Information theoretic results for stationary time series and the Gaussian-generalized von Mises time series

Riccardo Gatto
Submitted on November 17, 2020
Revised on January 20, 2021

Abstract This chapter presents some novel information theoretic results for the analysis of stationary time series in frequency domain. In particular, the spectral distribution that corresponds to the most uncertain or unpredictable time series with some values of the autocovariance function fixed, is the *generalized von Mises spectral distribution*. It is thus a maximum entropy spectral distribution and the corresponding stationary time series is called the *generalized von Mises time series*. The generalized von Mises distribution is used in directional statistics for modelling planar directions that follow a multimodal distribution. Furthermore, the *Gaussian-generalized von Mises times series* is presented as the stationary time series that maximizes entropies in frequency and time domains, respectively referred to as *spectral* and *temporal entropies*. Parameter estimation and some computational aspects with this time series are briefly analyzed.

1 Introduction

Nonstationary data typically have mean, variance and covariances that change significantly over time. It is consequently difficult to make reliable predictions or forecasts directly from these data. Thus, nonstationary data are transformed to stationary data, viz. data that possess constant mean, constant variance and constant covariance between any two observations that are separated by any fixed time lag. Stationary data are often analyzed in frequency domain, where the spectral distribution plays the

Riccardo Gatto
University of Bern, Department of Mathematics an Statistics, Institute of Mathematical Statistics and Actuarial Science, Alpeneggstrasse 22, 3012 Bern, Switzerland
e-mail: gatto@stat.unibe.ch

2010 Mathematics Subject Classification
62M20 Inference from stochastic processes: prediction; filtering
62H11 Multivariate analysis: directional data; spatial statistics

central role: it characterizes the correlations between the values of the time series and it allows for linear predictions. The analysis in frequency domain is particularly interesting for the identification of periodicities of the data. The first developments of the theory of stationary processes appeared at the end of the 19-th century with the analysis of data in frequency domain, which is called the spectral analysis. The alternative analysis in time domain, viz. based on the covariance function, appeared only later. The first statistical theory for periodic phenomena was developed by Fisher (1929). Other early leading contributions to the theory of stationary processes are Cramér (1942), Rice (1944 and 1945) as well as the volumes Cramér and Leadbetter (1967) and Yaglom (1962). A more recent volume on stationary processes is Lindgren (2012) and an historical review can be found in Brillinger (1993). This chapter provides various information theoretic results for spectral distributions of stationary processes with discrete time, i.e. stationary time series. It recasts the generalized von Mises (GvM) distribution, which was introduced in directional statistics as a model for planar directions, in the context of the spectral analysis of time series. It shows that the spectral distribution that corresponds to the most uncertain or unpredictable time series and whose autocovariance function agrees with some few first predetermined values, for example estimated from a sample, is the GvM spectral distribution. It is thus a maximum entropy spectral distribution and the corresponding stationary time series can be called the generalized von Mises time series. The Gaussian stationary time series with GvM spectral distribution, called Gaussian-GvM, follows the maximal entropy principle w.r.t. time and frequency. Although some estimation and other computational aspects are briefly analyzed, this chapter is only a first study of the Gaussian-GvM time series.

Let $\{X_j\}_{j\in\mathbb{Z}}$ be a complex-valued time series whose elements belong to a common Hilbert space $\mathcal{L}_2$ of square integrable random variables, thus $\mathsf{E}\left[|X_j|^2\right] < \infty, \forall j \in \mathbb{Z}$. Its autocovariance function (a.c.v.f.) is given by $\psi(j+r,j) = \operatorname{cov}\left(X_{j+r}, X_j\right) = \mathsf{E}\left[X_{j+r}\overline{X_j}\right] - \mathsf{E}[X_{j+r}]\mathsf{E}\left[\overline{X_j}\right], \forall j,r \in \mathbb{Z}$. We assume that the time series is weakly stationary, which will be shortened to stationary, precisely that $\mathsf{E}[X_j]$ and $\psi(j+r,j)$ do not depend on j, $\forall j,r \in \mathbb{Z}$. In this case we denote $\mu = \mathsf{E}[X_j]$, $\psi(r) = \psi(r,0) = \psi(j+r,j)$, $\forall j,r \in \mathbb{Z}$, and $\sigma^2 = \psi(0)$, for some $\sigma \in (0,\infty)$. A stronger type of stationarity is the strict stationarity, which requires that the double finite dimensional distributions (f.d.d.) of the time series are invariant after a fixed time shift, i.e. $\forall j_1 < \ldots < j_n \in \mathbb{Z}$, $r \in \mathbb{Z}$ and $n \geq 1$,

$$\left(U_{j_1}, \ldots, U_{j_n}, V_{j_1}, \ldots, V_{j_n}\right) \sim \left(U_{j_1+r}, \ldots, U_{j_n+r}, V_{j_1+r}, \ldots, V_{j_n+r}\right), \tag{1}$$

where $U_j = \operatorname{Re} X_j$ and $V_j = \operatorname{Im} X_j$, $\forall j \in \mathbb{Z}$. As usual, $E_1 \sim E_2$ means that the random elements E_1 and E_2 follow the same distribution. Stationary time series can be analyzed in frequency domain, precisely through the spectral distribution. A spectral distribution function (d.f.) is any nondecreasing function F_σ over $[-\pi,\pi]$ that satisfies $F_\sigma(-\pi) = 0$ and $F_\sigma(\pi) = \sigma^2$. This d.f. relates to the a.c.v.f. through the equation

$$\psi(r) = \int_{(-\pi,\pi]} e^{ir\theta} dF_\sigma(\theta),\ \forall r \in \mathbb{Z}.$$

The simplest nontrivial stationary time series $\{X_j\}_{j\in\mathbb{Z}}$ is called white noise if it has mean zero and a.c.v.f.

$$\psi(r) = \begin{cases} \sigma^2, & \text{if } r = 0, \\ 0, & \text{if } r = \pm1, \pm2, \ldots, \end{cases}$$

for some $\sigma > 0$. All frequencies of $\{X_j\}_{j\in\mathbb{Z}}$ are equally represented, because its spectral density is the uniform one with total mass σ^2, namely $f_\sigma(\theta) = \sigma^2/(2\pi)$, $\forall\theta \in (-\pi, \pi]$. The term white noise originates from the fact that white color reflects all visible wave frequencies of light. Real-valued time series are used in many applied sciences; refer e.g. to Brockwell and Davis (1991) or Chatfield (2013). However, complex-valued time series are often preferred representations of bivariate signals, mainly because their compact formulation. They have been applied in various technical domains, such as magnetic resonance imaging (cf. e.g. Rowe, 2005) or oceanography (cf. e.g. Gonella, 1972).

Spectral distributions of complex-valued time series can be viewed as rescaled circular distributions. For real-valued time series, the spectral distribution is a rescaled axially symmetric circular distribution. We recall that a circular distribution is a probability distribution over the circle that is used for modelling planar directions as well as periodic phenomena. During the last two decades, there has been a considerable amount of theoretical and applied research on circular distributions. Two major references are Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001). A short introduction is Gatto and Jammalamadaka (2015) and a recent review is Pewsey and García-Portugués (2020).

Let $k \in \{1, 2, \ldots\}$. A class of circular distributions that possess various theoretical properties has densities given by

$$f_1^{(k)}(\theta \mid \mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)} \exp\left\{\sum_{j=1}^{k} \kappa_j \cos j(\theta - \mu_j)\right\}, \tag{2}$$

$\forall\theta \in (-\pi, \pi]$ (or any other interval of length 2π), where $\mu_j \in (-\pi/j, \pi/j]$, $\kappa_j \geq 0$, for $j = 1, \ldots, k$,

$$G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi} \int_0^{2\pi} \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta_1) + \ldots + \kappa_k \cos k(\theta + \delta_{k-1})\} d\theta,$$

and where $\delta_j = (\mu_1 - \mu_{j+1}) \bmod (2\pi/(j+1))$, for $j = 1, \ldots, k-1$, whenever $k \geq 2$. The circular density (2) for $k \geq 2$ was thoroughly analyzed by Gatto and Jammalamadaka (2007) and Gatto (2009), who called it “generalized von Mises density of order

