Rényi entropy and complexity measure for multivariate skew-normal distributions and related families

Javier E. Contreras-Reyes

Abstract

Recent studies determined that the skew-normal Shannon entropy corresponds to the difference between the common gaussian Shannon entropy and a term that depends on the skewness parameter. This allows to identify the departure from normality of a perturbed distribution. In this paper, we provide the Rényi entropy and complexity measure for a novel, flexible class of multivariate skew-normal distributions and their related families, as a characteristic form of the skew-gaussian Shannon entropy. We give closed expressions considering a more general class of closed skew-normal distributions and the weighted moments estimation method. In addition, closed expressions of Rényi entropy are presented for univariate truncated skew-normal and multivariate extended skew-normal distributions. Finally, additional inequalities for skew-normal and extended skew-normal Rényi and Shannon entropies are reported.

Keywords: skew-normal; Rényi entropy; complexity; weighted moments; Jensen’s inequality

1 Introduction

The family of skew-normal distributions was discussed formally and popularized by Azzalini (1985) and since then it has been discussed extensively in the literature, which includes a wide variety of skewed models in addition to having normal distribution as a special case (Azzalini and Dalla-Valle, 1996; Azzalini and Capitanio, 1999; Azzalini, 2013). In many applications, the probability distribution function of some observed variables can be skewed, such as for example; meteorology (Flecher et al., 2010; Arellano-Valle et al., 2013), seismology (Contreras-Reyes and Arellano-Valle, 2012), and many other applications (Genton, 2004). More recently, a study by Contreras-Reyes and Arellano-Valle (2012) and Arellano-Valle et al. (2013) computes the Kullback-Leibler divergence measure for multivariate skew-normal distribution and Shannon entropy for the full class of multivariate skew-elliptical distributions, respectively. They highlight that Kullback-Leibler information measure should be represented by a quadratic form, including a non-analytical expected value. In addition, they gave the Kullback-Leibler divergence of a multivariate skew-normal distribution with respect to multivariate normal distribution. Information measure applications dealing with skewed data have been performed by Contreras-Reyes and Arellano-Valle (2012), Arellano-Valle et al. (2013) and Contreras-Reyes (2014).

In this work, we focus on the Rényi entropy (Rényi, 1970) as a characteristic form of the Shannon entropy to give a closed expression for multivariate skew-normal (MSN) density and, in LMC complexity measure (López-Ruiz et al., 1995; Anteneodo and Plastino, 1996; Feldman and Crutchfield, 1998) derived by the difference between the nonextensive Rényi entropy and Shannon entropy (Yamano, 2004). Then we study the multivariate extended skew-normal (MESN) case created by generalising the skew-normal distribution and adding a fourth real parameter, \( \tau \) (Capitanio et al., 2003; Azzalini, 2013). This last distribution is flexible enough to accommodate skewness and heavy tails. Finally, we compute the Rényi entropy and complexity measure for univariate truncated skew-normal density (TSN). This class of distribution is a good choice for those applications where observed variables can be simultaneously skewed and restricted to a fixed interval, especially for environmental and biological variables (Flecher et al., 2010; Carpi et al., 2011; Arellano-Valle et al., 2013). Some proofs are presented in Appendix.

1 Corresponding author. Division of Fisheries Research, Fisheries Development Institute, Blanco 839, Valparaiso, Chile. Tel.: +56-032-2151-513. E-mail: jecontrr@mat.puc.cl.
2 Rényi entropy and complexity measure

Let $X$ be a random variable with probability density function (pdf) $f(x)$, $x \in \mathbb{R}^k$. The $\alpha$th-order Rényi entropy (Rényi, 1970) measure of $X$ is given by

$$H_\alpha(X) = \begin{cases} \frac{\log \int f^\alpha(x)dx}{1-\alpha}, & 1 < \alpha < \infty; \\ -\int f(x)\log f(x)dx, & \alpha = 1. \end{cases}$$ (1)

Golshani and Pasha (2010) provide some important properties of Rényi entropy: 1. $H_\alpha(X)$ can be negative, 2. $H_\alpha(X)$ is invariant under a location transformation, 3. $H_\alpha(X)$ is not invariant under a scale transformation, and 4. for any $\alpha_1 < \alpha_2$, we have $H_{\alpha_1}(X) \geq H_{\alpha_2}(X)$ for all $X$, and are equal if and only if $X$ is uniformly distributed.

From (1), the Shannon entropy is obtained by the limit $H_1(X) = \lim_{\alpha \to 1} H_\alpha(X)$ by applying the l’Hôpital’s rule to $H_\alpha(X)$ with respect to $\alpha$ (Rényi, 1970). This measure is the expected value of $g(x) = -\log f(x)$, which satisfies $g(1) = 0$ and $g(0) = \infty$. See Cover and Thomas (2006) for additional properties and related measures. For example, let $X$ be normal distribution $N(\xi, \Omega)$ with mean $\xi$ and covariance matrix $\Omega$, with determinant $|\Omega|$, $|\Omega| > 0$. The Rényi and Shannon entropies of $X$ are given by

$$H_\alpha(X) = \begin{cases} \frac{1}{2}\log \{(2\pi)^k |\Omega| \} + k \log \frac{e}{\alpha}, & 1 < \alpha < \infty; \\ \frac{1}{2}\log \{(2\pi)^k |\Omega| \}, & \alpha = 1, \end{cases}$$ (2)

respectively (Dembo et al., 1991; Cover and Thomas, 2006).

