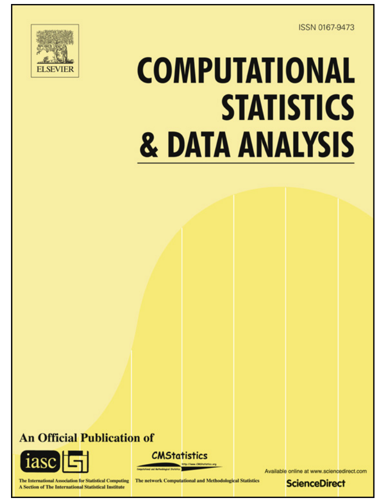
Journal Pre-proof



Entropy-based test for generalized Gaussian distributions

Mehmet Siddik Cadirci, Dafydd Evans, Nikolai Leonenko and Vitalii Makogin

PII: S0167-9473(22)00082-2
DOI: https://doi.org/10.1016/j.csda.2022.107502
Reference: COMSTA 107502

To appear in: *Computational Statistics and Data Analysis*

Received date: 10 August 2021
Revised date: 6 April 2022
Accepted date: 7 April 2022

Entropy-based test for generalized Gaussian distributions

Mehmet Siddik Cadirci, Dafydd Evans, Nikolai Leonenko

School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, Wales, UK, CF24 4AG

Vitalii Makogin*

Institute of Stochastics, Ulm University, Ulm, 08069 Germany.

Abstract

The proof of L^2 consistency for the kth nearest neighbour distance estimator of the Shannon entropy for an arbitrary fixed $k \geq 1$ is provided. It is constructed the non-parametric test of goodness-of-fit for a class of introduced generalized multivariate Gaussian distributions based on a maximum entropy principle. The theoretical results are followed by numerical studies on simulated samples. It is shown that increasing of k improves the power of the introduced goodness of fit tests. The asymptotic normality of the test statistics is experimentally proven.

Keywords: Maximum entropy principle, generalized Gaussian distribution, Shannon entropy, nearest neighbour estimator of entropy, goodness-of-fit test

2010 MSC: 62G05, 62G10, 62H12, 62H15, 28D20

We propose a non-parametric test of goodness-of-fit for a class of generalized multivariate Gaussian distributions. Our approach is based on the estimation of the differential (Shannon) entropy

$$H(f) = -\int_{\mathbb{R}^m} f(x) \log f(x) \mathrm{d}x, \tag{1}$$

where f is a density function of a continuous random vector $X \in \mathbb{R}^m$.

We use entropy estimators based on nearest neighbour distances. These were first studied by (Kozachenko and Leonenko, 1987) and subsequently by (Tsybakov and Van der Meulen, 1996; Evans et al., 2002; Goria et al., 2005; Leonenko et al., 2008; Leonenko and Pronzato, 2010; Evans, 2008; Penrose and Yukich, 2011; Delattre and Fournier, 2017; Gao et al., 2018; Bulinski and Kozhevin, 2019; Bulinski and Dimitrov, 2019) and Berrett et al. (2019). Nearest neighbour estimators (NNE) are particularly attractive because they are computationally efficient and generalized easily to the multivariate case. For an overview of non-parametric techniques of entropy estimation, see (Beirlant et al., 1997).

*Corresponding author

Email addresses: cadircims@cardiff.ac.uk (Mehmet Siddik Cadirci), evansd8@cardiff.ac.uk (Dafydd Evans), leonenkon@cardiff.ac.uk (Nikolai Leonenko), vitalii.makogin@uni-ulm.de (Vitalii Makogin)

Berrett et al. (2019) have shown that subject to certain regularity conditions, as $k \to \infty$, the k-NNE of Shannon entropy is efficient only for $m \leq 3$, and present a bias-corrected estimator for dimensions $m \geq 4$. In this paper, we focus on the conventional k-NNE with fixed $k \geq 1$, and prove its L^2 consistency. Note that the asymptotic variance of k-NNE decreases rapidly up to $k = 3$ only, see (Berrett et al., 2019, Table 1), and this asymptotic inflation is distribution-free, which leads to the conjecture that $k = 3$ is the most interesting case for any $m \geq 1$. Our computational study confirms that the increasing of k improves the test's power.

Entropy-based tests of goodness-of-fit exploit the so-called maximum entropy principle, see (Vasicek, 1976; Kapur, 1989). Choi (2008) introduced an entropy-based normality based on the fact that normal densities posses the largest Shannon entropy among all densities with the same variance, see also Dudewicz and Van Der Meulen (1981); Goria et al. (2005); Evans (2008) and the references therein. This paper proposes a new entropy-based test of a generalized normality based on the maximum Shannon entropy principle for the generalized multivariate Gaussian distribution. Our test statistic uses the k-th nearest neighbour estimators of entropy and a moments estimators of order $s > 0$. This methodology can be applied in image analysis, statistical physics, image processing, classifications of data, pattern recognition and machine learning. Provided Monte-Carlo simulations show that the test statistics are asymptotically normally distributed.

The goodness of fit testing for Gaussian data is well-studied, see the review of Ebner and Henze (2020), and (González-Manteiga et al., 2016) among recent papers. Note that for high-dimensional data, a dimension reduction is recommended before applying our methodology, see cf. (Shin and Artemiou, 2017).

The paper is organized as follows: we introduce the multivariate generalized Gaussian distribution in Section 1 followed by a maximum entropy principle for them established in Section 2. In Section 3, we discuss the state of the art for k-NNE, and present the associated goodness-of-fit statistics in Section 4. In Section 5, we prove the L^2 consistency of $k-$NNE of the Shannon entropy with some auxiliary material on entropy bounds deferred to Appendix A. Numerical results are included in Section 6.

1. The generalized Gaussian distribution

The *multivariate exponential power distribution* $MEP_m(s, \mu, \Sigma)$ on $\mathbb{R}^m$ has the density function (Solaro, 2004)

$$f(x\,; m, s, \mu, \Sigma) = \frac{\beta_1(m,s)}{\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}\left[(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right]^{s/2}\right), \tag{2}$$

where

$$\beta_1(m,s) = \frac{\Gamma(m/2+1)}{\pi^{m/2}\Gamma(m/s+1)2^{m/s}},$$

$\mu \in \mathbb{R}^m$ is a mean vector, Σ is an $m \times m$ positive definite matrix, $s > 0$ is a shape parameter (Solaro, 2004), and variance-covariance matrix $V = \beta_2 \Sigma$ where

$$\beta_2(m,s) = \frac{2^{2/s}\Gamma\left[(m+2)/s\right]}{m\Gamma(m/s)}. \tag{3}$$

Note that $s = 2$ corresponds to the multivariate normal distribution $N(\mu, \Sigma)$ on $\mathbb{R}^m$, while $s = 1$ corresponds to the multivariate Laplace distribution. Multivariate exponential power distribution was introduced by De Simoni (1968) and studied by Kano (1994) and Gómez et al. (1998). MEP_m distributions belong to the elliptic family of multivariate distributions, and particularly to symmetric Kotz type distributions, see (Fang et al., 1990) for details.

Taking μ to be the null vector and Σ to be the identity matrix, we obtain the *isotropic exponential power distribution* $IEP_m(s)$ on $\mathbb{R}^m$,

$$f(x\,;m,s) = \frac{\Gamma(m/2+1)}{\Gamma(m/s+1)\pi^{m/2}2^{m/s}} \exp\left(-\frac{1}{2}\|x\|^s\right), \tag{4}$$

$x \in \mathbb{R}^m$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$, i.e., $\|x\| = \sqrt{x_1^2 + \cdots + x_m^2}$, for $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$.

Applying the scaling $x \mapsto (2\tau)^{-1/s}x$ for $\tau > 0$ yields the *generalized Gaussian* distributions $GG_\tau(m,s)$ on $\mathbb{R}^m$, with density functions

$$f_c(x;m,s) = c(m,s)\exp\left(-\tau\|x\|^s\right), x \in \mathbb{R}^m, \tag{5}$$

where

$$c(m,s) = \frac{\Gamma(m/2+1)\tau^{m/s}}{\Gamma(m/s+1)\pi^{m/2}},$$

and taking $\tau = 1/s$ yields the canonical distribution $GG(m,s)$ with the density

$$f(x;m,s) = c_0(m,s)\exp\left(-\frac{\|x\|^s}{s}\right), x \in \mathbb{R}^m, \tag{6}$$

where

$$c_0(m,s) = \frac{\Gamma(m/2+1)}{\Gamma(m/s+1)\pi^{m/2}s^{m/s}}.$$

A random vector $X \in \mathbb{R}^m$ is called *isotropic* if its density f can be written as $f(x) = \tilde{f}(\|x\|)$ for some function $\tilde{f} : \mathbb{R} \to [0,\infty)$ called the *radial density*. If X is isotropic and $g : \mathbb{R} \to \mathbb{R}$ is a Borel function, it is easy to show that

$$\mathbb{E}\left[g(\|X\|)\right] = \int_{\mathbb{R}^m} g(\|x\|)f(x)\,dx = \frac{2\pi^{m/2}}{\Gamma(m/2)}\int_0^\infty g(r)\tilde{f}(r)r^{m-1}\,dr \tag{7}$$

provided the integrals exist. In particular, the moments of order $s > 0$ are given by

$$\mathbb{E}(\|X\|^s) = \frac{2\pi^{m/2}}{\Gamma(m/2)} \int_0^\infty r^{m+s-1} \tilde{f}(r)\, dr \tag{8}$$

provided the integrals exist.

Lemma 1. If $X \sim GG_\tau(m,s)$, then $\mathbb{E}(\|X\|^s) = m/(s\tau)$.

Proof. If $X \sim GG_\tau(m,s)$ then X is isotropic and has the radial density function

$$\tilde{f}(r) = \frac{\Gamma(m/2+1)\tau^{m/s}}{\Gamma(m/s+1)\pi^{m/2}} \exp(-\tau r^s).$$

Hence by (8) we have

$$\mathbb{E}(\|X\|^s) = \frac{2\pi^{m/2}}{\Gamma(m/2)} \int_0^\infty r^s \tilde{f}(r) r^{m-1}\, dr = \frac{m\tau^{m/s}}{\Gamma(m/s+1)} \int_0^\infty r^{m+s-1} \exp(-\tau r^s)\, dr$$

and changing the variable of integration to $t = \tau r^s$ yields

$$\mathbb{E}(\|X\|^s) = \frac{m}{s\tau\,\Gamma(m/s+1)} \int_0^\infty t^{m/s} e^{-t} dt = \frac{m}{s\tau}.$$

□

The above formulas can be derived from the results on symmetric Kotz type distributions, see (Fang et al., 1990, Section 3.2.3).

2. A maximum entropy principle for $GG_\tau(m,s)$

It is well known, (Kapur, 1989), that among all distributions on $\mathbb{R}^m$ whose densities f are supported on the whole of $\mathbb{R}^m$ and whose mean and covariance matrix are fixed at zero and Σ respectively, the Shannon entropy $H(f)$ is maximized by the multivariate Gaussian distribution $N(0,\Sigma)$ on $\mathbb{R}^m$, and thus

$$H(f) \le \log\left[(2\pi e)^{m/2}\sqrt{\det\Sigma}\right]. \tag{9}$$

We now prove an analogous result for the generalized Gaussian distribution.

Theorem 1. Let $X \in \mathbb{R}^m$ be a random vector, whose density f is supported on the whole $\mathbb{R}^m$, and for which there exists some $s > 0$ such that $\mathbb{E}(\|X\|^s) < \infty$. Then $H(f) < +\infty$ and satisfies

$$H(f) \le \frac{m}{s} \log\left(c_1(m,s)\mathbb{E}\|X\|^s\right),$$

where

$$c_1(m,s) = \left(\frac{\pi^{m/2}\Gamma(m/s+1)}{\Gamma(m/2+1)}\right)^{s/m} \left(\frac{se}{m}\right)$$

with equality if and only if $X \sim GG_\tau(m,s)$ with $\tau = m/(s\mathbb{E}\|X\|^s)$.

