Entropy-based goodness-of-fit tests for multivariate distributions

Mehmet Siddik CADIRCI

2021

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

School of Mathematics
Ysgol Mathemateg

Abstract

Entropy is one of the most basic and significant descriptors of a probability distribution. It is still a commonly used measure of uncertainty and randomness in information theory and mathematical statistics. We study statistical inference for Shannon and Rényi's entropy-related functionals of multivariate Gaussian and Student-t distributions. This thesis investigates three themes. First, we provide a non-parametric test of goodness-of-fit for a class of multivariate generalized Gaussian distributions based on maximum entropy principle via using the k-th nearest neighbour (NN) distance estimator of the Shannon entropy. Its asymptotic unbiasedness and consistency are demonstrated. Second, we show a class of estimators of the Rényi entropy based on an independent identical distribution sample drawn from an unknown distribution f on $\mathbb{R}^m$. We focus on the maximum Rényi entropy principle for multivariate Student-t and Pearson type II distributions. We also consider the entropy-based test for multivariate Student-t distribution using the k-th *NN* distances estimator of entropy and employ the properties of entropy estimators derived from *NN* distances. Third, we introduce a new classes of unimodal rotational invariant directional distributions, which generalize von Mises-Fisher distribution. We propose three types of distributions in which one of them represents the axial data. We provide all of the formula together with a short computational study of parameter estimators for each new type via the method of moments and method of maximum likelihood. We also offer the goodness-of-fit test to detect that the sample entries follow one of the introduced generalized von Mises-Fisher distribution based on the maximum entropy principle using the k-th *NN* distances estimator of Shannon entropy and to prove its L^2-consistence.

"Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."

— Lewis Carroll, *Alice in Wonderland*

"To visualise 43*-dimensional space I simply visualise n-dimensional space and let n be* 43*"*

— Unknown

"I don't believe it. Prove it to me and I still won't believe it."

— Douglas Adams

Acknowledgements

Interdependence is a higher value than independence. First and most importantly, I would like to enlarge special thanks to my academic supervisor, Professor Nikolai Leonenko, for giving me intellectual freedom in my work and unwavering support throughout my research program. Over the last four years, I am grateful for giving me inspiration and wisdom by encouragement and support he gave me. During Covid-19, which affects all generation life all over the world, we are not able to work face-to-face. He always encouraged me; without his guidance and feedback, my research would not have been manageable.

Many special thanks also to Dr Dafydd Evans who have mentored me during our remarkable discussions about understanding and solving the simulation section. Many thanks again Dr Vitalii Makogin from Ulm University in Germany. I appreciate his assistance for sharing his knowledge and offering collaborative work on understanding the generalized von Mises-Fisher distribution. I am grateful for Prof. Katja Schladitz's help with the real data, Martin Gurka and Sebastian Nissle (Institut für verbundwerkstoffe, Kaiserslautern) for permission to use the tomographic images, Prof. Claudia Redenbach for providing the data set of fiber directions.

In addition to the mentioned crucial individuals above, I offer many thanks to my family and dear wife, Tugce Busra. They are always believing, supporting and encouraging me to reach my dreams. Heartfelt thanks to my parents, Ismail and Adile, who have always supported me in dealing with living abroad during this challenging period. Finally, I offer many thanks to older brothers, Aydin, Mehmet and Musa. They encouraged me to embark on this life journey in the first place.

Finally, I would like to thank my dear friends and entire academic staff from the School of Mathematics for their sincere, sympathetic and welcoming environment. Many special thanks to my peers in the School of Mathematics, principally Jack Noonan, who always gave me exceptional advice on some technical aspects of simulation, Clement Twumasi, Nikoleta Glynatsi, Álvaro Torras Casas, Geraint Palmer, Hassan Izanloo and Bernabe Gomez Perez.

Dissemination of Work

Publications

- 2020: M. S. Cadirci, D. Evans, N. N. Leonenko and V. Makogin. **Entropy-based test for generalized Gaussian distributions**. ArXiv preprint arXiv:2010.06284.— *in progress*.
- 2020: N. N. Leonenko, V. Makogin and M. S. Cadirci. **The entropy based goodness-of-fit tests for generalized von Mises-Fisher distributions and beyond**. ArXiv preprint arXiv:2010.10918— *submitted to Electronic Journal of Statistics*.
- 2020: M. S. Cadirci, D. Evans, N. N. Leonenko and O. Seleznjev. **Statistical tests based on Rényi entropy.** ArXiv preprint arXiv: 2106.00453 — *in progress*.

Presentations

- 2018: **University of Wales Intercollegiate Colloquium in Mathematics:** "**Entropy-based test for generalized Gaussian distributions**", *Gregynog, UK.*
- 2018: "**Goodness-of fit test for multivariate Student distribution based on Rényi entropy**", *SIAM Meeting, Cardiff University.*
- 2019: "**Entropy-based test for generalized Gaussian distributions**", *2019 Welsh Mathematical Colloquim, Gregynog Hall.*
- 2019: "**Statistical tests based on Rényi entropy**", *Cardiff University*.
- 2020: "**The entropy based goodness-of-fit tests for generalized von Mises-Fisher distributions and beyond**", *Cardiff University.*

x

- 2021: "**Entropy-based test for generalized Gaussian distributions**". *OR/Statistics Seminar, Cardiff University (online)*.

Contents

List of Figures

List of Tables

— Chapter 1 —

Introduction

The concept of entropy plays a central role in information theory, and has found a wide array of uses in other disciplines, including statistics, probability and combinatorics. The differential entropy or Shannon (or Bolzmann-Gibbs) entropy of a random vector $X \in \mathbb{R}^m$ with probability density function (pdf) f is defined as

$$H = H_1 = H(X) = H(f) = -\mathbb{E}\{\log f(X)\} = -\int_{\mathbb{R}^m} f(x) \log f(x) \mathrm{d}x, \tag{1.1}$$

where log is the natural logarithm. It is assumed that (1.1) is well defined and finite. It is formally set $0/0 = 0, \quad 0 \log 0 = 0$. For the probability density f it is defined $S = S(f) = \{x \in \mathbb{R}^m : f(x) > 0\}$ as its support. Clearly, the integral (1.1) is taken over $S(f)$. Introduced by the highly influential Shannon [110], it represents the average information content of an observation, and it is usually thought of as a measure of unpredictability.

More generally, Rényi entropy for a random vector $X \in \mathbb{R}^m$ with pdf f is defined as

$$H_q = H_q(X) = H_q(f) = \frac{1}{1-q} \log \mathbb{E}[f^{q-1}(X)] = \frac{1}{1-q} \log \int_{\mathbb{R}^m} f^q(x) \mathrm{d}x, \tag{1.2}$$

where $q \neq 1, q > 0$, see [103]. The Rényi entropy satisfies $H_q(X) \leq H_{q'}(X)$ for $q > q'$, and as $q \to 1$, the Rényi entropy converges to the Shannon entropy, $H_q(X) \to H_1(X)$.

Beside the Shannon and Rényi entropies, other entropy definitions (e.g. [119], [53], etc.) are studied in the mathematical literature, but the Shannon and Rényi entropies are the only ones possessing all the desired properties of an information measure, including *subadditivity*;

$$H(X, Y) \leq H(X) + H(Y) \tag{1.3}$$

with equality if and only if X and Y are independent, where

$$H(X,Y) = -\int_{\mathbb{R}^{m_1+m_2}} f(x,y)\log f(x,y)\mathrm{d}x\mathrm{d}y,$$

and $f(x,y)$ is the joint probability density of random vectors $X \in \mathbb{R}^{m_1}$, $m_1 \geq 1$ and $Y \in \mathbb{R}^{m_2}$, $m_2 \geq 1$.

The entropies are of interest in the work of non-linear Fokker-Planck equations [41], [123], as well as non-linear Markor processes [69].

Differential entropy (1.1) can be both positive or negative and can even be minus infinity ($H(f) = -\infty$, if $f(x) = (r/x)(-\log x)^{-(r+1)}$ on $0 \leq x \leq \frac{1}{e}$, $m = 1$, $0 < r \leq 1$, see Appendix A. For an overview of the properties of entropies (1.1) and (1.2) see, for example, [26] or [61].

Importantly, given the constraints on certain moments and the support set, one can find the distribution that maximises the Shannon or Rényi entropies [67]. This leads to the principle of maximum entropy, which has been applied in physics and Bayesian statistics, etc.

In statistical contexts, it is often the estimation of entropy that is of primary interest, for instance in goodness-of-fit tests of normality [122] or uniformity [27], tests of independence [49], etc.

There is an extensive literature dealing with entropy estimates, and in their classification, it will be used the scheme offered by [10, 58] or [94].

Non-parametric methods of entropy estimation in the univariate or multivariate cases include sample spacings, plug-in estimates based on a consistent estimate of density, splitting data estimates, cross-validation estimates, and estimates based on the partitioning of the observation space, (see [4, 16, 25, 32, 48, 51, 60, 65, 72, 95, 112, 117, 121], among others).

The estimation of entropy via the k-th nearest neighbour distances proposed by [71] is particularly attractive as a starting point, both because it can be generalized easily into multivariate cases, since it only relies on the evaluation of the k-th nearest neighbour distances, it is straightforward to compute.

There have been previous studies of the k-th nearest neighbour distances estimator of the Shannon and Rényi entropies, including [71] in the case of $k = 1$, as well as [39, 43, 49, 75, 113, 115, 120] and [14, 19, 30, 73, 81, 115], among others.

Modern efficient multivariate entropy estimation theory via the k-th nearest neighbour distances was developed recently by [13]. They studied the weighted averages of the estimators proposed by [71], based on the k-th nearest neighbour distances of a sample of N independent and identically distributed random vectors in $\mathbb{R}^m$. A careful choice of weights enabled them to obtain an efficient estimator in arbitrary dimensions, given sufficiently smoothness of density. They also investigated the case, when $k \to \infty$.

Berrett, Samworth, Yuan, et al. in [13] have shown that subject to certain regularity conditions, such as $k \to \infty$, the k-NNE of Shannon entropy is efficient only for $m \leq 3$, and presents a bias-corrected estimator for dimensions $m \geq 4$. In this thesis, the conventional k-NNE is focused with fixed $k \geq 1$, for which the asymptotic variance decreases rapidly up to $k = 3$ only, see [12, Table 1]. It is interesting to note that this asymptotic inflation is distribution-free, which leads to the conjecture that $k = 3$ is the most interesting case for any $m \geq 1$. For case $q = 2$, the different method of estimation of the quadratic Rényi entropy proposed by [64] is used. The new method of investigation of entropy estimates proposed by [9, 97] and [98] (see also the references therein).

For a discrete random variable X which takes values $\{l_K, K \geq 1\}$, the Shannon entropy is defined as

$$H(X) = H(f) = -\sum_{K} f_K \log f_K,$$

where $f_K = \mathbb{P}(X = l_K)$, $K \geq 1$, while the Rényi entropy is

$$H_q(X) = \frac{1}{1-q} \log \sum_{K} f_K^q, \quad q > 0, \quad q \neq 1.$$

A brief overview of the statistical estimate of entropies of discrete distributions can be found in [130] (see also [94]).

1.1 The research aims and objectives of this thesis

- To revise the maximum Shannon entropy principle for multivariate generalized Gaussian, Student and Pearson type II distributions.

- To prove the convergence in the mean-square of the k-th nearest neighbour estimate of the Shannon entropy for multivariate densities with possible unbounded support for arbitrary fixed $k \geq 1$. Penrose and Yukich [99] proved the statement only for $k = 1$.

- To construct the non-parametric test of goodness-of-fit for a multivariate generalized Gaussian distribution based on the maximum entropy principle. To support the test by the Monte Carlo simulation.

- To prove the convergence in the mean-square of the k-th nearest neighbour estimate of the Rényi entropy for multivariate densities with possible unbounded support for arbitrary fixed $k \geq 1$.

- To construct the non-parametric test of goodness-of-fit for multivariate Student and Pearson type II (or Barenblatt) distributions based on the maximum entropy principle. To support the test by the Monte Carlo simulation.

- To prove the maximum entropy principle for generalized von Mises-Fisher distribution on sphere.

- To prove the convergence of the k-th nearest neighbour estimate of the Shannon entropy of spherical distributions.

- To develop the goodness-of-fit test for the generalized von Mises-Fisher distribution. To support the test by the Monte Carlo simulation. To apply the test to real life data.

1.2 The organization of the thesis:

The chapters in this thesis are organized as follows:

- Chapter 1 contains the review of the Shannon and Rényi entropies and statistical methods of their estimation.

- Chapter 2 presents a non-parametric test of goodness-of-fit for a class of multivariate generalized Gaussian distributions based on the maximum Shannon entropy principle and the k-th nearest neighbour distances method of Shannon entropy. It contains the proof of L^2 consistency of the k-th nearest neighbour distance estimate for arbitrary fixed $k \geq 1$. In [99], the consistency is proven for $k = 1$ only. The results will be supported by a Monte Carlo simulation.

- Chapter 3 presents a non-parametric test of goodness-of-fit for a classes of multivariate Student and Pearson type II (or Barenblatt) distributions based on the maximum Rényi entropy principle and consistency of the k-th nearest neighbour distance estimate for arbitrary fixed $k \geq 1$. The results will be supported by a Monte Carlo simulation.

- Chapter 4 introduces three generalizations of the von Mises-Fisher distribution on a sphere and the maximum Shannon entropy principle for them. Based on the L^2-consistency of the k-th nearest neighbour estimate of Shannon entropy based on the sample of directional data, the goodness-of-fit tests are constructed for three classes of generalized von Mises-Fisher distributions. The results will be supported by a Monte-Carlo simulation and applied to real data (of local fiber directions in a glassfibre reinforced composite material).

- Chapter 5 summarises the results of the previous chapters, and indicates possible directions of future work, identifying further research questions that have arisen.

- The thesis contains several Appendices.

The work in this thesis has been submitted as three peer-reviewed papers:

- Cadirci et al. (2020) [21] which corresponds to work in Chapter 2.

- Cadirci et al. (2021) [22] which corresponds to work in Chapter 3.

- Leonenko et al. (2020) [74] which corresponds to work in Chapter 4.

— Chapter 2 —

Entropy-based test of goodness-of-fit for generalized Gaussian distributions

This chapter proposes a new entropy-based goodness-of-fit test based on the maximum Shannon entropy principle for the generalized multivariate Gaussian distribution. The chapter is structured as follows:

- Section 2.1 defines the k-th nearest neighbour estimate of the Shannon entropy.
- Section 2.2 introduces the multivariate generalized Gaussian distribution.
- Section 2.3 reviews a maximum entropy principle for the generalized Gaussian distribution.
- Section 2.4 presents the associated goodness-of-fit statistic for the generalized Gaussian distribution.
- Section 2.5 includes the numerical results with some auxiliary material deferred to Appendix A.

2.1 Entropy estimation

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

$$\widehat{H}_{N,k} = \frac{1}{N}\sum_{i=1}^{N}\log\left[\rho_k^m(X_i,\mathcal{X}_N)V_m(N-1)e^{-\psi(k)}\right], \tag{2.1}$$

where $\psi(x)=\Gamma'(x)/\Gamma(x)=\sum_{j=1}^{k-1}\frac{1}{j}-\gamma$ 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

$$\widehat{H}_{N,1} = \frac{m}{N}\sum_{i=1}^{N}\log\rho_1(X_i,\mathcal{X}_N)+\log V_m+\gamma+\log(N-1), \tag{2.2}$$

where $\gamma=-\psi(1)\approx 0.577216$ is the Euler-Mascheroni constant. The estimator (2.2) was introduced by [71] while the general estimator (2.1) was first considered by [49].

The main properties of (2.1) have been studied by [12, 13, 18, 19, 30, 44, 75, 76, 98].

Convergence in mean-square for the case $k=1$ was proved by [99, Theorem 2.1.ii].

Theorem 2.1.1 ([99]). *Suppose that $\mathbb{E}(\|X\|^\alpha)<\infty$ for some $\alpha>0$ and $f(x)\le M$ for some $M>0$. Then*

$$\mathbb{E}\left[\widehat{H}_{N,1}-H(f)\right]^2\to 0 \quad \text{as } N\to\infty.$$

Remark 2.1.1. *The condition of boundedness for the density f is not explicitly stated by Penrose and Yukich (2013)[99, Theorem 2.4.ii]. In Appendix A(called lower bound on Shannon entropy) we give an example of a density f with bounded support and for which $H(f)$ is unbounded.*

In this chapter, the analogous result for arbitrary $k\ge 1$ is proved. To this end, (2.1) is written 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

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

First, the following theorem of [99, Theorem 3.1] is required.

Theorem 2.1.2. *Let $k\ge 1$ and $q=1$ or $q=2$, and suppose there exists $p\ge q$ such that*

$$\sup_{N\ge k}\mathbb{E}\left|l\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathcal{X}_N\right)\right|^p<\infty. \tag{2.3}$$

Then, L^q convergence can be written,

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

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

Theorem 2.1.3 (Main theorem). *Suppose that $\mathbb{E}\|X\|^\alpha < \infty$ for some $\alpha>0$ and $f(x)\le M$ for some $M>0$. Then for any fixed $k\ge 1$,*

$$\mathbb{E}\left[\widehat{H}_{N,k} - H(f)\right]^2 \to 0 \quad \text{as } N\to\infty. \tag{2.4}$$

Proof. From applying the Theorem 2.1.2. Firstly, it is shown that

$$H(f) = \int_{\mathbb{R}^m} \mathbb{E}\left[l(0,\mathscr{P}_{f(x)})\right] f(x)\,dx,$$

where

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

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 $\rho_k(0,\mathscr{P}_\lambda)$ is the distance from 0 to the k-th point of $\mathscr{P}_\lambda$, and thus has Erlang distribution with parameters k and $\lambda|B_t(0)| = \lambda V_m t^m$, that is

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

Then,

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

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

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

Second, the condition (2.3) is checked. Note that for every $\delta \in (0,1)$ and $p > 1$ there exists $C > 0$ such that [99, p.2206]

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

Then because

$$\begin{aligned}|l(x,\mathscr{X})|^p &= \left|\log V_m - \psi(k) + \log\rho_k^m(x,\mathscr{X})\right|^p \\ &\le \left|\log V_m - \psi(k)\right|^p + \left|\log\rho_k^m(x,\mathscr{X})\right|^p\end{aligned}$$

we have

$$\begin{aligned}&\frac{1}{2^{p-1}}\mathbb{E}\left|l\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right|^p \\ &\le |\log V_m - \psi(k)|^p + \mathbb{E}\left|\log\rho_k^m\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right|^p \\ &\le |\log V_m - \psi(k)|^p \\ &+ C\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\mathbb{1}_{[0,1]}\left[\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right] \qquad (2.5) \\ &+ C\mathbb{E}\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\mathbb{1}_{[1,\infty)}\left[\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right]. \qquad (2.6)\end{aligned}$$

Term (2.5) is finite because

$$\begin{aligned}&\sup_{N\ge k}\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\mathbb{1}_{[0,1]}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right) \\ &\le \sup_{N\ge k}\mathbb{E}\rho_1^{-\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right) < \infty, \qquad (2.7)\end{aligned}$$

where (2.7) is ensured by [99, Lemma 7.5] since f is bounded and $\delta \in (0,m)$.

Let $r_c(f) := \sup\{r \ge 0 : \mathbb{E}\|X_1\|^r < \infty\}$. In the proof of [98, Theorem 2.3], it is seen that if $r_c(f) > 0$ and $0 < \delta < mr_c(f)(m + r_c(f))^{-1}$, then

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

Thus, term (2.6) is finite. □

Remark 2.1.2. *For $k = 1$, [99] used exponential distribution. For $k > 1$, the Erlang distribution is used.*

2.2 The generalized Gaussian distribution

Testing for multivariate normality is a topic of ongoing interest, see [37] for a review of new developments. We denote $a^T b = \Sigma_{j=1}^m a_j b_j$ as the scalar product of vectors $a, b \in \mathbb{R}^m$.

The *multivariate exponential power distribution* $MEP_m(s, \mu, \Sigma)$ on $\mathbb{R}^m$ [114] has the density function

$$f(x\,; m, s, \mu, \Sigma) = \frac{\Gamma(m/2+1)}{\pi^{m/2}\Gamma(m/s+1)2^{m/s}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}\Big[(x-\mu)^T\Sigma^{-1}(x-\mu)\Big]^{s/2}\right), \tag{2.8}$$

where $\mu \in \mathbb{R}^m$ is the mean vector, Σ is an $m \times m$ positive definite matrix, $s > 0$ is a shape parameter [114], and variance-covariance matrix $V = \beta\Sigma$ where

$$\beta(m, s) = \frac{2^{2/s}\Gamma\big[(m+2)/s\big]}{m\Gamma(m/s)}. \tag{2.9}$$

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. 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 $\mathbb{R}^m$,

$$f(x\,; m, s) = \frac{\Gamma(m/2+1)}{\Gamma(m/s+1)\pi^{m/2}2^{m/s}} \exp\left(-\frac{1}{2}\|x\|^s\right), \tag{2.10}$$

$x \in \mathbb{R}^m$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$.

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

$$f_c(x; m, s) = c(m, s)\exp(-\tau\|x\|^s),\, x \in \mathbb{R}^m, \tag{2.11}$$

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

$$f(x; m, s) = c_0(m, s)\exp\left(-\frac{\|x\|^s}{s}\right), x \in \mathbb{R}^m, \tag{2.12}$$

where

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

Moments

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*, where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^m$. If X is isotropic and $g : \mathbb{R} \to \mathbb{R}$ is a Borel function, [99]

$$\mathbb{E}[g(\|X\|)] = \int_{\mathbb{R}^m} g(\|x\|) f(x)\, dx = \frac{2\pi^{m/2}}{\Gamma(m/2)} \int_0^\infty g(r) \tilde{f}(r) r^{m-1}\, dr \tag{2.13}$$

provided the integrals exist. In particular, the moments of order $s > 0$ are given in [75] and [116] by

$$\mathbb{E}(\|X\|^s) = \frac{2\pi^{m/2}}{\Gamma(m/2)} \int_0^\infty r^{m+s-1} \tilde{f}(r)\, dr \tag{2.14}$$

provided the integrals exist.

Lemma 2.2.1. *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 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 (2.14) it is obtained that

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

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

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

□

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

It is well known [66] 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 differential entropy $H(f)$ is maximised by the multivariate Gaussian distribution $N(0,\Sigma)$ on $\mathbb{R}^m$, and thus

$$H(f) \le \log\left[(2\pi e)^{m/2}\sqrt{\det\Sigma}\right]. \tag{2.15}$$

An analogous result for the generalized Gaussian distribution is now proven.

Theorem 2.3.1. *Let $X \in \mathbb{R}^m$ be a random vector, whose density f is supported on the whole of $\mathbb{R}^m$, and for which there exists some $s > 0$ such that $\mathbb{E}(\|X\|^s) < \infty$. Then $H(f)$ is finite and satisfies*

$$H(f) \leq \frac{m}{s} \log\big(c_1(m,s)\mathbb{E}\|X\|^s\big),$$

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, it is observed that

$$H(f) \leq -\int_{\mathbb{R}^m} f(x) \log f^*(x)\, dx, \tag{2.16}$$

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 $\mathbb{E}\|Z\|^s = \mathbb{E}\|X\|^s$) it is achieved that

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

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

$$\begin{aligned}
&-\int_{\mathbb{R}^m} f(x)\log f^*(x)dz \\
&= \tau \int_{\mathbb{R}^m} \|x\|^s f(x)\,dx - (\log c(m,s)) \int_{\mathbb{R}^m} f(x)\,dx \\
&= \tau\, \mathbb{E}\|X\|^s - \log c(m,s) \\
&= \frac{m}{s} - \log c(m,s) \quad \text{by Lemma 2.2.1.}
\end{aligned}$$

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 (2.16) is valid as well.

Thus by (2.16) and substituting for $c(m,s)$ it is obtained that

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

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

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

Therefore,

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

□

Remark 2.3.1. *Theorem* 2.3.1 *was proved for $m = 1$ in [127] and [107, p.103-104]. For $m \ge 1$, some statements of Theorem 2.3.1 were also proved using other methods in [84].*

2.4 A test statistic for $GG(m,s)$

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

1. $\operatorname{supp}(f) = \mathbb{R}^m$,
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 2.4.1. *The density functions of the $GG_\tau(m,s)$ belong to $\mathscr{K}$ for all $m \geq 1$, $s > 0$, $c > 0$ and $k \geq 1$.*

Proof. The statement follows from Theorem 2.1.3, which applies because f is bounded, and Lemma 2.2.1. □

Let $X \in \mathbb{R}^m$ be a random vector with density $f \in \mathscr{K}$, and let $s > 0$ be fixed. Based on a random sample $X_1, X_2, \dots$ from the distribution of X, the maximum entropy principle proved in Section 2.3 is used to test the hypothesis $X \sim GG(m,s)$ against a suitable alternative. By Theorem 2.3.1, 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

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

The entropy $H(X)$ is estimated by the k-th nearest neighbour estimator

$$\widehat{H}_{N,k} = \frac{m}{N}\sum_{i=1}^{N}\log\rho_k(X_i,\mathscr{X}_N) + \log V_m + \log(N-1) - \psi(k),$$

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

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

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 $k \in \{1, \ldots, N-1\}$ it is obtained that

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

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 negative. Thus, the hypothesis $X \sim GG(m,s)$ is rejected whenever $T_{N,k} \geq t_{N,k,\alpha}$, where $t_{N,k,\alpha} = t_{N,k,\alpha}(m,s)$ 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 $T_{N,k}$ and $\bar{X}_N^{(s)}$ are intractable, even though the *asymptotic* behaviour of $\widehat{H}_{N,k}$ can be revealed by applying results of [30, 98] or [13] and the asymptotic behaviour of $\bar{X}_N^{(s)}$ by the delta method. Thus, the Monte Carlo simulation is used to investigate the distribution of $T_{N,k} = T_{N,k}(m,s)$ for different combinations of parameter values.

Remark 2.4.1. *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*

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

2.5 Numerical results

To investigate the behaviour of $T_{N,k}(m,s)$, random samples are generated not only from the $GG(m,s)$ distribution, but also from the *multivariate Student distribution* $ST(m, \nu)$, $\nu > 0$ on $\mathbb{R}^m$ which has density function

$$f(x : m, \nu) = c_S |\Sigma|^{-1/2} \left(1 + \frac{1}{\nu}(x-a)^T \Sigma^{-1}(x-a)\right)^{-\frac{\nu+m}{2}}, \quad x \in \mathbb{R}^m,$$

where

$$c_S(m,\nu) = \frac{\Gamma[(\nu+m)/2]}{(\pi\nu)^{m/2}\Gamma(\nu/2)}.$$

This is achieved via the following stochastic representations [114].

Lemma 2.5.1. *1. 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 $V \sim Gamma(m/s, 2)$.*

2. For $X \sim ST(m,\nu)$ we have $X \stackrel{d}{=} Z/\sqrt{G}$, where $Z \sim N(0, I_m)$ and $G \sim Gamma(\nu/2, \nu/2)$.

For the case $m = 2$, the generated points are put on scatter plots for different values of s and ν, see Figure 2.1 for $GG(m,s)$ and Figure 2.2 for $ST(m,\nu)$. One can observe that visually distributions $GG(m,s)$ and $ST(m,\nu)$ are hardly distinguishable. Therefore, our goodness-of-fit test is applied for detecting of the generalized Gaussian distribution.

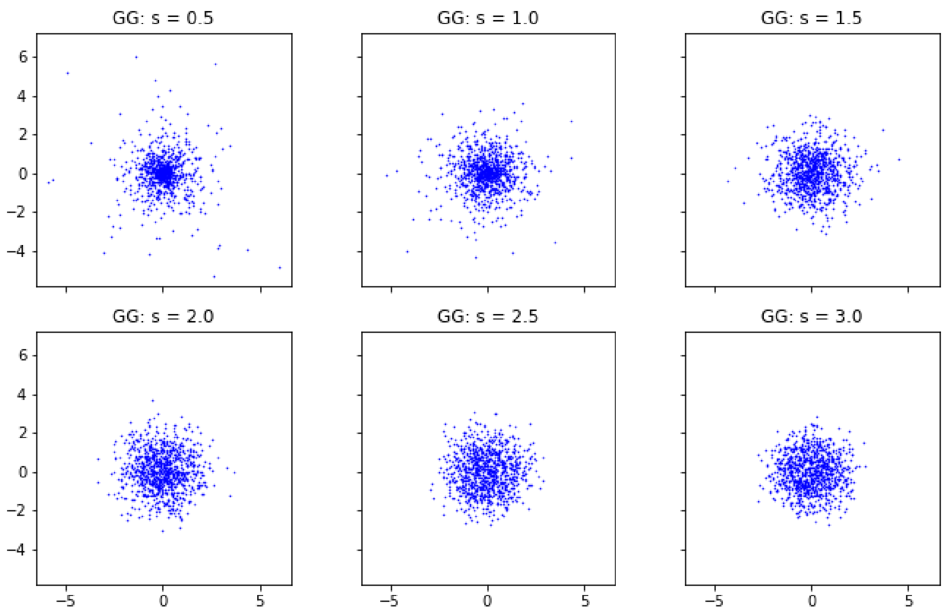


Figure 2.1: Scatter plots of 1000 random points from $GG(m,s)$ with $m = 2$

2.5.1 Empirical distribution of $GG(m,s)$

$N = 10^6$ points are generated from the $GG(m,s)$ distribution for different values of s. For the purpose of comparison the scaling $X \mapsto X/\sigma$ is applied where

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

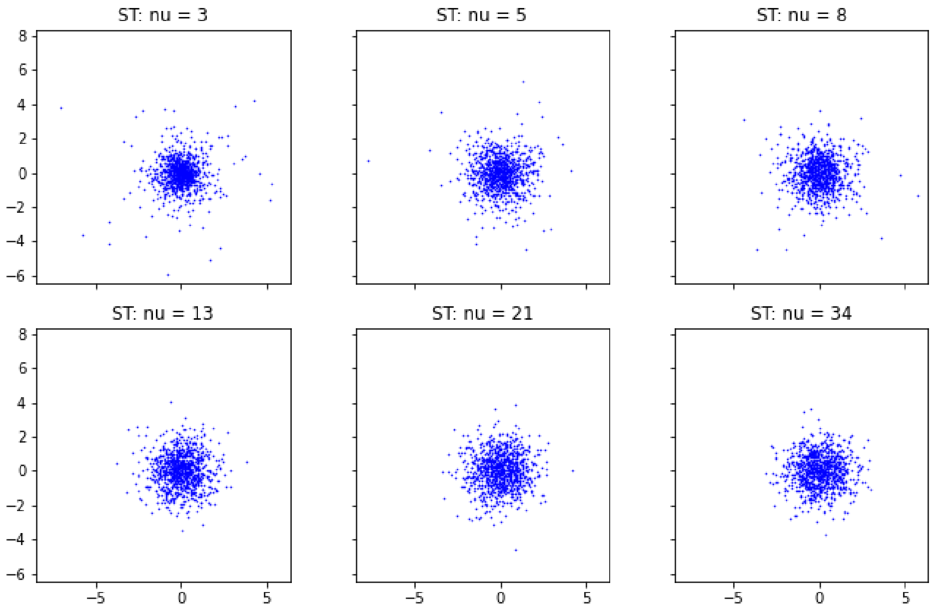
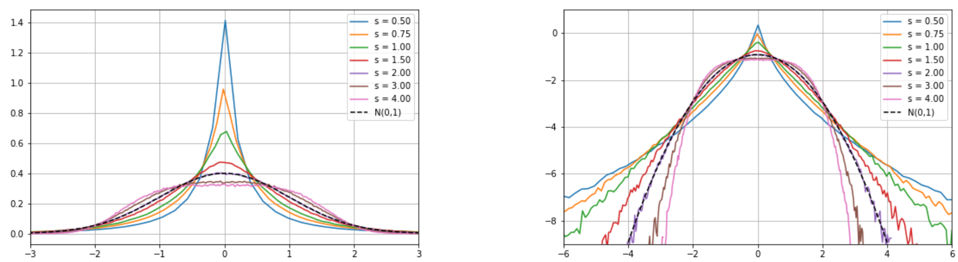


Figure 2.2: Scatter plots of 1000 random points from $ST(m, \nu)$ with $m = 2$

is the variance of the $GG(m, s)$ distribution. The test results are shown in Figure 2.3.



(a) (b)

Figure 2.3: The figure (a) contains empirical distribution of $GG(m, s)$ for $m = 1$ and different values of s. The figure (b) depicts the log-density function.

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

For fixed (N,k) and (m,s), a sample of size N from the $GG(m,s)$ distribution is generated and the empirical value of $T_{N,k}(m,s)$ is recorded, repeating this $M = 100$ times. This yields a sample realisation $\{T_1, T_2, \ldots, T_M\}$ from the distribution of $T_{N,k}(m,s)$, from which we estimate its mean and variance by

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

Figure 2.4: Consistency of $T_{N,k}(m,s)$ for different values of k ($M = 100$ repetitions).

Figures 2.4 demonstrates how $\bar{T}_{N,k}(m,s)$ approaches 0 as the number of sample size, N, increases for various values of $m \in \{1,2,3\}$, $s \in \{0.5, 1, 1.5, 2, 2.5\}$ and $k = 1,2,3$ corresponding to the standard error $S_{N,k}$. The mean statistics for $k = 1,2,3$ are shown in Figures 2.4.

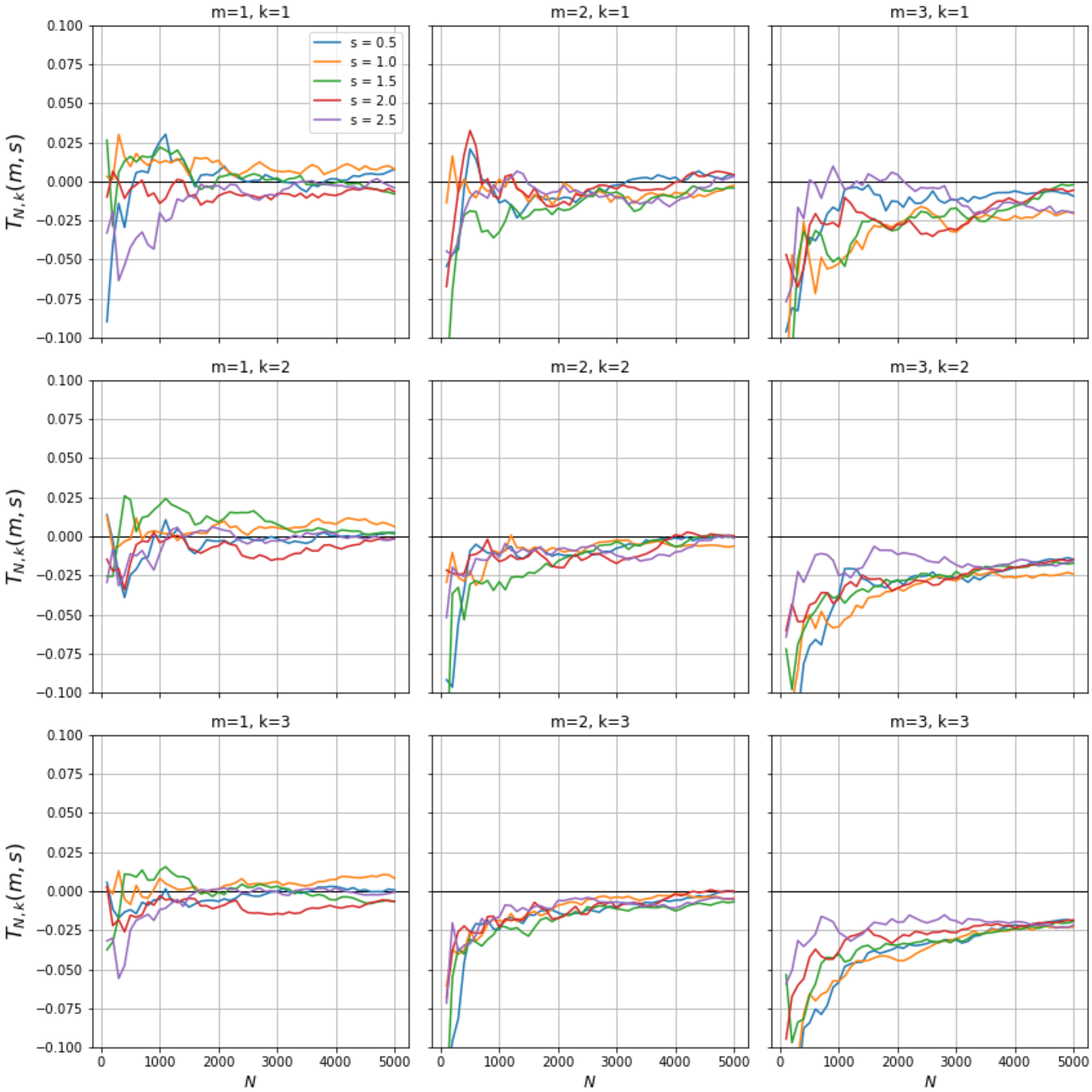


Figure 2.5: Consistency of $T_{N,k}(m,s)$ for different values of s ($M = 100$ repetitions).

Figure 2.5 shows how $\bar{T}_{N,k}(m,s)$ approaches 0 as N increases for various values of m, s and k. This experiment is repeated for the null distribution of $T_{N,k}(m,s)$. It can also be seen that the statistic of $T_{N,k}(m,s)$ increases with parameter value k, and decreases with the dimension m.

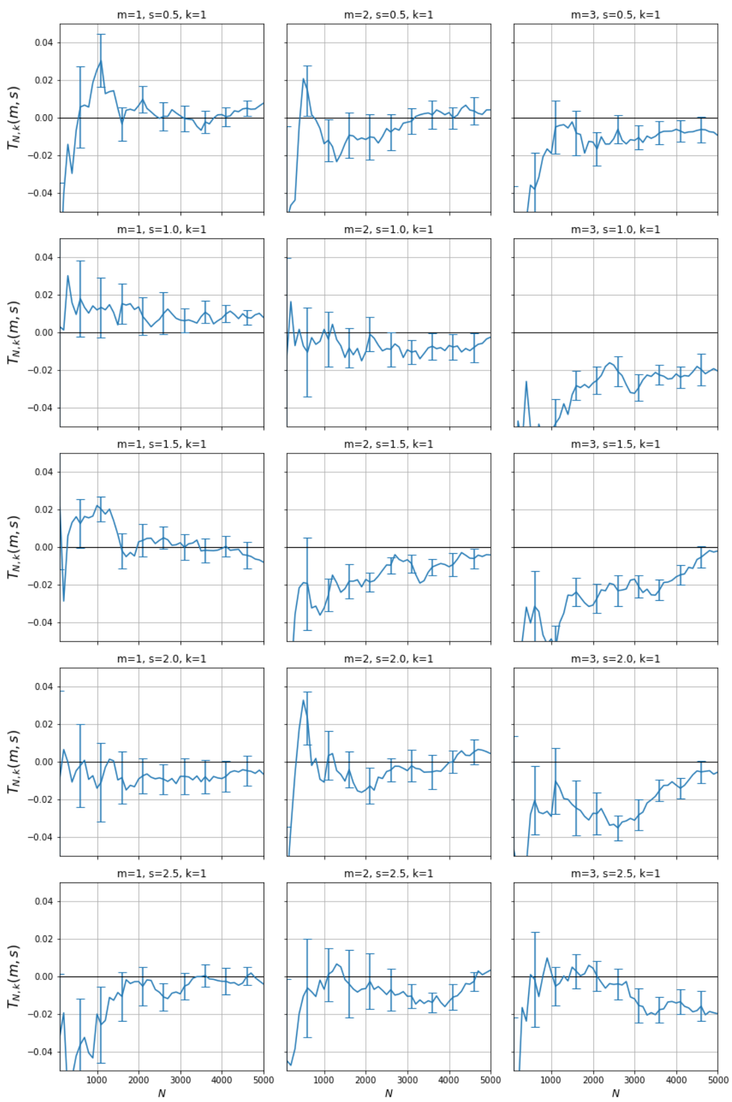


Figure 2.6: Consistency of $T_{N,k}(m,s)$ for $k=1$ ($M=100$ repetitions).

