathbb{R}$ represents a skewness/shape parameter, f and G are the respective pdf and cumulative distribution function (cdf) of symmetrical continuous distributions and $w(z; \eta)$ is an odd function of z , with $w(0; \eta) = 0$ for any fixed value of η . Furthermore, we assume that $w(z; \eta_0) = 0$ for all z and some value η_0 of η (typically $\eta_0 = 0$), so that $f(z; \eta_0) = f(z)$, thus recovering symmetry.

The notation $Z \sim SS(\eta; f, G, w)$ expresses that random variable Z has a distribution with the pdf given by (5). If $f(z) = \phi(z)$ represents the pdf of the standardized normal distribution, denoted by $N(0, 1)$, then (5) becomes a family of skew-symmetric distributions generated by the normal kernel, the GSN family. In this case, $Z \sim GSN(\eta; G, w)$ emerges. An important property of the GSN random variable Z is that all its moments are finite. In particular, it possesses the same even moments of $Z_0 \sim N(0, 1)$. For instance, $E(Z^2) = 1$, and so, $Var(Z) = 1 - \mu_z^2$, where $\mu_z = E(Z)$. The most popular GSN distribution is Skew-Normal (SN) [23], for which $w(z; \eta) = \eta z$ and $G(z) = \Phi(z)$ is the cdf of the standardized normal distribution. Therefore, $Z \sim SN(\eta)$ expresses that Z follows an SN distribution. The location-scale extension of the SS pdf in (5) follows by applying the Jacobian method to the linear random variable $X = \mu + \sigma Z$, where $\mu \in \mathbb{R}$ and $\sigma > 0$. In this case, we state that X follows an SS distribution with location parameter μ , scale parameter σ and shape/skewness parameter η and obtains $X \sim SS(\mu, \sigma^2, \eta; f, G, w)$. Furthermore, we write $X \sim GSN(\mu, \sigma^2, \eta; G, w)$ if $f = \phi$, $X \sim GSN(\mu, \sigma^2, \eta; G)$ if $f = \phi$ and $w(z; \eta) = \eta z$; and $X \sim SN(\mu, \sigma^2, \eta)$ if $f = \phi$, $w(z; \eta) = \eta z$ and $G = \Phi$.

Two other members of the GSN family that have been studied recently are the Skew-Normal-Cauchy (SNC) distribution [24,25], which follows from (5) by taking $f(z) = \phi(z)$, $w(z; \eta) = \eta z$ and $G(z) = 1/2 + (1/\pi)\arctan(z)$, and the Modified Skew-Normal (MSN) distribution [15], for which $f(z) = \phi(z)$, $w(z; \eta) = \eta z/\sqrt{1+z^2}$ and $G(z) = \Phi(z)$. Nadarajah and Kotz [24] recall that the SNC distribution appears to attain a higher degree of sharpness than the normal distribution, i.e., disparity exists from the common normal distribution produced by the skewness parameter η . A random variable Z with the SNC or MSN distribution is denoted, respectively, by $Z \sim SNC(\eta)$ or $Z \sim MSN(\eta)$ and by $X \sim SNC(\mu, \sigma^2, \eta; G)$ or $X \sim MSN(\mu, \sigma^2, \eta; G)$ for their respective location-scale extensions.

We consider the SE for the GSN subclass, i.e., the SE of $Z \sim \text{GSN}(\eta; G, w)$. Thus, assuming a normal kernel in (5), we get the GSN-SE given by:

$$H(Z) = H(Z_0) - E(\log [2G\{w(Z; \eta)\}]), \quad (6)$$

where $H(Z_0) = (1/2) \log(2\pi e)$ is the SE of Z_0 . It is assumed that a specific skewness value η_0 exists so that $w(z; \eta_0) = 0$ and so that $G\{w(z; \eta_0)\} = 1/2$, thus recovering symmetry at $\eta = \eta_0$. Therefore, at $\eta = \eta_0$, Z and Z_0 have the same distribution and thus the same SE.

Let $\mu_z = E(Z)$ and $\sigma_z^2 = \text{Var}(Z)$ be the mean and variance of $Z \sim \text{GSN}(\eta; G, w)$, respectively, which must constitute functions of the skewness parameter η . Since $\sigma_z^2 = 1 - \mu_z^2$ and $H(Z'_0) = \log \sigma_z + H(Z_0)$, we get from (2) that the negentropy of Z becomes:

$$N(Z) = \frac{1}{2} \log(1 - \mu_z^2) + E(\log [2G\{w(Z; \eta)\}]). \quad (7)$$

Since at $\eta = \eta_0$, we have by symmetry $\mu_z = 0$ and $w(Z; \eta_0) = 0$, so the negentropy in this case is null, as expected. Clearly, SE and negentropy depend on the choice of the functions $G(\cdot)$ and $w(\cdot; \eta)$. In this paper, we consider both families of GSN distributions for which $\eta_0 = 0$, with $w(z; 0) = w(0; \eta) = 0$ and $w(-z; \eta) = w(z; -\eta)$, thus following that $-Z \sim \text{GSN}(-\eta; G, w)$ and recovering the normality at $\eta = 0$. Examples of this type of functions are $w(z; \eta) = \eta u(z)$ and $w(z; \eta) = u(\eta z)$ for some odd function $u(z)$, with $u(0) = 0$. In this case, recalling that $\eta Z \stackrel{d}{=} \tau Z_\tau$, where $Z_\tau \sim \text{GSN}(\tau; G, w)$, $\tau = |\eta|$, and “ $\stackrel{d}{=}$ ” denotes equality in the distribution, we obtain:

$$E(\log [2G\{w(Z; \eta)\}]) = E(\log [2G\{w(Z_\tau; \tau)\}]), \quad (8)$$

thus $H(Z) = H(Z_\tau)$ and $N(Z) = N(Z_\tau)$. That is, the entropy and negentropy of $Z \sim \text{SN}(\eta)$ depend on the skewness parameter η only through its absolute value $\tau = |\eta|$.

We have interest in both KL and J divergences for a GSN distribution with respect to the normal distribution. that is, assuming in (3) and (4) that $Z_1 = Z \sim \text{GSN}(\eta; G, w)$ and $Z_2 = Z_0$. In this case, remembering that $\eta Z \stackrel{d}{=} \tau Z_\tau$, where $Z_\tau \sim \text{GSN}(\tau; G, w)$ and $\tau = |\eta|$, we have $K(Z, Z_0) = K(Z_\tau, Z_0)$ and $K(Z_0, Z) = K(Z_0, Z_\tau)$, with:

$$K(Z_\tau, Z_0) = E(\log [2G\{w(Z_\tau; \tau)\}]) \quad \text{and} \quad K(Z_0, Z_\tau) = -E(\log [2G\{w(Z_0; \tau)\}]). \quad (9)$$

Therefore, $J(Z, Z_0) = J(Z_\tau, Z_0)$, with:

$$J(Z_\tau, Z_0) = E(\log [2G\{w(Z_\tau; \tau)\}]) - E(\log [2G\{w(Z_0; \tau)\}]). \quad (10)$$

We also develop asymptotic expansions of the J divergence for the SN and MSN distributions from the normal distribution. To do this, we consider the following preliminary result, the proof of which stems from (9) and (10) by using the Taylor expansion of $\zeta(z; \tau) = \log[2G\{w(z; \tau)\}]$ at $z = 0$ and also because of the facts that (a) all moments of $Z_\tau \sim \text{GSN}(\tau; G, w)$ are finite and (b) Z_τ and Z_0 contain the same even moments.