Proof. Let X and Z be two random vectors whose density functions, f and f^* respectively, are supported on the whole of $\mathbb{R}^m$, and for which there exists some $s > 0$ with $\mathbb{E}\|X\|^s = \mathbb{E}\|Z\|^s < \infty$. First, we observe that

$$H(f) \leq -\int_{\mathbb{R}^m} f(x) \log f^*(x)\, dx, \tag{10}$$

with equality if and only if $f = f^*$ almost everywhere. This follows by Jensen's inequality,

$$\begin{aligned} &-\int_{\mathbb{R}^m} f(x) \log f(x)\, dx + \int_{\mathbb{R}^m} f(x) \log f^*(x) dx \\ &= \int_{\mathbb{R}^m} f(x) \log \left(\frac{f^*(x)}{f(x)} \right) dx \leq \log \left(\int_{\mathbb{R}^m} f(x) \left(\frac{f^*(x)}{f(x)} \right) dx \right) = \log \left(\int_{\mathbb{R}^m} f^*(x) dx \right) = 0, \end{aligned}$$

assuming that both integrals $-\int_{\mathbb{R}^m} f(x) \log f(x)\, dx$ and $\int_{\mathbb{R}^m} f(x) \log f^*(x) dx$ are finite.

If $Z \sim GG_\tau(m, s)$ with $\tau = \dfrac{m}{s\mathbb{E}\|X\|^s}$ (which ensures that $\mathbb{E}\|Z\|^s = \mathbb{E}\|X\|^s$) we have

$$f^*(x) = c(m, s) \exp(-\tau \|x\|^s),$$

where

$$c(m, s) = \frac{\Gamma(m/2 + 1)\tau^{m/s}}{\Gamma(m/s + 1)\pi^{m/2}}.$$

For this case, $-\log f^*(x) = \tau\|x\|^s - \log c(m, s)$ and hence

$$\begin{aligned} -\int_{\mathbb{R}^m} f(x) \log f^*(x) dz &= \tau \int_{\mathbb{R}^m} \|x\|^s f(x)\, dx - (\log c(m, s)) \int_{\mathbb{R}^m} f(x)\, dx \\ &= \tau\, \mathbb{E}\|X\|^s - \log c(m, s) = \frac{m}{s} - \log c(m, s) \quad \text{by Lemma 1.} \end{aligned}$$

Therefore $\int_{\mathbb{R}^m} f(x) \log f^*(x) dx$ is finite under existence of $\mathbb{E}\|X\|^s$ and the right-hand side of (10) is finite. If $H(f) = -\infty$, inequality (10) is valid as well.

Thus, by (10) and substituting for $c(m, s)$, we obtain

$$H(f) \leq \frac{m}{s} - \log \left[\frac{\tau^{m/s}\Gamma(m/2 + 1)}{\pi^{m/2}\Gamma(m/s + 1)} \right] = \frac{m}{s} \log \left[\left(\frac{\Gamma(m/s + 1)}{\Gamma(m/2 + 1)} \right)^{\frac{s}{m}} \left(\frac{e\pi^{\frac{s}{2}}}{\tau} \right) \right],$$

and substituting for $\mathbb{E}\|X\|^s = m/(s\tau)$ completes the proof in the case $H(f) < +\infty$.

Consider $H_M(f) := -\int_{A_M} f(x) \log f(x) dx$, where $A_M = \{x \in \mathbb{R}^m, |\log f(x)| \leq M\}$. Denote by $C_M := \int_{A_M} f(x) dx \leq 1$ and $C_M^* := \int_{A_M} f^*(x) dx \leq 1$. Then by Jensen's inequality we have

$$\begin{aligned} &-\int_{A_M} f(x) \log f(x)\, dx + \int_{A_M} f(x) \log f^*(x) dx \\ &= C_M \int_{A_M} \frac{f(x)}{C_M} \log \left(\frac{f^*(x)}{f(x)} \right) dx \leq C_M \log \left(\int_{A_M} \frac{f^*(x)}{C_M} dx \right) = C_M \log C_M^* - C_M \log C_M. \end{aligned}$$

Therefore,

$$\begin{aligned} H(f) &\leq \limsup_{M \to \infty} \left(-\int_{A_M} f(x) \log f(x)\, dx \right) \\ &\leq \limsup_{M \to \infty} \left(-\int_{A_M} f(x) \log f^*(x)\, dx + C_M \log \frac{C_M^*}{C_m} \right) < +\infty. \end{aligned}$$

□

Remark 1. Theorem 1 was proved for $m = 1$ in (Wyner and Ziv, 1969) and (Rosenblatt, 2000, p.103-104). For $m \geq 1$, some statements of Theorem 1 were also proved using other methods by Lutwak et al. (2007).

The similar to Theorem 1 result holds for the general multivariate exponential power distribution $MEP_m(s, \mu, \Sigma)$.

Theorem 2. Let $X \in \mathbb{R}^m$ be a random vector, whose density f is supported on the whole of $\mathbb{R}^m$. Let there exist some $s > 0$, $\mu \in \mathbb{R}^m$ and $m \times m$ positive definite matrix Σ such that $M_s := \mathbb{E}\left((X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu)\right)^{s/2} < \infty$. Then $H(f)$ is finite and satisfies

$$H(f) \leq \log \frac{\sqrt{\det \Sigma}}{\beta_1(m, s)} + \frac{1}{2} M_s \tag{11}$$

with equality if and only if $X \sim MEP_m(s, \mu, \Sigma)$.

Remark 2. Let $\lambda_1, \ldots, \lambda_m$ be eigenvalues of Σ. The moment M_s exists if and only if $\mathbb{E}\|X\|^s < \infty$ due to inequalities $\min_{i=1}^m \lambda_i^{-1}\|X\|^2 \leq (X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu) \leq \max_{i=1}^m \lambda_i^{-1}\|X\|^2$. Therefore, the entropy of X is maximized on an isotropic exponential power distribution *IEP*. Inequality (11) can be used for a goodness of fit test for a family of *MEP* distributions with standard estimators $\hat{\mu}$ and $\hat{V}$ for the mean value μ and the variance-covariance matrix $V = \beta_2\Sigma$, respectively. In the following, we consider the isotropic case in detail.

Aulogiaris and Zografos (2004) prove the analogous maximum entropy principle for the symmetric Kotz type distributions, but fixing the moments of two types $\mathbb{E}(\|X\|^s)$ and $\mathbb{E}(\log\|X\|)$.

3. Entropy estimation

In this section, we discuss the statistical estimation of the Shannon entropy for a wide class of continuous distributions. Let $k \geq 1$ and $N > k$, and let $\mathcal{X}_N = \{X_1, \ldots, X_N\}$ be a set of independent and identically distributed random vectors in $\mathbb{R}^m$ with a common density function f. Let F be a finite subset of $\mathcal{X}_N$ having cardinality at least k, and let $\rho_k(x, F)$ denote the Euclidean distance between a point x and its kth nearest neighbour in the set $F \setminus \{x\}$. The *kth nearest neighbour estimator* (k-NNE) of the Shannon entropy $H(f)$ is defined to be

$$\widehat{H}_{N,k} = \frac{1}{N}\sum_{i=1}^{N} \log\left[\rho_k^m(X_i, \mathcal{X}_N) V_m (N-1) e^{-\psi(k)}\right], \tag{12}$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function and $V_m = \pi^{m/2}/\Gamma(m/2+1)$ is the volume of the unit ball in $\mathbb{R}^m$. For $k = 1$, this reduces to

$$\widehat{H}_{N,1} = \frac{m}{N}\sum_{i=1}^{N} \log \rho_1(X_i, \mathcal{X}_N) + \log V_m + \gamma + \log(N-1), \tag{13}$$

where $\gamma = -\psi(1) \approx 0.577216$ is the Euler-Mascheroni constant. The estimator (13) was introduced by Kozachenko and Leonenko (1987) while the general estimator (12) was first considered by Goria et al. (2005). The main properties of (12) have been studied by Leonenko et al. (2008); Leonenko and Pronzato (2010); Penrose and Yukich (2011); Delattre and Fournier (2017); Gao et al. (2018); Bulinski and Dimitrov (2021, 2019); Berrett et al. (2019) and Berrett and Samworth (2019).

Particularly, the proof of NNE and k-NNE consistency developed as follows. Pioneering paper (Kozachenko and Leonenko, 1987) states that $\mathbb{E}\widehat{H}_{N,1} \to H(f)$ if $\int_{\mathbb{R}^m} |\log f(x)|^{1+\varepsilon} f(x)dx < \infty$ and $\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} |\log \rho_1(x,y)|^{1+\varepsilon} f(x)f(y)dxdy < \infty$ for some $\varepsilon > 0$. Later, Leonenko et al. (2008) prove that if f is bounded and $\int_{\mathbb{R}^m} f^{\varepsilon}(x)dx < \infty$ for some $\varepsilon < 1$, then $\mathbb{E}\widehat{H}_{N,1} \to H(f)$ in L^2 as $N \to \infty$.

Other researchers could not significantly improve these quite restrictive conditions. Thus, if f is bounded and has compact support, then $\widehat{H}_{N,1}$ is L_2-consistent, and for unbounded samples, f needs to be strictly (in some sense) positive. A new approach based on the limit theory of Poisson point processes allows breaking this wall.

Many new results and methods on the nearest neighbour estimate of entropy can be found in the book by Biau and Devroye (2015). In particular, they show that for $k = 1$, $\widehat{H}_{N,1} \to H(f)$ in probability as $N \to \infty$ if a density f is bounded and $\int_{R^m} f(x)\log^2(f(x)+1)dx < \infty$. Moreover, the recent paper (Devroye and Győfi, 2021) shows that $\mathbb{E}\left\{\left|\widehat{H}_{N,1} - H(f)\right|\right\} \to 0$ as $N \to \infty$, if and only if $\mathbb{E}\left\{\log\left(\|X\|+1\right)\right\} < \infty$. And for compactly supported densities $\widehat{H}_{N,1} \to H(f)$, $N \to \infty$ almost surely. Note that in the same spirit Lord et al. (2018) introduced and studied the geometric k-nearest neighbour estimation of entropy and mutual information.

In this paper, we prove the convergence in mean-square of $\widehat{H}_{N,k}$ for the arbitrary $k \geq 1$.

Theorem 3 (Main theorem)**.** Suppose that $\mathbb{E}\|X\|^{\alpha} < \infty$ for some $\alpha > 0$ and $f(x) \leq M$ for some $M > 0$. Then for any fixed $k \geq 1$,

$$\mathbb{E}\left[\widehat{H}_{N,k} - H(f)\right]^2 \to 0 \quad \text{as } N \to \infty. \tag{14}$$

The case $k = 1$ in Theorem 3 was proved by Penrose and Yukich (2013, Theorem 2.4.ii). Note that estimators $\widehat{H}_{N,k}$ from (12) and $\widehat{H}_{N,1}$ from (13) have several non-matching terms due to the different distributions (Erlang and Poisson) arising in the proof of their consistency.

Remark 3. The condition of boundedness for the density f is not explicitly stated in (Penrose and Yukich, 2013, Theorem 2.4.ii). In Appendix A, we give an example of a density f with bounded support and for which $H(f)$ is unbounded.

4. A test statistic for $GG(m, s)$

Let $k \geq 1$ be fixed and $\mathcal{K}$ be the class of density functions f on $\mathbb{R}^m$ such that

1. $\text{supp}(f) = \mathbb{R}^m$,
2. $\mathbb{E}(\|X\|^s) < \infty$ for some $s > 0$,
3. $\mathbb{E}(\widehat{H}_{N,k}) \to H(f)$ as $N \to \infty$, and
4. $\widehat{H}_{N,k} \to H(f)$ in probability as $N \to \infty$.

Proposition 1. The density functions of the $GG_\tau(m, s)$ belong to $\mathcal{K}$ for all $m \geq 1$, $s > 0$, $\tau > 0$ and $k \geq 1$.