Figure 2.6 shows its behaviour for $m \in \{1,2,3\}$, $s \in \{0.5, 1, 1.5, 2, 2.5\}$ and $k=1$ with error bars corresponding to the standard error $S_{N,k}/\sqrt{M}$. From these data it is 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-th nearest neighbour estimators observed in [13].

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

For various values of s_0 and s_1, samples from the $GG(m, s_1)$ distribution are generated and the behaviour of $T_{N,k}(m, s_0)$ is examined as N increases. The results are shown in Figure 2.7 and Figure 2.8. When $s_0 \neq s_1$ we see that the statistic approaches a strictly negative value, and that this becomes increasingly negative as the difference between s_0 and s_1 increases.

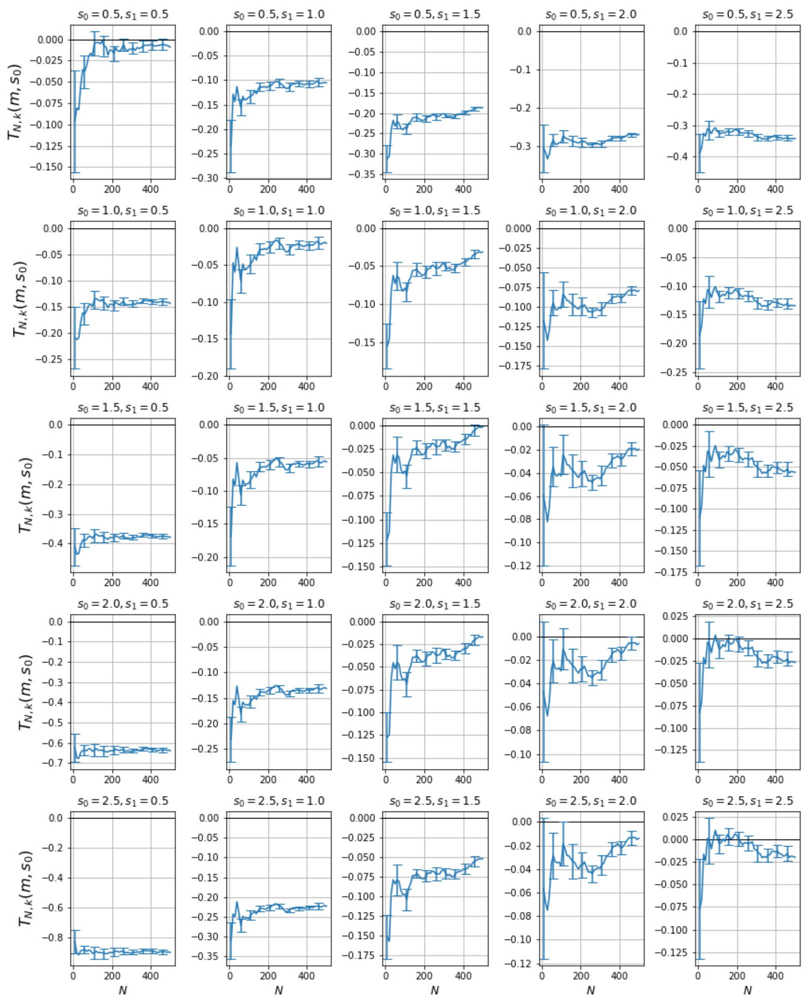


Figure 2.7: The behaviour of $T_{N,k}(m, s_0)$ with $k = 1$ on data from the $GG(m, s_1)$ distribution with $m = 2$.

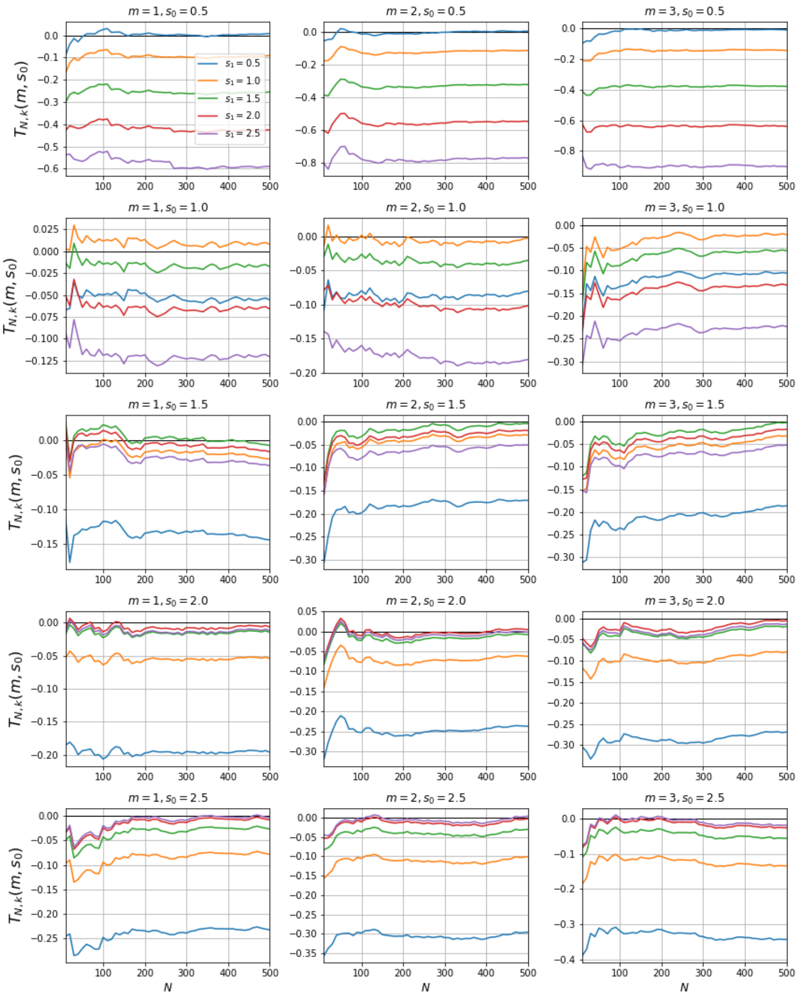


Figure 2.8: The behaviour of $T_{N,k}(m, s_0)$ with $k = 1$ on data from the $GG(m, s_1)$ distribution for different values of m

Asymptotic behaviour of $T_{N,k}(m, s)$ on data from $ST(m, \nu)$

For various values of s, we generate samples from the $ST(m, \nu)$ distribution are generated and the behaviour of $T_{N,k}(m, s)$ is examined as N increases. The outcomes of the test are shown in Figure 2.9.

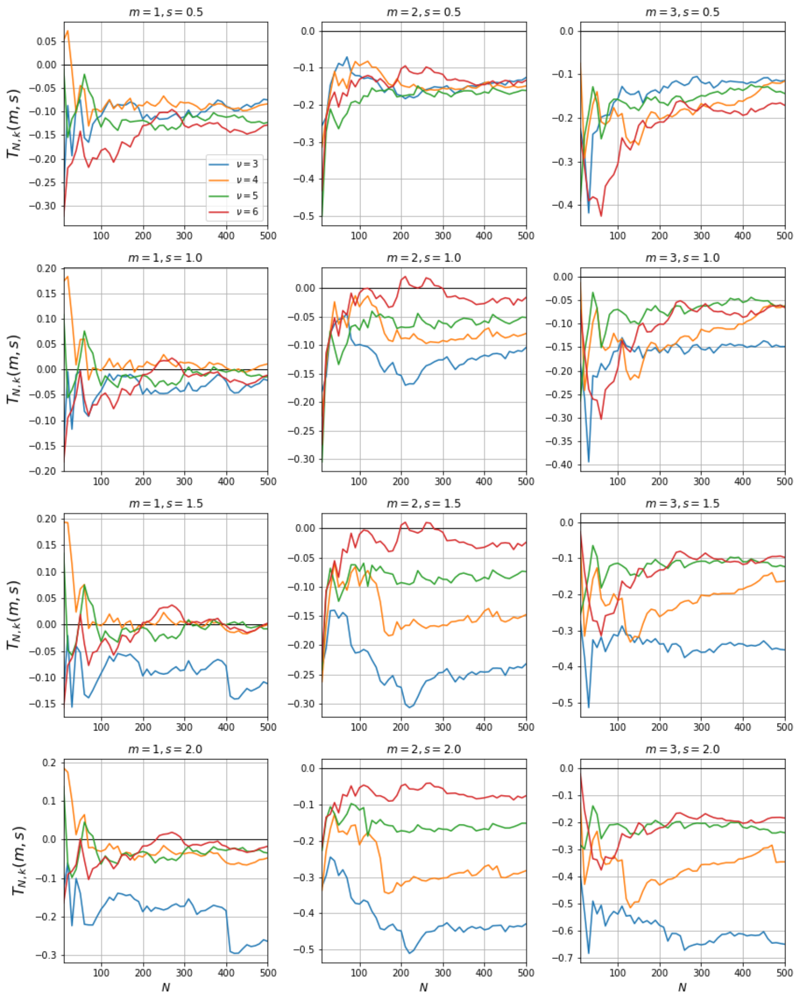


Figure 2.9: The behaviour of $T_{N,k}(m, s_0)$ with $k = 1$ on data from the $ST(m, \nu)$ distribution for different values of m.

2.5.3 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 different values of (N,k) and (m,s), $N_T = 1000$ samples from the $GG(m,s)$ distribution are generated and the corresponding values of $T_{N,k}(m,s)$ is recorded, repeating this $M = 100$ times. To each of these 10 samples from the distribution of $T_{N,k}(m,s)$ then the Shapiro-Wilk test is applied for normality [111] and the p-value is recorded returned by the test.

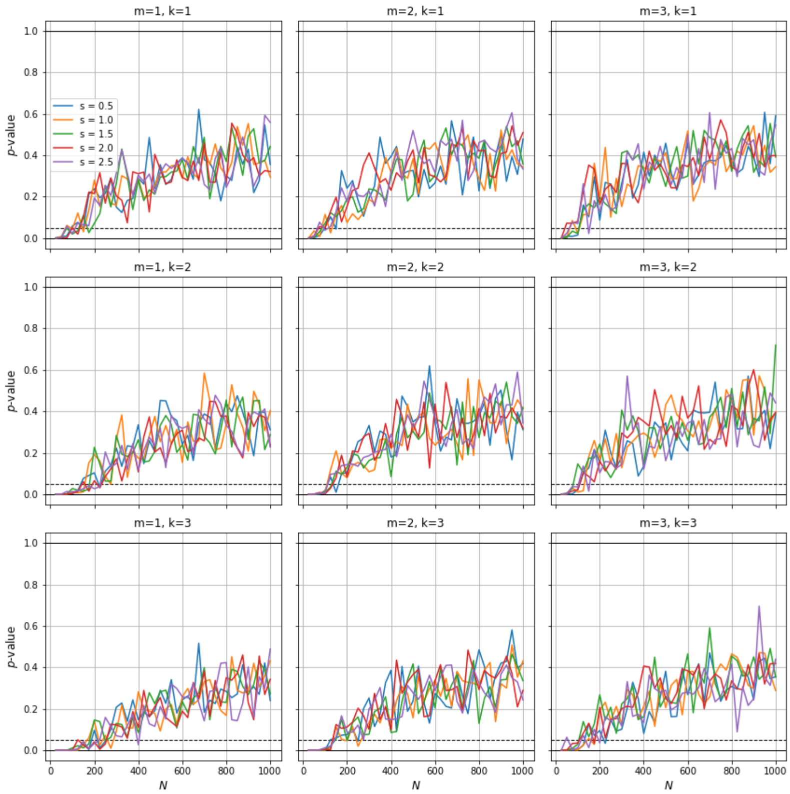


Figure 2.10: Shapiro-Wilk p-values as N increases for different values of m, s and k ($M = 100$ repetitions).

Figure 2.10 shows how these p-values behave as N increases, for various values of m, s and k. The plots suggest that the normal hypothesis cannot be rejected for samples of size $N = 200$ or more. Figure 2.11 shows how the p-values behave for $k = 1$, with error bars corresponding to the standard error across the $M = 100$ repetitions.

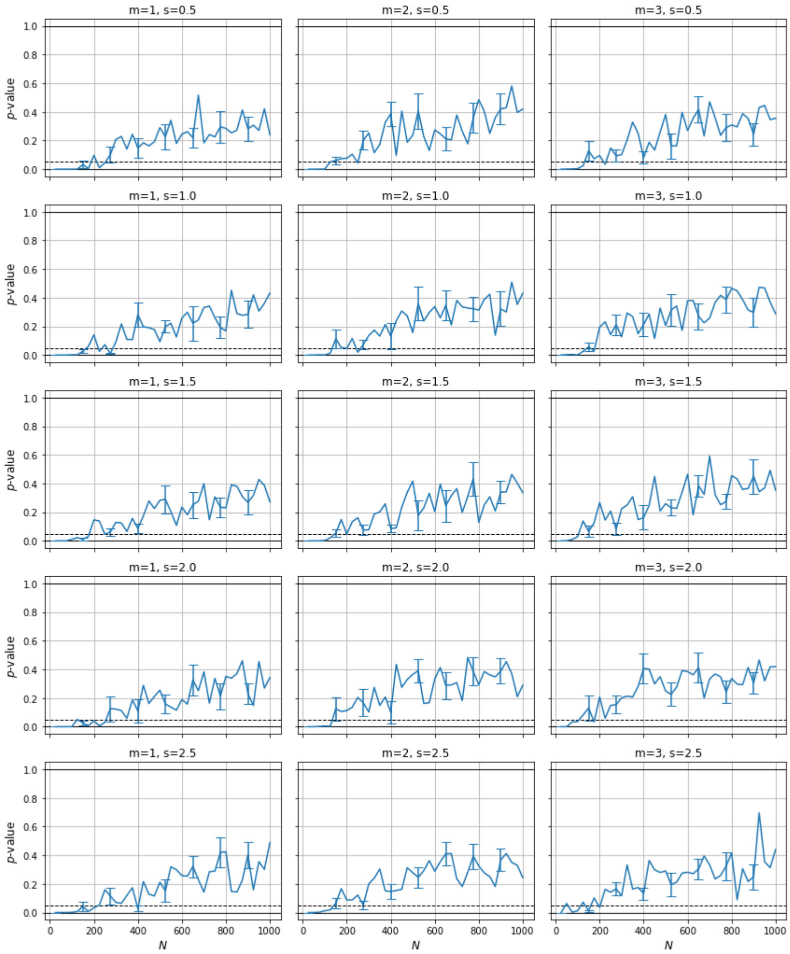


Figure 2.11: Shapiro-Wilk p-values as N increases for different values of m and s with $k = 1$ ($M = 100$ repetitions).

— Chapter 3 —

Goodness-of-fit test for multivariate Student and Pearson type II distributions

Entropy and its various generalizations are important in many fields including mathematical statistics, communication theory, physics and computer science, for characterizing the amount of information in a probability distribution. This chapter defines that a class of estimators of the Rényi entropy based on the independent identically distribution sample drawn from an unknown distribution f in $\mathbb{R}^m$. The chapter also develops the goodness-of-fit test to determine that sample appearance follows one of the established multivariate Student t distribution based on the maximum entropy principle. L^2- consistency is also proved using the k-th nearest neighbours estimator of Rényi entropy. The advantage of the estimator is indicated via theoretical and numerical considerations. This chapter is structured as follows:

- Section 3.1 reviews the maximum Rényi entropy principles for the Rényi entropy and clarifies the multivariate Student and Pearson type II distributions.

- Section 3.2 reviews the nearest-neighbour estimators for the Rényi entropy. The L^2- convergences of estimator are proven for arbitrary fixed $k \geq 2$.

- Section 3.3 proposes the goodness-of-fit test for the multivariate Student and Pearson type II distributions.

- Section 3.4 includes the numerical results with some auxiliary material.

3.1 Maximum entropy principle

Let $X \in \mathbb{R}^m$ be a random vector that has a density function $f(x)$ with respect to Lebesgue measure on $\mathbb{R}^m$. The Rényi entropy of order $q \in (0,1) \cup (1,\infty)$ of this distribution is

$$H_q(f) = \frac{1}{1-q} \log \int_{\mathbb{R}^m} f^q(x)\,dx, \tag{3.1}$$

which is continuous and non-increasing in q. If the support $S = \{x \in \mathbb{R}^m : f(x) > 0\}$ of the distribution has finite Lebesgue measure $|S|$, then

$$\lim_{q\to 0} H_q(f) = \log |S|,$$

otherwise $H_q(f) \to \infty$ as $q \to 0$. Note also that

$$\lim_{q\to 1} H_q(f) = H_1(f) = -\int_S f(x) \log f(x)\,dx.$$

Let $a \in \mathbb{R}^m$ and let Σ be a symmetric positive definite $m \times m$ matrix.

The multivariate Gaussian distribution $N_m(a,\Sigma)$

The *multivariate Gaussian* distribution $N_m(a,\Sigma)$ has density function

$$f^G_{a,\Sigma}(x) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-a)^T\,\Sigma^{-1}(x-a)\right), \quad x \in \mathbb{R}^m. \tag{3.2}$$

For $X \sim N_m(a,\Sigma)$, we have $a = \mathbb{E}(X)$ and $\Sigma =$ Cov(X), where Cov$(X) = \mathbb{E}[(X - a)(X-a)^T]$ is the covariance matrix of the distribution.

The multivariate Student distribution $T_m(a,\Sigma,\nu) = ST(m,\nu)$

For $\nu > 0$, the *multivariate Student distribution* $T_m(a,\Sigma,\nu)$ on $\mathbb{R}^m$ has density function

$$f^S_{a,\Sigma,\nu}(x) = f(x : m, \nu) = c_S |\Sigma|^{-1/2} \left(1 + \frac{1}{\nu}(x-a)^T \Sigma^{-1}(x-a)\right)^{-\frac{\nu+m}{2}}, \quad x \in \mathbb{R}^m,$$

where

$$c_S(m,\nu) = \frac{\Gamma[(\nu+m)/2]}{(\pi\nu)^{m/2}\Gamma(\nu/2)}. \tag{3.3}$$

For $X \sim T_m(a,\Sigma,\nu)$, we have $a = \mathbb{E}(X)$ when $\nu > 1$ and $\Sigma = (1-2/\nu)$Cov(X) when $\nu > 2$, see [62]. It is known that from [50], $f^S_{a,\Sigma,\nu}(x) \to f^G_{a,\Sigma}(x)$ converges pointwise as $\nu \to \infty$.

The multivariate Pearson type II distribution $P_m(a, \Sigma, \xi)$

For $\xi > 0$, the *multivariate Pearson Type II* distribution $P_m(a, \Sigma, \xi)$ on $\mathbb{R}^m$, also known as the *Barenblatt* distribution, has density function

$$f^P_{a,\Sigma,\xi}(x) = c_B|\Sigma|^{-1/2}\left[1-(x-a)^T\Sigma^{-1}(x-a)\right]_+^{\xi}, \quad x \in \mathbb{R}^m$$

where $t_+ = \max\{t, 0\}$ and

$$c_B(m,\xi) = \frac{\Gamma(m/2+\xi+1)}{\pi^{m/2}\Gamma(\xi+1)}. \tag{3.4}$$

For $X \sim P_m(a, \Sigma, \xi)$, we have $a = \mathbb{E}(X)$ and $\Sigma = (m + 2\xi + 2)\mathrm{Cov}(X)$. It is known that $f^P_{a,\Sigma,\xi}(x) \to f_{a,\Sigma}(x)$ as $\xi \to \infty$.

Remark 3.1.1. *If the covariance matrix C is diagonal, the Pearson Type II distribution belongs to the class of time-dependent distributions*

$$u(x,t) = c(\beta,\gamma)t^{-\alpha m}\left(1-\left(\frac{\|x\|}{ct^{\alpha}}\right)^{\beta}\right)_+^{\gamma}$$

with $c > 0$, $\mathrm{supp}\{u(x,t)\} = \{x \in \mathbb{R}^m : \|x\| < ct^{\alpha}\}$ and

$$c(\beta,\gamma) = \beta\Gamma\left(\frac{m}{2}\right)/\left[2c^m\pi^{\frac{m}{2}}B\left(\frac{m}{\beta},\gamma+1\right)\right],$$

which are known as Barenblatt solutions *of the source-type non-linear diffusion equations $u'_t = \Delta(u^q)$ where $q > 1$, Δ is the Laplacian and $\gamma = 1/(q-1)$. For details see [41, 123] or [29].*

3.1.1 Rényi entropy

The Rényi entropy of the multivariate Gaussian distribution $N_m(a, \Sigma)$, see [56], is

$$\begin{aligned} H_q(f^G_{a,\Sigma}) &= \log\left[(2\pi)^{m/2}|\Sigma|^{1/2}\right] - \frac{m}{2(1-q)}\log q \\ &= H_1(f^G_{a,\Sigma}) - \frac{m}{2}\left(1+\frac{\log q}{1-q}\right) \end{aligned}$$

where $H_1(f^G_{a,\Sigma}) = \log\left[(2\pi e)^{m/2}|\Sigma|^{1/2}\right]$ is the differential entropy of $N_m(a, \Sigma)$.

From [132], the Rényi entropy of the multivariate Student distribution $T_m(a, \Sigma, \nu)$ is

$$H_q\left(f^S_{a,\Sigma,\nu}\right) = \frac{1}{2}\log|\Sigma| + c'_S(m,q,\nu) \tag{3.5}$$

where

$$c_S'(m,q,\nu) = \frac{1}{1-q}\log\left(\frac{B\left(q\left(\frac{\nu+m}{2}\right)-\frac{m}{2},\frac{m}{2}\right)}{B\left(\frac{\nu}{2},\frac{m}{2}\right)^q}\right) + \frac{m}{2}\log(\pi\nu) - \log\Gamma\left(\frac{m}{2}\right).$$

Likewise, from [132] the Rényi entropy of the multivariate Pearson Type II distribution $P_m(a,\Sigma,\xi)$ is

$$H_q\left(f^P_{a,\Sigma,\xi}\right) = \frac{1}{2}\log|\Sigma| + c_B'(m,q,\xi), \tag{3.6}$$

where

$$c_B'(m,q,\xi) = \frac{1}{1-q}\log\left(\frac{B\left(q\xi+1,\frac{m}{2}\right)}{B\left(\xi+1,\frac{m}{2}\right)^q}\right) + \frac{m}{2}\log(\pi) - \log\Gamma\left(\frac{m}{2}\right).$$

3.1.2 Maximum entropy principle

Definition 3.1.1. *Let $\mathscr{K}$ be the class of density functions supported on $\mathbb{R}^m$ and subject to the constraints*

$$\int_{\mathbb{R}^m} x f(x)\,dx = a \quad \textit{and} \quad \int_{\mathbb{R}^m} (x-a)(x-a)^T f(x)\,dx = C,$$

where $a \in \mathbb{R}^m$ and C is a symmetric and positive definite $m \times m$ matrix.

It is well-known that the differential entropy H_1 is uniquely maximized by the multivariate normal distribution $N_m(a,\Sigma)$, that is

$$H_1(f) \le H_1(f^G_{a,\Sigma}) = \log\left[(2\pi e)^{m/2}|\Sigma|^{1/2}\right]$$

with equality if and only if $f = f^G_{a,\Sigma}$ almost everywhere. The following result is discussed by [56, 70, 83] and [62] among others.

Theorem 3.1.1 (Maximum Rényi entropy).

(1) For $m/(m+2) < q < 1$, $H_q(f)$ is uniquely maximized over $\mathscr{K}$ by the multivariate Student distribution $T_m(a,\Sigma,\nu)$ with $\nu = 1/(1-q) - m$ and $\Sigma = (1-2/\nu)C$.

(2) For $q > 1$, $H_q(f)$ is uniquely maximized over $\mathscr{K}$ by the multivariate Pearson Type II distribution $P_m(a,\Sigma,\xi)$ with $\xi = 1/(q-1)$ and $\Sigma = (2\xi+m+2)C$.

Applying (3.5) and (3.6) yields the following.

Corollary 3.1.1.1.

(1) For $m/(m+2) < q < 1$, the maximum value of H_q is

$$H_q^{max} = \frac{1}{2}\log|\Sigma| + c'_S(m,q,\nu)$$

with $\nu = 1/(1-q) - m$ *and* $\Sigma = (1-2/\nu)C$.

(2) For $q > 1$, the maximum value of H_q is

$$H_q^{maxx} = \frac{1}{2}\log|\Sigma| + c'_B(m,q,\xi)$$

with $\xi = 1/(q-1)$ *and* $\Sigma = (2\xi + m + 2)\,C$.

3.2 Statistical estimation of the Rényi entropy

The results on the statistical estimation of the Rényi entropy are stated due to [75] and [98]. Extensions of these results can be found in [13, 18, 30, 97] and [44].

Let $X \in \mathbb{R}^m$ be a random vector with density function f, and let $G_q(f)$ denote the expected value of $f^{q-1}(X)$,

$$G_q(f) = \mathbb{E}\big[f^{q-1}(X)\big] = \int_{\mathbb{R}^m} f^q(x)\,dx$$

so that $H_q(f) = \frac{1}{1-q}\log G_q(f)$.

Let $X_1, X_2, \dots$ be independent random vectors from the distribution of X, and for $k, N \in \mathbb{N}$ where $k < N$, let $\rho_{i,k,N}$ denote the *k-th nearest neighbour distance* of X_i in the sample $X_1, X_2, \dots, X_N$, defined to be the k-th order statistic of the $N-1$ distances $\|X_i - X_j\|$ with $j \neq i$,

$$\rho_{i,1,N} \leq \rho_{i,2,N} \leq \cdots \leq \rho_{i,N-1,N}.$$

The expectation $G_q(f) = \mathbb{E}(f^{q-1(X)})$ is estimated by the sample mean

$$\hat{G}_{k,N,q} = \frac{1}{N}\sum_{i=1}^{N}\big(\zeta_{i,k,N}\big)^{1-q},$$

where

$$\zeta_{i,k,N} = (N-1)C_k V_m \rho_{i,k,N}^m \quad \text{with} \quad C_k = \left[\frac{\Gamma(k)}{\Gamma(k+1-q)}\right]^{\frac{1}{1-q}}$$

and $V_m = \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2}+1)}$ is the volume of the unit ball in $\mathbb{R}^m$.

Convergence results

For $r > 0$, the *r-moment* of a density function f is

$$M_r(f) = \mathbb{E}(\|X\|^r) = \int_{\mathbb{R}^m} \|x\|^r f(x)\,dx,$$

and the *critical moment* of f is

$$r_c(f) = \sup\{r > 0 : M_r(f) < \infty\}$$

so that $M_r(f) < \infty$ if and only if $r < r_c(f)$.

Theorem 3.2.1 was formulated without proof in [75], the proof of it is presented below.

Theorem 3.2.1. *Let $0 < q < 1$ and $k \geq 1$ be fixed.*

1. *If $G = G_q(f) < \infty$ and*

$$r_c(f) > \frac{m(1-q)}{q}, \tag{3.7}$$

then

$$\mathbb{E}\left[\hat{G}_{k,N,q}\right] \to G_q \quad \text{as } N \to \infty. \tag{3.8}$$

2. *If $G_q(f) < \infty$, $q > \frac{1}{2}$ and*

$$r_c(f) > \frac{2m(1-q)}{2q-1}, \tag{3.9}$$

then

$$\mathbb{E}\left[\hat{G}_{k,N,q} - G_q\right]^2 \to 0 \quad \text{as } N \to \infty. \tag{3.10}$$

Remark 3.2.1. *If $G_q < \infty$ for $q \in \left(1, \frac{k+1}{2}\right)$ then*

$$\mathbb{E}\left[\hat{G}_{k,N,q}\right] \to G_q \quad \text{and} \quad \mathbb{E}\left[\hat{G}_{k,N,q} - G_q\right]^2 \to 0 \quad \text{as } N \to \infty,$$

see [75].

Remark 3.2.2. *If $G_q(f) < \infty$ for $q \in (0,1)$ and $f(x) = O(\|x\|^{-\beta})$ as $\|x\| \to \infty$ for some $\beta > m$, then $r_c(f) = \beta - m$ and condition (3.8) is automatically satisfied, see [99] for a discussion and counterexamples showing that conditions (3.7) and (3.9) cannot be omitted in general.*

Proof of Theorem 3.2.1.

Let us write

$$\hat{G}_{k,N,q} = \frac{1}{N}\sum_{i=1}^{N}\left[(N-1)^{1/m}(C_k V_k)^{1/m}\rho_{i,k,N-1}\right]^{(1-q)m}.$$

It is showed that the method proposed by [9] for $k=1$, in fact works for any fixed $k \geq 1$. By Theorem 2.1 of [99], the uniform integrability condition

$$\sup_N \mathbb{E}\left[\left\{(N-1)(C_k V_k)\rho^m_{i,k,N-1}\right\}^{(1-q)p}\right] < \infty \tag{3.11}$$

for some $p>1$ (statement 1) or some $p>2$ (statement 2) ensures the L_p convergence of $\hat{G}_{N,k.q}$ to I_q as $N \to \infty$. Because a bound on left-hand side of (3.11) is needed to obtain, the results can be used on the subadditivity of Euclidean functionals defined on the nearest-neighbors graph [128].

The following Lemma 3.3 from [98] is used, see also [128, p.85].

Lemma 3.2.1 ([98]). *Let* $0 < s < m$.

If $r_c(f) > \frac{ms}{m-s}$, *then*

$$\sum_{j=1}^{\infty} 2^{js}\left[P(A_j)\right]^{\frac{m-s}{m}} < \infty, \quad \textit{where} \quad P(A_j) = \int_{A_j} f(x)\,dx \textit{ and}$$
$$A_j = b(0,2^{j+1}) \setminus b(0,2^j) \textit{ for } j = 1,2,\ldots$$

with $b(0,R) = \{x \in \mathbb{R}^m : \|x\| \leq R\}$ *and* $A_0 = b(0,2)$.

Continuing the proof of Theorem 3.2.1, let $b=(1-q)mp$ and $0 < 1-b/m < 1$, one gets $p > 1/(1-q)$.

By exchangeability,

$$\begin{aligned}
&\mathbb{E}\left[(N-1)^{1/m}(C_k V_m)^{1/m}\rho_{i,k,N-1}\right]^b && (3.12)\\
&\quad = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left[(N-1)^{1/m}(C_k V_m)^{1/m}\rho_{i,k,N-1}\right]^b\right)\\
&\quad = \frac{(N-1)^{b/m}}{N}(C_k V_m)^{b/m}\mathbb{E}\left(\sum_{i=1}^{N}\rho^b_{i,k,N-1}\right)\\
&\quad \leq (C_k V_m)^{b/m}(N-1)^{b/m-1}\mathbb{E}(\mathscr{L}^b_k(\mathscr{X}_N)),
\end{aligned}$$

where $\mathscr{X}_N = \{X_1, X_2, \ldots, X_N\}$, and for any finite point set $\mathscr{X} \subset \mathbb{R}^m$ and $b>0$:

$$\mathscr{L}^b_k(\mathscr{X}) = \sum_{x \in \mathscr{X}} \mathscr{D}^b_k(x, \mathscr{X}),$$

where $\mathscr{D}_k(x,\mathscr{X})$ denotes the Euclidean distance from x to its k-nearest neighbour in the point set $\mathscr{X} \setminus \{x\}$ when $\mathrm{card}(\mathscr{X}) \geq k$; set $\mathscr{D}_k(x,\mathscr{X}) = 0$ if $\mathrm{card}(\mathscr{X}) \leq k$.

The function $\mathscr{X} \mapsto \mathscr{L}_k^b(\mathscr{X})$ satisfies the subadditivity relation

$$\mathscr{L}_k^b(\mathscr{X} \cap \mathscr{Y}) \leq \mathscr{L}_k^b(\mathscr{X}) + \mathscr{L}_k^b(\mathscr{Y}) + U_k t^b \tag{3.13}$$

for all $t > 0$ and finite $\mathscr{X}$ and $\mathscr{Y}$ contained in $[0,t]^m$, where $U_k = 2km^{b/2}$, $b > 0$. Indeed, if $\mathscr{X}$ has more than k elements, the k-nearest neighbour distances of points in $\mathscr{X}$ can only become smaller when we add some other set $\mathscr{Y}$. Hence, (3.13) holds with $U_k = 0$ if $\mathscr{X}$ and $\mathscr{Y}$ have more than k elements. If $\mathscr{X}$ has k elements or less, then $\mathscr{L}_k^b(\mathscr{X})$ is zero, but when the set $\mathscr{Y}$ is added, it is gained at most k new edges from points in $\mathscr{X}$ in the nearest neighbours graph, and each of these is of length most $t\sqrt{m}$ (for more details, see [[128, pp.101-103]).

Let $s(N)$ be the largest $j \in N$ such that $\mathscr{X}_N = \{X_1, X_2, \ldots, X_N\} \cap A_j$ is not empty. Using ideas from [128, p.87]:

$$\mathscr{X}_N \cap \left(\bigcup_{j=0}^{s(N)} A_j \right) = \bigcup_{j=0}^{s(N)} (X_N \cap A_j),$$

and by the subadditivity property,

$$\mathscr{L}_k^b(\mathscr{X}_N) \leq \mathscr{L}_k^b\{X_N \cap A_{s(N)}\} + \mathscr{L}_k^b\left(\mathscr{X}_N \cap \left\{ \bigcup_{j=0}^{s(N)-1} A_j \right\} \right) + U_k 2^{(s(N)+1)b}.$$

Applying subadditivity in the same way to the second term on the right yields

$$\begin{aligned} &\mathscr{L}_k^b\left(\mathscr{X}_N \cap \left\{ \bigcup_{j=0}^{s(N)-1} A_j \right\} \right) \\ &\leq \mathscr{L}_k^b(\mathscr{X}_N \cap A_{s(N)-1}) + \mathscr{L}_k^b\left(\mathscr{X}_N \cap \left\{ \bigcup_{j=0}^{s(N)-2} A_j \right\} \right) + U_k \left(2^{s(N)}\right)^b. \end{aligned}$$

Repeatedly applying subadditivity, it can be arrived at

$$\begin{aligned} \mathscr{L}_k^b(X_1, \ldots, X_N) &\leq \sum_{j=0}^{s(N)} \mathscr{L}_k^b(\mathscr{X}_N \cap A_j) + 2^{b+bs(N)} \frac{U_k}{1-2^{-b}} \\ &\leq \sum_{j=0}^{s(N)} \mathscr{L}_k^b(\mathscr{X}_N \cap A_j) + 2^{bs(N)} M_k \\ &\leq \sum_{j=0}^{s(N)} \mathscr{L}_k^b(\mathscr{X}_N \cap A_j) + M_k \max_{1 \leq i \leq N} \|X_i\|^b \end{aligned} \tag{3.14}$$

for some constant M_k depending on m, k and b.

From (3.13) and (3.14), it is obtained that

$$\begin{aligned}
&\mathbb{E}\left((N-1)^{1/m}(C_kV_m)^{1/m}\rho^{(i)}_{k,N-1}\right)^b \\
&\quad\leq\ (C_kV_m)^{b/m}(N-1)^{b/m-1}\mathbb{E}\left(\sum_{j=0}^{s(N)}\mathscr{L}^b_k(\mathscr{X}_N\cap A_j)\right) \\
&\qquad + W_k\,\mathbb{E}\left((N-1)^{b/m-1}\max_{1\leq i\leq N}\|X_i\|^b\right) \qquad (3.15)
\end{aligned}$$

for some constant W_k depending on m, k and b.

Using Lemma 3.3 of [128], it is achieved:

$$L^b_k(\mathscr{X})\leq L(\text{diam}\mathscr{X})^b(\ \text{card}\mathscr{X})^{1-b/m} \qquad (3.16)$$

for some constant L. Following [98], by Jensen's inequality and using the fact that $\text{diam}(A_j)=2^j$ from (3.15) and (3.16), we obtain that

$$(N-1)^{b/m-1}\mathbb{E}\left(\sum_{j=0}^{s(N)}L^b_k(\mathscr{X}_N\cap A_j)\right)\leq L_1\sum_{j=0}^{s(N)}2^{jb}\left[\mathbb{P}(X_1\in A_j)\right]^{1-b/m}, \qquad (3.17)$$

where $L_1>0$ is a constant.

Recall our assumptions that $0<\alpha<\frac{m}{\ell}$ where $\ell\in\{1,2\}$ and $\alpha=(1-q)m$, and also that $r_c(f)>\frac{lm\alpha}{m-l\alpha}$. Setting $s=b$ in Lemma 3.2.1, we see that the left hand side of (3.17) is finite, so the first term on the right hand side of (3.15) is bounded by a constant which is independent of N.

