

Article

Bounds on Rényi and Shannon Entropies for Finite Mixtures of Multivariate Skew-Normal Distributions: Application to Swordfish (*Xiphias gladius Linnaeus*)

Javier E. Contreras-Reyes^{1,2,*} and Daniel Devia Cortés¹

¹ Instituto de Fomento Pesquero (IFOP), Blanco 839, Valparaíso 2361827, Chile; ddeviac@gmail.com

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

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

Academic Editor: Antonio M. Scarfone

Received: 03 August 2016; Accepted: 21 October 2016; Published: 26 October 2016

Abstract: Mixture models are in high demand for machine-learning analysis due to their computational tractability, and because they serve as a good approximation for continuous densities. Predominantly, entropy applications have been developed in the context of a mixture of normal densities. In this paper, we consider a novel class of skew-normal mixture models, whose components capture skewness due to their flexibility. We find upper and lower bounds for Shannon and Rényi entropies for this model. Using such a pair of bounds, a confidence interval for the approximate entropy value can be calculated. Simulation studies are then applied to a swordfish (*Xiphias gladius Linnaeus*) length dataset.

Keywords: skew-normal; finite mixtures; Shannon entropy; Rényi entropy; swordfish

1. Introduction

Mixture models are in high demand for machine-learning analysis, due to their computational tractability and for offering a good approximation for continuous densities [1]. In addition, mixture models are an important statistical tool for many applications in clustering [2,3], discriminant analysis [4], image processing and satellite imaging [5,6]. Celeux and Soromenho [2] consider a Maximum Likelihood (ML)-based entropy criterion to estimate the number of clusters arising from a mixture model, and compare it with the classical Akaike (AIC) and Bayesian (BIC) information criteria. Carreira-Perpiñán [6] deal with the problem of finding all modes of multi-dimensional data, assuming a mixture of normal densities. Specifically, he uses the gradient as a mode locator and for controlling the significance modes thus obtained, by measuring the sparseness of the densities mixture via the entropy. However, so far, no analytical expressions, which consider bounds of Shannon entropy for the normal mixture entropy, exist. Similarly, in the case of Kullback–Leibler divergence, an analytic evaluation of the differential entropy is also impossible. Thus, approximate calculations become inevitable [7–9]. Jenssen et al. [3] use the Rényi entropy [10] as a similarity measure between clusters. They consider the Parzen window density estimation for differential Rényi entropy clustering to identify the *worst* cluster and subsequently reduce the overall number of clusters by one.

Predominantly, the entropy applications mentioned above have been developed in the normal context, but several results of both Shannon and Rényi entropies for various multivariate distributions (see, e.g., [11,12]) actually exist. Here, we consider the novel class of finite mixture of multivariate skew-normal mixture (FMSN) models [13]. This class provides some advantages over the normal mixtures. For instance, the normal components allow an arbitrarily close modeling of any distribution by increasing the number of components, and, in the context of supervised learning, groups of observations represented by asymmetrically distributed data can lead to wrong classification [14]. The components of skew-normal mixture models, however, capture skewness due to their flexibility.

Several examples, including references of this fact, can be found in [14–17]. Frühwirth-Schnatter and Pyne [14] studied high-dimensional flow cytometric data for Alzheimer’s disease treatment. A trivariate diffuse large B-cell lymphoma dataset is studied by [15] to cluster the cells into three groups. Lee and McLachlan [16] consider several applications, among them: (i) the clustering of a flow cytometric dataset derived from a hematopoietic stem cell transplant (HSCT) experiment, where each sample was stained with four fluorescent markers; (ii) the variables body mass index, lean body mass, and body fat percentage of a real dataset, concerning biomedical measurements for Australian athletes; (iii) a portfolio of three shares listed on the Australian Stock exchange in a value-at-risk (VaR) analysis; (iv) geometric measurements taken from X-ray images of wheat kernels in a discriminant analysis application; and (v) an image segmentation analysis. In addition, Lin et al. [17] studied the distribution of peripheral blood samples in transplanted organs.

In this paper, we calculate the bounds for Shannon and Rényi entropies for the skew-normal mixture model. The maximum entropy theorem and Jensen’s inequality are considered for the Shannon entropy case. Using such a pair of bounds, a confidence interval for the approximate entropy value can be calculated. Simulation studies are then applied to a swordfish length dataset.

The paper is organized as follows. Section 2 presents definitions of multivariate skew-normal (SN) and FMSN distributions, as well as previous results of Shannon and Rényi entropies for SN distributions. Section 3 presents the main results: the computation of upper and lower bounds of these information measures for FMSN distributions. Section 4 reports numerical results of some simulated examples and a real-world application of swordfish data. This paper ends with a discussion in Section 5. Some proofs are presented in Appendix A.

2. Preliminary Material

2.1. Skew-Normal Distribution

The multivariate skew-normal distribution has been introduced in [18]. This class of flexible distributions regulates the skewness, allowing for a continuous variation from normality to skew-normality. Below, we use a slight variant of the original definition. Consider a random vector $\mathbf{Z} \in \mathbb{R}^d$ with skew-normal distribution, location vector $\boldsymbol{\zeta} \in \mathbb{R}^d$, dispersion matrix $\boldsymbol{\Omega} \in \mathbb{R}^{d \times d}$ and shape/skewness parameter $\boldsymbol{\eta} \in \mathbb{R}^d$, denoted by $\mathbf{Z} \sim SN_d(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\boldsymbol{\zeta}, \boldsymbol{\Omega}, \boldsymbol{\eta})$, if its probability density function (pdf) is

$$f(\mathbf{z}; \boldsymbol{\theta}) = 2\phi_d(\mathbf{z}; \boldsymbol{\zeta}, \boldsymbol{\Omega})\Phi_1(\boldsymbol{\eta}^\top(\mathbf{z} - \boldsymbol{\zeta}); 0, 1), \quad (1)$$

where $\phi_d(\mathbf{z}; \boldsymbol{\zeta}, \boldsymbol{\Omega})$ is the pdf of the d -variate $N_d(\boldsymbol{\zeta}, \boldsymbol{\Omega})$ distribution, and $\Phi_1(\boldsymbol{\eta}^\top(\mathbf{z} - \boldsymbol{\zeta}); 0, 1)$ the univariate $N_1(0, 1)$ cumulative distribution function. where $|\boldsymbol{\Omega}|$ ($|\boldsymbol{\Omega}| > 0$) represents the determinant of $\boldsymbol{\Omega}$. The stochastic representation of \mathbf{Z} is given by

$$\mathbf{Z} \stackrel{d}{=} \boldsymbol{\zeta} + \delta|U_0| + \mathbf{U}, \quad (2)$$

where $U_0 \sim N(0, 1)$ and $\mathbf{U} \sim N_d(0, \boldsymbol{\Omega} - \delta\delta^\top)$, $\delta = \boldsymbol{\Omega}\boldsymbol{\eta}/\sqrt{1 + \boldsymbol{\eta}^\top\boldsymbol{\Omega}\boldsymbol{\eta}}$, $|\delta| < 1$, which are independent. $|U_0|$ represents the absolute value of U_0 , i.e., it is half-normal distributed. From Equation (2), Azzalini and Capitanio [19] derived the mean vector and covariance matrix of \mathbf{Z} :

$$E[\mathbf{Z}] = \boldsymbol{\zeta} + \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}, \quad (3)$$

$$Var[\mathbf{Z}] = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^\top. \quad (4)$$

2.2. Finite Mixtures of Skew-Normal Distributions

Let us consider the definition of [14] for finite mixtures of skew-normal distributions. The pdf of an m -component mixture model with parameter vector set $\tilde{\theta} = (\tilde{\xi}, \tilde{\Omega}, \tilde{\eta})$: $\tilde{\xi} = (\xi_1, \dots, \xi_m)$ a set of m location vector parameters, $\tilde{\Omega} = (\Omega_1, \dots, \Omega_m)$ a set of m dispersion matrices, $\tilde{\eta} = (\eta_1, \dots, \eta_m)$ a set of shape vector parameters; and with m mixing weights, $\pi = (\pi_1, \dots, \pi_m)$ is

$$p(\mathbf{y}; \tilde{\theta}, \pi) = \sum_{i=1}^m \pi_i f(\mathbf{y}; \theta_i), \tag{5}$$

where $\pi_i \geq 0$, $\sum_{i=1}^m \pi_i = 1$, and $f(\mathbf{y}; \theta_i)$ are defined as in Equation (1) for a known $\theta_i = (\xi_i, \Omega_i, \eta_i)$, $i = 1, \dots, m$. Additional details about the log-likelihood function of an FMSN model are described in [13]. Let $\mathbf{S} = (S_1, \dots, S_n)$ be a set of n latent allocations for the distribution of observations \mathbf{y} , $p(\mathbf{y}; \tilde{\theta}, \pi) = \prod_{j=1}^n p(\mathbf{y}; \tilde{\theta}, S_j)$, where $\Pr(S_j = i | \pi) = \pi_i$. Then, an equivalent stochastic representation to each j -th component density as in (2) is obtained:

$$\mathbf{Y}_j | (S_j = i) \stackrel{d}{=} \xi_i + \delta_i |U_{0j}| + \sqrt{\Omega_i - \delta_i \delta_i^\top} \mathbf{U}_j, \quad j = 1, \dots, n, \tag{6}$$

where U_{0j} and \mathbf{U}_j are mutually independent and standardized one- and d -dimensional normal distributed, respectively; and $\delta_i = \Omega_i \eta_i / \sqrt{1 + \eta_i^\top \Omega_i \eta_i}$, $i = 1, \dots, m$. Considering the stochastic representation (6), and the first and second moments of the i -th component of \mathbf{Y} , Equations (3) and (4), respectively; we obtain the first two moments of \mathbf{Y} :

$$E[\mathbf{Y}] = \sum_{i=1}^m \pi_i \left(\xi_i + \sqrt{\frac{2}{\pi}} \delta_i \right), \tag{7}$$

$$Var[\mathbf{Y}] = \sum_{i=1}^m \pi_i \left[\Omega_i - \frac{2}{\pi} \delta_i \delta_i^\top + \mu_i \mu_i^\top \right], \tag{8}$$

with $\mu_i = \xi_i + \sqrt{\frac{2}{\pi}} \delta_i - E[\mathbf{Y}]$, $i = 1, \dots, m$ (see, e.g., [6]).

