mathematics

MDPI

Article

An Information-Theoretic Approach for Multivariate Skew-*t* Distributions and Applications †

Salah H. Abid [1], Uday J. Quaez [1] and Javier E. Contreras-Reyes [2,*]

[1] Department of Mathematics, Education College, Al-Mustansiriya University, Baghdad 14022, Iraq; abidsalah@uomustansiriyah.edu.iq (S.H.A.); uoday1977@gmail.com (U.J.Q.)
[2] Instituto de Estadística, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso 2360102, Chile
* Correspondence: jecontrr@uc.cl; Tel.: +56-9-7960-8218

† This work corresponds to a revised, corrected and extended version of the pre-print: Abid, S.H., Quaez, U.J. (2019). Renyi and Shannon Entropies of Finite Mixtures of Multivariate Skew t-distributions. arXiv:1901.10569.

Abstract: Shannon and Rényi entropies are two important measures of uncertainty for data analysis. These entropies have been studied for multivariate Student-*t* and skew-normal distributions. In this paper, we extend the Rényi entropy to multivariate skew-*t* and finite mixture of multivariate skew-*t* (FMST) distributions. This class of flexible distributions allows handling asymmetry and tail weight behavior simultaneously. We find upper and lower bounds of Rényi entropy for these families. Numerical simulations illustrate the results for several scenarios: symmetry/asymmetry and light/heavy-tails. Finally, we present applications of our findings to a swordfish length-weight dataset to illustrate the behavior of entropies of the FMST distribution. Comparisons with the counterparts—the finite mixture of multivariate skew-normal and normal distributions—are also presented.

Keywords: skew-*t*; finite mixtures; skewness; heavy-tails; Shannon entropy; Rényi entropy; swordfish data

check for updates

Citation: Abid, S.H.; Quaez, U.J.; Contreras-Reyes, J.E. An Information-Theoretic Approach for Multivariate Skew-*t* Distributions and Applications. *Mathematics* **2021**, *9*, 146. https://doi.org/10.3390/math9020146

Received: 17 December 2020
Accepted: 8 January 2021
Published: 11 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Finite mixture models have used in the analysis of heterogeneous datasets due to its flexibility [1]. These models provide important applications in many scientific fields such as density estimation [2], data mining and pattern recognition [3], image processing and satellite imaging [4], medicine and genetics [4,5], fisheries [6,7], astronomy [2], and more. Specifically, Carreira-Perpiñán [2] considered the mixture of normal densities to find the modes of multi-dimensional data via the Shannon entropy. Because no analytical expressions exist for Shannon entropy of normal mixture of densities, the authors of [2] considered bounds to approximates the Shannon entropy [8]. Given that these applications have been developed in the normal mixture of densities context, several calculations of Shannon and Rényi entropies for non-normal distributions exist.

Azzalini and Dalla-Valle [9] and Azzalini and Capitanio [10] introduced the multivariate skew-normal and skew-*t* distributions as an alternative to multivariate normal distribution to deal with skewness and heavy-tailness in the data, respectively. Lin et al. [11] proposed a development of finite mixture of skew-*t* (FMST) distributions, and more recently [4] provided an overview of developments of FMST distributions. Arellano-Valle et al. [12] and Contreras-Reyes [13] presented the mutual information and Shannon entropy for multivariate skew-normal and skew-*t* distributions, respectively. They highlighted that the calculation of mutual information index and Shannon entropy include the negentropy concept represented by a one-dimensional integral. Contreras-Reyes [14] discussed the values of Rényi entropy of multivariate skew-normal and extended skew-normal distributions. Contreras-Reyes and Cortés [6] considered bounds to approximate the Rényi entropy for finite mixture of multivariate skew-normal (FMSN) distributions by using lower and upper

bounds. More recently, the authors of [15,16] presented the bounds of Rényi entropy of multivariate skew-normal-Cauchy and skew-Laplace distributions, and [17] considered the Kullback–Leibler divergence based on multivariate skew-t densities [13] as a classifier for an SAR image system.

In this paper, we propose an information-theoretic approach for FMST distributions. The explicit expressions of Shannon and Rényi entropies of skew-t distribution are derived. We give bounds for Rényi entropies of FMST models. An approximate value of these entropies can be calculated. Simulation studies illustrate the behavior of Rényi entropy approximations for a given order α, skewness, and freedom degrees of the proposed mixture model. Numerical simulations illustrates the results for several scenarios: symmetry/asymmetry and light/heavy-tails. Finally, a real data application of a swordfish length-weight dataset is revisited from the work in [6], where the skew-normal model is compared with the proposed skew-t one.

The paper is organized as follows. Section 2 gives the propositions, lemmas, and numerical simulations for computation of Shannon and Rényi entropies of multivariate skew-t random variables. Section 3 provides bounds to approximate Shannon and Rényi entropies for FMST distributions and a numerical application of swordfish data. Finally, Section 4 concludes the study.

2. Multivariate Skew-t Distribution

The multivariate t-distribution corresponds to a generalization of the Student-t-distribution. Let $\mathbf{x} \in \mathbb{R}^d$ be a random vector, $d \in \mathbb{Z}_+$, is multivariate t-distributed [18], denoted as $\mathbf{X} \sim T_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$, whose probability density function (pdf) is given by

$$t_d(\mathbf{x}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = B_d(\nu)|\boldsymbol{\Omega}|^{-1/2}\left(1 + \frac{Q(\mathbf{x})}{\nu}\right), \tag{1}$$

where $Q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\xi})^\top \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\xi})$, $\mu \in \mathbb{R}^d$ is a location parameter, $\boldsymbol{\Omega} \in \mathbb{R}^{d\times d}$ is a dispersion matrix, $\nu \in \mathbb{R}$ is the degree of freedom parameter,

$$B_\ell(\nu) = \frac{\Gamma\left(\frac{\nu+\ell}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\ell/2}}, \quad \nu > 0, \quad \forall \ell \in \mathbb{Z}, \tag{2}$$

and $\Gamma(\cdot)$ is the gamma function.

The multivariate normal distribution is obtained from the limit $T_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) \to N_d(\boldsymbol{\xi}, \boldsymbol{\Omega})$, as $\nu \to \infty$. In the special case where $\nu = 1$, $\boldsymbol{\xi} = \mathbf{0}$ and $\boldsymbol{\Omega} = \boldsymbol{I}_d$, $t_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ leads to the multivariate Cauchy distribution. The mean and variance of random vector $\mathbf{x}$ are

$$E[\mathbf{X}] = \boldsymbol{\xi}, \quad \nu > 1, \tag{3}$$

$$V[\mathbf{X}] = \left(\frac{B_{-2}(\nu)}{2\pi}\right)\boldsymbol{\Omega}, \quad \nu > 2, \tag{4}$$

respectively.

An extension of multivariate skew-normal distribution is the skew-t one by incorporating a degree of freedom parameter representing the presence of atypical observations in empirical data [10,12,19]. We considered the definition of skew-t distribution given in [12]. Let $\mathbf{y} \in \mathbb{R}^d$ be a multivariate skew-t random variable with location vector $\boldsymbol{\xi} \in \mathbb{R}^d$, dispersion matrix $\boldsymbol{\Omega} \in \mathbb{R}^{d\times d}$, asymmetry vector $\boldsymbol{\eta} \in \mathbb{R}^d$, and $\nu > 0$ degrees of freedom. We denoted $\mathbf{y}$ by $\mathbf{Y} \sim ST_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}, \nu)$ if has pdf given by

$$f(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}, \nu) = 2t_d(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)T\left(\boldsymbol{\eta}^\top \boldsymbol{\Omega}^{-1/2}(\mathbf{y} - \boldsymbol{\xi})\sqrt{\frac{\nu + d}{\nu + Q(\mathbf{y})}}; \nu + d\right), \tag{5}$$

where $\bar{\boldsymbol{\eta}} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\eta}$ and $T(\mathbf{x}; \nu + d) \equiv T_1(\mathbf{x}; 0, 1, \nu + d)$ is the univariate Student-t cumulative distribution function (cdf) of $\mathbf{x}$.