For a non-negative random variable $Z>0$, it is know that

$$\mathbb{E}(Z)=\int_0^\infty\mathbb{P}(Z>z)\,dz,$$

so the second term in (3.15) is bounded by

$$\begin{aligned}
&W_k\int_0^\infty\mathbb{P}\left(\max_{1\leq i\leq N}\|X_i\|^b>u\cdot N^{1-b/m}\right)du \\
&\quad\leq\ W_k\left[1+N\int_1^\infty\mathbb{P}\left(\|X_1\|^b>\left(u^{\frac{mb}{m-b}}N\right)^{1-b/m}\right)du\right]. \qquad (3.18)
\end{aligned}$$

By the Markov inequality $\mathbb{P}(Z>a)\leq\frac{\mathbb{E}|Z|}{a}$ for $a>0$, we get for $u\geq 1$.

$$\begin{aligned}
\mathbb{P}\left(\|X_1\|^b>\left(u^{\frac{m}{m-b}}N\right)^{1-b/m}\right)\ &=\ \mathbb{P}\left(\|X_1\|^{mb/(m-b)}>u^{m/(m-b)}N\right) \\
&\leq\ \mathbb{E}\|X_1\|^{mb/(m-b)}\frac{1}{u^{m/(m-b)}N}. \qquad (3.19)
\end{aligned}$$

From (3.18) and (3.19), it is seen that the second term in (3.15) is bounded by

$$W_k \left[1 + \int_1^\infty \mathbb{E}\|X_1\|^{mb/(m-b)} \frac{1}{u^{m/(m-b)}} \, du \right]$$

which is independent of N, because p can be chosen to ensure that $0 < 1 - b/m < 1$, and

$$\mathbb{E}\|X_1\|^{\frac{mp(1-q)}{1-p(1-q)}} < \infty \quad \text{or} \quad r_c(f) > \frac{mp(1-q)}{1-p(1-q)},$$

which is consistent with conditions of Theorem 2.1. Note that the function $h(p,q) = \frac{p(1-p)m}{1-(1-q)p}$ is such that $h(1,q)$ gives the right-hand side of (3.8) and $h(2,q)$ gives the right-hand side of (3.10). Moreover, if $r_c(f) > h(1,q)$ for some $q < 1$ (resp. $r_c(f) > h(2,q)$) for some q satisfying $1/2 < q < 1$, we also have $r_c(f) > h(p,q)$ for some $p > 1$ (resp. $r_c(f) > h(p,q)$ for $p > 2$).

3.3 Hypothesis tests

The class $\mathscr{K}$ is restricted to only those distributions which satisfy the following conditions: for any fixed $k \geq 1$ and $q > 1/2$,

$$\mathbb{E}(\hat{H}_{N,k,q}) \to H_q \quad \text{as } N \to \infty, \text{ and}$$
$$\hat{H}_{N,k,q} \to H_q \quad \text{in probability as } N \to \infty.$$

By the Theorem 3.2.1, it is known that $\mathscr{K}$ contains $T_m(a, \Sigma, \nu)$ for all $\nu > 2$ and $P_m(a, \Sigma, \xi)$ for all $\xi > 0$.

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed random vectors with common density $f \in \mathscr{K}$, and let $\hat{C}_N$ be the sample covariance matrix,

$$\hat{C}_N = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})^T.$$

(1) To test the hypothesis $X \sim T_m(a, \Sigma, \nu_0)$ where $\nu_0 > 2$, the test statistic is defined by

$$W_{N,k}^S(m, \nu) = H_q^{\max} - \hat{H}_{N,k}(m,q), \tag{3.20}$$

where

$$H_q^{\max} = \frac{1}{2} \log |\hat{\Sigma}_N| + c_S'(m, q, \nu)$$

with $q = 1 - 1/(\nu + m)$ and $\hat{\Sigma}_N = (1 - 2/\nu)\hat{C}_N$.

(2) To test the hypothesis $X \sim P_m(a, \Sigma, \xi_0)$ where $\xi_0 > 0$, the test statistic is defined by

$$W^P_{N,k}(m, \xi) = H^{\max}_q - \hat{H}_{N,k}(m, q), \tag{3.21}$$

where

$$H^{\max}_q = \frac{1}{2} \log |\hat{\Sigma}_N| + c'_P(m, q, \xi)$$

with $q = 1 + 1/\xi$ and $\hat{\Sigma}_N = (2\xi + m + 2)\hat{C}_N$.

By the law of large numbers, $\hat{C}_N \to C$ in probability as $N \to \infty$, so by Slutsky's theorem, for any fixed $k \geq 1$, it is obtained that

$$\lim_{N\to\infty} W^S_{N,k}(m, \nu) \xrightarrow{P} \begin{cases} 0 & \text{if } X \sim T_m(a, \Sigma, \nu), \\ c > 0 & \text{otherwise,} \end{cases}$$

and

$$\lim_{N\to\infty} W^P_{N,k}(m, \xi) \xrightarrow{P} \begin{cases} 0 & \text{if } X \sim P_m(a, \Sigma, \xi), \\ c > 0 & \text{otherwise,} \end{cases}$$

where symbol " $\xrightarrow{P}$ " stands for convergence in probability and c is a constant that depends on the distribution of X.

The distributions of $W^P_{N,k}(m, \nu)$ when $X \sim T_m(a, \Sigma, \nu)$ and $W^P_{N,k}(m, \xi)$ when $X \sim P_m(a, \Sigma, \xi)$ are unknown. An analytical derivation of these distributions seems difficult, because the random variables $\hat{H}_{N,k}$ and $\hat{C}_N$ are not independent and their covariance appears to be intractable, despite the fact that the asymptotic distribution of $\hat{H}_{N,k}$ can be revealed by applying the results of [24, 30, 98] and [13] and that of $\hat{C}_N$ by the delta method. In the next section, these null distributions are investigated using Monte Carlo methods.

3.4 Numerical experiments

3.4.1 Random samples

Random samples from $T_m(a, \Sigma, \nu)$ and $P_m(a, \Sigma, \xi)$ can be generated according to the stochastic representation

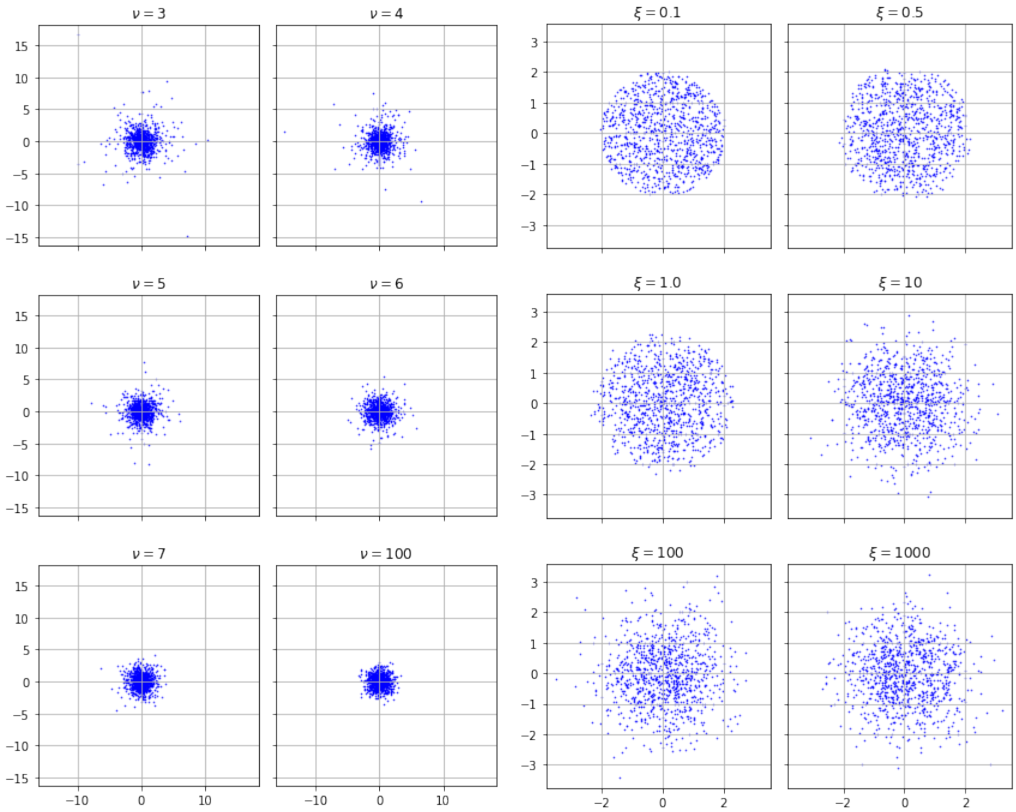
$$X = RBU + a,$$

where R is the distribution of the radial distance $\left[(X - a)^T \Sigma^{-1} (X - a)\right]^{1/2}$, B is an $m \times m$ matrix with $B^T B = \Sigma$ and U is uniformly distributed on the surface of a unit m-sphere S^{m-1}. In particular,

- $R^2 \sim \text{InvGamma}(m/2, m/2)$ yields $X \sim T_m(a, \Sigma, \nu)$;
- $R^2 \sim \text{Beta}(m/2, \xi + 1)$ yields $X \sim P_m(a, \Sigma, \xi)$.

Let I_m be the $m \times m$ identity matrix. The distributions are investigated:

- $T_m(\nu) = T_m\big(0, I_m,\ \nu\big)$ for $\nu > 2$ and
- $P_m(\xi) = P_m\big(0, I_m,\ \xi\big)$ for $\xi > 1$.



(a) Scatter plots for $T_2(\nu)$ (b) Scatter plots for $P_2(\xi)$

Figure 3.1: Scatter plots for the bivariate Student and Pearson II distributions.

For the case $m = 2$, points on scatter plots are generated for different values of ξ and ν, see Figure 3.1a for $T_m(\alpha, \Sigma, \nu)$ and Figure 3.1b for $P_m(\alpha, \Sigma, \xi)$. One can observe that as ν and ξ increase the $T_m(\alpha, \Sigma, \nu)$ and $P_m(\alpha, \Sigma, \xi)$ distributions converge to multivariate Gaussian and Uniform distributions respectively.

3.4.2 Consistency

To investigate the consistency of $W^S_{N,k}(m,\nu)$ for various values of m and ν, $M = 100$ random samples of size N from the $T_m(\nu)$ distribution are generated with N increasing from $N = 500$ to $N = 5000$ in steps of 500, and recording the value of $W^S_{N,k}(m,\nu)$ for $k = 1,2,3$ each time. The mean values of the statistics for $k = 1$

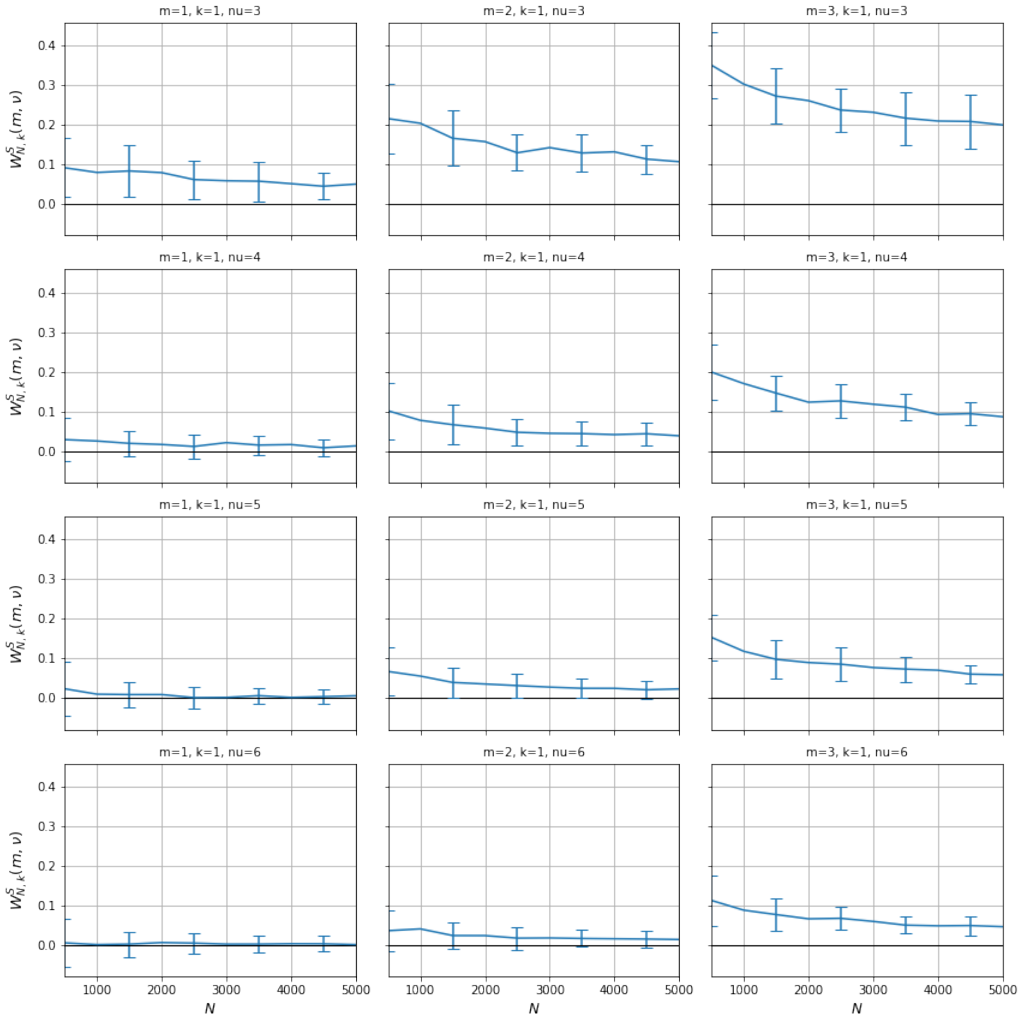


Figure 3.2: The asymptotic behaviour of $W^S_{N,k}(m,\nu)$ as $N \to \infty$ for $k = 1$. The statistic $W^S_{N,k}(m,\nu)$ increases with ν and decreases with m.

are shown in Figure 3.2, where the length of the error bars is equal to the standard deviation of the statistics around their mean values. The mean statistics for $k = 1,2,3$ are shown in Figure 3.3, where it is evident that the statistic of $W^S_{N,k}(m,\nu)$ increases with the parameter value ν, and decreases with the dimension m.

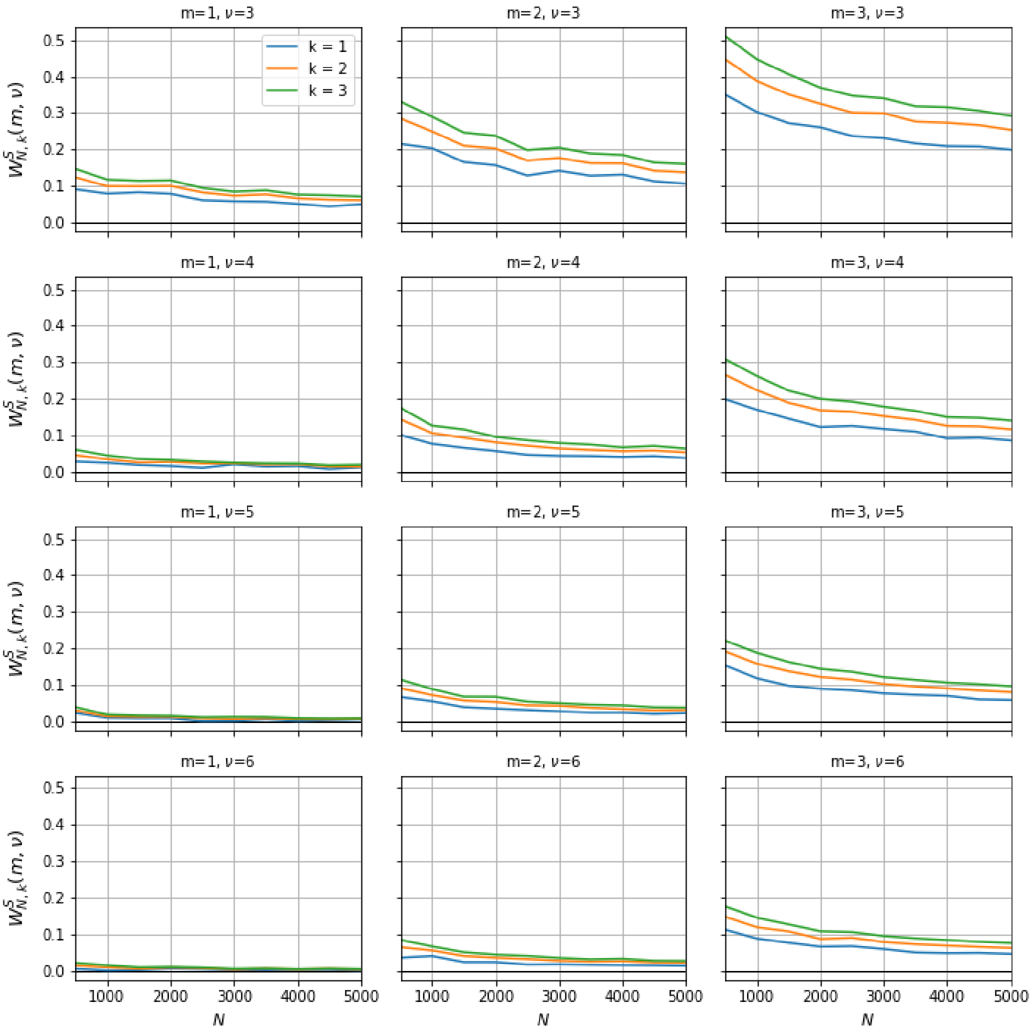


Figure 3.3: The asymptotic behaviour of $W^S_{N,k}(m,\nu)$ as $N \to \infty$. The statistic of $W^S_{N,k}(m,\nu)$ increases with ν and decreases with m.

For fixed N and (m,ν), a sample of size N from $T_m(\nu)$ distribution is generated and the empirical value of $W^S_{N,k}(m,\nu)$ is recorded for a fixed k, and this is repeated $M = 10^3$ times. This yields a sample realisation $\{W_1, W_2, \ldots, W_M\}$ from the distribution of $W^S_{N,k}(m,\nu)$, from which its mean and variance are estimated by

$$\bar{W}^S_{N,k}(m,\nu) = \frac{1}{M}\sum_{j=1}^{M} W_j \quad and \quad S^2_{N,k}(m,\nu) = \frac{1}{M-1}\sum_{j=1}^{M}(W_j - \bar{W}^S_{N,k})^2.$$

Figure 3.2 shows how $\bar{W}^S_{N,k}(m,\nu)$ approach 0 as N increases for various value of $m \in (1,2,3)$, $\nu \in (3,4,5,6)$ and $k = 1,2,3$ with error bar corresponding to the standard error $S_{N,k}(m, \nu)$. From these data, it is observed that the empirical variance is decreasing when k increases. From the other hand, the bias or mean $\bar{W}^S_{N,k}(m, \nu)$ is smaller for values of k or m. These result confirm the variance reduction of k-th nearest neighbour estimators observed in [13].

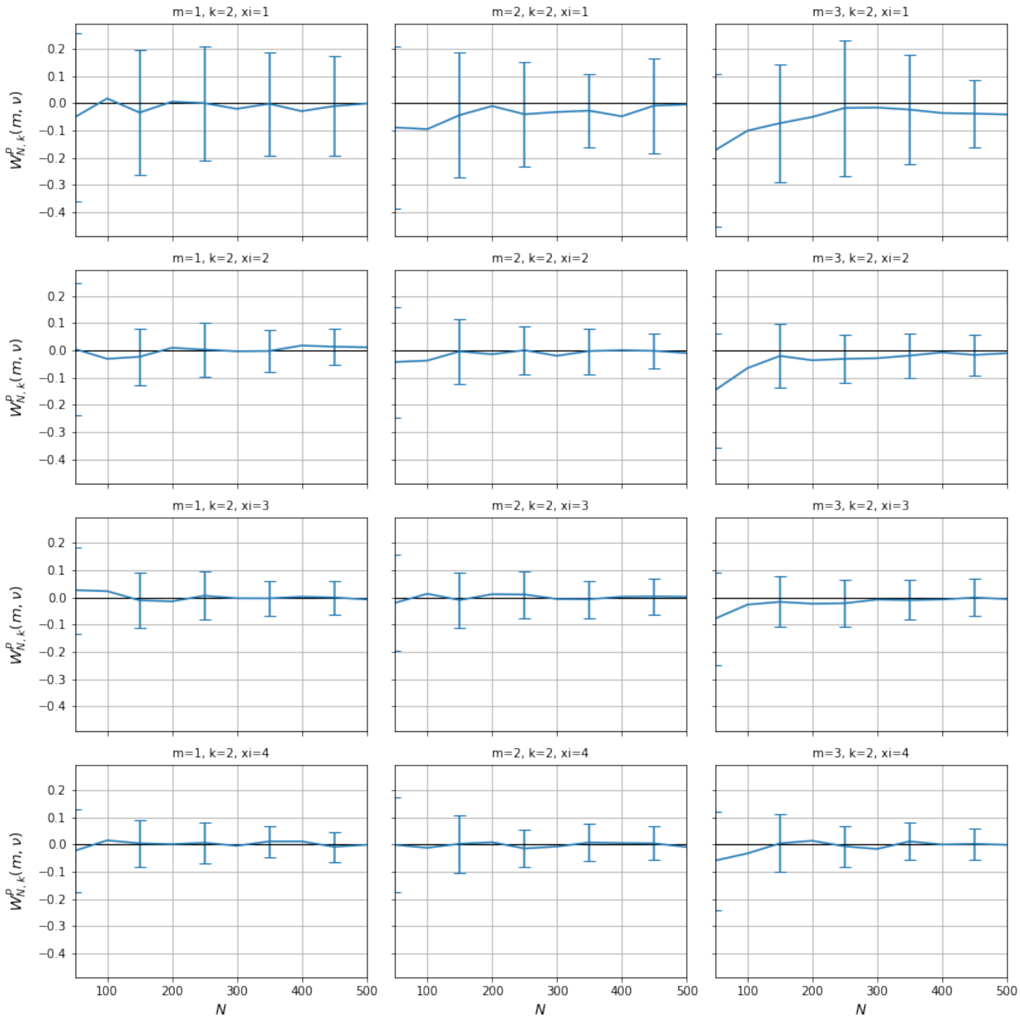


Figure 3.4: Asymptotic behaviour of $W^P_{N,k}(m, \xi)$ as $N \to \infty$ for $k = 2$.

The experiment for $W^P_{N,k}(m, \xi)$ is repeated but this time with samples increasing in size from $N = 50$ to $N = 500$ in steps of 50. The mean values of the statistics for $k = 2$ are shown in Figure 3.4, where length of the error bars are equal to the standard deviation of the statistics their mean values. The mean statistics for $k = 1, 2, 3$ are shown in Figure 3.5: note that these are only defined for $k > 1/\xi$. The convergence of $W^P_{N,k}(m, \xi)$ is evidently much faster than that of $W^S_{N,k}(m, \nu)$, perhaps because the support of $P_m(\xi)$ is bounded for any finite $\xi > 0$ while the support of $T_m(\nu)$ is unbounded.

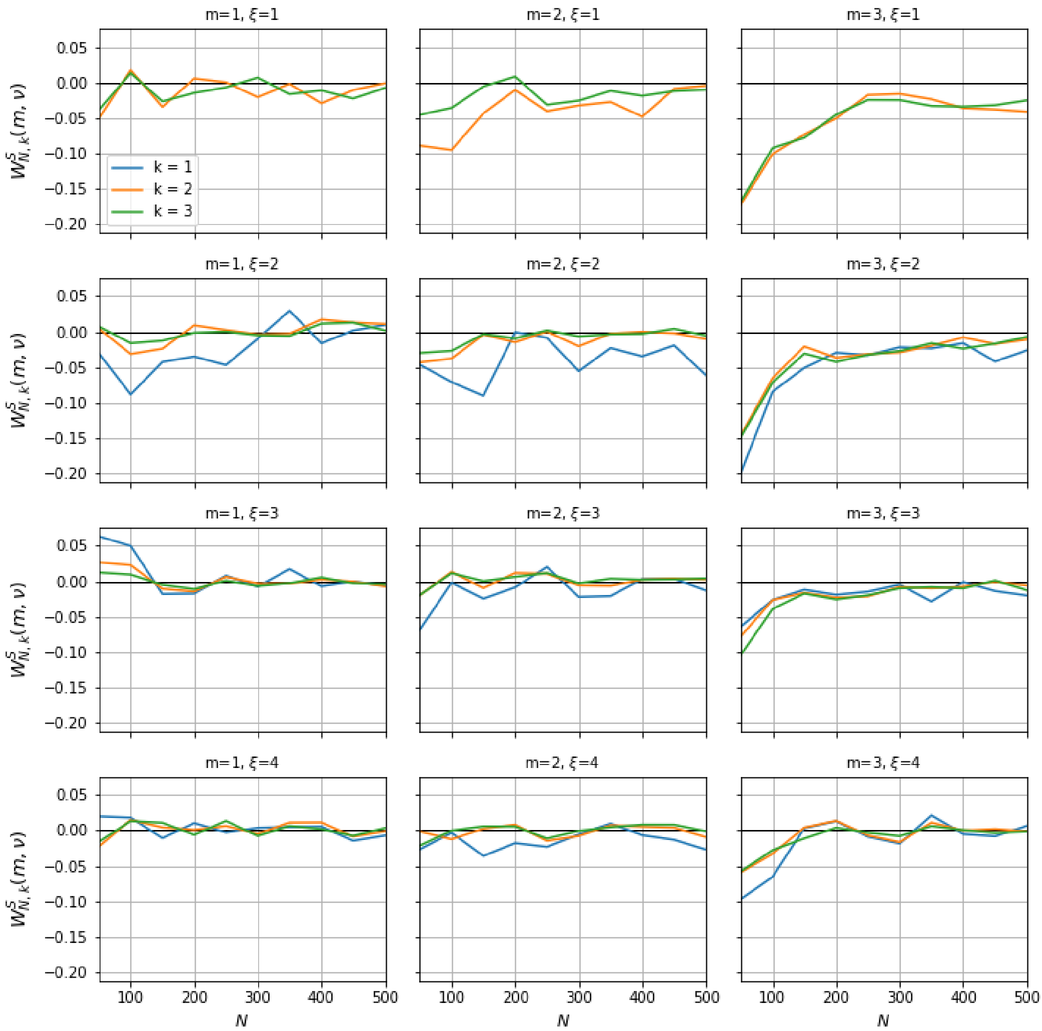


Figure 3.5: Asymptotic behaviour of $W^P_{N,k}(m,\xi)$ as $N \to \infty$. Note that the statistic is defined only for $k > 1/\xi$.

Rates of convergence

In Figure 3.6 shows the convergence of $W^S_{N,k}(m,\nu)$ with $m = 2$, $k = 1$ and $\nu = 5$ as $N \to \infty$ together with the corresponding plot of $\log W^S_{N,k}(m,\nu)$ against $\log N$, which suggests an empirical convergence rate of approximately $O(N^{-1/2})$ as $N \to \infty$.

The experiment for $W^P_{N,k}(m,\xi)$ is repeated with $m = 2$, $k = 2$ and $\xi = 2$. The results are shown in Figure 3.7, which in this case suggests an empirical convergence rate of approximately $O(N^{-2/3})$ as $N \to \infty$.

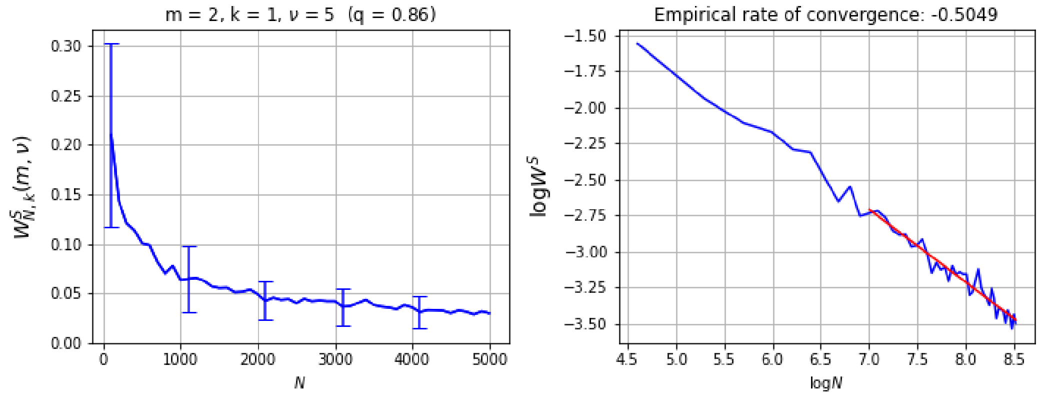


Figure 3.6: Asymptotic behaviour of $W^S_{N,k}(m,\nu)$ with $m=2$, $k=1$ and $\nu=5$.

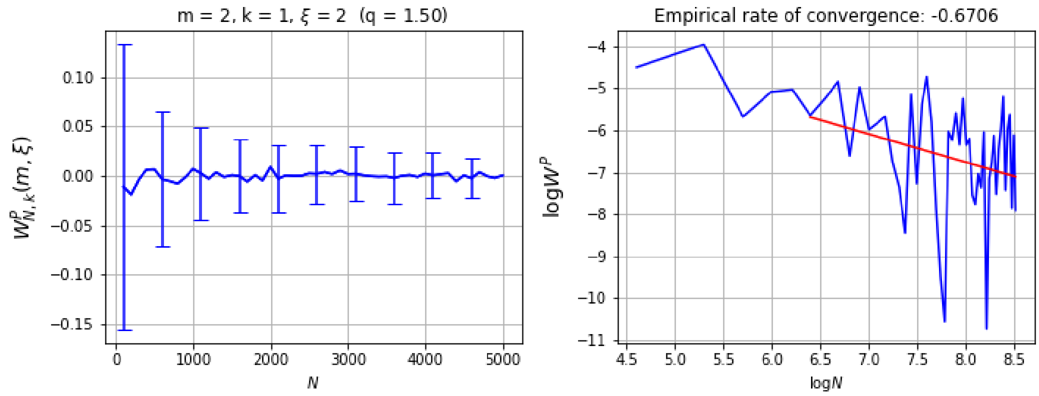


Figure 3.7: Asymptotic behaviour of $W^P_{N,k}(m,\xi)$ with $m=2$, $k=1$ and $\xi=2$.

3.4.3 Empirical distribution of the test statistics

For different values of (N,k) and (m,ν), $n=100$ random samples of size N from the $T_m(\nu)$ distribution is generated, each time recording the value of $W^S_{N,k}(m,\nu)$. The Shapiro-Wilk test is applied for normality to this random sample from the null distribution of $W^S_{N,k}(m,\nu)$, and is recorded the probability value computed by the test of Shapiro and Wilk [111]. This process is repeated $M=1000$ times.

Figure 3.8 shows how the mean probability value behaves as N increases for various values of m, k and ν. The experiment is repeated for the null distribution of $W^P_{N,k}(m,\xi)$. Figure 3.9 shows how the mean probability value behaves as N increases for various values of m, k and ξ.

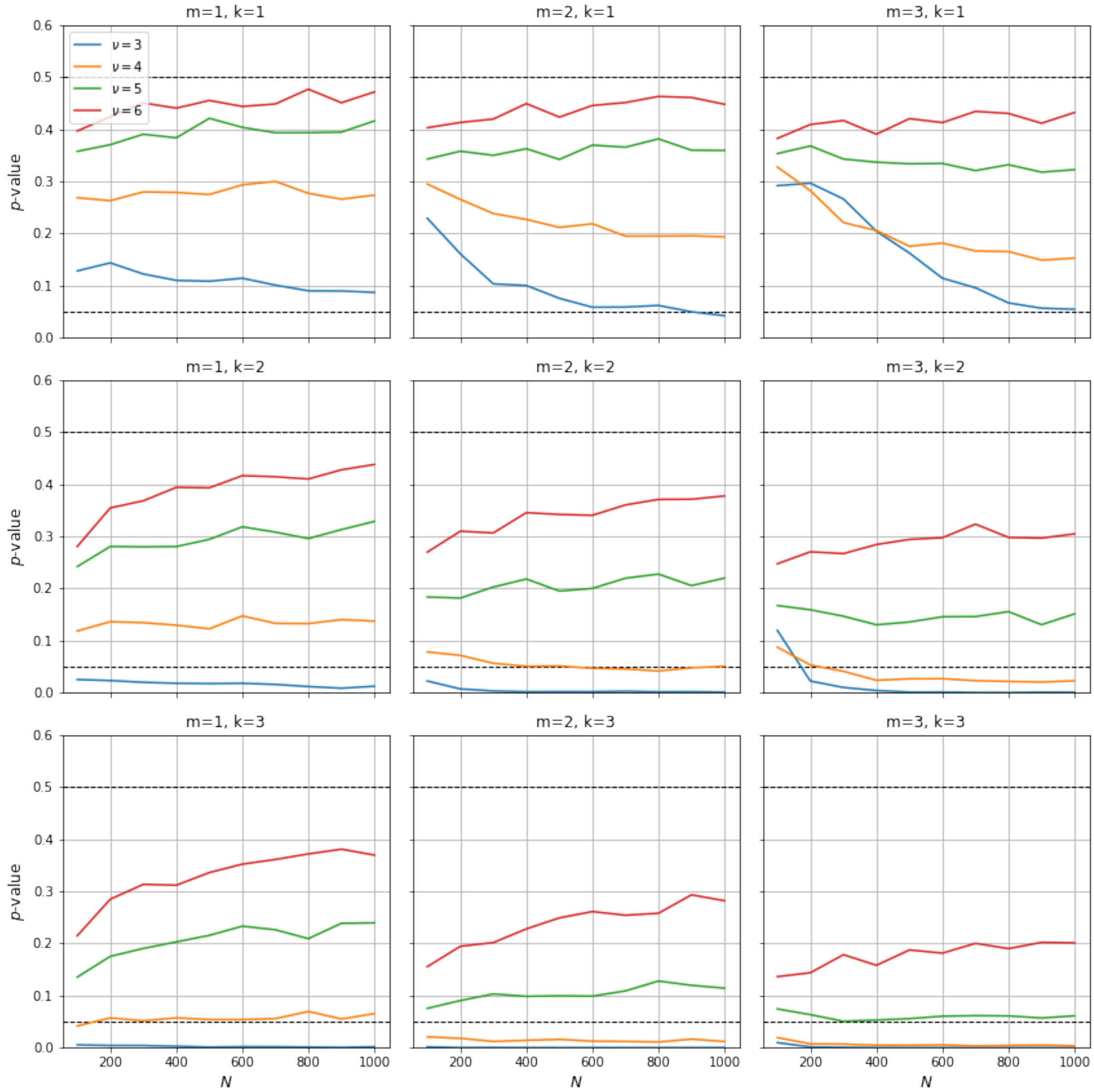


Figure 3.8: Shapiro-Wilk probability values for $W^S_{N,k}(m,\nu)$ as N increases for different values of m, k and ν (100 repetitions).

Figure 3.9: Shapiro-Wilk probability values for $W^{p}_{N,k}(m,\xi)$ as N increases for different values of m, k and ξ (100 repetitions). Note that the statistic is only defined for $k > 1/\xi$.

3.4.4 Point estimation

For fix ν_0 and ξ_0, we generate random points from $T_m(a, \Sigma, \nu)$ and $P_m(\alpha, \Sigma, \xi)$ for $k = 1$ and $m = 3$ then compute the test statistics of $W^S_{N,k}(m, \nu)$ and $W^P_{N,k}(m, \xi)$ for different values of ν and ξ. The point estimator of $\hat{\nu}$ and $\hat{\xi}$ are 4.7 and 2.8 respectively. Figures 3.10 and 3.11 illustrate the point estimator for ν_{true} and ξ_{true} that can be computed by

- $\hat{\nu} = \text{argmin}_{\nu>2}\, W^S_{N,k}(m, \nu)$ and
- $\hat{\xi} = \text{argmin}_{\xi>1}\, W^P_{N,k}(m, \xi)$ respectively.

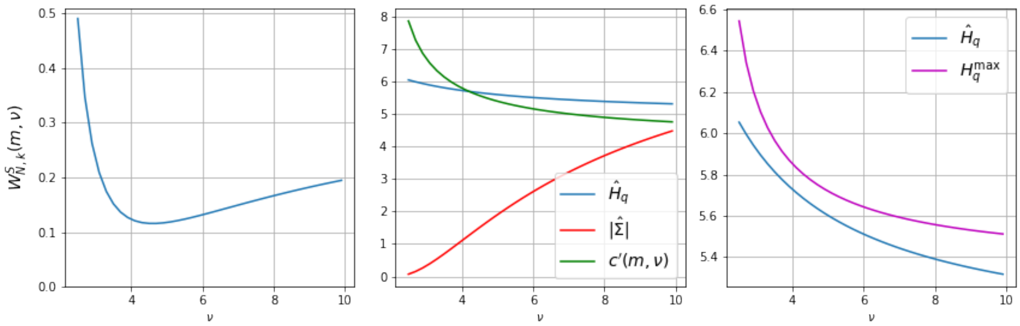


Figure 3.10: $W^S_{N,k}(m, \nu)$ with $N = 10^4$, $k = 1$, $m = 3$ and $\nu_0 = 4$ ($q_0 = 0.86$). The point estimate is $\hat{\nu} = 4.7$.

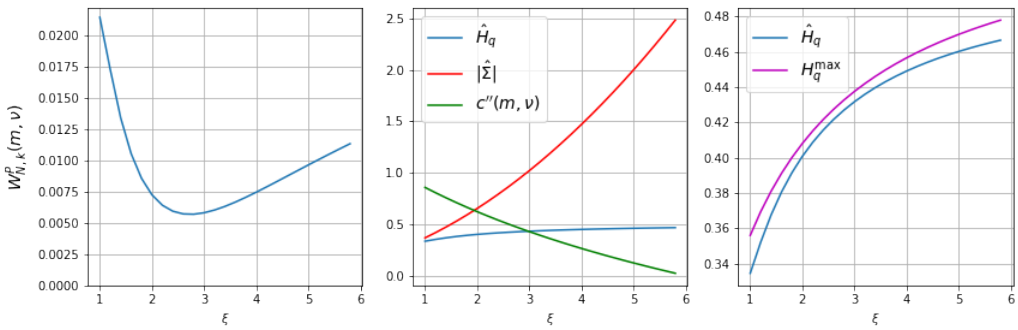


Figure 3.11: $W^P_{N,k}(m, \xi)$ with $N = 10^4$, $k = 1$, $m = 3$ and $\xi_0 = 3$ ($q_0 = 1.33$).The point estimate is $\hat{\xi} = 2.8$.

— Chapter 4 —

The entropy-based goodness-of-fit tests for generalized von Mises-Fisher distributions and beyond

This chapter introduces the new classes of unimodal rotational invariant directional distributions, which generalize von Mises-Fisher distribution. Three types of distributions are proposed and one of them represents axial data. Each new type provides formulas and short computational study of parameter estimators by method of moments and method of maximum likelihood. The main task of the chapter is to develop the goodness-of-fit test to detect that sample entries follows one of the introduced generalized von Mises-Fisher distribution based on the maximum entropy principle. The chapter uses k-th nearest neighbor distances estimator of Shannon entropy and proves its L^2-consistence. The chapter also examines the behaviour of the test statistics, finds critical values and compute power of the test on simulated samples. The goodness-of-fit test is applied to local fiber directions in a glass fibre reinforced composite material and the samples which follows axial generalized von Mises-Fisher distribution are detected.

This chapter is organised as follows:

- Section 4.2 reviews the basic facts for von Mises-Fisher distribution.
- Section 4.3 introduces three types of generalized von Mises-Fisher distributions.
- Section 4.4 provides the Shannon entropies of generalized von Mises-Fisher distributions and presents the maximum entropy principle for them. The sta-

tistical estimation of an entropy is also discussed and the L^2 convergence of k-nearest neighbor estimators are proven.