Proof. The statement follows from Theorem 3, which applies because f is bounded, and Lemma 1. $\square$

Let $X \in \mathbb{R}^m$ be a random vector with a density $f \in \mathcal{K}$, and let $s > 0$ be fixed. Based on a random sample $X_1, X_2, \dots$ from the distribution of X, we use the maximum entropy principle proved in Section 2 to test the hypothesis $X \sim GG(m, s)$ against a suitable alternative. By Theorem 1, if $X \sim GG(m, s)$ then

$$H(X) = \frac{m}{s} \log \mathbb{E}\|X\|^s + \frac{m}{s} \log c_1(m, s),$$

where

$$c_1(m, s) = \left(\frac{\pi^{m/2}\Gamma(m/s + 1)}{\Gamma(m/2 + 1)} \right)^{s/m} \left(\frac{se}{m} \right).$$

We estimate the entropy $H(X)$ by the kth nearest neighbour estimator $\widehat{H}_{N,k}$ from (12) and the moment $\mathbb{E}\|X\|^s$ by the sample moment

$$\bar{X}_N^{(s)} = \frac{1}{N} \sum_{i=1}^{N} \|X_i\|^s.$$

Our test statistic $T_{N,k} = T_{N,k}(m, s)$ is then

$$T_{N,k} = \frac{m}{s} \log c_1(m, s) + \frac{m}{s} \log \bar{X}_N^{(s)} - \widehat{H}_{N,k}.$$

By the law of large numbers, $\bar{X}_N^{(s)} \to \mathbb{E}\|X\|^s$ in probability as $N \to \infty$. Hence by Slutsky's theorem, if $X \sim GG(m, s)$ then for any fixed $k \in \{1, \dots, N-1\}$ we have

$$T_{N,k} \to 0 \quad \text{in probability as } N \to \infty.$$

Otherwise, by the maximum entropy principle it must be that $T_{N,k} \to \xi$ in probability as $N \to \infty$, where the constant $\xi = \xi(m, s, k)$ is strictly positive. Thus, we reject the hypothesis $X \sim GG(m, s)$ whenever $T_{N,k} \geq t_{N,k,\alpha}$, where $t_{N,k,\alpha} = t_{N,k,\alpha}(m, s)$ is a so-called *critical value* of the test statistic $T_{N,k}(m, s)$ at significance level α, which is a solution of

$$\mathbb{P}_{H_0}\big(T_{N,k} \geq t\big) = \alpha.$$

An analytical derivation of the distribution of $T_{N,k}$ when $X \sim GG(m, s)$ is difficult because the covariances of $\widehat{H}_{N,k}$ and $\bar{X}_N^{(s)}$ are intractable, even though the *asymptotic* behaviour of $\widehat{H}_{N,k}$ can be revealed by applying results of Penrose and Yukich (2011); Delattre and Fournier (2017) or Berrett et al. (2019), and the asymptotic behaviour of $\bar{X}_N^{(s)}$ by the delta method. Thus, we use Monte-Carlo simulation to investigate the distribution of $T_{N,k} = T_{N,k}(m,s)$ for different combinations of parameter values.

Remark 4. The test statistic $T_{N,k}$ is scale-invariant: if $Y = aX$ for some $a > 0$, then $\widehat{H}_{N,k}(Y) = \log(a^m) + \widehat{H}_{N,k}(X)$ and $\bar{Y}_N^{(s)} = a^s \bar{X}_N^{(s)}$, and hence

$$\begin{aligned} T_{N,k}(Y) &= \frac{m}{s}\log c_1(m,s) + \frac{m}{s}\log \bar{Y}_N^{(s)} - \widehat{H}_{N,k}(Y) \\ &= \frac{m}{s}\log c_1(m,s) + \frac{m}{s}\log \bar{X}_N^{(s)} + \frac{m}{s}\log(a^s) \\ &- \log(a^m) - \widehat{H}_{N,k}(X) \\ &= \frac{m}{s}\log c_1(m,s) + \frac{m}{s}\log \bar{X}_N^{(s)} - \widehat{H}_{N,k}(X) = T_{N,k}(X). \end{aligned}$$

5. L^2 consistency of the kth nearest neighbour estimator

In this section, we prove Theorem 3 for arbitrary $k \geq 1$. To this end we write (12) as

$$\widehat{H}_{N,k} = \frac{1}{N}\sum_{x\in\mathcal{X}_N} l\left(N^{\frac{1}{m}}x, N^{\frac{1}{m}}\mathcal{X}_N\right),$$

where

$$l(x,\mathcal{X}) := \log\left(\rho_k^m(x,\mathcal{X})V_m e^{-\psi(k)}\right), x \in \mathbb{R}^m.$$

First, we require the following corollary of (Penrose and Yukich, 2013, Theorem 3.1).

Theorem 4. Let $k \geq 1$ and $q = 1$ or $q = 2$, and suppose there exists $p \geq q$ such that

$$\sup_{N\geq k} \mathbb{E}\left| l\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathcal{X}_N\right)\right|^p < \infty. \tag{15}$$

Then we have L^q convergence,

$$\frac{1}{N}\sum_{x\in\mathcal{X}_N} l\left(N^{\frac{1}{m}}x, N^{\frac{1}{m}}\mathcal{X}_N\right) \to \int_{\mathbb{R}^m} \mathbb{E}\left[l(0,\mathcal{P}_{f(x)})\right] f(x)\,dx$$

as $N \to \infty$, where f is the density function of X_1 and $\mathcal{P}_\lambda$ denotes a homogeneous Poisson point process of intensity $\lambda > 0$ on $\mathbb{R}^m$.

Proof of Theorem 3. We apply Theorem 4. First, we show that $H(f) = \int_{\mathbb{R}^m} \mathbb{E}\big[l(0,\mathcal{P}_{f(x)})\big] f(x)\,dx$, where

$$l\left(0,\mathcal{P}_\lambda\right) = m\log\rho_k(0,\mathcal{P}_\lambda) + \log V_m - \psi(k).$$

Denote by $B_t(0)$ the (Euclidean) ball of radius t centred at 0 i.e, $B_t(0) = \{y \in \mathbb{R}^m, \|y\| \le t\}$. The random variable $\rho_k(0, \mathcal{P}_\lambda)$ is the distance from 0 to the kth point of $\mathcal{P}_\lambda$, and thus has Erlang distribution with parameters k and $\lambda|B_t(0)| = \lambda V_m t^m$, that is

$$\begin{aligned}\mathbb{P}\big(\rho_k(0, \mathcal{P}_\lambda) \le t\big) = \mathbb{P}\big(|\mathcal{P}_\lambda \cap B_t(0)| \ge k\big) &= 1 - \sum_{j=0}^{k-1} \frac{1}{j!}\big(\lambda|B_t(0)|\big)^j e^{-\lambda|B_t(0)|} \\ &= 1 - \sum_{j=0}^{k-1} \frac{1}{j!}(\lambda V_m t^m)^j e^{-\lambda V_m t^m} \qquad (t \ge 0).\end{aligned}$$

Then

$$\begin{aligned} m\mathbb{E}[\log \rho_k(0, \mathcal{P}_\lambda)] &= \int_0^\infty \log t^m \frac{(\lambda V_m)^k (t^m)^{(k-1)}}{(k-1)!} e^{-\lambda V_m t^m} m t^{m-1} dt \\ &= -\log(\lambda V_m) + \int_0^\infty \log y \frac{y^{k-1}}{(k-1)!} e^{-y} dy \\ &= -\log \lambda - \log V_m + \psi(k). \end{aligned}$$

Hence $\mathbb{E}\big[l(0, \mathcal{P}_\lambda)\big] = -\log \lambda$ and thus $H(f)$ equals

$$-\int_{\mathbb{R}^m} f(x) \log f(x)\, dx = \int_{\mathbb{R}^m} \mathbb{E}\big[l(0, \mathcal{P}_{f(x)})\big] f(x)\, dx.$$

Second, we check condition (15). Note that for every $\delta \in (0,1)$ and $p > 1$ there exists $C > 0$ such that

$$|\log t|^p \le C t^{-\delta} \mathbb{1}_{[0,1]}(t) + C t^{\delta} \mathbb{1}_{[1,\infty)}(t), \qquad t > 0.$$

Then because

$$|l(x, \mathcal{X})|^p = \big|\log V_m - \psi(k) + \log \rho_k^m(x, \mathcal{X})\big|^p \le \big|\log V_m - \psi(k)\big|^p + \big|\log \rho_k^m(x, \mathcal{X})\big|^p,$$

we have

$$\begin{aligned} \frac{1}{2^{p-1}} \mathbb{E}\left| l\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right)\right|^p &\le |\log V_m - \psi(k)|^p + \mathbb{E}\left|\log \rho_k^m \left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right)\right|^p \\ &\le |\log V_m - \psi(k)|^p \\ &+ C\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right) \mathbb{1}_{[0,1]}\left[\rho_k^\delta\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right)\right] \qquad (16)\\ &+ C\mathbb{E}\rho_k^{\delta}\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right) \mathbb{1}_{[1,\infty)}\left[\rho_k^\delta\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right)\right]. \qquad (17) \end{aligned}$$

Term (16) is finite because

$$\begin{aligned} &\sup_{N \ge k} \mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right) \mathbb{1}_{[0,1]}\left(\rho_k^\delta\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right)\right) \\ &\le \sup_{N \ge k} \mathbb{E}\rho_1^{-\delta}\left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right) < \infty, \qquad (18) \end{aligned}$$

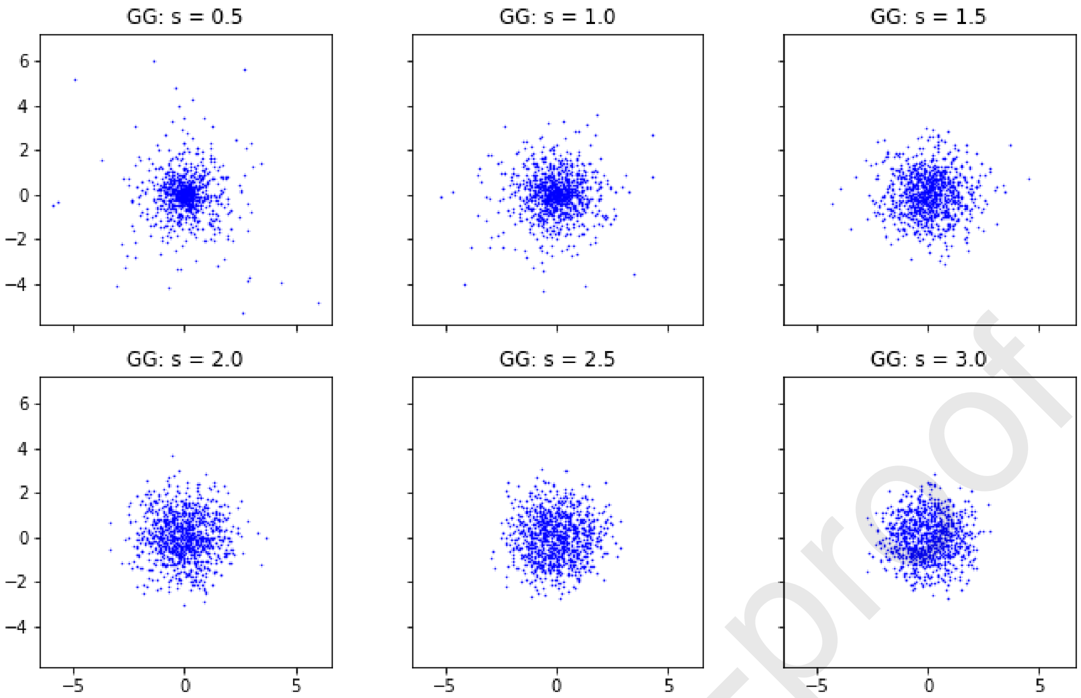


Figure 1: Scatter plots for $GG(m, s)$ with $m = 2$

where (18) is ensured by (Penrose and Yukich, 2013, Lemma 7.5) since f is bounded and $\delta \in (0, m)$.

Let $r_c(f) := \sup\{r \geq 0 : \mathbb{E}\|X_1\|^r < \infty\}$. In the proof of (Penrose and Yukich, 2011, Theorem 2.3) we see that if $r_c(f) > 0$ and $0 < \delta < mr_c(f)(m + r_c(f))^{-1}$, then

$$\sup_{N \geq k} \mathbb{E}\rho_k^\delta \left(N^{\frac{1}{m}} X_1, N^{\frac{1}{m}} \mathcal{X}_N\right) < \infty.$$

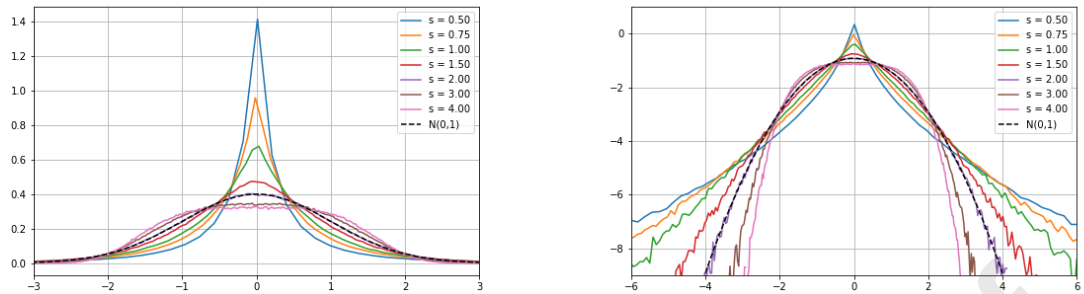
Thus, term (17) is finite. □

6. Numerical results

To investigate the behaviour of the statistic $T_{N,k}(m, s)$, we generate random samples from the $GG(m, s)$ distribution. This is achieved via the following stochastic representations (Solaro, 2004).

