Exponential expansions for approximation of probability distributions

Anna Maria Gambaro[1] [ID]

Abstract

This work analyses expansions of exponential form for approximating probability density functions, through the utilization of diverse orthogonal polynomial bases. Notably, exponential expansions ensure the maintenance of positive probabilities regardless of the degree of skewness and kurtosis inherent in the true density function. In particular, we introduce novel findings concerning the convergence of this series towards the true density function, employing mathematical tools of functional statistics. In particular, we show that the exponential expansion is a Fourier series of the true probability with respect to a given orthonormal basis of the so called Bayesian Hilbert space. Furthermore, we present a numerical technique for estimating the coefficients of the expansion, based on the first n exact moments of the corresponding true distribution. Finally, we provide numerical examples that effectively demonstrate the efficiency and straightforward implementability of our proposed approach.

Keywords Exponential expansion series · Convergence of density expansions · Bayes Hilbert space · Gram–Charlier Series

JEL Classification C02 · C14

1 Introduction

Polynomial-based expansions have been used in literature to approximate probability density functions on both closed and infinite domains, see for instance (Kendall et al. 1987; Schoutens 2000) and Provost (2005). This area of study often referred to as the *moments' problem* and traces its origins to the 19th century, with the pioneering work of Gram (1883). In the existing body of literature, researchers have concentrated on assessing the utility and efficacy of specific orthogonal polynomial

✉ Anna Maria Gambaro
  annamaria.gambaro@uniupo.it

1 Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone 18, 28100 Novara, Italy

Springer

bases for the estimation of various distributions. They typically opt for classical polynomial families as the foundation for expansion. These classical polynomials are often preferred due to their weight functions, which are the derivatives of the positive Borel measure that define the orthogonality properties of these polynomials. These weight functions bear a resemblance to canonical distributions. Among the orthogonal polynomials for which density approximations have been established in the literature, notable examples include: Hermite Polynomials, utilized in Edgeworth and A-type Gram–Charlier Expansions, Legendre Polynomials, Jacobi Polynomials and Laguerre Polynomials. These choices, as detailed by Gram (1883), Edgeworth (1905), Charlier (1914), Cramer (1946), Alexits (1961), and Kendall et al. (1987), are guided by the weight functions that resemble normal, uniform, beta, and gamma distributions, respectively. Importantly, these polynomials can be employed to derive moment-based expressions for specific density functions and are also recognized as eigenfunctions of a hypergeometric-type second-order differential equation, as noted in Schoutens (2000). More recently, by employing an A-type Gram–Charlier expansion methodology, Heston and Rossi (2016) show that Hermite polynomials cannot span fat-tailed distributions defined on the entire real line, commonly encountered in financial contexts. To overcome this deficiency, they introduce an analogous set of orthogonal polynomials based on the standardized logistic density.

However, the A-type Gram–Charlier and Edgeworth polynomial-based expansion methods proposed in the literature do not guarantee the positiveness of the truncated series, which therefore does not constitute a valid probability density function (PDF). To avoid this problem, Jondeau and Rockinger (2001), and Flamouris and Giamouridis (2002) have suggested to impose a non negative constraint in the estimation of the PDF approximated through the A-Gram–Charlier series. However, this constraint can lead to serious biases in the estimate of the probability density. To encompass the above drawbacks, Muscolino and Ricciardi (1999) and Rompolis and Tzavalis (2008) propose, in engineering and financial contexts respectively, to approximate PDF using the C-type Gram–Charlier expansion (Charlier 1928). This expansion is of an exponential form that always guarantees positive truncated probabilities for any degree of skewness and kurtosis of the true PDF. Moreover, Muscolino and Ricciardi (1999) propose a simple linear system to estimate the expansion coefficients, given the first n exact moments of the corresponding distributions.

In this work, we extend the existing literature in two main directions. Firstly, we highlight an interesting and promising link between the exponential expansion approach for approximating PDFs and the theory of *Bayesian Hilbert spaces*, extensively studied in functional statistics, see for instance (Egozcue et al. 2006; van den Boogaart et al. 2011, 2014) and Bongiorno and Goia (2016). Using the centered-log-ratio transformation, proposed for the first time by Atchinson for discrete probability distributions, Egozcue et al. (2006) and van den Boogaart et al. (2014) show that the subspace of square-log-integrable densities can be mapped to the Hilbert space of square integrable functions $\mathcal{L}^2$ with the ordinary notions of addition, scalar multiplication, and inner product. Our study establishes that an exponential expansion

corresponds to the Fourier series of the true density with respect to a given transformed orthogonal basis[1] of this Hilbert space, called in literature *Bayesian Hilbert space* or *Bayes space*. The proposed exponential expansion includes the C-type Gram–Charlier series as a particular case, adopting a specif orthogonal basis for the Bayes space, that is the (transformed) Hermite polynomial basis. However, it is important to highlight that when the density is not confined to a finite interval, the truncated Fourier series might not consistently correspond to a finite measure, as discussed in van den Boogaart et al. (2011, 2014). Consequently, its normalization as a valid probability measure could be impeded. A notable example is observed in the case of truncated Hermite series, which might lack integrability of the normalization constant for true densities with very fat tails. In this study, we address this issue by demonstrating that the choice of an appropriate basis of the Bayes space, such as the (transformed) Logistic polynomials, can mitigate this problem.

Secondly, we study the moment-based estimation of the coefficients of the Fourier series. In fact, it is often the case that the exact moments of a continuous-type statistic can be explicitly determined, while its density function either resists numerical evaluation or presents mathematical complexities. A notable example is found in stochastic volatility models widely utilized in finance, such as the Heston model. Alternatively, when employing real data, the estimation of moments up to a specific order may be more accurate compared to the direct estimation of the density using interpolation or kernel methods. Consequently, density approximations could be constructed based on the first *n* moments of corresponding distributions. Techniques proposed by Muscolino and Ricciardi (1999) and Rompolis and Tzavalis (2008) present a linear system to approximate the coefficients of the C-type Gram–Charlier expansion via these moments. Our second contribution lies in extending their approach to encompass any orthogonal polynomial basis of the *Bayes space*, while retaining the straightforward implementation, that is a simple linear system of equations.

A considerable body of literature focuses on inversion methods for the characteristic function or the Laplace transform of PDFs. Some of these methods, such as the COS method by Fang and Oosterlee (2009), as well as the Laplace transform methods by Cai et al. (2014) and Cai and Shi (2014), offer improved numerical accuracy, particularly in the tails of distributions. Our method provides two distinct advantages compared to numerical inversion methods. Firstly, our proposed approach provides an explicit analytical formula for the series coefficients and the approximated PDF. Secondly, our method does not require knowledge of the closed-form characteristic function or Laplace transform of the true PDF. Instead, to construct the N-th order analytical approximation of the PDF, our method requires only knowledge of the first $2N - 2$ moments of the true PDF.

Finally, we propose three illustrative numerical examples, employing Variance Gamma, Normal Inverse Gaussian and Heston densities as true probability distributions. Our results indicate that especially in the case of extreme values of skewness and kurtosis, the adoption of a Logistic polynomial basis makes the exponential expansion more robust with respect to the truncation of the PDF domain. This can enhance the

---

[1] Any orthogonal basis of $\mathcal{L}^2$ can be mapped to an orthogonal basis of the Bayes space using the inverse centered-log-ratio transformation.

convergence of the exponential series, improving the integrability of the normalization constant, a critical factor that renders any truncated series a legitimate and valid PDF.

The work is organized as follows. In Sect. 2, we shall begin with an overview of Bayesian Hilbert spaces and introducing the useful notation. In Sect. 3, we present our proposed exponential expansion for approximating a given PDF and we discuss the convergence of the series. The moment-based estimation of the expansion coefficients is presented in Sect. 4. Then, we illustrate the numerical examples in Sect. 5 and we end with concluding remarks in Sect. 6.

2 An overview of Bayesian Hilbert spaces

In this section, we offer a concise overview of the concept of Bayes space. For a more comprehensive illustration, readers are referred to the works of Egozcue et al. (2006), van den Boogaart et al. (2011, 2014) and Bongiorno et al. (2014).

We define the following set of functions:

$$
\mathcal{B}^2(I, \nu) := \left\{ f(x) = c \, \exp(\phi(x)) \, | 0 < c < \infty, \int_I \phi(x)^2 \nu(x) dx < \infty \right\}, \quad (1)
$$

where $I \subseteq \mathbb{R}$ is the domain of the function $f \in \mathcal{B}^2(I, \nu)$. For simplicity of exposition, we assume that the strictly positive weighting function $\nu(x)$ is a PDF on I, although this is not necessary. Moreover, each member of $\mathcal{B}^2(I, \nu)$ is the exponential of a function from $\mathcal{L}^2(I, \nu)$, that is the set of square integrable functions, under our chosen reference weight ν. In the rest of this section, we simplify the notation as $\mathcal{B}^2 = \mathcal{B}^2(I, \nu)$ and $\mathcal{L}^2 = \mathcal{L}^2(I, \nu)$, implicitly considering I and ν as the domain and the reference weighting function, unless otherwise specified.