- Section 4.5 devotes to the maximum likelihood estimators and estimators by method of moments for generalized von Mises-Fisher distributions.
- Section 4.6 develops the goodness-of-fit test based on maximum entropy principle for generalized von Mises-Fisher distributions.
- Section 4.7 shows the results of numerical experiments based on simulated samples.
- Section 4.8 contains the application of the theory for real data set.

4.1 Introduction

Directional distributions characterize randomness to unit vectors (directions). Spherical data sets appear in a wide range of problems arising from Earth sciences [102], biology [91], and material science [33]. Directional data are important in cosmology and astrophysics, for instance, in results of the Laser Interferometer Gravitational-Wave Observatory [1] and the Alpha Magnetic Spectrometer on the International Space Station [3]. Further applications and the modern state art of statistical theory on directional data can be found in [77, 100] and references therein. The following papers provide the most recent developments in directional statistics [52, 54, 55, 63, 77, 92, 104, 105, 106, 108, 109, 124, 125].

In this chapter, some group of random unit vectors are considered with values on sphere $\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$, which have the absolutely continuous directional distributions with respect to the uniform distribution on $\mathbb{S}^{d-1}$. The scalar product of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ is denoted by $\mathbf{a}^T\mathbf{b}$ and the Euclidean norm of $\mathbf{a} \in \mathbb{R}^d$ is denoted by $\|\mathbf{a}\|$.

The von Mises-Fisher distribution is a fundamental isotropic distribution which is widely used in directional statistics [86, p. 168]. It belongs to exponential family of distributions, is rotational invariant and has a density proportional to $\exp(\kappa\boldsymbol{\mu}^T\mathbf{x}), \mathbf{x} \in \mathbb{S}^{d-1}$, such that random vectors are concentrated with rate $\kappa \in \mathbb{R}$ along direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$. The rotating invariant family of distributions is now being researched extensively [28, 35, 45, 93].

Among several important properties of the von Mises-Fisher distribution we focus on maximum entropy characterization, that is, the von Mises-Fisher distribution has maximum entropy in the class of continuous distributions on $\mathbb{S}^{d-1}$ with a given value of $\mathbb{E}\mathbf{X}$ [85]. The von Mises-Fisher distribution is widely used for analysis of neutrino arrival directions recorded by the IceCube Neutrino Observatory, see e.g. [23, 31, 78] and arrival directions of ultrahigh energy cosmic rays recorded by the Pierre Auger Observatory, see e.g. [5, 20, 68].

There are several generalizations, including Fisher-Bingham distribution with a density proportional to $\exp(\kappa\boldsymbol{\mu}^T\mathbf{x}+\mathbf{x}^TA\mathbf{x}), \mathbf{x}\in\mathbb{S}^{d-1}$ [85], and generalized von Mises-Fisher distribution of order k (GvMFk) introduced by [46], having the density proportional to $\exp\left(\sum_{j=1}^{k}\kappa_j(\boldsymbol{\mu}_j\mathbf{x})^{r_j}\right)$, where $\boldsymbol{\mu}_j\in\mathbb{S}^{d-1}, \kappa_j\in\mathbb{R}, r_j\in\mathbb{N}$, $j=1,\dots,k$ and $r_1\leq\ldots\leq r_k$.

This chapter introduces the new generalization of von Mises-Fisher distribution, which stays in exponential family and rotational invariant with one mode. In contrast to generalized von Mises-Fisher distribution of order k with integer powers $r\in\mathbb{N}$, densities with arbitrary positive power $r\in\mathbb{R}_+$ are considered. The motivation of such choice is to provide the analogue of a generalized Gaussian distribution for random vectors on a unit sphere. To accomplish this, the following three types of distributions of order $\alpha\in\mathbb{R}_+$ are introduced, whose densities f are proportional to

$$\begin{aligned}
\text{Type I, } \mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu}): &\qquad f(\mathbf{x})\propto\exp\left(\frac{\kappa}{\alpha}(\boldsymbol{\mu}^T\mathbf{x})^{<\alpha>}\right), \mathbf{x}\in\mathbb{S}^{d-1},\\
\text{Type II, } \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu}): &\qquad f(\mathbf{x})\propto\exp\left(\frac{\kappa}{2^\alpha\alpha}\|\mathbf{x}-\boldsymbol{\mu}\|^{2\alpha}\right), \mathbf{x}\in\mathbb{S}^{d-1},\\
\text{Axial Type, } \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu}): &\qquad f(\mathbf{x})\propto\exp\left(\frac{\kappa}{\alpha}|\boldsymbol{\mu}^T\mathbf{x}|^{\alpha}\right), \mathbf{x}\in\mathbb{S}^{d-1},
\end{aligned}$$

where $\kappa\in\mathbb{R}$ is a concentration parameter, and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$ is a mean direction parameter. The whole chapter denotes $x^{<\alpha>}$ by $|x|^\alpha\mathrm{sgn}(x), x\in\mathbb{R}$.

Beside the study of the properties, simulations and parameter estimation for distributions $\mathrm{GvMF}_{j,d}, (j=1,2,3)$, the goodness-of-fit test based on the estimation of the Shannon entropy and independent identically distributed (i.i.d.) sample are developed. These tests exploit the so-called maximum entropy principle, which is also proven in [83] as spherical analogous of results.

In addition, the entropy estimators $\hat{H}_{N,k}$ derived from k-th nearest neighbor distances is employed. Starting from the pioneering paper of Kozachenko and Leonenko [71], which proves by direct probability methods the consistency of $\hat{H}_{N,1}$ for random vectors with values in Euclidean space, a large number of authors consider to extend the set of admissible distributions and improve the convergence of $\hat{H}_{N,k}$, see [13, 18, 30, 38, 39, 44, 49, 75, 76], and the references therein. In the papers of [79] and

[90] , the k-th nearest neighbor entropy estimation are generalized for hyperspherical distributions.

Unlike the above mentioned works, the limit theory for point processes with a fixed k allows to prove the L^p-consistency of functionals of k-th nearest neighbor distances for a wider class of distributions. The nearest neigbors method of estimation of the Shannon entropy for manifolds, including spheres, was developed by [99]. In this chapter, the aim is to continue their work and prove the L^2-consistency of $\hat{H}_{N,k}$, as $N \to \infty$ with arbitrary $k \geq 1$ and for a random vector on a Riemannian manifold if its density is bounded and has compact support, see Theorem 4.4.6. Therefore, it is shown that $\hat{H}_{N,k}$ is a consistent estimator for the samples from the introduced generalized von Mises-Fisher distributions.

From the recent papers, [13] is mentioned, where the efficient entropy estimation is provided via the weighted k-nearest neighbour distances with $k = k_N$ depending on sample size N. Moreover, Berrett and Samworth [12] introduced a non-parametric entropy-based test of independence for multidimensional data. Lund and Jammalamadaka [82] considered the entropy-based test of goodness-of-fit for the von Mises distribution on the circle and use a different entropy estimate. The study in this chapter is motivated, particularly, by the work of Chapter 1, where the entropy-based goodness-of-fit test for generalized Gaussian distribution is given.

The theoretical results are verified by computational study on simulated samples and the inflation of variances of $\hat{H}_{N,k}$ is shown as k growths, which confirms the conclusion of [13]. Moreover, the evidence of generalized von Mises-Fisher distributions is detected in the real world data by the presented entropy-based goodness-of-fit test. Particularly, the parts in 3D images of a glass fibre reinforced composite material are examined, where fiber directions follow a generalized von Mises-Fisher distribution of axial type.

4.2 Preliminaries

This section provides some known facts needed for the sequel. Let $\sigma(d\mathbf{x})$ be spherical measure on the sphere $\mathbb{S}^{d-1}$. It can be written in polar coordinates $\mathbf{x} = (1, \mathbf{u}), \mathbf{u} \in \mathbb{S}^{d-1}$ as $\sigma(d\mathbf{x}) = 2^{-1}\pi^{-d/2}\Gamma(d/2)\,d\mathbf{u}$. Further, Lemma 2.5.1 from [40] is used for computation of integrals with respect to σ. Namely, let $g : \mathbb{R} \to \mathbb{R}_+$ be a non-negative Borel function and $\mathbf{a} \in \mathbb{S}^{d-1}$, then

$$\int_{\mathbf{x}^T\mathbf{x}=1} g(\mathbf{a}^T\mathbf{x})\sigma(d\mathbf{x}) = \frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)} \int_{-1}^{1} g(y)(1-y^2)^{\frac{d-3}{2}}\,dy. \tag{4.1}$$

The parameters $\kappa \in \mathbb{R}$, $\mu \in \mathbb{S}^{d-1}$, $d \geq 2$ are taken and further the probability densities is considered with respect to the measure σ.

Definition 4.2.1. *A unit random vector* $\mathbf{X}$ *has the* $(d-1)$*-dimensional von Mises-Fisher distribution* $\mathrm{vMF}_{1,d}(\boldsymbol{\mu}, \kappa)$ *if its probability density function is*

$$f_{\mathbf{X}}(\mathbf{x}) = (\kappa/2)^{d/2-1}\,(2\pi^{d/2} I_{d/2-1}(\kappa))^{-1}\exp(\kappa\boldsymbol{\mu}^T\mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^{d-1},$$

where I_ν *is the modified Bessel function of order* $\nu \geq 0$, *see e.g., [59, (A5)].*

Note that the modified Bessel function of the I_ν (see 9.6.18 in [36]) is given by

$$I_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\nu}\gamma(\nu + 1/2)} \int_0^\pi e^{\pm z\cos\nu} \sin^{2\nu} \nu d\,\nu.$$

In the case $d = 3$, von Mises-Fisher distribution $M_{1,3}(\boldsymbol{\mu}, \kappa)$ is called *Fisher distribution* and its density simplifies to $\kappa/(4\pi\sinh\kappa)\exp(\kappa\boldsymbol{\mu}^T\mathbf{x}), \mathbf{x} \in \mathbb{S}^2$. The density of $\mathrm{vMF}_{1,d}$ can be written in the alternative form. We say that a random vector $\mathbf{X}$ has the *von Mises-Fisher distribution* $\mathrm{vMF}_{2,d}(\boldsymbol{\mu}, \kappa)$ if its density function is $f_{\mathbf{X}}(\mathbf{x}) = e^{\kappa}\,(\kappa/2)^{d/2-1}\,(2\pi^{d/2} I_{d/2-1}(\kappa))^{-1}\exp\left(-\kappa\|\mathbf{x}-\boldsymbol{\mu}\|^2/2\right), \mathbf{x} \in \mathbb{S}^{d-1}$. Indeed, $\frac{\kappa}{2}\|\mathbf{x}-\boldsymbol{\mu}\|^2 = \frac{\kappa}{2}\|\mathbf{x}\|^2 + \frac{\kappa}{2}\|\boldsymbol{\mu}\|^2 - \kappa\boldsymbol{\mu}^T\mathbf{x} = \kappa - \kappa\boldsymbol{\mu}^T\mathbf{x}$ for $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{d-1}$. Let us recall the standard directional statistics.

Definition 4.2.2. *Let* $\mathbf{X}$ *be random vector with values in* $\mathbb{S}^{d-1}$ *and* $\mathbf{E}\mathbf{X} \neq \mathbf{0}$. *A mean direction of* $\mathbf{X}$ *is a vector* $\mathbf{E}\mathbf{X}/\|\mathbf{E}\mathbf{X}\|$. *A mean resultant length is* $\|\mathbf{E}\mathbf{X}\|$.

The mean resultant length is invariant and the mean direction is equivalent under rotation. Formally, let $U \in SO(d)$ be a rotation matrix, $SO(d)$ is that we will derive the basis functions for their irreducible representations which will be used to obtain the corresponding characters and to demonstrate their orthogonality in [86], then $\|\mathbf{E}U\mathbf{X}\| = \|\mathbf{E}\mathbf{X}\|$ and $\mathbf{E}U\mathbf{X}/\|\mathbf{E}U\mathbf{X}\| = U\mathbf{E}\mathbf{X}/\|\mathbf{E}\mathbf{X}\|$.

Consider the class of distributions on $\mathbb{S}^{d-1}$ with rotational symmetry, that is their distribution functions have a form $f(\mathbf{x}) = g(\boldsymbol{\mu}^T\mathbf{x})$, $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{d-1}$, e.g. [15]. Such random vectors $\mathbf{X}$ posses a *tangent-normal decomposition*

$$\mathbf{X} = (\boldsymbol{\mu}^T\mathbf{X})\boldsymbol{\mu} + \sqrt{1-(\boldsymbol{\mu}^T\mathbf{X})^2}\mathbf{Y}, \tag{4.2}$$

where $\boldsymbol{\mu}^T\mathbf{X}$ and $\mathbf{Y}$ are independent, $\boldsymbol{\mu} \perp \mathbf{Y}$, and $\mathbf{Y}$ is uniformly distributed on the tangent space $\mathbb{S}_{\mu}^{d-1} := \{\mathbf{y} \in \mathbb{S}^{d-1} | \boldsymbol{\mu}^T\mathbf{y} = 0\}$. It follows from (4.2), that the mean resultant length is $\|\mathbf{E}\mathbf{X}\| = \mathbf{E}[\boldsymbol{\mu}^T\mathbf{X}]$ and the mean direction equals $\boldsymbol{\mu}$.

4.3 Generalized von Mises-Fisher distributions

This section introduces the generalizations of the von Mises-Fisher distribution. $\kappa \in \mathbb{R}$ is named a concentration parameter and $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ is named a mean direction parameter.

Definition 4.3.1. *A unit random vector* $\boldsymbol{X}$ *has the* $(d-1)$*-dimensional Type I generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$ *of order* $\alpha > 0$ *if its probability density function with respect to the uniform distribution is*

$$f_{\boldsymbol{X}}(\mathbf{x}) = c_{1,d}(\kappa,\alpha)\exp\left(\frac{\kappa}{\alpha}(\boldsymbol{\mu}^T\mathbf{x})^{<\alpha>}\right), \mathbf{x} \in \mathbb{S}^{d-1}, \tag{4.3}$$

where

$$c_{1,d}(\kappa,\alpha) = \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)}\int_0^1 \left(e^{\frac{\kappa}{\alpha}y^\alpha} + e^{-\frac{\kappa}{\alpha}y^\alpha}\right)(1-y^2)^{\frac{d-3}{2}}dy\right)^{-1}. \tag{4.4}$$

As an analogue of von Mises-Fisher distribution in the form $\mathrm{vMF}_{2,d}$, the following class is introduced.

Definition 4.3.2. *A unit random vector* $\boldsymbol{X}$ *has the* $(d-1)$*-dimensional type II generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$ *of order* $\alpha > 0$ *if its probability density function with respect to the uniform distribution is*

$$f_{\boldsymbol{X}}(\mathbf{x}) = c_{2,d}(\kappa,\alpha)\exp\left(-\frac{\kappa}{2^\alpha\alpha}\|\mathbf{x}-\boldsymbol{\mu}\|^{2\alpha}\right), \mathbf{x} \in \mathbb{S}^{d-1}, \tag{4.5}$$

where

$$c_{2,d}(\kappa,\alpha) = \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)}\int_0^1 \left(e^{-\frac{\kappa}{\alpha}(1-y)^\alpha} + e^{-\frac{\kappa}{\alpha}(1+y)^\alpha}\right)(1-y^2)^{\frac{d-3}{2}}dy\right)^{-1}. \tag{4.6}$$

In the case of $\alpha = 1$, the introduced distributions $\mathrm{GvMF}_{1,d}$ and $\mathrm{GvMF}_{2,d}$ become the von Mises-Fisher distributions $\mathrm{vMF}_{1,d}$ and $\mathrm{vMF}_{2,d}$ respectively.

If we do not distinguish opposite directions we deal with axial. Commonly used technique in this case is to consider symmetric density functions f such that $f(\mathbf{x}) = f(-\mathbf{x}), \mathbf{x} \in \mathbb{S}^{d-1}$. Since the motivation is to stay in the class of rotational invariant densities and to generalize the von Mises-Fisher distribution, the following model is proposed for an axial data.

Definition 4.3.3. *A unit random vector* $\boldsymbol{X}$ *has the* $(d-1)$*-dimensional axial generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ *(or distribution of axial type) of order* $\alpha > 0$ *if its probability density function with respect to the uniform distribution is*

$$f_{\boldsymbol{X}}(\mathbf{x}) = c_{3,d}(\kappa,\alpha)\exp\left(\frac{\kappa}{\alpha}|\boldsymbol{\mu}^T\mathbf{x}|^\alpha\right), \mathbf{x} \in \mathbb{S}^{d-1}, \tag{4.7}$$

where $\kappa \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ *and*

$$c_{3,d}(\kappa,\alpha) = \left(\frac{4\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^1 e^{\frac{\kappa}{\alpha} y^\alpha} (1-y^2)^{\frac{d-3}{2}} dy \right)^{-1}. \tag{4.8}$$

Remark 4.3.1. *Let us check that* $\int_{\mathbb{S}^{d-1}} f_{\boldsymbol{X}}(\mathbf{x})\sigma(d\mathbf{x}) = 1$. *Then, the constant* $c_{1,d}(\kappa,\alpha)$ *equals*

$$\begin{aligned}
&\left(\int_{\mathbb{S}^{d-1}} \exp\left(\frac{\kappa}{\alpha} (\boldsymbol{\mu}^T \mathbf{x})^{<\alpha>} \right) \sigma(d\mathbf{x}) \right)^{-1} \\
&\overset{(4.1)}{=} \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^1 \exp\left(\frac{\kappa}{\alpha} y^{<\alpha>} \right) (1-y^2)^{\frac{d-3}{2}} dy \right)^{-1} \\
&= \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^1 \left(e^{\frac{\kappa}{\alpha} y^\alpha} + e^{-\frac{\kappa}{\alpha} y^\alpha} \right) (1-y^2)^{\frac{d-3}{2}} dy \right)^{-1}.
\end{aligned}$$

Similarly, the constant $c_{2,d}(\kappa,\alpha)$ *equals*

$$\begin{aligned}
&\left(\int_{\mathbb{S}^{d-1}} \exp\left(-\frac{\kappa}{2^\alpha \alpha} \|\mathbf{x}-\boldsymbol{\mu}\|^{2\alpha} \right) \sigma(d\mathbf{x}) \right)^{-1} = \left(\int_{\mathbb{S}^{d-1}} \exp\left(-\frac{\kappa}{\alpha} \left(1-\boldsymbol{\mu}^T \mathbf{x}\right)^\alpha \right) \sigma(d\mathbf{x}) \right)^{-1} \\
&\overset{(4.1)}{=} \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^1 \exp\left(-\frac{\kappa}{\alpha} (1-y)^\alpha \right) (1-y^2)^{\frac{d-3}{2}} dy \right)^{-1} \\
&= \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_0^1 \left(e^{-\frac{\kappa}{\alpha}(1-y)^\alpha} + e^{-\frac{\kappa}{\alpha}(1+y)^\alpha} \right) (1-y^2)^{\frac{d-3}{2}} dy \right)^{-1}.
\end{aligned}$$

Remark 4.3.2. *As usual for axial distributions, parameter* $\boldsymbol{\mu}$ *is defined up to a sign, in a sense that* $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ *and* $\mathrm{GvMF}_{3,d}(\alpha,\kappa,-\boldsymbol{\mu})$ *are equal. In the case* $\alpha = 2$, *the generalized von Mises-Fisher distribution of axial type reduces to the Watson distribution.*

Remark 4.3.3. *For* $d = 3$ *and* $\alpha = 1$, *one can represent an expansion of the von Mises-Fisher* $\mathrm{vMF}_{1,3}(\boldsymbol{\mu},\kappa)$ *density into the series of orthogonal functions* $Y_l^m, -l \le m \le l, l = 0,1,2,\ldots$, *on the sphere (real spherical harmonics, see e.g., [59, p. 437], e.g., [87, S. 13.2]). For example, [59, (5)] gives*

$$f(\mathbf{x}) = \sum_{l=0}^{\infty} \sqrt{\frac{2l+1}{4\pi}} \frac{I_{l+1/2}(\kappa)}{I_{1/2}(\kappa)} Y_l^0(\boldsymbol{\mu}^T \mathbf{x}), \mathbf{x} \in \mathbb{S}^2,$$

which can be used potentially for computational purposes. However, it is difficult for general α *to express coefficients of expansions in terms of some known special functions (this is true even for* $\alpha = 2$, *see formulae (7) and (8) in [59]).*

Note that the orthogonality of spherical harmonics can be defined as

$$\int_{S^{d-1}} Y_e^m(\gamma,\theta)Y_{e'}^{m'}(\gamma,\theta)\sigma(d\gamma,d\theta)=\delta_e^{e'}\delta_m^{m'},$$

where $Y_e^m, -e \le m \le e, \quad e=0,1,2,\dots,$ on the sphere.

4.3.1 Moments

This section considers the moments of $\mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu}), j=1,2,3$ distributions. Denote by

$$A_1(\kappa,\alpha,\beta)=\int_0^1 e^{\frac{\kappa}{\alpha}y^\alpha}y^\beta(1-y^2)^{\frac{d-3}{2}}dy, \tag{4.9}$$

$$A_2(\kappa,\alpha,\beta)=\int_0^2 e^{\frac{-\kappa}{\alpha}y^\alpha}a(2-y)^{\frac{d-3}{2}}y^{\frac{d-3}{2}+\beta}dy. \tag{4.10}$$

Proposition 4.3.1. *Let* $\beta\ge 0$ *and* $\mathbf{X}\sim\mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$*, then*

$$\mathbb{E}\left((\boldsymbol{\mu}^T\mathbf{X})^{<\beta>}\right)=\frac{A_1(\kappa,\alpha,\beta)-A_1(-\kappa,\alpha,\beta)}{A_1(\kappa,\alpha,0)+A_1(-\kappa,\alpha,0)}. \tag{4.11}$$

Proof. Let f be the density of the form (4.3), then

$$\begin{aligned}\mathbb{E}\left((\boldsymbol{\mu}^T\mathbf{X})^{<\beta>}\right)&=\int_{\mathbb{S}^{d-1}}(\boldsymbol{\mu}^T\mathbf{x})^{<\beta>}f(\mathbf{x})\sigma(d\mathbf{x})\\ &\overset{(4.1)}{=} c_{1,d}(\kappa,\alpha)\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)}\int_{-1}^1 y^{<\beta>}\exp\left(\frac{\kappa}{\alpha}y^{<\alpha>}\right)(1-y^2)^{\frac{d-3}{2}}dy\\ &=c_{1,d}(\kappa,\alpha)\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)}\int_0^1\left(e^{\frac{\kappa}{\alpha}y^\alpha}-e^{-\frac{\kappa}{\alpha}y^\alpha}\right)y^\beta(1-y^2)^{\frac{d-3}{2}}dy\\ &=\frac{\int_0^1\left(e^{\frac{\kappa}{\alpha}y^\alpha}-e^{-\frac{\kappa}{\alpha}y^\alpha}\right)y^\beta(1-y^2)^{\frac{d-3}{2}}dy}{\int_0^1\left(e^{\frac{\kappa}{\alpha}y^\alpha}+e^{-\frac{\kappa}{\alpha}y^\alpha}\right)(1-y^2)^{\frac{d-3}{2}}dy}.\end{aligned}$$

□

Particularly, the mean direction of $\mathbf{X}\sim\mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$ is $\boldsymbol{\mu}$ and its mean resultant length equals

$$||\mathbb{E}[\mathbf{X}]||=\mathbb{E}[\boldsymbol{\mu}^T\mathbf{X}]=\frac{A_1(\kappa,\alpha,1)-A_1(-\kappa,\alpha,1)}{A_1(\kappa,\alpha,0)+A_1(-\kappa,\alpha,0)}. \tag{4.12}$$

Proposition 4.3.2. *Let* $\beta\ge 0$ *and* $\mathbf{X}\sim\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$*, then*

$$\mathbb{E}||\mathbf{X}-\boldsymbol{\mu}||^{2\beta}=2^\beta\frac{A_2(\kappa,\alpha,\beta)}{A_2(\kappa,\alpha,0)}. \tag{4.13}$$

Proof. Let f be the density of the form (4.5), then

$$\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\beta} = \int_{\mathbb{S}^{d-1}} \|\mathbf{x}-\boldsymbol{\mu}\|^{2\beta} f(\mathbf{x})\sigma(d\mathbf{x}) = \int_{\mathbb{S}^{d-1}} (2-2\boldsymbol{\mu}^T\mathbf{x})^{\beta} f(\mathbf{x})\sigma(d\mathbf{x})$$
$$\overset{(4.1)}{=} c_{2,d}(\kappa,\alpha)\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^{1} (2-2y)^{\beta} \exp\left(-\frac{\kappa}{\alpha}(1-y)^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}} dy$$
$$= 2^{\beta} \frac{\int_{-1}^{1}(1-y)^{\beta} e^{-\frac{\kappa}{\alpha}(1-y)^{\alpha}}(1-y^2)^{\frac{d-3}{2}} dy}{\int_{-1}^{1} e^{-\frac{\kappa}{\alpha}(1-y)^{\alpha}}(1-y^2)^{\frac{d-3}{2}} dy}$$
$$= 2^{\beta} \frac{\int_0^2 e^{-\frac{\kappa}{\alpha}z^{\alpha}}(2-z)^{\frac{d-3}{2}} z^{\frac{d-3}{2}+\beta} dz}{\int_0^2 e^{-\frac{\kappa}{\alpha}z^{\alpha}}(2-z)^{\frac{d-3}{2}} z^{\frac{d-3}{2}} dz}.$$

□

Particularly, the mean direction of $\mathbf{X} \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$ is $\boldsymbol{\mu}$ and its mean resultant length equals

$$\|\mathbb{E}[\mathbf{X}]\| = \mathbb{E}[\boldsymbol{\mu}^T\mathbf{X}] = 1 - \frac{A_2(\kappa,\alpha,1)}{A_2(\kappa,\alpha,0)}. \tag{4.14}$$

Proposition 4.3.3. *Let $\beta \geq 0$ and $\mathbf{X} \sim \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$, then*

$$\mathbb{E}\left(|\boldsymbol{\mu}^T\mathbf{X}|^{\beta}\right) = \frac{A_1(\kappa,\alpha,1)}{A_1(\kappa,\alpha,0)}. \tag{4.15}$$

Proof. Let f be the density of the form (4.7), then

$$\mathbb{E}\left(|\boldsymbol{\mu}^T\mathbf{X}|^{\beta}\right) = \int_{\mathbb{S}^{d-1}} |\boldsymbol{\mu}^T\mathbf{x}|^{\beta} f(\mathbf{x})\sigma(d\mathbf{x})$$
$$\overset{(4.1)}{=} c_{3,d}(\kappa,\alpha)\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^{1} |y|^{\beta} \exp\left(\frac{\kappa}{\alpha}|y|^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}} dy$$
$$= \frac{\int_0^1 e^{\frac{\kappa}{\alpha}y^{\alpha}} y^{\beta}(1-y^2)^{\frac{d-3}{2}} dy}{\int_0^1 e^{\frac{\kappa}{\alpha}y^{\alpha}}(1-y^2)^{\frac{d-3}{2}} dy}.$$

□

Note that the mean direction of $\mathbf{X} \sim \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ is not defined and its mean resultant length $\|\mathbb{E}[\mathbf{X}]\| = \mathbb{E}[\boldsymbol{\mu}^T\mathbf{X}]$ equals 0.

4.4 Entropy

This section finds the entropy of generalized von Mises-Fisher distributions, and shows the maximum entropy principle for them. Then, the statistical estimation of an entropy is discussed and the L^2 convergence of the k-th nearest neigbour estimator is proven for random variables on compact manifolds.

4.4.1 Maximum entropy principle for generalized von-Mises Fisher distributions

Recall that the entropy of a continuous random vector $\mathbf{X} \in \mathbb{S}^{d-1}$ with a density f is

$$H(\mathbf{X}) = -\int_{\mathbb{S}^{d-1}} (\log f(\mathbf{x})) f(\mathbf{x}) \sigma(d\mathbf{x}). \tag{4.16}$$

For a density version f, its support is denoted by supp$f = \{\mathbf{x} \in \mathbb{S}^{d-1} : f(\mathbf{x}) > 0\}$. Clearly, the integral in $H(\mathbf{X})$ is taken over supp$f$.

Theorem 4.4.1. *Let* $\mathbf{X}_j \sim \mathrm{GvMF}_{j,d}(\alpha, \kappa, \boldsymbol{\mu})$, $j = 1, 2, 3$, *then*

$$H(\mathbf{X}_1) = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \mathbb{E}\left((\boldsymbol{\mu}^T \mathbf{X}_1)^{<\alpha>}\right), \tag{4.17}$$

$$H(\mathbf{X}_2) = -\log c_{2,d}(\kappa, \alpha) + \frac{\kappa}{2^\alpha \alpha} \mathbb{E}\|\mathbf{X}_2 - \boldsymbol{\mu}\|^{2\alpha} \tag{4.18}$$

$$H(\mathbf{X}_3) = -\log c_{3,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \mathbb{E}|\boldsymbol{\mu}^T \mathbf{X}_3|^\alpha. \tag{4.19}$$

Proof. Let $\mathbf{X_1}$ have density f_1, then the entropy of $\mathbf{X_1}$ equals

$$-\int_{\mathbb{S}^{d-1}} (\log f_1(\mathbf{x})) f_1(\mathbf{x}) \sigma(d\mathbf{x}) = -\log c_{1,d}(\kappa, \alpha) \int_{\mathbb{S}^{d-1}} f_1(\mathbf{x}) \sigma(d\mathbf{x})$$
$$-\frac{\kappa}{\alpha} \int_{\mathbb{S}^{d-1}} (\boldsymbol{\mu}^T \mathbf{x})^{<\alpha>} f_1(\mathbf{x}) \sigma(d\mathbf{x}) = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \mathbb{E}\left((\boldsymbol{\mu}^T \mathbf{X}_1)^{<\alpha>}\right).$$

For $\mathbf{X_2}$ with density f_2 it follows that

$$H(\mathbf{X_2}) = -\int_{\mathbb{S}^{d-1}} (\log f_2(\mathbf{x})) f_2(\mathbf{x}) \sigma(d\mathbf{x}) = -\log c_{2,d}(\kappa, \alpha) \int_{\mathbb{S}^{d-1}} f_2(\mathbf{x}) \sigma(d\mathbf{x})$$
$$+\frac{\kappa}{2^\alpha \alpha} \int_{\mathbb{S}^{d-1}} \|\mathbf{x} - \boldsymbol{\mu}\|^{2\alpha} f(\mathbf{x}) \sigma(d\mathbf{x}) = -\log c_{2,d}(\kappa, \alpha) + \frac{\kappa}{2^\alpha \alpha} \mathbb{E}\|\mathbf{X}_2 - \boldsymbol{\mu}\|^{2\alpha}.$$

For $\mathbf{X_3}$ with density f_3

$$H(\mathbf{X_3}) = -\int_{\mathbb{S}^{d-1}} (\log f_3(\mathbf{x})) f_3(\mathbf{x}) \sigma(d\mathbf{x}) = -\log c_{3,d}(\kappa, \alpha) \int_{\mathbb{S}^{d-1}} f_3(\mathbf{x}) \sigma(d\mathbf{x})$$
$$-\frac{\kappa}{\alpha} \int_{\mathbb{S}^{d-1}} \|(\boldsymbol{\mu}^T \mathbf{X}_3)\|^\alpha f(\mathbf{x}) \sigma(d\mathbf{x}) = -\log c_{3,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \mathbb{E}|\boldsymbol{\mu}^T \mathbf{X}_3|^\alpha.$$

$\square$

Theorem 4.4.2. *Let a unit random vector* $\mathbf{Z} \in \mathbb{S}^{d-1}$ *have a generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{1,d}(\alpha, \kappa, \boldsymbol{\mu})$*. Then* $\mathbf{Z}$ *has the maximum entropy value over all continuous random variables* $\mathbf{X}$ *on* $\mathbb{S}^{d-1}$ *with*

$$\mathbb{E}\left((\boldsymbol{\mu}^T \mathbf{X})^{<\alpha>}\right) = \mathbb{E}\left((\boldsymbol{\mu}^T \mathbf{Z})^{<\alpha>}\right). \tag{4.20}$$

Proof. Let $\mathbf{X}$ be a random unit vector on $\mathbb{S}^{d-1}, d \geq 2$ such that (4.20) holds true. Let f and f^* be the densities of $\mathbf{X}$ and $\mathbf{Z}$ respectively. By Jensen's inequality,

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \log f^*(\mathbf{x}) \sigma(d\mathbf{x}) - \int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \log f(\mathbf{x}) \sigma(d\mathbf{x}) \\
&= \int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \log \frac{f^*(\mathbf{x})}{f(\mathbf{x})} \sigma(d\mathbf{x}) \leq \log\left(\int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \frac{f^*(\mathbf{x})}{f(\mathbf{x})} \sigma(d\mathbf{x})\right) = 0
\end{aligned}$$

with equality if and only if $f = f^*$ almost everywhere with respect to the Lebesgue measure on $\mathbb{S}^{d-1}$. So,

$$H(\mathbf{X}) = -\int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \log f(\mathbf{x}) \sigma(d\mathbf{x}) \leq -\int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \log f^*(\mathbf{x}) \sigma(d\mathbf{x}). \tag{4.21}$$

In this case $f^*(\mathbf{x}) = \log c_{1,d}(\kappa, \alpha) + \frac{\kappa}{\alpha} |\boldsymbol{\mu}^T \mathbf{x}|^{<\alpha>}, \mathbf{x} \in \mathbb{S}^{d-1}$ and hence

$$\begin{aligned}
H(\mathbf{X}) &\leq -\int_{\mathbb{S}^{d-1}} f(\mathbf{x}) \log f^*(\mathbf{x}) \sigma(d\mathbf{x}) \\
&= -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \int_{\mathbb{S}^{d-1}} |\boldsymbol{\mu}^T \mathbf{x}|^{<\alpha>} f(\mathbf{x}) \sigma(d\mathbf{x}) \\
&= -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \mathbb{E}\left[(\boldsymbol{\mu}^T \mathbf{X})^{<\alpha>}\right] = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \mathbb{E}\left[(\boldsymbol{\mu}^T \mathbf{Z})^{<\alpha>}\right] \\
&= -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \int_{\mathbb{S}^{d-1}} (\boldsymbol{\mu}^T \mathbf{x})^{<\alpha>} f^*(\mathbf{x}) \sigma(d\mathbf{x}) \\
&= -\int_{\mathbb{S}^{d-1}} f^*(\mathbf{x}) \log f^*(\mathbf{x}) \sigma(d\mathbf{x}) = H(\mathbf{Z}).
\end{aligned}$$

□

The maximum entropy principle for generalized von Mises-Fisher distribution of Type II has the following form.

Theorem 4.4.3. *Let a unit random vector* $\mathbf{Z} \in \mathbb{S}^{d-1}$ *have a generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{2,d}(\alpha, \kappa, \boldsymbol{\mu})$*. Then* $\mathbf{Z}$ *has the maximum entropy value over all continuous random variables* $\mathbf{X}$ *on* $\mathbb{S}^{d-1}$ *with*

$$\mathbb{E}\|\mathbf{X} - \boldsymbol{\mu}\|^{2\alpha} = \mathbb{E}\|\mathbf{Z} - \boldsymbol{\mu}\|^{2\alpha}. \tag{4.22}$$

Proof. Let f and f^* be the densities of $\mathbf{X}$ and $\mathbf{Z}$ respectively. The proof is similar to Theorem 4.4.2. Indeed, $f^*(\mathbf{x}) = \log c_{2,d}(\kappa,\alpha) - \frac{\kappa}{2^\alpha \alpha}||\mathbf{x}-\boldsymbol{\mu}||^{2\alpha}, \mathbf{x} \in \mathbb{S}^{d-1}$ and hence

$$\begin{aligned}
H(\mathbf{X}) &\le -\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) \\
&= -\log c_{2,d}(\kappa,\alpha) + \frac{\kappa}{2^\alpha\alpha}\int_{\mathbb{S}^{d-1}} ||\mathbf{x}-\boldsymbol{\mu}||^{2\alpha} f(\mathbf{x})\sigma(d\mathbf{x}) \\
&= -\log c_{2,d}(\kappa,\alpha) + \frac{\kappa}{2^\alpha\alpha}\mathbb{E}||\mathbf{X}-\boldsymbol{\mu}||^{2\alpha} = -\log c_{2,d}(\kappa,\alpha) + \frac{\kappa}{2^\alpha\alpha}\mathbb{E}||\mathbf{Z}-\boldsymbol{\mu}||^{2\alpha} \\
&= -\log c_{2,d}(\kappa,\alpha) + \frac{\kappa}{2^\alpha\alpha}\int_{\mathbb{S}^{d-1}} ||\mathbf{x}-\boldsymbol{\mu}||^{2\alpha} f^*(\mathbf{x})\sigma(d\mathbf{x}) \\
&= -\int_{\mathbb{S}^{d-1}} f^*(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) = H(\mathbf{Z}).
\end{aligned}$$

□

Theorem 4.4.4. *Let a unit random vector* $\mathbf{Z} \in \mathbb{S}^{d-1}$ *have an axial generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$*. Then* $\mathbf{Z}$ *has the maximum entropy value over all continuous random variables* $\mathbf{X}$ *on* $\mathbb{S}^{d-1}$ *with*

$$\mathbb{E}|\boldsymbol{\mu}^T\mathbf{X}|^\alpha = \mathbb{E}|\boldsymbol{\mu}^T\mathbf{Z}|^\alpha. \tag{4.23}$$

Proof. The proof is similar to Theorems 4.4.2 and 4.4.3. In this case, $f^*(\mathbf{x}) = \log c_{3,d}(\kappa,\alpha) + \frac{\kappa}{\alpha}|\boldsymbol{\mu}^T\mathbf{x}|^\alpha, \mathbf{x} \in \mathbb{S}^{d-1}$ and

$$\begin{aligned}
H(\mathbf{X}) &\le -\log c_{3,d}(\kappa,\alpha) - \frac{\kappa}{\alpha}\int_{\mathbb{S}^{d-1}} |\boldsymbol{\mu}^T\mathbf{x}|^\alpha f(\mathbf{x})\sigma(d\mathbf{x}) \\
&= -\log c_{3,d}(\kappa,\alpha) - \frac{\kappa}{\alpha}\mathbb{E}|\boldsymbol{\mu}^T\mathbf{X}|^\alpha = -\log c_{3,d}(\kappa,\alpha) - \frac{\kappa}{\alpha}\mathbb{E}|\boldsymbol{\mu}^T\mathbf{Z}|^\alpha = H(\mathbf{Z}).
\end{aligned}$$

□

4.4.2 Entropy estimation

This section gives the method of an entropy estimation for unit random vectors. Actually, the phase-state is extended to the arbitrary compact Riemannian manifold.