The multivariate Student-t distribution is a special case of the multivariate skew-t one and is related to a multivariate skew-normal random variable through equation

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\xi} + V^{-1/2}\mathbf{Z}_0, \tag{6}$$

where $\mathbf{Z}_0 = \boldsymbol{\Omega}^{-1/2}(\mathbf{Y} - \boldsymbol{\xi}) \sim SN_d(\mathbf{0}, \boldsymbol{I}_d, \boldsymbol{\eta})$ and $V \sim \chi^2/\nu$, and they are independent, where SN_d and χ^2 denotes the d-variate skew-normal [9] and univariate chi-square distribution [10], respectively. The multivariate skew-t distribution yields the multivariate skew-normal and Student-t ones as $\nu \to \infty$ and when $\boldsymbol{\eta} = \mathbf{0}$, respectively [4].

Considering the m-cumulant function

$$E(V^{-m/2}) = \frac{B_{-m}(\nu)}{(2\pi)^{m/2}}, \quad m \geq 1, \quad \nu > m,$$

and Equations (2)–(4), the mean vector and covariance matrix of $\mathbf{Y}$ are, respectively, derived by the authors of [10] in the following forms,

$$E[\mathbf{Y}] = \boldsymbol{\xi} + \frac{B_{-1}(\nu)}{\pi}\bar{\boldsymbol{\eta}}, \quad \nu > 1, \tag{7}$$

$$V[\mathbf{Y}] = \boldsymbol{\xi}\boldsymbol{\xi}^\top + \left(\frac{B_{-2}(\nu)}{2\pi}\right)\boldsymbol{\Omega} + \left(\frac{B_{-1}(\nu)}{\pi}\right)^2(\boldsymbol{\xi}\boldsymbol{\delta}^\top + \boldsymbol{\delta}\boldsymbol{\xi}^\top), \quad \nu > 2, \tag{8}$$

where $\boldsymbol{\delta} = \boldsymbol{\Omega}\boldsymbol{\eta}/\sqrt{1 + \boldsymbol{\eta}^\top\boldsymbol{\Omega}\boldsymbol{\eta}}$.

2.1. Entropies

Shannon [20] proposed an entropy to quantify the uncertainty of a system and posteriorly, Rényi [21] generalized the Shannon entropy for any probability distribution related to a discrete or continuous random variable. In this section, we derive these measures for the skew-t case.

Consider a pdf f associated with a random variable $\mathbf{y} \in \mathbb{R}^d$, denoted by $H(\mathbf{Y})$, the well-known Shannon entropy [22]. It is defined by

$$H(\mathbf{Z}) = -\int_{\mathbb{R}^d} f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}. \tag{9}$$

Proposition 1. *The Shannon entropy of* $\mathbf{Y} \sim ST_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}, \nu)$ *is*

$$H(\mathbf{Y}) = H(\mathbf{X}) - E\left[\log\left\{2T\left(\frac{\sqrt{\nu + d}\,\|\bar{\boldsymbol{\eta}}\|\,W_H}{\sqrt{\nu + d - 1 + W_H^2}}; \nu + k\right)\right\}\right], \tag{10}$$

where $W_H \sim ST_1(0, 1, \|\bar{\boldsymbol{\eta}}\|, \nu + d - 1)$, $\|\bar{\boldsymbol{\eta}}\| = \bar{\boldsymbol{\eta}}^\top\bar{\boldsymbol{\eta}} = (\boldsymbol{\eta}^\top\boldsymbol{\Omega}\boldsymbol{\eta})^{1/2}$, *and* $H(\mathbf{X})$ *is the Student-t Shannon entropy of* $\mathbf{X} \sim T_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ *given by*

$$H(\mathbf{X}) = \frac{1}{2}\log|\boldsymbol{\Omega}| - \log B_d(\nu) + \Psi_d(\nu), \tag{11}$$

where

$$\Psi_d(\nu) = \left(\frac{\nu + d}{2}\right)\left[\psi\left(\frac{\nu + d}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right], \tag{12}$$

$B_d(\nu)$ *is defined in (2) and* $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ *is the digamma function.*

The proof for the Shannon entropy of $\mathbf{Y}$ in (10) is shown in [12]. The same entropy, but for $\mathbf{X}$ in (11), was first presented in [23].

An extension of Shannon entropy is the αth Rényi order given by [21]. The Rényi entropy of $\mathbf{Y}$ is denoted by $R_\alpha(\mathbf{Y})$ and defined by

$$R_\alpha(\mathbf{Y}) = \frac{1}{1-\alpha}\log\int_{\mathbb{R}^d}[f(\mathbf{y})]^\alpha d\mathbf{y}, \tag{13}$$

with $0 < \alpha < \infty$, $\alpha \neq 1$, and where normalization to unity as given by $\int_{\mathbb{R}^d} f(\mathbf{y})d\mathbf{y} = 1$ [24]. The Shannon entropy is obtained from $R_\alpha(\mathbf{Y})$ as $\alpha \to 1$ [14,21]. The Rényi entropy is invariant under a location transformation, but not invariant under a scale transformation. For any $\alpha_1 < \alpha_2$, we have $R_{\alpha_1}(\mathbf{Y}) \geq R_{\alpha_2}(\mathbf{Y})$, and $R_{\alpha_1}(\mathbf{Y}) = R_{\alpha_2}(\mathbf{Y})$ if and only if the system is uniformly distributed.

In order to compute the Rényi entropy of a multivariate skew-t random variate, we need the following preliminary results that involve the pdf (5).

Lemma 1. *Let* $\mathbf{Y} \sim ST_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}, \nu)$. *Then,*

$$\begin{aligned}
&E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\boldsymbol{\Omega}^{-1/2}(\mathbf{y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{y})}};\nu+d\right)\right\}\right]\\
&= E\left[\log\left\{2T\left(\frac{\sqrt{\nu+d}\|\bar{\boldsymbol{\eta}}\|X_0}{\sqrt{\nu+d-1+X_0^2}};\nu+d\right)\right\}\left\{2T\left(\frac{\sqrt{\nu+1}\|\bar{\boldsymbol{\eta}}\|X_0}{\sqrt{\nu+X_0^2}};\nu+d\right)\right\}\right],
\end{aligned}$$

where $X_0 \sim T_1(0, 1, \nu+d-1)$.