k" (GvM_k). Let us denote a circular random variable θ with that density as $\theta \sim \mathrm{GvM}_k(\mu_1, \dots, \mu_k, \kappa_1, \dots, \kappa_k)$. The GvM_1 density is the well-known circular normal or von Mises (vM) density, which represents within circular statistics what the normal distribution represents in linear statistics. It is given by $f_1^{(1)}(\theta \mid \mu_1, \kappa_1) = \{2\pi I_0(\kappa_1)\}^{-1} \exp\{\kappa_1 \cos(\theta - \mu_1)\}$, $\forall \theta \in (-\pi, \pi]$, where $\mu_1 \in (-\pi, \pi]$, $\kappa_1 \geq 0$ and where $I_n(z) = (2\pi)^{-1} \int_0^{2\pi} \cos n\theta \, \exp\{z \cos \theta\} \mathrm{d}\theta$, $\forall z \in \mathbb{C}$, is the modified Bessel function of the first kind and integer order n (see e.g. Abramowitz and Stegun, 1972, p. 376). Compared to the vM, which is axially symmetric and unimodal whenever $\kappa_1 > 0$, the GvM_2 distribution allows for substantially higher adjustability, in particular in terms of asymmetry and bimodality. This makes it a practical circular distribution that has found various applications. Some recent ones are: Zhang et al. (2018), in meteorology, Lin and Dong (2019), in oceanography, Astfalck et al. (2018), in offshore engineering, and in Christmas (2014), in signal processing. The GvM_k spectral density is given by $f_\sigma^{(k)} = \sigma^2 f_1^{(k)}$, for some $\sigma \in (0, \infty)$: it is the GvM_k circular density $f_1^{(k)}$ given in (2) that is rescaled to have any desired total mass σ^2. When the GvM_k spectral density is axially symmetric around the null axis, then the corresponding time series $\{X_j\}_{j \in \mathbb{Z}}$ is real-valued. As shown in Salvador and Gatto (2021a), the GvM_2 density with $\kappa_1, \kappa_2 > 0$ is axially symmetric iff $\delta_1 = 0$ or $\delta_1 = \pi/2$. In both cases, the axis of symmetry has angle μ_1 with respect to (w.r.t.) the null direction. The GvM_2 spectral density has a practical role time series because of its uni- and bimodal shape. A complete analysis of the number of modes of the GvM_2 distribution is given in Salvador and Gatto (2021b). Note that in some situations a three-parameter version of the GvM_2 distribution introduced by Kim and SenGupta (2013) appears sufficient to model both asymmetric and bimodal data. The densities of this subclass are obtained by setting $\delta_1 = \pi/4$ and $k = 2$ in the GvM_k density (2). However, this subclass does not possess the optimality properties of the GvM_2 distribution that are presented in Section 2. It is worth mentioning that the GvM spectral distribution has many similarities with the exponential model of Bloomfield (1973), which is a truncated Fourier series of the logarithm of some spectral distribution. Bloomfield motivates the low truncation of the Fourier series by the fact that "the logarithm of an estimated spectral density function is often found to be a fairly well-behaved function". A closely related reference is Healy and Tukey (1963). However, Bloomfield's model is given for real-valued time series only.

The estimation of the spectral distribution is an important problem in the analysis of stationary time series. Information theoretic quantities like Kullback-Leibler's information (cf. Kullback and Leibler, 1951) or Shannon's entropy (cf. Shannon, 1948) are very useful in this context. These quantities are defined for probability distributions but they can be considered for distributions with finite mass. These are spectral distributions and we assume them absolutely continuous. Thus, let f_σ and g_σ be two spectral densities whose integrals over $(-\pi, \pi]$ are both equal to σ^2. The spectral Kullback-Leibler information of f_σ w.r.t. g_σ is given by

$$I(f_\sigma | g_\sigma) = \int_{-\pi}^{\pi} \log \frac{f_\sigma(\theta)}{g_\sigma(\theta)} f_\sigma(\theta) \mathrm{d}\theta = \sigma^2 I(f_1 | g_1), \tag{3}$$

where $0 \log 0 = 0$ is assumed and where the support of f_σ is included in the support of g_σ, otherwise $I(f_\sigma|g_\sigma) = \infty$. It follows from Gibbs inequality that $I(f_\sigma|g_\sigma)$ is nonnegative, precisely $I(f_\sigma|g_\sigma) \geq 0$, for all possible spectral densities f_σ and g_σ, with equality iff $f_1 = g_1$ a.e. The Kullback-Leibler information is also called relative entropy, Kullback-Leibler divergence or distance, eventhough it is not a metric. Thus (3) is a measure of divergence for distributions with same total mass σ^2. Shannon's entropy can be defined for the spectral density f_σ by

$$S(f_\sigma) = -\int_{-\pi}^{\pi} \log \frac{f_\sigma(\theta)}{(2\pi)^{-1}\sigma^2} f_\sigma(\theta)\mathrm{d}\theta = -I(f_\sigma|u_\sigma) = -\sigma^2 I(f_1|u_1), \tag{4}$$

where u_σ is the uniform density with total mass σ^2 over $(-\pi,\pi]$, viz. $u_\sigma = \sigma^2/(2\pi)\mathrm{I}_{(-\pi,\pi]}$, I_A denoting the indicator of set A. Shannon's entropy of the circular density f_1 over $(-\pi,\pi]$ is originally defined as $-\int_{-\pi}^{\pi} \log f_1(\theta) f_1(\theta)\mathrm{d}\theta = -(2\pi)^{-1} - I(f_1|u_1)$. It measures the uncertainty inherent in the probability distribution with density f_1. Equivalently, $S(f_1)$ measures the expected amount of information gained on obtaining an observation from f_1, based on the principle that the rarer an event, the more informative its occurrence. The spectral entropy defined in (4) slightly differs the original formula of Shannon's entropy for probability distributions: inside the logarithm, f_σ is divided by the uniform density with total mass σ^2. With this modification the spectral entropy becomes scale invariant w.r.t. σ^2, just like the spectral Kullback-Leibler information (3). The spectral entropy satisfies $S(f_\sigma) \leq 0$, with equality iff $f_\sigma = u_\sigma$ a.e. This follows from Gibbs inequality.

The topics of the next sections of this chapter are the following. Section 2 provides information theoretic results for spectral distributions and introduces the related GvM and the Gaussian-GvM time series. Section 2.1 gives general definitions and concepts. Section 2.2 provides the optimal spectral distributions under constraints on the a.c.v.f. The GvM spectral distribution maximizes Shannon's entropy under constraints on the first few values of the a.c.v.f. Section 2.3 motivates the Gaussian-GvM time series from the fact that it follows the maximal entropy principle in both time and frequency domains. Section 3 provides some computational aspects. Section 3.1 gives some series expansions for integral functions appearing in the context of the GvM_2 time series. An estimator for the parameters of the GvM spectral distribution is presented in Section 3.2. Section 3.3 provides an expansion for the GvM spectral d.f. Some short concluding remarks are given in Section 4.

2 The GvM and the Gaussian-GvM time series

Section 2.1 summarizes central results of time series and defines the GvM time series. Section 2.2 provides information theoretic results for spectral distributions. An important result is that the GvM spectral distribution maximizes the entropy under constraints on the a.c.v.f. Section 2.3 proposes the Gaussian-GvM time series based on the fact that it follows the maximal entropy principle in both time and frequency domains, under the same constraints.

2.1 General considerations

Two central theorems of spectral analysis of time series are the following. The first one is Herglotz theorem:

$\psi : \mathbb{Z} \to \mathbb{C}$ is nonnegative definite (n.n.d.)[1] $\Leftrightarrow \psi(r) = \int_{(-\pi,\pi]} \mathrm{e}^{\mathrm{i}r\theta} \mathrm{d}F_\sigma(\theta)$, $\forall r \in \mathbb{Z}$, *for some d.f. F_σ over $[-\pi,\pi]$, with $F_\sigma(-\pi) = 0$ and $\sigma^2 = F_\sigma(\pi) \in (0,\infty)$.*

The second theorem is a characterization of the a.c.v.f.:

$\psi : \mathbb{Z} \to \mathbb{C}$ is the a.c.v.f. of a (strictly) stationary complex-valued time series $\Leftrightarrow$ ψ is n.n.d.

These two theorems can be found at p. 117-119 of Brockwell and Davis (1991). They tell that if we consider the spectral d.f. $F_\sigma^{(k)} = \sigma^2 F_1^{(k)}$, where $F_1^{(k)}$ is the GvM$_k$ d.f. with density $f_1^{(k)}$ given by (2), then there exists a stationary time series $\{X_j\}_{j\in\mathbb{Z}}$ with spectral d.f. $F_\sigma^{(k)}$ and density $f_\sigma^{(k)} = \sigma f_1^{(k)}$ that we call GvM or, more precisely, GvM$_k$ time series. Thus the GvM$_k$ time series is stationary by definition, it has variance $F_\sigma^{(k)}(\pi) = \sigma^2$ and it is generally complex-valued, unless the GvM$_k$ spectral distribution is axially symmetric around the null direction.

The complex-valued GvM$_k$ stationary time series $\{X_j\}_{j\in\mathbb{Z}}$ can be chosen with mean zero, variance σ^2 and Gaussian, meaning that the double f.d.d. given in (1) are Gaussian. In this case, the distribution of $\{X_j\}_{j\in\mathbb{Z}}$ is however not entirely determined by its a.c.v.f. $\psi^{(k)}$ or, alternatively, by its spectral d.f. $F_\sigma^{(k)}$. (The formula for the a.c.v.f. is given later in Corollary 1.4.) In order to entirely determine this distribution, one also needs the so-called pseudo-covariance $\mathsf{E}[X_{j+r}X_j]$, $\forall j, r \in \mathbb{Z}$. So an arbitrary Gaussian, with mean zero and (weakly) stationary time series $\{X_j\}_{j\in\mathbb{Z}}$ is not necessarily strictly stationary: $\{X_j\}_{j\in\mathbb{Z}}$ is strictly stationary iff the covariance $\mathsf{E}\left[X_{j+r}\overline{X_j}\right]$ and the pseudo-covariance $\mathsf{E}[X_{j+r}X_j]$ do not depend on $j \in \mathbb{Z}$, $\forall r \in \mathbb{Z}$. This is indeed equivalent to the independence on $j \in \mathbb{Z}$ of

$$\begin{aligned}&\psi_{UU}(r) = \mathsf{E}[U_{j+r}U_j],\ \psi_{VV}(r) = \mathsf{E}[V_{j+r}V_j],\\ &\psi_{UV}(r) = \mathsf{E}[U_{j+r}V_j]\ \text{ and } \psi_{VU}(r) = \mathsf{E}[V_{j+r}U_j],\ \forall r \in \mathbb{Z},\end{aligned} \tag{5}$$

where $U_j = \operatorname{Re} X_j$ and $V_j = \operatorname{Im} X_j$, $\forall j \in \mathbb{Z}$. Note that under this independence on $j \in \mathbb{Z}$, we have $\psi_{VU}(r) = \psi_{UV}(-r)$, $\forall r \in \mathbb{Z}$. However, according Herglotz theorem, if the a.c.v.f. $\psi^{(k)}$ is obtained by Fourier inversion of the GvM$_k$ spectral density, then