Let $m,d \in \mathbb{N}$, $m \le d$, and $\mathscr{M}$ be a $m-$dimensional C^1 manifold embeded in $\mathbb{R}^d$ with the atlas $((U_i,g_i), i \in I_0)$, i.e., for each $y \in \mathscr{M}$ there exists an open subset U_i of R^m and a continuously differentiable injection $g_i : U_i \to \mathbb{R}^d$, such that $y \in g_i(U) \subset \mathscr{M}$, and g_i is an open map from U_i to $\mathscr{M}$, and the linear map $g_i'(u)$ has full rank for all $u \in U_i$.

For bounded measurable $h : \mathscr{M} \to \mathbb{R}$, the integral $\int_{\mathscr{M}} h(y)\nu(dy)$ is defined by

$$\int_{\mathscr{M}} h(y)\nu(dy) = \sum_{i\in I_0}\int_{U_i} h(g_i(x))\psi_i(g_i(x))\sqrt{\det(J_{g_i}(x))^T(J_{g_i}(x))}dx,$$

where ν is a $\sigma-$finite measure on $\mathscr{M}$, J_{g_i} is the Jacobian of g_i and $\{\psi_i, i \in I_0\}$ is the partition of unity, see [99, pp. 3-4] and [11, Chapter 2] for more detailed setting.

Let $f : \mathscr{M} \to \mathbb{R}_+$ be a probability density of independent random elements $X, X_i, i \in \mathbb{N}$ with values in $\mathscr{M}$, i.e., $\int_{\mathscr{M}} f(x)\nu(dx) = 1$. Denote by $\mathscr{X}_N = \{X_1,\dots,X_N\}$, $N \geq k$ the samples of the first N elements. The entropy of X equals

$$H(X) = -\int_{\mathscr{M}} \log(f(x))f(x)\nu(dx).$$

Let F be a finite subset of $\{X_i, i \geq k\}$ and $\rho_k(x,F)$ be the Euclidean distance between x and its k-th nearest neighbor in $F \setminus \{x\}$. Let $\gamma \approx 0.5772$ be the Euler-Mascheroni constant.

Definition 4.4.1. *The k-th nearest neighbour estimator of the entropy $H(X)$ is defined by*

$$\widehat{H}_{N,k}(\mathscr{X}_N) = \frac{1}{N}\sum_{i=1}^{N}\log\left(\rho_k^m(X_i,\mathscr{X}_N)V_m(N-1)e^{-\psi(k)}\right), \tag{4.24}$$

where

$$\psi(k) = \sum_{j=1}^{k-1}\frac{1}{j} - \gamma, \quad V_m = \frac{\pi^{m/2}}{\Gamma(1+m/2)}.$$

Recall that for $k = 1$, $\psi(1) = -\gamma$ with $\gamma = 0.57726...$ is the Euler constant, while for $k \geq 1$, $\psi(k) = -\gamma + A_{k-1}$, with $A_0 = 0$, and $A_j = \sum_{i=1}^{j}\frac{1}{i}$, so that $\exp\{\psi(k)\}/k \to 1$ as $k \to \infty$. For example, if $k = 1$, then the nearest neighbor distances estimator (NNE) is

$$\widehat{H}_{N,1}(\mathscr{X}_N) = \frac{m}{N}\sum_{i=1}^{N}\log\rho_1(X_i,\mathscr{X}_N) + \log V_m + \gamma + \log(N-1). \tag{4.25}$$

Let us start the L^2-consistency of $\widehat{H}_{N,k}$ by writing down the particular case of theorem of [99, Theorem 3.1] for the functional

$$\xi(x,\mathscr{X}) := \log\left(e^{-\psi(k)}V_m\rho_k^m(x,\mathscr{X})\right). \tag{4.26}$$

Theorem 4.4.5. *Let $k \geq 1$ put $q = 1$ or $q = 2$. Suppose there exists $p \geq q$ such that*

$$\sup_{N\geq k}\mathbb{E}\left|\xi\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right|^p < \infty. \tag{4.27}$$

Then as $N \to \infty$ we have L^q convergence

$$\frac{1}{N} \sum_{x \in \mathscr{X}_N} \xi\left(N^{\frac{1}{m}} x, N^{\frac{1}{m}} \mathscr{X}_N\right) \to \int_{\mathscr{M}} \mathbb{E}[\xi\left(\mathbf{0}, \mathscr{P}_{f(x)}\right)] f(x) \nu(dx), \tag{4.28}$$

where $\mathscr{P}_\lambda$ denotes a homogeneous Poisson point process of intensity $\lambda > 0$ in $\mathbb{R}^m$ (embedded in $\mathbb{R}^d$).

For the bounded random variables $X_i, i \geq 1$ and $\rho_k(x, \mathscr{X}_N)$, we generalize Lemma 7.8 from [99], which was proved for the case $k = 1$.

Lemma 4.4.1. *Let f is bounded and has compact support on $\mathscr{M}$, then for any $\delta \in (0, m)$*

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

Proof. The proof is very similar to [99, Lemma 7.8]. Recall that $\mathscr{M}$ has the atlas $((U_i, g_i), i \in I_0)$, where $I_0 = \{1, \ldots, i_0\}$, and there exist $\delta_i, x_i, i \in I_0$ such that $\mathscr{M} \in \cup_{i \in I_0} B_{\delta_i}(y_i)$.

Denote $A_i = B_{\delta_i} \setminus \cup_{j<i} B_{\delta_j}(y_j)$. Since supp$(f)$ is bounded then there exist $i_0 \in \mathbb{N}$ and constant $C > 0$ such that

$$\begin{aligned}
\mathbb{E}\left[N^{\frac{\delta}{m}} \rho_k^\delta(X_1, \mathscr{X}_N)\right] &= N^{\frac{\delta}{m}-1} \mathbb{E}\left(\sum_{x \in \mathscr{X}_N} \rho_k^\delta(x, \mathscr{X}_N)\right) \\
&\leq N^{\frac{\delta}{m}-1} \left[\sum_{i=1}^{i_0} \sum_{x \in A_i \cap \mathscr{X}_N} \rho_k^\delta(x, A_i \cap \mathscr{X}_N) + C\right].
\end{aligned} \tag{4.29}$$

Now, we prove that for all finite $\mathscr{Y} \subset A_i$

$$\sum_{x \in \mathscr{Y}} \rho_k^\delta(x, \mathscr{Y}) \leq C_i \left[\mathrm{card}(\mathscr{Y})\right]^{1-\frac{\delta}{m}},$$

where $C_i > 0$. Let $\mathscr{Y} \subset A_i$ and $y_j \in \mathscr{Y}$ be a the j-th nearest neighbor of $x \in \mathscr{Y}$. Taking $z_j \in \mathscr{Y}$ such that $g_i^{-1}(z_j)$ to be j-th nearest neighbor of $g_i^{-1}(x)$ in $g_i^{-1}(\mathscr{Y})$, it is obtained from [99, Lemma 4.1] that

$$\begin{aligned}
\rho_k(x, \mathscr{Y}) &= \max\{\|y_1 - x\|, \ldots, \|y_k - x\|\} \leq \max\{\|z_1 - x\|, \ldots, \|z_k - x\|\} \\
&\leq C_i \max\left\{\left\|g_i^{-1}(z_1) - g_i^{-1}(x)\right\|, \ldots, \left\|g_i^{-1}(z_k) - g_i^{-1}(x)\right\|\right\} \\
&= C_i \rho_k(g_i^{-1}(x), g_i^{-1}(\mathscr{Y})).
\end{aligned}$$

Thus, from [129, Lemma 3.3] we have for any $\delta \in (0,m)$

$$\sum_{x \in A_i \cap \mathscr{X}_N} \rho_k^\delta(x, A_i \cap \mathscr{X}_N)$$
$$\leq C_i[\mathrm{diam}(g_i^{-1}(A_i \cap \mathscr{X}_N))]^\delta [\mathrm{card}(g_i^{-1}(A_i \cap \mathscr{X}_N))]^{1-\frac{\delta}{m}} \leq \tilde{C}_i N^{1-\frac{\delta}{m}}.$$

Hence, the right hand side of (4.29) is bounded above uniformly. □

Now we prove the L^2 convergence of the k-th nearest neighbor estimator

$$\widehat{H}_{N,k}(\mathscr{X}_N) = \frac{1}{N} \sum_{x \in \mathscr{X}_N} \xi\left(N^{\frac{1}{m}} x, N^{\frac{1}{m}} \mathscr{X}_N\right), \tag{4.30}$$

which is an extension from the case $k = 1$ to $k \geq 1$ of Theorem 2.4 from [99].

Theorem 4.4.6. *Suppose f is bounded and has compact support. Then for every fixed $k \geq 1$*

$$\mathbb{E}\left[\widehat{H}_{N,k}(\mathscr{X}_N) - H(X)\right]^2 \to 0 \quad \textit{as } N \to \infty. \tag{4.31}$$

Proof. Theorem 4.4.5 is applied. First, $\mathbb{E}[\xi(\mathbf{0}, \mathscr{P}_\lambda)]$ is computed, where

$$\xi\left(\mathbf{0}, \mathscr{P}_{f(x)}\right) = \log V_m - \psi(k) + m \log \rho_k(\mathbf{0}, \mathscr{P}_\lambda).$$

The random variable $\rho_k(\mathbf{0}, \mathscr{P}_\lambda)$ is the distance to the kth point of $\mathscr{P}_\lambda$ from $\mathbf{0}$ and $\{\rho_k(\mathbf{0}, \mathscr{P}_\lambda) \leq t\} = \{[x \in \mathscr{P}_\lambda, x \in B_t(\mathbf{0})(t)] \geq k\}$. Random variable $[x \in \mathscr{P}_\lambda, x \in B_t(\mathbf{0})(t)]$ has Erland distribution. Therefore, $\rho_k(\mathbf{0}, \mathscr{P}_\lambda)$ also has the Erlang distribution with parameters k and $\lambda|B_t(\mathbf{0})| = \lambda t^m V_m$, that is

$$\mathbb{P}(\rho_k(\mathbf{0}, \mathscr{P}_\lambda) \leq t) = \mathbb{P}(\mathscr{P}_\lambda \cap B_t(\mathbf{0}) \geq k)$$
$$= 1 - \sum_{j=0}^{k-1} \frac{1}{j!} e^{-\lambda|B_t(\mathbf{0})|}(\lambda|B_t(\mathbf{0})|)^j = 1 - \sum_{j=0}^{k-1} \frac{1}{j!} e^{-\lambda t^m V_m}(\lambda t^m V_m)^j, t \geq 0.$$

Then

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

Thus, repeating the lines of proof of Theorem 2.1.3 in Chapter 2, it is obtained that

$$\int_{\mathscr{M}} \mathbb{E}[\xi\left(\mathbf{0}, \mathscr{P}_{f(x)}\right)] f(x) \nu(dx) = -\int_{\mathscr{M}} (\log f(x)) f(x) \nu(dx) = H(X).$$

Second, condition (4.27) is checked. Note that for every $\delta \in (0,1)$ and $p > 1$ there exists $C > 0$ such that $|\log t|^p \leq Ct^{-\delta}\mathbb{1}_{[0,1]}(t) + Ct^{\delta}\mathbb{1}_{[1,\infty)}(t)$, $t > 0$. Then

$$
\begin{aligned}
&\mathbb{E}\left|\xi\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right|^p \leq 2^{p-1}\left|\log V_m - \psi(k)\right|^p \\
&+ 2^{p-1}\mathbb{E}\left|\log \rho_k^m\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right|^p \leq 2^{p-1}\left|\log V_m - \psi(k)\right|^p \\
&+ 2^{p-1}C\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\mathbb{1}_{[0,1]}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right) \qquad (4.32)\\
&+ 2^{p-1}C\mathbb{E}\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\mathbb{1}_{[1,\infty)}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right). \qquad (4.33)
\end{aligned}
$$

Term (4.32) is finite because

$$
\begin{aligned}
&\sup_{N\geq k}\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\mathbb{1}_{[0,1]}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right)\right)\\
&\leq \sup_{N\geq k}\mathbb{E}\rho_1^{-\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right) < \infty, \qquad (4.34)
\end{aligned}
$$

where (4.34) is ensured by [99, Lemma 7.5] if f is bounded and $\delta \in (0, m)$. Hence, for (4.27) to be satisfied, it remains to show that

$$
\sup_{N\geq k}\mathbb{E}\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathscr{X}_N\right) < \infty. \qquad (4.35)
$$

Thus, applying Lemma 4.4.1 we get that (4.35) holds true if $0 < \delta < m$. □

The 2-dimensional sphere $\mathbb{S}^2$ is a compact manifold with $d = 3$ $m = 2$ and $\nu = \sigma$. Thus, Theorem 4.4.3 is valid for all bounded densities on $\mathbb{S}^2$, the k-th nearest neighbour estimator has the form

$$
\widehat{H}_{N,k}(\mathscr{X}_N) = \frac{2}{N}\sum_{i=1}^{N}\log\rho_k(X_i, \mathscr{X}_N) - \psi(k) + \log(N-1) + \log\pi,
$$

and $\widehat{H}_{N,k}(\mathscr{X}_N) \to H(X)$ in $L^2(\Omega)$. This yields that $\widehat{H}_{N,k}(\mathscr{X}_N)$ is a consistent estimator of the Shannon entropy.

4.5 Estimation of parameters

4.5.1 Fisher's maximum likelihood estimation

Let $\mathscr{X}_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a random sample. The log-likelihood $l(\mathscr{X}_N)$ is written down for random samples from the introduced generalized von Mises-Fisher distributions.

Lemma 4.5.1. *Let* $\mathscr{X}_{j,N} \sim \mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu})$, $j = 1,2,3$ *then*

$$l(\mathscr{X}_{1,N}) = N \log c_{1,d}(\kappa,\alpha) + \frac{\kappa}{\alpha}\sum_{i=1}^{N}(\boldsymbol{\mu}^T \mathbf{x}_i)^{<\alpha>}, \tag{4.36}$$

$$l(\mathscr{X}_{2,N}) = N \log c_{2,d}(\kappa,\alpha) - \frac{\kappa}{2^\alpha \alpha}\sum_{i=1}^{N}||\mathbf{x}_i - \boldsymbol{\mu}||^{2\alpha}, \tag{4.37}$$

$$l(\mathscr{X}_{3,N}) = N \log c_{3,d}(\kappa,\alpha) + \frac{\kappa}{\alpha}\sum_{i=1}^{N}|\boldsymbol{\mu}^T \mathbf{x}_i|^{\alpha}. \tag{4.38}$$

Proof. The statements follow from the direct calculation of $l(\mathscr{X}_N)$. □

For each case, numerical methods are used to find the maximum likelihood estimates $(\hat{\boldsymbol{\mu}}_L, \hat{\kappa}_L, \hat{\alpha}_L)$ of $(\boldsymbol{\mu},\kappa,\alpha)$ which maximize the log-likelihoods (4.36)-(4.38).

The problem becomes easier when parameter α is known. In such a case, the maximum likelihood estimates is derived taking derivatives of (4.36)-(4.38). The estimator of μ and κ is defined as

- Let $\mathscr{X}_N \sim \mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$, then $\hat{\boldsymbol{\mu}}_L = \arg\max_{\boldsymbol{\mu}\in\mathbb{S}^{d-1}} \sum_{i=1}^{N}(\boldsymbol{\mu}^T \mathbf{x}_i)^{<\alpha>}$,

$$\frac{A_1(\hat{\kappa}_L,\alpha,\alpha) - A_1(-\hat{\kappa}_L,\alpha,\alpha)}{A_1(\hat{\kappa}_L,\alpha,0) + A_1(-\hat{\kappa}_L,\alpha,0)} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\mu}_L^T \mathbf{x}_i)^{<\alpha>}.$$

- Let $\mathscr{X}_N \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, then $\hat{\boldsymbol{\mu}}_L = \arg\min_{\boldsymbol{\mu}\in\mathbb{S}^{d-1}} \sum_{i=1}^{N}||\mathbf{x}_i - \boldsymbol{\mu}||^{2\alpha}$,

$$2^\alpha \frac{A_2(\hat{\kappa}_L,\alpha,\alpha)}{A_2(\hat{\kappa}_L,\alpha,0)} = \sum_{i=1}^{N}||\mathbf{x}_i - \hat{\boldsymbol{\mu}}_L||^{2\alpha}.$$

- Let $\mathscr{X}_N \sim \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$, then $\hat{\boldsymbol{\mu}}_L = \arg\max_{\boldsymbol{\mu}\in\mathbb{S}^{d-1}} \sum_{i=1}^{N}|\boldsymbol{\mu}^T \mathbf{x}_i|^{\alpha}$,

$$\frac{A_1(\hat{\kappa}_L,\alpha,\alpha)}{A_1(\hat{\kappa}_L,\alpha,0)} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\boldsymbol{\mu}}_L^T \mathbf{x}_i|^{\alpha}.$$

4.5.2 Method of moments and generilized von Mises-Fisher distributions of I and II types

This section considers parameter estimation of generilized von Mises-Fisher distributions based on moments estimation. In the case of non-axial random vector $\mathbf{X} \in \mathbb{S}^{d-1}$, it is assumed that $\|\mathbb{E}\mathbf{X}\| \neq 0$. It is known from Definition 4.2.2 that $\hat{\boldsymbol{\mu}} := \bar{\mathscr{X}}_N / \|\bar{\mathscr{X}}_N\|$, where $\bar{\mathscr{X}}_N = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i$, is the natural estimator for mean direction parameter $\boldsymbol{\mu}$. In order to find estimates for parameters α and κ at least two more moment statistics are needed. A standard approach involves the use of resultant length. For the second relations, $\mathbb{E}(\operatorname{sgn}\{\mathbf{X}_1^T \mathbb{E}\mathbf{X}_1\})$ and $\mathbb{E}(\|\mathbf{x}_2 - \boldsymbol{\mu}\|^4)$ are chosen for vectors $\mathbf{X}_j \sim \mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu})(j = 1,2)$. Applying (4.12) and (4.14), one can get the estimators $\hat{\kappa}$, $\hat{\alpha}$ as a solution of the following equations

$$\frac{A_1(\hat{\kappa},\hat{\alpha},1) - A_1(-\hat{\kappa},\hat{\alpha},1)}{A_1(\hat{\kappa},\hat{\alpha},0) + A_1(-\hat{\kappa},\hat{\alpha},0)} = \|\bar{\mathscr{X}}_{1,N}\|$$
$$\frac{A_1(\hat{\kappa},\hat{\alpha},1) - A_1(-\hat{\kappa},\hat{\alpha},1)}{A_1(\hat{\kappa},\hat{\alpha},0) + A_1(-\hat{\kappa},\hat{\alpha},0)} = \frac{\sum_{i=1}^N \operatorname{sgn}(x_i^T)\bar{\mathscr{X}}_{1,N}}{N\|\bar{\mathscr{X}}_{1,N}\|};$$
$$\frac{A_2(\hat{\kappa},\hat{\alpha},1)}{A_2(\hat{\kappa},\hat{\alpha},0)} = 1 - \|\bar{\mathscr{X}}_{2,N}\|, \quad \frac{A_2(\hat{\kappa},\hat{\alpha},1)}{A_2(\hat{\kappa},\hat{\alpha},0)} = \frac{1}{4N}\sum_{i=1}^N \left\| x_i - \frac{\bar{\mathscr{X}}_{2,N}}{\|\bar{\mathscr{X}}_{2,N}\|}\right\|^4$$

for the samples $\mathscr{X}_{1,N} \sim \mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$ and $\mathscr{X}_{2,N} \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, respectively.

Remark 4.5.1. *If the parameter α is known, the problem of moment estimation can be reduced to the solution of one equation. Namely,*

$$\frac{A_1(\hat{\kappa},\alpha,1) - A_1(-\hat{\kappa},\alpha,1)}{A_1(\hat{\kappa},\alpha,0) + A_1(-\hat{\kappa},\alpha,0)} = \|\bar{\mathscr{X}}_{1,N}\|, \textit{ and } \quad \frac{A_2(\hat{\kappa},\alpha,1)}{A_2(\hat{\kappa},\alpha,0)} = 1 - \|\bar{\mathscr{X}}_{2,N}\|.$$

In the case of a symmetrically distributed random vector $\mathbf{X} \in \mathbb{S}^{d-1}, \mathbb{E}\mathbf{X} = 0$ and the value of $\mathbb{E}\mathbf{X}/\|\mathbb{E}\mathbf{X}\|$ is not defined. Recall the tangent-normal decomposition (4.2) of a random vector $\mathbf{X} \in \mathbb{S}^{d-1}$, that is $\mathbf{X} = \boldsymbol{\mu}\xi + \sqrt{1-\xi^2}\mathbf{Y}$, where $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ is a mean direction parameter, ξ is a random variable on $[-1,1]$ independent of a uniformly distributed random vector $Y \in \mathbb{S}^{d-2}$ such that $\boldsymbol{\mu} \perp \mathbf{Y}$.

To find relations which determine the parameter $\boldsymbol{\mu}$ and distribution ξ we consider an orientation tensor $T(\mathbf{X})$ given by

$$T(\mathbf{X}) = \mathbf{X}\mathbf{X}^T = \xi^2\boldsymbol{\mu}\boldsymbol{\mu}^T + \xi\sqrt{1-\xi^2}\left(\boldsymbol{\mu}\mathbf{Y}^T + \mathbf{Y}\boldsymbol{\mu}^T\right) + (1-\xi^2)\mathbf{Y}\mathbf{Y}^T. \tag{4.39}$$

Therefore, the mean orientation tensor is

$$\mathbb{E}T(\mathbf{X}) = \mathbb{E}[\mathbf{X}\mathbf{X}^T] = \boldsymbol{\mu}\boldsymbol{\mu}^T\mathbb{E}\xi^2 + (1-\mathbb{E}\xi^2)\mathbb{E}\mathbf{Y}\mathbf{Y}^T. \tag{4.40}$$

Theorem 4.5.1. *Let a random vector* $\mathbf{X}$ *has a representation as above, i.e.,* $\mathbf{X} = \boldsymbol{\mu}\xi + \sqrt{1-\xi^2}\mathbf{Y}$. *Then*

$$\boldsymbol{\mu}\boldsymbol{\mu}^T = \sqrt{\frac{d-1}{d\mathbb{E}[\mathbf{X}(\mathbb{E}T(\mathbf{X}))\mathbf{X}^T]-1}}\left(\mathbb{E}T(\mathbf{X}) - \frac{1}{d}I_d\right) + \frac{1}{d}I_d, \tag{4.41}$$

$$\mathbb{E}\xi^2 = \frac{1}{d} + \sqrt{\frac{d-1}{d}}\sqrt{\mathbb{E}[\mathbf{X}(\mathbb{E}T(\mathbf{X}))\mathbf{X}^T] - \frac{1}{d}}, \tag{4.42}$$

$$\mathbb{E}\xi^4 = \mathbb{E}\left[\mathbf{X}^T\mathbb{E}T(\mathbf{X})\mathbf{X} - \frac{1-\mathbb{E}\xi^2}{d-1}\right]^2 \frac{d-1}{d\mathbb{E}[\mathbf{X}(\mathbb{E}T(\mathbf{X}))\mathbf{X}^T]-1}. \tag{4.43}$$

where I_d *is* $d \times d$ *identity matrix.*

Proof. Let $U_\mu \in SO(d)$, such that $\boldsymbol{\mu} = U_\mu \mathbf{e}_x$, where $\mathbf{e}_x = (1,0,\ldots,0)^T$. Denote by $\tilde{\mathbf{Y}} = U_\mu^{-1}\mathbf{Y}$. The vector $\tilde{\mathbf{Y}}$ is uniformly distributed on $\mathbb{S}^{d-2}$ with the first coordinate equal 0. Then $\mathbf{X} = U_\mu\left(\xi\mathbf{e}_x + \sqrt{1-\xi^2}\tilde{\mathbf{Y}}\right)$ and

$$\mathbf{E}\mathbf{X}\mathbf{X}^T = U_\mu\left(\mathbf{e}_x\mathbf{e}_x^T\mathbb{E}\xi^2 + (1-\mathbb{E}\xi^2)\mathbb{E}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T\right)U_\mu^{-1}.$$

It follows from the symmetry that $\mathbb{E}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T = \frac{1}{d-1}\begin{pmatrix} 0 & 0 \\ 0 & I_{d-1}\end{pmatrix}$. Therefore,

$$\begin{aligned}\mathbb{E}\mathbf{X}\mathbf{X}^T &= U_\mu\left(\mathbf{e}_x\mathbf{e}_x^T\mathbb{E}\xi^2 + \frac{1-\mathbb{E}\xi^2}{d-1}(I_d - \mathbf{e}_x\mathbf{e}_x^T)\right)U_\mu^T \\ &= \mathbb{E}\xi^2\boldsymbol{\mu}\boldsymbol{\mu}^T + \frac{1-\mathbb{E}\xi^2}{d-1}(I_d - \boldsymbol{\mu}\boldsymbol{\mu}^T).\end{aligned} \tag{4.44}$$

Thus, $\boldsymbol{\mu}\boldsymbol{\mu}^T = \frac{d-1}{d\mathbb{E}\xi^2-1}\left(\mathbb{E}\mathbf{X}\mathbf{X}^T - \frac{1-\mathbb{E}\xi^2}{d-1}\right)$. Then consider $\mathbf{X}^T\mathbb{E}T(\mathbf{X})\mathbf{X}$. From (4.44), it follows that

$$\mathbf{X}^T\mathbb{E}T(\mathbf{X})\mathbf{X} = \mathbf{X}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\mathbf{X}\mathbb{E}\xi^2 + \frac{1-\mathbb{E}\xi^2}{d-1}(1-\mathbf{X}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\mathbf{X}) = \xi^2\mathbb{E}\xi^2 + \frac{1-\mathbb{E}\xi^2}{d-1}(1-\xi^2).$$

and $\mathbb{E}[\mathbf{X}^T\mathbb{E}T(\mathbf{X})\mathbf{X}] = (\mathbb{E}\xi^2)^2 + \frac{(1-\mathbb{E}\xi^2)^2}{d-1}$. This yields

$$\mathbb{E}\xi^2 = \frac{1}{d} + \sqrt{\frac{d-1}{d}}\sqrt{\mathbb{E}[\mathbf{X}(\mathbb{E}\mathbf{X}\mathbf{X}^T)\mathbf{X}^T] - \frac{1}{d}}$$

and

$$\boldsymbol{\mu}\boldsymbol{\mu}^T = \sqrt{\frac{d-1}{d\mathbb{E}[\mathbf{X}(\mathbb{E}\mathbf{X}\mathbf{X}^T)\mathbf{X}^T]-1}}\left(\mathbb{E}\mathbf{X}\mathbf{X}^T - \frac{1}{d}I_d\right) + \frac{1}{d}I_d.$$

Finally,

$$\mathbb{E}\left[\mathbf{X}^T\mathbb{E}T(\mathbf{X})\mathbf{X} - \frac{1-\mathbb{E}\xi^2}{d-1}\right]^2 = \left(\frac{d\mathbb{E}\xi^2-1}{d-1}\right)^2\mathbb{E}\xi^4.$$

□

Note that for an axial vector $\mathbf{X}$, the random variable ξ has a symmetric distribution on $[-1,1]$, therefore $\mathbb{E}\xi = 0$. So, if ξ has two-dimensional parametric distribution, one get from Theorem 4.5.1 the parameter estimates for ξ.

Theorem 4.5.1 is applied for the random sample $\mathscr{X}_N = \{\mathbf{x}_1,\dots,\mathbf{x}_N\}$ from $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$. Denote by $\bar{T} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i{}^T$. Then $\mathbb{E}\xi^2 = \mathbb{E}(\mu^T\mathbf{x}_1)^2$ and $\mathbb{E}\xi^4 = \mathbb{E}(\mu^T\mathbf{x}_1)^4$ are given in Proposition 4.3.3.

Corollary 4.5.1.1. *Let* $\mathscr{X}_N \sim \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ *and denote by* $\bar{T} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i{}^T$ *and* $\bar{V} = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i\bar{T}\mathbf{x}_i{}^T$. *Then* (4.41)-(4.43) *hold true and*

$$\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^T = \sqrt{\frac{d-1}{d\hat{V}-1}}\left(\hat{T}-\frac{1}{d}I_d\right)+\frac{1}{d}I_d, \quad \frac{A_1(\hat{\kappa},\hat{\alpha},2)}{A_1(\kappa,\alpha,0)} = \frac{1}{d}+\sqrt{\frac{d-1}{d}}\sqrt{\hat{V}-\frac{1}{d}},$$
$$\frac{A_1(\kappa,\alpha,4)}{A_1(\kappa,\alpha,0)} = \frac{d-1}{(d\hat{V}-1)N}\sum_{i=1}^N\left(\mathbf{x}_i\bar{T}\mathbf{x}_i{}^T-\frac{1}{d}-\frac{\sqrt{d\hat{V}-1}}{d\sqrt{d-1}}\right)^2.$$

Thus, if $\mathbf{X} \sim \mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$, then the estimator for κ is the solution of equation

$$\|\bar{\mathbf{x}}\| = A_{1,d}(\hat{\kappa},\alpha,1).$$

Lemma 4.5.2. *Let* $\alpha > 0, \beta > 0$. *Then for any* $\kappa > 0$ *and* $R \in (0,1)$ *there exists a unique solution of equation* $A_{1,d}(\kappa,\alpha,1) = R$.

Proof. Obviously, $A_{1,d}(\cdot,\alpha,1)$ is a continuous function.

If $\mathbf{X} \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, then $\hat{\kappa}$ is the solution of □

$$\|\bar{\mathbf{x}}\| = \bar{R} = 1-\frac{A_{2,d}(\hat{\kappa},\alpha,1)}{A_{2,d}(\hat{\kappa},\alpha,0)}.$$

Remark 4.5.2. *Let us write down other further characteristics of* $X \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$ *in terms of function* $A_{2,d}$. *Then*

$$c_{2,d}(\kappa,\alpha) = \left(\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)}A_{2,d}(\kappa,\alpha,0)\right)^{-1},$$

$$\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^2 = 2\frac{A_{2,d}(\hat{\kappa},\alpha,1)}{A_{2,d}(\hat{\kappa},\alpha,0)}, \quad \mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^\alpha = 2^\alpha\frac{A_{2,d}(\hat{\kappa},\alpha,\alpha)}{A_{2,d}(\hat{\kappa},\alpha,0)}.$$

4.6 Goodness-of-fit test based on the maximum entropy principle

This section provides the statistical test for verification that a random sample follows a generalized von Mises-Fisher distribution. The methodology for all three introduced distributions is very similar. For simplicity, detailed explanation for the Type II distribution is provided.

4.6.1 Type II

Denote by $\mathrm{GvMF}_{2,d}$ the class of generalized von Mises-Fisher distributions $\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, $\alpha>0$, $\kappa>0$ and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$. Let $\mathscr{X}_N=\{\mathbf{x}_1,\ldots,\mathbf{x}_N,\}$ be a random sample of vectors on a sphere $\mathbb{S}^{d-1}$ and $\mathbf{x}_j \overset{d}{=} \mathbf{X}, j=1,\ldots,N$ with unknown distribution.

Let $\mathbf{Z}\sim \mathrm{GvMF}_2(\alpha,\kappa,\boldsymbol{\mu})$. From (4.10), (4.13) and Theorem 4.4.3 it is known that $H(\mathbf{Z})\geq H(\mathbf{X})$ for all continuous random vectors $\mathbf{X}\in\mathbb{S}^{d-1}$ with

$$\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}=\mathbb{E}\|\mathbf{Z}-\boldsymbol{\mu}\|^{2\alpha}=2^{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}.$$

Using Theorem 4.4.1, we get

$$\inf_{\substack{\alpha,\kappa>0,\\ \boldsymbol{\mu}\in\mathbb{S}^{d-1}}}\left\{-\log c_{2,d}(\kappa,\alpha)+\frac{\kappa}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\,\middle|\,\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}=2^{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\right\}\tag{4.45}$$

does not exceed $H(\mathbf{X})$. Moreover, equality in (4.45) appears if and only if $\mathbf{X}$ belongs to some distribution from the family $\mathrm{GvMF}_{2,d}$. The unobservable value of $\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}$ is substituted by its statistical counterpart $\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_1-\boldsymbol{\mu}\|^{2\alpha}$ and the statistics $S_2(\mathscr{X}_N)$ is defined by

$$\inf_{\substack{\alpha,\kappa>0,\\ \boldsymbol{\mu}\in\mathbb{S}^{d-1}}}\left\{-\log c_{2,d}(\kappa,\alpha)+\frac{\kappa}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\,\middle|\,\frac{\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}}{N2^{\alpha}}=\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\right\}.$$

Consider the value under inf in $S_2(\mathscr{X}_N)$. Under the condition

$$\frac{1}{N2^{\alpha}}\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}=\frac{A_{2,d}(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)},$$

it follows that

$$-\log c_{2,d}(\kappa,\alpha)+\frac{\kappa}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}=-\left(\log c_{2,d}(\kappa,\alpha)-\frac{\kappa}{\alpha2^{\alpha}}\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}\right)=-\frac{l_2(\mathscr{X}_N)}{N}.$$

Then,

$$S_2(\mathscr{X}_N) = -\frac{1}{N} \sup_{\substack{\alpha,\kappa>0,\\ \mu\in\mathbb{S}^{d-1}}} \left\{ l(\mathscr{X}_N) \middle| \frac{\sum_{i=1}^N \|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}}{N2^\alpha} = \frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)} \right\}.$$

Let us consider unconditional maximization of log-likelihood $l_2(\mathscr{X}_N)$. Partial derivative with respect to κ equals

$$\begin{aligned}\frac{\partial l_2(\mathscr{X}_N)}{\partial\kappa} &= \frac{\partial}{\partial\kappa}\left(\log\frac{\Gamma\left(\frac{d-1}{2}\right)}{2\pi^{\frac{d-1}{2}}} - \log A_2(\kappa,\alpha,0) - \frac{\kappa}{\alpha 2^\alpha}\frac{1}{N}\sum_{i=1}^N \|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}\right)\\ &= \frac{1}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)} - \frac{1}{\alpha 2^\alpha N}\sum_{i=1}^N \|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha},\end{aligned}$$

where $\frac{\partial}{\partial\kappa}A_2(\kappa,\alpha,0) = -\frac{1}{\alpha}A_2(\kappa,\alpha,\alpha)$ is used. Thus, the supremum in $S_2(\mathscr{X}_N)$ with respect to κ coincides with the unconditional supremum of $l_2(\mathscr{X}_N)$ and

$$S_2(\mathscr{X}_N) = -\frac{1}{N}\sup_{\substack{\alpha,\kappa>0,\\ \mu\in\mathbb{S}^{d-1}}} l_2(\mathscr{X}_N). \tag{4.46}$$

Let Θ_0 be a compact subset of $\mathbb{R}_+^2$ large enough to contain all values of parameters (α,κ) appearing in practice. Consider the following hypotheses

- $H_{2,0} : \mathscr{X}_N \sim \mathrm{GvMF}_{2,d}$, for some $(\alpha,\kappa)\in\Theta_0$, and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$,
- $H_{2,1} : \mathscr{X}_N \nsim \mathrm{GvMF}_{2,d}$ for all $(\alpha,\kappa)\in\Theta_0$ and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$.

Since Θ_0 is compact, maximum likelihood estimators $\hat\alpha_L$, $\hat\kappa_L$ are consistent. Theorem 4.4.6 is proven that the k-th nearest neighbor estimator $\widehat{H}_{N,k}$ of $H(\mathbf{X})$ is L^2- consistent for any $k\in\mathbb{N}$. Thus, $H_{2,0}$ vs. $H_{2,1}$ are tested with the statistic

$$\hat T^L_{2,k}(\mathscr{X}_N) := -\log c_{2,d}(\hat\kappa_L,\hat\alpha_L) + \frac{\hat\kappa_L}{\hat\alpha_L}\frac{A_2(\hat\kappa_L,\hat\alpha_L,\hat\alpha_L)}{A_2(\hat\kappa_L,\hat\alpha_L,0)} - \widehat{H}_{N,k} \tag{4.47}$$

which tends in probability to 0, as $N\to\infty$. H_0 with level of significance β is rejected if $|\hat T^L_{2,k}(\mathscr{X}_N)| \ge x_\beta$, where x_β is a critical value determined by $\mathbb{P}_{H_0}(|\hat T^L_{2,k}(\mathscr{X}_N)| \ge x_\beta) \le \beta$. The usage of maximum likelihood estimates will yields the higher power of the test.

Remark 4.6.1. *It is easy to see that maximum likelihood estimates of α,κ, estimator $\widehat{H}_{N,k}$ and the statistics $\hat T^L_{2,k}$ are rotational invariant.*

Furthermore, by Slutsky's theorem, the maximum likelihood estimates of α, κ in (4.47) *can be replaced by any consistent estimates $\hat{\alpha}, \hat{\kappa}$. Indeed, if $\mathbf{x}_1 \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$ under hypothesis $H_{2,0}$, then*

$$2^{\hat{\alpha}} \frac{A_2(\hat{\kappa},\hat{\alpha},\hat{\alpha})}{A_2(\hat{\kappa},\hat{\alpha},0)} \xrightarrow[N\to\infty]{P} 2^{\alpha} \frac{A_2(\kappa,\alpha,\alpha)}{A_{2,d}(\kappa,\alpha,0)} = \mathbb{E}\|\mathbf{x}_1 - \boldsymbol{\mu}\|^{\alpha}$$

and

$$\hat{T}_{2,k}(\mathscr{X}_N) := -\log c_{2,d}(\hat{\kappa},\hat{\alpha}) + \frac{\hat{\kappa}}{\hat{\alpha}} \frac{A_2(\hat{\kappa},\hat{\alpha},\hat{\alpha})}{A_2(\hat{\kappa},\hat{\alpha},0)} - \widehat{H}_{N,k} \xrightarrow[\text{under } H_0]{P} H(\mathbf{x}_1) - H(\mathbf{x}_1) = 0$$

as $N \to \infty$.

The critical values x_β can be found by Monte Carlo simulations of test statistics $\hat{T}^L_{2,N}$ or $\hat{T}_{2,N}$.

4.6.2 Type I and axial data

The goodness-of-fit test for the axial generalized von-Mises distribution and the distribution of the Type I are constructed similarly to Type II distributions. Let Θ_0 be a compact subset of $\mathbb{R}^2_+$ large enough to contain all values parameters (α,κ) appearing in practice. Let $j = 1,3$ and consider the following hypotheses

- $H_{j,0} : \mathscr{X} \sim \mathrm{GvMF}_{j,d}$ for some $(\alpha,\kappa) \in \Theta_0$ and $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$,
- $H_{j,1} : \mathscr{X} \not\sim \mathrm{GvMF}_{j,d}$ for all $(\alpha,\kappa) \in \Theta_0$ and $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$.