Proof. We directly get

$$\begin{aligned}
&E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\boldsymbol{\Omega}^{-1/2}(\mathbf{Y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{Y})}};\nu+d\right)\right\}\right]\\
&= \int_{\mathbb{R}^d}\log\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\boldsymbol{\Omega}^{-1/2}(\mathbf{y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{y})}};\nu+d\right)\right\}\\
&\times 2t_d(\mathbf{y};\boldsymbol{\xi},\boldsymbol{\Omega},\nu)T\left(\bar{\boldsymbol{\eta}}^\top\boldsymbol{\Omega}^{-1/2}(\mathbf{y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{y})}};\nu+d\right)d\mathbf{y}.
\end{aligned}$$

Using the change of variables $\mathbf{z}_0 = \boldsymbol{\Omega}^{-1/2}(\mathbf{y}-\boldsymbol{\xi})$, we get that $Z_0 \sim ST_d(\mathbf{0}, \boldsymbol{I}_d, \boldsymbol{\eta}, \nu)$ and the determinant of Jacobian matrix $|\boldsymbol{\Omega}|^{-1/2}$. Therefore,

$$\begin{aligned}
&E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\boldsymbol{\Omega}^{-1/2}(\mathbf{Y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{Y})}};\nu+d\right)\right\}\right]\\
&= \int_{\mathbb{R}^d}\log\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\mathbf{z}_0\sqrt{\frac{\nu+d}{\nu+\mathbf{z}_0^\top\mathbf{z}_0}};\nu+d\right)\right\}\\
&\times 2|\boldsymbol{\Omega}|^{-1/2}t_d(\mathbf{z}_0;\mathbf{0},\boldsymbol{I}_d,\nu)T\left(\bar{\boldsymbol{\eta}}^\top\mathbf{z}_0\sqrt{\frac{\nu+d}{\nu+\mathbf{z}_0^\top\mathbf{z}_0}};\nu+d\right)d\mathbf{y}\\
&= E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\mathbf{Z}_0\sqrt{\frac{\nu+d}{\nu+\mathbf{Z}_0^\top\mathbf{Z}_0}};\nu+d\right)\right\}\right].
\end{aligned}$$

Lemma 3 of [12] implies that

$$E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^{\top}\boldsymbol{\Omega}^{-1/2}(\mathbf{Y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{Y})}};\nu+d\right)\right\}\right]$$
$$= E\left[\log\left\{2T\left(\frac{\sqrt{\nu+d}\|\bar{\boldsymbol{\eta}}\|W_R}{\sqrt{\nu+d-1+W_R^2}};\nu+d\right)\right\}\right],$$

with $W_R \sim ST_1(0,1,\|\bar{\boldsymbol{\eta}}\|,\nu+d-1)$ and $\|\bar{\boldsymbol{\eta}}\| = \bar{\boldsymbol{\eta}}^{\top}\bar{\boldsymbol{\eta}}$. The latter result yields the proof. □

Lemma 1 allows us to compute an expected value of a d-variate random variable $\mathbf{Y}$ as an expected value of a univariate Student-t one X_0. This result is also applied in the computation of Shannon entropy of $\mathbf{Y}$; however, the latter depends on a univariate skew-t. Then, we can represent the expected value of (10) with respect to a univariate Student-t random variable as follows,

Proposition 2. *Let* $\mathbf{X} \sim T_d(\boldsymbol{\xi},\boldsymbol{\Omega},\nu)$ *and* $\mathbf{Y} \sim ST_d(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)$*, the Shannon entropy of* $\mathbf{Y}$ *can be written as*

$$H(\mathbf{Y}) = H(\mathbf{X}) - N(X_0), \tag{14}$$

where

$$N(X_0) = E\left[\log\left\{2T\left(\frac{\sqrt{\nu+d}\|\bar{\boldsymbol{\eta}}\|X_0}{\sqrt{\nu+d-1+X_0^2}};\nu+d\right)\right\}\left\{2T\left(\frac{\sqrt{\nu+1}\|\bar{\boldsymbol{\eta}}\|X_0}{\sqrt{\nu+X_0^2}};\nu+d\right)\right\}\right],$$

and $X_0 \sim T_1(0,1,\nu+d-1)$.

Proof. By computing the natural logarithm and expectation in both sides of pdf (5), we have

$$E[\log\{f(\mathbf{Y};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)\}] = E[\log\{t_d(\mathbf{Y};\boldsymbol{\xi},\boldsymbol{\Omega},\nu)\}]$$
$$+E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^{\top}\boldsymbol{\Omega}^{-1/2}(\mathbf{Y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{Y})}};\nu+d\right)\right\}\right].$$

Using Lemma 1 and Proposition 1 of [25], we obtain

$$H(\mathbf{Y}) = H(\mathbf{X}) - E\left[\log\left\{2T\left(\bar{\boldsymbol{\eta}}^{\top}\boldsymbol{\Omega}^{-1/2}(\mathbf{Y}-\boldsymbol{\xi})\sqrt{\frac{\nu+d}{\nu+Q(\mathbf{Y})}};\nu+d\right)\right\}\right].$$

Then, Lemma 1 gives us the required result for this proof. □

Lemma 1 and Proposition 2 decompose the expected value into two parts; however, it is still necessary to solve the integral (13) that involves parameter α. The next Lemma uses the previous results to obtain the skew-t Rényi entropy.

Lemma 2. *Let* $\mathbf{Y} \sim ST_d(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)$*. Then,*

$$\int_{\mathbb{R}^d}[f(\mathbf{y};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)]^{\alpha}d\mathbf{y} = C_{\alpha,d}(\boldsymbol{\Omega},\nu)\,E\left[\left\{2T\left(\frac{\sqrt{\nu+d}\|\bar{\boldsymbol{\eta}}\|X_0}{\sqrt{\nu+d-1+X_0^2}};\nu+d\right)\right\}^{\alpha}\right], \tag{15}$$

where

$$C_{\alpha,d}(\boldsymbol{\Omega},\nu) = |\boldsymbol{\Omega}|^{-\alpha/2}[B_d(\nu)]^{\alpha}B_{-d}(\alpha[\nu+d])\left(\frac{\nu}{\alpha[\nu+d]}\right)^{d/2},$$

and $X_0 \sim T_1(0,1,\beta)$, $\beta = \alpha(\nu+d)-d$, $0<\alpha<\infty$, $\alpha \neq 1$.

Proof. Using the change of variables $\mathbf{z}_0 = \boldsymbol{\Omega}^{-1/2}(\mathbf{y}-\boldsymbol{\xi})$ associated with Jacobian matrix determinant $|\boldsymbol{\Omega}|^{-1/2}$ in Equation (5), we get

$$\begin{aligned}
&\int_{\mathbb{R}^d}[f(\mathbf{y};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)]^\alpha d\mathbf{y}\\
&=\left(\frac{B_d(\nu)}{|\boldsymbol{\Omega}|^{1/2}}\right)^\alpha \int_{\mathbb{R}^d}\left(1+\frac{\mathbf{z}_0^\top\mathbf{z}_0}{\nu}\right)^{-\alpha\left(\frac{\nu+d}{2}\right)}\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\mathbf{z}_0\sqrt{\frac{\nu+d}{\nu+\mathbf{z}_0^\top\mathbf{z}_0}};\nu+d\right)\right\}^\alpha d\mathbf{z}_0.
\end{aligned}$$

By replacing $\beta=\alpha(\nu+d)-d$ and using the change of variables $\mathbf{u}=\mathbf{z}_0\sqrt{\frac{\beta}{\nu}}$, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d}[f(\mathbf{y};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)]^\alpha d\mathbf{y}\\
&=\left(\frac{B_d(\nu)}{|\boldsymbol{\Omega}|^{1/2}}\right)^\alpha\left(\frac{\nu}{\beta}\right)^{\frac{d}{2}}\int_{\mathbb{R}^d}\left(1+\frac{\mathbf{u}^\top\mathbf{u}}{\beta}\right)^{-\alpha\left(\frac{u+d}{2}\right)}\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\mathbf{u}\sqrt{\frac{\nu+d}{\beta+\mathbf{u}^\top\mathbf{u}}};\nu+d\right)\right\}^\alpha d\mathbf{u}\\
&=\left(\frac{B_d(\nu)}{|\boldsymbol{\Omega}|^{1/2}}\right)^\alpha\left(\frac{\nu}{\beta}\right)^{\frac{d}{2}}\frac{1}{B_d(\nu)}E\left[\left\{2T\left(\bar{\boldsymbol{\eta}}^\top\mathbf{U}\sqrt{\frac{\nu+d}{\beta+\mathbf{U}^\top\mathbf{U}}};\nu+d\right)\right\}^\alpha\right],
\end{aligned}$$

where $\mathbf{U}\sim T_d(\mathbf{0},\boldsymbol{I}_d,\beta)$. The proof is completed by using Lemma 3 of [12]. □

By taking natural logarithm and multiplying by $(1-\alpha)^{-1}$ in both sides of Equation (15), the Lemma 2 yields the final expression of skew-*t* Rényi entropy as is presented in the next Proposition.