Lemma 1. Consider the composite function $\zeta(z; \tau) = \log[2G\{w(z; \tau)\}]$, $\tau > 0$, by assuming that both functions $G(z)$ and $w(z; \tau)$ are infinitely differentiable; hence, also $\zeta(z; \tau)$ is infinitely differentiable at $z = 0$. If $Z_\tau \sim \text{GSN}(\tau; G, u)$, then:

$$\begin{aligned} K(Z_\tau, Z_0) &= \sum_{k=1}^{\infty} \frac{\zeta^{(k)}(0; \tau)}{k!} E(Z_\tau^k), \\ J(Z_\tau, Z_0) &= \sum_{k=1}^{\infty} \frac{\zeta^{(2k-1)}(0; \tau)}{(2k-1)!} E(Z_\tau^{2k-1}), \end{aligned} \quad (11)$$

where $\zeta^{(k)}(z; \tau)$ is the k -th derivative of $\zeta(z; \tau)$. Moreover, from (11), the expressions (6) and (7) for the SE and negentropy of the GSN distributions have the forms:

$$\begin{aligned} H(Z_\tau) &= \frac{1}{2} \log(2\pi e) - \sum_{k=1}^{\infty} \frac{\zeta^{(k)}(0; \tau)}{k!} E(Z_\tau^k), \\ N(Z_\tau) &= \frac{1}{2} \log(1 - \mu_\tau^2) + \sum_{k=1}^{\infty} \frac{\zeta^{(k)}(0; \tau)}{k!} E(Z_\tau^k), \end{aligned}$$

respectively, where $\mu_\tau = E(Z_\tau)$.

Notice in Lemma 1 that the coefficient $\zeta^{(k)}(0; \tau)$ depends on the derivatives of $G(z)$ and $w(z; \tau)$ at $z = 0$, which change for different GSN distributions. Moreover, since the expansion of $\zeta(z; \tau)$ emerges around $z = 0$ by assuming a fixed τ , the approximations may not be reasonable for some values of τ .

3.1. Skew-Normal Distribution

If $Z \sim SN(\eta)$ or $Z \sim SN(0, 1, \eta)$ represents an SN random variable, then its pdf is:

$$f(z; \eta) = 2\phi(z)\Phi(\eta z), \quad z \in \mathbb{R}.$$

Clearly, if $\eta = \eta_0 = 0$, then (12) reduces to the $N(0, 1)$ -pdf. The SN random variable Z can be conveniently represented as a linear combination of half-normal and normal variables through the following stochastic representation [26]:

$$Z \stackrel{d}{=} \delta |U_0| + \sqrt{1 - \delta^2} U, \quad (12)$$

where $\delta = \eta / \sqrt{1 + \eta^2}$, U_0 and U are independent and identically distributed with a unit normal distribution. In particular, since the half-normal random variable $|U_0|$ has mean $b = \sqrt{2/\pi}$ and variance one, it follows from (12) that the mean and variance of Z_τ , $\tau = |\eta|$, are given by:

$$E(Z_\tau) = b\delta_\tau \quad \text{and} \quad \text{Var}(Z_\tau) = 1 - (b\delta_\tau)^2, \quad (13)$$

where $\delta_\tau = \tau / \sqrt{1 + \tau^2}$.

In the SN case, $G(z) = \Phi(z)$ and $w(z; \eta) = \eta z$, which are both infinitely differentiable functions at $z = 0$. Consequently, the function $\zeta(z; \tau) = \zeta_0(\tau z) = \log\{2\Phi(\tau z)\}$ is also infinitely differentiable at $z = 0$, thus admitting a Taylor expansion about zero. Therefore, since $\zeta^{(k)}(z; \tau) = \tau^k \zeta_k(\tau z)$, where $\zeta_k(x)$ is the k -th derivative of $\zeta_0(x) = \log\{2\Phi(x)\}$, the expansion (11) in Lemma 1 of $E\{\zeta(Z_\tau; \tau)\} = E\{\zeta_0(\tau Z_\tau)\}$, $Z_\tau \sim SN(\tau)$, becomes:

$$E\{\zeta_0(\tau Z_\tau)\} = \sum_{k=1}^{\infty} \frac{\tau^k \kappa_k \mu_k}{k!}, \quad (14)$$

where $\kappa_k = \zeta_k(0)$ and $\mu_k = E(Z_\tau^k)$ (see Appendix A).

In summary, since the even moments of $Z_\tau \sim SN(\tau)$ are also the even moments of Z_0 , Equation (14) can be rewritten as:

$$E\{\zeta_0(\tau Z_\tau)\} = \underbrace{\sum_{k=1}^{\infty} \frac{\tau^{2k} \kappa_{2k}}{(2k)!} \mu_{2k}}_{E\{\zeta_0(\tau Z_0)\}} + \underbrace{\sum_{k=1}^{\infty} \frac{\tau^{2k-1} \kappa_{2k-1}}{(2k-1)!} \mu_{2k-1}}_{J(Z_\tau, Z_0)}.$$

Hence, considering also Equation (13), we can compute for the SN case the results for the KL and J divergences, SE and negentropy given in Lemma 1 using the following Proposition 1.

Proposition 1. Let $Z_\tau \sim SN(\tau)$ and $Z_0 \sim N(0, 1)$. Then:

$$\begin{aligned} K(Z_\tau, Z_0) &= \sum_{k=1}^{\infty} \frac{\tau^k \kappa_k \mu_k}{k!}, \\ J(Z_\tau, Z_0) &= b \sum_{k=1}^{\infty} \frac{\kappa_{2k-1} \delta_\tau^{2k-1}}{(2k-1)!} \sum_{m=1}^k a_k(m) \tau^{2m-1}, \\ H(Z_\tau) &= \frac{1}{2} \log(2\pi e) - K(Z_\tau, Z_0), \\ N(Z_\tau) &= \frac{1}{2} \log \{1 - (b\delta_\tau)^2\} + K(Z_\tau, Z_0), \end{aligned} \quad (15)$$

where the coefficients $a_k(m)$, $k = 1, 2, \dots$, are given in the Appendix A.

To gain a more complete analysis of the behavior of these series, we need appropriate forms for the calculation of the coefficients $\kappa_k = \zeta_k(0)$, $k = 1, 2, \dots$ (see Appendix A).

3.2. Modified Skew-Normal Distribution

The pdf for a random variable Z with MSN distribution, denoted by $Z \sim MSN(\eta)$, is given by:

$$f(z) = 2\phi(z)\Phi\{\eta u(z)\}, \quad z \in \mathbb{R},$$

where $u(z) = z/\sqrt{1+z^2}$. Similarly to the SN case, the MSN random variable $Z_\tau \sim MSN(\tau)$, $\tau = |\eta|$, has even moments equal to the corresponding even moments of the standardized normal random variable Z_0 [15], i.e., $\mu_{2k} = (2k-1)!! = (2k)!/(2^k k!)$ (for odd moments $\mu_{2k-1}(\tau)$, $k = 1, 2, \dots$, see Appendix A).

In the MSN case, $G(w) = \Phi(w)$ and $w(z; \tau) = \tau u(z) = \tau z/\sqrt{1+z^2}$, both of which are infinitely differentiable at $z = 0$. Thus, in Lemma 1, we have $\zeta(z; \tau) = \zeta_0\{\tau u(z)\}$, where $\zeta_0(x) = \log\{2\Phi(x)\}$ is also infinitely differentiable at $z = 0$. Thus, the series expansion of $E\{\zeta(Z_\tau; \tau)\} = E[\zeta_0\{\tau u(Z_\tau)\}]$, $Z_\tau \sim MSN(\tau)$, can be obtained from (11) for which we need the derivatives of the composite function $\zeta_0\{\tau u(z)\} = \log[2\Phi\{\tau u(z)\}]$. Another way to obtain these derivatives is to define random variable $Z_\tau^* = u(Z_\tau) = Z_\tau/\sqrt{1+Z_\tau^2}$ and using (14) with Z_τ and $\mu_k = E(Z_\tau)$ replaced by Z_τ^* and $\mu_k^* = E\{(Z_\tau^*)^k\}$, respectively. Thus, we obtain the series expansion:

$$E\{\zeta_0(\tau Z_\tau^*)\} = \sum_{k=1}^{\infty} \frac{\tau^k \kappa_k \mu_k^*}{k!}.$$

From Lemma 1, the KL and J divergences, SE and negentropy for the MSN case can be computed using the following Proposition 2.