For testing $H_{1,0}$ vs. $H_{1,1}$ the statistic $\hat{T}_{1,N}$ is used given by

$$\hat{T}_{1,k}(\mathscr{X}_N) := -\log c_{1,d}(\hat{\kappa},\hat{\alpha}) - \frac{\hat{\kappa}}{\hat{\alpha}} \frac{A_1(\hat{\kappa},\hat{\alpha},\hat{\alpha}) - A_1(-\hat{\kappa},\hat{\alpha},\hat{\alpha})}{A_1(\hat{\kappa},\hat{\alpha},0) + A_1(-\hat{\kappa},\hat{\alpha},0)} - \widehat{H}_{N,k}, \tag{4.48}$$

where $\hat{\alpha}, \hat{\kappa}$ are some consistent estimates of α, κ.

For the axial distribution, $H_{3,0}$ vs. $H_{3,1}$ are tested by the test statistic $\hat{T}_{3,N}$ given by

$$\hat{T}_{3,k}(\mathscr{X}_N) := -\log c_{3,d}(\hat{\kappa},\hat{\alpha}) - \frac{\hat{\kappa}}{\hat{\alpha}} \frac{A_1(\hat{\kappa},\hat{\alpha},\hat{\alpha})}{A_1(\hat{\kappa},\hat{\alpha},0)} - \widehat{H}_{N,k}, \tag{4.49}$$

where $\hat{\alpha}, \hat{\kappa}$ are some consistent estimates of α, κ.

Remark 4.6.2. *Note that our goodness-of-fit tests do not detect some particular generalized von Mises-Fisher distribution but tell whether a sample belongs to the parametric family* $\mathrm{GvMF}_{j,d}, j = 1,2,3$.

4.7 Numerical experiments

This section provides the method for simulation of $\mathrm{GvMF}_{j,d}$ distributed random vectors and study the behaviour of the test statistic $\hat{T}_{j,k}$ on simulated samples $j = 1,2,3$.

4.7.1 Simulation

Let $X_j \sim \mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu}), j = 1,2,3$. Due to the tangent-normal decomposition

$$\mathbf{X}_j = (\boldsymbol{\mu}^T\mathbf{X}_j)\boldsymbol{\mu} + \sqrt{1-(\boldsymbol{\mu}^T\mathbf{X}_j)^2}\mathbf{Y_j},$$

where $\mathbf{Y_j}, j = 1,2,3$ are orthogonal to $\boldsymbol{\mu}$ and uniformly distributed on $\mathbb{S}^{d-2}$. Therefore, in order to simulate $\mathbf{X}_j$ it can be easily simulated random vectors $\mathbf{Y_j}$ and independent random variables $\boldsymbol{\mu}^T\mathbf{X}_j$. Let us find the distributions of $\boldsymbol{\mu}^T\mathbf{X}_j$ $j = 1,2,3$. The probability densities f_j of $\boldsymbol{\mu}^T\mathbf{X}_j$ are given in [77, (2.22)] or can be found by applying (4.1).

Lemma 4.7.1. *The random variables $\boldsymbol{\mu}^T\mathbf{X}_1, \boldsymbol{\mu}^T\mathbf{X}_2$, and $\boldsymbol{\mu}^T\mathbf{X}_3$ have probability densities f_1, f_2, and f_3 respectively, given by*

$$f_1(y) = \frac{2\pi^{\frac{d-1}{2}} c_{1,d}(\kappa,\alpha)}{\Gamma\left(\frac{d-1}{2}\right)} \exp\left(\frac{\kappa}{\alpha} y^{<\alpha>}\right)(1-y^2)^{\frac{d-3}{2}}, y \in [-1,1], \tag{4.50}$$

$$f_2(y) = \frac{2\pi^{\frac{d-1}{2}} c_{2,d}(\kappa,\alpha)}{\Gamma\left(\frac{d-1}{2}\right)} \exp\left(-\frac{\kappa}{\alpha}(1-y)^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}}, y \in [-1,1], \tag{4.51}$$

$$f_3(y) = \frac{2\pi^{\frac{d-1}{2}} c_{3,d}(\kappa,\alpha)}{\Gamma\left(\frac{d-1}{2}\right)} \exp\left(\frac{\kappa}{\alpha}|y|^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}}, y \in [-1,1]. \tag{4.52}$$

Proof. Consider $\boldsymbol{\mu}^T\mathbf{X}_1$. It follows from (4.1) that

$$\begin{aligned}\mathbb{P}(\boldsymbol{\mu}^T\mathbf{X}_1 \le u) = c_{1,d}(\kappa,\alpha)\int_{\|\mathbf{x}\|=1} \mathbb{1}\{\boldsymbol{\mu}^T\mathbf{x} \le u\} \exp\left(\frac{\kappa}{\alpha}|\boldsymbol{\mu}^T\mathbf{x}|^{<\alpha>}\right)\sigma(d\mathbf{x}) \\ = c_{1,d}(\kappa,\alpha)\frac{2\pi^{\frac{d-1}{2}}}{\Gamma\left(\frac{d-1}{2}\right)}\int_{-1}^{1} \mathbb{1}\{y \le u\} \exp\left(\frac{\kappa}{\alpha}|y|^{<\alpha>}\right)dy = \int_{-1}^{u} f_1(y)dy.\end{aligned}$$

The cases of $\boldsymbol{\mu}^T\mathbf{X}_2$ and $\boldsymbol{\mu}^T\mathbf{X}_3$ are similar. □

Applying described procedure of simulation several samples of generalized von Mises-Fisher distributions are obtained with 1000 entries on 2-dimensional sphere deferred in the Appendix C the locations of samples entries on a unit sphere and the corresponding histograms and probability densities of random variables $\boldsymbol{\mu}^T\mathbf{X}_i, (i =$

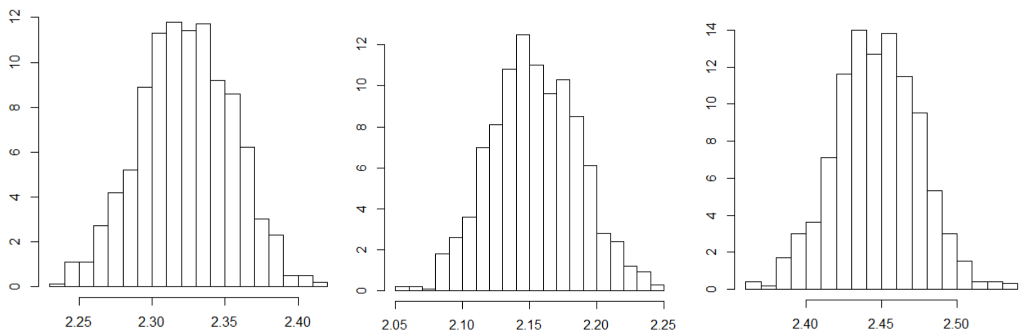
$1, 2, 3$). One can observe that larger values of parameter κ corresponds to more concentrated samples along direction μ.

A short computational study of the parameter estimation methods is provided from sections 4.5.1 and 4.5.2 in the Appendix C.

The method of moments is preferable for Types I and II, if the computing power plays a decisive role. For axial data this method has no such advantages because we need to operate with an orientation tensor. The experiments show that the speed of convergence $\hat{\alpha} \to \alpha$ and $\hat{\kappa} \to \kappa$ depend on the values of α and κ and the speed of convergence of $\hat{\kappa} \to \kappa$ is more quickly than $\hat{\alpha} \to \alpha$. The section can also conclude that the errors of maximum likelihood estimators are generally less than the estimators of moment. Comparing the errors by the distribution type, it is observed that samples of Type II very often carry the smallest error.

4.7.2 Entropy estimation

This section applies k-NNE estimator (4.24) to the simulated set of samples. The estimator $\widehat{H}_{N,k}(\mathscr{X}_N)$ is computed and let us denote the sample variances of entropy estimation as $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha, \kappa))$ for $k = 1, 2, 3, 4, 5$, $\kappa \in \{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7\}$, and $\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3\}$. Figure 4.1 illustrates the distribution of $\widehat{H}_{N,k}(\mathscr{X}_N)$ with $k = 3$ by histograms for samples simulated from distributions $\mathrm{GvMF}_{j,3}(1.5, 2, \cdot)$, $j = 1, 2, 3$.



(a) Type I (b) Type II (c) Axial type

Figure 4.1: Histograms of $\hat{H}_{N,3}(\mathscr{X}_N)$, $\mathscr{X}_N \sim \mathrm{GvMF}_{j,3}(\alpha, \kappa, \cdot)$ with $\alpha = 1.5$ and $\kappa = 2$.

In order to choose the right value of k the sample variances $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha, \kappa))$ is compared for $k = 1, 2, 3, 4, 5$. The minimum and maximum values of $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha, \kappa))$ with respect to α and κ are presented in Table 4.1. These results confirm the conclusion in [13], that is, the asymptotic variance of $\widehat{H}_{N,k}$ decreases rapidly up to $k = 3$.

One can observe that k-th nearest neighbor estimates depend on values α and κ. Although, the sample variances are quite small for all examined values of α and κ and sample size $N = 1000$.

Table 4.1: Sample variance $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha,\kappa))$ for distribution of Type I

Distribution		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\mathrm{GvMF}_{1,3}(\alpha,\kappa,\cdot)$	$\min_{\alpha,\kappa}(\mathrm{sVar})$	0.00214	0.00092	0.00058	0.00042	0.00034
	$\max_{\alpha,\kappa}(\mathrm{sVar})$	0.00388	0.00239	0.00208	0.00185	0.00152
$\mathrm{GvMF}_{2,3}(\alpha,\kappa,\cdot)$	$\min_{\alpha,\kappa}(\mathrm{sVar})$	0.00212	0.00093	0.00059	0.00043	0.00034
	$\max_{\alpha,\kappa}(\mathrm{sVar})$	0.00404	0.00304	0.00275	0.00152	0.00257
$\mathrm{GvMF}_{3,3}(\alpha,\kappa,\cdot)$	$\min_{\alpha,\kappa}(\mathrm{sVar})$	0.00212	0.00093	0.00059	0.00043	0.00034
	$\max_{\alpha,\kappa}(\mathrm{sVar})$	0.00334	0.00207	0.00170	0.00152	0.00144

Thus, we choose $k = 3$ for computations in the next sections.

4.7.3 Test statistic

This section presents our study of the goodness-of-fit tests from Section 4.6 and their test statistics $\hat{T}_{j,k}(\mathscr{X}_N), j = 1,2,3$ from (4.47), (4.48) and (4.49) with $k = 3$ and $N = 1000$. We compute $\hat{T}^L_{j,k}(\mathscr{X}_N)$ and $\hat{T}^M_{j,k}(\mathscr{X}_N)$ separately with maximum likelihood estimates and moment estimates of parameters α, κ, respectively.

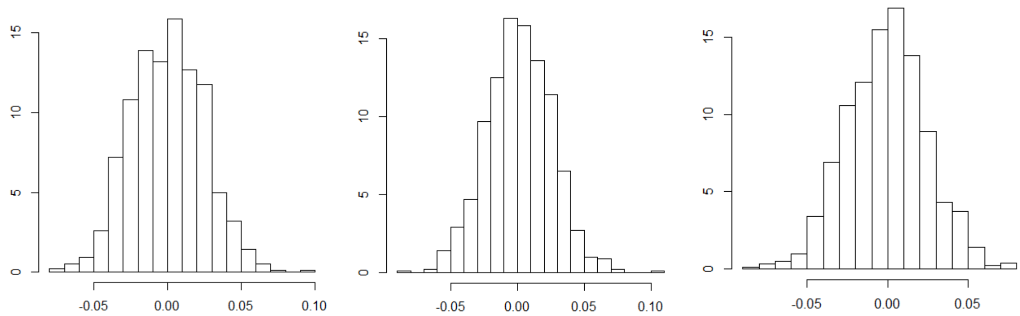
For comparison of different types of estimates, the fact that we look on the sample variances $\mathrm{sVar}(\hat{T}^M_{j,3}(\alpha,\kappa))$ and $\mathrm{sVar}(\hat{T}^L_{j,3}(\alpha,\kappa))$ of $\hat{T}^M_{j,3}(\mathscr{X}_N)$ and $\hat{T}^L_{j,3}(\mathscr{X}_N)$, respectively, for all combinations of parameters $\kappa \in \{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7\}$ and $\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3\}$. The minimum and maximum values of $\mathrm{sVar}(\hat{T}^M_{j,3}(\alpha,\kappa))$ and $\mathrm{sVar}(\hat{T}^L_{j,3}(\alpha,\kappa))$ are presented in Table 4.2. Numbers in this table demonstrate significant benefits of the maximum likelihood estimator over the method of moments for distributions of I and II types. For axial data, one can also prefer $\hat{T}^L_{j,k}$. Additionally, one can observe from Table 4.2 and tables with errors of estimates $\hat{\kappa}^L, \hat{\kappa}^M, \hat{\alpha}^L, \hat{\alpha}^M$ that the statistics $\hat{T}^M_{j,k}$ and $\hat{T}^L_{j,k}$ are much more accurate than estimators of parameters and they have small variances even for small α, κ in contrast to $\hat{\alpha}, \hat{\kappa}$, whose deviations are large.

Section 4.6 chooses the two-sided test with rejection criteria $|\hat{T}^L_{j,k}| > x_\beta$. This choice is confirmed by histograms $\hat{T}^L_{j,3}$ with $\alpha = 1.5$ and $\kappa = 2$, see Figure 4.2.

It is seen that the statistics $\hat{T}^L_{j,3}$ have approximately symmetric distribution with mode at 0. Therefore, the rejection region is put as $(-\infty, -x_\beta) \cup [x_\beta, +\infty)$, where critical values x_β are obtained as a sample quantiles $\mathbb{P}_{H_0}(|\hat{T}_{i,3}| > x_{\beta,j}) \leq \beta, j = 1,2,3$. The corresponding values of $x_{\beta,j}$ with significance level $\beta = 0.05$ are presented in Table 4.3 for $\hat{T}^L_{1,3}$, in Table 4.4 for $\hat{T}^L_{2,3}$, and in Table 4.5 for $\hat{T}^L_{3,3}$.

Table 4.2: Sample variances $\mathrm{sVar}(\hat{T}_{j,3}(\alpha,\kappa))$

	The method of moments			Maximum likelihood method		
Distribution	$\mathrm{GvMF}_{1,3}$	$\mathrm{GvMF}_{2,3}$	$\mathrm{GvMF}_{3,3}$	$\mathrm{GvMF}_{1,3}$	$\mathrm{GvMF}_{2,3}$	$\mathrm{GvMF}_{3,3}$
$\min_{\alpha,\kappa}(\mathrm{sVar})$	0.000563	0.000538	0.000573	0.000558	0.000535	0.000544
$\max_{\alpha,\kappa}(\mathrm{sVar})$	0.001014	0.001558	0.000734	0.000665	0.000688	0.000667



(a) $\mathrm{GvMF}_{1,3}$ (b) $\mathrm{GvMF}_{2,3}$ (c) $\mathrm{GvMF}_{3,3}$

Figure 4.2: Histograms of $\hat{T}^L_{j,3}(\mathscr{X}_N), \mathscr{X}_N \sim \mathrm{GvMF}_{j,3}(\alpha,\kappa,\cdot)$ with $\alpha = 1.5$ and $\kappa = 2$.

Table 4.3: Critical values $x_{\beta,1}$ for test statistic $\hat{T}^L_{1,3}$ and $\beta = 0.05$ with respect to α (rows) and κ (columns), multiplied by 10^2.

	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**	**7**
0.5	4.884	4.626	5.128	5.281	5.069	4.960	4.726	4.999	5.132	5.360	5.301
1	4.727	4.792	4.879	4.736	4.757	4.843	4.920	5.217	4.951	5.278	5.091
1.5	4.935	4.914	4.657	4.731	4.745	4.879	4.849	4.995	5.152	4.990	5.009
2	4.916	4.873	4.839	4.982	4.801	4.987	4.752	4.934	4.829	5.076	5.010
2.5	4.916	4.921	5.276	4.932	4.854	4.852	4.711	4.700	4.860	4.983	5.373
3	4.687	4.821	4.783	4.626	4.926	4.704	4.656	4.690	4.631	4.706	4.844

Table 4.4: Critical values $x_{\beta,2}$ for test statistic $\hat{T}^L_{2,3}$ and $\beta = 0.05$ with respect to α (rows) and κ (columns), multiplied by 10^2.

	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**	**7**
0.5	4.921	4.795	4.750	4.962	4.711	5.015	4.742	5.182	5.430	5.374	5.388
1	4.718	4.688	4.858	4.937	4.996	4.829	4.870	5.065	5.234	5.287	5.449
1.5	4.698	4.916	4.943	4.714	4.824	4.988	4.964	4.613	4.975	5.267	5.263
2	4.920	4.989	4.852	4.611	4.890	4.869	5.147	5.037	4.941	5.423	5.217
2.5	5.066	4.753	5.034	4.805	4.768	4.930	4.973	5.346	5.283	5.138	5.074
3	4.713	4.417	4.731	4.907	5.001	5.007	4.870	5.023	5.015	5.083	5.116

Table 4.5: Critical values $x_{\beta,3}$ for test statistic $\hat{T}^L_{3,3}$ and $\beta = 0.05$ with respect to α (rows) and κ (columns), multiplied by 10^2.

	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**	**7**
0.5	4.932	5.095	4.843	4.863	5.040	4.995	5.000	5.477	5.810	5.219	5.527
1	4.956	4.793	4.636	4.912	4.896	4.873	4.959	5.251	5.196	5.565	5.917
1.5	4.620	4.873	4.781	4.892	4.839	5.006	4.806	4.824	5.016	5.239	5.370
2	4.727	5.049	4.804	4.567	4.811	4.651	4.842	4.999	4.836	5.162	5.191
2.5	4.925	4.799	4.938	4.831	4.735	4.834	5.024	4.739	4.917	4.899	4.703
3	4.960	4.901	4.904	4.649	5.037	4.843	4.898	4.852	4.828	4.895	4.974

The study of the goodness-of-fit test's power is also provided for the samples from Fisher-Bingham distribution. We put $\boldsymbol{\mu}_1 = (1,0,0)^T$ and $\boldsymbol{\mu}_2 = (0, \sqrt{2}/2, \sqrt{2}/2)^T$ and consider the following series of hypotheses.

For Type I:

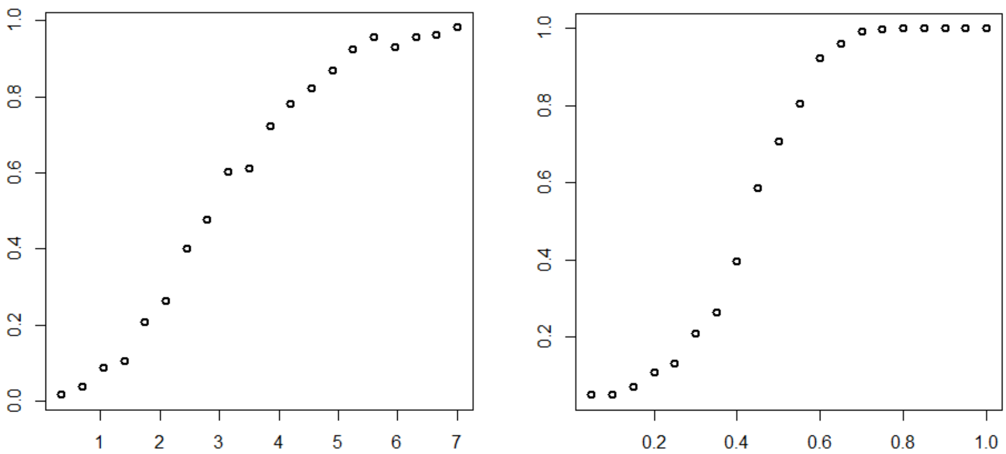
- $H_0^1 : \mathscr{X}_N \sim \mathrm{GvMF}_{1,3}$,
- $H_{1,j}^1$: $\mathscr{X}$ has the Fisher-Bingham distribution with density $\propto \exp(3\boldsymbol{\mu}_1^T\mathbf{x} + 0.35j(\boldsymbol{\mu}_2^T\mathbf{x})^2)$, $\mathbf{x} \in \mathbb{S}^2$, $j = 1,\ldots,20$.

For axial type:

- $H_0^2 : \mathscr{X}_N \sim \mathrm{GvMF}_{3,3}$,
- $H_{1,j}^2 : \mathscr{X}_N$ has the Fisher-Bingham distribution with density $\propto \exp(0.05j(\boldsymbol{\mu}_1^T\mathbf{x}) + 6(\boldsymbol{\mu}_2^T\mathbf{x})^2)$, $\mathbf{x} \in \mathbb{S}^2$, $j = 1,\ldots,20$.

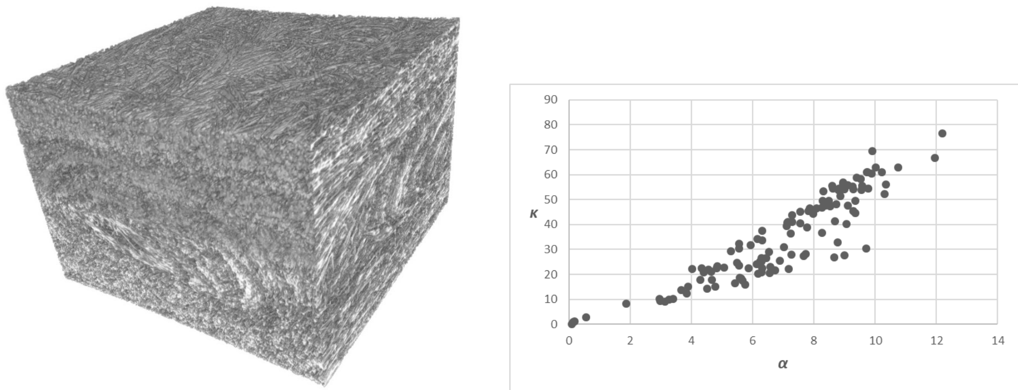
For each $j = 1,\ldots,20$, 500 samples $\mathscr{X}_{j,N}^1$ are simulated under $H_{1,j}^1$ and $\mathscr{X}_{j,N}^2$ under $H_{1,j}^2$ with sample size $N = 1000$. The simulation procedure from the R package is used "Directional" in [118]. In order to simplify replacements, H_0^1 and H_0^2 are rejected if $|\hat{T}^L_{1,3}(\mathscr{X}_{j,N}^1)| > x_\beta^{(1)}$ and $|\hat{T}^L_{2,3}(\mathscr{X}_{j,N}^2)| > x_\beta^{(2)}$ respectively, where critical values $x_\beta^{(1)} = 0.05373$ and $x_\beta^{(2)} = 0.05917$ are taken as maximum of x_β from Tables 4.3 and 4.5 for significance level $\beta = 0.05$.

The ratios of rejections H_0^1 and H_0^2 are presented in Figure 4.3.



(a) Type I, $0.35j$ on the $x-$axis. (b) Axial type, $0.05j$ on the $x-$axis.

Figure 4.3: Powers of goodness-of-fit tests H_0^1 vs $H_{1,j}^1$ (left) and H_0^2 vs $H_{1,j}^2$ (right), $j = 1, \ldots, 20$.



(a) (b)

Figure 4.4: Testing on a glass fibre reinforced composite material.(a) $3D$ image of a glass fibre reinforced (b) Maximum likelihood estimates of α composite material. $970 \times 1469 \times 1217$ and κ for each subsample $\mathscr{X}_1$ voxels, spacing: 4μm.

4.8 Application to a real data set

This section applies the introduced goodness-of-fit test to the data set consists of fiber directions in a glass fibre reinforced composite material. The 3D-images of a fibre composite obtained by micro computed tomography and are provided by the Institute for Composite Materials (IVW) in Kaiserslautern, Germany, see Figure 4.4a (left). The detailed description of the material can be found in [126] and it was the object of studies in [34] and [33], where the regions of anomaly behaviour of the fibres were found. The data set is provided by Prof. Claudia Redenbach (TU Kaiserslautern) and consists of local direction of fibres estimated by the tools of MAVI software [42]. Each data set entry $Y_k, k = ([1,97] \times [1,80] \times [1,64]) \cap \mathbb{N}^3$ is the average of fibre local directions in small observation windows $\tilde{W}$ with $75 \times 75 \times 75$ voxels each. Note that some of such windows can be empty or they might contain not enough material for direction computation. J_W is denoted the collection of indexes k such that Y_k is non-empty. In the cosidered data set $|J_W| = 430741$ and its precise construction is given in [34].

The estimating procedure of directions in MAVI software produces vectors on a unit sphere which are not necessarily symmetrically distributed. However, the fibres are not oriented, therefore an axial distribution of their directions is expected. The symmetrization of original sample is proposed by $\mathscr{X}_N = \{\mathbf{X_k} = \mathbf{Y_k}\xi_{\mathbf{k}}, \mathbf{k} \in J_W\}$, where $\xi_{\mathbf{k}}, \mathbf{k} \in J_W$ are i.i.d random variables with $\mathbb{P}(\xi_{\mathbf{k}} = +1) = \mathbb{P}(\xi_{\mathbf{k}} = -1) = \frac{1}{2}$. The whole material into blocks $W_{\mathbf{l}}$ are separated, each of size $16 \times 15 \times 16$, such that $J_{\mathbf{l}} = J_W \cap [l_1, l_1 + 16) \times [l_2, l_2 + 15) \times [l_3, l_2 + 16)$, $\mathbf{l} = (l_1, l_2, l_3)$ and consider subsamples $\mathscr{X}_{\mathbf{l}} = \{\mathbf{X_k}, \mathbf{k} \in J_{\mathbf{l}}\}$ with simple sizes $2736 \leq |\mathscr{X}_{\mathbf{l}}| \leq 3745$. For each subsample $\mathscr{X}_{\mathbf{l}}$ the introduced goodness-of-fit test is provided for distributions $\mathrm{GvMF}_{3,3}$, i.e., the following is tested

- $H_{0,\mathbf{l}} : \mathscr{X}_{\mathbf{l}} \sim \mathrm{GvMF}_{3,3}$
- $H_{1,\mathbf{l}} : \mathscr{X}_{\mathbf{l}} \not\sim \mathrm{GvMF}_{3,3}$.

At first, the maximum likelihood estimation of parameters α and κ (for the variety of their values $\hat{\alpha}_{\mathbf{l}}$ and $\hat{\kappa}_{\mathbf{l}}$ see Figure 4.4b) is provided. Second, we need to simulate the samples of statistics $\hat{T}_{3,k}(\mathscr{X}_{\mathbf{l}})$ under hypothesis $\mathscr{X}_{\mathbf{l}} \sim \mathrm{GvMF}_{3,3}(\hat{\alpha}_{\mathbf{l}}, \hat{\kappa}_{\mathbf{l}}, \cdot)$ based on samples sizes $|\mathscr{X}_{\mathbf{l}}|$. Unfortunately, our computational resources was limited and we have to group simulations with close values of $\hat{\alpha}_{\mathbf{l}}$ and $\hat{\kappa}_{\mathbf{l}}$. One can observe that the majority of $\hat{\alpha}_{\mathbf{l}}$ belongs to the interval $[3,11]$ and the ratios $\frac{\hat{\kappa}_{\mathbf{l}}}{\hat{\alpha}_{\mathbf{l}}}$ are mostly in $[3,7]$. Therefore, we simulate 800 samples $\mathscr{Y}_N \sim \mathrm{GvMF}_{3,3}(\alpha, \kappa, \cdot)$ each of size $N = 3500$ for all combinations of $\alpha \in \{4,6,8,10\}$ and $\frac{\kappa}{\alpha} \in \{4,6\}$ to obtain the corresponding empirical distributions of $\hat{T}_{3,3}(\mathscr{Y}_N)$.

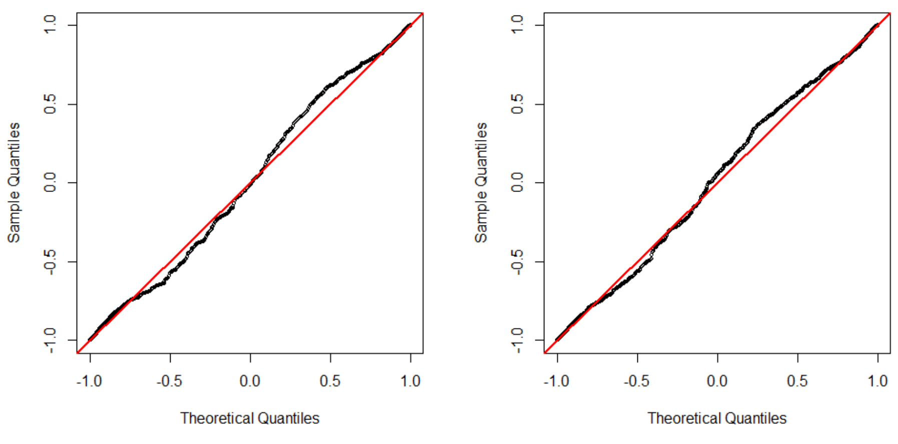
Then, the statistics values of $\hat{T}_{3,3}(\mathscr{X}_{\mathbf{l}})$ are computed for each $\mathbf{l}$ and their p-values. It is obtained that the goodness-of-fit test rejects almost all hypotheses $H_{0,\mathbf{l}}$ with significance level 0.05, and detects 3 regions with directional distributions $\mathrm{GvMF}_{3,3}$ (see Table 4.6 for the samples $\mathscr{X}_{\mathbf{l}}$ with p—values greater than 0.01). In order to illustrate how tight the fitted distributions are, two blocks $W_{\mathbf{l}}$ are presented with $\mathbf{l} = (49,61,1)$

and $\mathbf{l} = (49, 61, 1)$, the QQ-plots for samples $\{\hat{\mu}_{\mathbf{l}}^T \mathbf{X}_{\mathbf{k}}\}$ and distribution f_3 defined in (4.52) with parameters $(\hat{\alpha}_{\mathbf{l}}, \hat{\kappa}_{\mathbf{l}}, \cdot)$, see Figure 4.5.

Table 4.6: Results of goodness-of-fit tests $H_{0,1}$ vs. $H_{1,1}$ for fiber directions in glass fibre reinforced composite material.

l_1	l_2	l_3	$\lvert\mathscr{X}_{\mathbf{l}}\rvert$	$\hat{\alpha}$	$\hat{\kappa}$	$\hat{H}_{N,3}$	$\hat{T}_{3,3}(\mathscr{X}_{\mathbf{l}})$	p−value
49	46	1	3434	8.80	53.90	0.4369	0.02344	0.0775
49	61	1	3222	8.53	47.62	0.7132	0.01976	0.1234
49	16	17	3474	8.84	53.63	0.4690	0.02987	0.0263
49	61	17	3364	10.22	60.91	0.4628	0.02946	0.0263
1	46	65	3319	7.25	36.45	1.0455	0.02057	0.1275

For each type of distributions $\mathrm{GvMF}_{1,3}, \mathrm{GvMF}_{2,3}, \mathrm{GvMF}_{3,3}$ 1000 samples are simulated with $N = 1000$ entries each for several values of $\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3\}$ and $\kappa \in \{0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, \}$. For each sample, the maximum likelihood estimates $\hat{\mu}_L, \hat{\alpha}_L, \hat{\kappa}_L$ and moment estimates $\hat{\mu}_M, \hat{\alpha}_M, \hat{\kappa}_M$ are computed. The sample mean square errors of $\hat{\alpha}_L$ and $\hat{\alpha}_M$ is presented in Tables C.1 (type I), C.3 (type II), and C.5 (axial type). The mean square errors of $\hat{\kappa}_L$ and $\hat{\kappa}_M$ can be found in Tables C.2 (Type I), C.4 (Type II), and C.6 (axial data). We group error values of $\hat{\kappa}_L$, $\hat{\kappa}_M$ and $\hat{\alpha}_L$, $\hat{\alpha}_M$ in order to decide which method is more appropriate for parameter estimation. If computing power is a factor, the method of moments is appropriate for Types I and II. This method has no advantages for axial data because it requires the usage of an orientation tensor.



(a) $\mathbf{l} = (49, 61, 1), \hat{\alpha}_{\mathbf{l}} = 8.53, \hat{\kappa}_{\mathbf{l}} = 47.62$

(b) $\mathbf{l} = (1, 46, 65), \hat{\alpha}_{\mathbf{l}} = 7.25, \hat{\kappa}_{\mathbf{l}} = 36.45$

Figure 4.5: QQ plots for samples $\hat{\mu}_{\mathbf{l}}^T \mathbf{X}_{\mathbf{l}}$ and distributions with density f_3 (4.52).

— Chapter 5 —

Conclusions

The main results from each chapter of the thesis will be presented here. This chapter also discusses some possible future works.

5.1 Research summary

Chapter 1 contains the review of the Shannon and Rényi entropies and the statistical methods of their estimation. It includes the properties of Shannon and Rényi entropy for discrete and continuous distributions.

The maximum Shannon entropy principle and the k-th nearest neighbour distances method of Shannon entropy are used in Chapter 2 to offer a non-parametric goodness-of-fit test for a class of multivariate generalized Gaussian distributions. Some basic explanations and notations which are associated with the multivariate generalized Gaussian distribution are defined in terms of entropy-based test. For an arbitrary fixed $k \geq 1$, the proof of L^2 consistency is presented for the k-th nearest neighbour distance estimator of the Shannon entropy. Based on the maximum entropy principle, a non-parametric goodness-of-fit test is constructed for a class of implemented generalized multivariate Gaussian distributions. In addition to, N random points from the generalized Gaussian distribution are generated to figure out the function of empirical probability, cumulative and log-density of generalized Gaussian distribution for different values of s parameter. The simulation procedure from the Python package is used. The asymptotic behaviour of test statistic is provided for different values of m-dimension, s-parameter and the k-th nearest neighbour distances as N tends to infinity. For a fixed (N, k) and (m, s), we generate a sample of size N from the generalized Gaussian distribution is generated and the empirical value of the test statistic is recorded, by repeating this $M = 250$ times. This yields a sample realisation from the distribution of test statistic, from which we estimate its mean and variance. Moreover, under the null hypothesis, data from the generalized Gaussian distribution are generated and the behaviour of the test statistic is examined where the value is approaching to zero. Under the alternative hypothesis, data from the multivariate Student-t distribution are generated and the behaviour of the test statistic is investigated where the value is approaching a constant. It is also defined that how p-values behave as the sample size increases. The simulation part suggests that the null hypothesis cannot be rejected for samples of size $N = 200$ or more. To each of these 200 samples from the distribution of the test statistic, then we apply the Shapiro-Wilk test is applied for normality [111] and the p-value returned by the test is recorded. One can visually observe that the distributions of the generalized Gaussian and multivariate Student-t distribution are hardly distinguishable. Therefore, the goodness-of-fit test for detecting the generalized Gaussian distribution is applied.

Chapter 3 introduces a class of the Rényi entropy based on an independent identically distributed sample that is drawn from an unknown distribution f in $\mathbb{R}^m$. Then, a non-parametric test of goodness-of-fit is presented for a classes of multivariate Student and Pearson type II (or Barenblatt) distributions based on the maximum Rényi entropy principle and consistency of the k-th nearest neighbour distance estimator for arbitrary fixed $k \geq 1$. The k-th nearest neighbours estimator of Rényi entropy is also used to prove L^2- accuracy. The asymptotic behaviour of the test statistics on data obtained from the multivariate Student and Pearson type II distributions are proposed in Chapter 3. The tests are supported by Monte-Carlo simulation.

In Chapter 4, the new classes of unimodal rotational invariant directional distributions that generalize the von Mises-Fisher distributions are introduced. This chapter proposes three different types of distributions, one of which is for axial data. Formulas and a brief computational analysis of parameter estimators are provided using the method of moments and the method of maximum likelihood for each new type, moreover some basic facts regarding the von Mises-Fisher distribution are described. The aim of Chapter 4 is to establish a goodness-of-fit test to decide if sample entries follow one of the introduced the generalized von Mises-Fisher distributions based on the concept of the maximum entropy. On simulated samples, we analyse the behaviour of the test statistics is analysed, critical values are found, and the power of the test is computed. In a glass fibre reinforced composite material, we use the goodness-of-fit test to identify samples that follow the axial generalized von Mises-Fisher distribution. Moreover, Chapter 4 is dedicated to the Shannon entropy of generalized von Mises-Fisher distributions and the maximum entropy principle. Then, the statistical estimation of entropy is discussed and the L^2 convergence of the k-th nearest neighbor estimator is proven for random variables on compact manifolds.

5.2 Future research directions

For further investigation and relevant research directions, there are some interesting questions which can be considered for future research study. It would be beneficial to use the k-th nearest neighbour method for the estimator Tsallis entropy and use the test goodness-of-fit for other exponentially family. Some of them are written in below.

Problem 1. To investigate the nearest neighbour estimates for the Tsallis entropy in [119]

$$H^T = H^T(X) = H_q^T(f) = \frac{1}{q-1}\left(1 - \mathbb{E}f^{q-1}(X)\right) = \frac{1}{q-1}\left(1 - \int_{\mathbb{R}^m} f^q(x)\,dx\right)$$

and variance of entropy for a random vector X with pdf f is defined as

$$Var\{-\log f(X)\} = E[\log f(X)]^2 - H(X)^2,$$

where the Shannon entropy

$$H(X) = E\{-\log f(X)\}.$$

Problem 2. To investigate entropy based tests for generalized Gaussian, Student and Pearson type II distribution using a weighted average of k-th nearest neighbour estimates or efficient k-th nearest neighbour estimates proposed by Berrett, Samworth and Yuan [13].

Problem 3. To construct the goodness-of-fit test for other classes of multivariate distributions such as multivariate Gamma, Exponential and normal inverse Gaussian.

Problem 4. To investigate the nearest neighbour estimates of (non-symmetric) statistical distances between two distributions with densities f and g, such as *Kullback-Leibler divergence*

$$K(f,g) = \int_{\mathbb{R}^m} f(x)\log\frac{f(x)}{g(x)}\,dx,$$

Bregman divergence

$$D_q(f,g) = \int_{\mathbb{R}^m}\left[g^q(x) + \frac{1}{q-1}f^q(x) - \frac{q}{q-1}f(x)g^{q-1}(x)\right]dx, \quad q \neq 1,$$

based on two samples.

Problem 5. To generalize the nearest neighbour method for dependent observations.

Appendices

— Appendix A —

Lower bound on Shannon entropy

Below, some essentials about lower bounds of Shannon entropy are presented. Firstly, it is shown that there exist densities such that $-\infty = H(f) < \infty$. An example of Gnedenko and Kolmogorov is modified [47, p.223]. For other examples, see [8].