We can map a member $f \in \mathcal{B}^2$ to its image $\phi \in \mathcal{L}^2$ applying the centered log-ratio (clr) transformation, defined in the compositional data analysis context by Aitchison (1986), that is

$$
\phi(x) = clr(f)(x) = \ln f(x) - \int_I \ln f(x) \, \nu(x) dx, \quad (2)
$$

the clr transformation is a shift of the logarithm of the function f, such that the zero integral constraint is satisfied, that is $\int_I clr(f)(x) \, \nu(x) dx = 0$. We say that two members of $\mathcal{B}^2$, $f_1(x) = c_1 \, \exp(\phi_1(x))$ and $f_2(x) = c_2 \, \exp(\phi_2(x))$, are equivalent (equal) if and only if $\phi_1(x) = \phi_2(x)$ for any $x \in I$; in other words, the normalization constants, c_1 and c_2, need not be the same and $\mathcal{B}^2$ is a set of equivalent classes of functions. Under this condition, the clr transform is an isomorphism, and then $\mathcal{B}^2$ is isomorphic to $\mathcal{L}^2$. In fact, as proved in van den Boogaart et al. (2014), the clr transform is linear, surjective and invertible with inverse $clr^{-1}(\phi) = \exp(\phi)$.

Then, we introduce addition and multiplication by real scalar. The set $B^2(I, \nu)$ equipped with these operations is established as a vector space, termed a Bayesian linear space, over the field $\mathbb{R}$. We define vector addition as in van den Boogaart et al.

(2011), $\oplus: \mathcal{B}^2 \times \mathcal{B}^2 \to \mathcal{B}^2$, between two elements $f_1, f_2 \in \mathcal{B}^2$ to be:

$$(f_1 \oplus f_2)(x) = f_1(x)f_2(x) = c_1 c_2 \exp(\phi_1(x) + \phi_2(x)) \in \mathcal{B}^2, \text{ for any } x \in I,$$

and likewise scalar multiplication, $\cdot : \mathbb{R} \times \mathcal{B}^2 \to \mathcal{B}^2$, of $f \in \mathcal{B}^2$ by $\alpha \in \mathbb{R}$ to be:

$$(\alpha \cdot f)(x) = (f(x))^\alpha = c^\alpha \exp(\alpha \phi(x)) \in \mathcal{B}^2, \text{ for any } x \in I.$$

Notably, the zero vector is simply any constant function, $c \exp(0)$. Vector subtraction is defined in the usual way, $f_1 \ominus f_2 = f_1 \oplus (-1) \cdot f_2$, that is

$$(f_1 \ominus f_2)(x) = \frac{c_1}{c_2} \exp(\phi_1(x) - \phi_2(x)) \in \mathcal{B}^2, \text{ for any } x \in I.$$

We note that subtraction is equivalent to the Radon-Nikodym derivative operation. We endow the vector space with an inner product defined as

$$\langle f_1, f_2 \rangle = \frac{1}{2} \int_I \int_I \ln \frac{f_1(x)}{f_1(y)} \ln \frac{f_2(x)}{f_2(y)} \nu(x)\nu(y)dxdy, \tag{3}$$

where $f_1, f_2 \in \mathcal{B}^2(I, \nu)$. We note that the normalization constants, c_1 and c_2, associated with f_1 and $f_2 \in \mathcal{B}^2$ play no role in the definition of the inner product. Since ν is a valid PDF, we can also write the inner product in (3) as

$$\langle f_1, f_2 \rangle = \mathbb{E}^\nu[\ln f_1 \ln f_2] - \mathbb{E}^\nu[\ln f_1]\mathbb{E}^\nu[\ln f_2],$$

where $\mathbb{E}^\nu[]$ represents the expected value under the probability measure induced by ν. Following (van den Boogaart et al. 2014), then, we can claim that $\mathcal{B}^2$ with inner product (3) forms a separable Hilbert space, which is referred to as a Bayesian Hilbert space. We shall sometimes briefly refer to it as a Bayesian space or Bayes space. The norm of $f \in \mathcal{B}^2$ can be taken as $||f|| = \sqrt{\langle f, f \rangle}$. Accordingly, we can define the distance, called Atchinson distance, between two members of $\mathcal{B}^2$, f and g, as

$$d_A(f, g) = ||f \ominus g|| = \sqrt{\langle f \ominus g, f \ominus g \rangle}, \tag{4}$$

which induces a metric on Bayes space. Moreover, since we add a metric to the Bayes space, Egozcue et al. (2006) and van den Boogaart et al. (2014) prove that clr is an isometry that preserves distances passing from the Bayes space $\mathcal{B}^2$ to the space of square integrable function $\mathcal{L}^2$:

$$d_A(f, g) = d_2(clr(f), clr(g)) = \int_I (clr(f)(x) - clr(g)(x))^2 \nu(x)dx, \tag{5}$$

where d_2 is the classical Euclidean distance between two members of $\mathcal{L}^2$. Then, $\mathcal{B}^2$ and $\mathcal{L}^2$ are isometrically isomorphic normed vector spaces and they can be identified with each other.

Finally, it is important to asses the relation between the members of the space $\mathcal{B}^2$ and the set of PDFs defined on I. Not all the possible PDFs are included in $\mathcal{B}^2$. In fact, to be included in $\mathcal{B}^2$ the PDF should be strictly positive, that is $p(x) > 0$ for any $x \in I$, and log-squared integrable with respect to the reference weighting function, that is $\int_I \ln p(x)^2 v(x) dx < \infty$. Moreover, in case of infinite domain I, $\mathcal{B}^2$ includes also functions that cannot be normalized, that corresponds to an infinite measures on I. We define $\mathcal{B}_P^2 \subseteq \mathcal{B}^2$ as the subset of elements in $\mathcal{B}^2$ that can be identified with a finite measure on I, that is

$$\mathcal{B}_P^2 := \left\{ f \in \mathcal{B}^2 \mid \int_I f(x) dx < \infty \right\}.$$

Then, we can apply the normalization operator $\mathcal{P} : \mathcal{B}_P^2 \to \mathcal{B}_P^2$

$$p(x) = \mathcal{P}(f)(x) = \frac{f(x)}{\int_I f(x) dx}, \tag{6}$$

then $p \in \mathcal{B}^2$ is a valid PDF on I and is equivalent to $f \in \mathcal{B}^2$.

3 Exponential expansion series for approximating PDFs

In this section, we present our proposed exponential expansion for approximating a strictly positive PDFs defined on a domain I. Let $\{\psi_j\}_{j \geq 0}$ be an orthonormal basis for $\mathcal{L}^2(I, v)$ space,[2] that is the space of squared integrable function with respect to the reference PDF v, as introduced in Sect. 2. For $I = \mathbb{R}$, examples of orthonormal basis are the (normalized) Hermite or Logistic polynomials, with weighting functions the standard Gaussian or the logistic PDFs, respectively. Then, we approximate the (strictly positive) PDF $p(x)$ on I by a truncated expansion series of the following form:

$$p_N(x) := C_0 \exp\left(\sum_{j=1}^N c_j \psi_j(x) \right), \tag{7}$$

for $N = 1, 2, 3, \ldots$ and with normalization constant

$$C_0 = \left(\int_I \exp\left(\sum_{j=1}^N c_j \psi_j(x) \right) dx \right)^{-1}. \tag{8}$$

---

[2] To be an orthonormal basis of $\mathcal{L}^2(I, v)$, the functions $\{\psi_j\}_{j \geq 0}$ must be square integrable, that is $\int_I (\psi_j(x))^2 v(x) dx < \infty$ for $j = 0, 1, 2, \ldots$, and orthonormal, that is $\int_I \psi_j(x) \psi_k(x) v(x) dx = \delta_{j,k}$ for $j, k = 0, 1, 2, \ldots$, where $\delta_{j,k}$ is the Kronecker delta function. Finally, the system $\{\psi_j\}_{j \geq 0}$ must be complete in $\mathcal{L}^2(I, v)$.