2.3. Entropies

Let \mathbf{X} be a random vector defined in \mathbb{R}^d and for all values of parameter $\theta \in \Theta$, where Θ is an open subset of the real line and $f(\mathbf{x}; \theta)$ is the pdf of \mathbf{x} . Let us consider the α th-order Rényi entropy [10] on variable \mathbf{x} :

$$\mathcal{R}_\alpha[\mathbf{X}; \theta] = \frac{1}{1 - \alpha} \ln \int_{\mathbb{R}^d} f(\mathbf{x}; \theta)^\alpha d\mathbf{x}, \quad 0 < \alpha < \infty, \alpha \neq 1, \tag{9}$$

and the Shannon entropy is obtained by the limit

$$\begin{aligned} \mathcal{H}[\mathbf{X}; \theta] &= \lim_{\alpha \rightarrow 1} \mathcal{R}_\alpha[\mathbf{X}; \theta] \\ &= -E[\ln f(\mathbf{x}; \theta)] \\ &= - \int_{\mathbb{R}^d} f(\mathbf{x}; \theta) \ln f(\mathbf{x}; \theta) d\mathbf{x}, \end{aligned} \tag{10}$$

by applying l'Hôpital's rule to $\mathcal{R}_\alpha[\mathbf{X}; \theta]$ with respect to α . Hereafter, we will refer to Equation (10) as the expected information of $f(\mathbf{x}; \theta)$ (for additional properties of the Shannon entropy, see [20]) and, $E[g(\mathbf{X})]$ denotes the expected information in \mathbf{X} of the random function $g(\mathbf{x}) = \ln f(\mathbf{x}; \theta)$. An important property of Rényi entropy is $\mathcal{R}_{\alpha_1}[\mathbf{X}; \theta] \leq \mathcal{R}_{\alpha_2}[\mathbf{X}; \theta]$, given that $\alpha_1 > \alpha_2$ (see, e.g., [11]). In addition, the Rényi entropy represents a generalization of the Shannon entropy and could be used to derive a continuous family of information measures.

According to [21], the negentropy of a living system is the entropy it exports to keep its own entropy low, and thus it lies at the intersection of entropy and life. In our case, the negentropy component of the Rényi and Shannon entropies measures the dispersion of the distribution of \mathbf{Z} from normal distribution [11]. The following Proposition of [11] presents these differences in terms of the deviation matrix and shape/skewness parameter.

Proposition 1. Let $\mathbf{Z} \sim SN_d(\boldsymbol{\theta})$, $\mathbf{Z}_N \sim N_d(\boldsymbol{\xi}, \boldsymbol{\Omega})$, and $\|\bar{\boldsymbol{\eta}}\| = \boldsymbol{\eta}^\top \boldsymbol{\Omega} \boldsymbol{\eta}$. Then,

(i) the Shannon entropy of \mathbf{Z} is

$$\mathcal{H}[\mathbf{Z}; \boldsymbol{\theta}] = \mathcal{H}[\mathbf{Z}_N; \boldsymbol{\theta}_N] - \mathcal{N}[\mathbf{Z}; \boldsymbol{\theta}],$$

where

$$\begin{aligned} \mathcal{H}[\mathbf{Z}_N; \boldsymbol{\theta}_N] &= \frac{1}{2} \ln[(2\pi e)^d |\boldsymbol{\Omega}|] \quad (\text{normal Shannon entropy}), \\ \mathcal{N}[\mathbf{Z}; \boldsymbol{\theta}] &= E[\ln\{2\Phi_1(\|\bar{\boldsymbol{\eta}}\|W)\}] \quad (\text{Shannon negentropy}), \end{aligned}$$

with $W \sim SN_1(\|\bar{\boldsymbol{\eta}}\|)$.

(ii) the α th-Rényi entropy of \mathbf{Z} , $\alpha = 2, 3, \dots$, is

$$\mathcal{R}_\alpha[\mathbf{Z}; \boldsymbol{\theta}] = \mathcal{R}_\alpha[\mathbf{Z}_N; \boldsymbol{\theta}_N] - \mathcal{N}_\alpha[\mathbf{Z}; \boldsymbol{\theta}],$$

where

$$\begin{aligned} \mathcal{R}_\alpha[\mathbf{Z}_N; \boldsymbol{\theta}_N] &= \frac{1}{2} \ln[(2\pi)^\alpha |\boldsymbol{\Omega}|] + \frac{d \ln \alpha}{2(\alpha - 1)} \quad (\text{normal Rényi entropy}), \\ \mathcal{N}_\alpha[\mathbf{Z}; \boldsymbol{\theta}] &= \frac{1}{\alpha - 1} \ln \left[2^\alpha \frac{\Phi_{\alpha+1}(0; 0, \bar{\boldsymbol{\Omega}})}{\Phi_1(0; 0, \omega)} \right] \quad (\text{Rényi negentropy}), \end{aligned}$$

with $\boldsymbol{\theta}_N = (\boldsymbol{\xi}, \boldsymbol{\Omega})$, $\bar{\boldsymbol{\Omega}} = \mathbf{I}_{\alpha+1} + \|\bar{\boldsymbol{\eta}}\|^2 \bar{\mathbf{D}}^\top \bar{\mathbf{D}}$, $\bar{\mathbf{D}} = (\mathbf{1}_\alpha, \|\bar{\boldsymbol{\eta}}\|)^\top$, $\mathbf{1}_\alpha$ is the α -dimensional vector of ones, and $\omega = 1 + \|\bar{\boldsymbol{\eta}}\|^4$.

From Proposition 1, we observe that the Shannon negentropies do not depend on the dispersion matrix $\boldsymbol{\Omega}$, but only on the shape parameter vector $\boldsymbol{\eta}$. However, Rényi entropy depends on $\boldsymbol{\Omega}$ and $\boldsymbol{\eta}$ parameters. Contreras-Reyes [11] show in Proposition 1 (ii) that the skew-normal Rényi entropy is the normal Rényi entropy [12], as $\boldsymbol{\eta} \rightarrow 0$.

3. Results

In this section, we present practical results of upper and lower bounds of Shannon and Rényi entropies for FMSN distributions in Sections 3.1 and 3.2, respectively, which ought to be considered in numerical simulations and real-world application (Section 4).

3.1. Shannon Entropy Bounds

As in the normal case, the Shannon entropy of mixture of skew-normal distributions does not have a closed form. However, the following proposition presents some lower and upper bounds as an approximation of the entropy of finite mixture of skew-normal densities.

Proposition 2. Consider the FMSN density of $(\mathbf{Y}; \tilde{\boldsymbol{\theta}}, \boldsymbol{\pi})$ defined in Equation (5). Then, the following inequalities are accomplished:

- (i) $A_{lower} \leq \mathcal{H}[\mathbf{Y}; \tilde{\boldsymbol{\theta}}] \leq A_{upper}$,
- (ii) $B_{lower} \leq \mathcal{H}[\mathbf{Y}; \tilde{\boldsymbol{\theta}}] \leq B_{upper}$,

where

$$\begin{aligned}
 A_{upper} &= \frac{1}{2} \ln \{ (2\pi e)^d |\Sigma| \}, \\
 A_{lower} &= A_{upper} - \sum_{i=1}^m \pi_i \mathcal{N}[\mathbf{Y}; \boldsymbol{\theta}_i], \\
 B_{upper} &= A_{lower} - \sum_{i=1}^m \pi_i \ln \pi_i, \\
 B_{lower} &= - \sum_{i=1}^m \pi_i \ln \left(\sum_{s=1}^m \pi_s \int_{-\infty}^{\infty} f(t; \boldsymbol{\eta}_i) f(t; \boldsymbol{\eta}_s) dt \right),
 \end{aligned}$$

with $\mathcal{N}[\mathbf{Y}; \boldsymbol{\theta}_i] = E [\ln \{ 2\Phi_1(\|\tilde{\boldsymbol{\eta}}_i\| W_i) \}] = \int_{-\infty}^{\infty} f(w_i; \|\tilde{\boldsymbol{\eta}}_i\|) \ln \{ 2\Phi_1(\|\tilde{\boldsymbol{\eta}}_i\| w_i) \} dw_i$, $W_i \sim SN_1(\|\tilde{\boldsymbol{\eta}}_i\|)$, $\|\tilde{\boldsymbol{\eta}}_i\| = \boldsymbol{\eta}_i^\top \boldsymbol{\Omega}_i \boldsymbol{\eta}_i$, and $\Sigma = \text{Var}[\mathbf{Y}]$ is defined by Equation (8).