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 $s = 1$*

$$\mathbb{E}X = \mathbb{E}|X| = \int_0^1 \left[\log^2 \frac{e}{x}\right]^{-1} dx = 1 - E_1(1) \simeq 0.40365..., \tag{A.2}$$

where

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

is the generalized exponential integral. Thus by Theorem 2.3.1 with $m = 1$ and $\alpha = 1$,

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

From the other hand,

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

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

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 [17]

$$\frac{1}{m}H(f) \geq \log \|f\|_\infty^{-1/m}. \tag{A.3}$$

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 pth moment $\|X\|_p = \{\mathbb{E}\|X\|^p\}^{1/p} < \infty$, $p \geq 1$, one has for a log-concave density f, see [88],

$$H(f) \geq \log \frac{2\|X - \mathbb{E}[X]\|_p}{[\Gamma(1+p)]^{1/p}}. \tag{A.4}$$

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

$$H(f) \geq \log \frac{2\|X\|_p}{[\Gamma(p+1)]^{1/p}}, \quad p > -1. \tag{A.5}$$

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

$$H(f) \geq \frac{m}{2} \log \frac{(\det \Sigma_x)^{1/m}}{c_3(m)}, \tag{A.6}$$

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

$$c_3(m) = \frac{e^2 m^2}{4\sqrt{2}(m+2)}. \tag{A.7}$$

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

$$c_3(m) = e^2/2. \tag{A.8}$$

The constant (A.8) is better than the constant (A.7) for $m \geq 5$.

The Shannon entropy, $H = -\sum_K p_K \log p_K$, can be easily be infinite, see [7]. Perhaps the simplest example is to consider the sum

$$Q(s) = \sum_{K=[e]}^{\infty} \frac{1}{K(\log K)^{1+s}},$$

which converge s for $s > 0$, diverges for $s \leq 0$ and $[e]$ is the integer part of e.

The corresponding probabilities are

$$p_K = \frac{1}{Q(s)K(\log K)^{1+s}}, \quad K \geq [e].$$

Then, the Shannon entropy

$$\begin{aligned} H(X) &= \log Q(s) + \frac{1}{Q(s)} \sum_K \frac{1}{K(\log K)^s} + \frac{1+s}{Q(s)} \sum_K \frac{\log(\log K)}{K(\log K)^{1+s}} \\ &= \log Q(s) + \frac{1}{Q(s)} \sum_K \frac{1}{K(\log K)^s} - \frac{dQ(s)/ds}{Q(s)}. \end{aligned}$$

The first and third terms converge for $s > 0$, but the second term converges only for $s > 1$. As a result, this probability distribution has infinite Shannon entropy over the entire range $s \in (0, 1]$.

Apart from the entire range $s \in (0, 1]$ above, there are many examples along similar line. For example, one could consider the sums,

$$Q_1(s) = \sum_{K=[e^e]}^{\infty} \frac{1}{K \log K (\log\log K)^{1+s}};$$

or

$$Q_2(s) = \sum_{K=[e^{e^e}]}^{\infty} \frac{1}{K \log K (\log\log K)(\log\log\log K)^{1+s}}.$$

Alternative conditions for consistency of nearest neighbour estimates of Shannon entropy

An alternative sets of conditions for consistency of the nearest neighbour estimates of Shannon entropy are presented after Bulinski and Dimitrov [18, 19]. These conditions are also proved for generalized Gaussian distribution.

For $k \geq 1$, let $f(x)$, $x \in \mathbb{R}^m$, be a density which satisfies the following conditions:

Condition A.0.1. *For some $\varepsilon > 0$ and $R > 0$,*

$$\int_{\mathbb{R}^m} [\sup_{r \in [0,R]} I_f(x,r)]^{\varepsilon} f(x) dx < \infty \tag{A.9}$$

$$\int_{\mathbb{R}^m} [\inf_{r\in[0,R]} I_f(x,r)]^{-\varepsilon} f(x)dx < \infty, \tag{A.10}$$

where

$$I_f(x,r) = \frac{\int_{\|x-y\|\leq r} f(y)dy}{[r^m V_m]}, \quad V_m = \frac{\pi^{m/2}}{\Gamma\left(\frac{m}{2}+1\right)}.$$

Condition A.0.2. *For some* $p > 1$

$$\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} |\log\|x-y\||^p f(x)f(y)dxdy < \infty. \tag{A.11}$$

From [18, 19], one can obtain the following results.

Theorem A.0.3. *Assume that Condition* $A.0.1$ *hold.*

1. *If Condition* $A.0.2$ *holds for some* $p > 1$*, then for any fixed* $k \in \{1,\ldots,N-1\}$

$$\mathbb{E}\widehat{H}_{N,k} \to H, \quad as \quad N \to \infty; \tag{A.12}$$

2. *If Condition* $A.0.2$ *holds for some* $p > 2$*, then for any fixed* $k \in \{1,\ldots,N-1\}$

$$\mathbb{E}\left[\widehat{H}_{N,k} - H\right]^2 \to 0 \quad as \quad N \to \infty. \tag{A.13}$$

Remark A.0.1. *Condition* (A_1)*,* $\int_{\mathbb{R}^m}[\sup_{r\in[0,R]} I_f(x,r)]^{\varepsilon} f(x)dx < \infty$*, in Theorem* $A.0.3$ *can be replaced by the following condition: for some* $M > 0$

$$f(x) \leq M < \infty, \quad x \in \mathbb{R}^m; \tag{A.14}$$

while the condition (A_2)*,* $\int_{\mathbb{R}^m}[\inf_{r\in[0,R]} I_f(x,r)]^{-\varepsilon} f(x)dx < \infty$*, in Theorem* $A.0.3$ *can be replaced by the following condition: for a fixed* $R > 0$*, there exists a constant* $c > 0$*, such that*

$$\inf_{r\in[0,R]} I_f(x,r) \geq cf(x), \quad x \in \mathbb{R}^m, \tag{A.15}$$

and for some $\varepsilon > 0$

$$\int_{\mathbb{R}^m} f^{1-\varepsilon}(x)dx < \infty. \tag{A.16}$$

Remark A.0.2. *Condition A.0.1 was introduced in [71] and [49], while condition* (A.15) *is considered in [30] for $k = 1$ (together with other conditions), see also in [39] and [38].*

Remark A.0.3. *For $k = 1$, it was proven by Bulinski and Dimitrov [18] that* (A.12), (A.13) *hold for multidimensional Gaussian distribution while [30] show that* (A.12) *and* (A.13) *hold for $GG(m,s)$ and Student distributions.*

Proposition A.0.1. *The density function $f(x) = f_0 \exp\{-\frac{\|x\|^s}{s}\}$, $x \in \mathbb{R}^m$ of $GG(m,s)$ distribution belongs to the class $\mathscr{K}$ for $k \geq 1$, $s > 0$, $m \geq 1$.*

Proof. It is easy to see that (A.14) for $p > 2$ (and hence (A.9)) and $A.0.2$ hold for $GG(m,s)$ distribution. To prove (A.10) we note that for $x, y \in \mathbb{R}^m$

$$\|y\|^2 = \|x\|^2 + \|y - x\|^2 + 2\langle x, y - x\rangle.$$

Using an elementary inequality:

$$(a + b)^\alpha \leq a^\alpha + b^\alpha, \quad a, b \geq 0, \quad \alpha \in [0, 1],$$

for $0 < s \leq 2$ it follows that:

$$\begin{aligned}\|y\|^s = \left(\|y\|^2\right)^{s/2} &\leq \left(\|x\|^2 + \|y - x\|^2 + 2|\langle x, y - x\rangle|\right)^{s/2} \\ &\leq \|x\|^s + \|y - x\|^s + 2^{s/2}|\langle x, y - x\rangle|^{s/2}.\end{aligned}$$

Therefore,

$$\begin{aligned}f(y) = c_0 \exp\{-\frac{1}{s}\|x\|^s\} &\geq c_0 \exp\left\{-\frac{1}{s}\left(\|x\|^s + \|y - x\|^s + 2^{s/2}|\langle x, y - x\rangle|^{s/2}\right)\right\} \\ &\geq f(x) \exp\left\{-\frac{1}{s}\|y - x\|^2 - \frac{1}{s}2^{s/2}|\langle x, y - x\rangle|^{s/2}\right\}.\end{aligned} \tag{A.17}$$

Let us fix an arbitrary $x \in \mathbb{R}^m$, $R > 0$, and for any $r \in (0, R)$ by (A.17) we obtain

$$\begin{aligned}\int_{\|x-y\| \leq r} f(y) dy &\geq f(x) \int_{\|x-y\| \leq r} \exp\left\{-\frac{1}{s}\|y - x\|^s\right\} \exp\left\{-\frac{1}{s}2^s|\langle x, y - x\rangle|^{s/2}\right\} dx \\ &\geq f(x) \int_{\|z\| \leq r} \exp\left\{-\frac{1}{s}\|z\|^s\right\} \exp\left\{-\frac{1}{s}2^s|\langle x, z\rangle|^{s/2}\right\} dz.\end{aligned} \tag{A.18}$$

Simple inequality: $e^u \geq 1 + u, u \in \mathbb{R}$, leads to the formula

$$\begin{aligned}\int_{\|z\| \leq r} e^{-\frac{1}{s}2^s|\langle x, z\rangle|^{s/2}} dz &\geq \int_{\|z\| \leq r} \left(1 - \frac{1}{s}2^s|\langle x, z\rangle|^{s/2}\right) dz \\ &= V_m r^m - \frac{1}{s}2^s \int_{\|z\| \leq r} |\langle x, z\rangle|^{s/2} dz.\end{aligned}$$

Using Cauchy-Schwartz inequality,

$$|\langle x,z\rangle|^2 \leq \|x\|\|z\|,$$

it follows that

$$\begin{aligned}
\int_{\|z\|\leq r} \exp\left\{-\frac{1}{s}2^s|\langle x,z\rangle|^{s/2}\right\} dz &\geq V_m r^m - \frac{1}{s}2^s\|x\|^s \int_{\|x\|\leq r} \|z\|^s dz \\
&= V_m r^m - \frac{1}{s}2^s\|x\|^s \int_{\|z\|\leq r} \|z\|^s dz \\
&= V_m r^m - \frac{1}{s}2^s\|x\|^s \frac{2\pi^{m/2}}{\Gamma(\frac{m}{2})} \frac{r^{m+\frac{s}{2}}}{\left(m+\frac{s}{2}\right)} \\
&= V_m r^m - \frac{2^s\pi^{m/2}}{s\Gamma(\frac{m}{2})\left(m+\frac{s}{2}\right)}\|x\|^s r^{m+\frac{s}{2}}.
\end{aligned} \tag{A.19}$$

From (A.18) and (A.19) we get

$$\int_{\|x-y\|\leq r} f(y)dy \geq f(x)\exp\left\{-\frac{1}{s}r^s\right\}\left[r^m V_m(1-c_* r^{s/2}\|x\|^s)\right], \tag{A.20}$$

where

$$c_8 = \frac{2^s}{s}\frac{m}{2\left(m+\frac{s}{2}\right)} = \frac{2^s}{s}\frac{m}{(2m+s)}$$

and

$$m_f(x,R) = \inf_{r\in[0,R]}\left[\left[\int_{\|x-y\|\leq r} f(y)dy\frac{1}{r^m V_m}\right]\geq f(x)e^{-\frac{1}{s}R^s}\left(1-R^2 c_*\|x\|^{s/2}\right) = F(x).\right.$$

Note that there exists $\varepsilon \in (0,1)$ such that for a fixed $R > 0$

$$\begin{aligned}
\int_{\mathbb{R}^m} \frac{1}{F(x)^\varepsilon} f(x)dx &= \int_{\mathbb{R}^m} f^{1-\varepsilon}(x)\frac{1}{\left(1-R^2c_*\|x\|^{s/2}\right)^\varepsilon}dx \\
&= c_0^{1-\varepsilon}e^{-\frac{1}{s}R^2}\int_{\mathbb{R}^m}\exp\left\{\|x(1-\varepsilon)^{1/s}\|^s\right\}\frac{dx}{\left(1-R^2c_*\|x\|^{s/2}\right)^\varepsilon} < \infty,
\end{aligned}$$

and (A.10) holds for $0 < s \leq 2$. For $s \geq 2$, we use an elementary inequality

$$(x+y)^\alpha \leq 2^{\alpha-1}(x^\alpha + y^\alpha), \quad \alpha \geq 1, \quad x,y \geq 0.$$

We have

$$\begin{aligned}\|y\|^s = (\|y\|^2)^{s/2} &\leq 2^{\frac{s}{2}-1}\left[\|x\|^s + \left(\|y-x\|^2 + 2|\langle x, y-x\rangle|\right)^{s/2}\right] \\ &\leq 2^{\frac{s-2}{2}}\left[\|x\|^s + 2^{\frac{s-2}{2}}\left(\|y-x\|^s + 2^{\frac{s}{2}}|\langle x, y-x\rangle|^{s/2}\right)\right] \\ &= 2^{\frac{s-2}{2}}\|x\|^s + 2^{s-2}\|y-x\|^s + 2^{\frac{3s-4}{2}}|\langle x, y-x\rangle|^{\frac{3s-4}{2}}.\end{aligned}$$

Therefore,

$$f(y) \geq c_0 \exp\left\{-\frac{1}{s}\|x\|^s 2^{\frac{s-2}{2}}\right\} \exp\left\{-\frac{1}{s}2^{s-2}\|y-x\|^s\right\} \exp\left\{-\frac{1}{s}2^{\frac{3s-4}{2}}|\langle x, y-x\rangle|^{\frac{3s-4}{2}}\right\}$$

and

$$\int_{\|x-y\|\leq r} f(y)dy \geq c_0 \exp\left\{-\frac{1}{s}\|x\|^s 2^{\frac{s-2}{2}}\right\} \exp\left\{-\frac{1}{s}2^{s-2}r^s\right\} r^m V_m\left(1 - c^*\|x\|^{3s-4}R^2\right),$$

where

$$c^* = \frac{m}{s}2^{\frac{2s-6}{2}}\frac{1}{(m+3s-4)}.$$

Thus,

$$m_f(x,R) \geq c_0 e^{\left(-\frac{1}{s}\|x\|^s 2^{\frac{s-2}{2}}\right)} e^{\left(-\frac{1}{s}2^{s-2}R^s\right)}\left(1 - c^*\|x\|^{3s-4}R^2\right) = F_1(x).$$

Note that there exists $\varepsilon \in (0,1)$ such that for a fixed $R > 0$:

$$\int_{\mathbb{R}^m} \frac{f(x)dx}{F_1(x)^\varepsilon} = c_0^{1-\varepsilon} e^{\left(-\frac{1}{s}2^{s-2}R^2\right)} \int_{\mathbb{R}^m} e^{\left(-\frac{1}{s}2^{\frac{s-2}{2}}\left|\|x(1-\varepsilon)^{1/s}\|\right|^s\right)} \frac{dx}{(1-R^2c^*\|x\|^{3s-4})} < \infty$$

and (A.10) holds for $s \geq 2$.

$\square$

— Appendix B —

Numerical result for Rényi entropy

This section contains some numerical experiments for Rényi entropy and shows empirical distribution of multivariate Student, Pearson type II distribution for different values of *degrees of freedom dof*, q and m dimensions.

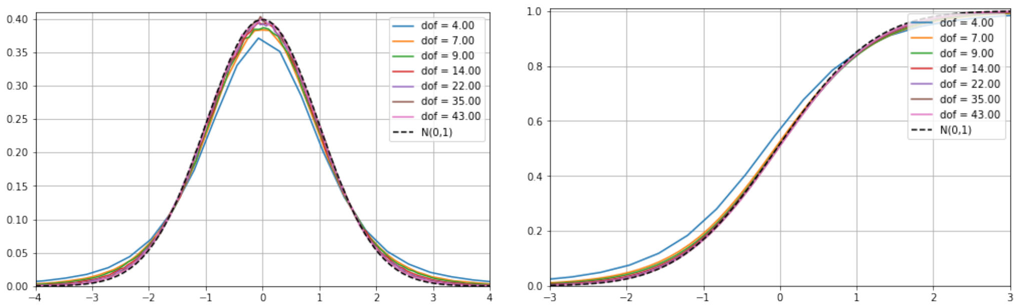


Figure B.1: Empirical distribution of $ST(m, \nu)$ for $m = 1$ and different values of dof.

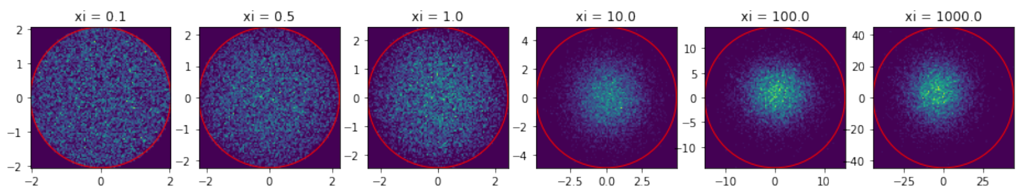
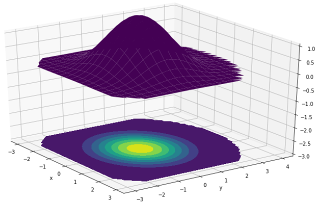
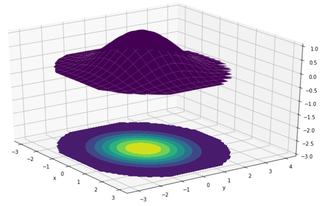


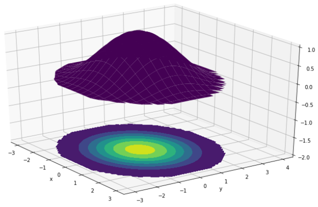
Figure B.2: Heatmaps for multivariate Pearson type II distribution as q increase the plots becomes uniform distribution, $m = 2$.



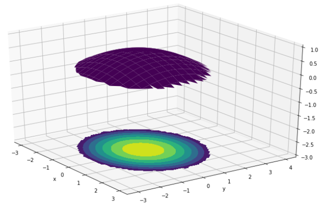
(a) q=1.2



(b) q=1.3



(c) q=1.4



(d) q=1.9

Figure B.4: Visulation of multivariate Pearson Type II distribution for $m = 3$ and different values of q.

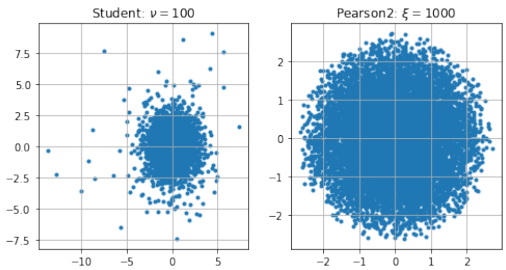


Figure B.3: Scatter plots for bivariate Student and Pearson type II distributions.

— Appendix C —

Numerical experiments for von Mises-Fisher distributions

In the Appendix C, tables of mean square errors of estimates of κ and α and plots with relations of the generalized von Mises-Fisher distributions are presented. The procedure of simulation obtaining several samples of generalized von Mises-Fisher distributions with 1000 entries on 2- dimensional sphere is also shown.

Table C.1: Mean square error of estimator of $\hat{\alpha}_M$ (top raw) and $\hat{\alpha}_L$ (bottom raw) for $\mathrm{GvMF}_{1,3}$ distribution, for different values of α (rows) and κ (columns) with aspect of Type I

	0.5	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**
0.5	0.01131	0.00309	0.00229	0.00334	0.00720	0.00234	0.00067	0.00044	0.00046
	0.01103	0.00294	0.00209	0.00290	0.00600	0.01028	0.01568	0.02000	0.02222
1	0.27260	0.03760	0.01778	0.01126	0.00766	0.00633	0.00686	0.00950	0.02210
	0.21929	0.03207	0.01518	0.00838	0.00569	0.00480	0.00508	0.00644	0.00850
1.5	2.89363	0.44954	0.12728	0.07166	0.04616	0.03330	0.02049	0.01535	0.01465
	2.75421	0.20300	0.07216	0.03613	0.02355	0.01725	0.01081	0.00904	0.00793
2	8.34069	2.90354	0.89834	0.44383	0.27728	0.16970	0.08593	0.05445	0.04011
	10.16847	1.30990	0.30051	0.13914	0.08601	0.05726	0.03554	0.02187	0.01650
2.5	8.68665	6.79656	3.96727	2.18844	1.23084	0.62982	0.35996	0.18528	0.13112
	15.44292	4.46642	1.21787	0.42840	0.23206	0.16395	0.08478	0.05083	0.03633
3	10.1863	8.74483	6.31636	6.33024	3.48461	2.84763	1.35829	0.62283	0.50274
	18.7129	8.33385	3.35711	1.24416	0.60680	0.36062	0.20265	0.11569	0.08699

Table C.2: Mean square error of estimator of $\hat{\kappa}_M$ (top raw) and $\hat{\kappa}_L$ (bottom raw) for GvMF$_{1,3}$ distribution, for different values of α (rows) and κ (columns) with regard to Type I

	0.5	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**
0.5	0.01899 0.01717	0.01699 0.01598	0.01898 0.01727	0.02602 0.02396	0.03563 0.03306	0.03419 0.04987	0.03749 0.07889	0.04394 0.10651	0.06683 0.13889
1	0.22319 0.11194	0.07076 0.05895	0.06534 0.05552	0.06965 0.05215	0.06449 0.05029	0.06737 0.05277	0.08770 0.06947	0.11201 0.08617	0.19431 0.10606
1.5	2.11049 1.19008	0.64790 0.19097	0.26637 0.14496	0.25165 0.12205	0.23415 0.11508	0.22145 0.11355	0.20812 0.10874	0.20797 0.12384	0.23012 0.12786
2	3.68444 3.19240	4.15468 0.92650	1.72159 0.39620	1.18887 0.29374	1.05376 0.27322	0.82951 0.24016	0.63399 0.24388	0.54231 0.20779	0.49912 0.19860
2.5	2.17772 4.02490	6.54718 2.52365	7.27056 1.19600	5.81158 0.61967	4.41608 0.50278	2.47149 0.50154	2.26686 0.41390	1.43704 0.35365	1.34411 0.35024
3	1.87215 4.29139	5.16488 3.63187	7.11001 2.47833	13.30717 1.40486	10.1227 0.99006	11.05648 0.80562	7.43739 0.74920	4.25858 0.62553	4.56458 0.61910

Let us illustrate the generalized von Mises-Fisher distribution on 2-dimensional sphere by several samples with 1000 entries. For all samples, mean direction $\boldsymbol{\mu} = (0, \sqrt{2/2}, \sqrt{2/2})$ is fixed. For different values of α and κ, locations of samples entries on a unit sphere are presented for Type I, see Figure C.2 (with value $\alpha = 0.5$) and Figure C.4 (with $\alpha = 1.5$); for Type 2, Figures C.6 and C.8 with $\alpha = 0.5$ and $\alpha = 1.5$, respectively, and the samples of axial data are presented in Figure C.10 (with $\alpha = 0.5$) and C.12 (with $\alpha = 1.5$). The corresponding histograms and probability densities of random variables $\boldsymbol{\mu}^T\mathbf{X}_i, i = 1, 2, 3$ can be found in Figures C.1 (with value $\alpha = 0.5$) and C.3 (with value $\alpha = 1.5$) for Type I, in Figures C.5 (with value $\alpha = 0.5$) and C.7 (with value $\alpha = 1.5$) for Type II, and in Figures C.9 (with value $\alpha = 0.5$) and C.11 (with value $\alpha = 1.5$) for axial data.

Table C.3: Mean square error of estimator of $\hat{\alpha}_M$ (top raw) and $\hat{\alpha}_L$ (bottom raw) for GvMF$_{2,3}$ distribution, for different values of α (rows) and κ (columns) with aspect of Type II

	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**
0.5	0.04439 0.03719	0.00892 0.00571	0.00384 0.00244	0.00240 0.00131	0.00200 0.00111	0.00176 0.00082	0.00174 0.00069	0.00208 0.00073	0.00112 0.00072	
1	0.11476 0.15498	0.02662 0.02641	0.01238 0.01200	0.00779 0.00721	0.00620 0.00555	0.00508 0.00469	0.00415 0.00369	0.00424 0.00357	0.00384 0.00341	
1.5	0.20015 0.30670	0.05429 0.05952	0.02785 0.02937	0.01742 0.01765	0.01432 0.01473	0.01291 0.01303	0.00981 0.01006	0.00956 0.01004	0.00925 0.00962	
2	0.30415 0.40585	0.09795 0.09406	0.05169 0.05020	0.03786 0.03618	0.02962 0.02640	0.02579 0.02466	0.02137 0.01993	0.02306 0.02191	0.02314 0.02198	
2.5	0.56165 0.51124	0.12324 0.12344	0.08668 0.06832	0.06643 0.05349	0.05643 0.04539	0.05414 0.04304	0.05298 0.04235	0.04757 0.03748	0.05055 0.04087	
3	0.71927 0.52430	0.18736 0.16199	0.14068 0.09598	0.11660 0.08051	0.11100 0.07280	0.11534 0.07200	0.09450 0.06493	0.10280 0.07062	0.10102 0.06307	

The mean square errors of $\hat{\kappa}_L$ and $\hat{\kappa}_M$ can be found in Tables C.4 (Type II), and C.6 (axial data). We group error values of $\hat{\kappa}_L$, $\hat{\kappa}_M$ and $\hat{\alpha}_L$, $\hat{\alpha}_M$ in order to decide which method is more appropriate for parameter estimation. The method of moments is preferable fot Types I and II, if the computing power plays a decisive role. For axial data, this method has no such advantage because it needs to operate with an orientation tensor.

Table C.4: Mean square error of estimator of $\hat{\kappa}_M$ (top raw) and $\hat{\kappa}_L$ (bottom raw) for GvMF$_{2,3}$ distribution, for different values of α (rows) and κ (columns) with aspect of Type II

	0.5	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**
0.5	0.00593 0.00585	0.00788 0.00619	0.01080 0.00815	0.01751 0.01132	0.03210 0.02080	0.05088 0.02932	0.14597 0.06803	0.39689 0.16595	0.26720 0.29599
1	0.00460 0.00504	0.00611 0.00602	0.00912 0.00885	0.01508 0.01432	0.02487 0.02260	0.03871 0.03695	0.08568 0.07654	0.18878 0.16276	0.31815 0.28533
1.5	0.00382 0.00433	0.00496 0.00497	0.00712 0.00715	0.01159 0.01170	0.02027 0.02050	0.03192 0.03221	0.07540 0.07713	0.16707 0.17453	0.28675 0.29832
2	0.00507 0.00594	0.00512 0.00515	0.00658 0.00663	0.01078 0.01090	0.01733 0.01729	0.02968 0.02992	0.07776 0.07590	0.17039 0.16294	0.33426 0.32125
2.5	0.00725 0.00797	0.00705 0.00761	0.00885 0.00847	0.01145 0.01155	0.01878 0.01850	0.03171 0.03021	0.08975 0.07865	0.19040 0.16314	0.36456 0.29699
3	0.01072 0.01004	0.01052 0.01033	0.01256 0.01179	0.01447 0.01420	0.02044 0.02058	0.03274 0.03070	0.09333 0.07711	0.24011 0.17748	0.45512 0.30782

It can be seen that the critical values of $x_{\beta,j}, j = 1, 2, 3$, are obtained as a sample quantiles for test statistic $\hat{T}^L_{j,3}$ with respect to significance level $\beta = 0.025$ are presented clearly in Table C.7 for $\hat{T}^L_{1,3}$, in Table C.8 for $\hat{T}^L_{2,3}$ and in Table C.9 for $\hat{T}^L_{3,3}$.

Table C.5: Mean square error of estimator of $\hat{\alpha}_M$ (top raw) and $\hat{\alpha}_L$ (bottom raw) for GvMF$_{3,3}$ distribution, for different values of α (rows) and κ (columns) with aspect of Type III

	0.5	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**
0.5	0.18987 0.16463	0.05705 0.04162	0.02550 0.02490	0.01188 0.01907	0.00814 0.01631	0.00512 0.01637	0.00236 0.02073	0.00128 0.02427	0.00045 0.02521
1	0.72857 1.99071	0.23713 0.21501	0.11809 0.08130	0.07490 0.04956	0.05300 0.03309	0.03753 0.02432	0.02852 0.02189	0.01925 0.01923	0.01203 0.02272
1.5	1.41746 7.55889	0.69305 1.11221	0.28007 0.28031	0.16580 0.14598	0.09577 0.08200	0.07204 0.06144	0.04872 0.04110	0.03733 0.03296	0.03224 0.02763
2	2.16193 11.8352	1.34087 4.29336	0.66370 1.22665	0.37039 0.40504	0.20955 0.20974	0.15980 0.16143	0.09065 0.08807	0.05605 0.05484	0.04741 0.04622
2.5	3.14653 17.0535	2.42570 7.75701	1.51950 3.11424	0.79157 1.16675	0.47397 0.52106	0.33253 0.34421	0.17312 0.18537	0.11837 0.12071	0.08002 0.08197
3	4.87133 17.9716	4.04307 10.7576	2.64681 5.75811	1.72811 2.78415	1.07167 1.28941	0.70276 0.74024	0.33226 0.35797	0.20734 0.22610	0.15167 0.15511

Table C.6: Mean square error of estimator of $\hat{\kappa}_M$ (top raw) and $\hat{\kappa}_L$ (bottom raw) for $\mathrm{GvMF}_{3,3}$ distribution, for different values of α (rows) and κ (columns) with aspect of Type III

	0.5	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**
0.5	0.06602 0.06010	0.05435 0.04156	0.05186 0.04885	0.04292 0.05101	0.04582 0.05939	0.04249 0.06571	0.05097 0.10154	0.05318 0.12560	0.06447 0.15621
1	0.14994 0.64104	0.12997 0.14258	0.13578 0.10734	0.14216 0.10834	0.14670 0.10368	0.12929 0.09266	0.14336 0.11498	0.14140 0.12950	0.14276 0.17055
1.5	0.22813 1.96361	0.22610 0.54431	0.21028 0.25810	0.22102 0.22148	0.19741 0.18410	0.19829 0.17962	0.21212 0.18814	0.23031 0.21156	0.24706 0.21623
2	0.15945 2.32459	0.29881 1.70401	0.35199 0.86157	0.36035 0.45843	0.31712 0.35382	0.33244 0.35891	0.31035 0.30941	0.30316 0.30274	0.32061 0.31870
2.5	0.15577 3.15115	0.37208 2.58508	0.55335 1.76622	0.55512 0.99471	0.51438 0.65380	0.51368 0.58423	0.48383 0.53832	0.49249 0.50869	0.42680 0.43904
3	0.14980 3.34039	0.44026 3.09945	0.67499 2.79798	0.86090 1.99287	0.84880 1.32447	0.85450 1.07203	0.69410 0.79730	0.67099 0.75204	0.66948 0.69418

Table C.7: Critical values $x_{\beta,1}$ for test statistic $\hat{T}^L_{1,3}$ and $\beta = 0.025$ with respect to α (rows) and κ (columns), multiplied by 10^2.

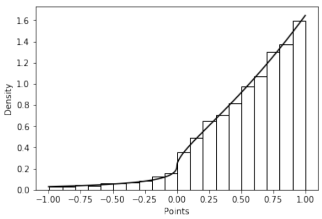
	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**	**7**
0.5	6.251	6.212	6.364	6.716	6.803	6.477	6.484	6.909	7.037	6.917	7.116
1	5.957	6.08	6.301	5.905	6.8	6.32	6.046	6.837	6.872	7.375	7.101
1.5	5.682	6.147	5.826	6.09	6.17	6.107	6.095	5.883	6.714	6.675	6.811
2	6.305	6.22	6.149	6.289	6.17	6.097	5.794	6.62	6.476	6.729	6.556
2.5	6.207	6.123	6.098	6.132	6.354	5.847	6.471	6.55	6.142	6.391	6.258
3	6.33	6.127	5.962	6.026	6.055	6.002	6.115	5.672	6.326	6.077	5.802

Table C.8: Critical values $x_{\beta,2}$ for test statistic $\hat{T}^L_{2,3}$ and $\beta = 0.025$ with respect to α (rows) and κ (columns), multiplied by 10^2.

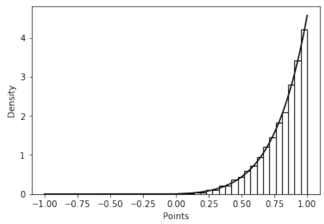
	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**	**7**
0.5	5.634	5.792	6.048	6.222	6.357	6.899	6.86	7.125	6.695	6.935	7.17
1	5.226	6.544	6.30	6.111	6.51	6.667	6.771	6.731	6.582	7.029	6.551
1.5	5.556	6.086	5.46	6.566	6.446	6.135	6.374	6.571	6.949	6.767	6.542
2	6.183	5.945	6.438	5.902	6.267	7.026	6.736	6.558	6.605	6.425	6.618
2.5	6.421	5.93	5.949	6.476	6.872	7.033	6.112	6.696	6.513	6.285	6.824
3	6.004	6.597	5.854	6.876	6.167	6.407	6.90	6.493	6.363	6.71	6.627

Table C.9: Critical values $x_{\beta,3}$ for test statistic $\hat{T}^{L}_{3,3}$ and $\beta = 0.025$ with respect to α (rows) and κ (columns), multiplied by 10^2.

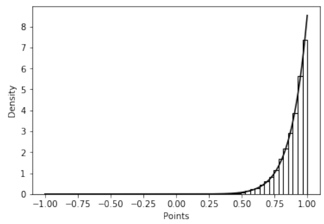
	0.1	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**4**	**5**	**6**	**7**
0.5	5.988	6.192	6.573	6.32	7.043	6.726	6.925	7.338	7.384	7.524	8524
1	6.286	6.178	6.052	6.626	5.632	6.731	6.549	7.223	6.533	7.491	8.313
1.5	5.773	5.846	5.879	6.16	6.845	6.214	6.157	6.437	7.153	6.832	7.11
2	5.901	5.959	5.917	6.057	6.414	6.46	6.344	6.732	6.477	6.329	6.478
2.5	6.409	5.725	6.305	6.237	6.037	6.068	6.167	6.145	6.115	6.769	6.817
3	6.188	5.981	5.974	6.158	6.261	5.941	5.961	5.762	5.883	6.348	6.652



(a) $\kappa = 1$

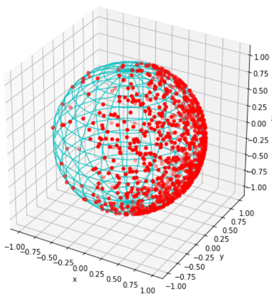


(b) $\kappa = 4$

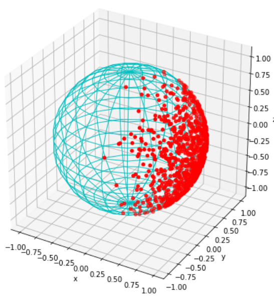


(c) $\kappa = 8$

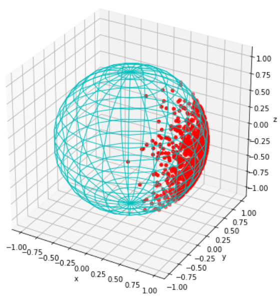
Figure C.1: Density f_1 from (4.50) for $\alpha = 0.5$ and different values of κ.



(a) $\kappa = 1$

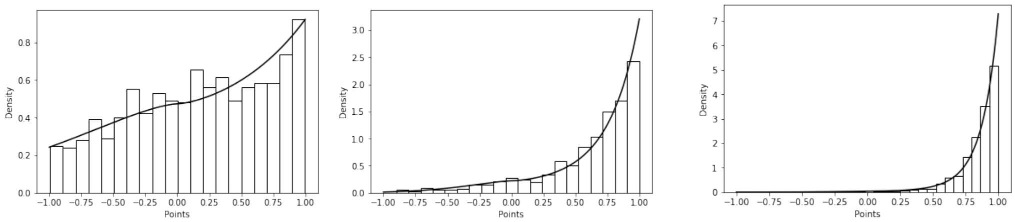


(b) $\kappa = 4$



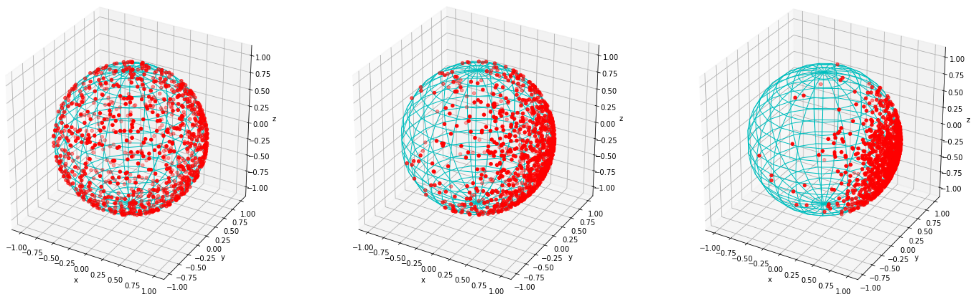
(c) $\kappa = 8$

Figure C.2: Realisations of $X \sim \mathrm{GvMF}_{1,3}(\alpha, \kappa, \mu)$ with $\alpha = 0.5$ and different values of κ.



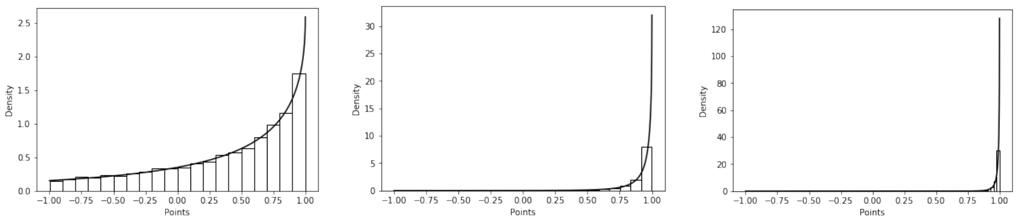
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.3: Density f_1 from (4.50) for $\alpha = 1.5$ and different values of κ.



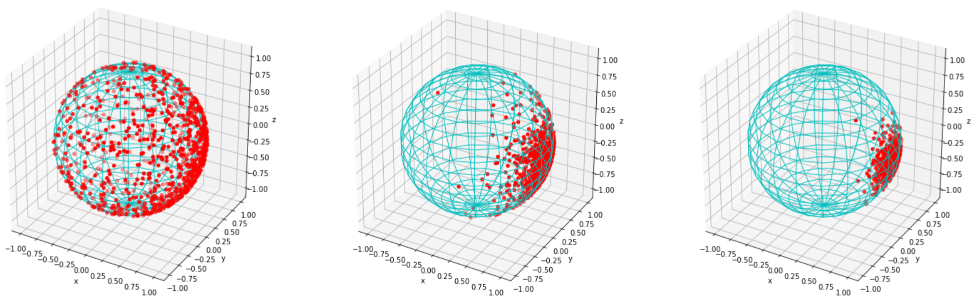
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.4: Realisations of $X \sim \mathrm{GvMF}_{1,3}(\alpha, \kappa, \mu)$ with $\alpha = 1.5$ and different values of κ.



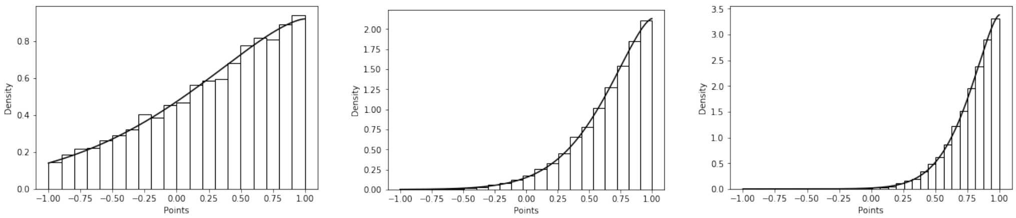
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.5: Density f_2 from (4.51) for $\alpha = 0.5$ and different values of κ.