[1] The function $f : \mathbb{R} \to \mathbb{C}$ is n.n.d. if $\sum_{i=1}^n \sum_{j=1}^n c_i \overline{c}_j f(x_i - x_j) \geq 0$, $\forall x_1, \ldots, x_n \in \mathbb{R}$, $c_1, \ldots, c_n \in \mathbb{C}$ and $n \geq 1$.
Any n.n.d. function f is Hermitian, i.e. $f(-x) = \overline{f(x)}$, $\forall x \in \mathbb{R}$.

it is n.n.d. By the above characterization of the a.c.v.f., a strictly stationary GvM_k time series always exists. The existence of a particular (precisely radially symmetric) strictly stationary Gaussian-GvM_k time series that satisfies some constraints on the a.c.v.f. is shown in Section 2.3.

Next, for any given Gaussian-GvM_k time series with spectral d.f. $F_\sigma^{(k)}$, there exists a spectral process $\{Z_\theta\}_{\theta\in[-\pi,\pi]}$ that is complex-valued and Gaussian. We remind that the process that modulates the harmonics, $\{Z_\theta\}_{\theta\in[-\pi,\pi]}$, is defined through the mean square stochastic integral

$$X_j = \int_{(-\pi,\pi]} \mathrm{e}^{\mathrm{i}\theta j} \mathrm{d}Z_\theta, \quad \text{a.s.}, \forall j \in \mathbb{Z}, \tag{6}$$

and by the following conditions: $\mathsf{E}[Z_\theta] = 0, \forall\theta \in [-\pi,\pi], \mathsf{E}\left[\left(Z_{\theta_2} - Z_{\theta_1}\right) \overline{\left(Z_{\theta_4} - Z_{\theta_3}\right)}\right] = 0, \forall -\pi \leq \theta_1 < \theta_2 < \theta_3 < \theta_4 \leq \pi$, viz. it has orthogonal increments, and

$$\mathsf{E}\left[\left|Z_{\theta_2} - Z_{\theta_1}\right|^2\right] = F_\sigma^{(k)}(\theta_2) - F_\sigma^{(k)}(\theta_1), \ \forall -\pi \leq \theta_1 < \theta_2 \leq \pi. \tag{7}$$

There are several reasons for considering the Gaussian-GvM time series. A practical one is that their simulation can be done with the algorithms presented in Chapter XI of Asmussen and Glynn (2007), where one of these algorithms makes use of the decomposition (6). A theoretical reason for considering normality is that it leads to a second maximal entropy principle, this one no longer in frequency domain but in time domain. We pursue this explanation on temporal entropy in Section 2.3.

2.2 Spectral Kullback-Leibler information and entropy

Let g_σ be the spectral density of some stationary time series with variance σ^2, for some $\sigma \in (0,\infty)$. For a chosen $k \in \{1,2,\ldots\}$, consider the a.c.v.f. conditions or constraints

$$C_k : \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}r\theta} g_\sigma(\theta)\mathrm{d}\theta = \psi_r, \quad \text{for } r = 1,\ldots,k, \tag{8}$$

for some $\psi_1,\ldots,\psi_k \in \mathbb{C}$ satisfying $|\psi_r| \leq \sigma^2$, for $r = 1,\ldots,k$, and such that the $(k+1)\times(k+1)$ matrix

$$\begin{pmatrix} \sigma^2 & \psi_1 & \ldots & \psi_k \\ \overline{\psi_1} & \sigma^2 & \ldots & \psi_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\psi_k} & \overline{\psi_{k-1}} & \ldots & \sigma^2 \end{pmatrix} \tag{9}$$

is n.n.d. One can re-express these conditions as

$$C_k : \int_{-\pi}^{\pi} \cos r\theta \, g_\sigma(\theta) \mathrm{d}\theta = \nu_r \quad \text{and} \quad \int_{-\pi}^{\pi} \sin r\theta \, g_\sigma(\theta) \mathrm{d}\theta = \xi_r, \quad \text{for } r = 1, \ldots, k, \tag{10}$$

where $\nu_r = \mathrm{Re}\, \psi_r$ and $\xi_r = \mathrm{Im}\, \psi_r$, giving thus $\sqrt{\nu_r^2 + \xi_r^2} \leq \sigma^2$, for $r = 1, \ldots, k$, and with n.n.d. matrix (9).

One encounters the two following practical problems. In an applied field where a specific spectral density h_σ is traditionally used,[2] one may search for the spectral density g_σ that satisfies C_k and that is the closest to the traditional density h_σ. Alternatively, the spectral density g_σ is unknown but the values of $\psi_1, \ldots, \psi_k$ are available, either because they constitute a priori knowledge about the time series or because they are obtained from a sample of the stationary time series. In this second case, the values of $\psi_1, \ldots, \psi_k$ can be obtained by taking them equal to the corresponding values of the empirical or sample a.c.v.f. For the sample $X_1, \ldots, X_n$ of the time series, the sample a.c.v.f. is given by

$$\hat{\psi}_n(r) = \frac{1}{n} \sum_{j=1}^{n-r} (X_{j+r} - M_n)\overline{(X_j - M_n)} \quad \text{and} \quad \hat{\psi}_n(-r) = \overline{\hat{\psi}_n(r)}, \quad \text{for } r = 0, \ldots, n-1, \tag{11}$$

where $M_n = n^{-1} \sum_{j=1}^{n} X_j$. Thus we set $\psi_r = \hat{\psi}_n(r)$, for $r = 1, \ldots, k$ and for $k \leq n-1$. Note that the matrix (9) is n.n.d. in this case. Note also that the sample a.c.v.f. is a biased estimator of the true a.c.v.f. (but asymptotically unbiased).

Theorem 1 below addresses the first of these two problems and it is the central part of this article. The second problem is addressed by Corollary 1. The following definitions are required. For $k = 1, 2, \ldots$ and for an arbitrary circular density g_1, define the following integral functions:

$$G_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k; g_1) =$$

$$\int_0^{2\pi} \cos r\theta \, \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta_1) + \ldots + \kappa_k \cos k(\theta + \delta_{k-1})\} \, g_1(\theta) \mathrm{d}\theta,$$

$$H_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k; g_1) =$$

$$\int_0^{2\pi} \sin r\theta \, \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta_1) + \ldots + \kappa_k \cos k(\theta + \delta_{k-1})\} g_1(\theta) \mathrm{d}\theta,$$

$$A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k; g_1) = \frac{G_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k; g_1)}{G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k; g_1)}$$

and

[2] Related comments are given in Section 4.

$$B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;g_1)=\frac{H_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;g_1)}{G_0^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;g_1)},$$

for $r=1,\ldots,k$, where $\delta_j=(\mu_1-\mu_{j+1})\mathrm{mod}(2\pi/(j+1))$, for $j=1,\ldots,k-1$ and $\kappa_1,\ldots,\kappa_k\geq 0$. For these constants we make the conventions that the arguments $\delta_1,\ldots,\delta_{k-1}$ vanish when $k=1$ and that the argument g_1 is omitted when equal to the circular uniform density u_1. For example, $G_0^{(1)}(\kappa_1)=(2\pi)^{-1}\int_0^{2\pi}\mathrm{e}^{\kappa_1\cos\theta}\mathrm{d}\theta=I_0(\kappa_1)$. Define the matrix of counter-clockwise rotation of angle θ as

$$\mathbf{R}(\theta)=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}. \quad (12)$$

Theorem 1 (Kullback-Leibler closest spectral distribution)
Let $\sigma\in(0,\infty)$ and let g_σ and h_σ be two spectral densities with total mass σ^2.

1. *The spectral density g_σ that satisfies C_k, given in and that is the closest to another spectral density h_σ, in the sense of minimizing the Kullback-Leibler information $I(g_\sigma|h_\sigma)$, is the exponential tilt of h_σ that takes the form*

$$g_\sigma(\theta)=\frac{1}{G_0^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;h_1)}\exp\left\{\sum_{j=1}^k\kappa_j\cos j(\theta-\mu_j)\right\}h_\sigma(\theta), \quad (13)$$

$\forall\theta\in(-\pi,\pi]$, where $\delta_j=(\mu_1-\mu_{j+1})\mathrm{mod}(2\pi/(j+1))$, for $j=1,\ldots,k-1$, $\mu_j\in(-\pi/j,\pi/j]$ and $\kappa_j\geq 0$, for $j=1,\ldots,k$. The values of these parameters are the solutions of

$$\begin{pmatrix}\nu_r\\ \xi_r\end{pmatrix}=\sigma^2\mathbf{R}(r\mu_1)\begin{pmatrix}A_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;h_1)\\ B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;h_1)\end{pmatrix}, \quad (14)$$

where $\mathbf{R}(r\mu_1)$ denotes the rotation matrix (12) at $\alpha=r\mu_1$, for $r=1,\ldots,k$, and where $\nu_1,\ldots,\nu_k$ and $\xi_1,\ldots,\xi_k$ are given by (10).