Lemma 2. For $X \sim GG(m, s)$, we have $X \stackrel{d}{=} UR$ where U is uniformly distributed on $\mathbb{S}^{m-1}$ and $R \stackrel{d}{=} V^{1/s}$ with $V \sim \text{Gamma}(m/s, 2)$.

For the case $m = 2$, we put the generated points on scatter plots for different values of s, see Figure 1.



(a) Empirical density function

(b) Empirical log–density function

Figure 2: Empirical distribution of $GG(m, s)$ for $m = 1$ and different values of s.

6.1. *Empirical distribution of $GG(m, s)$*

We generate $N = 10^6$ points from the $GG(m, s)$ distribution for different values of s. For the purpose of comparison, we apply the scaling $X \mapsto X/\sigma$, where

$$\sigma^2 = \frac{2^{2/s}\Gamma\big[(m+2)/s\big]}{m\Gamma(m/s)}$$

is the variance of the $GG(m, s)$ distribution. The results are shown in Figure 2.

6.2. *Asymptotic behaviour of $T_{N,k}(m, s)$ as $N \to \infty$.*

For fixed N and (m, s), we generate a sample of size N from the $GG(m, s)$ distribution and record the empirical value of $T_{N,k}(m, s)$ for a fixed k, repeating this $M = 10^3$ times. This yields a sample realization $\{T_1, T_2, \dots, T_M\}$ from the distribution of $T_{N,k}(m, s)$, from which we estimate its mean and variance by

$$\bar{T}_{N,k}(m, s) = \frac{1}{M}\sum_{j=1}^{M} T_i \quad \text{and} \quad S^2_{N,k}(m, s) = \frac{1}{M-1}\sum_{j=1}^{M}(T_i - \bar{T}_{N,k})^2.$$

In Figure 3, we show how $\bar{T}_{N,k}(m, s)$ approaches 0 as N increases for various values of $m \in \{2, 3\}$, $s \in \{0.5, 1.5, 2.5\}$ and $k = 1, 2, 3$ with error bars corresponding to the standard error $S_{N,k}(m, s)$. From these data, we observe that the empirical variance is decreasing when k increases. From the other hand, the bias or mean $\bar{T}_{N,k}(m, s)$ is smaller for smaller values of k or m. These results confirm the variance reduction of k-nearest neighbour estimators observed in Berrett et al. (2019).

Moreover, we study the rate of convergence of $\bar{T}_{N,k}(m, s)$ and $S_{N,k}(m, s)$ with respect to N. We examine the model

$$\log|\bar{T}_{N,k}(m, s)| = \alpha_{m,s,k} + \beta_{m,s,k}\log N - \frac{1}{2}\log N, \quad S^2_{N,k}(m, s) = \frac{\sigma_k^2}{N},$$

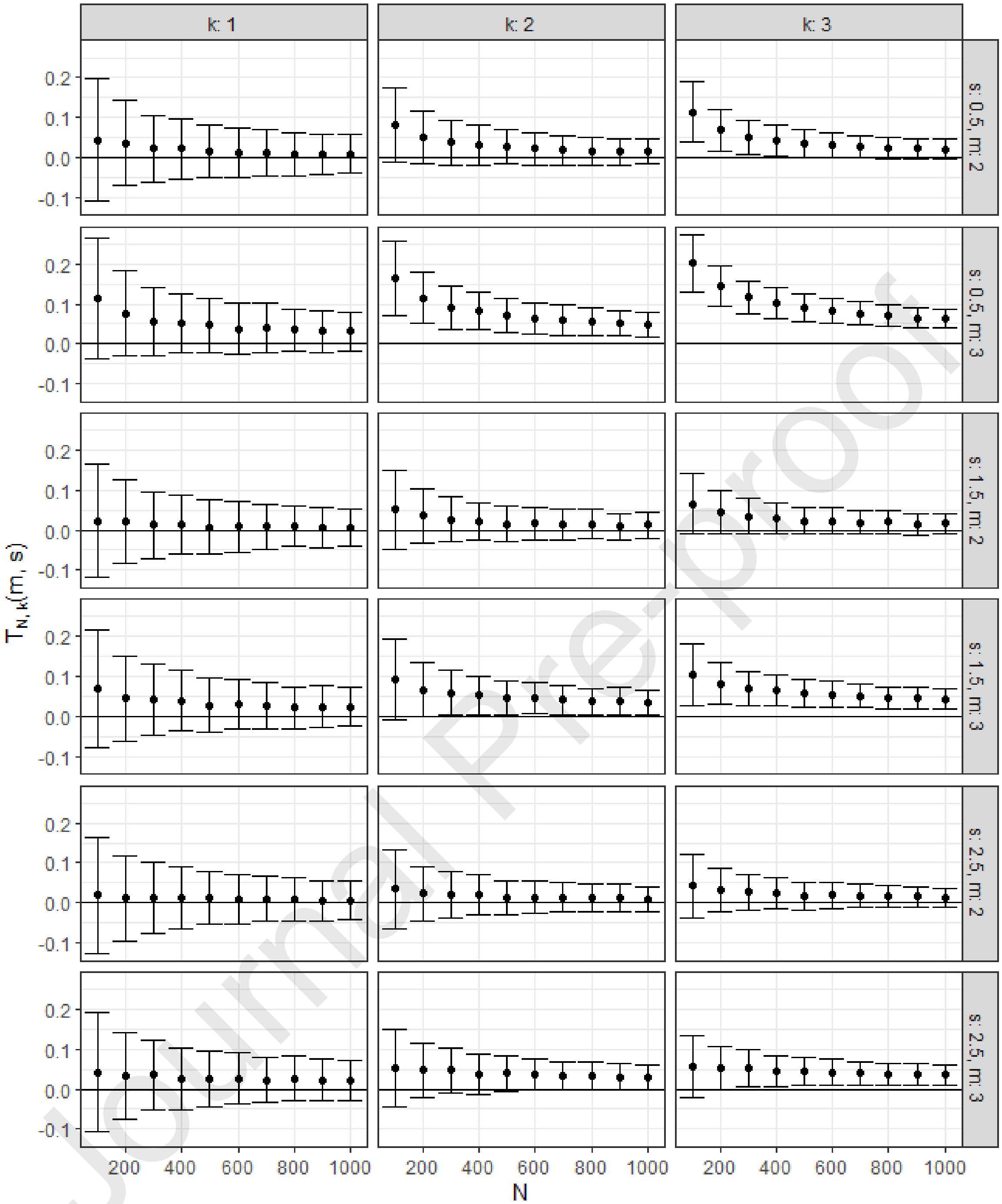


Figure 3: Consistency of $T_{N,k}(m, s)$ for different values of k, m and s ($M = 10^3$ repetitions).

based on Monte-Carlo simulations with $M = 10^3$ and $s \in \{0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 6\}$, $m = 1, 2, 3$, $k = 1, 2, 3$ and $N = \{200, 250, 300, \ldots, 1000\}$. We apply the standard linear regression method for estimation of α and β. The values of β presented in Table 1 show that for $m = 1, s \in (0, 6]$,

	$m=1$			$m=2$			$m=3$		
s	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
0.5	-0.5927	-0.5949	-0.5695	-0.4048	-0.2908	-0.2505	-0.0274	-0.0158	-0.0142
0.75	-0.9096	-0.7919	-0.6755	-0.3325	-0.2681	-0.2446	-0.0395	-0.0060	0.0161
1	-0.7831	-0.8072	-0.5907	-0.2570	-0.1509	-0.1587	0.0236	0.0369	0.0545
1.5	-0.4684	-0.3141	-0.4933	-0.1627	-0.1818	-0.1569	0.0980	0.1087	0.1103
2	-0.6773	-0.8744	-0.7892	-0.2581	-0.1591	-0.1066	0.1027	0.1355	0.1652
2.5	-0.5324	-0.2288	-0.4225	-0.0812	-0.0398	-0.0564	0.1720	0.1897	0.2417
3	-0.0982	-0.2166	-0.2609	-0.0732	0.0200	0.0491	0.1296	0.2073	0.2813
4	0.4087	-1.1497	-0.5968	-0.0849	0.0067	0.0706	0.3756	0.4163	0.5182
6	-1.2897	-0.3403	-0.5030	0.1171	0.1165	0.1810	0.5713	1.0758	1.9151

Table 1: Slope values β in Log-Log regression $\log|\mathbb{E}\hat{T}_{N,k}(m,s)| = \alpha_{m,s,k} + \beta_{m,s,k}\log N - \frac{1}{2}\log N$.

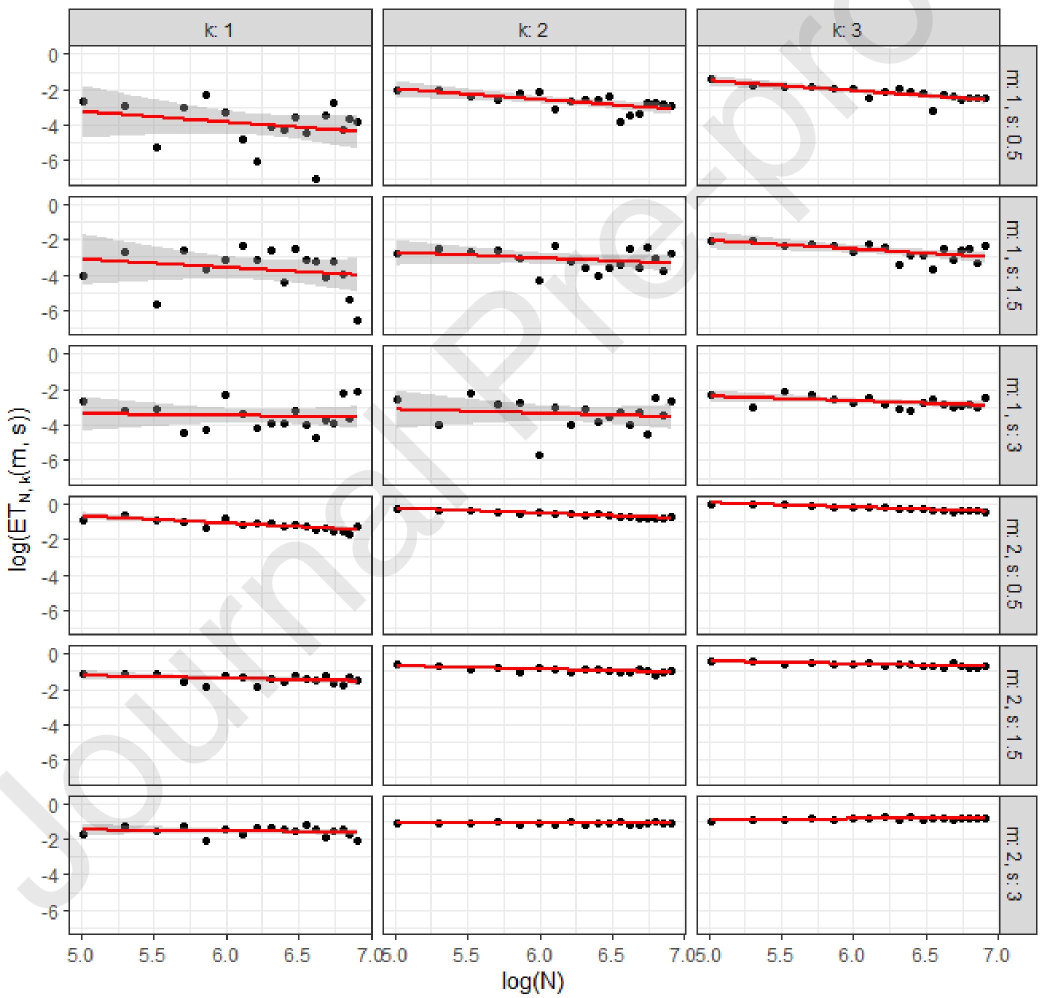


Figure 4: Log-Log regression $\log|\mathbb{E}\hat{T}_{N,k}(m,s)| = \alpha_{m,s,k} + \beta_{m,s,k}\log N - \frac{1}{2}\log N$.