Proposition 3. *The Rényi entropy of* $\mathbf{Y}\sim ST_d(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)$ *is*

$$R_\alpha(\mathbf{Y}) = R_\alpha(\mathbf{X}) + \frac{1}{1-\alpha}\log\left(E\left[\left\{2T\left(\frac{\sqrt{\nu+d}\|\bar{\boldsymbol{\eta}}\|W_R}{\sqrt{\alpha(\nu+d)-1+W_R^2}};\nu+d\right)\right\}^\alpha\right]\right), \tag{16}$$

where $R_\alpha(\mathbf{X})$ *is the Student-t Rényi entropy of* $\mathbf{X}\sim T_d(\boldsymbol{\xi},\boldsymbol{\Omega},\nu)$ *given by*

$$R_\alpha(\mathbf{X}) = H(\mathbf{X}) + \frac{1}{1-\alpha}\log[C_{\alpha,d}(\boldsymbol{\Omega},\nu)], \tag{17}$$

$H(\mathbf{X})$ *is the Student-t Shannon entropy of* $\mathbf{X}$ *given in (11),* $W_R\sim T_1(0,1,\beta)$, $\beta=\alpha(\nu+d)-d$, *and* $C_{\alpha,d}(\boldsymbol{\Omega},\nu)$ *is given in Lemma 2.*

It is easy to observe that skew-*t* Shannon entropy is obtained from (16) by taking the limit as α converges to 1.

2.2. Computational Implementation and Numerical Simulations

All numerical computations were made with `R` software [26]. The integrals of skew-*t* Rényi entropies of Equation (16) were evaluated using the `integrate` function of `R` software's QUADPACK routine [27]. This method allows to integrate the Student-*t* cdf of Equation (16) in the interval $(-\infty,\infty)$. This section illustrates the relationship between parameters α, $\boldsymbol{\eta}$, and ν with Rényi entropy for $d=1$, 2, 3, and 4 dimensions. Consider $\mathbf{Y}\sim ST_d(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\eta},\nu)$ with the following cases.

(a) $d=1$, $\xi=0$, $\Omega=1.5$, and $\eta=0.3$.

(b) $d=2$, $\boldsymbol{\xi}=\mathbf{0}$, $\boldsymbol{\Omega}=\begin{pmatrix}0.7 & 0.3\\ 0.3 & 3\end{pmatrix}$, and $\boldsymbol{\eta}=\begin{pmatrix}0.3 & 2\end{pmatrix}^\top$.

(c) $d = 3$, $\boldsymbol{\xi} = \mathbf{0}$, $\boldsymbol{\Omega} = \boldsymbol{I}_3$, and $\boldsymbol{\eta} = \begin{pmatrix} 0.3 & 2 & 0.3 \end{pmatrix}^\top$.

(d) $d = 4$, $\boldsymbol{\xi} = \mathbf{0}$, $\boldsymbol{\Omega} = \boldsymbol{I}_4$, and $\boldsymbol{\eta} = \begin{pmatrix} 0.3 & 2 & 0.3 & -0.5 \end{pmatrix}^\top$.

Figure 1 shows the relationship between Rényi entropy of $\mathbf{Y}$ and $\nu = 3, \ldots, 30$, $\alpha = 2, 3, 4, 5, 6, 8, 10$, and $d = 1, 2, 3, 4$, related to cases (a)–(d). The Rényi entropies converge to a finite value for all α and ν, and increase for increments of d. The dispersion matrix $\boldsymbol{\Omega}$ plays an important role in Rényi entropy, mainly by the increment of matrix dimension related to d. Moreover, the Rényi entropy decreases when α increases, as mentioned in the properties of $R_\alpha(\mathbf{Y})$. The Rényi entropies decrease when ν increases, yielding the respective Student-t Rényi entropies ($\nu \to \infty$).

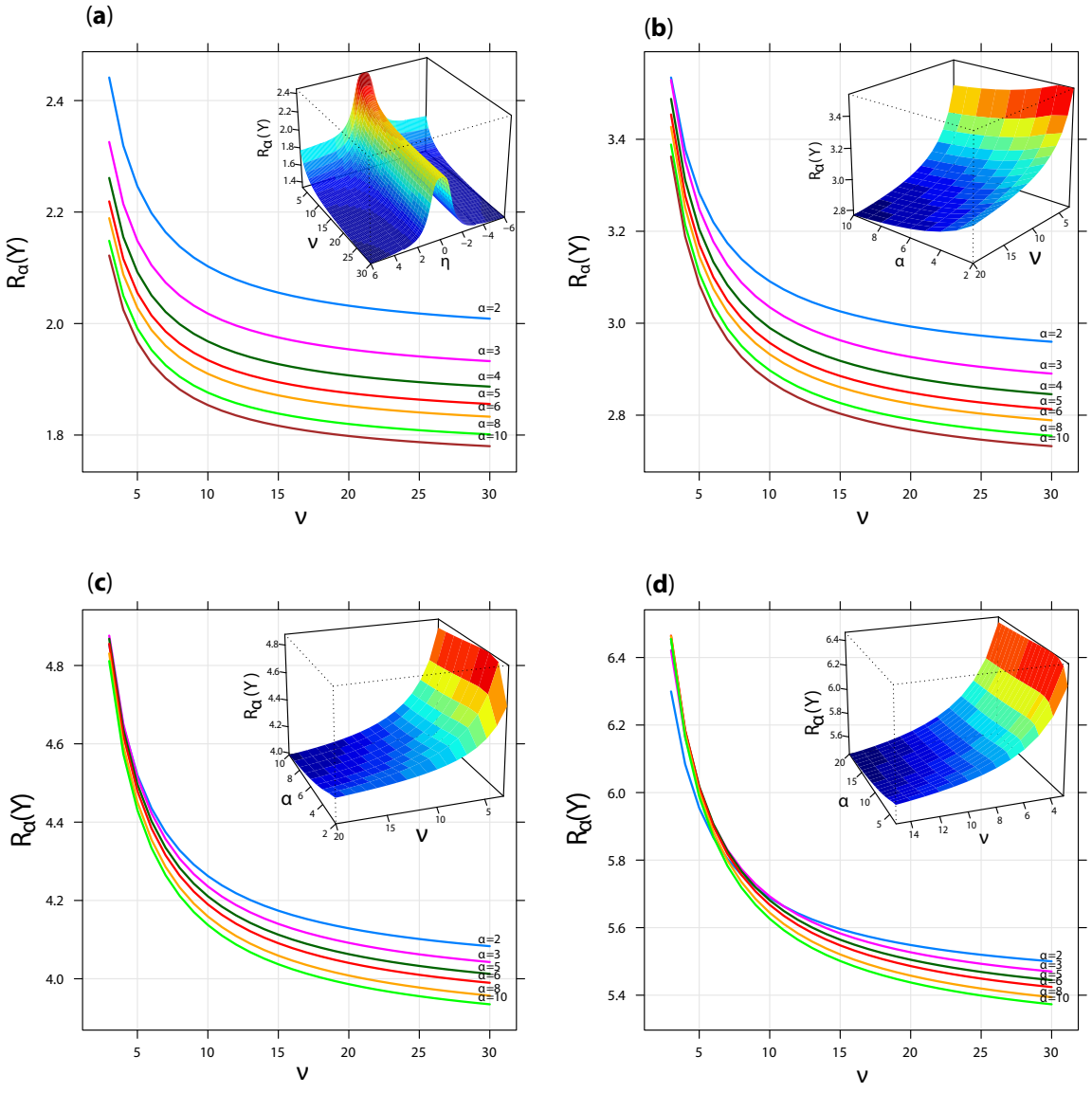


Figure 1. Skew-t Rényi entropy [$R_\alpha(\mathbf{Y})$] versus degree of freedom parameter ($\nu = 3, \ldots, 30$) for cases (**a**–**d**) described above. Each line corresponds to order $\alpha = 2, 3, 4, 5, 6, 8, 10$. 3d-subplot of case (**a**) corresponds to $R_\alpha(Y)$ versus ν and η, whereas the 3d-subplots of cases (**b**–**d**) are $R_\alpha(\mathbf{Y})$ versus ν and α.