Another important concept is the statistical complexity, that measures the randomness and structural correlations of a known system (Carpi et al., 2011). López-Ruiz et al. (1995) proposed a measure of statistical complexity (LMC) in order to determine the disequilibrium of the system attributed to entropy measure (Anteneodo and Plastino, 1996). LMC measure is defined as

$$C_1(Z) = H_1(Z) e^{-H_2(Z)},$$ (3)

where $H_1(Z)$ and $H_2(Z)$ are the Shannon and quadratic ($\alpha = 2$) Rényi entropies of a random variable $Z$. Yamano (2004) provide a nonextensive entropy instead of an additive Shannon entropy in (3), characterised as a difference between the $\alpha$th-order Rényi entropy and quadratic Rényi entropy as

$$C_\alpha(Z) = e^{H_\alpha(Z) - H_2(Z)}.$$ (4)

Note that $C_\alpha$ reflects the shape of the distribution density function of $Z$ and takes unity for all distributions when $\alpha = 2$. In addition, $C_\alpha$ satisfies a great variety of interesting mathematical and physical properties. Let us just recall here the following properties: 1. $C_\alpha(Z) > 1$, $\forall \alpha \leq 2$, and, $0 < C_\alpha(Z) \leq 1$, $\forall \alpha > 2$; 2. $C_\alpha$ is invariant under a location and scale transformation in the distribution, i.e, $C_\alpha(Z) = C_\alpha(Z_0)$, $Z_0 = a^T (Z - b)$, $a > 0$ and $b \in \mathbb{R}$; and 3. is invariant under replications of the original distribution of $Z$.

3 Main results

3.1 Closed skew-normal distributions

The closed skew-normal distribution (CSN) has interesting properties inherited from the Gaussian distribution. Concerning the definition of González-Fariñas et al. (2004) and Flecher et al. (2009), let $Y \in \mathbb{R}^k$ be a random vector with distribution $CSN_{k,s}(\xi, \Sigma, \nu, \Delta)$ and density function

$$f_{k,s}(y) = C_{s,k} \phi_k(y; \xi, \Sigma) \Phi_s(D^T(y - \xi); \nu, \Delta),$$ (5)

where $C_{s,k}^{-1} = \Phi_s(0; \nu, \Delta + D^T \Sigma D)$, $\xi \in \mathbb{R}^k$, $\nu \in \mathbb{R}^s$, $\Sigma \in \mathbb{R}^{k \times k}$ and $\Delta \in \mathbb{R}^{s \times s}$ are both covariance matrices, $D \in \mathbb{R}^{k \times s}$, $D^T$ denotes the transposed $D$ matrix, $\phi_k(y; \xi, \Sigma)$ and $\Phi_s(y; \xi, \Sigma)$ are the probability function
where $\Sigma_\nu = \nu^T \Sigma \nu$, $D_\nu = D^T \Sigma D$ and $\Delta_{\nu} = \Delta^T D^T \Sigma D$. A particular case of (6), is the standardised variable $z_0 = \Sigma^{-1} (y - \xi)$. In this case, Eq. (5) is rewritten as

$$f_{k,s}(z_0) = \tilde{C}_{s,D} \phi_{s}(z_0) \Phi(s^T z_0; \nu, \Delta),$$

where $\tilde{C}_{s,D} = \phi_{s}(0; \nu) = D^T \Sigma D$, and $\tilde{D} = \Sigma^{1/2}$. Given that the CSN distribution is closed under translations and by property (6), $Z_0 \sim CSN_{k,s}(0, I_k, \tilde{D}^T, \nu, \Delta)$. For the moment generating function of the CSN distribution, see González-Farías et al. (2004). For more details about closed skew-normal distributions, see Flecher et al. (2009) and Genton (2004).

**Proposition 1** (Flecher et al., 2009). Let $Y$ be a $CSN_{k,s}(\xi, \Sigma, D, 0, \Delta)$, $r$ a positive integer and $h(y) = h(y_1, \ldots, y_k)$ be any real valued function such that $E[h(Y)]$ is finite, then

$$E[h(Y) \{ \Phi_k(Y; 0, I_k) \}] = \eta_k(y) \Phi_k(r_{k+s}(0; \nu_+, \Delta_+ + D^T \Sigma D_+),$$

where $\eta_k(y) = C_{s,D}^{-1} E[h(Y)]$, $C_{s,D}^{-1} = \phi_{s}(0; \nu) = D^T \Sigma D$, and $Y_+ \sim CSN_{k,r+s}(\xi, \Sigma, D_+, \nu_+, \Delta_+)$ with $D_+ = (D^T, 0)$, $D_+ = (D^T, 0)$, $\nu_+ = (-\xi, \ldots, -\xi, 0)$, $\Delta_+ = (I, \ldots, I, \Delta)$. Here $E[h(Y)]$ denotes the expected information in $Y$ of the random function $h(Y)$.