Proposition 1 *Let us assume that $p \in \mathcal{B}^2(I, v)$. Then, the truncated series p_N defined in* (7) *with coefficients*

$$c_j = \int_I clr(p)(x) \, \psi_j(x) \, v(x) dx, \tag{9}$$

converges to p in the Bayes space, that is the Atchinson distance defined in (4) *converges to zero,*

$$d_A(p, \hat{p}_N) \to 0, \text{ for } N \to \infty. \tag{10}$$

Proof Let $\{\psi_j\}_{j \geq 0}$ be an orthonormal basis for $\mathcal{L}^2$ with $\psi_0(x)$ a constant function, then $\{g_j\}_{j \geq 1}$, with $g_j = (clr)^{-1}(\psi_j) = \exp(\psi_j)$ is an orthonormal basis for $\mathcal{B}^2$. In fact, orthogonality and completeness are guaranteed by the clr-isometry. Given a basis $\{g_j\}_{j \geq 1}$ of $\mathcal{B}^2$ any $p \in \mathcal{B}^2(I, v)$ can be represented by its Fourier coefficients in the basis. The truncated Fourier representation at N terms is

$$p_N = \bigoplus_{j=1}^{N} c_j \cdot g_j = \prod_{j=1}^{N} \exp(\psi_j)^{c_j} = \exp\left(\sum_{j=1}^{N} c_j \psi_j\right), \text{ with } c_j = \langle p, g_j \rangle,$$

where the addition operation $\oplus$ and the multiplication by scalar $\cdot$ in Bayes space are defined in Sect. 2. We recall that the function p_N defined above is equivalent to the one defined in (7) regardless of the constant C_0. Moreover, by the clr-isometry, the coefficients c_j defined in (9) are exactly the Fourier coefficients, that is $c_j = \langle p, g_j \rangle$. Since $\mathcal{B}^2$ is isometrically isomorphic to $\mathcal{L}^2$, then $clr(p_N)$ converges to $clr(p) = \sum_{j=1}^{N} c_j \psi_j$ in the $\mathcal{L}^2$ sense, that is

$$d_A(p, p_N) = d_2\left(clr(p), \sum_{j=1}^{N} c_j \psi_j\right) \to 0, \text{ for } N \to \infty.$$

$\square$

The rate of convergence of orthogonal polynomial expansion for square integrable functions is discussed in literature, see for instance (Schwartz 1967) and Walter (1977) for Hermite polynomials and Efromovich (2010) and Abilov et al. (2009) and references therein for a more general classes of orthogonal polynomial basis. In particular, the rate of convergence of the polynomial expansion is related to the differentiability order of the function. Proposition 1 establishes that the convergence rate for $clr(p_N)$ in $\mathcal{L}^2$ remains valid for p_N in the Hilbert space $\mathcal{B}^2$, equipped with the Atchinson distance as defined in Eq. (4) and the orthogonal basis obtained as the inverse transform clr^{-1} of the polynomial basis.

Using the normalization operator $\mathcal{P}$ defined in (6), we can rewrite the truncated series in (7) as

$$p_N(x) = \mathcal{P}(f_N)(x),$$

$$f_N(x) = \exp\left(\sum_{j=1}^{N} c_j\, \psi_j(x)\right),\tag{11}$$

where $f_N \in \mathcal{B}_P^2$. In the case of infinite domain I, although the Proposition 1 remains valid, it is not possible to guarantee a priori that for each N, f_N corresponds to a finite measure on I, which implies that $p_N(x)$ is a valid PDF. Indeed, in Bayes spaces, p_N represents a class of equivalent functions independently of the constant C_0. This problem can be solved in practice by truncating the I domain of the referenced PDF. In Sect. 5, we present a heuristic rule of thumb to strike a balance between domain truncation and accuracy in replicating the tails of the actual distribution. Moreover, we show that adopting an appropriate orthonormal basis of the Bayes space, such as the (transformed) logistic polynomials, the exponential truncated series is more robust with respect to an enlargement of the domain I, reproducing better the tails of the distribution.

The overview of Bayes Hilbert spaces in Sect. 2, along with the convergence proof in this section, is presented in a one-dimensional setting for convenience. However, the exposition, and particularly Proposition 1, can be readily generalized to a multidimensional setting. In fact, if $\{\psi_{1,j}\}_{j\geq 0}$ and $\{\psi_{2,j}\}_{j\geq 0}$ are orthonormal bases of $\mathcal{L}^2(I_1 \subseteq \mathbb{R}, \nu_1)$ and $\mathcal{L}^2(I_2 \subseteq \mathbb{R}, \nu_2)$, respectively, then $\{\psi_{n,m}\}_{n\geq 0, m\geq 0}$ with $\psi_{n,m}(x_1, x_2) := \psi_{1,n}(x_1)\psi_{2,m}(x_2)$ for $(x_1, x_2) \in I_1 \times I_2$ is an orthonormal basis for $\mathcal{L}^2(I_1 \times I_2 \subseteq \mathbb{R}^2, \nu_1\,\nu_2)$. The above procedure to build the orthonormal basis in two dimensions can be iterated to achieve any desired dimensionality. Then in a multidimensional setting the truncated expansion series of the pdf $p(\mathbf{x})$ with $\mathbf{x} \in I \subseteq \mathbb{R}^d$ is given by

$$p^N(\mathbf{x}) = \exp\left(\sum_{j_1=1}^{N} \cdots \sum_{j_d=1}^{N} c_{j_1,\dots,j_d}\, \psi_{j_1,\dots,j_d}(\mathbf{x})\right),$$

with

$$c_{j_1,\dots,j_d} = \int_I clr(p)(\mathbf{x})\psi_{j_1,\dots,j_d}(\mathbf{x})\prod_{i=1}^{d} \nu_i(x_i)d\mathbf{x}.\tag{12}$$

4 A moment-based estimation of coefficients

In this section, we present a moment-based technique to estimate the coefficients of the Fourier series c_j, defined in Proposition 1. In fact, it is often the case that the exact moments of a continuous-type statistic can be explicitly determined, while its density function either resists numerical evaluation or presents mathematical complexities. A notable example is found in stochastic volatility models widely utilized in finance, such as the Heston model. Consequently, density approximations could be constructed based on the first n moments of the corresponding distribution. Techniques proposed

by Muscolino and Ricciardi (1999) and Rompolis and Tzavalis (2008) present a linear system to approximate the coefficients of the C-type Gram–Charlier expansion via these moments. We extend their approach to encompass any orthogonal polynomial basis of the Bayes space, while retaining the straightforward implementation, that is a simple linear system of equations.

Let $\{h_j\}_{j \geq 0}$ be the set of (normalized) Hermite polynomials, that forms an orthonormal basis for $\mathcal{L}^2(\mathbb{R}, \omega)$, where ω is the Gaussian density function with mean $m \in \mathbb{R}$ and standard deviation $\sigma > 0$, that is $\omega(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$. In particular, $h_j(x^*) = \frac{(-1)^j}{\sqrt{j!}} \frac{1}{\eta(x^*)} \frac{d^j \eta(x^*)}{d(x^*)^j}$, with $x^* = \frac{x-m}{\sigma}$ and η is the standard normal PDF. Then, let assume that $\{\varphi_j(x^*)\}_{j \geq 0}$ is a set of polynomial functions that represents an orthonormal basis of $\mathcal{L}^2(\mathbb{R}, \nu)$, with ν a valid positive weighting function on $\mathbb{R}$.

By Ackerer and Filipović (2020), any polynomial φ_n with $n = 0, 1, 2, \ldots$ on $\mathbb{R}$ can be expressed as a linear combination of Hermite polynomials in the following way

$$\varphi_n(x^*) = \sum_{j=0}^{n} q_{n,j} h_j(x^*), \tag{13}$$

where the vector $q_{n,j}$ for $j = 0, 1, 2, ..$ is the representation of φ_n in the basis $\{h_j\}_{j \geq 0}$. Let $H^n \in \mathbb{R}^{(n+1) \times (n+1)}$ denotes the matrix, whose element (i, j) is given by the coefficient in front of the monomial x^{j-1} in the $(i-1)^{th}$ Hermite polynomial h_{i-1} for $i, j = 1, 2, ..., n+1$. Define similarly the matrix $B^n \in \mathbb{R}^{(n+1) \times (n+1)}$, with respect to the polynomial basis $\{\varphi_j\}_{j \geq 0}$. The matrices H^n and B^n are widely recognized in the literature, particularly for the polynomial bases most commonly employed. Then, the coefficients $q_{n,j}$ can be obtained solving the following linear system

$$B^n = H^n Q^n, \tag{14}$$

where Q^n is the matrix, whose elements are $q_{i-1,j-1}$ for $i, j = 1, 2, ..., n+1$. We note that B^n, H^n and Q^n are upper triangular matrices, then the system (14) is a triangular system of equations that can be solved very efficiently.

By taking into account of the property of the derivative of the Hermite polynomials and the change of basis formula (13), we obtain the following linear system for approximating the coefficients $\{c_j\}_{j=1}^{N}$ of the expansion defined in (7),

$$A^N \hat{c}^N = b^N,$$
$$A_{i,n}^N = \sum_{j=0}^{n} \sqrt{j} \, q_{n,j} \, \tilde{A}_{i,j}^N, \text{ for } i, n = 1, 2, ..., N. \tag{15}$$

The matrix $\tilde{A}^N$ and the vector b^N are obtained as

$$\tilde{A}_{i,j}^N = \sum_{k=0}^{i+j-2} \frac{1}{k!} \Delta_{i-1,j-1,k} \, m_k^h,$$

with

$$\Delta_{p,q,r} = \prod_{a=p,q,r} \frac{\Gamma(a+1)}{\Gamma(b-a+1)}, \quad \text{if } p+q+r \text{ is even}, \ b = (p+q+r)/2 \geq p, q, r$$

$$\Delta_{p,q,r} = 0, \quad \text{otherwise}$$

and

$$b_i^N = -\sqrt{i-1} \, m_{i-1}^h, \ \text{for } i = 1, 2, ..., N$$

where m_k^h is the exact $k-th$ normalized Hermite moment of $p(x)$, that is

$$m_k^h = \int_I h_k(x^*) \, p(x) dx.$$

The (non-normalized) $k-th$ Hermite moment, that is $\sqrt{k!} \, m_k^h$, can be calculated knowing the first k exact moments of $p(x)$, see (Rompolis and Tzavalis 2008) equations (2) and (3). For the detailed derivation of the matrix $\tilde{A}^N$ and the vector b^N in (15), we refer the interested reader to Muscolino and Ricciardi (1999).