For the case $m = 1$, Contreras-Reyes and Arellano-Valle [22] consider the upper bound of the property (i) of Proposition 2 to approximate the Shannon entropy of an SN distribution using the property (ii) of Proposition 2. In this Proposition 2(ii), the left side includes an integral related to a product of two skew-normal densities. When $i = s$, these integrals correspond to an L^2 -norm and are represented by the quadratic Rényi entropy [11]. For the case $i \neq s$, the integral does not have explicit form and requires numerical methods to be computed. Moreover, the right side corresponds to the sum of the entropy of a multinomial density with parameters π_1, \dots, π_m and a second term based on the weights and shape parameters of the skew-normal density. Other refinements can be found in [9]. These are suitable for cases of several components (for example, $m \gg 5$), i.e., a skew-normal density consisting of several and well separated clusters.

A lower bound can be found for $\mathcal{H}[\mathbf{Y}; \tilde{\boldsymbol{\theta}}]$ using the L^2 -norm of an FMSN density and Jensen’s inequality [20]:

$$\mathcal{H}[\mathbf{Y}; \tilde{\boldsymbol{\theta}}] \geq -2 \ln \|p(\mathbf{y}; \tilde{\boldsymbol{\theta}}, \boldsymbol{\pi})\|_2 = - \ln \int_{\mathbb{R}^d} p(\mathbf{y}; \tilde{\boldsymbol{\theta}}, \boldsymbol{\pi})^2 d\mathbf{y}. \tag{11}$$

Considering Equation (9), Equation (11) and Proposition 2(i)–(ii), we obtain the additional inequality $\mathcal{R}_2[\mathbf{Y}; \tilde{\boldsymbol{\theta}}] \leq \mathcal{H}[\mathbf{Y}; \tilde{\boldsymbol{\theta}}]$ and

$$B_{lower} < A_{lower} < B_{upper} < A_{upper}. \tag{12}$$

The next section provides upper and lower bounds for Rényi entropy of FMSN random vectors.

3.2. Rényi Entropy Bounds

For the sake of simplicity, we define the following function in terms of Rényi entropy and α as

$$\mathcal{P}_\alpha[\mathbf{Y}; \tilde{\boldsymbol{\theta}}] = e^{(1-\alpha)\mathcal{R}_\alpha[\mathbf{Y}; \tilde{\boldsymbol{\theta}}]}, \quad 0 < \alpha < \infty, \alpha \neq 1,$$

for the calculus of the bounds of $\int_{\mathbb{R}^d} p(\mathbf{y}; \tilde{\boldsymbol{\theta}}, \boldsymbol{\pi})^\alpha d\mathbf{y}$. By applying the function $\ln(\cdot)/(1 - \alpha)$ to these integrals, we have the Rényi entropy of FMSN density.

As in the Shannon entropy case, the Rényi entropy can be upper bounded in terms of the dispersion matrix of the finite mixture random variable. Sánchez-Moreno et al. [23] derived a multidimensional upper bound using a variational approach,

$$\mathcal{R}_\alpha[\mathbf{Y}; \tilde{\boldsymbol{\theta}}] \leq \frac{d}{2} \ln \left(\frac{|\Sigma|}{d} \right) + \mathcal{F}_d(\alpha), \tag{13}$$

with $\Sigma = Var[\mathbf{Y}]$ defined in Equation (8),

$$\mathcal{F}_d(\alpha) = \begin{cases} \frac{d}{2} \ln \left(\frac{\pi b}{\alpha-1} \right) + \frac{1}{\alpha-1} \ln \left(\frac{b}{2\alpha} \right) + \ln \Gamma \left(\frac{\alpha}{\alpha-1} \right) - \ln \Gamma \left(\frac{b}{2(\alpha-1)} \right), & \text{if } \alpha > 1, \\ \frac{d}{2} \ln \left(\frac{\pi b}{1-\alpha} \right) + \frac{\alpha}{\alpha-1} \ln \left(\frac{b}{2\alpha} \right) - \ln \Gamma \left(\frac{\alpha}{1-\alpha} \right) + \ln \Gamma \left(\frac{b}{2(1-\alpha)} \right), & \text{if } \alpha \in \left(\frac{d}{d+2}, 1 \right), \\ \mathcal{H}[\mathbf{W}_0; \boldsymbol{\theta}_0], & \text{if } \alpha = 1, \end{cases}$$

$b = (2 + d)\alpha - d$, $\boldsymbol{\theta}_0 = (0, \mathbf{I}_d)$, and $\mathbf{W}_0 \sim N_d(0, \mathbf{I}_d)$ (\mathbf{I}_d denotes the d -dimensional identity matrix). $\mathcal{H}[\mathbf{W}_0; \boldsymbol{\theta}_0]$ is obtained using property (i) of Proposition 1.

The right side of the inequality (13) is equivalent to the maximum Shannon entropy of Proposition 2. The first term depends on the dispersion matrix and the shape parameters, and the second only depends on the α th order and dimension d .

The next Lemma presents a useful result to compute the lower bound for Rényi entropy of an FMSN random vector \mathbf{Y} in terms of each component.

Lemma 1. Consider the FMSN density of $(\mathbf{Y}; \boldsymbol{\theta}, \boldsymbol{\pi})$ defined in Equation (5). Then,

$$\int_{\mathbb{R}^d} p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\pi})^\alpha d\mathbf{y} \geq \mathcal{P}_\alpha[\mathbf{Y}; \boldsymbol{\theta}_m] + \sum_{i=1}^{m-1} \left\{ \left(\sum_{k=1}^i \pi_k \right)^\alpha \left(\mathcal{P}_\alpha[\mathbf{Y}; \boldsymbol{\theta}_i] - \mathcal{P}_\alpha[\mathbf{Y}; \boldsymbol{\theta}_{i+1}] \right) \right\},$$

with $0 < \alpha < \infty$, $\alpha \neq 1$, and $m > 1$.

4. Numerical Results

4.1. Simulations

To study the behavior of the Shannon entropy bounds of Proposition 2 and the Rényi entropy bounds of Equation (13) and Lemma 1, some examples are simulated for the cases $d = 1, 2$ and 3 :

Example 1: $d = 1, m = 2, \boldsymbol{\pi} = (0.3, 0.7), \tilde{\boldsymbol{\xi}} = (0.5, 5), \tilde{\boldsymbol{\Omega}} = (3.5, 6)$, and $\tilde{\boldsymbol{\eta}} = (0.5, 3.5)$;

Example 2: [24] $d = 1, m = 3, \boldsymbol{\pi} = (0.5, 0.2, 0.3), \tilde{\boldsymbol{\xi}} = (2, 20, 35), \tilde{\boldsymbol{\Omega}} = (9, 16, 9)$, and $\tilde{\boldsymbol{\eta}} = (5, 3, 6)$;

Example 3: [24] $d = 2, m = 2, \boldsymbol{\pi} = (0.65, 0.35)$,

$$\tilde{\boldsymbol{\xi}} = \left(\left(\begin{matrix} 5 \\ 7 \end{matrix} \right), \left(\begin{matrix} 2 \\ 5 \end{matrix} \right) \right), \tilde{\boldsymbol{\Omega}} = \left(\left(\begin{matrix} 0.18 & 0.6 \\ 0.6 & 4 \end{matrix} \right), \left(\begin{matrix} 0.15 & 1.15 \\ 1.15 & 4 \end{matrix} \right) \right), \text{ and}$$

$$\tilde{\boldsymbol{\eta}} = \left(\left(\begin{matrix} 0.69 \\ 0.64 \end{matrix} \right), \left(\begin{matrix} 4.3 \\ 2.7 \end{matrix} \right) \right);$$

Example 4: [24] $d = 2, m = 3, \boldsymbol{\pi} = (0.25, 0.5, 0.25)$,

$$\tilde{\boldsymbol{\xi}} = \left(\left(\begin{matrix} 0 \\ 0 \end{matrix} \right), \left(\begin{matrix} 5 \\ 5 \end{matrix} \right), \left(\begin{matrix} 2 \\ 8 \end{matrix} \right) \right), \tilde{\boldsymbol{\Omega}} = \left(\left(\begin{matrix} 3 & 1 \\ 1 & 3 \end{matrix} \right), \left(\begin{matrix} 2 & 1 \\ 1 & 2 \end{matrix} \right), \left(\begin{matrix} 0.15 & 1.15 \\ 1.15 & 40 \end{matrix} \right) \right), \text{ and}$$

$$\tilde{\boldsymbol{\eta}} = \left(\left(\begin{matrix} 4 \\ 4 \end{matrix} \right), \left(\begin{matrix} 2 \\ 2 \end{matrix} \right), \left(\begin{matrix} 4.3 \\ 2.7 \end{matrix} \right) \right);$$

Example 5: [2] $d = 3, m = 3, \pi = (0.22, 0.36, 0.42),$

$$\tilde{\xi} = \left(\left(\begin{array}{c} 10 \\ 12 \\ 10 \end{array} \right), \left(\begin{array}{c} 8.5 \\ 10.5 \\ 8.5 \end{array} \right), \left(\begin{array}{c} 12 \\ 14 \\ 12 \end{array} \right) \right), \tilde{\Omega} = \left(\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right), \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right), \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) \right),$$

$$\text{and } \tilde{\eta} = \left(\left(\begin{array}{c} 4 \\ 0 \\ 1 \end{array} \right), \left(\begin{array}{c} 2 \\ 1 \\ 3 \end{array} \right), \left(\begin{array}{c} 4 \\ 2 \\ 2 \end{array} \right) \right);$$