2. *For any spectral density g_σ that satisfies C_k, the minimal Kullback-Leibler information of g_σ w.r.t. h_σ is given by*

$$-\sigma^2\log G_0^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;h_1)+\sum_{r=1}^k\kappa_r(\nu_r\cos r\mu_r+\xi_r\sin r\mu_r)\leq I(g_\sigma|h_\sigma), \quad (15)$$

with equality iff g_σ is a.e. given by (13), where the values of the parameters $\mu_j\in(-\pi/j,\pi/j]$ and $\kappa_j\geq 0$, for $j=1,\ldots,k$, are solutions of (14).

Theorem 1 is a rather direct consequence or generalization of Theorem 2.1 of Gatto (2009), in which the trigonometric moments are replaced by the a.c.v.f. and the circular distribution is replaced by the spectral distribution. Indeed, along with the generalization of the circular distribution to the spectral distribution, the a.c.v.f. of a stationary time series generalizes the trigonometric moment. Precisely, the r-th trigonometric moment of the circular random variable θ with density g_1 is given by

$$\varphi_r = \gamma_r + \mathrm{i}\sigma_r = \mathsf{E}\left[\mathrm{e}^{\mathrm{i}r\theta}\right] = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}r\theta} g_1(\theta)\mathrm{d}\theta, \tag{16}$$

for some $\gamma_r, \sigma_r \in \mathbb{R}$ and $\forall r \in \mathbb{Z}$, whereas the a.c.v.f. of the stationary time series with the spectral density $g_\sigma = \sigma^2 g_1$ is given by

$$\psi(r) = \sigma^2\varphi_r = \sigma^2(\gamma_r + \mathrm{i}\sigma_r),\ \forall r \in \mathbb{Z}. \tag{17}$$

Clearly, $\psi(0) = \sigma^2$ and $|\psi(r)| \leq \psi(0)$, $\forall r \in \mathbb{Z}$. The claim that (17) is indeed the a.c.v.f. of a stationary time series is rigorously justified by the above mentioned Herglotz theorem and characterization of the a.c.v.f.

The existence and the unicity of the solution to (14), i.e. of the parameter values satisfying C_k, can be justified by the fact (14) can be reparametrized in terms of the saddlepoint equation (or exponential tilting equation) given by (14) of Gatto (2009). This is the saddlepoint equation of a distribution with bounded domain. In this case, the solution, called saddlepoint, exists and it is unique. These facts are well know in the theory of large deviations.

In the context of the justification of Theorem 1.1, we can note that an equivalent expression to (14) is given by

$$\begin{aligned}\psi_r &= \sigma^2\mathrm{e}^{\mathrm{i}r\mu_1}\\ &\quad\cdot\left\{A_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;h_1) + \mathrm{i}B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k;h_1)\right\},\end{aligned}$$

for $r = 1,\ldots,k$, which can be seen equivalent to C_k.

When analyzing a time series with periodic components, leading for example to certain monthly or weekly constraints, then the set of k constraints C_k may no longer be appropriate. Instead of it, one may still need the constraints in the form given in (8) but exclusively for r limited to some subset of $\{1,\ldots,k\}$, which is possibly different than $\{1,\ldots,j\}, \forall j \in \{1,\ldots,k\}$. Theorem 1 can be easily generalized to this situation. For simplicity, assume that only the l-th constraint must be removed from C_k, for some $l \in \{1,\ldots,k-1\}$, and thus assume $k \geq 2$. Then Theorem 1 has to be adapted by setting $\kappa_l = 0$ in (13) and by removing the equation (14) whenever $r = l$. In addition, if $l = 1$, then μ_1 appearing in δ_j given just after (13) and appearing in (14) must be replaced by μ_m, with m arbitrary selected in $\{2,\ldots,k\}$. Similar adaptations could be considered for the next results of this chapter, essentialy to Corollary 1 and to Theorem 2. Writing the results of this chapter in the most general form would have negative repercussions on readability; cf. Gatto (2009).

A major consequence of Theorem 1 is that the GvM$_k$ spectral distribution is a maximum entropy distribution. This fact and related results are given in Corollary 1.

Corollary 1 (Maximal Shannon's spectral entropy distribution and GvM a.c.v.f.)

Let $\sigma \in (0,\infty)$ *and* g_σ *a spectral density with total mass* σ^2.

1. *The spectral density* g_σ *that maximizes Shannon's entropy* $S(g_\sigma)$ *under* C_k*, given in (8), is the* $\mathrm{GvM}_{\sigma,k}(\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$ *density, i.e.* $f_\sigma^{(k)}(\cdot|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$*, where* $\mu_j \in (-\pi/j,\pi/j]$ *and* $\kappa_j \geq 0$*, for* $j = 1,\ldots,k$*. The values of these parameters are determined by (14).*
2. *If* g_σ *is a spectral density satisfying* C_k*, then its entropy is bounded from above as follows,*
$$S(g_\sigma) \leq \sigma^2 \log G_0^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k) - \sum_{r=1}^{k} \kappa_r(\nu_r \cos r\mu_r + \xi_r \sin r\mu_r),$$
with equality iff $g_\sigma = f_\sigma^{(k)}(\cdot|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$ *a.e. The values of the parameters are determined by (14) with* $h_1 = u_1$*, i.e. the circular uniform density, where* $\nu_1,\ldots,\nu_k$ *and* $\xi_1,\ldots,\xi_k$ *are given by (10).*
3. *The entropy of the* $\mathrm{GvM}_{\sigma,k}(\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$ *spectral density is given by*
$$\begin{aligned} S\left(f_\sigma^{(k)}\right) = \sigma^2 \Bigg\{ & \log G_0^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k) \\ & - \kappa_1 A_1^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k) \\ & - \sum_{r=2}^{k} \kappa_r \big[A_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k)\cos r\delta_{r-1} \\ & - B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k)\sin r\delta_{r-1} \big] \Bigg\}, \end{aligned}$$
where $\sum_{r=2}^{k}$ *vanishes whenever* $k < 2$.
4. *The a.c.v.f.* $\psi_\sigma^{(k)}$ *of the* $\mathrm{GvM}_{\sigma,k}(\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$ *spectral distribution can be obtained by*
$$\begin{pmatrix} \mathrm{Re}\,\psi_\sigma^{(k)}(r) \\ \mathrm{Im}\,\psi_\sigma^{(k)}(r) \end{pmatrix} = \sigma^2 \mathbf{R}(r\mu_1) \begin{pmatrix} A_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k) \\ B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k) \end{pmatrix}$$
and $\psi_\sigma^{(k)}(-r) = \overline{\psi_\sigma^{(k)}(r)}$*, for* $r = 1,2,\ldots$.

Corollary 1 can be obtained from Theorem 1 as follows. Theorem 1.1 and the relation between Kullback-Leibler information and entropy (4) tell that the GvM_k spectral distribution maximizes the entropy, under the given constraints on the a.c.v.f. The upper bound for the entropy of a spectral distribution satisfying the given constraints is provided by Theorem 1.2. Thus, by considering $h_1 = u_1$ in Theorem 1, we obtain parts 1 and 2 of Corollary 1. Part 3 is a consequence of part 2. It is obtained by replacing ν_r and ξ_r, for $r = 1,\ldots,k$, that appear in the upper bound of the entropy, by expressions depending on the parameters of the GvM_k distribution, through the identity (14).

In the practice, when partial prior information in the form of C_k is available and it is desired to determine the most noninformative spectral distribution that satisfies the known prior information, then the GvM$_k$ spectral distribution is the optimal choice. It is in fact the most credible distribution or the one that nature would have generated, when some prior information and only that information would be available. Maximal entropy distributions are important in many contexts: in statistical mechanics, the choice of a maximum entropy distribution subject to constraints is a classical approach referred to as the maximum entropy principle. One can find various studies on spectral distributions with maximal entropy. It is explained in Section 3.2 that the autoregressive model of order k (AR(k)) maximizes an alternative entropy among all stationary time series satisfying C_k. Franke (1985) showed that the autoregressive and moving average time series (ARMA) maximizes that entropy among all stationary time series satisfying these same constraints and additional constraints on the impulse responses. Further properties on these optimal ARMA time series can be found in Huang (1990). Other references on spectral distributions with maximal entropy are, for instance: Burg (1978), Kay and Marple (1981) and Laeri (1990).

The simplest situation is the following.