Proposition 2. Let $Z_\tau \sim MSN(\tau)$ and $Z_0 \sim N(0, 1)$. Then:

$$\begin{aligned} K(Z_\tau, Z_0) &= \sum_{k=1}^{\infty} \frac{\tau^k \kappa_k}{k!} \mu_k^*, \\ J(Z_\tau, Z_0) &= \sum_{k=1}^{\infty} \frac{\tau^{2k-1} \kappa_{2k-1}}{(2k-1)!} \mu_{2k-1}^*, \\ H(Z_\tau) &= \frac{1}{2} \log(2\pi e) - K(Z_\tau, Z_0), \\ N(Z_\tau) &= \frac{1}{2} \log(1 - [b\{2\zeta_1(\tau) - 1\}]^2) + K(Z_\tau, Z_0). \end{aligned} \quad (16)$$

In order to compute the quantities given by Proposition 2, we need to calculate the new moments $\mu_k^* = E\{(Z_\tau^*)^k\}$, $k = 1, 2, \dots$. Since $Z_\tau^* = Z_\tau/\sqrt{1+Z_\tau^2}$ is a random variable limited to the interval

$(-1, 1)$, all its moments are finite. In particular, Z_τ^* clearly has the same even moments as $Z_0^* = Z_0 / \sqrt{1 + Z_0^2}$. Moreover, from the Jacobian method, the pdf of Z_τ^* becomes:

$$f^*(u) = \frac{2}{(1-u^2)^{3/2}} \phi\left(\frac{u}{\sqrt{1-u^2}}\right) \Phi(\tau u), \quad u \in (-1, 1).$$

Hence, the k -th moment of Z_τ^* is:

$$\mu_k^* = \int_{-1}^1 u^k f^*(u) du, \quad k = 1, 2, \dots,$$

which must be computed numerically.

3.3. J Divergence between SN and MSN Distributions

In the previous sections, SN and MSN distributions were compared with the normal distribution by means of the J divergence measure. As a byproduct, we were also computing the J divergence between the SN and MSN distributions, both with the same skewness parameter. This allows measuring the distance between these distributions with different $w(z; \eta)$'s. For this, we consider in Equation (4) that $Z_1 \sim SN(\tau)$ and $Z_2 \sim MSN(\tau)$ and define the random variables $Z_i^* = Z_i / \sqrt{1 + Z_i^2}$ for $i = 0, 1, 2$, where $Z_0 \sim N(0, 1)$. Let $\mu_{i,k} = E(Z_i^k)$ and $\mu_{i,k}^* = E\{(Z_i^*)^k\}$, $i = 0, 1, 2$. Recall that $\mu_{i,2k} = \mu_{0,2k}$ and $\mu_{i,2k}^* = \mu_{0,2k}^*$ for all $k = 1, 2, \dots$. Thus, using (4) and then the Taylor expansion of $\zeta_0(x) = \log\{\Phi(x)\}$ around $x = 0$, Proposition 3 is obtained:

Proposition 3. Let $Z_0 \sim N(0, 1)$, $Z_1 \sim SN(\tau)$ and $Z_2 \sim MSN(\tau)$. Define the random variables $Z_i^* = Z_i / \sqrt{1 + Z_i^2}$, $i = 1, 2$. Then:

$$\begin{aligned} J(Z_1, Z_2) &= E\{\zeta_0(\tau Z_1)\} - E\{\zeta_0(\tau Z_1^*)\} + E\{\zeta_0(\tau Z_2^*)\} - E\{\zeta_0(\tau Z_2)\} \\ &= J(Z_1, Z_0) - J(Z_1^*, Z_0) + J(Z_2^*, Z_0) - J(Z_2, Z_0), \end{aligned}$$

where as before:

$$\begin{aligned} J(Z_i, Z_0) &= \sum_{k=1}^{\infty} \tau^{2k-1} \frac{\mu_{i,2k-1} \kappa_{2k-1}}{(2k-1)!}, \\ J(Z_i^*, Z_0) &= \sum_{k=1}^{\infty} \tau^{2k-1} \frac{\mu_{i,2k-1}^* \kappa_{2k-1}}{(2k-1)!}, \quad i = 1, 2. \end{aligned}$$

Proposition 3 indicates that J divergence between SN and MSN distributions is decomposed to the divergences of the normal distribution with each of these distributions, which depends only on their odd moments and cumulants.

4. Asymptotic Tests

Let $f(x; \theta)$, $x \in \mathcal{X}$, $\theta \in \Theta$, be the pdf of a regular parametric class of distributions, i.e., for which the sample space $\mathcal{X}$ does not depend on θ , the parametric space Θ is an open subset of $\mathbb{R}^p$, and the regularity conditions (i)–(iii) stated in Salicrú et al. [27] are satisfied. As in Salicrú et al. [27], we denote the KL divergence between $f(x; \theta)$ and $f(x; \theta')$, $\theta, \theta' \in \Theta$, by:

$$K(\theta, \theta') = \int_{\mathcal{X}} f(x; \theta) \log \left\{ \frac{f(x; \theta)}{f(x; \theta')} \right\} dx.$$

Consider the partition $\theta = (\theta_1, \theta_2)$, where $\theta_1 \in \Theta_1 \subset \mathbb{R}^r$ and $\theta_2 \in \Theta_2 = \Theta \cap \Theta_1^c \subset \mathbb{R}^{p-r}$. Let $\theta' = (\theta_1, \theta_2^0)$ and consider the null hypothesis $H_0 : \theta_2 = \theta_2^0$ for a known $\theta_2^0 \in \Theta \cap \Theta_1^c$. Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ and $\hat{\theta}' = (\hat{\theta}_1, \hat{\theta}_2^0)$ be the (unrestricted) MLE of θ and θ' , respectively, both based on a random sample of size n from X with pdf $f(x; \theta)$. Under these conditions, we have from Part (b) of Theorem 2 presented in Salicrú et al. [27] that:

$$2nK(\hat{\theta}, \hat{\theta}') = 2n \int_{\mathcal{X}} f(x; \hat{\theta}) \log \left\{ \frac{f(x; \hat{\theta})}{f(x; \hat{\theta}')} \right\} dx \xrightarrow[n \rightarrow \infty]{d} \chi_{p-r}^2, \quad \forall \theta \in \Theta, \quad (17)$$

where " $\xrightarrow{d}$ " denotes convergence in distribution and χ_s^2 denotes the chi-squared distribution function with s degrees of freedom. From (17), the above null hypothesis can be tested by the statistic $2nK(\hat{\theta}, \hat{\theta}')$, which is asymptotically chi-squared distributed with $p - r$ degrees of freedom. Specifically, for large values of n , if we observe $K(\hat{\theta}, \hat{\theta}') = K_0$, then H_0 is rejected at level α if $P(\chi_{p-r}^2 > 2nK_0) \leq \alpha$.

4.1. One-Sample Case: Test for Normality

The result in (17) can be applied for example to construct a normality test from the KL divergence between a regular GSN distribution and the normal distribution. Specifically, consider a random sample $X_1, \dots, X_n$ from $X \sim \text{GSN}(\mu, \sigma^2, \eta, G, w)$ and the null hypothesis $H_0 : \eta = \eta_0$ under which $G\{w(z; \eta_0)\} = G(0) = 1/2$; thus, the GSN random variable X becomes a $N(\mu, \sigma^2)$ random variable. Let $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\eta})$ be the MLE of $\theta = (\mu, \sigma^2, \eta)$ and $\hat{\theta}' = (\hat{\mu}, \hat{\sigma}^2, \eta_0)$. Therefore, under $H_0 : \eta = \eta_0$, we have:

$$2nK(\hat{\theta}, \hat{\theta}') = 2nK(Z_{\hat{\tau}}, Z_0) \xrightarrow[n \rightarrow \infty]{d} \chi_1^2, \quad (18)$$

where $K(Z_{\hat{\tau}}, Z_0)$ is the MLE of $K(Z_{\tau}, Z_0)$, which is defined in Equation (11) of Lemma 1 and depends only on $\hat{\tau} = |\hat{\eta}|$. As stated in the Introduction, normality is typically obtained from the GSN class at $\eta_0 = 0$ or equivalently $\tau_0 = |\eta_0| = 0$.