Example 6: [25] $d = 3, m = 4, \pi = (0.125, 0.19, 0.135, 0.55),$

$$\tilde{\xi} = \left(\left(\begin{array}{c} 420 \\ 360 \\ 425 \end{array} \right), \left(\begin{array}{c} 160 \\ 570 \\ 200 \end{array} \right), \left(\begin{array}{c} 320 \\ 540 \\ 260 \end{array} \right), \left(\begin{array}{c} 530 \\ 80 \\ 450 \end{array} \right) \right),$$

$$\tilde{\Omega} = \left(\left(\begin{array}{ccc} 9160 & 5580 & 7000 \\ 5580 & 12105 & 7160 \\ 7000 & 7160 & 7250 \end{array} \right), \left(\begin{array}{ccc} 3870 & 1810 & 1770 \\ 1810 & 2900 & 1270 \\ 1770 & 1270 & 1320 \end{array} \right), \left(\begin{array}{ccc} 1695 & 1190 & 2280 \\ 1190 & 2780 & 2010 \\ 2280 & 2010 & 3720 \end{array} \right), \right.$$

$$\left. \left(\begin{array}{ccc} 1590 & 590 & 15 \\ 590 & 2425 & 415 \\ 15 & 415 & 1870 \end{array} \right) \right), \text{ and } \tilde{\eta} = \left(\left(\begin{array}{c} 4.8 \\ 17 \\ 50 \end{array} \right), \left(\begin{array}{c} 4 \\ 80 \\ 60 \end{array} \right), \left(\begin{array}{c} 40 \\ 8 \\ 10 \end{array} \right), \left(\begin{array}{c} 60 \\ 90 \\ 6 \end{array} \right) \right).$$

Figure 1 presents the examples mentioned in the settings above. Examples 1 and 2 are represented in histogram plots and examples 3 and 4 in contour plots, according to [24]. Examples 5 and 6 are represented in 3D plots, according to [25]. For all simulations, a sample of $n = 500$ generations is considered, and then fixed using the function `smsn.mix` from `mixsmsn` package, developed by [24] and implemented in an R environment [26]. Prates et al. [24] implemented routines for ML estimation via the Expectation Maximization (EM)-type algorithm in FMSN models (among several others).

For each example, Table 1 summarizes the four Shannon, as well as the Rényi entropy bounds for $\alpha = 2, \dots, 5$ and $m = 2, \dots, 6$. Shannon and Rényi entropies are compared with AIC and BIC criteria (see e.g. [27]), misclassification (MC) rates and consistency scores: normal skill score (NSS), Heidke skill score (HSS), and Hanssen–Kuipers (HK) (see e.g., [28]). All these indicators show an optimal performance of model fit if they are near 1; except MC, which ideally should be close to 0 (i.e., $100(1 - \text{MC}) \approx 100\%$). For all examples, it is worth pointing out that these criteria are optimal for minimum AIC and BIC values (marked in gray). For examples 1–3, the misclassification rates are close to 0, and, for examples, 4–6 less than 0.46. This is because of the complexity of systems with high dimensions and parsimonious models fit (excess of parameters).

The information measures illustrated a similar effect. It can be seen that inequalities given in (12) are accomplished in the Shannon entropy case and the information increase for more parsimonious systems, where these 3D systems are characterized by a bigger set of components and dispersion matrices with large elements. With respect to Rényi entropies, the lower and upper bounds rather slowly increase with more components in examples 1–3, but rise faster with more components in examples 4–6. However, in examples 4–6, the lower and upper bounds are maximum for large α . Therefore, the Rényi information criterion is suitable for model fits with accurate classification of observations, i.e., incorrect performance of Rényi entropy is related to inadequate selection of components in complex systems. Additionally, the Rényi entropy of FMSN is localized between the upper and lower bounds, and an approximation should be given by the mean of these bounds.

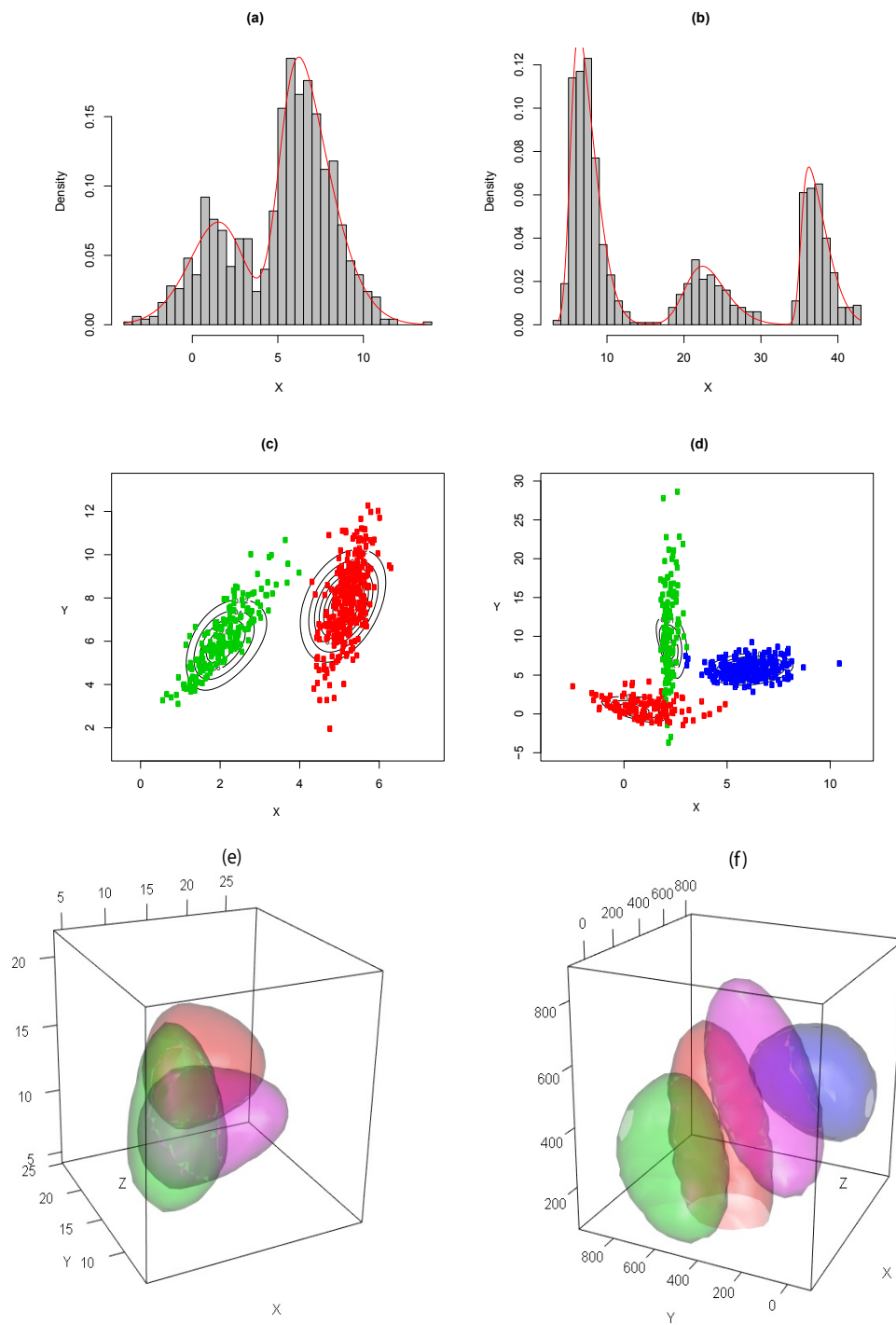


Figure 1. Finite mixtures of skew-normal (FMSN) density simulations using samples of 500 generations and the settings given by the examples (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, and (f) 6 of Section 4.1.

Table 1. Simulated Shannon and Rényi entropy bounds. Rényi entropy bounds are computed for $\alpha = 2, \dots, 5$. For each model and number of clusters m , the Akaike (AIC) and Bayesian (BIC) information criteria, misclassification (MC), normal skill (NSS), Heidke skill (HSS), and Hanssen–Kuipers (HK) scores appear. The shaded regions highlight the smallest AIC and BIC values.