*Example 1 (*vM *spectrum)* Corollary 1.3 with $k = 1$ yields the entropy of the vM spectral distribution,

$$S\left(f_\sigma^{(1)}\right) = \sigma^2 \left\{\log G_0^{(1)}(\kappa_1) - \kappa_1 A_1^{(1)}(\kappa_1)\right\} = \sigma^2 \left\{\log I_0(\kappa_1) - \kappa_1 \frac{I_1(\kappa_1)}{I_0(\kappa_1)}\right\},$$

for $\kappa_1 \geq 0$. By noting that $B_r^{(1)}(\kappa_1) = 0$, for $r = 1, 2, \ldots$, Corollary 1.4 with $k = 1$ gives the a.c.v.f. of the vM spectral distribution as

$$\begin{pmatrix} \mathrm{Re}\,\psi^{(1)}(r) \\ \mathrm{Im}\,\psi^{(1)}(r) \end{pmatrix} = \sigma^2 A_r^{(1)}(\kappa_1) \begin{pmatrix} \cos r\mu_1 \\ \sin r\mu_1 \end{pmatrix} = \sigma^2 \frac{I_r(\kappa_1)}{I_0(\kappa_1)} \begin{pmatrix} \cos r\mu_1 \\ \sin r\mu_1 \end{pmatrix}, \tag{18}$$

and $\psi^{(1)}(-r) = \overline{\psi^{(1)}(r)}$, for $r = 1, 2, \ldots$. When $\kappa_1 > 0$, the vM spectral distribution is axially symmetric about the origin iff $\mu_1 = 0$. In other terms and according to (18), the GvM$_1$ or vM time series is real-valued iff $\mu_1 = 0$.

2.3 Temporal entropy

This section provides a strictly stationary Gaussian-GvM time series that follows the maximal entropy principle in the time domain, in addition to the maximal entropy principle in frequency domain, under common constraints on the a.c.v.f. Consider the complex-valued Gaussian time series $\{X_j\}_{j\in\mathbb{Z}}$ in $\mathcal{L}_2$ that is strictly stationary with mean zero. This time series is introduced at the end of Section 2.1. Define $U_j = \mathrm{Re}\, X_j$ and $V_j = \mathrm{Im}\, X_j$, $\forall j \in \mathbb{Z}$. Let $n \geq 1$ and $j_1 < \ldots < j_n \in \mathbb{Z}$. Consider the random vector $(U_{j_1}, \ldots, U_{j_n}, V_{j_1}, \ldots, V_{j_n})$ and denote by $p_{j_1,\ldots,j_n}$ its joint density.

Thus $p_{j_1,\ldots,j_n}$ is the $2n$-dimensional normal density with mean zero and $2n \times 2n$ covariance matrix

$$\Sigma_{j_1,\ldots,j_n} = \operatorname{var}\left(\left(U_{j_1},\ldots,U_{j_n},V_{j_1},\ldots,V_{j_n}\right)\right) = \mathsf{E}\left[\begin{pmatrix} \mathbf{U}\mathbf{U}^\top & \mathbf{U}\mathbf{V}^\top \\ \mathbf{V}\mathbf{U}^\top & \mathbf{V}\mathbf{V}^\top \end{pmatrix}\right], \tag{19}$$

where $\mathbf{U} = \left(U_{j_1},\ldots,U_{j_n}\right)^\top$ and $\mathbf{V} = \left(V_{j_1},\ldots,V_{j_n}\right)^\top$. According to (5), the elements of $\Sigma_{j_1,\ldots,j_n}$ are given by

$$\mathsf{E}[U_{j_l}U_{j_m}] = \psi_{UU}(j_l - j_m),\ \mathsf{E}[V_{j_l}V_{j_m}] = \psi_{VV}(j_l - j_m),$$
$$\mathsf{E}[U_{j_l}V_{j_m}] = \psi_{UV}(j_l - j_m) \text{ and } \mathsf{E}[V_{j_l}U_{j_m}] = \psi_{VU}(j_l - j_m),$$

with $\psi_{VU}(j_l - j_m) = \psi_{UV}(j_m - j_l)$, for $l, m = 1,\ldots,n$. Because $\Sigma_{j_1,\ldots,j_n}$ depends on $j_1,\ldots,j_n$ only through $l_1 = j_2 - j_1,\ldots,l_{n-1} = j_n - j_{n-1}$, we consider the alternative notation $\Sigma^{l_1,\ldots,l_{n-1}} = \Sigma_{j_1,\ldots,j_n}$.

An important subclass of complex-valued normal random vectors is given by the radially symmetric ones, which is obtained by setting the mean and the pseudo-covariance matrix equal to zero. That is, the Gaussian vector $\mathbf{X} = (X_{j_1},\ldots,X_{j_n})^\top$, where $X_{j_l} = U_{j_l} + \mathrm{i}V_{j_l}$, for $l = 1,\ldots,n$, is radially symmetric iff $\mathsf{E}[\mathbf{X}] = \mathbf{0}$ and $\mathsf{E}\left[\mathbf{X}\mathbf{X}^\top\right] = \mathbf{0}$. A radially symmetric complex normal random vector $\mathbf{X}$ is characterized by the fact that, $\forall\theta \in (-\pi,\pi]$, $\mathrm{e}^{\mathrm{i}\theta}\mathbf{X} \sim \mathbf{X}$. Because these vectors and the related processes are often used in signal processing, we consider them in this section.

Generally, by assuming neither stationarity nor normality, one defines the temporal entropy of the complex-valued time series $\{X_j\}_{j\in\mathbb{Z}}$ at times $j_1 < \ldots < j_n \in \mathbb{Z}$ in terms of Shannon's entropy of $\left(U_{j_1},\ldots,U_{j_n},\ V_{j_1},\ldots,V_{j_n}\right)$, precisely as

$$T_{j_1,\ldots,j_n} = -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \log p_{j_1,\ldots,j_n}(u_1,\ldots,u_n,v_1,\ldots,v_n) \\ p_{j_1,\ldots,j_n}(u_1,\ldots,u_n,v_1,\ldots,v_n)\mathrm{d}u_1\ldots\mathrm{d}u_n\mathrm{d}v_1\ldots\mathrm{d}v_n, \tag{20}$$

whenever the density $p_{j_1,\ldots,j_n}$ exists. Under strict stationarity, the temporal entropy (20) becomes invariant under time shift and we can thus define the alternative notation $T^{l_1,\ldots,l_{n-1}} = T_{j_1,\ldots,j_n}$.

Let us now mention two known and important information theoretic results for the Gaussian distribution. The first one is the formula of the Gaussian entropy:

if $p_{j_1,\ldots,j_n}$ is the $2n$-dimensional Gaussian density with arbitrary mean and covariance matrix $\Sigma_{j_1,\ldots,j_n}$, then the temporal entropy (20) is given by

$$T_{j_1,\ldots,j_n} = \{1 + \log(2\pi)\}\, n + \frac{1}{2}\log\det\Sigma_{j_1,\ldots,j_n}. \tag{21}$$

The second result is the maximum entropy property of the Gaussian distribution:

among random vectors $\left(U_{j_1},\ldots,U_{j_n},V_{j_1},\ldots,V_{j_n}\right)$ having arbitrary density with fixed covariance matrix $\Sigma_{j_1,\ldots,j_n}$, the one that is normally distributed maximizes Shannon's entropy (20). The maximum of the entropy is given by (21).

We now consider the constraints on the a.c.v.f. (8) and search for the (strictly) stationary time series, with mean and pseudo-covariances null, that maximizes the temporal entropy.

Theorem 2 (Maximal Shannon's temporal temporal entropy distribution) *Consider the class of complex-valued and stationary time series $\{X_j\}_{j\in\mathbb{Z}}$ with mean null, variance σ^2, for some $\sigma \in (0,\infty)$, and pseudo-covariances null. Denote by ψ the a.c.v.f. of $\{X_j\}_{j\in\mathbb{Z}}$, $\nu = \operatorname{Re}\psi$ and $\xi = \operatorname{Im}\psi$.*

1. *If the a.c.v.f. ψ satisfies C_k given in (8) or in (10), thus $\psi(1) = \psi_1 = \nu_1 + \mathrm{i}\,\xi_1, \dots, \psi(k) = \psi_k = \nu_k + \mathrm{i}\,\xi_k$, then the time series $\{X_j\}_{j\in\mathbb{Z}}$ in the above class that maximizes Shannon's temporal entropy (20) with $n = k+1$ and $j_1 = 1, \dots, j_{k+1} = k+1$ is the one for which the corresponding double f.d.d. (1) with $j_1 = 1, \dots, j_{k+1} = k+1$ is Gaussian, with mean zero and with $2(k+1)\times 2(k+1)$ covariance matrix $\Sigma(k) = \Sigma^{1,\dots,1}$ given by (19) with*
$$\psi_{UU}(r) = \psi_{VV}(r) = \frac{\nu_r}{2} \ \ \textit{and} \ \ \psi_{UV}(r) = \psi_{VU}(-r) = -\frac{\xi_r}{2},$$
for $r = 1, \dots, k$.
2. *The corresponding value of the temporal entropy is given by*
$$T(k) = \{1 + \log(2\pi)\}\,(1 + k) + \frac{1}{2}\log\det\Sigma(k).$$

Proof *1.a.* This initial part of the proof shows that for any a.c.v.f. ψ, there exists a complex-valued Gaussian time series that is strictly stationary and radially symmetric. Let $n \geq 1$, $u_j, v_j \in \mathbb{R}$, $c_j = u_j - \mathrm{i}v_j$, for $j = 1, \dots, n$, let $j_1 < \dots < j_n \in \mathbb{Z}$, $\mathbf{u} = (u_1, \dots, u_n)^\top$, $\mathbf{v} = (v_1, \dots, v_n)^\top \in \mathbb{R}^n$ and define $q(\mathbf{u},\mathbf{v}) = 1/2\sum_{l=1}^n\sum_{m=1}^n c_l\overline{c_m}\psi(j_l - j_m)$. Then $q(\mathbf{u},\mathbf{v}) \geq 0$ implies
$$\begin{aligned} q(\mathbf{u},\mathbf{v}) &= \frac{1}{2}\sum_{l=1}^n\sum_{m=1}^n (u_l - \mathrm{i}v_l)(u_m + \mathrm{i}v_m)\{\nu(j_l - j_m) + \mathrm{i}\xi(j_l - j_m)\} \\ &= \frac{1}{2}\sum_{l=1}^n\sum_{m=1}^n (u_l u_m + v_l v_m)\nu(j_l - j_m) - (u_l v_m - v_l u_m)\xi(j_l - j_m). \qquad (22) \end{aligned}$$