Finally, the true PDF $p(x)$ is approximated by the following exponential truncated series

$$\hat{p}_n(x) = \hat{C}_0 \exp\left(\sum_{j=1}^{N} \hat{c}_j^N \varphi_j(x^*)\right),$$

$$\hat{C}_0 = \left(\int_I \exp\left(\sum_{j=1}^{N} \hat{c}_j^N \varphi_j(x^*)\right) dx\right)^{-1}, \tag{16}$$

where $x^* = \frac{x-m_1}{\sigma}$, $\sigma = \sqrt{m_2 - m_1^2}$ and m_1, m_2 are the first two exact moments of $p(x)$.

The extension to a multidimensional setting of the method described above for the analytical estimation of the series coefficients $\hat{c}^N$ is not trivial. In Appendix B, we discuss the extension of the method for PDFs in two variables, using the product of Hermite polynomials.

5 Numerical examples

In this section, we apply our proposed approximation (16) to three PDFs: Variance Gamma (VG), Normal Inverse Gaussian (NIG) and Heston models. The three examples represents different levels of skewness and excess kurtosis, reported in Table 1. For the Variance Gamma model example, we use the parameters defined in Heston and Rossi (2016), that gives skewness equal to zero and excess kurtosis of 2. For the Heston model, we use parameters defined in Rompolis and Tzavalis (2008), that produce a

Table 1 Skewness and excess kurtosis of the three PDFs

Model	Skewness	Excess kurtosis
VG	0	2
NIG	0.2	1
Heston	−1.2	2.5

negative skewness of −1.2 and excess kurtosis of 2.5. For the NIG model, we choose the parameters such that the values of skewness and kurtosis are intermediate between VG and Heston, the parameters values and the specification for the NIG model are reported in Appendix A. As orthonormal basis for the Bayes Space $\mathcal{B}^2$, we use the transformed[3] Hermite or Logistic polynomials.

In particular, we study the convergence of the approximated PDF $\hat{p}_N$ defined in (16) to the true PDF, as the number of terms N of the truncated series increases. As a benchmark for the numerical estimation of the true PDF, we use the COS method proposed in Fang and Oosterlee (2009), with a grid size equal 2^{12}. We emphasize that, while the COS method offers enhanced numerical accuracy in PDF estimation, a significant advantage of our proposed approach lies in providing an explicit analytical formula for the coefficients $\{c_j\}_{j \geq 1}$ and the approximated PDF.

Firstly, we show the convergence of the estimated coefficient $\hat{c}^N$ obtained solving the system (15), to the exact coefficients of the Fourier series $\{c_j\}_{j \geq 1}$ defined in (9). The benchmark value for the coefficients $\{c_j\}_{j \geq 1}$ is obtained by numerically calculating the integral in Eq. (9), with $clr(p)(x)$ estimated via COS method. The series of coefficients $\{c_j\}_{j \geq 1}$ is square summable, that is, $\sum_{j=1}^{\infty} c_j^2 < \infty$, then we verify that the following distance converges to zero as N increases,

$$d_2(\hat{c}^N, c) = \sum_{j=1}^{N} \left(\hat{c}_j^N - c_j \right)^2,\tag{17}$$

with $\hat{c}_j^N = 0$ for $j > N$. Figures 1, 2 and 3 illustrate the plots of $d_2(\hat{c}^N, c)$ as the number of terms N increases. Results show a very good convergence of both the Hermite and the Logistic coefficients for the VG and the NIG model, while the estimated coefficients struggles to converge in the case of the Heston model, which is the example with the most extreme values of skewness and kurtosis.

The convergence of the approximation $\hat{p}_N$ to the benchmark PDFs is illustrated using three type of distances. Firstly, we adopt the Atchinson distance defined in (4) and used in Proposition 1, that is

$$d_A(\hat{p}_N, p) = \int_I \left(clr(\hat{p}_N)(x) - clr(p)(x) \right)^2 \nu(x)dx.\tag{18}$$

---

[3] As detailed in the proof of Proposition 1, an orthonormal basis for $\mathcal{B}^2$ is obtained applying the inverse clr transform to a basis for $\mathcal{L}^2$.
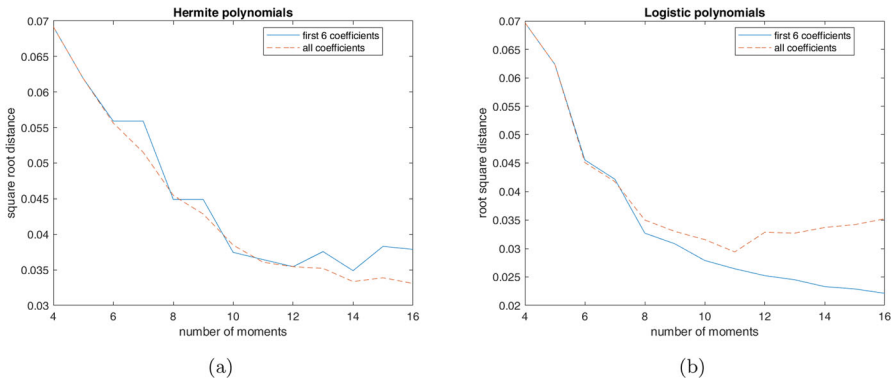
Fig. 1 These figures show the convergence of the sequence of coefficients $\hat{c}^N$ of the VG model, using distance defined in (17), for the Hermite basis, **a** on the left, and for Logistic basis, **b** on the right
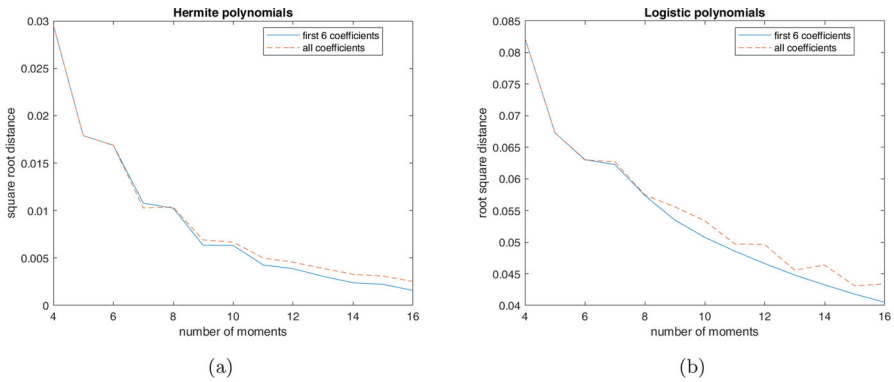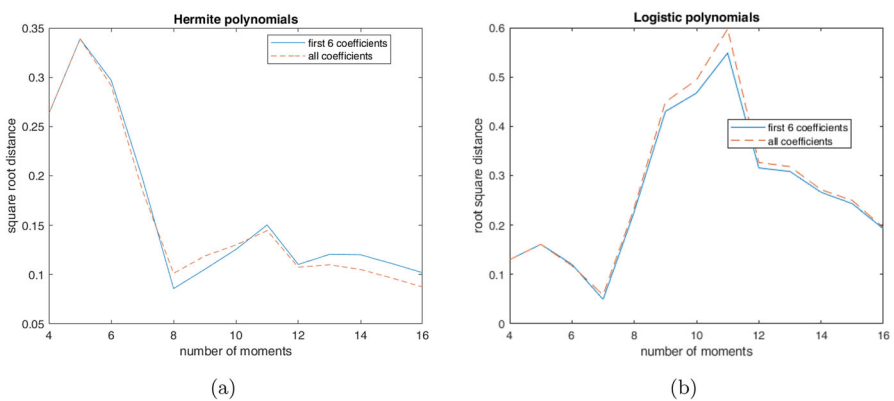

Fig. 2 These figures show the convergence of the sequence of coefficients $\hat{c}^N$ of the NIG model, using distance defined in (17), for the Hermite basis, **a** on the left, and for Logistic basis, **b** on the right

Fig. 3 These figures show the convergence of the sequence of coefficients $\hat{c}^N$ of the Heston model, using distance defined in (17), for the Hermite basis, **a** on the left, and for Logistic basis, **b** on the right

Then, we use also the $\mathcal{L}^2$ distance between the logarithms of the two pdfs, that is

$$d_2(\ln \hat{p}_N, \ln p) = \int_I \left(\ln \hat{p}_N(x) - \ln p(x)\right)^2 dx. \tag{19}$$