Azzalini [20], Arellano-Valle and Azzalini [28] and Azzalini and Capitanio [23] recall the singularity of SN FIM at $\eta = 0$, preventing the asymptotic distribution of the above statistic tests. As suggested by Azzalini [20], a solution to recover the non-singularity of the information matrix under the symmetry hypothesis comes from the use of the so-called centered parametrization defined in terms of the mean, variance and the skewness parameters of the SN distribution (see also [28,29]). Otherwise, the FIM of the MSN model is non-singular at $\eta = 0$ [15]. Thus, this model satisfies all the standard regularity conditions of Salicrú et al. [27], leading to consistence and asymptotic normality of the MLEs under the null hypothesis of normality. Therefore, the MSN model serves to test the null hypothesis of normality using (18). Hence, the symmetry null hypothesis $H_0 : \tau = 0$ is rejected at level α if $P(\chi_1^2 > 2nK_0) \leq \alpha$, with $K_0 = K(Z_{\hat{\tau}}, Z_0)$.

4.2. Two-Sample Case

Consider two independent samples of sizes n_1 and n_2 from X_1 and X_2 , respectively; where $\theta, \theta' \in \Theta \subset \mathbb{R}^p$, and X_1 and X_2 have pdf's $f(x; \theta_1)$ and $f(x; \theta_2)$, respectively. Suppose partition $\theta_i = (\theta_{i1}, \theta_{i2})$, $i = 1, 2$, and assume $\theta_{21} = \theta_{11} \in \Theta_1 \subset \mathbb{R}^r$, so that $\theta_{i2} \in \Theta \cap \Theta_1^c \subset \mathbb{R}^{p-r}$, $i = 1, 2$. Let $\hat{\theta}_i = (\hat{\theta}_{i1}, \hat{\theta}_{i2})$ be the MLE of $\theta_i = (\theta_{i1}, \theta_{i2})$, $i = 1, 2$, which correspond to the MLE of the full model parameters (θ_1, θ_2) under null hypothesis $H_0 : \theta_{21} = \theta_{11}$. Thus, Part (b) of Corollary 1 in Salicrú et al. [27] establishes that if the null hypothesis $H_0 : \theta_{22} = \theta_{12}$ holds and $\frac{n_1}{n_1 + n_2} \xrightarrow[n_1, n_2 \rightarrow \infty]{} \lambda$, with $0 < \lambda < 1$, then:

$$\frac{2n_1 n_2}{n_1 + n_2} K(\hat{\theta}_1, \hat{\theta}_2) = \frac{2n_1 n_2}{n_1 + n_2} \int_{-\infty}^{\infty} f(x; \hat{\theta}_1) \log \left\{ \frac{f(x; \hat{\theta}_1)}{f(x; \hat{\theta}_2)} \right\} dx \xrightarrow[n_1, n_2 \rightarrow \infty]{d} \chi_{p-r}^2. \quad (19)$$

Thus, a test of level α for the above homogeneity null hypothesis consists of rejecting H_0 if:

$$\frac{2n_1n_2}{n_1 + n_2} K(\hat{\theta}_1, \hat{\theta}_2) > \chi_{p-r, 1-\alpha}^2$$

where $\chi_{p-r, \alpha}^2$ is the α -th percentile of the χ_{p-r}^2 -distribution.

Contreras-Reyes and Arellano-Valle [6] considered the result of Kupperman [30] to develop an asymptotic test of complete homogeneity in terms of the J divergence between two SN distributions. The SN distribution satisfies all the aforementioned regularity conditions when skewness parameter $\eta \neq 0$. Thus, considering this condition, we can also apply (17) and (19) to obtain, respectively, asymptotic tests with one or two samples of other hypotheses not covered by Kupperman's test.

5. Simulations

In this section, we study the behavior of the series expansions of the SE and negentropy for the SN and MSN distributions. In both cases, we compare the SE and negentropies obtained from their series expansions with their corresponding "exact" versions computed from the Quadpack numerical integration method of Piessens et al. [31]. More precisely, the "exact" expected values $E\{\zeta_0(\tau Z_\tau)\}$ and $E\{\zeta_0(\tau Z_\tau^*)\}$ are computed using the Quadpack method as in Arellano-Valle et al. [16], Contreras-Reyes and Arellano-Valle [6] or Contreras-Reyes [18]. From the series expansions, the SE and negentropies were carried out for $k = 12$ as in Withers and Nadarajah [19]. However, they tend to converge for $k = 4$ as in the Gram–Charlier and Edgeworth expansion methods (see, e.g., Hyvärinen et al. [17] and Stehlík et al. [1], respectively). All proposed methods are implemented with R software [32].

From Figure 1, we observe that the approximations by series expansions are better in the MSN case (Panels C and D) than in the SN case (Panels A and B). Furthermore, that series expansion approximations are quite exact for small to moderate values of the skewness parameter τ ; more specifically, for $0 \leq \tau \leq 2$ in the SN case, and $0 \leq \tau \leq 4$ in the MSN case. Additionally, Panels A and C show that the SE decreases as τ increases, while Panels B and D indicate that the negentropy increases with τ . Finally, as expected in both GSN models, the SE is less than or equal to the SE of the normal model, namely $H(Z_0) \approx 1.418$ [6,33].

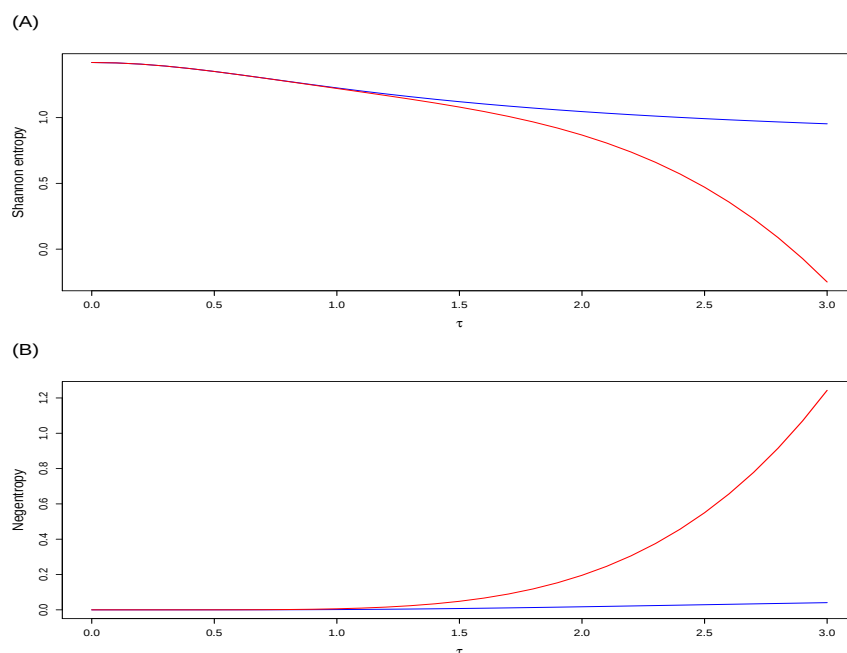


Figure 1. Cont.