Define $\mathbf{U} = (U_{j_1}, \dots, U_{j_n})^\top$ and $\mathbf{V} = (V_{j_1}, \dots, V_{j_n})^\top$. Assume $\left(\mathbf{U}^\top, \mathbf{V}^\top\right)$ normally distributed with mean zero and covariance matrix $\Sigma_{j_1,\dots,j_n}$, as in (19). A particular choice of $\Sigma_{j_1,\dots,j_n}$ can be obtained by setting
$$\varphi(\mathbf{u},\mathbf{v}) = \mathsf{E}\left[\exp\left\{\mathrm{i}\left(\mathbf{u}^\top, \mathbf{v}^\top\right)\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix}\right\}\right] = \exp\left\{-\frac{1}{2}q(\mathbf{u},\mathbf{v})\right\},$$
leading to
$$\left(\mathbf{u}^\top, \mathbf{v}^\top\right)\Sigma_{j_1,\dots,j_n}\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\end{pmatrix} = q(\mathbf{u},\mathbf{v}).$$

This, (22) and (19) yield

$$\begin{aligned} \mathsf{E}[U_{j_l}U_{j_m}] &= \frac{1}{2}\nu(j_l - j_m),\ \mathsf{E}[V_{j_l}V_{j_m}] = \frac{1}{2}\nu(j_l - j_m), \\ \mathsf{E}[U_{j_l}V_{j_m}] &= -\frac{1}{2}\xi(j_l - j_m),\ \mathsf{E}[V_{j_l}U_{j_m}] = \frac{1}{2}\xi(j_l - j_m) \end{aligned} \tag{23}$$

and therefore $\xi(j_l - j_m) = -\xi(j_m - j_l)$, for $l, m = 1, \ldots, n$. Define $\mathbf{X} = (X_{j_1}, \ldots, X_{j_n})^\top$, where $X_{j_l} = U_{j_l} + \mathrm{i}V_{j_l}$, for $l = 1, \ldots, n$. We obtain the covariance matrix

$$\begin{aligned} \mathrm{var}(\mathbf{X}) = \mathsf{E}\left[\mathbf{X}\overline{\mathbf{X}}^\top\right] &= \mathsf{E}\left[(\mathbf{U} + \mathrm{i}\mathbf{V})(\mathbf{U} - \mathrm{i}\mathbf{V})^\top\right] = \mathsf{E}\left[\mathbf{U}\mathbf{U}^\top + \mathbf{V}\mathbf{V}^\top + \mathrm{i}\left(\mathbf{V}\mathbf{U}^\top - \mathbf{U}\mathbf{V}^\top\right)\right] \\ &= \frac{1}{2}\left(\nu(j_l - j_m) + \nu(j_l - j_m) + \mathrm{i}\left[\xi(j_l - j_m) - \{-\xi(j_l - j_m)\}\right]\right)_{l,m=1,\ldots,n} \\ &= (\nu(j_l - j_m) + \mathrm{i}\xi(j_l - j_m))_{l,m=1,\ldots,n} = (\psi(j_l - j_m))_{l,m=1,\ldots,n} \end{aligned}$$

and the pseudo-covariance matrix

$$\begin{aligned} \mathsf{E}\left[\mathbf{X}\mathbf{X}^\top\right] &= \mathsf{E}\left[(\mathbf{U} + \mathrm{i}\mathbf{V})(\mathbf{U} + \mathrm{i}\mathbf{V})^\top\right] = \mathsf{E}\left[\mathbf{U}\mathbf{U}^\top - \mathbf{V}\mathbf{V}^\top + \mathrm{i}\left(\mathbf{V}\mathbf{U}^\top + \mathbf{U}\mathbf{V}^\top\right)\right] \\ &= \frac{1}{2}\left(\nu(j_l - j_m) - \nu(j_l - j_m) + \mathrm{i}\left[\xi(j_l - j_m) + \{-\xi(j_l - j_m)\}\right]\right)_{l,m=1,\ldots,n} \\ &= (0 + \mathrm{i}0)_{l,m=1,\ldots,n} = \mathbf{0}, \end{aligned}$$

as desired. We have thus established the existence of a complex-valued Gaussian time series $\{X_t\}_{t\in\mathbb{R}}$ that is strictly stationary and radially symmetric.

1.b. Consider $n = k + 1$ and $j_1 = 1, \ldots, j_{k+1} = k + 1$. Under C_k, $\mathrm{var}(\mathbf{X})$ is entirely determined: it is the $(k+1) \times (k+1)$ n.n.d. and Toeplitz matrix (9). The pseudo-covariance matrix and the mean vector are null and thus also determined. We know from (23) that, for $r = 1, \ldots, k+1$,

$$\psi_{UU}(r) = \psi_{UV}(r) = \frac{1}{2}\nu(r) = \frac{\nu_r}{2} \ \text{ and } \ \psi_{UV}(r) = \psi_{VU}(-r) = -\frac{1}{2}\xi(r) = -\frac{\xi_r}{2}.$$

So the covariance matrix of $(\mathbf{U}^\top, \mathbf{V}^\top)$ is entirely determined by C_k and it is the $2(k+1) \times 2(k+1)$ matrix $\Sigma_{1,\ldots,k+1} = \Sigma^{1,\ldots,1}$. Clearly, $\mathsf{E}\left[(\mathbf{U}^\top, \mathbf{V}^\top)\right] = 0$. The second information theoretic result for the Gaussian distribution, just above, concludes the proof Theorem 2.1.

2. The second information theoretic result for the Gaussian distribution, viz. (21), leads directly to the entropy formula in Theorem 2.2. □

So when $\{X_j\}_{j\in\mathbb{Z}}$ is the strictly stationary Gaussian-GvM time series, both spectral and temporal Shannon's entropies are maximized under the constraints C_k.

3 Some computational aspects

The following computational aspects are presented in this section: the computation of the integral functions of the GvM_2 time series in Section 3.1, the estimation of the GvM_k spectral distribution in Section 3.2 and the computation of the GvM_k spectral d.f. in Section 3.3.

3.1 Integral functions of the GvM_2 time series

Some series expansions for some of the integral functions appearing with the GvM_k spectral distribution are provided. Indeed, the results of Section 2 require the constants or integral functions $G_r^{(k)}$, for $r = 0, \ldots, k$, and $H_r^{(k)}$, for $r = 1, \ldots, k$. They are integrals over a bounded domain of smooth integrands and therefore numerical integration should perform well. Alternatively, one can evaluate these integral functions by series expansions. Gatto (2009) provides some of these expansions and in particular for $k = 2$, reported below. Define

$$\mathrm{ep}_r = \begin{cases} 1, \text{ if } r \text{ is even and positive,} \\ 0, \text{ otherwise.} \end{cases}$$

Let $\delta \in [0, \pi)$ and $\kappa_1, \kappa_2 \geq 0$. Then the following expansions hold for $r = 0, 1, \ldots$,

$$\begin{aligned} G_r^{(2)}(\delta, \kappa_1, \kappa_2) = {} & I_0(\kappa_1) I_{\frac{r}{2}}(\kappa_2) \cos r\delta \; \mathrm{ep}_r + I_0(\kappa_2) I_r(\kappa_1) \\ & + \sum_{j=1}^{\infty} \cos 2j\delta \; I_j(\kappa_2) \left\{ I_{2j+r}(\kappa_1) + I_{|2j-r|}(\kappa_1) \right\}, \end{aligned} \tag{24}$$

and

$$\begin{aligned} H_r^{(2)}(\delta, \kappa_1, \kappa_2) = & - I_0(\kappa_1) I_{\frac{r}{2}}(\kappa_2) \sin r\delta \; \mathrm{ep}_r \\ & + \sum_{j=1}^{\infty} \sin 2j\delta \; I_j(\kappa_2) \left\{ I_{2j+r}(\kappa_1) - I_{|2j-r|}(\kappa_1) \right\}. \end{aligned} \tag{25}$$

In these expansions we can see that G_r and H_r inherit the asymptotic behavior of the Bessel function I_r, for large r. It follows from Abramowitz and Stegun (1972, 9.6.10, p. 375) that $I_r(z) = (z/2)^r \{r\Gamma(r)\}^{-1} \left\{1 + \mathrm{O}\left(r^{-1}\right)\right\}$, as $r \to \infty$. This and the Stirling approximation yield $I_r(z) = (2\pi r)^{-1/2} \{\mathrm{e}z/(2r)\}^r \left\{1 + \mathrm{O}\left(r^{-1}\right)\right\}$, as $r \to \infty$. Hence I_r decreases rapidly to zero as r increases and this behavior is transmitted to G_r and H_r.

3.2 Estimation of the GvM spectral distribution

This section concerns the estimation problem: it reviews some classical results of spectral estimation and it presents an estimator to the parameters of the GvM spectral distribution.