### 3.2 Skew-normal distribution

A special case of CSN density is the multivariate normal (MN) density when $D = 0$. When $s = 1$, the MN density function is obtained (Azzalini and Dalla Valle, 1996; Azzalini and Capitanio, 1999; Azzalini, 2013). For simplicity, a slight variant of the original definition is considered here. In this work it is postulated that a random vector $Z \in \mathbb{R}^k$ has a skew-normal distribution with location vector $\xi \in \mathbb{R}^k$, dispersion matrix $\Omega \in \mathbb{R}^{k \times k}$ and shape/skewness parameter $\eta \in \mathbb{R}^k$, denoted by $Z \sim SN_k(\xi, \Omega, \eta)$, if its probability density function is

$$f_Z(z) = 2\phi_k(z; \xi, \Omega)\Phi(\eta^T (z - \xi)), \quad z \in \mathbb{R}^k,$$

where $\phi_k(z; \xi, \Omega)$ is the probability density function of the $k$-variate $N_k(\xi, \Omega)$ distribution, and $\Phi$ is the univariate $N(0, 1)$ cumulative distribution function.

The mean vector and the covariance matrix (Azzalini and Capitanio, 1999; Contreras-Reyes and Arellano-Valle, 2012) are $E[Z] = \xi + \sqrt{\frac{\eta^T \delta}{\sigma^2}}$ and $\text{Var}[Z] = \Omega - \frac{\eta^T \delta \delta^T}{\sigma^2}$, respectively; where $\delta = \Omega \eta / \sqrt{1 + \eta^T \Omega \eta}$.

**Proposition 2** Let $Z$ be a $SN_k(\xi, \Omega, \eta)$ with $z \in \mathbb{R}^k$. Then:

$$\int f^n(z) dz = \psi_{\alpha,k}(\Omega) \frac{\Phi_{\alpha+1}(0; 0, \Omega_+)}{\Phi_1(0; 0, \sigma^2)} \alpha \in \mathbb{N}, \alpha > 1,$$

where $\psi_{\alpha,k}(\Omega) = 2^\alpha \alpha^{-k/2} \left(\frac{(2\pi)^k |\Omega|^{1-\alpha/2}}{(1-\alpha)^2} \right)$, $\Omega_+ = I_{\alpha+1} + \|\Omega\|^{1/2} \Omega_1^T \Omega_1$, $\Omega_1 = (I_1, \|\Omega\|^{1/2} \eta) \eta$, $\eta = \alpha^{1/2} \Omega^{1/2} \eta$. $\eta = \alpha^{-1/2} \Omega^{-1/2} \eta$.

By (1) and (10), the Rényi entropy of a random variable $Z \sim SN_k(\xi, \Omega, \eta)$ is retrieved. Taking $\eta = 0$ in (10), the Rényi entropy of the normal distribution given by (2) is obtained. Proposition 1 allows the computing of the expected value of the cdf of a multivariate standardised normal density. Considering the standarised CSN variable in (7), the Proposition 2 is solved by (8), by setting $\nu = 0$ and $\Delta = I_k$, with
\( k = s = 1 \). However, the case \( \boldsymbol{\nu} \neq \mathbf{0} \) and \( \Delta \neq I_k, k > 1 \), is still an open problem and, it is useful to find the Rényi entropy for CSN distributions. By (1) and (7), the Shannon entropy for CSN distributions is rewritten as

\[
\mathcal{H}_1(Y) = -\mathbb{E}[\log \{ f_{k,s}(Y) \}] = \frac{1}{2} \log |\Sigma| - \log \bar{C}_{s,\nu} - \mathbb{E}[\log \{ \phi_k(Z_0)\Phi(D^\top Z_0; \nu, \Delta) \}]
\]

where \( Z_N \sim N(\mathbf{0}, I_k) \) and \( \mathcal{H}_1(Z_N) = (1/2)\log(2\pi e) \).

**Corollary 1** Let \( Z \sim SN_k(\xi, \Omega, \eta) \) and \( Z_N \sim N_k(\xi, \Omega) \). Then,

(i) \( \mathcal{H}_\alpha(Z) = \mathcal{H}_\alpha(Z_N) - N_\alpha(Z), \quad \alpha \in \mathbb{N}, \, \alpha > 1 \),

where \( N_\alpha(Z) = (\alpha - 1)^{-1} \log \left\{ 2^\alpha \Phi_{\alpha+1}(\mathbf{0}; 0, \Omega_+) / \Phi(0; 0, \sigma^2) \right\} \) is the so-called Negentropy, \( \mathcal{H}_\alpha(Z_N) \) is given by (2), and \( \Omega_+ \) and \( \sigma^2 \) are defined as in Proposition 2.

(ii) \( \lim_{\alpha \to 1} N_\alpha(Z) = \mathbb{E}[\log \{ 2\Phi(||\eta||W) \}] \).

(iii) \( \mathcal{H}_1(Z) = \mathcal{H}_1(Z_N) - \mathbb{E}[\log \{ 2\Phi(||\eta||W) \}] \),

where \( \mathcal{H}_1(Z_N) \) is given by (2), \( W \sim SN_1(0, 1, ||\eta||) \) with \( ||\eta|| = \bar{\eta}^\top \eta \) and \( \bar{\eta} = \Omega^{1/2} \eta \).