Subplot of case (a) considered negative and positive values of η ($-6 \leq \eta \leq 6$), producing a symmetry of $R_\alpha(\mathbf{Y})$ with respect to this parameter. However, as mentioned below, $R_\alpha(\mathbf{Y})$ decreases for increments of ν. Subplots of cases (b)–(d) cannot represent the

behavior of $\boldsymbol{\eta}$ given this parameter's dimension $d = 2, 3, 4$ of. For these cases, order α is considered and the Rényi entropy decreases when α increases.

3. Application to Finite Mixtures of Multivariate Skew-*t* Distributions

Let us consider the definitions of [5,7,28] for FMST distributions. The pdf of an m-component mixture model with parameter vector set $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\Lambda}}, \tilde{\nu})$: $\tilde{\boldsymbol{\xi}} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)$, a set of m location vector parameters $\tilde{\boldsymbol{\Omega}} = (\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_m)$, a set of m dispersion matrices $\tilde{\boldsymbol{\Lambda}} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_m)$, a set of shape vector parameters $\tilde{\nu} = (\nu_1, \ldots, \nu_m)$, a set of degree-of-freedom parameters, and with m mixing weights $\mathbf{p} = (p_1, \ldots, p_m)$ is

$$p(\tilde{\mathbf{y}}; \tilde{\boldsymbol{\theta}}, \boldsymbol{\pi}) = \sum_{j=1}^{m} p_j f(\mathbf{y}_j; \boldsymbol{\theta}_j), \tag{18}$$

where $p_j \geq 0$, $\sum_{j=1}^{m} p_j = 1$, $\tilde{\mathbf{y}} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$, and $f(\mathbf{y}_j; \boldsymbol{\theta}_j)$ are defined as in (5) for a known $\boldsymbol{\theta}_j = (\boldsymbol{\xi}_j, \boldsymbol{\Omega}_j, \boldsymbol{\eta}_j, \nu_j)$, $j = 1, \ldots, m$. If $\tilde{\mathbf{Y}}$ has pdf (18), it is denoted as $\tilde{\mathbf{Y}} \sim FMST_d(\tilde{\boldsymbol{\theta}}, \mathbf{p})$.

Let $\mathbf{S}_i = (S_{i,1}, \ldots, S_{i,m})$ with $\sum_{j=1}^{m} S_{ij} = 1$ be a set of m latent indicators of observations $\tilde{\mathbf{y}}$, $i = 1, \ldots, n$, where j corresponds to a binary index such that $S_{ij} = 1$, if $\mathbf{Y}_i$ is from group j, and $S_{ij} = 0$ otherwise. Then, indicators $\mathbf{S}_1, \ldots, \mathbf{S}_n$ are independent and each one with multinomial distribution given by

$$p(\boldsymbol{s}_i) = p_1^{s_{i,1}} p_2^{s_{i,2}} \cdots \left(1 - \sum_{j=1}^{m-1} p_j\right)^{s_{i,m}}$$

and denoted as $\mathbf{S}_i \sim M(1; p_1, \ldots, p_m)$. Proposition 1 of the work in [28] and (6) allow obtaining a hierarchical representation to each j-th component pdf (see Section 3.1 of [28] for details). This hierarchical representation is useful for parameter estimation based on EM framework for FMST distributions.

The first and second moments of the j-th component $\mathbf{Y}_j$, (7) and (8), respectively, get the first two moments of $\tilde{\mathbf{Y}}$:

$$E[\tilde{\mathbf{Y}}] = \sum_{j=1}^{m} p_j \left(\boldsymbol{\xi}_j + \frac{B_{-1}(\nu)}{\pi} \bar{\boldsymbol{\eta}}_j\right), \tag{19}$$

$$V[\tilde{\mathbf{Y}}] = \sum_{j=1}^{m} p_j \left[\boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top + \left(\frac{B_{-2}(\nu)}{2\pi}\right) \boldsymbol{\Omega}_j + \left(\frac{B_{-1}(\nu)}{\pi}\right)^2 (\boldsymbol{\xi}_j \boldsymbol{\delta}_j^\top + \boldsymbol{\delta}_j \boldsymbol{\xi}_j^\top) + \boldsymbol{\mu}_j \boldsymbol{\mu}_j^\top\right], \tag{20}$$

where

$$\boldsymbol{\mu}_j = \boldsymbol{\xi}_j + \left(\frac{B_{-1}(\nu)}{\pi}\right) \bar{\boldsymbol{\eta}}_j - E[\tilde{\mathbf{Y}}],$$

$\bar{\boldsymbol{\eta}}_j = \boldsymbol{\Omega}_j^{1/2} \boldsymbol{\eta}_j$, and

$$\boldsymbol{\delta}_j = \frac{\boldsymbol{\Omega}_j \boldsymbol{\eta}_j}{\sqrt{1 + \boldsymbol{\eta}_j^\top \boldsymbol{\Omega}_j \boldsymbol{\eta}_j}}, \quad j = 1, \ldots, m;$$

see, e.g., in [2,6].

Dehesa et al. [24] obtained an upper bound of Rényi entropy using a variational approach that expresses the Rényi entropy of a finite mixture random variable in terms of the dispersion matrix. Specifically, the result provided by [29] allows obtaining a lower bound for Rényi entropy of an FMST in terms of each component (see also [6]), as is presented next.

$$\frac{\log[E_\alpha(\tilde{\mathbf{Y}})]}{1 - \alpha} \leq R_\alpha(\tilde{\mathbf{Y}}) \leq \frac{d}{2} \left[\log\left(\frac{V[\tilde{\mathbf{Y}}]}{d}\right) + F_\alpha(d)\right], \tag{21}$$

with

$$E_\alpha(\tilde{\mathbf{Y}}) = e^{(1-\alpha)R_\alpha(\mathbf{Y}_m)} + \sum_{j=1}^{m-1}\left\{\left(\sum_{k=1}^{j} p_k\right)^{\alpha}\left(e^{(1-\alpha)R_\alpha(\mathbf{Y}_j)} - e^{(1-\alpha)R_\alpha(\mathbf{Y}_{j+1})}\right)\right\}$$

and

$$F_\alpha(d) = \begin{cases} \log\left(\frac{\pi(\alpha(2+d)-d)}{\alpha-1}\right) + \frac{2}{d(\alpha-1)}\log\left(\frac{\alpha(2+d)-d}{2\alpha}\right) + \frac{2}{d}\log\left\{\frac{\Gamma\left(\frac{\alpha}{\alpha-1}\right)}{\Gamma\left(\frac{\alpha(2+d)-d}{2(\alpha-1)}\right)}\right\}, & \text{if } \alpha > 1, \\ \log\left(\frac{\pi(\alpha(2+d)-d)}{1-\alpha}\right) + \frac{2\alpha}{d(\alpha-1)}\log\left(\frac{\alpha(2+d)-d}{2\alpha}\right) + \frac{2}{d}\log\left\{\frac{\Gamma\left(\frac{\alpha(2+d)-d}{2(1-\alpha)}\right)}{\Gamma\left(\frac{\alpha}{1-\alpha}\right)}\right\}, & \text{if } \frac{d}{d+2} \le \alpha \le 1, \\ \log(2\pi e), & \text{if } \alpha = 1. \end{cases}$$

On the left side of the inequality (21), the lower bound depends on Rényi entropy of each mixture component. A proof of this lower bound is available in Lemma 1 of [6], obtained from Proposition 1 (B1) of [29]. We can see that the right side of the inequality (21) depends on the dispersion matrix, the shape parameters, the αth order and dimension d. A proof of this upper bound is available in Section 3 of [24]. For the case $\alpha = 1$, $F_\alpha(d)$ is related to Shannon entropy of a multivariate standardized normal random variable.