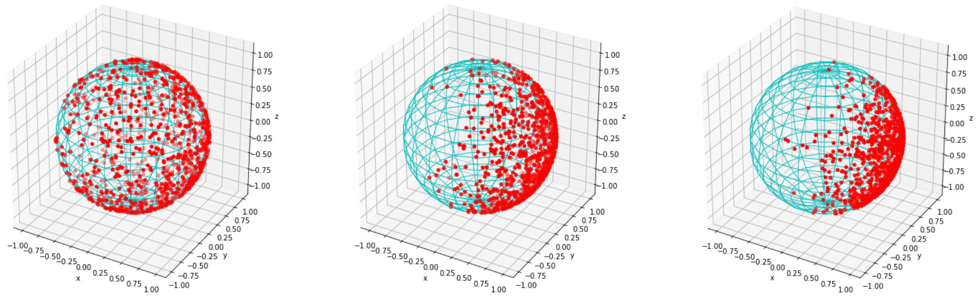
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.6: Realisations of $X \sim \mathrm{GvMF}_{2,3}(\alpha, \kappa, \mu)$ with $\alpha = 0.5$ and different values of κ.



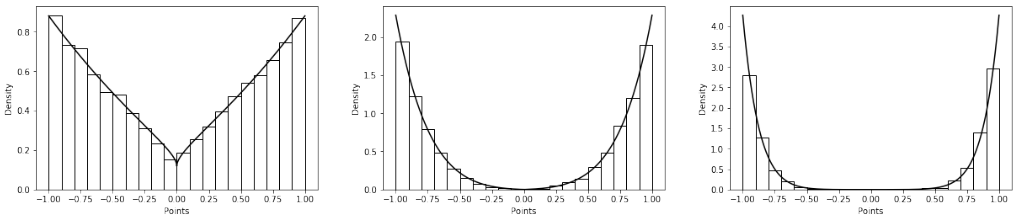
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.7: Density f_2 from (4.51) for $\alpha = 1.5$ and different values of κ.



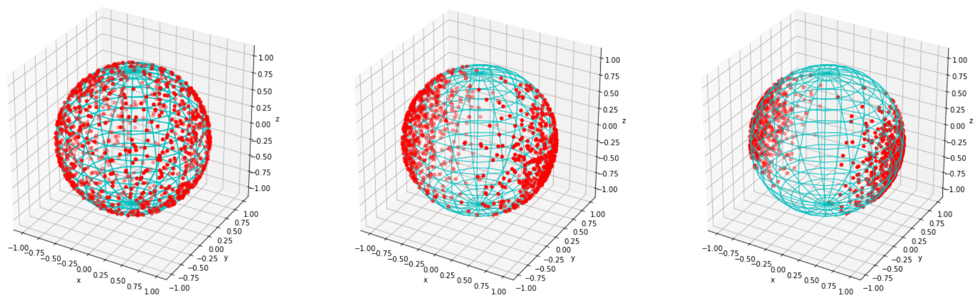
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.8: Realisations of $X \sim \mathrm{GvMF}_{2,3}(\alpha, \kappa, \mu)$ with $\alpha = 1.5$ and different values of κ.



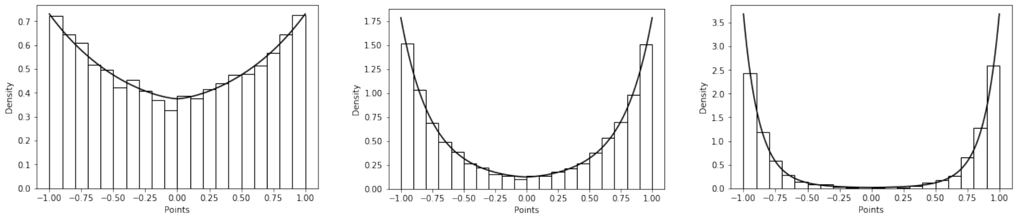
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.9: Density f_3 from (4.52) for $\alpha = 0.5$ and different values of κ.



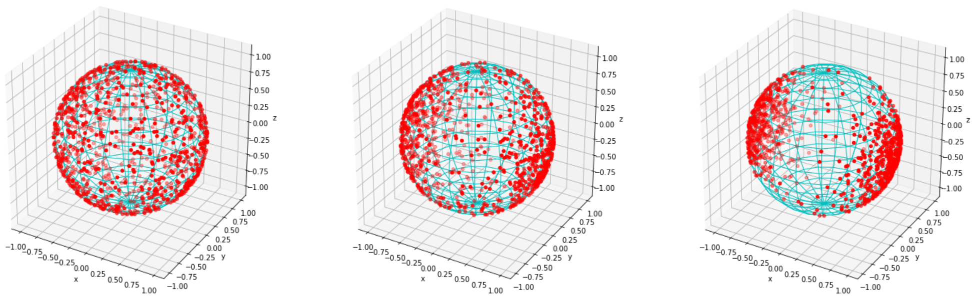
(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.10: Realisations of $X \sim \mathrm{GvMF}_{3,3}(\alpha, \kappa, \mu)$ with $\alpha = 0.5$ and different values of κ.



(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.11: Density f_3 from (4.52) for $\alpha = 1.5$ and different values of κ.



(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

Figure C.12: Realisations of $X \sim \mathrm{GvMF}_{3,3}(\alpha, \kappa, \mu)$ with $\alpha = 1.5$ and different values of κ.

— Appendix D —

Historical perspective of Entropy

The concept of entropy has an exciting long history. The notion of entropy way first defined by German physicist Rudolph Clausius in 1865 as a general law of physics, which now is known as the second law of thermodynamic. Researchers in that time were forced to understand and predict macroscopic observable phenomenon since the atomic (microscopic) nature of matter was not discovered yet. Clausius showed that there exist an exact differential function $dS = dQ/T$ where dS is the differential change in entropy resulting from an infinitesimal flow of heat dQ at temperature T, [89]. Clausius also showed that in any equilibrium state, the observable thermodynamic state parameters (such as mole number, pressure and volume) attain values for which the entropy is maximum subject to any physical constraints, see [6].

In 1876, the physicists, Joshua Willard Gibss formulated the principles of statistical mechanics based on the Clausius ideas. Later in the 19th and early 20th centuries, James Clerk Maxwell and Ludwing Boltzmann developed these concepts further. Boltzmann was able to describe his entropy function by the equation

$$H = k_B \log W \quad \text{or} \quad H = k_B \log P \tag{D.1}$$

where k_B denotes the Boltzmann's constant and $P = 1/W$ and W is the number of microstates that are consistent with the given equilibrium microstate. The constant k_B was not written by Boltzmann himself because of the Planck's reading of [96]. Boltzmann considered the entropy as a measure of statistical disorder or "mixedupness". Then, American mathematical physicist J. Willard Gibbs refined this concept into a common formula for statistical mechanical entropy which is considered as one of the foundations of statistical mechanics theory. The general formula for the entropy of a system that was obtained by Gibbs' systematic development of statistical thermodynamics is given by

$$H = -k_B \sum p_K \log p_K \tag{D.2}$$

where p_K denotes the probability of a system being in various $\{K\}$ microstates. The next significant development in entropy appeared in 1948 from an unexpected area of communication theory. The concept of information entropy was published in a seminal paper "A mathematical Theory of Communication" by an American mathematician Claude Elwood Shannon, known as the father of information of Shannon theory [110]. In that year, Claude Shannon developed a complete theoretical framework for quantifying the performance of communication links. Shannon entropy is a measure of the uncertainty or disorder associated with a random variable. Primarily, Shannon entropy measures the expected value of the information contained in a message. Any communication link contains a source, transmitter, a physical link and a receiver. Such a link will be used to transmit a message of interest. Shannon defined as a following form:

$$H = -K \sum p_K \log p_K \tag{D.3}$$

where K denotes a positive constant, Shannon interpreted H as the average uncertainty. Shannon also proposed that the entropy help us to maximize of the bit transfer rate under a quality constraint. Jaynes (1950) suggested to use the entropy measure for radio interferometric image deconvolution in order to choose from a group of possible solutions that include minimum information or following the definition of Shannon entropy that has maximum entropy. A great amount of work has been achieved in the last 30 years using entropy for the general question of data filtering and deconvolution. Significant developments and applications of theory have been established by Irving S. Reed, David E. Muller and Fumitada Itakura in 1960 and 1966 respectively, [6].

The concept of entropy in other areas science

In molecular biology, using the concept of information entropy in molecular biology allows us to distinguish information- coding regions between random ones in ensembles of genomes and quantify the information content. The application of information theory in molecular biology indicates the association of regulatory molecules accompanied by their binding sites and protein-protein interactions. Besides, information theory demonstrates the recognition of the polymorphism nature of many viral proteins based on drug design by maximizing the information shared between the target and drug ensembles, see [2].

In psychology, entropy has many practical applications in psychological sciences. As indicated in from information theory, the concept of entropy supplies a useful framework for explaining the nature and psychological impact of uncertainty. Hirsh [57] found out that the entropy model of uncertainty enables us to understand the competition between behavioural affordance and competing perceptual. Psychological entropy emerges conversely in associated with the integrity of the existence of an individual in the world. The entropy-based model also demonstrates the critical importance of uncertainty management for productivity, well-being and individual's survival within the physical context and a broader evolutionary. Entropy model of uncertainty plays an essential role in psychological sciences. It presents a critical adaptive challenge for any organism and appears with activity in heightened noradrenaline release and the anterior cingulate cortex. Finally, the entropy-based framework shows how cognitive and behavioural consequences of heightened uncertainty can be defined.

In finance, the concept and relevant principles of entropy have been used in the area of finance for an extended period. Entropy usually can be defined as an essential tool in portfolio selection and asset pricing. The entropy concept in portfolio selection was adjusted initially in [101]. Entropy was first used as a measure of risk in the field of portfolio selection. It was replaced with variance in typical mean-variance models by some scholars. In contrast, entropy was added to the original portfolio models and optimized the new models later by some other scholars. Moreover, entropy can also be used as a measure of risk for applying fuzzy portfolio selection situation. Finally, entropy can be used as a measure of capital increment, portfolio diversification and option pricing in finance [131].

It is well-known that the concept of entropy has various explanations and is measured in a separate system in other areas. **In classical physics**, entropy aims to point out the proportional of the quantity of energy to do physical movements. It appears to be key concept in the Second Law of thermodynamics that describes an isolated system; any activity increases the entropy. In other words, entropy is a significant physical concept originated from this law. It also helps to measure the quantity of order, disorder and chaos. **In quantum mechanics**, the notion of entropy was extended by von Neumann entropy to the quantum system with the density matrix's help. **In a dynamical system**, entropy quantifies the exponential complexity of the average flow of information per unit of time. **In the theory of probability**, the entropy of random variable measures the uncertainty. **In sociology**, entropy is the natural decay of structures, such as organization, law and convention [80].

Bibliography

[1] Abadie, J., Abbott, B., Abbott, R., Abernathy, M. et al. (2011). Directional limits on persistent gravitational waves using LIGO S5 science data. *Physical Review Letters*, *107*(27), 271102.

[2] Adami, C. (2004). Information theory in molecular biology. *Physics of Life Reviews*, *1*(1), 3–22.

[3] Aguilar, M., Cavasonza, L. A., Ambrosi, G., Arruda, L., Attig, N. et al. (2020). The Alpha Magnetic Spectrometer (AMS) on the international space station: Part ii–Results from the first seven years. *Physics Reports*.

[4] Ahmad, I., & Lin, P.-E. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions (corresp.) *IEEE Transactions on Information Theory*, *22*(3), 372–375.

[5] Anchordoqui, L. A., Bergman, D. R., Bertaina, M. E. et al. (2020). Performance and science reach of the Probe of Extreme Multimessenger Astrophysics for ultrahigh-energy particles. *Physical Review D*, *101*(2), 023012.

[6] Ayres, R. U. (1997). *Information, Entropy, and Progress: A New Evolutionary Paradigm*. Springer Science & Business Media.

[7] Baccetti, V., & Visser, M. (2013). Infinite Shannon entropy. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*(04), P04010.

[8] Barron, A. R. (1986). Entropy and the central limit theorem. *Ann. Prob.*, *14*(1), 336–342.

[9] Baryshnikov, Y., Penrose, M. D., & Yukich, J. (2009). Gaussian limits for generalized spacings. *Annals of Applied Probability*, *19*(1), 158–185.

[10] 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*(1), 17–39.

[11] Berger, M., & Gostiaux, B. (1988). Graduate texts in mathematics–differential geometry: Manifolds, curves and surfaces.

[12] Berrett, T. B., & Samworth, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika*, *106*(3), 547–566.

[13] Berrett, T. B., Samworth, R. J., Yuan, M. et al. (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *The Annals of Statistics*, *47*(1), 288–318.

[14] Biau, G., & Devroye, L. (2015). *Lectures on the Nearest Neighbor Method* (Vol. 246). Springer.

[15] Bingham, M., & Mardia, K. (1975). Maximum likelihood characterization of the von Mises distribution. *A Modern Course on Statistical Distributions in Scientific Work* (pp. 387–398). Springer.

[16] Birgé, L., Massart, P. et al. (1995). Estimation of integral functionals of a density. *The Annals of Statistics*, *23*(1), 11–29.

[17] 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*(8), 4940–4954.

[18] Bulinski, A., & Dimitrov, D. (2019). Statistical estimation of the Shannon entropy. *Acta Mathematica Sinica, English Series*, *35*(1), 17–46.

[19] Bulinski, A., & Dimitrov, D. (2021). Statistical estimation of the Kullback–Leibler divergence. *Mathematics*, *9*(5), 544.

[20] Caccianiga, L. (2019). Anisotropies of the highest energy cosmic-ray events recorded by the Pierre Auger Observatory in 15 years of operation. *36th International Cosmic Ray Conference*, *358*, 206.

[21] Cadirci, M. S., Evans, D., Leonenko, N., & Makogin, V. (2020). Entropy-based test for generalized Gaussian distributions. *arXiv preprint arXiv:2010.06284*.

[22] Cadirci, M. S., Evans, D., Leonenko, N., & Seleznjev, O. (2021). Statistical tests based on Rényi entropy estimation. *arXiv preprint arXiv:2106.00453*.

[23] Carpio, J., & Gago, A. (2017). Roadmap for searching cosmic rays correlated with the extraterrestrial neutrinos seen at IceCube. *Physical Review D*, *95*(12), 123009.

[24] Chatterjee, S. (2008). A new method of normal approximation. *Annals of Probability*, *36*(4), 1584–1610.

[25] Chesneau, C., Navarro, F., & Serea, O. S. (2017). A note on the adaptive estimation of the differential entropy by wavelet methods. *Commentationes Mathematicae Universitatis Carolinae*, *58*(1).

[26] Cover, T. M., & Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons.

[27] Cressie, N. (1976). On the logarithms of high-order spacings. *Biometrika*, *63*(2), 343–355.

[28] Cutting, C., Paindaveine, D., Verdebout, T. et al. (2020). On the power of axial tests of uniformity on spheres. *Electronic Journal of Statistics*, *14*(1), 2123–2154.

[29] De Gregorio, A., & Garra, R. (2020). Alternative probabilistic representations of Barenblatt-type solutions. *Modern Stochastics: Theory and Applications* (pp. 1–16).

[30] Delattre, S., & Fournier, N. (2017). On the Kozachenko–Leonenko entropy estimator. *Journal of Statistical Planning and Inference*, *185*, 69–93.

[31] Denton, P. B., & Tamborra, I. (2018). Exploring the properties of choked gamma-ray bursts with Icecubeâs high-energy neutrinos. *The Astrophysical Journal*, *855*(1), 37.

[32] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., & Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 508–539.

[33] Dresvyanskiy, D., Karaseva, T., Makogin, V., Mitrofanov, S., Redenbach, C., & Spodarev, E. (2020). Detecting anomalies in fibre systems using 3-dimensional image data. *Statistics and Computing*, *30*(4), 817–837.

[34] Dresvyanskiy, D., Karaseva, T., Mitrofanov, S., Redenbach, C., Schwaar, S., Makogin, V., & Spodarev, E. (2019). Application of clustering methods to anomaly detection in fibrous media. *IOP Conference Series: Materials Science and Engineering*, *537*(2), 022001.

[35] Duerinckx, M., & Ley, C. (2012). Maximum likelihood characterization of rotationally symmetric distributions on the sphere. *Sankhya A*, *74*(2), 249–262.

[36] Dyn, C. F. (2001). M. abramowitz, ia stegun, eds. handbook of mathematical functions with for-mulas, graphs, and mathematical tables, dover (1972). rk ahuja, tl magnanti, jb orlin, tl magnanti, network flows: Theory, al-gorithms, and applications, prentice hall (1993). r. albert, al barabási,“statistical mechanics of complex networks”. rev. mod. *Phys. Rev. E*, *64*, 036106.

[37] Ebner, B., & Henze, N. (2020). Tests for multivariate normality—a critical review with emphasis on weighted l^2 -statistics. *Test*, 1–48.

[38] Evans, D. (2008). A law of large numbers for nearest neighbour statistics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *464*(2100), 3175–3192.

[39] 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*(2028), 2839–2849.

[40] Fang, K.-T., & Zhang, Y.-T. (1990). *Generalized Multivariate Analysis* (Vol. 19). Science Press Beijing; Springer-Verlag, Berlin.

[41] Frank, T. D. (2005). *Nonlinear Fokker-Planck Equations: Fundamentals and Applications*. Springer Science & Business Media.

[42] Fraunhofer, I. (2005). Mavi–Modular algorithms for volume images. *Fraunhofer ITWM, Kaiserslautern*.

[43] Gao, W., Oh, S., & Viswanath, P. (2018a). Breaking the bandwidth barrier: Geometrical adaptive entropy estimation. *IEEE Transactions on Information Theory*, *64*(5), 3313–3330.

[44] Gao, W., Oh, S., & Viswanath, P. (2018b). Demystifying fixed k-nearest neighbor information estimators. *IEEE Transactions on Information Theory*, *64*(8), 5629–5661.

[45] García-Portugués, E., Paindaveine, D., & Verdebout, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *Journal of the American Statistical Association*, *115*(532), 1873–1887.

[46] Gatto, R., & Jammalamadaka, S. (2007). The generalized von Mises distribution. *Statistical Methodology*, *4*(3), 341–353.

[47] Gnedenko, B., & Kolmogorov, A. (1954). *Limit Distributions of Sum of Independent Random Variables*. Addison-Wesley Educational Publishers Inc.

[48] Godavarti, M., & Hero, A. (2004). Convergence of differential entropies. *IEEE Transactions on Information Theory*, *50*(1), 171–176.

[49] 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*(3), 277–297.

[50] Grigelionis, B. (2013). *Student's t-distribution and related stochastic processes*. Springer.

[51] Hall, P., & Morton, S. C. (1993). On the estimation of entropy. *Annals of the Institute of Statistical Mathematics*, *45*(1), 69–88.

[52] Hallin, M., Paindaveine, D. et al. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *The Annals of Statistics*, *30*(4), 1103–1133.

[53] Havrda, J., & Charvát, F. (1967). Quantification method of classification processes. Concept of structural α-entropy. *Kybernetika*, *3*(1), 30–35.

[54] Hayakawa, T. (1990). On tests for the mean direction of the Langevin distribution. *Annals of the Institute of Statistical Mathematics*, *42*(2), 359–373.

[55] Hayakawa, T., & Puri, M. (1985). Asymptotic expansions of the distributions of some test statistics. *Annals of the Institute of Statistical Mathematics*, *37*(1), 95–108.

[56] Heyde, C., & Leonenko, N. N. (2005). Student processes. *Advances in Applied Probability*, *37*(2), 342–365.

[57] Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304.

[58] Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., & Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, *441*(1), 1–46.

[59] Jammalamadaka, S., & Terdik, G. H. (2019). Harmonic analysis and distribution-free inference for spherical distributions. *Journal of Multivariate Analysis*, *171*, 436–451.

[60] Joe, H. (1989). Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, *84*(405), 157–164.

[61] Johnson, O. (2004). *Information Theory and The Central Limit Theorem*. World Scientific.

[62] Johnson, O., & Vignat, C. (2007). Some results concerning maximum Rényi entropy distributions. *Annales de l'IHP Probabilités et Statistiques*, *43*(3), 339–351.

[63] Jupp, P. E. (2015). Copulae on products of compact Riemannian manifolds. *Journal of Multivariate Analysis*, *140*, 92–98.

[64] Källberg, D., Leonenko, N., & Seleznjev, O. (2014). Statistical estimation of quadratic Rényi entropy for a stationary m-dependent sequence. *Journal of Nonparametric Statistics*, *26*(2), 385–411.

[65] Kandasamy, K., Krishnamurthy, A., Poczos, B., Wasserman, L. A., & Robins, J. M. (2015). Nonparametric von Mises Estimators for Entropies, Divergences and Mutual Informations. *NIPS*, *15*, 397–405.

[66] Kapur, J. N. (1989). *Maximum-Entropy Models in Science and Engineering*. John Wiley & Sons.

[67] Kapur, J. N. (1994). *Measures of Information and Their Applications*. Wiley-Interscience.

[68] Khanin, A., & Mortlock, D. J. (2016). A Bayesian analysis of the 69 highest energy cosmic rays detected by the Pierre Auger Observatory. *Monthly Notices of the Royal Astronomical Society*, *460*(3), 2765–2778.

[69] Kolokoltsov, V. N. (2010). *Nonlinear Markov Processes and Kinetic Equations* (Vol. 182). Cambridge University Press.

[70] Kotz, S., & Nadarajah, S. (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press.

[71] Kozachenko, L. F., & Leonenko, N. N. (1987). A statistical estimate for the entropy of a random vector. *Problems Infor. Transmiss*, *23*(2), 9–16.

[72] Krishnamurthy, A., Kandasamy, K., Poczos, B., & Wasserman, L. (2014). Nonparametric estimation of Rényi divergence and friends. *International Conference on Machine Learning*, 919–927.

[73] Learned-Miller, E. G., & Fisher III, J. W. (2004). ICA using spacings estimates of entropy. *Journal of Machine Learning Research*, *4*.

[74] Leonenko, N., Makogin, V., & Cadirci, M. S. (2020). The entropy based goodness-of-fit tests for generalized von Mises-Fisher distributions and beyond. *arXiv preprint arXiv:2010.10918*.

[75] Leonenko, N., Pronzato, L., Savani, V. et al. (2008). A class of Rényi information estimators for multidimensional densities. *The Annals of Statistics*, *36*(5), 2153–2182.

[76] Leonenko, N. N., & Pronzato, L. (2010). Correction: A class of Rényi information estimators for multidimensional densities. *Annals of Statistics*, *38*(6), 3837–3838.

[77] Ley, C., & Verdebout, T. (2017). *Modern Directional Statistics*. CRC Press.

[78] Leyton, M., Dye, S., & Monroe, J. (2017). Exploring the hidden interior of the Earth with directional neutrino measurements. *Nature Communications*, *8*(1), 1–11.

[79] Li, S., Mnatsakanov, R. M., & Andrew, M. E. (2011). K-nearest neighbor based consistent entropy estimation for hyperspherical distributions. *Entropy*, *13*(3), 650–667.

[80] Liu, Y., Liu, C., & Wang, D. (2011). Understanding atmospheric behaviour in terms of entropy: A review of applications of the second law of thermodynamics to meteorology. *Entropy*, *13*(1), 211–240.

[81] Lord, W. M., Sun, J., & Bollt, E. M. (2018). Geometric k-nearest neighbor estimation of entropy and mutual information. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(3), 033114.

[82] Lund, U., & Rao Jammalamadaka, S. (2000). An entropy-based test for goodness-of-fit of the von Mises distribution. *Journal of Statistical Computation and Simulation*, *67*(4), 319–332.

[83] Lutwak, E., Yang, D., Zhang, G. et al. (2004). Moment-entropy inequalities. *The Annals of Probability*, *32*(1B), 757–774.

[84] Lutwak, E., Yang, D., & Zhang, G. (2007). Moment-entropy inequalities for a random vector. *IEEE Transactions on Information Theory*, *53*(4), 1603–1607.

[85] Mardia, K. (1975). Statistics of directional data. *Journal of the Royal Statistical Society*, *37*, 349–393.

[86] Mardia, K., & Jupp, P. (2009). *Directional Statistics* (Vol. 494). John Wiley & Sons.

[87] Marinucci, D., & Peccati, G. (2011). *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications* (Vol. 389). Cambridge University Press.

[88] Marsiglietti, A., & Kostina, V. (2018). A lower bound on the differential entropy of log-concave random vectors with applications. *Entropy*, *20*(3), 185.

[89] Michalowicz, J. V., Nichols, J. M., & Bucholtz, F. (2013). *Handbook of Differential Entropy*. Crc Press.

[90] Misra, N., Singh, H., & Hnizdo, V. (2010). Nearest neighbor estimates of entropy for multivariate circular distributions. *Entropy*, *12*(5), 1125–1144.

[91] Mouritsen, H., & Mouritsen, O. (2000). A mathematical expectation model for bird navigation based on the clock-and-compass strategy. *Journal of Theoretical Biology*, *207*(2), 283–291.

[92] Paindaveine, D., Verdebout, T. et al. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *The Annals of Statistics*, *45*(2), 800–832.

[93] Paindaveine, D., Verdebout, T. et al. (2020). Inference for spherical location under high concentration. *Annals of Statistics*, *48*(5), 2982–2998.

[94] Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*, *15*(6), 1191–1253.

[95] Paninski, L., & Yajima, M. (2008). Undersmoothed kernel entropy estimators. *IEEE Transactions on Information Theory*, *54*(9), 4384–4388.

[96] Partington, J. R. (1949). *An Advanced Treatise on Physical Chemistry.* (tech. rep.). Longmans, Green,

[97] Penrose, M., Yukich, J. et al. (2003). Weak laws of large numbers in geometric probability. *The Annals of Applied Probability*, *13*(1), 277–303.

[98] Penrose, M., & Yukich, J. (2011). Laws of large numbers and nearest neighbor distances. *Advances in Directional and Linear Statistics* (pp. 189–199). Springer.

[99] Penrose, M., Yukich, J. et al. (2013). Limit theory for point processes in manifolds. *The Annals of Applied Probability*, *23*(6), 2161–2211.

[100] Pewsey, A., & García-Portugués, E. (2021). Recent advances in directional statistics. *TEST*, 1–58.

[101] Philippatos, G. C., & Wilson, C. J. (1972). Entropy, market risk, and the selection of efficient portfolios. *Applied Economics*, *4*(3), 209–220.

[102] Pischiutta, M., Rovelli, A., Salvini, F., Di Giulio, G., & Ben-Zion, Y. (2013). Directional resonance variations across the Pernicana Fault, Mt Etna, in relation to brittle deformation fields. *Geophysical Journal International*, *193*(2), 986–996.

[103] Rényi, A. et al. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*.

[104] Rivest, L.-P. (1986). Modified Kent's statistics for testing goodness-of-fit for the fisher distribution in small concentrated samples. *Statistics & Probability Letters*, *4*(1), 1–4.

[105] Rivest, L.-P. et al. (1989). Spherical regression for concentrated Fisher-von Mises distributions. *The Annals of Statistics*, *17*(1), 307–317.

[106] Rivest, L.-P., Duchesne, T., Nicosia, A., & Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 445–463.

[107] Rosenblatt, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer, New York.

[108] Rosenthal, M., Wu, W., Klassen, E., & Srivastava, A. (2014). Spherical regression models using projective linear transformations. *Journal of the American Statistical Association*, *109*(508), 1615–1624.

[109] SenGupta, A., Kim, S., & Arnold, B. C. (2013). Inverse circular–circular regression. *Journal of Multivariate Analysis*, *119*, 200–208.

[110] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

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

[112] Shwartz, S., Zibulevsky, M., & Schechner, Y. Y. (2005). Fast kernel entropy estimation and optimization. *Signal Processing*, *85*(5), 1045–1058.

[113] Singh, H., Misra, N., Hnizdo, V., Fedorowicz, A., & Demchuk, E. (2003). Nearest neighbor estimates of entropy. *American Journal of Mathematical and Management Sciences*, *23*(3-4), 301–321.

[114] Solaro, N. (2004). Random variate generation from multivariate exponential power distribution. *Statistica & Applicazioni*, *2*(2), 25–44.

[115] Sricharan, K., Raich, R., & Hero, A. O. (2012). Estimation of nonlinear functionals of densities with confidence. *IEEE Transactions on Information Theory*, *58*(7), 4135–4159.

[116] Stein, E. M., & Weiss, G. (1972). Introduction to Fourier Analysis on Euclidean Spaces (pms-32).

[117] Stowell, D., & Plumbley, M. D. (2009). Fast multidimensional entropy estimation by k-d partitioning. *IEEE Signal Processing Letters*, *16*(6), 537–540.

[118] Tsagris, M., Athineou, G., Sajib, A., Amson, E., & Waldstein, M. (2020). Directional: Directional statistics. r package version 4.4.

[119] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, *52*(1), 479–487.

[120] 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.

[121] Van Es, B. (1992). Estimating functionals related to a density by a class of statistics based on spacings. *Scandinavian Journal of Statistics*, 61–72.

[122] Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(1), 54–59.

[123] Vázquez, J. L. (2007). *The Porous Medium Equation: Mathematical Theory*. Oxford University Press.

[124] Watamori, Y. (1996). Statistical inference of Langevin distribution for directional data. *Hiroshima Mathematical Journal*, *26*(1), 25–74.

[125] Watson, G. S. (1984). The theory of concentrated Langevin distributions. *Journal of Multivariate Analysis*, *14*(1), 74–82.

[126] Wirjadi, O., Godehardt, M., Schladitz, K., Wagner, B., Rack, A., Gurka, M., Nissle, S., & Noll, A. (2014). Characterization of multilayer structures in fiber reinforced polymer employing synchrotron and laboratory X-ray CT. *International Journal of Materials Research*, *105*(7), 645–654.

[127] Wyner, A. D., & Ziv, J. (1969). On communication of analog data from a bounded source space. *Bell System Technical Journal*, *48*(10), 3139–3172.

[128] Yukich, J. (1998). *Probability Theory of Classical Euclidean Optimization Problems*. Springer, Berlin.

[129] Yukich, J. (2006). *Probability Theory of Classical Euclidean Optimization Problems*. Springer.

[130] Zhang, Z. (2017). *Statistical Implications of Turing's Formula*. Wiley Online Library.

[131] Zhou, R., Cai, R., & Tong, G. (2013). Applications of entropy in finance: A review. *Entropy*, *15*(11), 4909–4931.

[132] Zografos, K., & Nadarajah, S. (2005). Expressions for Rényi and Shannon entropies for multivariate distributions. *Statistics & Probability Letters*, *71*(1), 71–84.

— Appendix E —

Numerical simulation codes

The python code below can be used to fit our simulation model to investigate the behaviour of test statistics. The data are generated sample points from multivariate generalized Gaussian distribution are generated using the Python software which was implemented for computing programme codes. Here, 5000 number of replications of experiments of different sizes ($N = 10, 100, 1000, vs$) are generated from several alternatives to normality and $\widehat{H}_{N,k}$ is computed from samples of different sample sizes generated from normal distributions with parameters $s = 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4$ and for the first, second and third nearest neighbour distances.

For an alternative hypothesis, the Student distribution has been considered. Between 100 and 1000 simulated random samples of size N have been generated and sample values of test statistics are evaluated using a Monte Carlo simulation and also each of the sample sizes has been repeated 5000 times. The aim is to demonstrate that this empirical distribution is not a Gaussian distribution.

For different values of (N, k) and (m, s), we generate $N_T = 1000$ samples from the $GG(m, s)$ distribution and record the corresponding values of $T_{N,k}(m, s)$, repeating this $M = 100$ times. To each of these 10 samples from the distribution of $T_{N,k}(m, s)$ we then apply the Shapiro-Wilk test for normality and record the p-value returned by the test.

```
import numpy as np
import scipy as sc
import scipy.stats as st
import matplotlib.pyplot as plt
import matplotlib.ticker as tkr
from sklearn.neighbors import KDTree
GR = (1+np.sqrt(5))/2 # aspect ratio for plots

# Multivariate Gaussian distribution
```

```
def create_points_GG(npts, dim=2, expo=2, std=True):
    '''
    If X = UR where U is uniformly distributed on the unit sphere in R^m
    and R = V^{1/s} where V ~ Gamma(m/s,2) then X ~ GG(m,s). From (Solaro 20
    '''
    # set mean vector and covariance matrix
    mvec = [0]*dim
    cmtx = np.identity(dim)
    # create isotropic normal vectors
    zpts = st.multivariate_normal.rvs(mvec, cmtx, npts)
    # project onto sphere
    upts = np.array([z/np.linalg.norm(z) for z in zpts])
    # create gamma values
    gvals = st.gamma.rvs(dim/expo, scale=2, size=npts)**(1/expo)
    # create points
    points = np.multiply(upts, gvals[:, np.newaxis]) if dim > 1 else
    np.reshape(np.multiply(upts, gvals),(npts,1))
    # standardise if required
    if std:
        sf = (2**(2/expo))*sc.special.gamma((dim+2)/expo)/
        (dim*sc.special.gamma(dim/expo))
        points = points/np.sqrt(sf)
    return points

# Multivariate Student-t distribution
def create_points_T(npts, dim=2, dof=10, std=False):
    '''If X = Z/\sqrt{G} where Z ~ N(0,I_m) and G ~
    Gamma(nu/2,2/nu) then  X ~ IST(m,nu)'''
    # set mean vector and covariance matrix
    mvec = [0]*dim
    cmtx = np.identity(dim)
    # create normal vectors
    zpts = np.random.multivariate_normal(mvec, cmtx, npts)
    # create gamma values
    gvals = np.tile(np.random.gamma(dof/2.0, 2.0/dof, npts),(dim,1)).T
    # create points
```

```
    points = zpts/np.sqrt(gvals)
    # standardise if required
    if std:
        sf = dof/(dof-2)
        points = points/np.sqrt(sf)
    return(points)
\begin{minted}{python}
def compute_constant(dim=2, expo=2):
    return (expo*np.e/dim)**(dim/expo)*np.pi**(dim/2)*sc.special.gamma
    (dim/expo+1)/sc.special.gamma(dim/2+1)

def compute_sample_moment(points, expo):
    # Euclidean norms
    norms = np.sqrt(np.sum(points**2, axis=1))
    # power-weighted norms
    pw_norms = norms**expo
    # value
    return np.mean(pw_norms)

def compute_near_neighbour_distances(points, nnmax):
    # search tree
    tree = KDTree(points)
    # extract distances
    dist, ind = tree.query(points, k=nnmax+1)
    # exclude zeroth neighbour (the point itself)
    return dist[:,1:]

def compute_entropy_estimates(points, nnmax=1):
    # dimensions
    npts, dim = points.shape
    # volume of unit ball
    vub = (np.pi**(dim/2))/(sc.special.gamma(dim/2 + 1))
    # digamma function values (scipy.special.digamma is slow)
    psi = -np.euler_gamma + np.array([0] +
    [1/i for i in range(1,nnmax)]).cumsum()
    # near neighbour distances
    nnd = compute_near_neighbour_distances(points, nnmax)
```

```
    # geometric means
    gmeans = sc.stats.mstats.gmean(nnd)
    # value
    return m*np.log(gmeans) + np.log(vub) + np.log(npts-1) - psi

def compute_statistics(points, expo=2, nnmax=1):
    # dimensions
    npts, dim = points.shape
    # entropy estimates
    eest = compute_entropy_estimates(points, nnmax)
    # moment estimate
    smom = compute_sample_moment(points, expo)
    # constant
    const = compute_constant(dim, expo)
    # value
    return eest - (dim/expo)*np.log(smom) - np.log(const)

# Set parameter values
# These are used further down!
# dimension
mvals = np.array([1,2,3])
# neighbours
kvals = np.array([1,2,3])
# exponent
svals = np.array([0.5, 1.0, 1.5, 2.0, 2.5])
# sample size
Nmin = 10, Nmax = 500, Ninc = 10
Nvals = np.arange(start=Nmin, stop=Nmax+1, step=Ninc)
# repetitions
nreps = 10

# Create data
def create_data(Nvals, mvals, svals, kvals, nreps):
    # init memory
    datacube = np.zeros(shape=(nreps,len(Nvals),len(mvals),
    len(svals), len(svals), len(kvals)))
    # info
```

```
    import datetime
    print('Started at: {}'.format(datetime.datetime.now()))
    # main loop
    for rep in range(nreps):
        # progress bar
        print('\r' + 'x'*(rep+1) + '-'*(nreps-rep-1), end='')
        # iterate over m-values
        for midx, m in enumerate(mvals):
            # iterate over sR-values (reality)
            for sRidx, sR in enumerate(svals):
                # create sample
                pts = create_points_GG(Nmax, dim=m, expo=sR)
                # iterate over subsamples
                for Nidx, N in enumerate(Nvals):
                    #iterate over sH-values (hypothesised)
                    for sHidx, sH in enumerate(svals):
                       datacube[rep,Nidx,midx,sRidx,sHidx:]=
                      compute_statistic(pts[:N], expo=sH,
                      nnmax=len(kvals))
    print('Ended at:   {}'.format(datetime.datetime.now()))
    return(datacube)
# check
if input("Generate data: are you sure? (y/n)") == "y":
    data  = create_data(Nvals, mvals, svals, kvals, nreps)

# save data
np.save('datacube1.npy', data)
data.shape

# load data
data2 = np.load('datacube1.npy')
data2.shape

# plot separately with errorbars (k fixed)
kval = 1
kidx = np.where(kvals==kval)[0][0]
width = 12
```

```
fig, axes = plt.subplots(nrows=len(svals), ncols=len(mvals),
sharex=True, sharey=True, figsize=(width,1.5*width))
for sidx, sval in enumerate(svals):
    for midx, mval in enumerate(mvals):
        ax = axes[sidx,midx]
        ax.axhline(y=0, linewidth=1, color='k')
        desc = st.describe(data[:,:,midx,sidx,sidx,kidx])
        ax.errorbar(x=Nvals, y=desc.mean, yerr=np.sqrt(desc.variance)/
        np.sqrt(nreps), capsize=5, errorevery=5)
        ax.set_title('m={}, s={}, k={}'.format(mval,sval,kval));
        if sidx == len(svals)-1: ax.set_xlabel('$N$', fontsize=12)
        if midx == 0: ax.set_ylabel('$T_{N,k}(m,s)$', fontsize=16)
        ax.set_xlim([Nmin,Nmax])
        ax.set_ylim([-0.05,0.05]) # tweak
        ax.grid(1);
fig.tight_layout()
plt.savefig('consistency-k={}.png'.format(kval))
plt.show();
```