(iv) \( \mathcal{H}_1(Z_N) - \log(4e) \leq \mathcal{H}_1(Z) \leq \mathcal{H}_1(Z_N), \forall \eta \).

Contreras-Reyes and Arellano-Vall (2012) defines the negentropy as the departure from normality of the distribution of \( Z \). Therefore, the skew-normal Rényi entropy corresponds to the difference between normal Rényi entropy and negentropy, that depends on the skewness parameter \( \eta \). On the other hand, by setting \( \boldsymbol{\nu} = \mathbf{0} \) and \( \Delta = I_k \) in (11) with \( k = s = 1 \), we obtain the property (ii) of Corollary 1. By property (iii), \( \mathcal{H}_1(Z) \geq \mathcal{H}_1(Z_N) - \log(4e) \geq -0.967 \) because, the minimum value of normal Shannon entropy is obtained for \( k = 1 \). In addition, from property (ii) and (iii), \( 0 \leq \mathbb{E}[\log \{ 2\Phi(||\eta||W) \}] \leq \log(4e) \approx 2.386, \forall \eta \). In Contreras-Reyes and Arellano-Vall (2012) is reported a maximum value of this expected value equal to 2.339, using numerical approximations. Considering (1), (4) and (10), the complexity measure for skew-normal distribution is obtained.

### 3.3 Extended skew-normal distributions

Consider a slight variant of the multivariate extended skew-normal (ESN) distribution proposed by Capitanio et al. (2003). Let \( Z \sim ESN_k(\xi, \Omega, \eta, \tau) \), with location vector \( \xi \in \mathbb{R}^k \), positive definite dispersion matrix \( \Omega \in \mathbb{R}^{k \times k} \), shape/skewness parameter \( \eta \in \mathbb{R}^k \), extended parameter \( \tau \in \mathbb{R} \), and with pdf given by:

\[
p(z) = \frac{1}{\Phi(\bar{\tau})} \phi_k(z; \xi, \Omega) \Phi \left( \eta^\top (z - \xi) + \bar{\tau} \right),
\]

where \( z \in \mathbb{R}^k \) and \( \bar{\tau} = \tau \sqrt{1 + \eta^\top \Omega \eta} \). Here \( \phi_k(z; \xi, \Omega) \) is the probability density function of \( N_k(\xi, \Omega) \), the \( k \)-variate distribution, and \( \Phi \) is the univariate \( N_1(0, 1) \) cumulative distribution function. From Arellano-Vall and Azzalini (2006), a stochastic representation of the \( ESN_k \) distribution is

\[
Z \overset{d}{=} W + \delta U,
\]

where \( \delta = \Omega \eta / \sqrt{1 + \eta^\top \Omega \eta}, \quad U \sim LTN_{(-\tau, \infty)}(0, 1) \), which is independent of \( W \sim N_k(\xi, \Sigma), \quad \Sigma = \Omega - \delta \delta^\top \), where \( LTN_{(-\tau, \infty)}(0, 1) \) represents the unit normal distribution truncated below the point \(-\tau\) and \( " \overset{d}{=} " \) denotes equality in terms of distribution. From the stochastic representation (13) it follows that \( Z \overset{d}{=} W + \delta W_\tau, \)
where \( W_\tau \overset{d}{=} (W_0 \mid W_0 + \tau > 0) \) and
\[
\left( \begin{array}{c} W_0 \\ W \end{array} \right) \sim N_{1+k} \left( \left( \begin{array}{c} 0 \\ \xi \end{array} \right), \left( \begin{array}{cc} 1 & 0^T \\ 0 & \Sigma \end{array} \right) \right).
\]

Note that \( W_0 \) and \( W \) are independent. This fact must be interpreted as the non-normality effect of \( W_0 \) on \( Z \), where \( W_0 \) is the so-called unobserved confounder variable. From (13), the first and second moments are
\[
E[Z] = \xi + \delta \zeta_1(\tau) \quad \text{and} \quad \text{Var}[Z] = \Omega - \zeta_1(\tau)\{\tau + \zeta_1(\tau)\} \delta \delta^T,
\]
respectively; where \( \zeta_1(z) = \phi(z)/\Phi(z) \) is the zeta function (Azzalini and Capitanio, 1999; Capitanio et al., 2003).

**Proposition 3** Let \( Z \) be a \( \text{ESN}_k(\xi, \Omega, \eta, \tau) \), \( z \in \mathbb{R}^k \). Then:
\[
\int f^\alpha(z)dz = \frac{\psi_{\alpha,k}(\Omega)}{2^\alpha \Phi^\alpha(\tau)} E[\Phi^\alpha(W)], \quad \alpha \in \mathbb{N}, \alpha > 1,
\]
where \( \psi_{\alpha,k}(\Omega) \) is defined as in Proposition 2 and \( W = \bar{\eta}^T Z_0 + \tilde{\tau} \sim \text{ESN}_1(\bar{\tau}, \|\bar{\eta}\|, \|\bar{\eta}\|, \tau) \) with \( \|\bar{\eta}\| = \bar{\eta}^T \bar{\eta} \).