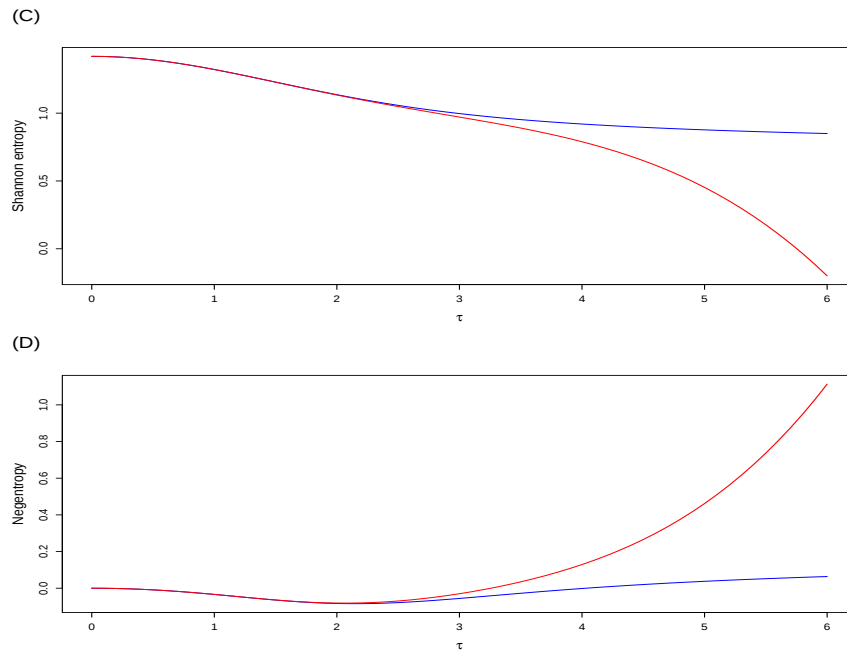


Figure 1. Shannon entropy and negentropy for the (A,B) Skew-Normal (SN) and (C,D) Modified Skew-Normal (MSN) cases. The blue and red lines correspond to numerical integration and cumulant expansion series methods, respectively.

Panel A of Figure 2 shows, respectively, the behavior of the KL divergences of the SN and MSN distributions from the normal one obtained from the expansions in series given in Equations (15) and (16). As in Figure 1, the KL divergence between the SN and normal distributions increases smoothly for values of $\tau \in [0, 2]$, but rises sharply for $\tau > 2$. Meanwhile, the increase in KL divergence between the MSN and normal distributions seems more stable, at least for $\tau \in [0, 5]$. Crucially, for $\tau = |\eta| \geq 2$, the SN model is close to its maximum level of asymmetry, while the MSN model does it for $\tau = |\eta| \geq 5$ (see [15] (Figure 2)).

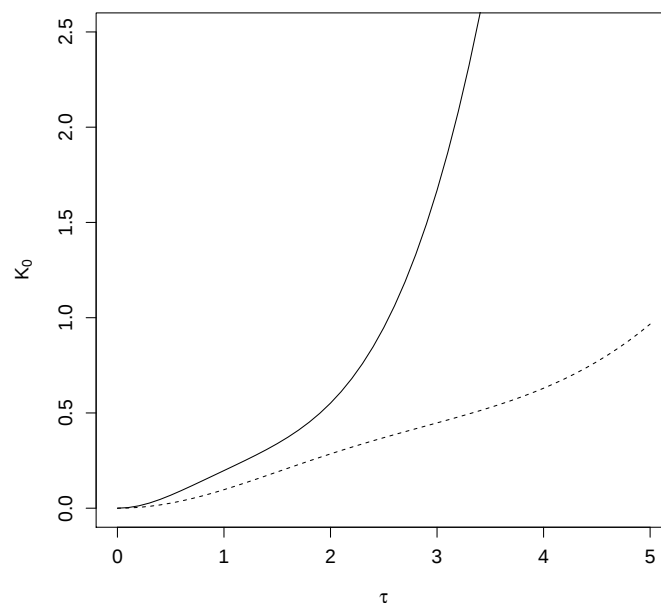


Figure 2. Kullback-Leibler (KL) divergence, $K_0 = K(Z_\tau, Z_0)$, between SN and normal (solid line) and MSN and normal (dotted line).

Table 1 presents the observed power of the asymptotic test of normality obtained from Equation (18) in Section 4.1, for different sample sizes and values of the skewness parameter. All these results were obtained from 2000 simulations for a nominal level of 5%. In each simulation, the MLE of $Z \sim MSN(\eta)$ was obtained by maximizing the log-likelihood function:

$$\ell(\eta) = \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n z_i^2 + \sum_{i=1}^n \log \zeta_0\{\eta u(z_i)\}, \quad (20)$$

for shape parameter η and a random sample of size n from Z [15]. Table 1 shows that the proposed test is considerably conservative since the observed rate of incorrect rejections of the normality hypothesis is always lower than the nominal level. The proposed test is also considerably more powerful in large samples ($n \geq 300$) and values of the skewness parameter far from zero ($|\eta| \geq 1.2$). As expected, the power of the test increases with sample size, particularly for small values of the skewness parameter (close to normality), given that statistic $2nK_0$ depends on n despite K_0 being small (Figure 2).

Table 1. Observed power (both in %) of the proposed normality test using the maximum likelihood estimator (MLE) of the MSN model from 2000 simulations for the nominal level of 5% and various values of shape parameter η and sample size n .

n/η	0	0.1	0.2	0.3	0.4	0.8	1.2	1.6	2
50	97.55	37.02	39.45	37.39	37.60	40.07	40.82	48.37	51.29
100	97.65	46.15	45.90	47.90	46.05	50.05	56.10	66.20	75.49
200	98.05	54.25	54.30	53.65	54.50	58.80	70.55	85.20	93.25
300	98.30	57.30	57.70	58.00	58.65	62.00	77.05	92.65	98.25
400	98.55	57.20	57.65	57.15	57.50	64.10	83.95	95.80	99.65
500	98.70	58.55	59.40	59.95	58.45	66.70	86.95	97.45	99.75

Now, we compare the proposed asymptotic test with two additional tests considered by Arrué et al. [15] for null hypothesis $H_0 : \eta = \eta_0$ versus $H_1 : \eta \neq \eta_0$: the Likelihood Ratio Test (LRT) (see Appendix A) and the asymptotic normality-based test. Since the regularity condition on MSN's FIM at $\eta = 0$ is satisfied, the authors proposed a distributional normal theory for testing H_0 , i.e., based on asymptotic normality of MLE given by $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N_3(0, i_{MSN}^{-1}(\theta_0))$, as $n \rightarrow \infty$, where $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\eta})^\top$ is the MLE of $\theta = (\mu, \sigma^2, \eta)^\top$, $\theta_0 = (\mu, \sigma^2, 0)^\top$ and $i_{MSN}^{-1}(\theta_0)$ is the inverse FIM component related to θ_0 . For asymptotic normality and LRT, they conclude that H_0 is rejected for large values of $\hat{\tau} = |\hat{\eta}|$, and for large values of n , the coverage rate increases when $\hat{\eta}$ exists (H_0 is rejected) (see [15] (Tables 3–5)). Analogously, in Table 6 of Arrué et al. [15], the coverage rate increase when $\hat{\eta}$ exists for large values of n .

6. Application to Condition Factor Time Series

To apply our results to a real-world problem, we considered the Condition Factor (CF) index [34], which serves as an important indicator of the fatness condition of fish [18]. The CF index,

$$CF = \frac{W}{\widehat{W}} (100\%),$$

of an individual of length L is computed in terms of the observed weight $W = W(L)$ and an estimation $\widehat{W} = \widehat{W}(L)$ obtained from the morphometric relationships of the expected weight $E(W)$ at length L . Then, the CF index is interpretable as food deficit ($<100\%$) and abundance ($>100\%$) conditions. The expected length-weight relationship is described through the non-linear relationship:

$$E(W) = \alpha L^\beta, \quad (21)$$

where α is the theoretical weight at length zero and β is the weight growth rate [35]. According to (21), $\hat{W}$ is computed as $\hat{W} = \hat{\alpha}L^{\hat{\beta}}$, where $\hat{\alpha}$ and $\hat{\beta}$ are obtained by fitting the non-linear regression induced by (21) to the length-weight data obtained from a sample of the species under study.

The CF index can be mainly affected by environmental factors such as El Niño (cold events) or La Niña (warm events). These effects are conductors of threshold biological processes due to the limitation of food. For these reasons, Contreras-Reyes [18] considered a threshold autoregressive model based on the stochastic representation (12) to model CF time series. That is, by assuming an SN distribution with skewness parameter η for the CF index [20], the condition $|\delta| < 1$ ensures the weak stationarity of the process. Additionally, when η is positive, CF values fall below 100% (food deficit). Otherwise, CF values are greater than 100% (food abundance).