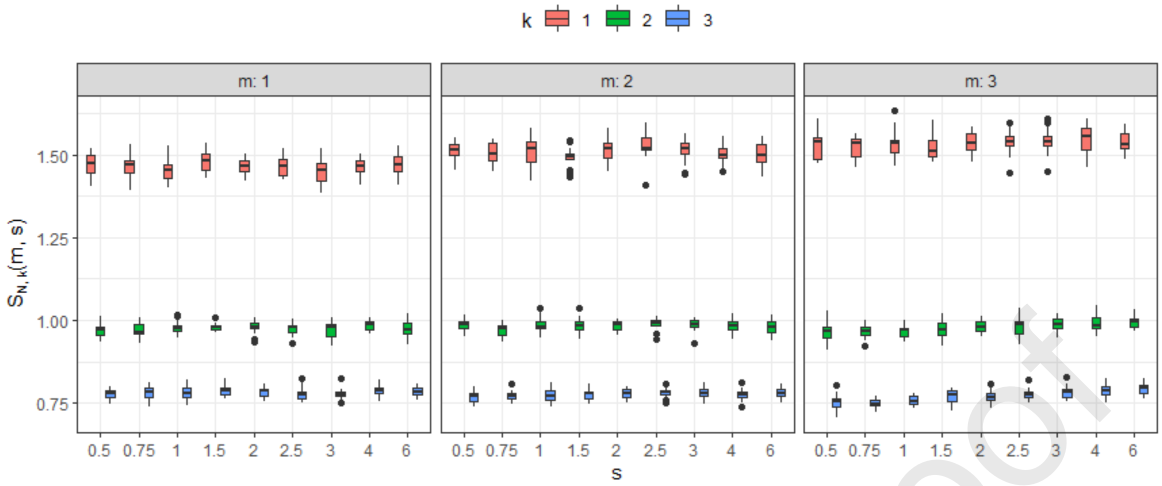


Figure 5: Values of $\sqrt{N}S_{N,k}(m,s)$.

$m = 2, s \in (0, 3]$, and $m = 3, s \in (0, 1]$ the decay of $\bar{T}_{N,k}(m,s)$ is faster or equal than $N^{-0.5}$. We can observe also, that the decay is faster as values of m, k, or s are smaller. These facts are illustrated by Figure 4 as well.

The values of $\sqrt{N}S_{N,k}(m,s)$ are presented as box-plots in Figure 5. We deduce that $\sqrt{N}S_{N,k}(m,s)$ is approximately a constant, σ_k, depending on k only. The mean values of σ_k are $\sigma_1 = 1.5014$, $\sigma_2 = 0.9788$, and $\sigma_3 = 0.7760$.

6.2.1. *Asymptotic behaviour of $T_{N,k}(m,s_0)$ on data from $GG(m,s_1)$*

For $m = 2$ and various values of s_0 and s_1 from the set $\{0.5, 1.5, 2.5\}$, we generate samples from the $GG(m,s_1)$ distribution and examine the behaviour of $T_{N,k}(m,s_0)$ as N increases. The results are shown in Figure 6.

6.3. *Empirical distribution of $T_{N,k}(m,s)$*

Numerical results suggest that the distribution of $T_{N,k}(m,s)$ is asymptotically normal as the sample size $N \to \infty$. For example, the histograms of $T_{N,k}(m,s)$ with $N = 1000$, $s = 1.5$, $m = 2$, $k = 1, 2, 3$ in Figure 7 have the Gaussian bell-shapes. The corresponding Q-Q plots in Figure 8 confirm that the distributions belong to a Gaussian family.

For different values of (N,k) and (m,s), we generate samples from the $GG(m,s)$ distribution and record the corresponding values of $T_{N,k}(m,s)$, repeating this $M = 1000$ times. To each of these samples from the distribution of $T_{N,k}(m,s)$ we then apply the Shapiro-Wilk test for normality (Shapiro and Wilk, 1965) and record the p-value returned by the test. Figure 9 shows how these p-values behave as

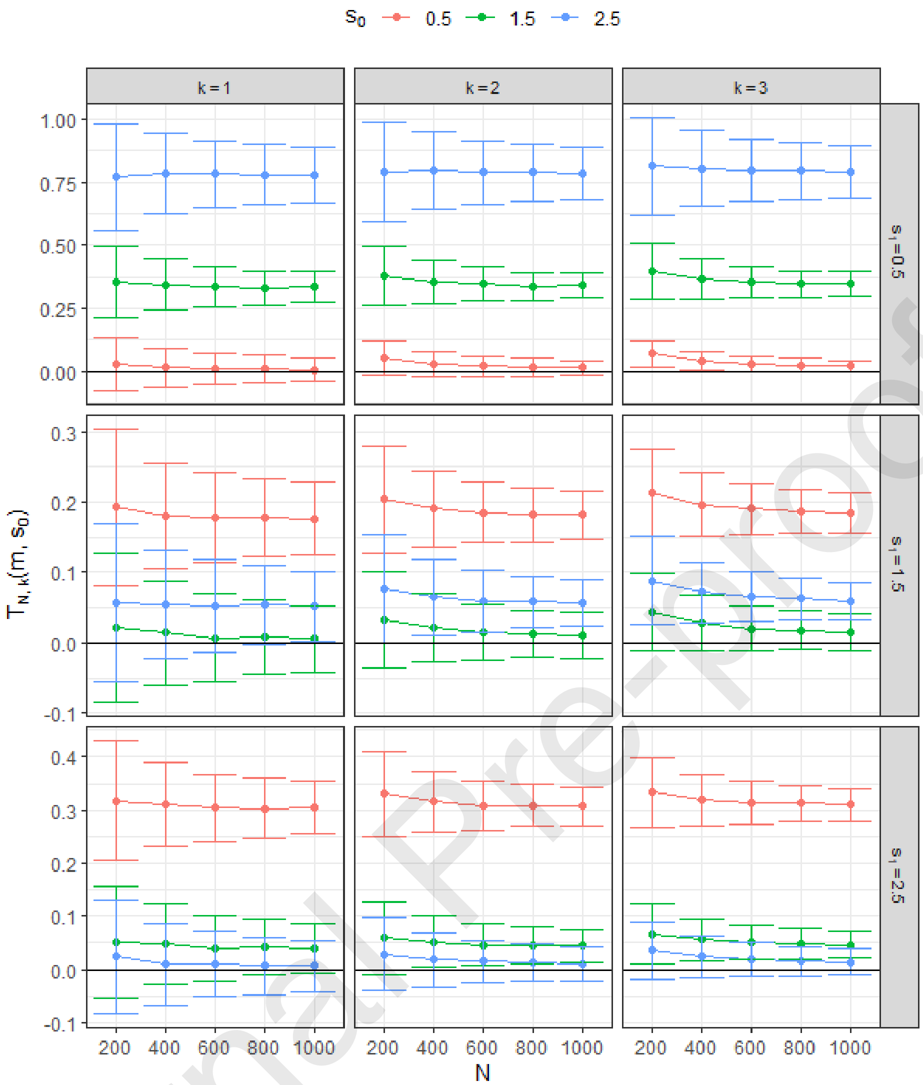


Figure 6: The behaviour of $T_{N,k}(m, s_0)$ with $m = 2$ on data from the $GG(m, s_1)$ distribution with $m = 2$.

N increases, for various values of m, s and k. The plots suggest that the normal hypothesis cannot be rejected for samples of size $N = 500$ or more.

160 6.4. Asymptotic distribution of $T_{N,k}(m, s)$

We obtain from simulations that the limiting distribution of $T_{N,k}(m, s)$ is Gaussian and the variance decay is of order N^{-2}. Therefore,

$$\sqrt{N}(T_{N,k}(m, s) - \mathbb{E}T_{N,k}(m, s)) \to N(0, \sigma_k^2), \quad N \to \infty$$

in distribution. As we have showen, $\mathbb{E}T_{N,k}(m, s)) \to 0$ for some combinations of k, m, s. Let $\hat{q}_\alpha$

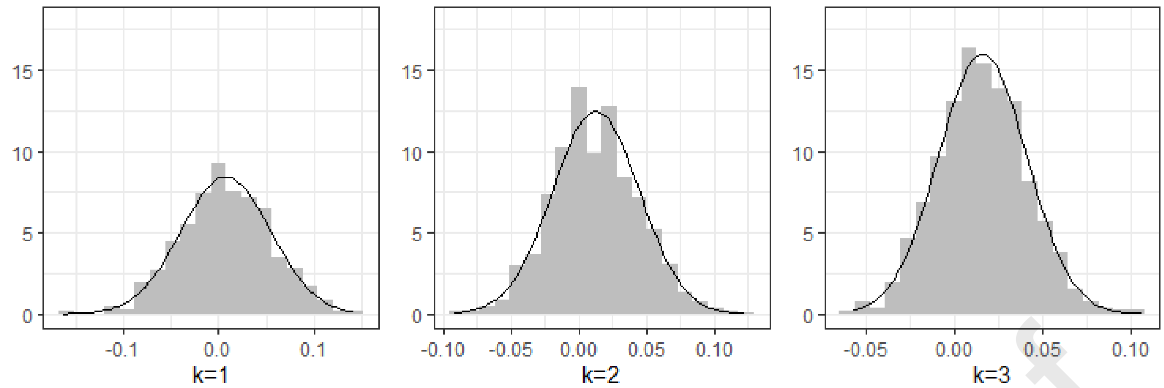


Figure 7: Empirical distributions of $T_{N,k}(m,s)$ with $m=2$ and the corresponding fitted normal curves.

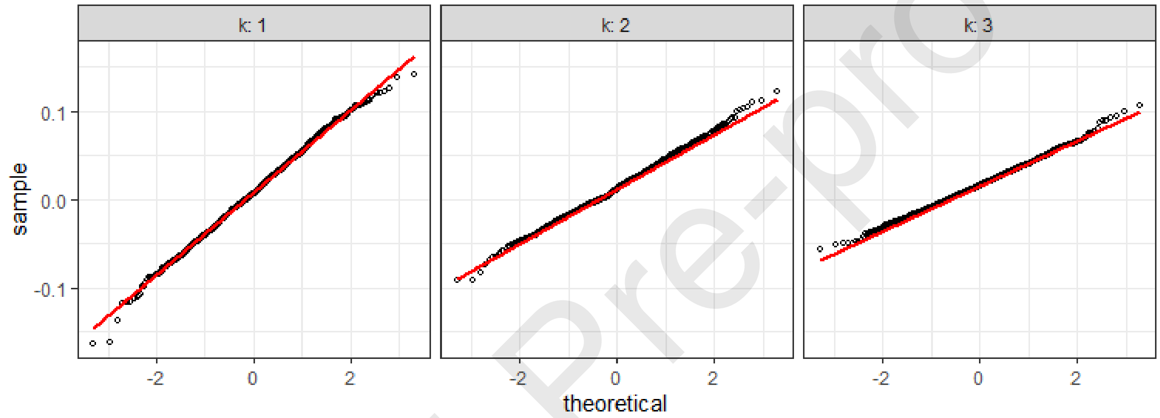


Figure 8: Q-Q plots of empirical distribution of $T_{N,k}(m,s)$ with $m=2$ and Gaussian distribution.