**Corollary 2** Let \( Z \sim \text{ESN}_k(\xi, \Omega, \eta, \tau) \), \( Z_N \sim N_k(\xi, \Omega) \) and \( W \) are defined as in Proposition 3. Then,

(i) \( \mathcal{H}_\alpha(Z) = \mathcal{H}_\alpha(Z_N) - N_\alpha(Z) \), \( \alpha \in \mathbb{N}, \alpha > 1 \),
where \( N_\alpha(Z) = \log E \left[ \left( \frac{\Phi(W)/\Phi(\tau)}{\alpha - 1} \right) \right] \), and \( \mathcal{H}_\alpha(Z_N) \) is given by (2).

(ii) \( \mathcal{H}_\alpha(Z) \leq \mathcal{H}_\alpha(Z_N) + \frac{\alpha}{1 - \alpha} \log \left( \frac{\Phi(\bar{\tau} + \delta \zeta_1(\tau))}{\Phi(\bar{\tau})} \right) \),
where \( \delta = \|\bar{\eta}\|^3/\sqrt{1 + \|\bar{\eta}\|^2} \).

(iii) \( \mathcal{H}_1(Z) = \mathcal{H}_1(Z_N) - E \left[ \log \left( \frac{\Phi(W)}{\Phi(\tau)} \right) \right] \).

(iv) \( \mathcal{H}_1(Z_N) + \log \Phi(\tau) - \Phi \left( \frac{\tilde{\tau}}{\sqrt{1 + \eta_\tau}} \right) \leq \mathcal{H}_1(Z) \leq \frac{1}{2} \log \left\{ (2\pi e)^k \left| (\Omega - \zeta_1(\tau)\{\tau + \zeta_1(\tau)\}) \delta \delta^T \right| \right\}, \forall \eta.

(v) \( \lim_{\alpha \to 1} N_\alpha(Z) = E \left[ \log \left( \frac{\Phi(W)}{\Phi(\tau)} \right) \right] \).

In Pourahmadi (2007) the behaviour of \( \zeta_1(\tau) \) is illustrated, \( \tau \in \mathbb{R} \). This function is strictly decreasing for any \( \tau \in \mathbb{R} \), tends to 0 when \( \tau \to +\infty \), and diverge when \( \tau \to -\infty \). For \( \tau = 0 \), the property (iv) of Corollary 2 becomes property (iii) of Corollary 1. By properties (iii) of Corollary 2 and (ii) of Corollary 1, the negentropy of a MESN random variable is always larger than the negentropy of a MSN random variable. Therefore, we obtain the following relationship among the Shannon entropies of MN \( (Z_N) \), MSN \( (Y) \), and MESN \( (Z) \) distributions: \( \mathcal{H}_1(Z_N) \geq \mathcal{H}_1(Y) \geq \mathcal{H}_1(Z) \). Considering (1), (4) and (15); the complexity measure for extended skew-normal distribution is obtained.

### 3.4 Truncated skew-normal distributions

The truncated skew-normal pdf given by Flecher et al. (2010), consider the variable \( Z \sim SN_1(\mu, \omega, \lambda) \) and the definition given in (9) for the case \( k = 1 \). Flecher et al. (2010) gives the expressions of the \( m \)-order moments and some weighted moments of truncated skew-normal distributions. We also consider the following definition based on (9) for a truncated skew-normal distribution random variable \( W \sim TSN(\mu, \omega, \lambda) \) with pdf
\[
f_w(w) = \frac{f_Z(w)}{|F_Z(w)|_a^b} I_{[a,b]}, \quad \text{with} \ -\infty \leq a < w \leq b \leq +\infty;
\]
where \( f_Z(z) \) is defined in (9) for \( k = 1 \) with \( \xi = \mu, \ \Omega = \omega, \ \eta = \lambda; \) \( F_Z(z) \) is the cdf of \( Z \) and \( [F_Z(w)]_a^b = F_Z(b) - F_Z(a). \) The following Remark allows the computation \( [F_Z(w)]_a^b \) in terms of the normal cdf and a bivariate integral term.

**Remark 1** Let \( Z \sim SN_1(\mu, \omega, \lambda), \) Azzalini (1985) and Genton (2004) gives the expression to compute the cdf of a skew-normal distribution \( Z \sim SN_1(\mu, \omega, \lambda) \) as follow

\[
F_Z(z) = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda z} \phi(s)\phi(t) \, dt \, ds. 
\]

(17)

The calculation of (17) can be obtained through the function \( T(z; \lambda) = \int_{z}^{\infty} \int_{0}^{\lambda z} \phi(s)\phi(t) \, dt \, ds \) (Owen, 1956) as

\[
F_Z(z) = \Phi(z) - 2T(z; \lambda). 
\]

(18)

Then, by replacing (18) in \([F_Z(w)]_a^b\) we obtain

\[
[F_Z(w)]_a^b = \Phi(b) - \Phi(a) - 2 \int_{a}^{b} \int_{0}^{\lambda s} \phi(s)\phi(t) \, dt \, ds. 
\]

Note in (18) that the behaviour of the left and right tails of a skew-normal distribution related to \( X \sim SN_1(\mu, \omega, \lambda) \) is \( T(a; \lambda) = T(b; \lambda) = 0 \), when \( a \to -\infty \) and \( b \to +\infty \), for any \( \lambda \) and; \( T(\mu; 0) = T(b; 0) = 0 \) for any truncations \( a \) and \( b \) parameters in the left and right tails, respectively (Owen, 1956).