Swordfish Data Analysis

We considered the dataset used in [6], which corresponds to a sample of 486 and 507 length-weight observations of swordfish males and females, respectively. The swordfish data were sampled in the south Pacific off northern Chile during 2011. The observations were obtained using the sampling program of the Instituto de Fomento Pesquero (IFOP, http://www.ifop.cl/). The dataset includes swordfish from 120 to 257 cm and 110 to 299 cm for males and females, respectively.

Following [6], the length-weight nonlinear function $w(l) = a\, l^b$ is considered to explain the increments in swordfish weight $w(l)$ in terms of l, where a and b are the theoretical weight at $l = 0$ and weight growth rate, respectively [30]. The authors obtained a good fitting of $w(l)$ after considering a log-transformed $w(l)$. Then, a two-column matrix formed by these variables is obtained for the clustering procedure. This means no collinearity problem exists, given the nonlinear relationship of length and weight variables. Therefore, the length-weight data are evaluated with the FMST model for $m = 1, \ldots, m^*$ (where m^* is the maximum age by gender) and $d = 2$. The FMST parameter estimates were computed using the `mixsmsn` `R` software's library [31].

Figure 2a,b shows the upper bounds for Rényi entropies, for $m^* = 9$ and 11 for males and females, respectively [6]. Upper bounds are obtained using the right side of inequality (21), where the parameters of each upper bound were replaced by their respective MLEs as plug-in type estimators [6,12,32]. In theses panels, the upper bounds for Rényi entropies are considered because it is necessary to detect the maximum information for given α and m. Importantly, the values increase when the number of components increases for both genders. Only for males, the Rényi entropies increase until $m = 8$ and then stabilize. For females, there exists a breakpoint at $m = 7$ components; however, the information still increases to $m > 7$. For both genders, we can see that information is maximized for $\alpha = 2$ (quadratic Rényi entropy) and there also exist some differences between α values. As in Section 2.2, the 3d-subplot illustrates similar behavior, where order α is considered and the Rényi entropy decreases when α increases.

Given that maximum information of the system is given at $\alpha = 2$, panel (c) of Figure 2 illustrates the upper and lower Rényi entropies and their average. Lower bounds are obtained using the left side of inequality (21), where the parameters of each lower bound were replaced by their respective MLEs as plug-in type estimators. Females presented the largest averages for each m. For both genders, the averages tend to be similar for all

components m. Panel (d) shows the estimated ν degree of freedom parameters with respect to average Rényi entropy. In general, males presented light-tails ($\nu \gg 30$), whereas females presented heavy-tails ($\nu \leq 20$), except in models with $m = 3$ and $m = 4$ components. For each gender, Table 1 summarizes these results in detail as described next.

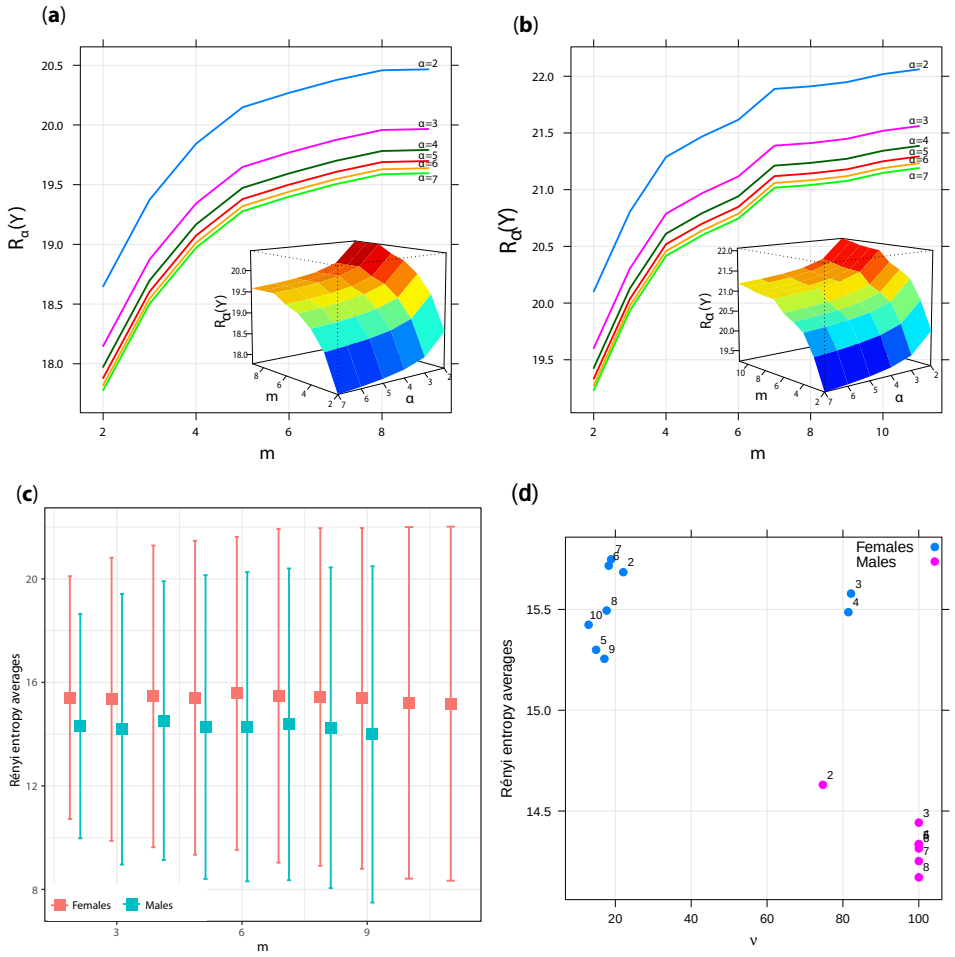


Figure 2. Upper bounds for Rényi entropies based on finite mixture of skew-t (FMST) distribution [$R_\alpha(\mathbf{Y})$] by number of components (m), for (**a**) males and (**b**) females, respectively. 3d-subplots correspond to $R_\alpha(\mathbf{Y})$ versus m and α. Panel (**c**) shows the whisker plots of lower and upper bounds for Rényi entropies based on FMST distribution, and their respective averages by gender and number of components m. Panel (**d**) shows a dispersion plot between $R_\alpha(\mathbf{Y})$ and ν by gender, where the number of each point corresponds to m.