This second type of distance is adopted in order to compare the distances obtained by using different reference PDFs ν, that are the Gaussian and Logistic. Furthermore, we check the convergence also using the $\mathcal{L}^1$ and the $\mathcal{L}^2$ distances directly on the approximated and true PDFs, that are

$$d_1(\hat{p}_N, p) = \int_I |p_N(x) - p(x)| \, dx, \tag{20}$$

and

$$d_2(\hat{p}_N, p) = \int_I \left(\hat{p}_N(x) - p(x)\right)^2 dx. \tag{21}$$

To guarantee the convergence of the approximation towards the true PDF, it is necessary that the two assumptions on the true PDF in Proposition 1 are respected. Firstly, that $p(x) > 0$ holds strictly for all $x \in I$, and secondly, that its logarithm is square integrable, i.e., $\int_I \ln p(x)^2 \nu(x) dx < \infty$. For all the three considered examples, these two assumptions are numerically achieved by the truncation of the infinite domain I.[4] Particularly delicate is the strict positivity assumption. In functional statistics, a common heuristic rule of thumb is that the absolute value of the clr of the true PDF must not be greater than 10, i.e., $|clr(p)(x)| < 10$ for all $x \in I$. This rule implies that, for negative values of $\ln p(x)$, $p(x) > 5 \times 10^{-5}$ for all $x \in I$. This establishes the tolerance of the proposed method towards the assumption of strict positivity. Then, we compare the truncated range I achieved using the rule just delineated with the range derived from the utilization of the first four exact cumulants, k_i for $i = 1, 2, 3, 4$, as adopted in Fang and Oosterlee (2009), that is

$$I = [k_1 - L\sqrt{k_2 + \sqrt{k_4}}, \ k_1 + L\sqrt{k_2 + \sqrt{k_4}}], \tag{22}$$

with $L > 0$, in general, the parameter L is fixed greater or equal to 3. In Fig. 4, we compare the clr of the three standardized PDFs in an interval I, fixed as in (22) with $L = 4$. We note that for the Heston model, that is the example with the most extreme values of skewness and kurtosis, the clr exceeds the tolerance value of -10.

Then, in the case of the VG and NIG models, the convergence analysis of both Hermite and Logistic series is conducted utilizing a domain I defined as in Eq. (22) with $L = 4$. For the VG and NIG models, the plots of different distances outlined in Eqs. (18), (19), (20), and (21) are presented in Figs. 5 and 6. These figures illustrate the behavior of the mentioned distances as the number of terms N increases. In case of Heston model, we adopt and compare two procedures. Firstly, we truncate the domain I in such a way that the clr does not exceed the tolerance, that is $clr(p)(x) > -10$

---

[4] We notably remark that the estimation method for the coefficients of series $\hat{c}^N$ does not depend on the truncated domain I.

Fig. 4 The figure reports the logarithm of the three considered PDFs: Variance Gamma, Normal Inverse Gaussian and Heston. The plot is obtained using the COS method described in Fang and Oosterlee (2009)
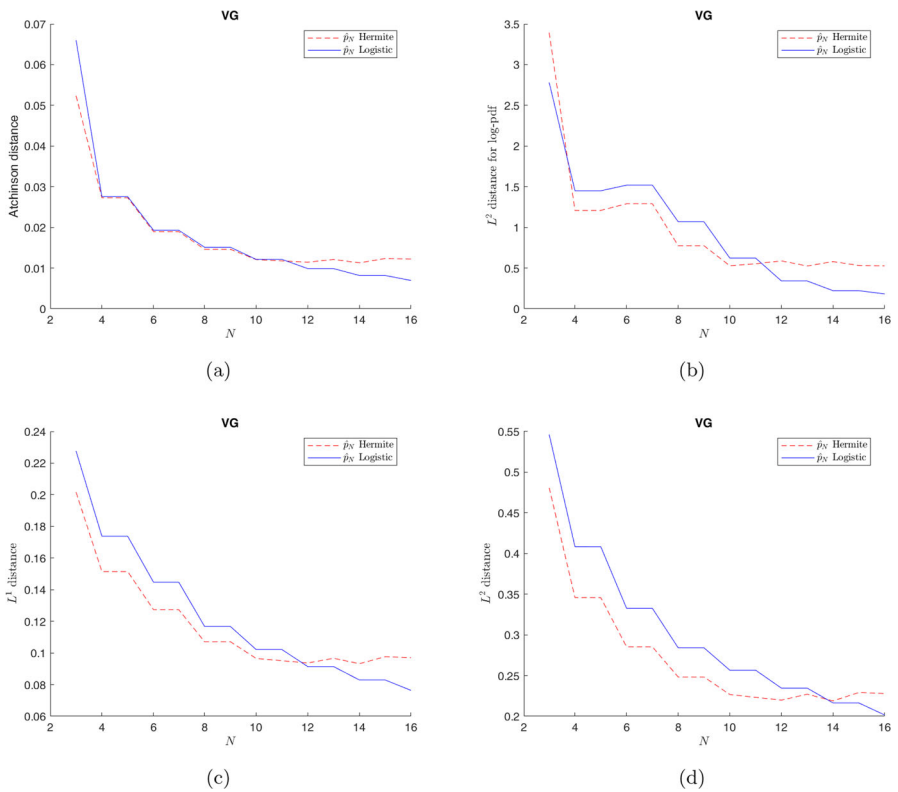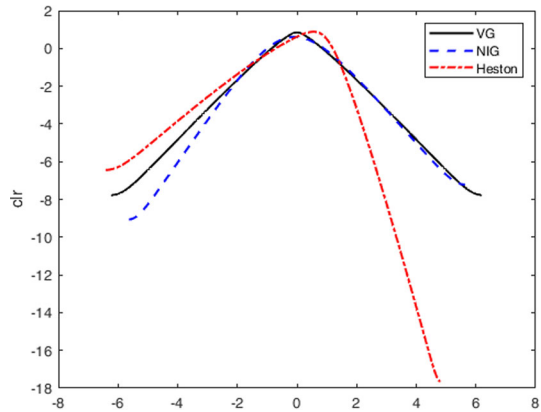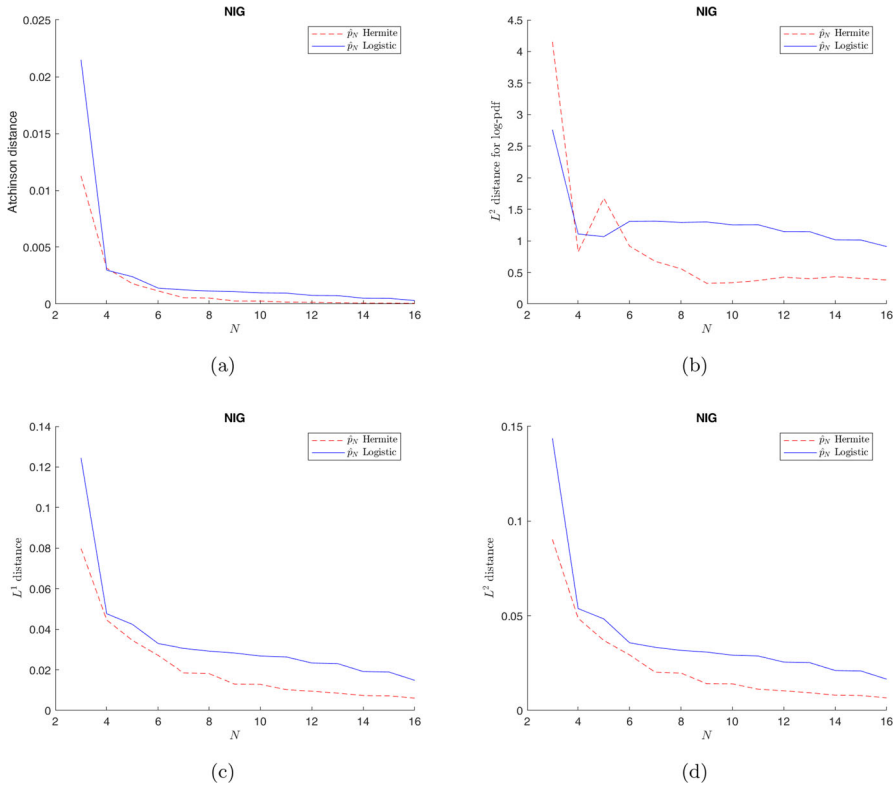

Fig. 5 These figures show, for the VG model, the convergence of the approximated PDF $\hat{p}_N$ towards the true PDF $p(x)$, using: the Atchinson distance defined in (18) (**a**-above left), the log-square distance defined in (19) (**b**-above right), the $\mathcal{L}^1$ distance defined in (20) (**c**-below left) and the $\mathcal{L}^2$ distance defined in (21) (**d**-below right)

Fig. 6 These figures show, for the NIG model, the convergence of the approximated PDF $\hat{p}_N$ towards the true PDF $p(x)$, using: the Atchinson distance defined in (18) (**a**-above left), the log-square distance defined in (19) (**b**-above right), the $\mathcal{L}^1$ distance defined in (20) (**c**-below left) and the $\mathcal{L}^2$ distance defined in (21) (**d**-below right)

for all $x \in I$. In this first case, both the Hermite and the logistic series converges towards the true PDF and, given a fixed number of terms N, the Hermite series is more accurate. Figure 7 reports the plots of the distances between the approximate and the true distribution, with the limited domain I, as the number of terms N increases. However due to the restricted domain, we lose information in tails of the distribution. Secondly, we adopt the original domain I with $L = 4$. In this second case, the Hermite series diverges and it cannot be used to approximate the distribution. The Logistic series, instead, remains stable and converges toward the Heston PDF. Then, we conclude that the Logistic series is more robust with respect to the enlargement of the domain I, reproducing better the tails of the distribution. In this second case, Fig. 8 reports the plots of the distances only for Logistic series, as the number of terms N increases.