be the empirical quantile of order α of $T_{N,k}(m,s)$, i.e., $\mathbb{P}(T_{N,k}(m,s) > \hat{q}_\alpha | X \sim GG(m,s)) = \alpha$. We compute the values of $\hat{q}_\alpha$ by Monte-Carlo simulations with $M = 1000$ repetitions and put them in Table 2 for the case $\alpha = 0.05$ and $m = 2,3$. Applying the convergence to a Gaussian distribution, we can approximate critical values by

$$\hat{q}_\alpha \approx q_\alpha^a := \frac{z_\alpha \sigma_k + \mu_{m,s,k}}{\sqrt{N}}, \tag{19}$$

where z_α is a quantile of a standard normal law, i.e. $1 - \Phi(z_\alpha) = \alpha$, where Φ is the cumulative distribution function of $N(0,1)$. Indeed,

$$\begin{aligned}
&\mathbb{P}(T_{N,k}(m,s) > \hat{q}_\alpha | X \sim GG(m,s)) \\
&= \mathbb{P}\left(\frac{\sqrt{N}(T_{N,k}(m,s) - \mu_{m,s,k})}{\sigma_k} > \frac{\sqrt{N}(\hat{q}_\alpha - \mu_{m,s,k})}{\sigma_k} | X \sim GG(m,s) \right) \\
&\approx 1 - \Phi\left(\frac{\sqrt{N}(\hat{q}_\alpha - \mu_{m,s,k})}{\sigma_k} \right) \approx 1 - \Phi\left(\frac{\sqrt{N}(q_\alpha^a - \mu_{m,s,k})}{\sigma_k} \right) = 1 - \Phi(z_\alpha).
\end{aligned}$$

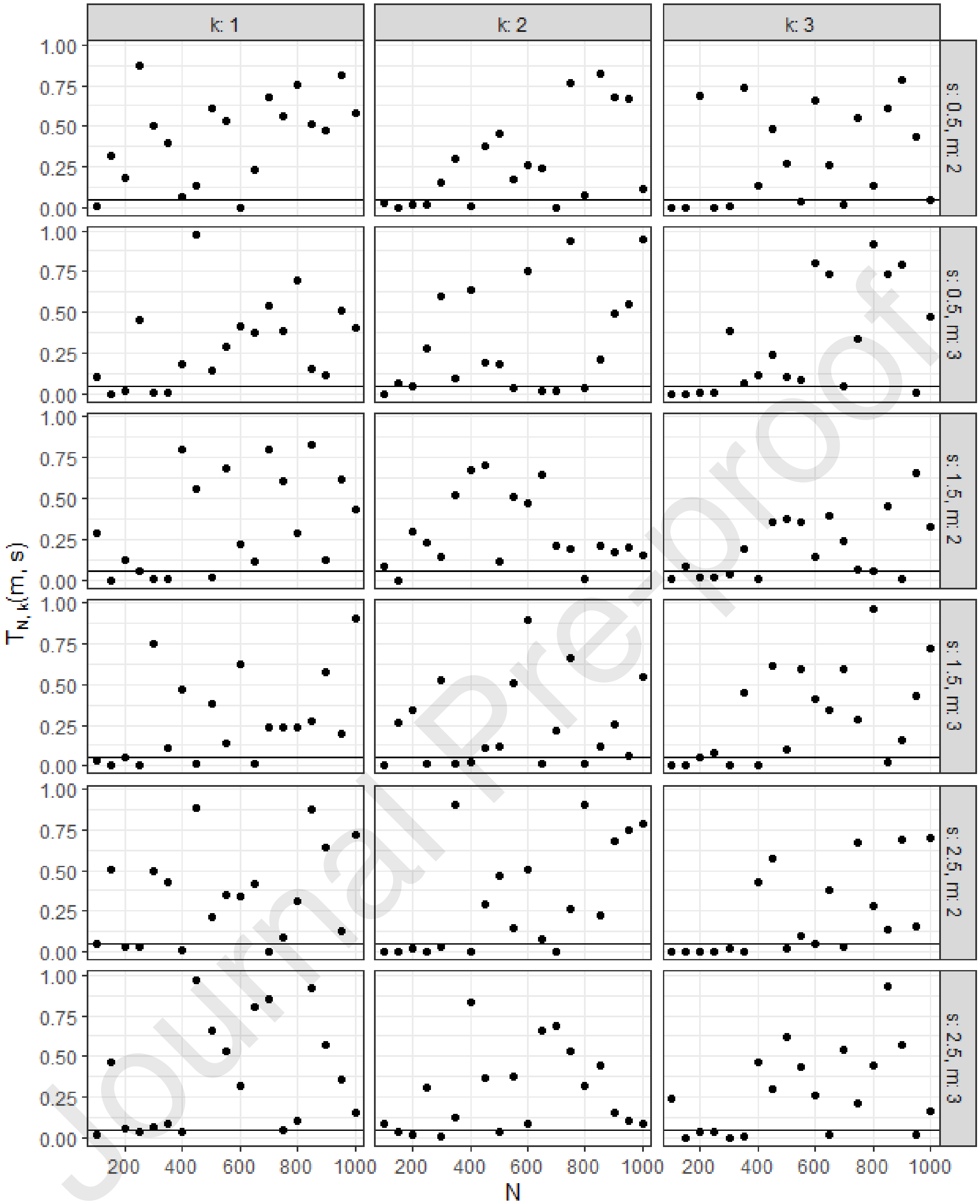


Figure 9: Shapiro-Wilk p-values as N increases for different values of m, s and k ($M = 1000$ repetitions).

We compare $\hat{q}_{0.05}$ and $q^a_{0.05}$ via the values of $\sqrt{N}(\hat{q}_{0.05} - z_{0.05}\sigma_k)$ and put them in Figure 10. We observe that these differences are bounded with small deviations for $N \in [400, 1000]$. We put $\mu_{m,s,k} =$

		$m=2$			$m=3$		
s	N	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
0.5	100	0.30793	0.24175	0.24822	0.36634	0.33337	0.32294
	200	0.21050	0.16076	0.15914	0.25667	0.22756	0.23070
	300	0.16142	0.13318	0.12532	0.19678	0.17996	0.18353
	400	0.14987	0.11753	0.10938	0.17643	0.16032	0.16511
	500	0.11999	0.09835	0.09504	0.16458	0.14597	0.14865
	600	0.11633	0.09082	0.08360	0.14080	0.12807	0.13425
	700	0.11017	0.08036	0.07208	0.14109	0.12629	0.12869
	800	0.09688	0.07448	0.06789	0.11715	0.11493	0.11691
	900	0.08891	0.07289	0.06407	0.11932	0.10248	0.10692
	1000	0.08601	0.06571	0.06167	0.11151	0.09898	0.10241
1.5	100	0.25888	0.21302	0.19695	0.31502	0.26798	0.24472
	200	0.20046	0.15151	0.14194	0.22006	0.18361	0.17399
	300	0.15430	0.12647	0.11199	0.18876	0.14910	0.14607
	400	0.13574	0.10170	0.09125	0.16631	0.13473	0.13628
	500	0.11918	0.08670	0.07895	0.14081	0.11325	0.11465
	600	0.11443	0.08309	0.07641	0.13692	0.11288	0.11011
	700	0.10600	0.07363	0.06833	0.12376	0.10341	0.09946
	800	0.09096	0.07745	0.06870	0.10651	0.09512	0.09291
	900	0.09028	0.06437	0.05851	0.11374	0.09236	0.09001
	1000	0.08621	0.06662	0.05823	0.10552	0.08675	0.08698
2.5	100	0.25793	0.19961	0.17907	0.30506	0.21805	0.18471
	200	0.19882	0.13606	0.12423	0.21229	0.16270	0.13955
	300	0.16689	0.11676	0.10050	0.18882	0.14901	0.13202
	400	0.13664	0.10649	0.09069	0.15503	0.11663	0.10525
	500	0.12374	0.08466	0.07320	0.14034	0.11124	0.10215
	600	0.10841	0.08133	0.07235	0.12984	0.10494	0.09515
	700	0.09978	0.07492	0.06539	0.12056	0.09383	0.08865
	800	0.09550	0.06888	0.06011	0.12023	0.09048	0.08349
	900	0.08935	0.06529	0.05861	0.10784	0.08451	0.07856
	1000	0.08663	0.06144	0.05170	0.10475	0.08237	0.07724

Table 2: Critical values of statistic $T_{N,k}(m,s)$ corresponding to significance level 0.05

$\max_{N\in[400,1000]}\sqrt{N}(\hat{q}_{0.05} - z_{0.05}\sigma_k)$ and present them in Table 3. In such case, $q^a_{0.05} \geq \hat{q}_{0.05}$ and the I-type error of a test $\mathbb{P}(T_{N,k}(m,s) > q^a_{0.05} | X \sim GG(m,s))$ based on approximated quantiles $q^a_{0.05}$ is less or equal 0.05 for $N \in [400, 1000]$, $s \in [0.5, 6]$, $k = 1, 2, 3$, and $m = 1, 2, 3$, which is confirmed by Figure 11a.

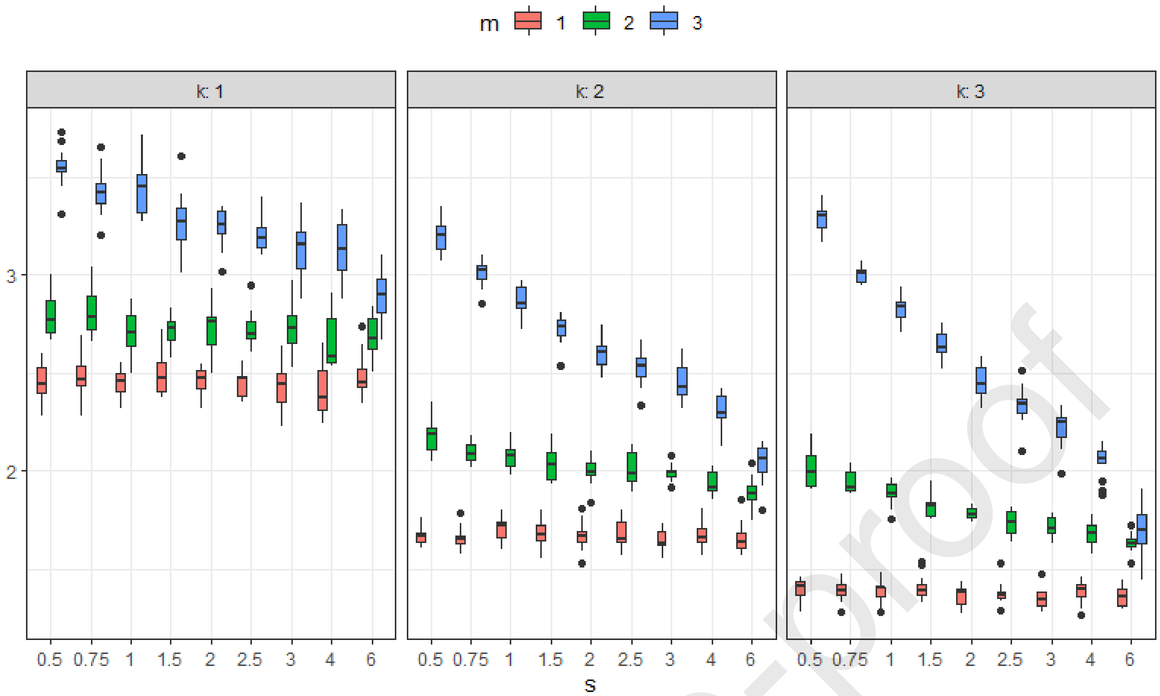


Figure 10: Values of $\sqrt{N}(\hat{q}_\alpha - z_\alpha \sigma_k)$ for $\alpha = 0.05$.