Example	m	MC	NSS	HSS	HK	AIC	BIC	\mathcal{H}				\mathcal{R}_α Lower				\mathcal{R}_α Upper			
								A_{lower}	A_{upper}	B_{lower}	B_{upper}	2	3	4	5	2	3	4	5
1	2	0.02	0.98	0.95	0.39	4831.88	4866.23	0.99	2.20	0.888	1.58	0.72	0.96	1.01	1.03	3.54	3.06	2.89	2.80
	3	0.03	0.97	0.93	0.42	4840.36	4894.34	1.89	2.79	1.58	2.92	0.58	0.84	0.91	0.94	3.52	3.04	2.87	2.78
	4	0.02	0.98	0.96	0.41	4846.62	4920.23	1.85	2.80	1.44	3.03	0.58	0.85	0.93	0.97	3.52	3.04	2.87	2.78
	5	0.61	0.39	0.01	0.01	4851.35	4944.59	2.15	2.91	1.93	3.59	0.50	0.63	0.65	0.65	3.50	3.01	2.85	2.76
	6	0.77	0.23	0.00	0.00	4858.87	4971.75	2.19	2.92	2.04	3.74	0.59	0.70	0.70	0.70	3.50	3.01	2.85	2.76
2	2	0.01	0.99	0.98	0.40	6581.52	6615.87	2.57	3.92	0.66	3.26	1.18	1.29	1.32	1.33	4.93	4.45	4.28	4.19
	3	0.00	1.00	1.00	0.62	6065.25	6119.23	2.96	4.25	0.75	3.99	1.06	1.32	1.38	1.40	4.97	4.49	4.32	4.23
	4	0.51	0.49	0.13	0.08	6071.55	6145.16	2.98	4.25	0.78	4.32	0.79	1.06	1.13	1.16	4.95	4.47	4.30	4.21
	5	0.60	0.40	0.00	0.00	6080.71	6173.95	3.28	4.26	1.55	4.81	0.83	1.06	1.11	1.12	4.95	4.47	4.30	4.21
	6	0.59	0.41	0.00	0.00	6090.82	6203.70	3.62	4.39	2.32	5.33	0.73	0.90	0.94	0.94	4.95	4.47	4.30	4.21
3	2	0.00	1.00	1.00	0.46	5766.82	5840.43	3.44	3.94	2.84	4.09	1.68	1.61	1.55	1.50	5.27	4.67	4.46	4.35
	3	1.00	0.00	-0.96	-0.49	5778.71	5891.59	3.66	4.45	1.84	4.72	0.79	0.95	0.98	0.99	6.28	5.69	5.48	5.36
	4	1.00	0.00	-0.98	-0.50	5785.43	5937.57	3.78	4.62	1.73	5.13	0.62	0.80	0.84	0.86	6.62	6.03	5.82	5.70
	5	1.00	0.00	-0.24	-0.19	5798.10	5989.51	3.89	4.66	1.92	5.31	0.72	0.84	0.85	0.84	6.71	6.11	5.90	5.79
	6	0.26	0.74	0.00	0.00	5758.53	5989.20	4.01	4.79	1.88	5.67	0.64	0.79	0.81	0.81	6.97	6.38	6.17	6.05
4	2	0.76	0.24	-0.11	-0.07	8758.63	8832.25	3.35	4.61	0.82	3.92	1.64	1.94	2.02	2.05	6.60	6.00	5.79	5.68
	3	0.30	0.70	0.14	0.05	8282.84	8395.72	4.11	4.84	2.06	5.15	1.36	1.53	1.56	1.56	7.24	6.65	6.44	6.32
	4	0.36	0.64	0.46	0.31	8295.26	8447.40	4.51	5.24	2.06	5.71	1.73	1.83	1.84	1.83	7.86	7.27	7.06	6.94
	5	0.36	0.64	0.47	0.32	8300.94	8492.34	4.41	5.24	1.78	5.79	1.27	1.39	1.40	1.40	7.86	7.27	7.06	6.95
	6	0.57	0.43	0.21	0.15	8246.46	8477.12	4.64	5.43	1.88	6.17	1.29	1.42	1.43	1.42	8.23	7.64	7.43	7.32
5	2	0.64	0.36	-0.03	-0.02	9650.23	9772.92	5.57	6.52	1.61	6.25	2.43	2.54	2.52	2.48	10.26	9.56	9.30	9.17
	3	0.45	0.55	0.02	0.01	9510.53	9697.02	5.66	6.73	1.28	6.65	1.84	2.07	2.13	2.15	10.53	9.82	9.57	9.43
	4	0.53	0.47	0.19	0.13	9513.73	9764.02	5.83	6.78	1.43	7.19	1.68	1.90	1.96	1.98	11.06	10.36	10.10	9.97
	5	0.92	0.08	-0.23	-0.17	9539.02	9853.11	5.99	6.90	1.52	7.49	1.60	1.78	1.80	1.79	11.41	10.70	10.45	10.31
	6	0.68	0.32	0.05	0.04	9550.44	9928.33	5.93	6.85	1.51	7.63	1.54	1.73	1.76	1.77	11.26	10.56	10.30	10.17
6	2	0.62	0.38	-0.04	-0.02	33479.01	33601.70	15.56	16.93	0.64	16.25	7.53	7.70	7.75	7.77	41.50	40.79	40.54	40.40
	3	1.00	0.00	-0.47	-0.32	33019.83	33206.33	16.88	18.18	0.75	17.84	7.45	7.71	7.79	7.82	45.26	44.55	44.30	44.16
	4	0.45	0.55	0.29	0.18	32417.80	32668.10	17.31	18.62	0.73	18.65	7.45	7.77	7.87	7.92	45.82	45.11	44.86	44.72
	5	0.93	0.07	-0.15	-0.12	32346.29	32660.39	17.31	18.63	0.71	18.78	6.95	7.13	7.15	7.15	46.61	45.91	45.65	45.52
	6	0.56	0.44	0.20	0.14	32458.05	32835.95	17.54	18.85	0.73	19.20	7.05	7.27	7.32	7.33	47.26	46.56	46.30	46.17

4.2. Application

Estimation of age from growth of swordfish (*Xiphias gladius Linnaeus*) is an important factor in assessing stock trends [29]. The swordfish belongs to highly migratory pelagic species and has been observed in tropical to temperate waters (between 5 and 27 °C), and in the western and eastern Pacific and Atlantic [30]. A more detailed description of this species can be found in [30].

Age and growth estimation of swordfish presents several difficulties [29]. For example, Cerna [30] describes age estimates obtained by cross sections of the second anal fin ray [31], which appears an expensive procedure for age estimation. Queele et al. [29] recall the inconclusive results obtained from the indirect validation test.

Roa-Ureta [32] maintain that since age is a latent variable, extracting growth information objectively is difficult. He estimates growth parameters using a likelihood function approach underlying a normal mixture model to be applied on the squat lobster length dataset, where age is unknown. The normal mixture model components are determined by AIC, which depends on the sample size and the number of parameters of the mixture.

This application is motivated by the determination of age–length relationship by sex group using information measures. This is presented in a framework format based on the following steps:

- (a) The matrix of data includes both length and weight ($d = 2$) for each observation. Because it is necessary to avoid colinearity, the length–weight regression is computed to show non-linear relationship among both columns.
- (b) Given that the number of components is unknown (age is unknown), the FMSN parameters are estimated considering the two-dimensional matrix of the last step for several values m .
- (c) The number of components is determined by the bounds of information measures developed in Section 3 and then compared with AIC and BIC criteria.
- (d) The observed (measures obtained from the procedure of [30]) and estimated (by selected mixture model) ages of all observations are compared using a misclassification analysis.

Section 4.1 describes the dataset used and Sections 4.2.2 and 4.2.3 describe the results for the steps mentioned.

4.2.1. Data and Software

The dataset used for evaluating the performance of our findings corresponds to a sample of, respectively, 486 and 507 swordfish male and female length observations. The samples were collected in the southeastern Pacific off Chile during 2011 and were obtained using the routine sampling program of the fishery conducted by the Instituto de Fomento Pesquero. All these fish were measured to the nearest centimeter and the range of observed lengths. The catch included fish between 120–257 cm for males, and 110–299 cm for females. As is described in Section 4.1, the FMSN parameter estimates were computed using the `mixsmsn` package.

4.2.2. Length–Weight Relationship

Following [33] (and references therein), we briefly describe the length–weight function. This function explains the increments in weight of species in terms of their length by the non-linear relationship

$$W(x) = \alpha x^\beta, \quad (14)$$

where $W(x)$ represents the observed weight at length x , α is the theoretical weight at length zero and β is the weight growth rate.

The model (14) is fitted to an empirical dataset, $(y_i, x_i) \in \mathbb{R}_+ \times \mathbb{R}_+$, $i = 1, \dots, n$. This can be described in terms of multiplicative structure the errors, $y_i = W(x_i)\varepsilon_i$, where ε_i are non-negative random errors and their transformations are given by $\varepsilon'_i = \log \varepsilon_i$. Here, we consider the residuals ε'_i iid and normal distributed, denoted by $N(0, \sigma^2)$, for a dispersion σ^2 parameter.

Figure 2 illustrates the linear regression fits of (14), for which we have a high value for the R^2 coefficient of determination for both sexes (Table 2). There exists a small number of observations of length classes larger than 210 and 250 cm for males and females, respectively, that tends to be isolated with respect to lighter weights. Given the good fitting of length–weight model, we can see that a non-linear relationship could be assumed between length and weight. Therefore, we consider a matrix with two columns constructed by these variables for the clustering modeling.

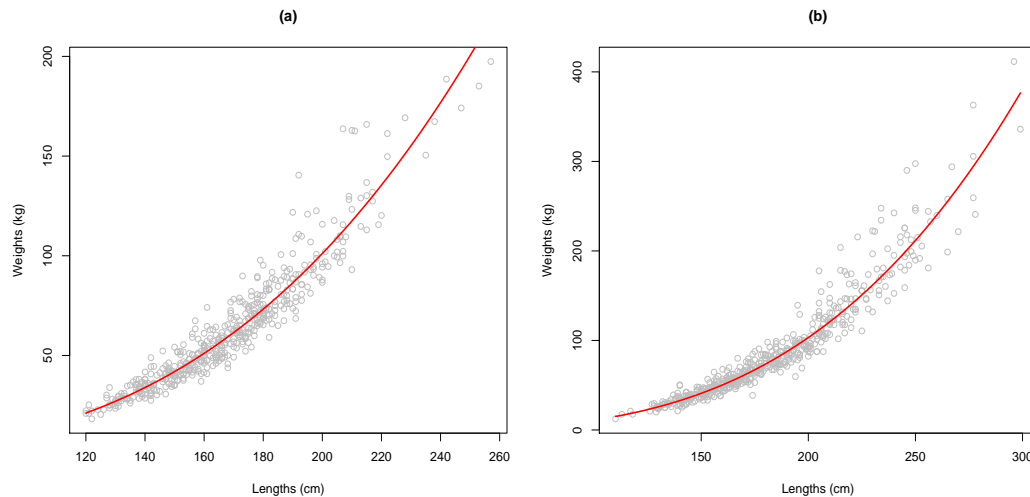


Figure 2. Length–weight log-transformed relationship and regression fit (red solid line) for (a) male and (b) female swordfish.