Figures 9, 10, 11 and 12 provide graphical comparisons between the true PDF (or log-PDF) $p(x)$ and the approximated PDF (or log-PDF) $\hat{p}_N$ using either $N=6$ or $N=16$ terms. In the context of the VG model, the Logistic series surpasses the Hermite series, particularly in accurately representing the tails of the distribution. Conversely, for the
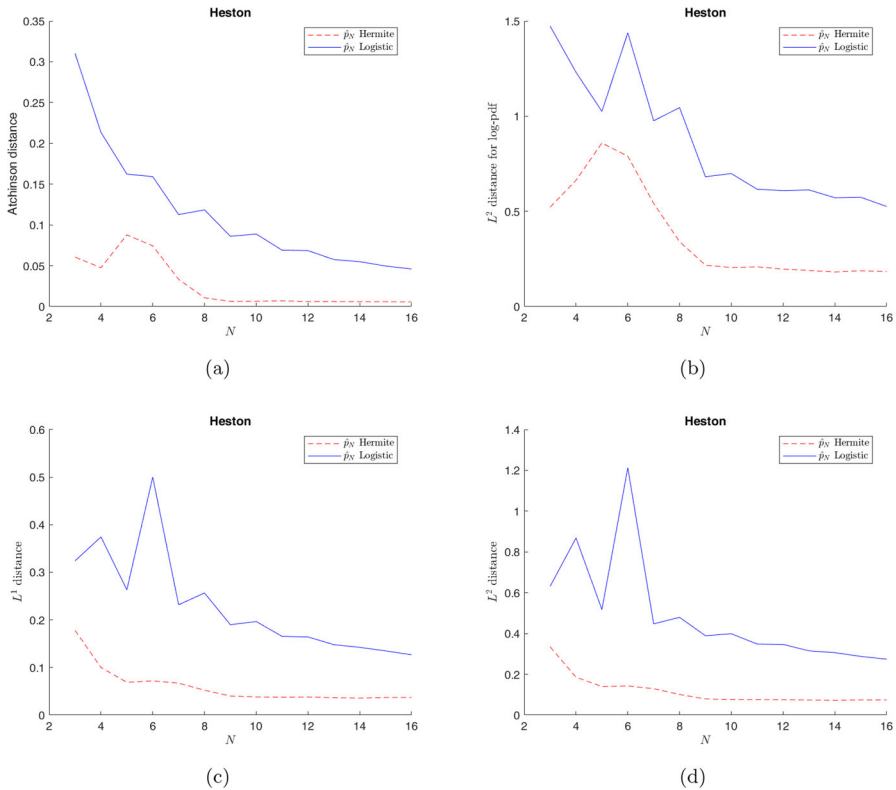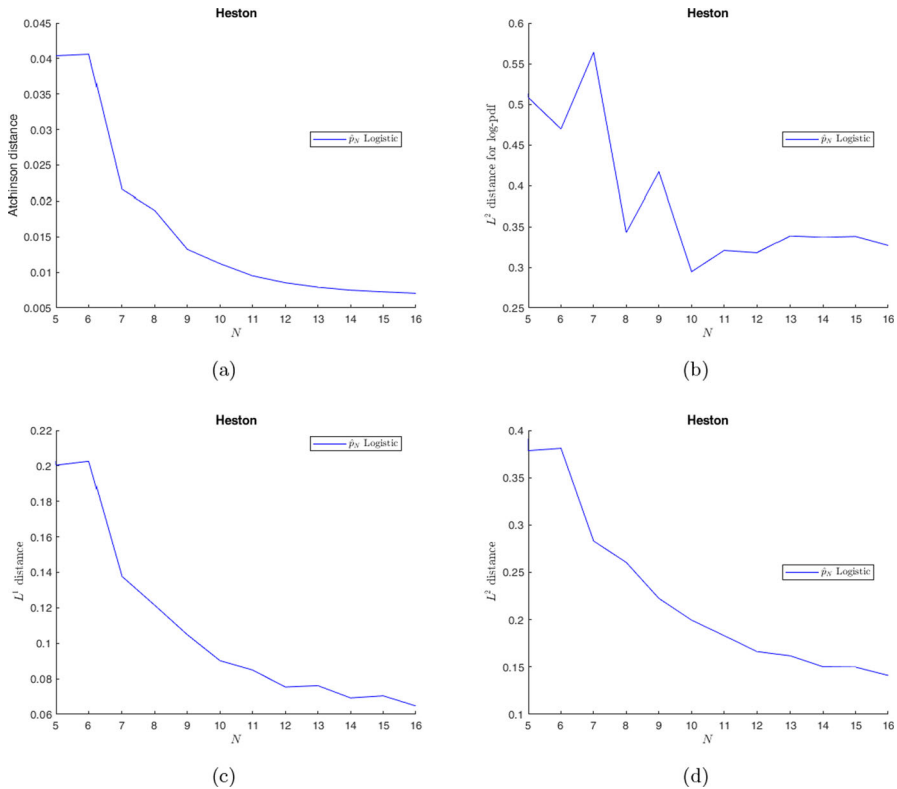
Fig. 7 These figures show, for the Heston model, the convergence of the approximated PDF $\hat{p}_N$ towards the true PDF $p(x)$, using: the Atchinson distance defined in (18) (**a**-above left), the log-pdf distance defined in (19) (**b**-above right), the $\mathcal{L}^1$ distance defined in (20) (**c**-below left) and the $\mathcal{L}^2$ distance defined in (21) (**d**-below right). The domain of the PDF I is truncated in such a way that $clr(p)(x) > -10$ for all $x \in I$

Table 2 For the three considered PDFs, the table reports the CPU time in seconds for the COS method with a grid size equal 2^{12} and the exponential expansion method with $N=16$ using Hermite or Logistic polynomials

Model	COS	Hermite pol	Logistic pol
VG	0.1475	0.0164	0.0614
NIG	0.3919	0.0205	0.0926
Heston	0.6552	0.0150	0.0492

NIG model, it appears that the Hermite series slightly outperforms the Logistic series. Concerning the Heston model, when a more constrained domain I is utilized, the Hermite series exhibits superior performance, as evidenced in Fig. 11. However, with an expanded domain I, the Logistic series approximation experiences a substantial improvement, as illustrated in Fig. 12, while the Hermite series diverges.

Finally, for the three considered PDFs, in Table 2, we report the CPU time in seconds for the COS method with a grid size equal 2^{12} and the exponential expansion

Fig. 8 These figures show, for the Heston model, the convergence of the approximated PDF $\hat{p}_N$ towards the true PDF $p(x)$, using: the Atchinson distance defined in (18) (**a**-above left), the log-square distance defined in (19) (**b**-above right), the $\mathcal{L}^1$ distance defined in (20) (**c**-below left) and the $\mathcal{L}^2$ distance defined in (21) (**d**-below right). The domain of the PDF I is defined using Eq. (22) with $L = 4$

method with N=16 using Hermite or Logistic polynomials. The numerical outcomes presented herein were generated on a computational system furnished with an Intel Core i3-9100F CPU clocked at 3.60 GHz and equipped with 32.0 GB of RAM. The results show that the proposed method is very fast and efficient.