	$m=1$			$m=2$			$m=3$		
s	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
0.5	0.1297	0.1520	0.1817	0.5279	0.7406	0.9112	1.2633	1.7412	2.1283
0.75	0.2231	0.1732	0.1937	0.5690	0.5685	0.7610	1.1847	1.4941	1.7934
1	0.0816	0.1858	0.2037	0.4041	0.5870	0.6840	1.2457	1.3621	1.6662
1.5	0.2501	0.1907	0.2566	0.3601	0.5805	0.6669	1.1384	1.1986	1.4742
2	0.0732	0.2003	0.1612	0.4628	0.4952	0.5567	0.8820	1.1371	1.3024
2.5	0.0864	0.1863	0.2502	0.4766	0.5198	0.5375	0.9311	1.0595	1.2391
3	0.1683	0.1225	0.1999	0.5035	0.4658	0.5067	0.8932	1.0124	1.0566
4	0.1808	0.1930	0.1842	0.4400	0.4114	0.5017	0.8679	0.8074	0.8689
6	0.2685	0.2398	0.1649	0.3680	0.4282	0.4472	0.6333	0.5403	0.6336

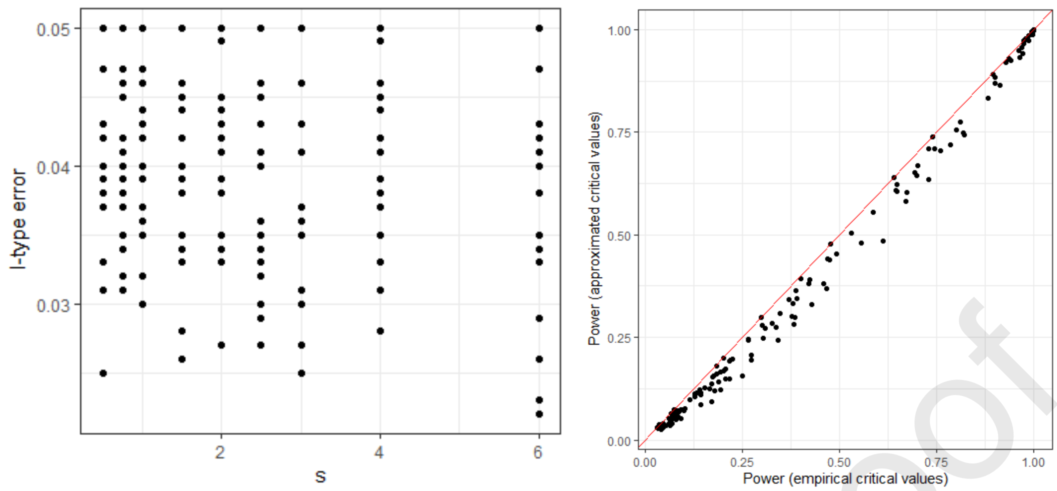
Table 3: Values of $\mu_{m,s,k}$ for $\alpha = 0.05$

6.5. Power of the goodness of fit test

We investigate also the power of the test to detect the alternative distribution. For this aim, we test

170 H_0: $X \sim GG(m, s)$ for a given $s = 1.5$ with unknown scale parameter τ_1, vs.

H_1: $X \sim GG(m, v)$ with $s \neq v > 0$ and unknown scale parameter τ_2



(a) (b)

Figure 11: (a) I-type error of the test based on approximated critical values. (b) Comparison of the test's power based on empirical and approximated critical values.

via test statistics $T_{N,k}(m,s)$ with the empirical critical values $\hat{q}_{0.05}$. We generate $M = 1000$ samples of size $N = 500, 1000$ under $GG(m,v)$ distribution, and compute $T_{N,k}(m,s)$ with $s = 1.5$ for each sample. The empirical power of the test is the ratio of number of rejections (when $T_{N,k}(m,s) > \hat{q}_{0.05}$) and M, see Figure 12. We definitely see, that the power is increasing with respect to k, N and the distance between $s = 1.5$ and v. Moreover, the power's grows for $v < s$ is significantly larger than for $v > s$. Thus, we recommend applying the k-nearest entropy neighbour estimator and test statistics with $k = 3$, which lead to the goodness of fit test with the maximum power.

We provide also the testing of H_0 vs. H_1 via rejection criteria $T_{N,k}(m,s) > q^a_{0.05}$ with approximated critical values $q^a_{0.05}$ given in (19). We compare the corresponding test's powers in Figure 11b and observe that the proposed approximation in (19) does not affect the test's power a lot. Hence, a researcher can avoid Monte Carlo simulations of empirical quantiles and use formula (19) for running our goodness of fit tests.

We apply the very same methodology on samples from the isotropic multivariate generalized Student t-distribution $GST(m,s,v)$, $v > 0$ on $\mathbb{R}^m$, which has the density function

$$f(x;m,s,v) = \frac{\Gamma\left[(v+m)/s\right]}{\Gamma(v/s)(v\pi)^{m/2}} \frac{\Gamma(m/2+1)}{\Gamma(m/s+1)} \left(1 + \frac{\|x\|^s}{v}\right)^{-(v+m)/2}, x \in \mathbb{R}^m,$$

see e.g. Lutwak et al. (2013). Namely, we test

H_0: $X \sim GG(m,s)$ for a given $s = 1.5$ with unknown scale parameter τ_1, vs.

H_1: $X \sim GST(m,s,v)$ with $v > 0$ and unknown scale parameter τ_2.

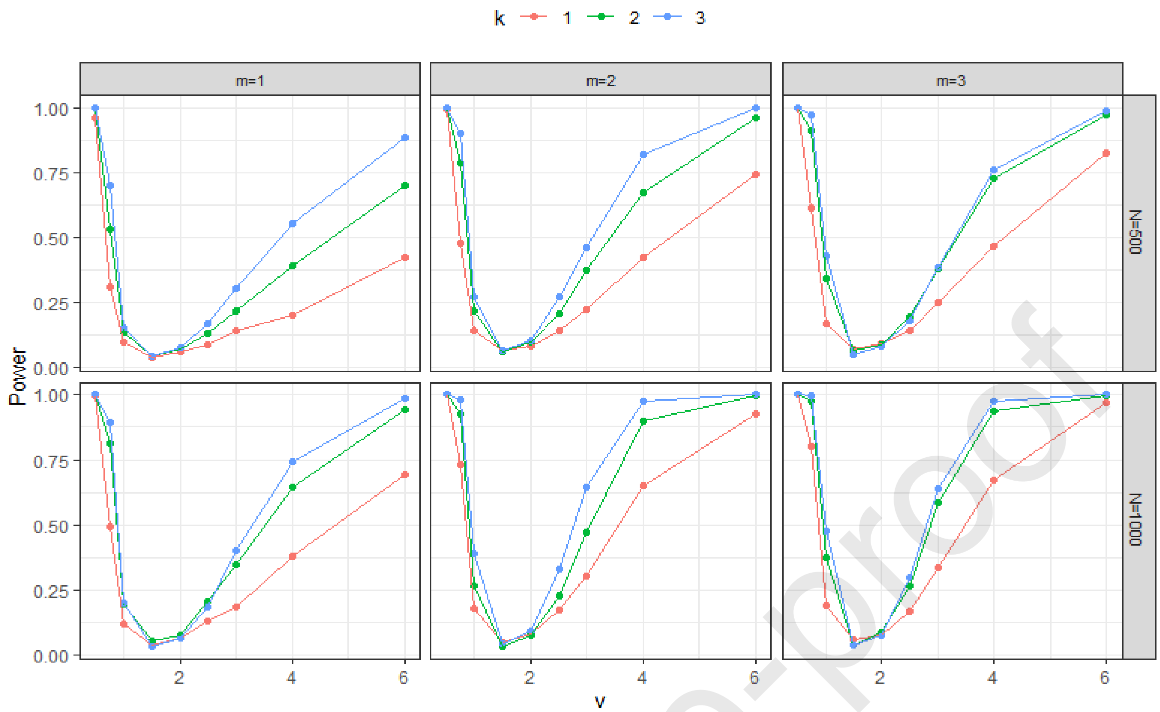


Figure 12: Empirical power function of the test for $s = 1.5$ on $GG(m, v)$ distributions ($M = 1000$ repetitions).

H_1: $X \sim GST(m, s, v)$ with $v > 0$ and unknown scale parameter τ_2.

In the Gaussian case, $GST(m, 2, v)$ is a multivariate t-distribution with v degrees of freedom. It is not hard to show the following stochastic representation.

Lemma 3. For $X \sim GST(m, s, v)$, we have $X \overset{d}{=} Z(V)^{-1/s}$, where $Z \sim GG(m, s)$ and $V \sim$ Gamma$(v/s, s/v)$.

Therefore, $GST(m, s, \infty) = GG(m, s)$ and the test's power must decrease with respect to v, which is confirmed by our Monte Carlo simulations. Their results are illustrated by Figure 13.

Acknowledgment

Nikolai Leonenko would like to thank Prof. Richard Samworth and Prof. Mathew Penrose for a fruitful discussion on a problem of entropy estimation on the Workshop 'Estimation of entropies and other functionals: Statistics meets information theory' on 9-11 September 2019, Cambridge (UK).

Vitalii Makogin is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 390879134. Nikolai Leonenko was partially supported by LMS grant 42997, ARC grant DP220101680, and the Isaac Newton Institute (Cambridge) Program "Fractional Differential Equations“.

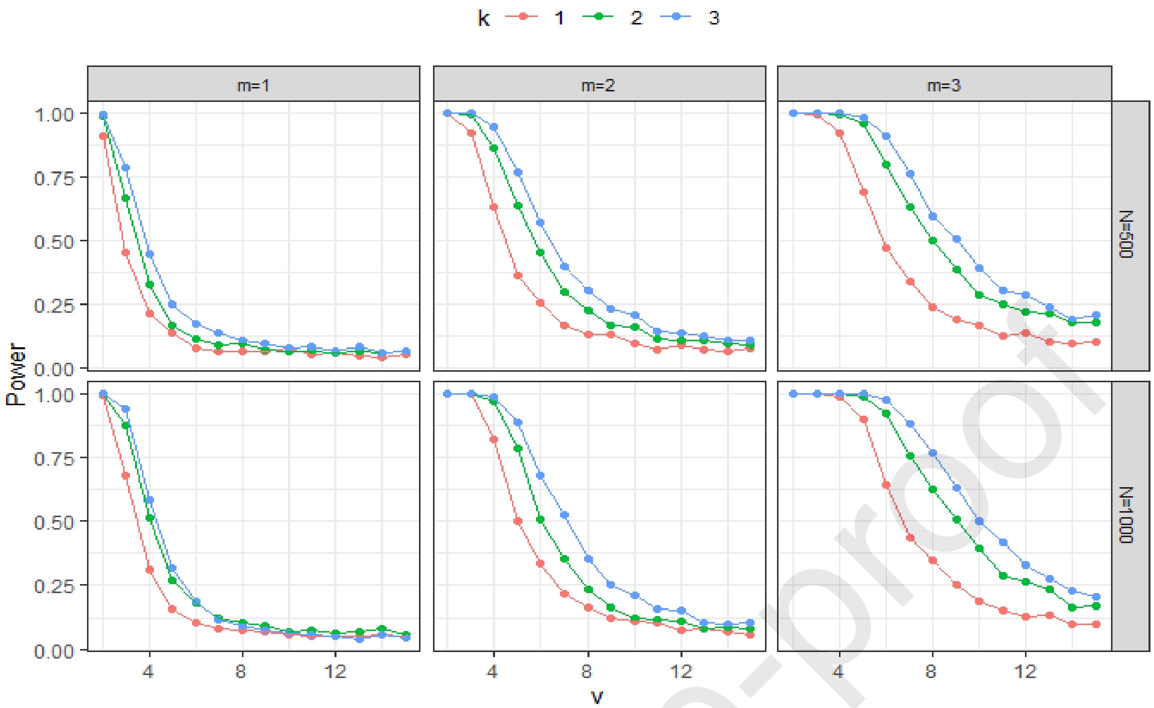


Figure 13: Empirical power function of the test for $s = 1.5$ on $GST(m, s, v)$ distributions ($M = 1000$ repetitions).

References

Aulogiaris, G., Zografos, K., 2004. A maximum entropy characterization of symmetric Kotz type and Burr multivariate distributions. Test 13, 65–83. URL: https://doi.org/10.1007/BF02603001, doi:10.1007/BF02603001.

Baccetti, V., Visser, M., 2013. Infinite Shannon entropy. J. Stat. Mech. Theory Exp. , P04010URL: https://doi.org/10.1088/1742-5468/2013/04/p04010, doi:10.1088/1742-5468/2013/04/p04010.

Barron, A.R., 1986. Entropy and the central limit theorem. Ann. Prob. 14, 336–342.

Beirlant, J., Dudewicz, E.J., Györfi, L., Van der Meulen, E.C., 1997. Nonparametric entropy estimation: An overview. International Journal of Mathematical and Statistical Sciences 6, 17–39.

Berrett, T.B., Samworth, R.J., 2019. Nonparametric independence testing via mutual information. Biometrika 106, 547–566.

Berrett, T.B., Samworth, R.J., Yuan, M., 2019. Efficient multivariate entropy estimation via k-nearest neighbour distances. The Annals of Statistics 47, 288–318.

Biau, G., Devroye, L., 2015. Lectures on the nearest neighbor method. Springer Series in the Data Sciences, Springer, Cham. URL: https://doi.org/10.1007/978-3-319-25388-6, doi:10.1007/978-3-319-25388-6.

Bobkov, S., Madiman, M., 2011. The entropy per coordinate of a random vector is highly constrained under convexity conditions. IEEE Transactions on Information Theory 57, 4940–4954.