A classical estimator of the spectral density is the periodogram, which is based on the discrete Fourier transform of the sample a.c.v.f., i.e. on

$$\Lambda_n(j) = \sum_{r=-(n-1)}^{n-1} \hat{\psi}_n(r) \mathrm{e}^{-\mathrm{i}\frac{2\pi j r}{n}},$$

for $j = \lfloor (n-1)/2 \rfloor, \ldots, -1, 1, \ldots, \lfloor n/2 \rfloor$, $\hat{\psi}_n$ being the sample a.c.v.f. (11), n the sample size and $\lfloor \cdot \rfloor$ the floor function. Because of its nonparametric nature, the periodogram is well-suited for detecting particular features, such as periodicities, that may not be identified by a parametric estimator. However, its irregular nature may not be desirable in some contexts and it does not result from an important optimality criterion.

One of the earliest studies on maximum entropy spectral distributions is Burg (1967), who considered $B(f_\sigma) = \int_{-\pi}^{\pi} \log f_\sigma(\theta) \mathrm{d}\theta$ as measure of entropy of the spectral density f_σ. This entropy is different than our adaptation of Shannon's entropy, viz. $S(f_\sigma)$ given in (4), but we can relate Burg's entropy to Kullback-Leibler's information by

$$B(f_\sigma) = 2\pi \left\{ \log \frac{\sigma^2}{2\pi} - \frac{1}{\sigma^2} I(u_\sigma | f_\sigma) \right\} = 2\pi \left\{ \log \frac{\sigma^2}{2\pi} - I(u_1 | f_1) \right\}. \tag{26}$$

Thus, maximizing Burg's entropy amounts to minimize the re-directed Kullback-Leibler information, instead of the usual Shannon's entropy. For real-valued time series, it turns out that the spectral density estimator that maximizes the entropy (26) subject to the constraints C_k in (8), with $\psi_r = \hat{\psi}_n(r)$, for $r = 1, \ldots, n-1$ and $k = n-1$, is equal to the autoregressive estimator of order k. This autoregressive estimator is given by the formula of the spectral density of the AR(k) model which has been fitted to the sample of n consecutive values of the time series. For more details refer e.g. to p. 365-366 of Brockwell and Davis (1991).

Estimators of the parameters of the GvM$_k$ spectral distribution can be obtained from the generalization of the trigonometric method of moments estimator for the GvM$_k$ circular distribution, which is introduced by Gatto (2008). This estimator is the circular version of the method of moments estimators. Consider the GvM$_k$ spectral distribution with unknown parameters $\mu_1, \ldots, \mu_k$ and $\kappa_1, \ldots, \kappa_k$, for some $k \leq n-1$. Consider the a.c.v.f. conditions C_k in which the spectral density g_σ is taken equal to the GvM$_k$ spectral density with variance σ^2, viz. σ^2 times the circular density (2), and in which the quantity $\psi_r \in \mathbb{C}$ is replaced by the sample a.c.v.f. at r, namely by $\hat{\psi}_n(r)$, for $r = 1, \ldots, k$, cf. (11). The resulting r-th equation can be re-expressed in a similar way to (14), precisely as

$$\begin{pmatrix} \operatorname{Re} \hat{\psi}_n(r) \\ \operatorname{Im} \hat{\psi}_n(r) \end{pmatrix} = \sigma^2 \mathbf{R}(r\mu_1) \begin{pmatrix} A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) \\ B_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) \end{pmatrix}, \tag{27}$$

for $r = 1, \ldots, k$. This gives a system of $2k$ real equations and $2k$ unknown real parameter. The values of $\mu_1, \delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k$ that solve this system of equations are the resulting estimators and they can be denoted $\hat{\mu}_1, \hat{\delta}_1, \ldots, \hat{\delta}_{k-1}, \hat{\kappa}_1, \ldots, \hat{\kappa}_k$. We now give two examples.

*Example 2 (*vM *spectrum)* When $k = 1$ we have the vM spectral distribution. Because $B_1^{(1)}(\kappa_1) = 0$, we obtain the two estimating equations

$$\begin{pmatrix} \operatorname{Re} \hat{\psi}_n(1) \\ \operatorname{Im} \hat{\psi}_n(1) \end{pmatrix} = \sigma^2 A_1^{(1)}(\kappa_1) \begin{pmatrix} \cos \mu_1 \\ \sin \mu_1 \end{pmatrix} = \sigma^2 \frac{I_1(\kappa_1)}{I_0(\kappa_1)} \begin{pmatrix} \cos \mu_1 \\ \sin \mu_1 \end{pmatrix},$$

with the two unknown parameters μ_1 and κ_1. The solutions are the estimators $\hat{\mu}_1$ and $\hat{\kappa}_1$. For $\kappa_1 > 0$, if $\mu_1 = 0$ is given, then we have axial symmetry about the null axis, thus real-valued time-series, and the two estimating equations reduce to the single equation

$$\frac{\hat{\psi}_n(1)}{\sigma^2} = A_1^{(1)}(\kappa_1), \tag{28}$$

whose solution is the estimator $\hat{\kappa}_1$. Amos (1974) showed that $A_1^{(1)}$ has positive derivative over $(0, \infty)$; for this derivative cf. e.g. p. 289 of Jammalamadaka and SenGupta (2001). It follows that $A_1^{(1)}$ is a strictly increasing and differentiable probability d.f. over $[0, \infty)$. So its inverse function is easily computed.

Example 3 (GvM_2 *spectrum)* When $k = 2$ we have the GvM_2 spectral distribution. The estimating equations (27) can be solved with the expansions of the constants given by (24) and (25). As previously mentioned, with $\kappa_1, \kappa_2 > 0$, the GvM_2 distribution is axially symmetric around the axis μ_1 iff $\delta_1 = \delta^{(1)} = 0$ or $\delta_1 = \delta^{(2)} = \pi/2$. With these values of δ_1 and with $\mu_1 = 0$, the axial symmetry is about the null axis and the time series becomes real-valued. We note that $B_r^{(2)}\left(\delta^{(j)}, \kappa_1, \kappa_2\right) = 0$, for $r = 1, 2$ and for the cases $j = 1, 2$. These equalities and $\operatorname{Im} \hat{\psi}_n(r) = 0$, for $r = 1, 2$, allow to simplify (27) to

$$\frac{\hat{\psi}_n(r)}{\sigma^2} = A_r^{(2)}\left(\delta^{(j)}, \kappa_1, \kappa_2\right),$$

for $r = 1, 2$ and for the two cases $j = 1, 2$. These estimating equations generalize the estimation equation (28) of the real-valued vM time series. For each one of these two cases, we have two equations and two unknown values, namely κ_1 and κ_2, giving the estimators $\hat{\kappa}_1$ and $\hat{\kappa}_2$.

We conclude this section on estimation by mentioning the test that the spectral density is a GvM one, precisely the null hypothesis

$$\mathrm{H}_0 : f_\sigma = f_\sigma^{(k)}(\cdot \mid \mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k),$$

where all parameters are specified. Anderson (1993) addresses this problem with the Cramér-von Mises and with the Kolmogorov-Smirnov criteria. Both criteria are based on

$$\sqrt{n}\{\hat{F}_n - F_\sigma^{(k)}(\cdot|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)\}, \tag{29}$$

where n is the sample size, $\hat{F}_n$ is the estimator of the spectral d.f. obtained by integration of the periodogram and where $F_\sigma^{(k)}(\cdot|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$ is the d.f. of $f_\sigma^{(k)}(\cdot|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)$. Under the main assumption that the time series admits the AR(∞) representation, the asymptotic distribution (29) of is obtained, of course not only for the GvM but for any specified spectral density. Note that Anderson (1993) considers real-valued time series only, but the results can be directly adapted to complex-valued time series. Section 3.3 provides a formula for the spectral GvM d.f. which appears in this goodness of fit problem.

3.3 GvM spectral distribution function

A formula for the GvM$_k$ spectral d.f. can be obtained in the form of a series as follows. Let $\psi_\sigma^{(k)}$ denote the a.c.v.f. and let $f_\sigma^{(k)}$ denote the spectral density of the GvM$_k$ time series with variance σ^2. It follows from $\mathrm{Re}\,\psi_\sigma^{(k)}(-r) = \mathrm{Re}\,\psi_\sigma^{(k)}(r)$ and $\mathrm{Im}\,\psi_\sigma^{(k)}(-r) = -\mathrm{Im}\,\psi_\sigma^{(k)}(r)$, for $r = 1, 2, \ldots$, and from (14) that

$$\begin{aligned} f_\sigma^{(k)}(\theta|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k) &= \frac{1}{2\pi}\sum_{r=-\infty}^{\infty}\psi_\sigma^{(k)}(r)\exp\{-\mathrm{i}r\theta\} \\ &= \frac{1}{2\pi}\left(1 + 2\sum_{r=1}^{\infty}\mathrm{Re}\,\psi_\sigma^{(k)}(r)\cos r\theta + \mathrm{Im}\,\psi_\sigma^{(k)}(r)\sin r\theta\right) \\ &= \frac{\sigma^2}{2\pi}\left\{1 + 2\sum_{r=1}^{\infty}(\cos r\theta, \sin r\theta)\mathbf{R}(r\mu_1)\begin{pmatrix} A_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k) \\ B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k)\end{pmatrix}\right\}, \end{aligned}$$