We applied hypothesis testing developed in Section 4 to monthly CF time series associated with anchovy from Chile's northern coast during the period 1990–2010, which were classified by length and sex, for length classes 12,...,18 cm and ALL (all length classes). Therefore, the sample size of each classification depends on the availability of the routine biological sampling program (see more details in [18]). CF were previously standardized, since the shape parameter η is not affected by a linear transformation of the CF [23]. Table 2 shows the $\hat{\eta}$'s assuming an SN and MSN distribution based on the MLE method of Azzalini [36] and Arrué et al. [15], respectively. For MSN, we considered the log-likelihood function of Equation (20). In both models, negative and positive values of $\hat{\eta}$ correspond to asymmetry to the right and left, respectively (see Contreras-Reyes [18] (Figure 5)). This means that CF of the above-mentioned classes are affected by extreme events. As expected, we find generally that for low values of the empirical skewness index, the shape parameter of both distributions is close to zero.

Table 2. Shape parameter estimates ($\hat{\eta}$) of SN (reported in [18]) and MSN models for each sex and length class L , together with its respective standard deviations (s.d). Sample size (n), empirical skewness ($\sqrt{\hat{b}_1}$) and kurtosis ($\hat{b}_2$), as well as the log-likelihood function $\ell(\hat{\eta})$ for each model fit are also reported.

Sex	L	n	$\sqrt{\hat{b}_1}$	$\hat{b}_2$	SN			MSN		
					$\hat{\eta}$	s.d	$\ell(\hat{\eta})$	$\hat{\eta}$	s.d	$\ell(\hat{\eta})$
Male	12	213	−0.220	3.723	−1.065	0.147	−301.066	−1.717	0.662	−235.783
	13	238	0.658	5.120	1.778	0.134	−332.638	2.617	0.447	−359.291
	14	260	−0.147	2.885	−1.086	0.166	−367.935	0.459	0.542	−342.881
	15	261	−0.030	2.755	−0.442	0.150	−369.834	0.065	0.386	450.489
	16	261	0.307	3.138	1.616	0.149	−367.542	2.565	0.415	−354.073
	17	221	−3.461	27.958	−3.092	0.080	−285.125	−3.597	0.524	−312.701
	18	180	0.264	3.001	1.368	0.201	−253.988	2.821	0.643	−269.193
	All	262	0.068	2.687	0.721	0.178	−371.192	−0.287	0.517	−131.335
Female	12	198	0.041	2.738	0.552	0.196	−280.434	−0.142	0.386	−209.551
	13	228	0.917	6.103	2.267	0.128	−315.305	3.101	0.524	−326.665
	14	260	0.190	2.907	1.242	0.157	−367.555	1.459	0.867	−331.373
	15	260	0.076	2.672	0.728	0.174	−368.344	−0.189	0.383	−425.212
	16	261	0.349	3.091	1.702	0.155	−367.160	2.631	0.475	−346.705
	17	246	−0.056	3.115	−0.689	0.149	−348.487	0.041	0.754	−373.068
	18	208	−0.539	4.349	−1.581	0.136	−291.484	−2.223	0.472	−313.160
	All	262	0.072	2.764	0.748	0.172	−371.174	−0.267	0.401	−110.108

Table 3. MSN Shannon entropy (H) and negentropy (N) for each sex and length class L using expansion series of cumulants. For each time series, the KL divergence $K_0 = K(Z_{\hat{\tau}}, Z_0)$, statistic $2nK_0$ of Equation (18), the Likelihood Ratio Test (LRT) statistic and its respective p -values are reported. All values reported consider estimates $\hat{\eta}$ (for $\hat{\tau} = |\hat{\eta}|$) and sample size n from Table 2.

Sex	L	H	N	K_0	Asymptotic Test		LRT	
					$2nK_0$	p -Value	Statistic	p -Value
Male	12	1.187	−0.074	0.232	98.950	<0.001	98.304	<0.001
	13	1.030	−0.063	0.389	185.087	<0.001	180.469	<0.001
	14	1.396	−0.008	0.023	11.699	0.001	11.693	0.001
	15	1.418	0.000	0.001	0.242	0.623	0.246	0.620
	16	1.038	−0.066	0.381	198.700	<0.001	193.567	<0.001
	17	0.873	0.050	0.546	241.472	<0.001	211.924	<0.001
	18	0.999	−0.047	0.420	151.341	<0.001	146.427	<0.001
	All	1.410	−0.003	0.009	4.682	0.031	4.673	0.030
Female	12	1.417	−0.001	0.002	0.873	0.350	0.903	0.342
	13	0.956	−0.019	0.463	211.320	<0.001	198.516	<0.001
	14	1.237	−0.062	0.183	94.892	<0.001	93.300	<0.001
	15	1.415	−0.001	0.004	2.027	0.155	2.071	0.150
	16	1.028	−0.062	0.391	204.117	<0.001	199.015	<0.001
	17	1.418	0.000	2e-04	0.091	0.763	0.089	0.765
	18	1.095	−0.080	0.324	134.967	<0.001	133.629	<0.001
	All	1.411	−0.003	0.008	4.058	0.044	4.112	0.042

The values of $\hat{\eta}$ obtained from the SN and MSN models are presented in Table 2. Since that SN model is not regular at $\eta = 0$, we used only the MSN model to perform the test of normality and LRT for each sample datum. The results of this analysis appear in Table 3 and are not analogous for all the length classes in both groups. In fact, for the group of males, the null hypothesis $H_0 : \tau = 0$ is not rejected, only in length class 15 (95% confidence level) and in class ALL (90% confidence level). In contrast, for the group of females, the null hypothesis is not rejected for length classes 12, 15, 17 (95% confidence level) and in class ALL (90% confidence level). For both tests, we obtained similar decisions on each time series.

According to Contreras-Reyes [18], the time series in which the shape parameter is close to zero or when the null hypothesis is not rejected are influenced simultaneously by both normal and extreme events as in the length class ALL, where all the fish population is included for the analysis. For length class 17 in males, for example, the CF is susceptible to some atypical events such as the moderate-strong El Niño event between 1991 and 1992 (high negative empirical skewness and high empirical kurtosis). For length class 13 in both sexes, the CF is susceptible to the strong El Niño event produced between 1997 and 1998.

7. Discussion

We have presented the methodology to compute the Shannon entropy, the negentropy and the Kullback–Leibler and Jeffrey’s divergences for a broad family of asymmetric distributions with the normal kernel called generalized skew-normal distributions. Our method considers asymptotic expansions regarding moments and cumulants for two particular cases: the skew-normal and modified skew-normal distributions. We then measured the degrees of disparity of these distributions from the normal distribution by using exact expressions for the negentropy in terms of moments and cumulants. Additionally, given the regularity conditions accomplished by the modified skew-normal distribution, normality was tested based on the modified skew-normal distribution. This test considered the asymptotic behavior of the Kullback–Leibler divergence, which is determined by the negentropy for normality disparity.

Numerical results showed that the Shannon entropy and negentropy of the modified skew-normal distribution are better approximated than the skew-normal one, at least for a wider range of the shape parameter. For small to moderate values of the asymmetry parameter, where the approximations are appropriate, we find that expansions series converge from the fourth moment/cumulant to greater, as in the Gram–Charlier and Edgeworth expansion methods [17]. For large values of the skewness parameter, where the expansions are inappropriate, the functions related to negentropy are not well approximated by Taylor expansions around zero, produced by a divergence in the moment and cumulant terms, i.e., the Taylor expansions for the expected values of the functions $\zeta_0(\tau Z_\tau)$ and $\zeta_0\{\tau u(Z_\tau)\}$ (SN and MSN case, respectively) if $\tau = |\eta|$ is too large. When this happens, the normal cdf, $\Phi(\tau Z_\tau)$ and $\Phi\{\tau u(Z_\tau)\}$ (SN and MSN case, respectively), tends to one, since according to the stochastic representation in (12), for large values of τ , the distribution of Z_τ converges to the standardized half-normal distribution [37].