As in [7], the Rényi entropies are compared using the AICand BICcriteria. Inequality (21) is accomplished and the Rényi information increases for parsimonious models (bigger set of m components). The average Rényi entropy rather slowly increases with components $m = 2, \ldots, 5$, but stabilizes at $m = 6$ components for males. A similar phenomenon occurs for females, where the average Rényi entropy is maximum for $m = 7$. Figure 3 illustrates the FMST fits for length-weight by gender. The older swordfish lengths present more variability than younger ones. The estimated parameters of FMST model related to males are

$$\boldsymbol{\pi} = (0.274, 0.320, 0.074, 0.178, 0.130, 0.025),$$

$$\widehat{\boldsymbol{\xi}} = \left(\begin{pmatrix} 167.42 \\ 58.10 \end{pmatrix}, \begin{pmatrix} 155.49 \\ 41.65 \end{pmatrix}, \begin{pmatrix} 197.62 \\ 102.32 \end{pmatrix}, \begin{pmatrix} 179.64 \\ 77.11 \end{pmatrix}, \begin{pmatrix} 138.87 \\ 33.15 \end{pmatrix}, \begin{pmatrix} 213.50 \\ 158.73 \end{pmatrix} \right),$$

$$\widehat{\boldsymbol{\Omega}} = \left(\begin{pmatrix} 9.03 & 4.74 \\ 4.74 & 9.57 \end{pmatrix}, \begin{pmatrix} 8.91 & 3.10 \\ 3.10 & 7.62 \end{pmatrix}, \begin{pmatrix} 11.35 & 7.02 \\ 7.02 & 18.29 \end{pmatrix}, \begin{pmatrix} 10.78 & 5.35 \\ 5.35 & 11.21 \end{pmatrix}, \begin{pmatrix} 7.58 & 3.74 \\ 3.74 & 4.84 \end{pmatrix}, \right.$$

$$\left(\begin{array}{cc} 20.66 & 11.35 \\ 11.35 & 14.00 \end{array} \right) \Bigg),$$

$$\widehat{\boldsymbol{\Lambda}} = \left(\left(\begin{array}{c} 0.53 \\ 0.72 \end{array} \right), \left(\begin{array}{c} -0.92 \\ 1.09 \end{array} \right), \left(\begin{array}{c} 0.94 \\ 1.38 \end{array} \right), \left(\begin{array}{c} 0.79 \\ 0.91 \end{array} \right), \left(\begin{array}{c} -0.90 \\ -0.63 \end{array} \right), \left(\begin{array}{c} 0.72 \\ 0.94 \end{array} \right) \right),$$

and $\widehat{\nu} = 100$.

The estimated parameters of FMST model related to females are

$$\boldsymbol{\pi} = (0.016, 0.295, 0.095, 0.238, 0.045, 0.217, 0.094),$$

$$\widehat{\boldsymbol{\xi}} = \left(\left(\begin{array}{c} 263.87 \\ 291.84 \end{array} \right), \left(\begin{array}{c} 192.68 \\ 82.77 \end{array} \right), \left(\begin{array}{c} 207.20 \\ 119.81 \end{array} \right), \left(\begin{array}{c} 151.80 \\ 41.75 \end{array} \right), \left(\begin{array}{c} 236.13 \\ 211.90 \end{array} \right), \left(\begin{array}{c} 160.97 \\ 51.30 \end{array} \right), \left(\begin{array}{c} 222.78 \\ 156.56 \end{array} \right) \right),$$

$$\widehat{\boldsymbol{\Omega}} = \left(\left(\begin{array}{cc} 16.69 & 13.22 \\ 13.22 & 53.03 \end{array} \right), \left(\begin{array}{cc} 11.53 & 5.17 \\ 5.17 & 13.51 \end{array} \right), \left(\begin{array}{cc} 8.54 & 3.76 \\ 3.76 & 13.97 \end{array} \right), \left(\begin{array}{cc} 11.65 & 6.02 \\ 6.02 & 7.96 \end{array} \right), \left(\begin{array}{cc} 16.35 & 7.45 \\ 7.45 & 22.68 \end{array} \right),$$

$$\left(\begin{array}{cc} 8.96 & 5.35 \\ 5.35 & 9.73 \end{array} \right), \left(\begin{array}{cc} 14.71 & 8.13 \\ 8.13 & 22.89 \end{array} \right) \Bigg),$$

$$\widehat{\boldsymbol{\Lambda}} = \left(\left(\begin{array}{c} -0.06 \\ 0.67 \end{array} \right), \left(\begin{array}{c} -0.83 \\ 0.64 \end{array} \right), \left(\begin{array}{c} 1.04 \\ 1.20 \end{array} \right), \left(\begin{array}{c} -0.79 \\ -0.56 \end{array} \right), \left(\begin{array}{c} 0.97 \\ 1.28 \end{array} \right), \left(\begin{array}{c} 0.84 \\ 0.92 \end{array} \right), \left(\begin{array}{c} 0.75 \\ 1.14 \end{array} \right) \right),$$

and $\widehat{\nu} = 20.849$.

Some differences to the FMSN model considered in [6] appear here, where for females the maximum Rényi entropy based on FMSN distributions is obtained for $m = 7$.

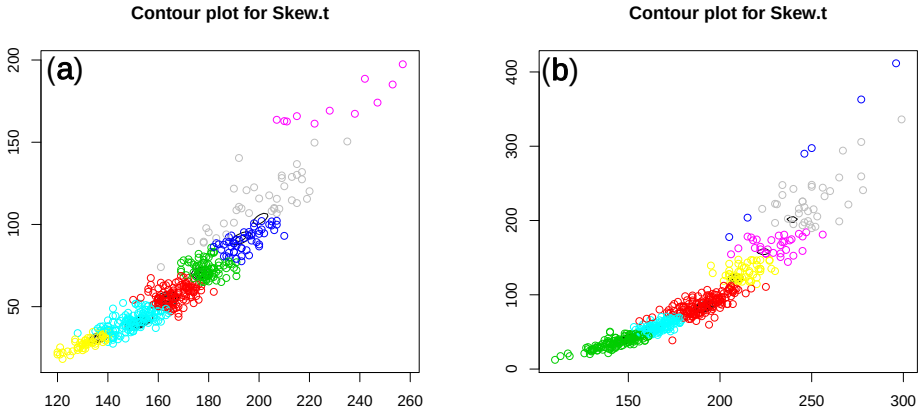


Figure 3. Selected FMST fits for (**a**) males ($m = 6$) and (**b**) females ($m = 7$). Each color is related to each FMST component.

Table 1. Summary of FMST models. Upper and lower bounds of $R_\alpha(\mathbf{Y})$ and their average are computed for $\alpha = 2$. For each FMST model and number of clusters m, the AICand BICcriteria are computed.

Gender	m	$\widehat{\nu}$	Upper $R_\alpha(\mathbf{Y})$	Lower $R_\alpha(\mathbf{Y})$	Average $R_\alpha(\mathbf{Y})$	AIC	BIC
Males	2	26.43	18.65	9.98	14.31	7747.17	7809.96
	3	100	19.39	9.05	14.22	7747.83	7844.11
	4	100	19.81	9.27	14.54	7747.76	7877.53
	5	100	20.07	8.44	14.26	7753.50	7916.77
	6	100	20.32	8.87	14.60	7756.28	7953.03
	7	100	20.43	8.38	14.41	7770.71	8000.95
	8	100	20.43	8.01	14.22	7775.86	8039.59
	9	100	20.52	7.82	14.17	7778.41	8075.64

Table 1. *Cont.*

Gender	m	$\widehat{\nu}$	Upper $R_\alpha(\mathbf{Y})$	Lower $R_\alpha(\mathbf{Y})$	Average $R_\alpha(\mathbf{Y})$	AIC	BIC
Females	2	20.09	20.10	10.72	15.41	8835.20	8898.62
	3	70.62	20.82	9.88	15.35	8842.99	8940.25
	4	85.97	21.30	9.60	15.45	8838.54	8969.62
	5	16.35	21.45	9.53	15.49	8847.64	9012.55
	6	17.68	21.76	9.25	15.51	8846.07	9044.81
	7	15.99	21.87	9.32	15.60	8850.85	9083.42
	8	17.23	21.91	8.82	15.36	8864.95	9131.34
	9	16.83	21.96	8.72	15.34	8874.23	9174.45
	10	20.02	22.04	8.68	15.36	8892.39	9226.44
	11	11.91	22.08	8.41	15.24	8902.36	9270.24

4. Conclusions and Final Remarks

We derive upper and lower bounds on the Rényi entropy of a multivariate skew-t random variable. Then, we extended these tools to the class of finite mixture of multivariate skew-t densities. Considering the average of these bounds, the approximate value of entropy can be calculated. Both entropies converge to finite value of a multivariate skew-t random variable and its mixture model for any values of α order, ν degrees of freedom parameter, and dimension d. Given that FMST Rényi entropies are localized between the upper and lower bounds, the average of these bounds can be used as an approximation of the FMST Rényi entropies. In addition, the FMST Rényi entropy bounds provide useful information about the data and could be considered as a criterion to choose the possible number of components in each gender-based group.