**Proposition 4** Let \( Z, W \in \mathbb{R} \) be a \( SN_1(\mu, \omega, \lambda) \) and \( TSN_1(\mu, \omega, \lambda) \), respectively, \( \lambda \neq 0 \). Then:

\[
\int_{a}^{b} f_W(w) \, dw = 2\psi_{\alpha,1}(\omega) \Phi_{\alpha+1}(0; 0, \Omega_+) \frac{[F_V(w)]_{a_0}^{b_0}}{([F_Z(z)]_a^b)^{\alpha}},
\]

(19)

where \( \psi_{\alpha,1}(\omega) \) is defined as in Proposition 2 with \( k = 1, \ \Omega = \omega, \ \Omega_+ = I_{\alpha+1} + \bar{\lambda}^2D_+^\top D_+ \), \( \bar{\lambda}^2 = \omega \lambda^2/\alpha, \)

\( D_+ = (I_{\alpha}, \bar{\lambda})^\top \) and \( V \sim CSN_{1,2}((0, \lambda^2, B_+), 0, I_2) \) with \( B_+ = (1, \lambda)^\top, \) \( a_0 = \lambda(\alpha - \mu)/\omega \) and \( b_0 = \lambda(\beta - \mu)/\omega. \)

By Lemma 2.2.1 of Genton (2004), \( F_V(v) \) is easily computable by a tri-variate normal cdf as

\[
F_V(v) = C_2 \Phi_3 \left[ \begin{array}{c} v \\ 0 \\ 0 \end{array} \right]; \begin{pmatrix} 0 & -\bar{\lambda}^2B_+ \\ \bar{\lambda}^2B_+^\top & I_2 + \bar{\lambda}^2B_+^\top B_+ \end{pmatrix}. 
\]

(19)

Considering (1), (4) and (19): the complexity measure for extended skew-normal distribution is obtained.

### 4 Conclusions

In this Letter, we have presented some solutions to compute the Rényi entropy with discrete \( \alpha \)-order and for a wide range of multivariate asymmetric distributions. Specifically, we find a closed expression for skew-normal, extended skew-normal, and truncated skew-normal distributions. Finally, additional inequalities for skew-normal and extended skew-normal entropies were reported.

**Appendix**

**Proof of Proposition 2.**

To compute the integral \( \int f^\alpha(z) \, dz \), we use the change of variables \( \Omega_\alpha = \alpha^{-1} \Omega \) and \( Z_0 = \Omega_\alpha^{-1/2}(Z - \xi), \)

\( Z_0 \sim SN_k(0, I_k, \eta), \) \( \eta = \Omega_\alpha^{1/2} \eta. \) We shall use the fact that \( |\Omega_\alpha| = \alpha^{-k}|\Omega| \) for \( k \)-dimensional matrices. Then, by Lemma 2 of Arellano-Valle et al. (2013), the integral \( \int f^\alpha(z) \, dz \) should be rewritten in terms of an expected value with respect to a multivariate standardized normal density as

\[
\int f^\alpha(z) \, dz = 2^\alpha |\Omega|^{-\alpha/2} |\Omega_\alpha|^{1/2} (2\pi)^{(1-\alpha)k/2} E[f^\alpha(\eta^\top Z_0)] 
\]
\[
W \sim SN_1(0, \|\eta\|^2, \|\eta\|) \quad \text{with} \quad \|\eta\| = \eta^\top \eta \quad \text{(Contreras-Reyes and Arellano-Valle, 2012; Arellano-Valle et al., 2013), i.e., the expected value} \quad \text{E}[\Phi^\alpha(\eta^\top Z_0)] \quad \text{is reduced from} \quad k \quad \text{dimensions to one dimension} \quad \text{(Arellano-Valle et al., 2013; Contreras-Reyes, 2014). By Proposition 1 and setting} \quad \xi = 0, \Sigma = \|\eta\|^2, \quad \text{D} = \|\eta\|, \quad r = \alpha, \quad \Delta = s = h(w) = 1; \quad \text{we obtain} \quad \Delta_+ = I_{\alpha+1} \quad \text{and} \quad D_+ = (1_\alpha, \|\eta\|)^\top. \quad \text{Therefore, the expected value of the integral is reduced to}
\]

\[
E[\Phi^\alpha(w)] = \frac{\Phi_{\alpha+1}(0;0, I_{\alpha+1} + \|\eta\|^2 D_+ D_+)}{\Phi_1(0;0,1 + \|\eta\|^2)}.
\]

**Proof of Corollary 1**

(i) Follows from (2) and Proposition 2.

(ii) See Proposition 2 of Arellano-Valle et al. (2013).

(iii) Right side: see Contreras-Reyes and Arellano-Valle (2012). Left side: consider the nonsymmetrical entropy of Liu (2009) given by \( S(u) = -\int f(u) \log \{\beta\(u\) f(u)\} du \), where \( f(u) \) is the probability density function of a normal variable \( u \). By choosing \( \beta(u) = 2\Phi(\eta^\top \Omega^{-1/2}(u - \xi)), \quad u = Z_N \), it follows that \( E[\log \beta(Z_N)] = \log 2 + \Phi(0) = (1/2) \log(4e) \) (see Proposition 4 of Azzalini and Dalla-Valle, 1996). Then, as \( E[\log \beta(Z)] \leq 2E[\log \beta(Z_N)] \), the result is obtained.