However, the normality test considered in the application used skewness parameters inside the appropriate range. Furthermore, we plan to investigate the negentropy of the modified skew-normal-Cauchy distribution or similar models. In addition, although the approximations are appropriate over the range of variation of the asymmetry admitted by both models, more work should be done in order to improve the asymptotic approximations for a greater range of the skewness parameter values. Besides, this is not an easy task since generally it is difficult to approximate KL divergences involving asymmetric and heavy-tailed distributions [38].

The statistical application related to condition factor time series of anchovies off northern Chile is given. The results show that the proposed methodology serves to detect non-normal events in these time series, which produces an empirical distribution with high presence of skewness [18]. The proposed test for normality is therefore useful to detect anomalies in condition factor time series, linked to food deficit (positive shape parameter) or food abundance (negative shape parameter) influenced by environmental conditions.

Acknowledgments: We are grateful to the Instituto de Fomento Pesquero (IFOP) for providing access to the data used in this work. The authors also thank Jaime Arrué for providing useful R codes of the MSN distribution. We are sincerely grateful to the four anonymous reviewers for their comments and suggestions that greatly improved an early version of this manuscript. Arellano-Valle's research was supported by Grant FONDECYT (Chile) Nos. 1120121 and 1150325. Contreras-Reyes's research was supported by Grant PIIC 069/2016 from UTFSM (Valparaíso, Chile) and by the CONICYT doctoral scholarship 2016 No. 21160618. Stehlík's research was supported by FONDECYT (Chile) No. 1151441 and LIT-2016-1-SEE-023. All R codes used in this paper are available by request to the corresponding author.

Author Contributions: Javier E. Contreras-Reyes conceived the experiments and analyzed the data; Reinaldo B. Arellano-Valle, Javier E. Contreras-Reyes and Milan Stehlík designed and performed the experiments, contributed reagents/analysis tools and wrote the paper; and Javier E. Contreras-Reyes contributed materials tools. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Moments of the Skew-Normal Distribution.

The moments $\mu_k = E(Z_\tau^k)$ are given by:

$$\begin{aligned}\mu_{2k} &= (2k-1)!! = \frac{(2k)!}{2^k k!} \quad (\text{even moments}), \\ \mu_{2k-1} &= \frac{(2k-1)! b \tau}{2^{k-1} (1+\tau^2)^{(2k-1)/2}} \sum_{j=0}^{k-1} \frac{j! (2\tau)^{2j}}{(2j+1)!(k-j-1)!} \quad (\text{odd moments; [26]}).\end{aligned}$$

From Proposition 2 in Martínez et al. [39], the odd moments can also be computed as:

$$\mu_{2k-1} = \frac{b \sum_{m=1}^k a_k(m) \tau^{2m-1}}{(1 + \tau^2)^{(2k-1)/2}},$$

where the coefficient $a_k(m)$ is computed iteratively as follows:

$$\begin{aligned} a_1(1) &= 1, \\ a_k(1) &= (2k-1)a_{k-1}(1), \quad k \geq 2, \\ a_k(m) &= 2(k-1)\{a_{k-1}(m) + a_{k-1}(m-1)\}, \quad 1 < m < k, \quad k \geq 2, \\ a_k(k) &= 2(k-1)a_{k-1}(k-1), \quad k \geq 2. \end{aligned}$$

Appendix A.2. Cumulants of the Skew-Normal Distribution

The coefficients $\kappa_k = \zeta_k(0)$, $k = 1, 2, \dots$, are related to the cumulants of the half-normal random variable $V \sim \sqrt{\chi_1^2}$ given by $K(t) = (1/2)t^2 + \zeta_0(t)$ (see also [21,23]). Let $K_m(0) = \frac{d^m}{dt^m} K(t) |_{t=0}$ be the m -th cumulant of V and clearly $K_1(0) = \zeta_1(0) = \kappa_1$, $K_2(0) = 1 + \zeta_2(0) = 1 + \kappa_2$, and $K_m(0) = \zeta_m(0) = \kappa_m$, $m = 3, 4, \dots$. Furthermore, from [21,23], it emerges that:

$$\begin{aligned} \zeta_1(x) &= \frac{\phi(x)}{\Phi(x)}, \\ \zeta_2(x) &= -\zeta_1(x)\{x + \zeta_1(x)\}, \\ \zeta_3(x) &= -\zeta_2(x)\{x + \zeta_1(x)\} - \zeta_1(x)\{1 + \zeta_2(x)\}, \\ \zeta_4(x) &= -\zeta_3(x)\{x + \zeta_1(x)\} - 2\zeta_2(x)\{1 + \zeta_2(x)\} - \zeta_1(x)\zeta_3(x), \\ \zeta_5(x) &= -\zeta_4(x)\{x + \zeta_1(x)\} - 3\zeta_3(x)\{1 + \zeta_2(x)\} - 3\zeta_2(x)\zeta_3(x) - \zeta_1(x)\zeta_4(x), \\ &\vdots \end{aligned}$$

Recalling that $b = \sqrt{2/\pi}$, the first five coefficients $\kappa_k = \zeta_k(0)$ are:

$$\begin{aligned} \kappa_1 &= b, \\ \kappa_2 &= -\kappa_1^2, \\ \kappa_3 &= -2\kappa_2\kappa_1 - \kappa_1, \\ \kappa_4 &= -2\kappa_3\kappa_1 - 2\kappa_2^2 - 2\kappa_2, \\ \kappa_5 &= -2\kappa_4\kappa_1 - 6\kappa_3\kappa_2 - 3\kappa_3, \\ &\vdots \end{aligned}$$

Thus, by letting $\kappa_0 = 1$, a recursive rule for these coefficients is obtained as follows:

$$\begin{aligned} \kappa_1 &= b, \\ \kappa_{2k} &= -(2k-2)\kappa_{2k-2} - 2 \sum_{i=1}^k \binom{2k-2}{i-1} \kappa_i \kappa_{2k-i} + \binom{2k-2}{k-1} \kappa_k^2, \\ \kappa_{2k+1} &= -(2k-1)\kappa_{2k-1} - 2 \sum_{i=1}^k \binom{2k-1}{i-1} \kappa_i \kappa_{2k-i+1}, \quad k = 1, 2, \dots \end{aligned}$$

Appendix A.3. Odd Moments of the Modified Skew-Normal Distribution

Recalling that $b = \sqrt{2/\pi}$, the odd moments can be computed as:

$$\mu_{2k-1}(\tau) = 2^{k-1}(k-1)!b\{2\tilde{\xi}_k(\tau) - 1\},$$

with:

$$\tilde{\xi}_k(\tau) = \int_0^\infty \Phi\{\tau u(\sqrt{x})\} \frac{x^{k-1}e^{-x/2}}{2^k\Gamma(k)} dx, \quad k = 1, 2, \dots,$$

where $\Gamma(\cdot)$ denotes the usual gamma function. Note that $\tilde{\xi}_k(\tau) = E[\Phi\{\tau u(\sqrt{X_k})\}]$, where $X_k \sim \chi_{2k}^2$; thus, $0 < 2\tilde{\xi}_k(\tau) - 1 < 1$ for all $k = 1, 2, \dots$ and $\tau > 0$. In particular, the first four moments are $\mu_1(\tau) = b\{2\tilde{\xi}_1(\tau) - 1\}$, $\mu_2 = 1$, $\mu_3(\tau) = 2b\{2\tilde{\xi}_2(\tau) - 1\}$, and $\mu_4 = 3$.

Appendix A.4. Likelihood Ratio Test

The Likelihood Ratio Test (LRT) statistic [40] for a null hypotheses $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \notin \Theta_0$, $\Theta_0 \subset \Theta$ (the parametric space), is given by:

$$\text{LRT} = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_0)\},$$

where $\hat{\theta}$ is the unrestricted MLE of θ , $\hat{\theta}_0$ is the MLE of θ under H_0 and the log-likelihood function $\ell(\theta)$ for MSN distributions is presented in (20). As before, normality is typically obtained from the GSN class at $\eta_0 = 0$. Because the MSN distribution satisfies the standard regularity conditions [15], the LRT statistic is asymptotically χ_s^2 distributed under H_0 , with $s = \dim(\Theta) - \dim(\Theta_0) = 1$ degrees of freedom [41]. Hence, the p -value associated with the LRT is computed as $1 - \chi_s^2(\text{LRT})$, where $\chi_s^2(\text{LRT})$ denotes the χ_s^2 -distribution function evaluated at the observed value of the LRT statistic.