6 Conclusions

In this work, we study the expansions of exponential form for the approximation of probability density functions, through the utilization of diverse orthogonal polynomial bases. Firstly, we introduce novel findings concerning the convergence of this series towards the true density function, employing mathematical tools of functional statistics. In particular, we show that the exponential expansion is a Fourier series of the true probability with respect to a given orthonormal basis of the so called Bayesian Hilbert space. Moreover, we extend to any orthogonal polynomial defined on the real line the numerical technique proposed by Muscolino and Ricciardi (1999) for the Hermite
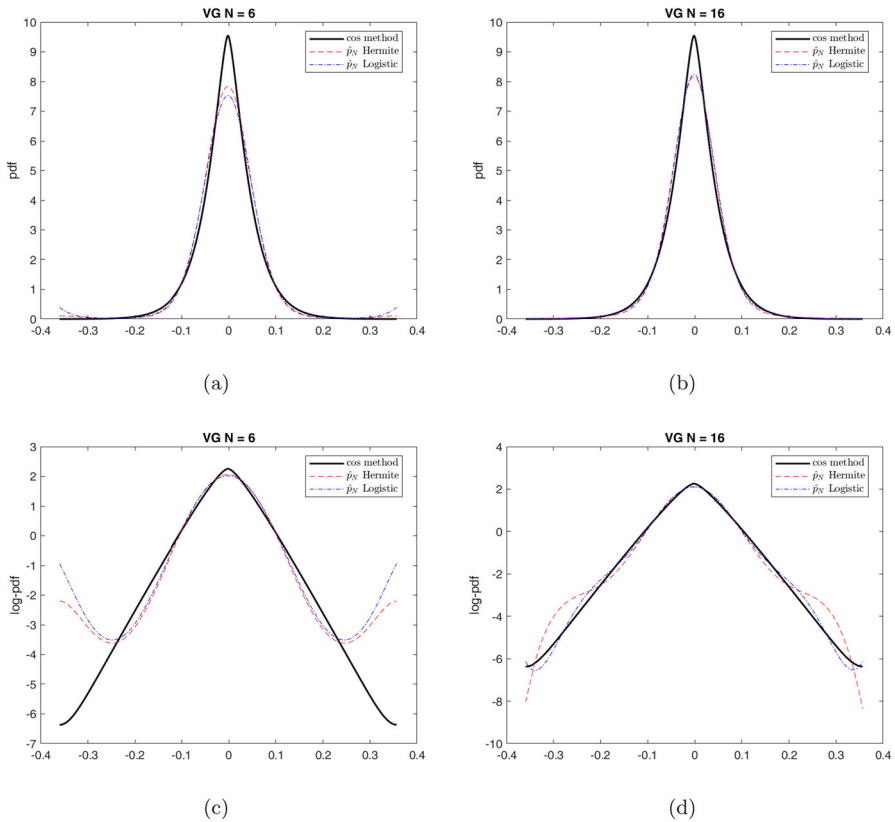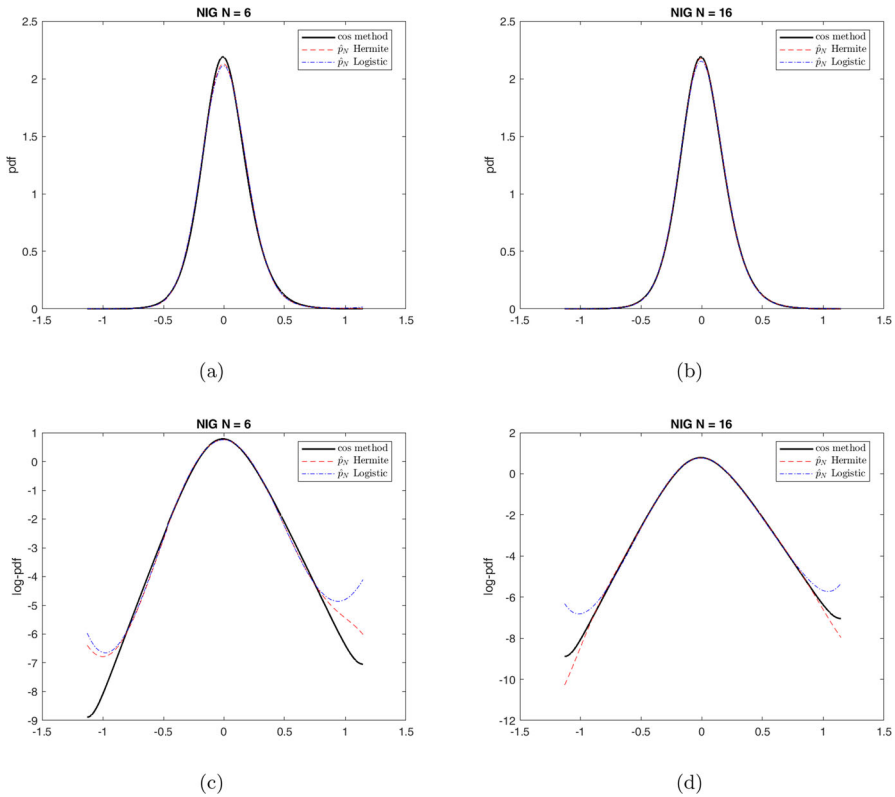
Fig. 9 For the VG model, these figures compare the true PDF (or log-PDF) $p(x)$ with the approximated PDF (or log-PDF) $\hat{p}_N$, using N=6 or N=16 number of terms

polynomials to estimate the coefficients of the expansion, using the first n moments of the corresponding true distribution. Finally, we illustrate numerical examples, applying our proposed approximation to three PDFs with different degrees of skewness and kurtosis and studying the convergence of the exponential series as the number of terms increases. Our results confirm the accuracy and straightforward implementation of our proposed approach. Furthermore, we show that adopting orthogonal basis different from Hermite polynomials, such as Logistic polynomials, can make the approximation more robust, when we enlarge the PDF domain, improving the accuracy of reproducing distribution tails.

6.1 Further researches

This preliminary study opens the way for several further researches. Firstly, continue the exploration of alternative bases could be beneficial. One approach could involve changing the reference weighting function, such as by considering a mixture of Gaussian distributions with non-zero skewness. Another possibility is to explore Riesz basis
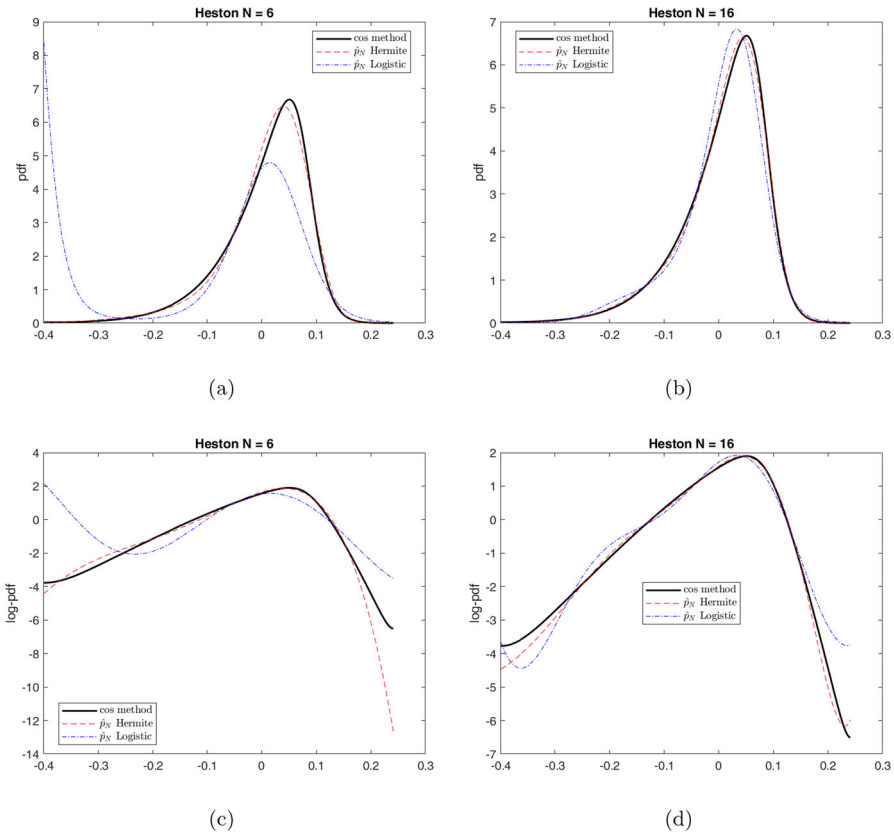
Fig. 10 For the NIG model, these figures compare the true PDF (or log-PDF) $p(x)$ with the approximated PDF (or log-PDF) $\hat{p}_N$, using $N=6$ or $N=16$ number of terms

functions (or bi-orthogonal bases), such as the class of B-splines, to approximate the centered-log-ratio transformation of the density function via orthogonal projections onto the basis span, see for instance (Ortiz-Gracia and Oosterlee 2013) and Kirkby (2015). In this case, an interesting target for further investigation is the estimation of projection coefficients in analytical form, given the first n exact moments of the corresponding true distribution, as done for orthonormal basis in the present paper. Secondly, it could be very interesting to analyse the robustness of the coefficient estimation with respect to noise in the estimated moments of the true PDF. Finally, a significant advantage of our proposed approach lies in providing an explicit analytical formula for the approximated PDFs. The explicit analytical formula can be used to calculate the derivatives of the PDF with respect to the model parameters, improving the results of different sensitivity analysis applications.
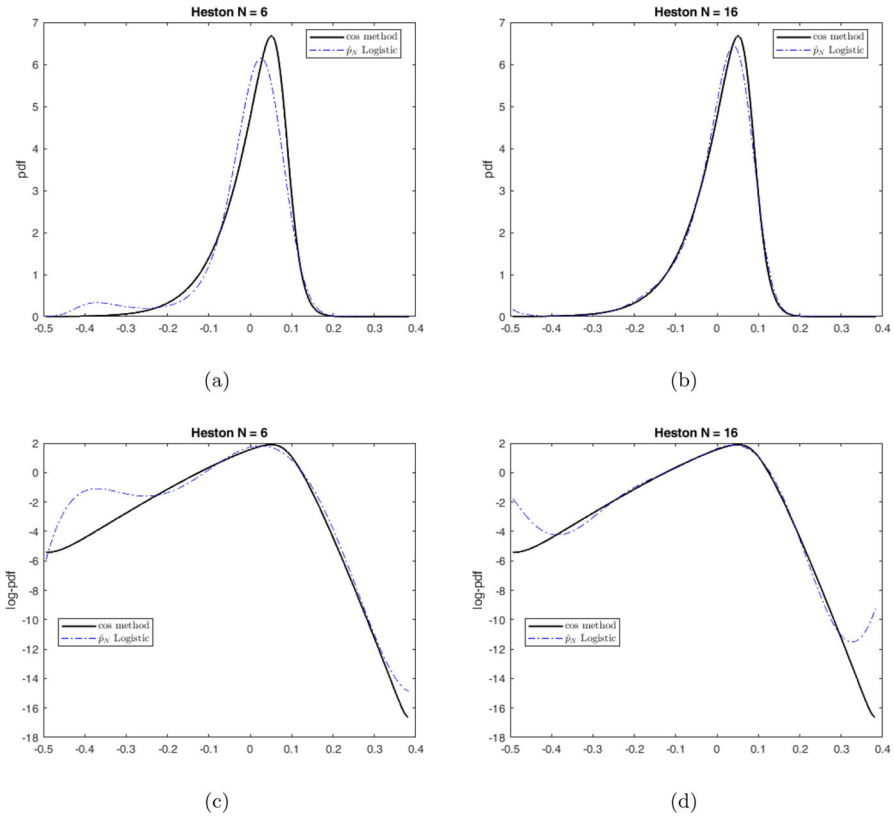
Fig. 11 For the Heston model, these figures compare the true PDF (or log-PDF) $p(x)$ with the approximated PDF (or log-PDF) $\hat{p}_N$, using $N=6$ or $N=16$ number of terms. The domain of the PDF I is truncated in such a way that $clr(p)(x) > -10$ for all $x \in I$