$\forall\theta \in (-\pi, \pi]$. Pointwise convergence is due to Dirichlet's theorem (cf. e.g. Pinkus and Zafrany, 1997, p. 47). The term by term integration of the Fourier series of a piecewise continuous function converges uniformly towards the integral of the original function (cf. e.g. Pinkus and Zafrany, 1997, p. 77). So the GvM$_k$ spectral d.f. admits the series representation

$$
\begin{aligned}
F_\sigma^{(k)}(\theta|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k) &= \int_{-\pi}^{\theta} f_\sigma(\alpha|\mu_1,\ldots,\mu_k,\kappa_1,\ldots,\kappa_k)\mathrm{d}\alpha \\
=&\frac{\sigma^2}{2\pi}\Bigg(\theta + 2\sum_{r=1}^{\infty}\frac{1}{r}\big[A_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k)\{\sin r(\theta-\mu_1)+(-1)^r\sin r\mu_1\} \\
&- B_r^{(k)}(\delta_1,\ldots,\delta_{k-1},\kappa_1,\ldots,\kappa_k)\{\cos r(\theta-\mu_1)-(-1)^r\cos r\mu_1\}\big]\Bigg),
\end{aligned}
$$

$\forall\theta \in (-\pi,\pi]$, where the convergence is uniform. The order of the r-th summand above is r times smaller than in the original Fourier series, so we can expect rapid convergence. When $k = 2$, we can use the series expansions (24) and (25) for the computation this d.f. This expansion of the spectral d.f. can be used in conjunction with (7) for the computation of the $\mathcal{L}_2$-norm of the increments of the spectral process of the GvM$_k$ time series. It may also be used in the testing problem (29).

4 Concluding remarks

This chapter presents an application of the GvM distribution of planar directional statistics to the analysis of stationary time series. The GvM and Gaussian-GvM time series are presented, together with some related results. Further results or methods could be developed. For example, the consistency of the estimators of the parameters of the GvM spectral distribution, as the sample of the time series augments, should be established. Another open research topic would concern simulation algorithms for the Gaussian-GvM time series. The GvM spectrum is motivated by theoretical considerations and we are aware that ad hoc spectra are often used in applied domains. For example, the Pierson-Moskowitz spectrum is widely used in naval construction for modelling of ocean waves. We refer to p. 315-316 of Lindgren (2012) for a list of commonly used spectra. (These spectra correspond to continuous time stationary models, but wrapping them around the unit circle gives the spectra of processes sampled at integer times, thus of time series.) We finally mention that the connection between spectral and circular distributions is well-known: the wrapped Cauchy circular distribution is the normalized spectrum of the AR(1) time series and the cardioid distribution is the normalized spectrum of the first order moving average (MA(1)) time series. This connection is exploited by Tanigichi et al. (2020), who construct new circular distributions starting from spectra. It is thus in the opposite direction that this chapter exploits this connection.

Acknowledgment The author is thankful to the editors and to two referees for various valuable comments and corrections.

References

1. Abramowitz, M., Stegun, I. E. (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications (9-th printing conform to the 10-th original printing).
2. Amos, D. E. (1974), “Computation of modified Bessel functions and their ratios”, *Mathematics of Computation*, 28, 239-251.
3. Anderson, T. W. (1993), “Goodness of fit tests for spectral distributions”, *The Annals of Statistics*, 830-847.
4. Asmussen, S., Glynn P. W. (2007), *Stochastic Simulation. Algorithms and Analysis*, Springer.
5. Astfalck, L., Cripps, E., Gosling, J., Hodkiewicz, M., Milne, I. (2018), “Expert elicitation of directional metocean parameters”, *Ocean Engineering*, 161, 268-276.
6. Bloomfield, P. (1973), “An exponential model for the spectrum of a scalar time series”, *Biometrika*, 60, 217-226.
7. Bogert, B. P., Healy, M. J. R., Tukey, J. W., Rosenblatt, M. (1963), “The quefrency analysis of time series for echoes: cepstrum, pseudoauto-covariance, cross-cepstrum and saphe cracking”, *Proceedings of the Symposium on Time Series Analysis*, editor Rosenblatt, Wiley and Sons, 209-243.
8. Brillinger, D. R. (1993), “The digital rainbow: some history and applications of numerical spectrum analysis”, *Canadian Journal of Statistics*, 21, 1-19.
9. Brockwell, P. J., Davis, R. A. (1991), *Time Series: Theory and Methods*, second edition, Springer.
10. Burg, J. P. (1967), “Maximum entropy spectral analysis”, unpublished presentation, 37-th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma.
11. Burg, J. P. (1978), “Maximum entropy spectral analysis”, *Modern Spectrum Analysis*, editor Childers. Wiley and Sons, 34-41.
12. Chatfield, C. (2013), *The Analysis of Time Series. Theory and Practice*, Springer.
13. Christmas, J. (2014), “Bayesian spectral analysis with Student-t noise”, *IEEE Transactions on Signal Processing*, 62, 2871-2878.
14. Cramér, H. (1942), “On harmonic analysis in certain functional spaces”, *Arkiv för Matematik, Astronomi och Fysik*, 28B, 12.
15. Cramér, H., Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes*, Wiley, New York.
16. Fisher, R. A. (1929), “Tests of significance in harmonic analysis”, *Proceedings of the Royal Society of London*, Series A, 125, 796, 54-59.
17. Gatto, R. (2008), “Some computational aspects of the generalized von Mises distribution”, *Statistics and Computing*, 18, 321-331.
18. Gatto, R. (2009), “Information theoretic results for circular distributions”, *Statistics*, 43, 409-421.
19. Gatto, R., Jammalamadaka, S. R. (2007), “The generalized von Mises distribution”, *Statistical Methodology*, 4, 341-353.
20. Gatto, R., Jammalamadaka, S. R. (2015), “Directional statistics: introduction”, *StatsRef: Statistics Reference Online*, editors Balakrishnan et al., Wiley and Sons, 1-8.
21. Gonella, J. (1972), “A rotary-component method for analysing meteorological and oceanographic vector time series”, *Deep-Sea Research*, 19, 833-846.
22. Huang, D. (1990), “On the maximal entropy property for ARMA processes and ARMA approximation”, *Advances in Applied Probability*, 612-626.
23. Jammalamadaka, S. R., SenGupta, A. (2001), *Topics in Circular Statistics*, World Scientific Press.
24. Kay, S. M., Marple, S. L. (1981), “Spectrum analysis: a modern perspective”, *Proceedings of the IEEE*, 69, 1380-1419.
25. Kim, S., SenGupta, A. (2013), “A three-parameter generalized von Mises distribution”, *Statistical Papers*, 54, 685-693.

26. Kullback, S. (1954), “Certain inequalities in information theory and the Cramér-Rao inequality”, *The Annals of Mathematical Statistics*, 25, 745-751.
27. Kullback, S., Leibler, R. A. (1951), “On information and sufficiency”, *The Annals of Mathematical Statistics*, 22, 79-86.
28. Laeri, F. (1990), “Some comments on maximum entropy spectral analysis of time series”, *Computers in Physics*, 4, 627-636.
29. Lin, Y., Dong, S. (2019), “Wave energy assessment based on trivariate distribution of significant wave height, mean period and direction”, *Applied Ocean Research*, 87, 47-63.
30. Lindgren, G. (2012), *Stationary Stochastic Processes: Theory and Applications*, Chapman and Hall/CRC.
31. Mardia, K. V., Jupp, P. E. (2000), *Directional Statistics*, Wiley.
32. Pewsey, A., García-Portugués, E. (2020), “Recent advances in directional statistics”, arXiv preprint.
33. Pinkus, A., Zafrany, S. (1997), *Fourier Series and Integral Transforms*, Cambridge University Press.
34. Rice, S. O. (1944), “Mathematical analysis of random noise”, *Bell Systems Technical Journal*, 23, 282-332.
35. Also in Rice (1954), *Selected Papers on Noise and Stochastic Processes*, editor Wax, Dover Publications, p. 133-295.
36. Rice, S. O. (1945), “Mathematical analysis of random noise”, *Bell Systems Technical Journal*, 24, 46-156.
37. Also in Rice (1954), “Mathematical analysis of random noise”, *Selected Papers on Noise and Stochastic Processes*, editor Wax, Dover Publications, p. 133-295.
38. Rowe, D. B. (2005), “Modeling both the magnitude and phase of complex-valued fMRI data”, *NeuroImage*, 25, 1310-1324.
39. Salvador, S., Gatto, R. (2021a), “Bayesian tests of symmetry for the generalized von Mises distribution”, preprint, Institute of Mathematical Statistics and Actuarial Science, University of Bern.
40. Salvador, S., Gatto, R. (2021b), “A Bayesian test of bimodality for the generalized von Mises distribution”, preprint, Institute of Mathematical Statistics and Actuarial Science, University of Bern.
41. Shannon, C. E. (1948), “A mathematical theory of communication”, *Bell System Technical Journal*, 27, 379-423, 623-656.
42. Taniguchi, M., Kato, S., Ogata, H., Pewsey, A. (2020), “Models for circular data from time series spectra”, *Journal of Time Series Analysis*, 41, 808-829.
43. Yaglom, A. M. (1962), *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall.
44. Zhang, L., Li, Q., Guo, Y., Yang, Z., Zhang, L. (2018), “An investigation of wind direction and speed in a featured wind farm using joint probability distribution methods”, *Sustainability*, 10.