Table 2. Summary of estimates $\alpha' = \log \alpha$ and β with their respective standard errors in parenthesis, for each length–weight log-transformed relationships of Equation (14) and sex.

Sex	Parameter	Estimate (SE)	t-Value	p-Value	$R^2(\%)$
Male	α'	−11.619 (0.202)	−57.53	< 0.01	92.6
	β	3.064 (0.040)	77.58	< 0.01	
Female	α'	−12.413 (0.176)	−70.43	< 0.01	94.7
	β	3.218 (0.034)	94.95	< 0.01	

4.2.3. Clustering and Model Selection

As in Section 4.1, the length–weight data is evaluated with the FMSN model for several values m depending on the maximum age by sex. Some authors reported that maximum age in males and females reaches 9 and 11 years, respectively [29–31]. One of the difficulties that anal-fins readers observed was that they could find multiple annuli and disappearance of the first annulus in older fish, and thus careful interpretation is important [29]. In addition, this species were aged as younger at given body lengths, i.e., it was difficult to find older fish by selectivity [27,30]. We take into account these facts to discuss the optimal number of clusters for the classification of lengths into age classes.

To reduce the scale of the plots, in Figure 3, the logarithmic of the average between upper and lower bounds for Shannon and Rényi entropies appears, for $m = 1, \dots, 9$ in males and $m = 1, \dots, 11$ in females. It is worth pointing out that the values related to Shannon entropy (panels (a) and (c)) increase when the number of components increases in both sexes. Panels (b) and (d) show that values related to Rényi entropies increase until $m = 7$ and then decrease. This means that Rényi entropy bounds provide information of the models and help us to determine a criterion to choose a possible

number of components on each sex group. There also exist some differences between α values, where the quadratic Rényi entropy ($\alpha = 2$) provides more information.

The results mentioned before are compared first with AIC and BIC criteria in Table 3. These criteria increase when the number of components increases, and minimum AIC and BIC values correspond to the simplest model $m = 2$. Table 3 also shows the misclassification (MC) rates and consistency scores considered in Section 4.1. All these indicators are applied over the assigned observations for each cluster and the observed age, for each FMSN and Finite Mixture of Normal (FMN) model. The values corresponding to $m = 7$ clusters, marked in gray, provide the best results. The model has a classification rate of 71% and 65% for males and females, respectively; and the highest values of NSS, HSS and HK scores. The best FMN model corresponded to $m = 6$ and eight for males and females, respectively, where its respective classification rates were 57% and 55%.

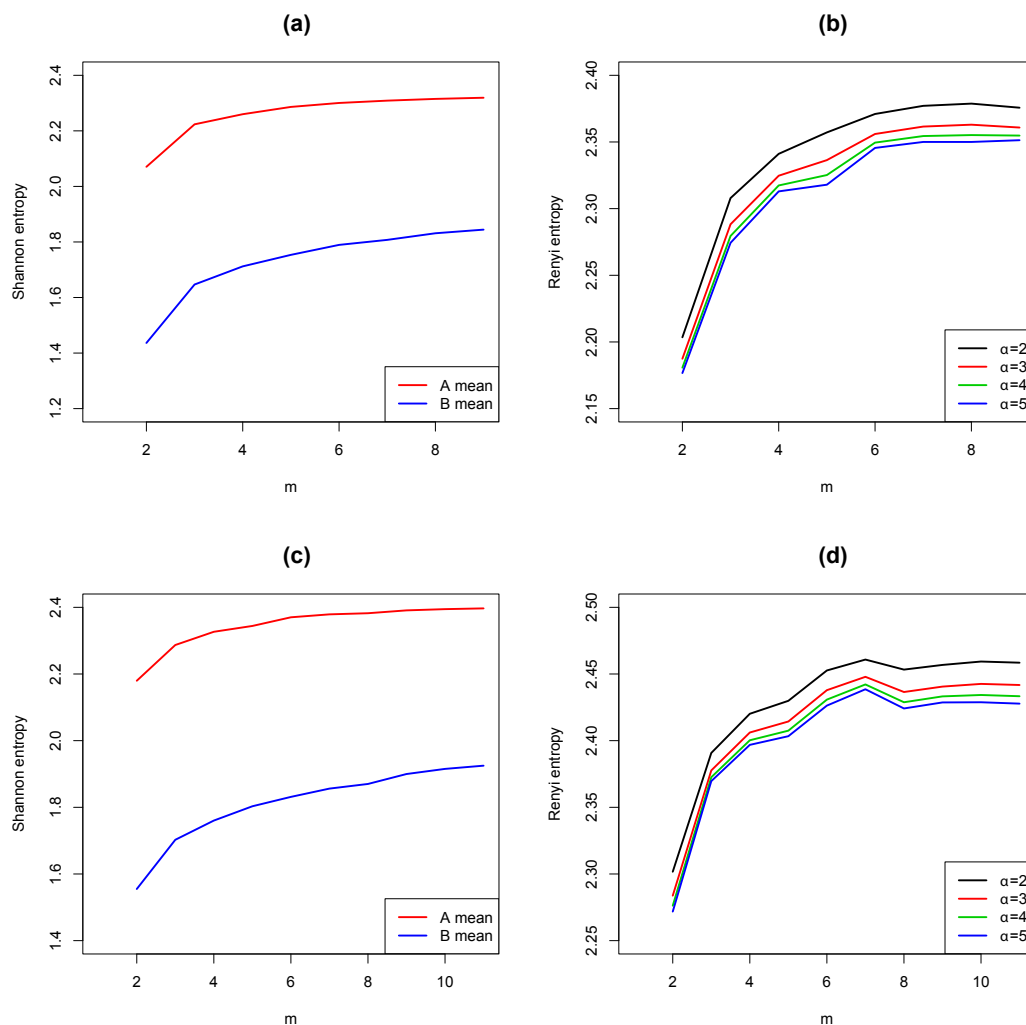


Figure 3. Logarithmic of the average between upper and lower bounds for Shannon and Rényi entropies, for (a,c) males and (b,d) females, respectively.

The FMSN fits for length–weight of males and females are shown in Figure 4. The lengths of the older species presents high variability compared to younger ones. The group of males has the parameters $\pi = (0.167, 0.117, 0.159, 0.257, 0.025, 0.084, 0.191)$,

$$\begin{aligned} \xi &= \left(\begin{pmatrix} 175.77 \\ 68.10 \end{pmatrix}, \begin{pmatrix} 192.98 \\ 93.99 \end{pmatrix}, \begin{pmatrix} 141.15 \\ 34.65 \end{pmatrix}, \begin{pmatrix} 155.55 \\ 41.27 \end{pmatrix}, \begin{pmatrix} 211.49 \\ 156.93 \end{pmatrix}, \begin{pmatrix} 201.31 \\ 105.25 \end{pmatrix}, \begin{pmatrix} 162.76 \\ 53.97 \end{pmatrix} \right), \\ \Omega &= \left(\begin{pmatrix} 6.47 & 2.05 \\ 2.05 & 7.90 \end{pmatrix}, \begin{pmatrix} 8.74 & 4.18 \\ 4.18 & 10.74 \end{pmatrix}, \begin{pmatrix} 8.45 & 4.09 \\ 4.09 & 5.28 \end{pmatrix}, \begin{pmatrix} 7.66 & 1.81 \\ 1.81 & 6.73 \end{pmatrix}, \begin{pmatrix} 21.63 & 12.29 \\ 12.29 & 14.91 \end{pmatrix}, \right. \\ &\quad \left. \begin{pmatrix} 14.16 & 7.83 \\ 7.83 & 18.12 \end{pmatrix}, \begin{pmatrix} 7.20 & 2.34 \\ 2.34 & 6.94 \end{pmatrix} \right), \text{ and} \\ \eta &= \left(\begin{pmatrix} 0.87 \\ 1.06 \end{pmatrix}, \begin{pmatrix} 0.62 \\ -1.35 \end{pmatrix}, \begin{pmatrix} -1.28 \\ -0.99 \end{pmatrix}, \begin{pmatrix} -1.12 \\ 1.23 \end{pmatrix}, \begin{pmatrix} 0.91 \\ 1.19 \end{pmatrix}, \begin{pmatrix} -0.83 \\ 0.87 \end{pmatrix}, \begin{pmatrix} 0.82 \\ 0.73 \end{pmatrix} \right); \end{aligned}$$

and for females, $\pi = (0.279, 0.058, 0.109, 0.070, 0.240, 0.015, 0.229)$,

$$\begin{aligned} \xi &= \left(\begin{pmatrix} 193.98 \\ 82.98 \end{pmatrix}, \begin{pmatrix} 236.10 \\ 205.41 \end{pmatrix}, \begin{pmatrix} 207.07 \\ 120.38 \end{pmatrix}, \begin{pmatrix} 221.49 \\ 158.33 \end{pmatrix}, \begin{pmatrix} 153.72 \\ 43.52 \end{pmatrix}, \begin{pmatrix} 264.01 \\ 283.60 \end{pmatrix}, \begin{pmatrix} 161.01 \\ 51.52 \end{pmatrix} \right), \\ \Omega &= \left(\begin{pmatrix} 11.23 & 4.31 \\ 4.31 & 13.57 \end{pmatrix}, \begin{pmatrix} 16.77 & 6.92 \\ 6.92 & 25.62 \end{pmatrix}, \begin{pmatrix} 9.68 & 4.03 \\ 4.03 & 15.32 \end{pmatrix}, \begin{pmatrix} 14.88 & 5.98 \\ 5.98 & 18.20 \end{pmatrix}, \begin{pmatrix} 13.01 & 6.91 \\ 6.91 & 9.03 \end{pmatrix}, \right. \\ &\quad \left. \begin{pmatrix} 16.87 & 13.23 \\ 13.23 & 58.63 \end{pmatrix}, \begin{pmatrix} 8.86 & 5.29 \\ 5.29 & 10.43 \end{pmatrix} \right), \text{ and} \\ \eta &= \left(\begin{pmatrix} -0.99 \\ 0.67 \end{pmatrix}, \begin{pmatrix} 1.13 \\ 1.25 \end{pmatrix}, \begin{pmatrix} 1.09 \\ 1.11 \end{pmatrix}, \begin{pmatrix} 1.18 \\ 1.33 \end{pmatrix}, \begin{pmatrix} -1.23 \\ -0.86 \end{pmatrix}, \begin{pmatrix} -0.20 \\ 1.49 \end{pmatrix}, \begin{pmatrix} 0.95 \\ 0.96 \end{pmatrix} \right). \end{aligned}$$