We present an application to 2-dimensional length-weight swordfish data. We compared our results with those obtained in [6]. Given that best results of the work in [6] were obtained using the FMSN Rényi entropies, the Rényi entropy bounds of FMST are compared with the FMSN model, rather than the simplest normal model. As in [6], AIC and BIC values increase when the m increases, where the minimum AIC and BIC values correspond to the simplest model with $m = 2$ components. This fact is related to data set complexity (high-dimensionality) and parsimonious models (large number of parameters). However, of all these models, the FMST model has smaller AIC and BIC values with respect to those obtained under the FMSN one (see Table 3 of [6]). The latter is produced by the presence of heavy-tails in distributions of female swordfish, as the FMST model is more flexible than the FMSN one.

Finally, we encourage researchers to use the proposed approach for real-world applications and data analysis, such as environmental [32] and biological [7,30] data.

Author Contributions: Conceptualization, S.H.A.; Data curation, U.J.Q. and J.E.C.-R.; Formal analysis, U.J.Q. and J.E.C.-R.; Investigation, S.H.A., U.J.Q. and J.E.C.-R.; Methodology, S.H.A. and J.E.C.-R.; Project administration, S.H.A.; Software, U.J.Q. and J.E.C.-R.; Supervision, S.H.A.; Validation, U.J.Q.; Writing—original draft, S.H.A. and U.J.Q.; Writing—review & editing, J.E.C.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FONDECYT (Chile) grant No. 11190116.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: We are grateful to the IFOP (Valparaíso, Chile) for providing access to the swordfish data used in this paper. The authors thank the editor and three anonymous referees for their helpful comments and suggestions. All R codes used in this paper are available upon request from the corresponding author.

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

References

1. McLachlan, G.; Peel, D. *Finite Mixture Models*; John Wiley & Sons: New York, NY, USA, 2000.
2. Carreira-Perpiñán, M.A. Mode-Finding for Mixtures of Gaussian Distributions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1318–1323. [CrossRef]
3. Celeux, G.; Soromenho, G. An entropy criterion for assessing the number of clusters in a mixture model. *J. Classif.* **1996**, *13*, 195–212. [CrossRef]
4. Lee, S.X.; McLachlan, G.J. Finite mixtures of multivariate skew t-distributions: some recent and new results. *Stat. Comput.* **2014**, *24*, 181–202. [CrossRef]
5. 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. [CrossRef] [PubMed]
6. 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. [CrossRef]
7. Contreras-Reyes, J.E.; Quintero, F.O.L.; Yáñez, A. Towards age determination of Southern King crab (*Lithodes Santolla*) off Southern Chile using flexible mixture modeling. *J. Mar. Sci. Eng.* **2018**, 6, 157. [CrossRef]
8. 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.
9. Azzalini, A.; Dalla-Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [CrossRef]
10. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. Roy. Stat. Soc. B* **2003**, *65*, 367–389. [CrossRef]
11. Lin, T.I.; Lee, J.C.; Hsieh, W.J. Robust mixture modeling using the skew t distribution. *Stat. Comput.* **2007**, *17*, 81–92. [CrossRef]
12. 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. [CrossRef]
13. Contreras-Reyes, J.E. Asymptotic form of the Kullback–Leibler divergence for multivariate asymmetric heavy-tailed distributions. *Phys. A* **2014**, *395*, 200–208. [CrossRef]
14. Contreras-Reyes, J.E. Rényi entropy and complexity measure for skew-gaussian distributions and related families. *Phys. A* **2015**, *433*, 84–91. [CrossRef]
15. Abid, S.H.; Quaez, U.J. Rényi Entropy for Mixture Model of Ultivariate Skew Normal-Cauchy distributions. *J. Theor. Appl. Inf. Technol.* **2019**, *97*, 3526–3539.
16. Abid, S.H.; Quaez, U.J. Rényi Entropy for Mixture Model of Multivariate Skew Laplace distributions. *J. Phys. Conf. Ser.* **2020**, *1591*, 012037. [CrossRef]
17. Ferreira, J.A.; Coêlho, H.; Nascimento, A.D. A family of divergence-based classifiers for Polarimetric Synthetic Aperture Radar (PolSAR) imagery vector and matrix features. *Int. J. Remote Sens.* **2021**, *42*, 1201–1229. [CrossRef]
18. Lin, P.E. Some Characterization of the Multivariate t Distribution. *J. Multivar. Anal.* **1972**, *2*, 339–344. [CrossRef]
19. Branco, M.; Dey, D. A general class of multivariate skew-elliptical distribution. *J. Multivar. Anal.* **2001**, *79*, 93–113. [CrossRef]
20. Shannon, C.E. A mathematical theory of communication. *Bell Sys. Tech. J.* **1948**, *27*, 379–423.
21. Rényi, A. On Measures of Entropy and Information. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
22. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley & Son, Inc.: New York, NY, USA, 2006.
23. Guerrero-Cusumano, J.L. A measure of total variability for the multivariate t distribution with applications to finance. *Inf. Sci.* **1996**, *92*, 47–63.
24. Dehesa, J.S.; Gálvez, F.J.; Porras, I. Bounds to density-dependent quantities of D-dimensional many-particle systems in position and momentum spaces: Applications to atomic systems. *Phys. Rev. A* **1989**, *40*, 35. [CrossRef]
25. Azzalini, A.; Regoli, G. Some Properties of Skew-symmetric Distributions. *Ann. Inst. Stat. Math.* **2012**, *64*, 857–879. [CrossRef]
26. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
27. Piessens, R.; deDoncker-Kapenga, E.; Uberhuber, C.; Kahaner, D. *Quadpack: A Subroutine Package for Automatic Integration*; Springer: Berlin, Germany, 1983.
28. Cabral, C.R.B.; Lachos, V.H.; Prates, M.O. Multivariate mixture modeling using skew-normal independent distributions. *Comput. Stat. Data Anal.* **2012**, *56*, 126–142. [CrossRef]
29. Bennett, G. Lower bounds for matrices. *Linear Algebra Appl.* **1986**, *82*, 81–98. [CrossRef]
30. Arellano-Valle, R.B.; Contreras-Reyes, J.E.; Stehlík, M. Generalized skew-normal negentropy and its application to fish condition factor time series. *Entropy* **2017**, *19*, 528. [CrossRef]
31. Prates, M.O.; Lachos, V.H.; Cabral. C. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *J. Stat. Soft.* **2013**, *54*, 1–20. [CrossRef]
32. Contreras-Reyes, J.E.; Maleki, M.; Cortés, D.D. Skew-Reflected-Gompertz information quantifiers with application to sea surface temperature records. *Mathematics* **2019**, *7*, 403. [CrossRef]