(iv) Follows from properties (i), (ii) and (1). \( \square \)

**Proof of Proposition 3**

By (12), \( \phi_k(y; \xi, \Omega) = 1_{\Omega}^{-1/2} \phi_k \left( \Omega^{-1/2}(y - \xi) \right) \), where \( \phi_k(z) \) is the probability density function of \( N_k(0, I_k) \). Then, as in (2), to compute the integral \( \int f^\alpha(z) dz \) we use the change of variables \( \Omega_\alpha = \alpha^{-1} \Omega \) and \( Z_0 = \Omega_\alpha^{-1/2}(Z - \xi) \). In this case, \( Z_0 \sim ESN_k(0, I_k, \eta, \tau) \) with \( \eta = \Omega_\alpha^{1/2} \eta \). We shall use the fact that \( |\Omega_\alpha| = \alpha^{-k} |\Omega| \) for \( k \)-dimensional matrices. Then, by Lemma 2 of Arellano-Valle et al. (2013), the integral \( \int f^\alpha(z) dz \) should be rewritten in terms of an expected value with respect to a multivariate standardized normal density as

\[
\int f^\alpha(z)dz = \frac{1}{\Phi^\alpha(\tau)} |\Omega|^{-1/2} \|\Omega_\alpha\|^{1/2} (2\pi)^{(1-\alpha)/2} E[\Phi^\alpha(|\eta^\top Z_0 + \tau|)]
\]

\[
= \frac{1}{\Phi^\alpha(\tau)} \alpha^{-k} (2\pi)^{(1-\alpha)/2} |\Omega|^{(1-\alpha)/2} E[\Phi^\alpha(W)].
\]

where \( W = \eta^\top Z_0 + \tau \sim ESN_1(\tau, \|\eta\|^2, \|\eta\|, \tau) \) with \( \|\eta\| = \eta^\top \eta \) (Contreras-Reyes and Arellano-Valle, 2012; Arellano-Valle et al., 2013), i.e., the expected value \( E[\Phi^\alpha(\eta^\top Z_0 + \tau)] \) is reduced from \( k \) dimensions to one dimension (Arellano-Valle et al., 2013; Contreras-Reyes, 2014). \( \square \)

**Proof of Corollary 2**

(i) From Proposition 3, we obtain directly

\[
H_\alpha(Z) = \frac{1}{1 - \alpha} \left( \psi_{\alpha,k}(\Omega) - \alpha \log \{ 2\Phi(\tau) \} + \log \{ E[\Phi^\alpha(W)] \} \right),
\]

\[
= H_\alpha(Z_N) + \frac{\alpha}{1 - \alpha} \log \left\{ \frac{1}{\Phi(\tau)} \right\} + \frac{1}{1 - \alpha} \log \{ E[\Phi^\alpha(W)] \}.
\]

(ii) Considering the Jensen’s inequality, we obtain \( E[\Phi^\alpha(W)] \geq \Phi^\alpha(E[\Phi^\alpha(W)]) \). Then, (ii) is straightforward from (14).

(iii) By (1), it follows that

\[
H_1(Z) = -\mathbb{E} \left\{ \log \left\{ \phi_k(Z_0) \frac{\Phi(\eta^\top Z_0 + \tau)}{\Phi(\tau)} \right\} \right\}
\]
\[ = H_1(Z_N) - E \left[ \log \left( \frac{\Phi(W)}{\Phi(\tau)} \right) \right], \]

where, as in Proposition 3, \( Z_0 = \Omega^{-1/2}(Z - \xi) \sim ESN_k(0, I_k, \eta, \tau) \) and \( W = \eta^\top Z_0 + \tilde{\tau} \sim ESN_1(\tilde{\tau}, ||\eta||^2, ||\eta||, \tau). \)

(iv) Right side: by Cover and Thomas (2006), for any density \( f(x) \) of a random vector \( X \in \mathbb{R}^k \) (not necessary normal) with zero mean and variance \( \Sigma = E[XX^\top] \), the Shannon entropy of \( X \) is maximized under normality as \( H_1(X) \leq (1/2) \log \{|(2\pi e)^k|\Sigma|\} \). Then, the result is obtained from (14). Left side: as in Corollary 1 (iii), by choosing \( \beta(u) = \Phi(\eta^\top \Omega^{-1/2}(u - \xi) + \tilde{\tau})/\Phi(\tau) \) in the nonsymmetrical entropy of Liu (2009) given by \( S(\beta(u)f(u)) \) \( d\mathbf{u} \), it follows that \( E[\log \beta(Z_N)] = -\log \Phi(\tau) + \Phi(\tilde{\tau}/\sqrt{1 + \eta^\top \eta}) \) (see Proposition 4 of Azzalini and Dalla-Valle, 1996). Then, as \( E[\log \beta(Z)] \leq E[\log \beta(Z_N)]/\Phi(\tau) \), the result is obtained.