Table 3. Summary of Finite Mixture of Skew-Normal (FMSN) and Finite Mixture of Normal (FMN) clustering. The shaded regions highlight the smallest Akaike (AIC) and Bayesian (BIC) information criteria values. For each model and number of clusters m the misclassification (MC), normal skill (NSS), Heidke skill (HSS), and Hanssen-Kuipers (HK) scores appear.

Model	m	Male						Female					
		MC	NSS	HSS	HK	AIC	BIC	MC	NSS	HSS	HK	AIC	BIC
FMSN	2	0.70	0.30	0.01	0.01	7742.24	7805.03	0.75	0.25	0.00	0.00	8834.89	8898.32
	3	0.77	0.23	-0.05	-0.04	7754.18	7850.46	0.87	0.13	-0.05	-0.04	8844.91	8942.17
	4	0.62	0.38	0.14	0.10	7741.22	7871.00	0.89	0.11	-0.09	-0.07	8838.63	8969.71
	5	0.42	0.58	0.42	0.30	7751.47	7914.73	0.90	0.10	-0.10	-0.08	8847.75	9012.67
	6	0.45	0.55	0.43	0.35	7760.76	7957.51	0.83	0.17	-0.04	-0.03	8864.74	9063.48
	7	0.29	0.71	0.61	0.46	7770.85	8001.10	0.35	0.65	0.56	0.46	8865.46	9098.03
	8	0.51	0.49	0.37	0.30	7783.31	8047.05	0.69	0.31	0.14	0.11	8879.20	9145.59
	9	0.65	0.35	0.22	0.18	7769.48	8066.70	0.49	0.51	0.42	0.35	8885.98	9186.20
	10	-	-	-	-	-	-	0.59	0.41	0.31	0.27	8897.56	9231.61
	11	-	-	-	-	-	-	0.65	0.35	0.26	0.22	8900.87	9268.75
	FMN	2	0.70	0.30	0.02	0.01	7818.79	7864.84	0.75	0.25	0.00	0.00	8914.32
3		0.78	0.22	-0.04	-0.03	7737.23	7808.40	0.88	0.12	-0.03	-0.03	8848.43	8920.31
4		0.52	0.48	0.28	0.20	7729.77	7826.05	0.87	0.13	-0.08	-0.06	8818.25	8915.51
5		0.43	0.57	0.43	0.32	7733.33	7854.73	0.82	0.18	-0.05	-0.04	8820.12	8942.74
6		0.43	0.57	0.46	0.36	7744.00	7890.52	0.70	0.30	0.11	0.09	8831.16	8979.16
7		0.53	0.47	0.35	0.29	7738.41	7910.04	0.66	0.34	0.17	0.14	8839.63	9013.00
8		0.52	0.48	0.36	0.29	7750.27	7947.02	0.45	0.55	0.47	0.39	8849.82	9048.56
9		0.85	0.15	-0.01	-0.01	7751.24	7973.11	0.50	0.50	0.41	0.35	8855.10	9079.21
10		-	-	-	-	-	-	0.78	0.22	0.11	0.10	8857.49	9106.97
11		-	-	-	-	-	-	0.62	0.38	0.29	0.25	8852.37	9127.22

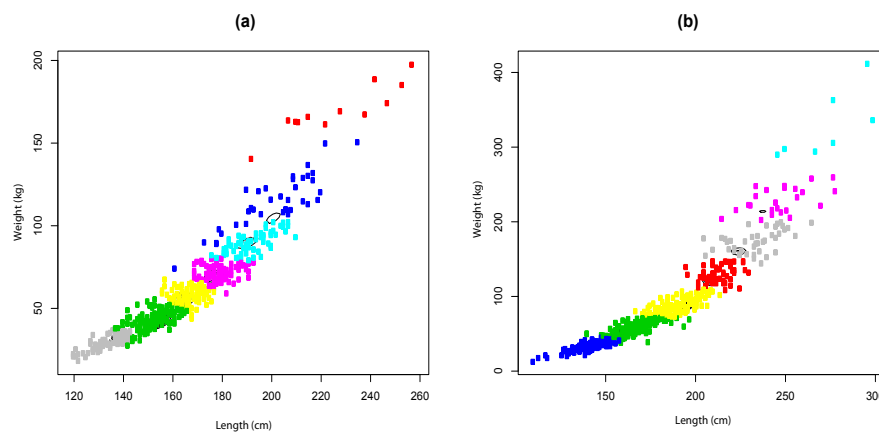


Figure 4. Selected Finite Mixture of Skew-Normal (FMSN) fits for (a) males and (b) females. Each color corresponds to each mixture component from a total of $m = 7$.

5. Conclusions

5.1. Methodology

In this paper, lower and upper bounds of the Shannon and Rényi entropies for FMSN distributions were derived. Using such a pair of bounds, some kind of confidence interval for the approximate entropy value can be calculated, where the average between these values can be used as an approximation of the entropy. We presented practical (bounds) and theoretical (bounds and asymptotic expression) results for Rényi entropy. In the case of practical results, the first upper bound deals only with the density parameters and the second one with the density and mixing weights parameters.

Inserting the ML estimation (fixed) parameters represents the simplest evaluation of these bounds [22]. However, between the lower and upper Rényi entropy bounds exists a considerable distance. For this reason, further research must consider the exact expression and asymptotic approximation presented in this paper. In addition, the Bayesian approach allows for a direct estimation of the entropies, depending on the accuracy of prior parameters, where performance can be substantially improved compared to ML or nonparametric estimators [34].

The results presented are valid for the skew-normal case, taking the shape parameters set $\eta = (0, \dots, 0)$, for integer values of α [11]. However, numerical algorithms can be applied for real values of α ($\alpha > 2$), but that requires more challenging computational work. In addition, Proposition 2 and Lemma 1 are also valid for other continuous densities where the Rényi entropies of the component exist. We hope the Rényi entropy developments in finite mixtures of densities can stimulate more research in the future, for more flexible densities such as Skew- t distribution [27].

5.2. Application

We applied two-dimensional length–weight data for the determination of swordfish age. We considered a length–weight dataset instead of the usual length (considered by [32]) to determine the number of clusters, and posteriorly, we compared it with the real observations obtained by the procedure of [30]. The best results were obtained using the Rényi entropy, as an average between upper and lower bounds, over Shannon entropy and information criteria. Additionally, the classification rates and consistency scores of FMSN models showed better results versus the FMN model.

Wrongly classified observations arise with older species because they produce higher variability in the length–weight relationship. Moreover, the age determination in these age classes is difficult to obtain for the reasons mentioned in Section 4.2. We encourage anal-fins readers to consider the

proposed methodology to compare their results with this statistical methodology, especially for the revision of older species data.

Supplementary Materials: The R codes of the upper and lower bounds of Shannon and Rényi entropies are available by request to the correspondence author. The swordfish data are available by request to the Instituto de Fomento Pesquero (IFOP, Valparaíso, Chile), website: <http://www.ifop.cl>.

Acknowledgments: We thank the Instituto de Fomento Pesquero (IFOP, Valparaíso, Chile), for providing the biological information used in this work. The research of J. Contreras-Reyes was supported by Comisión Nacional de Investigación Científico y Tecnológico (CONICYT, Santiago, Chile) doctoral scholarship 2016 No. 21160618 (Res. Ex. 4128/2016). We would like to thank the editor and three anonymous reviewers for their helpful comments, suggestions, and valuable discussion of this work.

Author Contributions: J. Contreras-Reyes and D. Devia Cortés conceived the experiments and analyzed the data; J. Contreras-Reyes designed and performed the experiments, contributed reagents/analysis tools and wrote the paper; and D. Devia Cortés contributed materials tools. All authors have read and approved the final manuscript.

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

Appendix A

Proof of Proposition 2.