Bulinski, A., Dimitrov, D., 2019. Statistical estimation of the Shannon entropy. Acta Mathematica Sinica, English Series 35, 17–46.

Bulinski, A., Dimitrov, D., 2021. Statistical estimation of the kullback–leibler divergence. Mathematics 9, 544.

Bulinski, A., Kozhevin, A., 2019. Statistical estimation of conditional Shannon entropy. ESAIM Probab. Stat. 23, 350–386. URL: https://doi.org/10.1051/ps/2018026, doi:10.1051/ps/2018026.

Choi, B., 2008. Improvement of goodness-of-fit test for normal distribution based on entropy and power comparison. Journal of Statistical Computation and Simulation 78, 781–788.

De Simoni, S., 1968. Su una estensione dello schema delle curve normali di ordine r alle variabili doppie. Statistica 37, 447–474.

Delattre, S., Fournier, N., 2017. On the Kozachenko–Leonenko entropy estimator. Journal of Statistical Planning and Inference 185, 69–93.

Devroye, L., Győfi, L., 2021. On the consistency of the Kozachenko-Leonenko entropy estimate. IEEE Transactions on Information Theory .

Dudewicz, E.J., Van Der Meulen, E.C., 1981. Entropy-based tests of uniformity. Journal of the American Statistical Association 76, 967–974.

Ebner, B., Henze, N., 2020. Rejoinder on: Tests for multivariate normality — a critical review with emphasis on weighted L^2-statistics. TEST 29, 911–913. URL: https://doi.org/10.1007/s11749-020-00744-w, doi:10.1007/s11749-020-00744-w.

Evans, D., 2008. A computationally efficient estimator for mutual information. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 464, 1203–1215.

Evans, D., Jones, A.J., Schmidt, W.M., 2002. Asymptotic moments of near–neighbour distance distributions. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 458, 2839–2849.

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric multivariate and related distributions. volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London. URL: https://doi.org/10.1007/978-1-4899-2937-2, doi:10.1007/978-1-4899-2937-2.

Fradelizi, M., Li, J., Madiman, M., 2020. Concentration of information content for convex measures. Electron. J. Probab. 25, 1–22. URL: https://doi.org/10.1214/20-ejp416, doi:10.1214/20-ejp416.

Gao, W., Oh, S., Viswanath, P., 2018. Demystifying fixed k-nearest neighbor information estimators. IEEE Transactions on Information Theory 64, 5629–5661.

Gnedenko, B.V., Kolmogorov, A.N., 1954. Limit distributions for sums of independent random variables. Addison-Wesley Publishing Company, Inc., Cambridge, Mass.

Gómez, E., Gomez-Viilegas, M., Marín, J.M., 1998. A multivariate generalization of the power exponential family of distributions. Communications in Statistics-Theory and Methods 27, 589–600.

González-Manteiga, W., Martínez-Miranda, M.D., Van Keilegom, I., 2016. Goodness-of-fit test in parametric mixed effects models based on estimation of the error distribution. Biometrika 103, 133–146. URL: https://doi.org/10.1093/biomet/asv061, doi:10.1093/biomet/asv061.

Goria, M.N., Leonenko, N.N., Mergel, V.V., Novi Inverardi, P.L., 2005. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. Journal of Nonparametric Statistics 17, 277–297.

Kano, Y., 1994. Consistency property of elliptic probability density functions. Journal of Multivariate Analysis 51, 139–147.

Kapur, J.N., 1989. Maximum-entropy models in science and engineering. John Wiley & Sons, Inc., New York.

Kozachenko, L.F., Leonenko, N.N., 1987. Sample estimate of the entropy of a random vector. Problems of Information Transmission 23, 95–101.

Leonenko, N.N., Pronzato, L., 2010. Correction: A class of Rényi information estimators for multidimensional densities. Annals of Statistics 38, 3837–3838.

Leonenko, N.N., Pronzato, L., Savani, V., 2008. A class of Rényi information estimators for multidimensional densities. The Annals of Statistics 36, 2153–2182.

Lord, W.M., Sun, J., Bollt, E.M., 2018. Geometric k-nearest neighbor estimation of entropy and mutual information. Chaos 28, 033114. URL: https://doi.org/10.1063/1.5011683, doi:10.1063/1.5011683.

Lutwak, E., Lv, S., Yang, D., Zhang, G., 2013. Affine moments of a random vector. IEEE Trans. Inform. Theory 59, 5592–5599. URL: https://doi.org/10.1109/TIT.2013.2258457, doi:10.1109/TIT.2013.2258457.

Lutwak, E., Yang, D., Zhang, G., 2007. Moment-entropy inequalities for a random vector. IEEE Transactions on Information Theory 53, 1603–1607.

Madiman, M., Nayar, P., Tkocz, T., 2021. Sharp moment-entropy inequalities and capacity bounds for symmetric log-concave distributions. IEEE Trans. Inform. Theory 67, 81–94. URL: https://doi.org/10.1109/TIT.2020.3032371, doi:10.1109/TIT.2020.3032371.

Marsiglietti, A., Kostina, V., 2018. A lower bound on the differential entropy of log-concave random vectors with applications. Entropy 20, 185.

Penrose, M.D., Yukich, J.E., 2011. Laws of large numbers and nearest neighbor distances, in: Advances in Directional and Linear Statistics. Springer, pp. 189–199.

Penrose, M.D., Yukich, J.E., 2013. Limit theory for point processes in manifolds. The Annals of Applied Probability 23, 2161–2211.

Rosenblatt, M., 2000. Gaussian and non-Gaussian linear time series and random fields. Springer Series in Statistics, Springer-Verlag, New York. URL: https://doi.org/10.1007/978-1-4612-1262-1, doi:10.1007/978-1-4612-1262-1.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.

Shin, S.J., Artemiou, A., 2017. Penalized principal logistic regression for sparse sufficient dimension reduction. Comput. Statist. Data Anal. 111, 48–58. URL: https://doi.org/10.1016/j.csda.2016.12.003, doi:10.1016/j.csda.2016.12.003.

Solaro, N., 2004. Random variate generation from multivariate exponential power distribution. Statistica & Applicazioni 2, 25–44.

Tsybakov, A.B., Van der Meulen, E., 1996. Root-n consistent estimators of entropy for densities with unbounded support. Scandinavian Journal of Statistics , 75–83.

Vasicek, O., 1976. A test for normality based on sample entropy. Journal of the Royal Statistical Society: Series B (Methodological) 38, 54–59.

Wyner, A.D., Ziv, J., 1969. On communication of analog data from a bounded source space. Bell System Technical Journal 48, 3139–3172.

Appendix A. Bounds on Shannon entropy

Below, we present some essentials about bounds of the Shannon entropy. First, we show that there exist densities such that $-\infty = H(f) < \infty$. We modify an example of (Gnedenko and Kolmogorov, 1954, p.223). For other examples, see Barron (1986).

Example 1. Let $m = 1$ and consider the density

$$f(x) = \left[x \log^2 \frac{e}{x}\right]^{-1} \mathbb{1}_{[0,1]}(x), \qquad x \in \mathbb{R}. \tag{A.1}$$

If X is random variable with density (A.1), then for $s = 1$

$$\mathbb{E}X = \mathbb{E}|X| = \int_0^1 \left[\log^2 \frac{e}{x}\right]^{-1} \mathrm{d}x = 1 - \mathrm{E}_1(1) \simeq 0.40365..., \tag{A.2}$$

where

$$\mathrm{E}_p(z) = z^{p-1}\Gamma(1-p,z) = z^{p-1}\int_z^\infty \frac{e^{-zt}}{t^p}\mathrm{d}t, \quad p > 0, \quad z \geq 0,$$

is the generalized exponential integral. Thus, by Theorem 1 with $m = 1$ and $s = 1$,

$$H(f) \leq \log\left[2e\mathbb{E}|X|\right] \simeq 0.8073.$$

From the other hand,

$$H(f) = -\int_0^1 \left[x \log^2 \frac{e}{x}\right]^{-1} \log\left[x \log^2 \frac{e}{x}\right]^{-1} \mathrm{d}x = -\infty.$$

Example 2. For $m \geq 2$, the similar properties has the density

$$f(x) = c_2(m)\left[\|x\|^m \log^2 \frac{e}{\|x\|}\right]^{-1} \mathbb{1}_{B_1(0)}(x), \quad x \in \mathbb{R}^m,$$

where $c_2(m) = \Gamma(\frac{m}{2})/(2\pi^{m/2})$. That is f has finite moments but $H(f) = -\infty$.

Second, we give an example of density with $H(f) = +\infty$.

Example 3. Let random variable X have density (A.1). Then the density of $Y = X^{-1}$ is

$$f(x) = \left[x \log^2(ex)\right]^{-1} \mathbb{1}_{[1,+\infty)}(x), \qquad x \in \mathbb{R} \tag{A.3}$$

and for any $s > 0$

$$\mathbb{E}|Y|^s = \int_1^{+\infty} \frac{x^s}{x(1+\log x)^2}dx = +\infty. \tag{A.4}$$

The entropy of f equals

$$H(f) = \int_1^{+\infty} \frac{\log x + 2\log(1+\log x)}{x(1+\log x)^2} dx = \int_1^{+\infty} \frac{dx}{x(1+\log x)} dx + \int_1^{+\infty} \frac{\log \frac{\log^2(ex)}{e}}{x\log^2(ex)} dx$$
$$= \int_1^{+\infty} \frac{dy}{y} dy + \frac{2}{\sqrt{e}} \int_{1/\sqrt{e}}^{+\infty} \frac{\log z}{z^2} dz = +\infty.$$

For further examples and conditions when an entropy is finite, see Baccetti and Visser (2013).

If a random vector X in $\mathbb{R}^m$ has a bounded density f with $\|f\|_\infty = \sup_{x\in\mathbb{R}^m} f(x) < \infty$, then there is a lower bound for its entropy (Bobkov and Madiman, 2011).

$$\frac{1}{m} H(f) \geq \log \|f\|_\infty^{-1/m}. \tag{A.5}$$

If, in addition, f is log-concave (that is, $\log f$ is concave), then

$$\log \|f\|_\infty^{-1/m} \leq \frac{1}{m} H(f) \leq 1 + \log \|f\|_\infty^{-1/m}.$$

Moreover, provided the existence of p-th moment $\mathbb{E}\|X\|^p < \infty$, $p \geq 1$, one has for a log-concave density f, see Marsiglietti and Kostina (2018),

$$H(f) \geq \frac{1}{p} \log \frac{2^p \mathbb{E}\|X - \mathbb{E}[X]\|^p}{\Gamma(1+p)}. \tag{A.6}$$

If $m = 1$, then for a symmetric log-concave random variable X

$$H(f) \geq \frac{1}{p} \log \frac{2^p \mathbb{E}\|X\|^p}{\Gamma(p+1)}, \quad p > -1. \tag{A.7}$$

If a symmetric log-concave random vector on $\mathbb{R}^m$ has finite second moments, then

$$H(f) \geq \frac{m}{2} \log \frac{(\det \Sigma_x)^{1/m}}{c_5(m)}, \tag{A.8}$$

where $\Sigma_x = \mathbb{E}\left[(X - \mathbb{E}X)(X - \mathbb{E}X)^T\right]$ denotes the the covariance matrix of X and

$$c_3(m) = \frac{e^2 m^2}{4\sqrt{2}(m+2)}. \tag{A.9}$$

Recently, bounds (A.6)–(A.8) have been improved by Madiman et al. (2021). Further useful inequalities on the Shannon entropy can be found in (Fradelizi et al., 2020).

Constant $c_3(m)$ can be improved in the case of unconditional random vectors. A function $f : \mathbb{R}^m \to \mathbb{R}^m$ is called *unconditional* if for every $(x_1, \ldots, x_m) \in \mathbb{R}^m$ and $(\varepsilon_1, \ldots, \varepsilon_m) \in \{-1, 1\}^m$, one has

$$f(\varepsilon_1 x_1, \ldots, \varepsilon_m x_m) = f(x_1, \ldots, x_m).$$

For example, the density of standard isotropic Gaussian vector is unconditional. Thus, if X is unconditional, symmetric, and log-concave, then

$$c_3(m) = e^2/2. \tag{A.10}$$

The constant (A.10) is better than the constant (A.9) for $m \geq 5$.