A Specification of the NIG model

The characteristic function of the NIG model is the following

$$\phi(u) = \exp\left(i\,u\,\mu\,dt + \frac{dt}{\varkappa} : \left(1 - \sqrt{1 + u^2\,\sigma^2\,\varkappa - 2i\,u\,\theta\,\varkappa}\right)\right)$$

where the parameters values are $\mu = 0$, $\theta = 0.05$, $\sigma = 0.2$ and $\varkappa = 0.3$.

B A moment-based estimation of coefficients for a PDF of two variables

In this section we discuss the moment-based estimation of the coefficients of the Fourier series c_j defined in formula (12) for the dimension $d=2$. Let us consider a

Fig. 12 For the Heston model, these figures compare the true PDF (or log-PDF) $p(x)$ with the approximated PDF (or log-PDF) $\hat{p}_N$, using N=6 or N=16 number of terms. The domain of the PDF I is defined using Eq. (22) with $L = 4$

(strictly positive) PDF $p(\mathbf{x})$ defined in $\mathbb{R}^2$. Then, as in Sect. 4, we consider $\{h_j\}_{j \geq 0}$ be the set of (normalized) Hermite polynomials, that forms an orthonormal basis for $\mathcal{L}^2(\mathbb{R}, \omega)$, where ω is the a standard Gaussian density function. The $\{h_j\, h_i\}_{j \geq 0, i \geq 0}$ is a orthogonal basis for $\mathcal{L}^2(\mathbb{R}^2, \omega\, \omega)$.

Following a similar procedure as the one dimension procedure of Muscolino and Ricciardi (1999) and in Sect. 4, by the integration by part formula and the property of the derivative of Hermite polynomials, that is $h'_j(x) = \sqrt{j}h_{j-1}(x)$, we obtain the following system of linear equations for $i, j = 1, ...N$

$$\sum_{n=1}^{N} \sum_{m=1}^{N} A_{i,j,n,m}\, \hat{c}_{n,m} = b_{i,j},$$

where

$$A_{i,j,n,m} = \sqrt{n} \sum_{k=0}^{i+n-2} \sum_{l=0}^{j+l-1} \frac{1}{k!l!} \Delta_{i-1,n-1,k} \ \Delta_{j-1,m,l} \ m_{k,l}^h,$$

$$b_{i,j} = -\sqrt{i-1} \ m_{i-2,j-1}^h,$$

with $m_{i,j}^h$ is the mixed normalized Hermite moment of $p(\mathbf{x})$, that is

$$m_{i,j}^h = \int h_i(x_1) h_j(x_2) p(\mathbf{x}) d\mathbf{x}.$$

Declarations

References

Abilov, V., Abilova, F., Kerimov, M.: Sharp estimates for the convergence rate of Fourier series in terms of orthogonal polynomials in $L_2((a, b), p(x))$. Comput. Math. Math. Phys. **49**, 927–941 (2009)

Ackerer, D., Filipović, D.: Option pricing with orthogonal polynomial expansions. Math. Finance **30**(1), 47–84 (2020)

Aitchison, J.: The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman and Hall, London (Reprinted in 2003 with additional material by Press Blackburn) (1986)

Alexits, G.: Convergence Problems of Orthogonal Series. International Series of Monographs in Pure and Applied Mathematics, vol. 20, pp. 63–170. Pergamon Press, Oxford (1961)

Bongiorno, E.G., Goia, A.: Classification methods for Hilbert data based on surrogate density. Comput. Stat. Data Anal. **99**, 204–222 (2016)

Bongiorno, E., Salinelli, E., Goia, A., Vieu, P.: Contributions in infinite-dimensional statistics and related topics. Soc. Editrice Esculapio srl (2014)

Cai, N., Shi, C.: A two-dimensional, two-sided Euler inversion algorithm with computable error bounds and its financial applications. Stoch. Syst. **4**(2), 404–448 (2014)

Cai, N., Kou, S.G., Liu, Z.: A two-sided Laplace inversion algorithm with computable error bounds and its applications in financial engineering. Adv. Appl. Probab. **46**(3), 766–789 (2014)

Charlier, C.: Frequency curves of type a in heterograde statistics. Ark. Mat. Ast. Fysik **9**, 1–17 (1914)

Charlier, C.: A new form of the frequency function. Maddalende fran Lunds Astronomiska Observatorium, Ser **II**, 51 (1928)

Cramer, H.: Mathematical Methods of Statistics, vol. 9, pp. 85–89. Princeton University Press, Princeton (1946)

Edgeworth, F.: The law of error. Proc. Camb. Philos. Soc. **20**, 36–65 (1905)

Efromovich, S.: Orthogonal series density estimation. WIREs Comput. Stat. **2**(4), 467–476 (2010)

Egozcue, J., Diaz-Barrero, J., Pawlowsky-Glahn, V.: Hilbert space of probability density functions based on Aitchison geometry. Acta Math. Sinica **22**, 1175–1182 (2006)

Fang, F., Oosterlee, C.W.: A novel pricing method for European options based on Fourier-cosine series expansions. SIAM J. Sci. Comput. **31**(2), 826–848 (2009)

Flamouris, D., Giamouridis, D.: Estimating implied pdfs from American options on futures: a new semiparametric approach. J. Future Mark. **22**(1), 1–30 (2002)

Gram, J.: On the development of real functions in rows by the method of least squares. J. Reine Angew. Math. **94**, 41–73 (1883)

Heston, S.L., Rossi, A.G.: A spanning series approach to options. Rev. Asset Pricing Stud. **7**(1), 2–42 (2016)

Jondeau, E., Rockinger, M.: Gram-Charlier densities. J. Econ. Dyn. Control **25**(10), 1457–1483 (2001)

Kirkby, J.L.: Efficient option pricing by frame duality with the fast Fourier transform. SIAM J. Financ. Math. **6**(1), 713–747 (2015)

Kendall, M.G., Stuat, A., Ord, J.K.: Advanced Theory of Statistics. Oxford University Press Inc, New York (1987)

Muscolino, G., Ricciardi, G.: Probability density function of mdof structural systems under non-normal delta-correlated inputs. Comput. Methods Appl. Mech. Eng. **168**(1), 121–133 (1999)

Ortiz-Gracia, L., Oosterlee, C.W.: Robust pricing of European options with wavelets and the characteristic function. SIAM J. Sci. Comput. **35**(5), B1055–B1084 (2013)

Provost, B.: Moment-based density approximants. Math. J. **9**(4), 727–756 (2005)

Rompolis, L.S., Tzavalis, E.: Recovering risk neutral densities from option prices: a new approach. J. Financ. Quant. Anal. **43**(4), 1037–1053 (2008)

Schoutens, W.: Stochastic Processes and Orthogonal Polynomials. In: Bickel, P., Diggle, P., Fienberg, S., Krickeberg, K. (eds.) Lecture Notes in Statistics, vol. 146. Springer, Berlin (2000)

Schwartz, S.C.: Estimation of probability density by an orthogonal series. Ann. Math. Stat. **38**(4), 1261–1265 (1967)

van den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V.: Bayes Hilbert spaces. Aust. N. Zeal. J. Stat. **56**(2), 171–194 (2014)

Van Den Boogaart, K.-G., Juan José, E., Vera, P.-G.: Bayes linear spaces. SORT-Stat. Oper. Res. Trans. **34**(2), 201–222 (2011)

Walter, G.G.: Properties of Hermite series estimation of probability density. Ann. Stat. **5**(6), 1258–1264 (1977)