- (i) For any finite mixture $f(\mathbf{x}; \tilde{\theta}, \boldsymbol{\pi}) = \sum_{i=1}^m \pi_i f(\mathbf{x}; \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is the associated parameter set of each i -th component $f(\mathbf{x}; \boldsymbol{\theta}_i)$, $\tilde{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m)$, $\pi_i \geq 0$, $\sum_{i=1}^m \pi_i = 1$, and $\mathbf{X} \in \mathbb{R}^d$ is not necessarily normal with non-zero location vector and dispersion matrix $\boldsymbol{\Lambda}$. Then,

$$\sum_{i=1}^m \pi_i \mathcal{H}[\mathbf{X}; \boldsymbol{\theta}_i] \leq \mathcal{H}[\mathbf{X}; \tilde{\theta}] \leq \frac{1}{2} \ln\{(2\pi e)^d |\boldsymbol{\Lambda}|\}. \tag{A1}$$

For a proof of Equation (A1), see pp. 27 and 663 of [20]. Basically, the fact that $g(t) = -\ln t$ is a concave function ($-g(t)$ is convex) allows the use of Jensen’s inequality. Then, considering the location vector (7) and dispersion matrix (8), we have the inequalities $\mathcal{H}[\mathbf{Y}; \tilde{\theta}] \leq A_{upper}$ and $\mathcal{H}[\mathbf{Y}; \tilde{\theta}] \geq \sum_{i=1}^m \pi_i \mathcal{H}[\mathbf{Y}; \boldsymbol{\theta}_i]$. Considering the property (i) of Proposition 1 and the condition $\sum_{i=1}^m \pi_i = 1$, we prove the left side of the inequality.

- (ii) Left side: by the property of log concavity for skew-normal densities [35] and employing Jensen’s inequality [20], the proof is analogous to Theorem 2 of [9]. Right side: see Theorem 3 of [9]. \square

Proof of Lemma 1. Consider the Proposition 1 (B1) of [36]. Let $p \geq 1$, then for a α -th-order, $1 < \alpha \leq p$, we have

$$\begin{aligned} p(\mathbf{y}; \tilde{\theta}, \boldsymbol{\pi})^\alpha &= \left[\sum_{i=1}^m \pi_i f(\mathbf{y}; \boldsymbol{\theta}_i) \right]^\alpha \\ &\stackrel{(B1)}{\geq} \left(\sum_{i=1}^m f(\mathbf{y}; \boldsymbol{\theta}_i)^p \right)^{\frac{\alpha}{p}-1} \left[\sum_{i=1}^{m-1} \left\{ i^{1-\frac{\alpha}{p}} \left(\sum_{k=1}^i \pi_k \right)^\alpha \left(f(\mathbf{y}; \boldsymbol{\theta}_i)^p - f(\mathbf{y}; \boldsymbol{\theta}_{i+1})^p \right) \right\} \right. \\ &\quad \left. + m^{1-\frac{\alpha}{p}} \left(\sum_{k=1}^m \pi_k \right)^\alpha f(\mathbf{y}; \boldsymbol{\theta}_m)^p \right]. \end{aligned} \tag{A2}$$

By choosing $p = \alpha$ related to condition (iii) of Proposition 1 of [36] in Equation (A2), the following equality holds

$$p(\mathbf{y}; \tilde{\theta}, \boldsymbol{\pi})^\alpha \geq f(\mathbf{y}; \boldsymbol{\theta}_m)^\alpha + \sum_{i=1}^{m-1} \left\{ \left(\sum_{k=1}^i \pi_k \right)^\alpha \left(f(\mathbf{y}; \boldsymbol{\theta}_i)^\alpha - f(\mathbf{y}; \boldsymbol{\theta}_{i+1})^\alpha \right) \right\}. \tag{A3}$$

The conditions (i), (ii) and (iv) of Proposition 1 of [36] can not be accomplished given the Rényi entropy conditions of Equation (9), and thus the equality in (A3) is not accomplished. Finally, integrating both sides of (A3) the result is obtained. \square

References

1. McLachlan, G.; Peel, D. *Finite Mixture Models*; John Wiley & Sons: New York, NY, USA, 2000.
2. Celeux, G.; Soromenho, G. An entropy criterion for assessing the number of clusters in a mixture model. *J. Classif.* **1996**, *13*, 195–212.
3. Jenssen, R.; Hild, K.E.; Erdogmus, D.; Principe, J.C.; Eltoft, T. Clustering using Rényi's entropy. *IEEE Proc. Int. Jt. Conf. Neural Netw.* **2003**, *1*, 523–528.
4. Amoud, H.; Snoussi, H.; Hewson, D.; Doussot, M.; Duchêne, J. Intrinsic mode entropy for nonlinear discriminant analysis. *IEEE Signal Process. Lett.* **2007**, *14*, 297–300.
5. Caillol, H.; Pieczynski, W.; Hillion, A. Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation. *IEEE Trans. Image Process.* **1997**, *6*, 425–440.
6. Carreira-Perpiñán, M.A. Mode-finding for mixtures of Gaussian distributions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1318–1323.
7. Durrieu, J.-L.; Thiran, J.; Kelly, F. Lower and upper bounds for approximation of the Kullback–Leibler divergence between Gaussian mixture models. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 4833–4836.
8. Nielsen, F.; Sun, K. Guaranteed bounds on the Kullback–Leibler divergence of univariate mixtures. *IEEE Signal Process. Lett.* **2016**, *23*, 1543–1546.
9. Huber, M.F.; Bailey, T.; Durrant-Whyte, H.; Hanebeck, U.D. On entropy approximation for Gaussian mixture random vectors. In Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Seoul, Korea, 20–22 August 2008; pp. 181–188.
10. Rényi, A. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; Neyman, J., Ed.; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
11. Contreras-Reyes, J.E. Rényi entropy and complexity measure for skew-gaussian distributions and related families. *Physica A* **2015**, *433*, 84–91.
12. Zografos, K.; Nadarajah, S. Expressions for Rényi and Shannon entropies for multivariate distributions. *Stat. Probab. Lett.* **2005**, *71*, 71–84.
13. Lin, T.I. Maximum likelihood estimation for multivariate skew normal mixture models. *J. Multivar. Anal.* **2009**, *100*, 257–265.
14. Frühwirth-Schnatter, S.; Pyne, S. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics* **2010**, *11*, 317–336.
15. Lee, S.X.; McLachlan, G.J. On mixtures of skew normal and skew t-distributions. *Adv. Data Anal. Classif.* **2013**, *7*, 241–266.
16. Lee, S.X.; McLachlan, G.J. Model-based clustering and classification with non-normal mixture distributions. *Stat. Meth. Appl.* **2013**, *22*, 427–454.
17. Lin, T.I.; Ho, H.J.; Lee, C.R. Flexible mixture modelling using the multivariate skew-t-normal distribution. *Stat. Comput.* **2014**, *24*, 531–546.
18. Azzalini, A.; Dalla-Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726.
19. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distributions. *J. R. Stat. Soc. Ser. B* **1999**, *61*, 579–602.
20. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley & Son, Inc.: New York, NY, USA, 2006.
21. Schrödinger, E. *What is Life—The Physical Aspect of the Living Cell*; Cambridge University Press: Cambridge, UK, 1944.
22. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback–Leibler divergence measure for multivariate skew-normal distributions. *Entropy* **2012**, *14*, 1606–1626.
23. Sánchez-Moreno, P.; Zozor, S.; Dehesa, J.S. Upper bounds on Shannon and Rényi entropies for central potentials. *J. Math. Phys.* **2011**, *52*, 022105.

24. Prates, M.O.; Lachos, V.H.; Cabral, C. Mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *J. Stat. Softw.* **2013**, *54*, 1–20.
25. Lee, S.X.; McLachlan, G.J. EMMIXskew: An R package for fitting mixtures of multivariate skew t-distributions via the EM algorithm. *J. Stat. Softw.* **2013**, *55*, 1–22.
26. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
27. Contreras-Reyes, J.E.; Arellano-Valle, R.B.; Canales, T.M. Comparing growth curves with asymmetric heavy-tailed errors: Application to the southern blue whiting (*Micromesistius australis*). *Fish. Res.* **2014**, *159*, 88–94.
28. Contreras-Reyes, J.E. Nonparametric assessment of aftershock clusters of the maule earthquake Mw = 8.8. *J. Data Sci.* **2013**, *11*, 623–638.
29. Quelle, P.; González, F.; Ruiz, M.; Valeiras, X.; Gutierrez, O.; Rodriguez-Marin, E.; Mejuto, J. An approach to age and growth of south Atlantic swordfish (*Xiphias gladius*) Stock. *Collect. Vol. Sci. Pap. ICCAT* **2014**, *70*, 1927–1944.
30. Cerna, J.F. Age and growth of the swordfish (*Xiphias gladius Linnaeus*, 1758) in the southeastern Pacific off Chile. *Lat. Am. J. Aquat. Res.* **2009**, *37*, 59–69.
31. Sun, C.L.; Wang, S.P.; Yeh, S.Z. Age and growth of the swordfish (*Xiphias gladius L.*) in the waters around Taiwan determined from anal-fin rays. *Fish. Bull.* **2002**, *100*, 822–835.
32. Roa-Ureta, R.H. A likelihood-based model of fish growth with multiple length frequency data. *J. Agric. Biol. Environ. Stat.* **2010**, *15*, 416–429.
33. Contreras-Reyes, J.E. Analyzing fish condition factor index through skew-gaussian information theory quantifiers. *Fluct. Noise Lett.* **2016**, *15*, 1650013.
34. Gupta, M.; Srivastava, S. Parametric Bayesian estimation of differential entropy and relative entropy. *Entropy* **2010**, *12*, 818–843.
35. Gupta, R.C.; Brown, N. Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Commun. Stat. Theory Methods* **2001**, *30*, 2427–2445.
36. Bennett, G. Lower bounds for matrices. *Linear Algebra Appl.* **1986**, *82*, 81–98.



© 2016 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/>).