In order to test normality, we considered the particular null hypothesis $H_0 : \eta = 0$ versus $H_1 : \eta \neq 0$, with the rest of the parameters not specified. Therefore, by (20), the LRT statistic is given by:

$$\text{LRT} = 2 \left[-\frac{1}{2} \sum_{i=1}^n \hat{z}_i^2 + \frac{1}{2} \sum_{i=1}^n \hat{z}_{i0}^2 + \sum_{i=1}^n \log \zeta_0\{\hat{\eta} u(\hat{z}_i)\} \right],$$

where $\hat{\eta}$ is the unrestricted MLE of η , $\hat{z}_i$ and $\hat{z}_{i0}$ are the unrestricted and restricted MLE of $z_i = (y_i - \mu)/\sigma$, respectively; and the p -value is computed as $1 - \chi_1^2(\text{LRT})$.

References

1. Stehlík, M.; Střelec, L.; Thulin, M. On robust testing for normality in chemometrics. *Chemom. Intell. Lab. Syst.* **2014**, *130*, 98–108.
2. Vidal, I.; Iglesias, P.; Branco, M.D.; Arellano-Valle, R.B. Bayesian Sensitivity Analysis and Model Comparison for Skew Elliptical Models. *J. Stat. Plan. Inference* **2006**, *136*, 3435–3457.
3. Stehlík, M. Distributions of exact tests in the exponential family. *Metrika* **2003**, *57*, 145–164.
4. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
5. Stehlík, M. Decompositions of information divergences: Recent development, open problems and applications. In Proceedings of the 9th International Conference on Mathematical Problems in Engineering, Aerospace and Sciences: ICNPAA 2012, Vienna, Austria, 10–14 July 2012; Volume 1493, pp. 972–976.
6. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback–Leibler divergence measure for Multivariate Skew-Normal Distributions. *Entropy* **2012**, *14*, 1606–1626.
7. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A* **1946**, *186*, 453–461.
8. Gómez-Villegas, M.A.; Main, P.; Navarro, H.; Susi, R. Assessing the effect of kurtosis deviations from Gaussianity on conditional distributions. *Appl. Math. Comput.* **2013**, *219*, 10499–10505.

9. Main, P.; Arevalillo, J.M.; Navarro, H. Local Effect of Asymmetry Deviations from Gaussianity Using Information-Based Measures. In Proceedings of the 2nd International Electronic Conference on Entropy and Its Applications, Santa Barbara, CA, USA, 15–30 November 2015.
10. Dette, H.; Ley, C.; Rubio, F.J. Natural (non-) informative priors for skew-symmetric distributions. *Scand. J. Stat.* **2017**, in press.
11. Gao, J.H.; Zhang, B. Estimation of seismic wavelets based on the multivariate scale mixture of Gaussians model. *Entropy* **2009**, *12*, 14–33.
12. Wang, Z.; Zhang, B.; Gao, J. The residual phase estimation of a seismic wavelet using a Rényi divergence-based criterion. *J. Appl. Geophys.* **2014**, *106*, 96–105.
13. Pires, C.A.; Ribeiro, A.F. Separation of the atmospheric variability into non-Gaussian multidimensional sources by projection pursuit techniques. *Clim. Dyn.* **2017**, *48*, 821–850.
14. Pires, C.A.; Hannachi, A. Independent Subspace Analysis of the Sea Surface Temperature Variability: Non-Gaussian Sources and Sensitivity to Sampling and Dimensionality. *Complexity* **2017**, doi:10.1155/2017/3076810.
15. Arrué, J.; Arellano-Valle, R.B.; Gómez, H.W. Bias reduction of maximum likelihood estimates for a modified skew-normal distribution. *J. Stat. Comput. Simul.* **2016**, *86*, 2967–2984.
16. Arellano-Valle, R.B.; Contreras-Reyes, J.E.; Genton, M.G. Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scand. J. Stat.* **2013**, *40*, 42–62.
17. Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
18. Contreras-Reyes, J.E. Analyzing fish condition factor index through skew-gaussian information theory quantifiers. *Fluct. Noise Lett.* **2016**, *15*, 1650013.
19. Withers, C.S.; Nadarajah, S. Negentropy as a function of cumulants. *Inf. Sci.* **2014**, *271*, 31–44.
20. Azzalini, A. A Class of Distributions which includes the Normal Ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
21. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distributions. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602.
22. Gupta, A.K.; Chang, F.C.; Huang, W.J. Some skew-symmetric models. *Random Oper. Stoch. Equ.* **2002**, *10*, 133–140.
23. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; Cambridge University Press: Cambridge, UK, 2013; Volume 3.
24. Nadarajah, S.; Kotz, S. Skewed distributions generated by the normal kernel. *Stat. Probab. Lett.* **2003**, *65*, 269–277.
25. Arrué, J.; Gómez, H.W.; Varela, H.; Bolfarine, H. On the skew-normal-Cauchy distribution. *Commun. Stat. A Theory* **2010**, *40*, 15–27.
26. Henze, N. A probabilistic representation of the ‘skew-normal’ distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.
27. Salicrú, M.; Menéndez, M.L.; Pardo, L.; Morales, D. On the applications of divergence type measures in testing statistical hypothesis. *J. Multivar. Anal.* **1994**, *51*, 372–391.
28. Arellano-Valle, R.B.; Azzalini, A. The centred parametrization for the multivariate skew-normal distribution. *J. Multivar. Anal.* **2009**, *100*, 816.
29. Chiogna, M. Some results on the scalar skew-normal distribution. *J. Ital. Stat. Soc.* **1998**, *1*, 1–13.
30. Kupperman, M. Further Applications of Information Theory to Multivariate Analysis and Statistical Inference. Ph.D. Thesis, George Washington University, Washington, DC, USA, January 1957; p. 270.
31. Piessens, R.; deDoncker-Kapenga, E.; Uberhuber, C.; Kahaner, D. *Quadpack: A Subroutine Package for Automatic Integration*; Springer: Berlin, Germany, 1983.
32. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015. Available online: <http://www.R-project.org> (accessed on 1 October 2016).
33. Contreras-Reyes, J.E. Rényi entropy and complexity measure for skew-gaussian distributions and related families. *Physica A* **2015**, *433*, 84–91.
34. Le Cren, E.D. The length–weight relationship and seasonal cycle in gonad weight and condition in the perch (*Perca fluviatilis*). *J. Anim. Ecol.* **1951**, *20*, 201–219.
35. Contreras-Reyes, J.E.; Cortés, D.D. Bounds on Rényi and Shannon Entropies for Finite Mixtures of Multivariate Skew-Normal Distributions: Application to Swordfish (*Xiphias gladius* Linnaeus). *Entropy* **2016**, *18*, 382.

36. Azzalini, A. *R package sn: The Skew-Normal and Skew-t Distributions (version 0.4-6)*; Università di Padova: Padua, Italy, 2010.
37. Arnold, B.C.; Beaver, R.J. Hidden truncation models. *Sankhya A* **2000**, *62*, 23–35.
38. Stehlík, M.; Somorčík, J.; Střelec, L.; Antoch, J. Approximation of information divergences for statistical learning with applications. *Math. Slovaca* **2017**, in press.
39. Martínez, E.H.; Varela, H.; Gómez, H.W.; Bolfarine, H. A note on the likelihood and moments of the skew-normal distribution. *Stat. Oper. Res. Trans.* **2008**, *32*, 57–66.
40. Chernoff, H. On the distribution of the likelihood ratio. *Ann. Math. Stat.* **1954**, *25*, 573–578.
41. Azzalini, A.; Arellano-Valle, R.B. Maximum penalized likelihood estimation for skew-normal and skew-*t* distributions. *J. Stat. Plan. Inference* **2013**, *143*, 419–433.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).