(v) Follows from properties (i), (iii) and (1). \( \square \)

**Proof of Proposition 4**

By (16), \( f_a^b f_W^b(w)dw = ([F_Z(z)]_a^b - \alpha) f_a^b f_Z^b(w)dw \) and, by Proposition 2, the integral \( \int f_Z^b(w)dw \) should be rewritten in terms of an expected value as

\[ \int_a^b f_Z^b(w)dw = \psi_{\alpha,1}(\omega)E[\Phi^\alpha(u)\mid a_0 < u \leq b_0], \]

where \( U \sim SN_1(0, \lambda^2, \lambda), \lambda^2 = \omega^2/\alpha, a_0 = \lambda(\omega - \mu)/\omega \) and \( b_0 = \lambda(\omega - \mu)/\omega \). Again, by Proposition 1 and setting \( \xi = 0, \Sigma = \lambda^2, r = \alpha, k = s = \Delta = h(u) = 1; \) we obtain \( \Delta_+ = I_{a+1}, D_+ = (1_\alpha, \lambda)^\top \) and \( \Omega_+ = I_{a+1} + \lambda^2 D_+^\top D_+ \). Then, the expected value is

\[ E[\Phi^\alpha(u)\mid a_0 < u \leq b_0] = 2\Phi_{a+1}(0, 0, \Omega_+)[F_V(v)]_a^b, \]

where \( F_V(v) \) is the cdf of a closed skew-normal variable \( V \sim CSN_{1,2}(0, \lambda^2, B_+, 0, I_2) \) with \( B_+ = (1, \lambda)^\top \) (see Proposition 3 of Flecher et al., 2010). \( \square \)

**References**

Anteneodo, C., Plastino, A.R. (1996). Some features of the López-Ruiz-Mancini-Calbet (LMC) statistical measure of complexity. *Phys. Lett. A* 223, 348-354.

Arellano-Valle, R.B., Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Stat.* 33, 561-574.

Arellano-Valle, R.B., Contreras-Reyes, J.E., Genton, M.G. (2013). Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scand. J. Stat.* 40, 42-62.

Azzalini, A. (1985). A Class of Distributions which includes the Normal Ones. *Scand. J. Stat.* 12, 171-178.

Azzalini, A. (2013). The Skew-Normal and Related Families. Vol. 3, Cambridge University Press.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *J. Roy. Stat. Soc. Ser. B* 61, 579-602.

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

Capitanio, A., Azzalini, A., Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scand. J. Stat.* 30, 129-144.

Carpi, L.C., Rosso, O.A., Saco, P.M., Ravetti, M.G. (2011). Analyzing complex networks evolution through Information Theory quantifiers. *Phys. Lett. A* 375, 801-804.
Contreras-Reyes, J.E. (2014). Asymptotic form of the Kullback-Leibler divergence for multivariate asymmetric heavy-tailed distributions. *Phys. A* 395, 200-208.

Contreras-Reyes, J.E., Arellano-Valle, R.B. (2012). Kullback-Leibler divergence measure for Multivariate Skew-Normal Distributions. *Entropy* 14, 1606-1626.

Cover, T.M., Thomas, J.A. (2006). Elements of information theory. 2nd edition. Wiley & Son, Inc., New York.

Dembo, A., Cover, T.M., Thomas, J.A. (1991). Information Theoretic Inequalities. *IEEE Trans. Inform. Theory* 37, 1501-1518.

Feldman, D.P., Crutchfield, J.P. (1998). Measures of statistical complexity: Why? *Phys. Lett. A* 238, 244-252.

Flecher, C., Naveau, P., Allard, D. (2009). Estimating the Closed Skew-Normal distributions parameters using weighted moments. *Stat. Prob. Lett.* 79, 1977-1984.

Flecher, C., Allard, D., Naveau, P. (2010). Truncated skew-normal distributions: moments, estimation by weighted moments and application to climatic data. *Metron* 68, 265-279.

Genton, M.G. (2004). Skew-elliptical distributions and their applications: A journey beyond normality. Edited Volume, Chapman & Hall/CRC, Boca Raton, FL, 416 pp.

Golshani, L., Pasha, E. (2010). Rényi entropy rate for Gaussian processes. *Inform. Sci.* 180, 1486-1491.

González-Farías, G., Domínguez-Molina, J., Gupta, A. (2004). Additive properties of skew normal random vectors. *J. Stat. Plann. Inference* 126, 521-534.

Liu, C.-S. (2009). Nonsymmetric entropy and maximum nonsymmetric entropy principle. *Chaos Soliton. Fract.* 40, 2469-2474.

López-Ruiz, R., Mancini, H.L., Calbet, X. (1995). A statistical measure of complexity. *Phys. Lett. A* 209, 321-326.

Owen, D.B. (1956). Tables for computing bivariate normal probabilities. *Ann. Math. Stat.* 27, 1075-1090.

Pourahmadi, M. (2007). Skew-Normal ARMA Models with Nonlinear Heteroscedastic Predictors. *Commun. Stat. A-Theor.* 36, 1803-1819.

Rényi, A. (1970). Probability theory. North-Holland, Amsterdam.

Yamano, T. (2004). A statistical measure of complexity with nonextensive entropy. *Phys. A* 340, 131-137.