A Framework for Mellin Kind Series Expansion Methods

Torgeir Brenn and Stian Normann Anfinsen

***Abstract*—Mellin kind statistics (MKS) is the framework which arises if the Fourier transform is replaced with the Mellin transform when computing the characteristic function from the probability density function. We may then proceed to retrieve logarithmic moments and cumulants, that have important applications in the analysis of heavy-tailed distribution models for non-negative random variables. In this paper we present a framework for series expansion methods based on MKS. The series expansions recently proposed in [1] are derived independently and in a different way, showing that the methods truly are Mellin kind analogies to the classical Gram-Charlier and Edgeworth series expansion. From this new approach, a novel series expansion is also derived. In achieving this we demonstrate the role of two differential operators, which are called Mellin derivatives in [2], but have not been used in the context of Mellin kind statistics before. Also, the Bell polynomials [3] are used in new ways to simplify the derivation and representation of the the Mellin kind Edgeworth series expansion. The underlying assumption of the nature of the observations which validates that series is also investigated. Finally, a thorough review of the performance of several probability density function estimation methods is conducted. This includes classical [4], [5] and recent methods [1], [6], [7] in addition to the novel series expansion presented in this paper. The comparison is based on synthesized data and sheds new light on the strengths and weaknesses of methods based on classical and Mellin kind statistics.**

***Index Terms*—Synthetic aperture radar, non-negative random variables, probability density function estimation, Mellin kind statistics, method of log-cumulants, Gram-Charlier series, Edgeworth series.**

NOMENCLATURE

CF	Characteristic function.
CGF	Cumulant generating function.
FT	Fourier transform.
IID	Independent and identically distributed.
GΓD	Generalized gamma distribution.
MoLC	Method of log-cumulants.
MKCF	Mellin kind characteristic function.
MKCGF	Mellin kind cumulant generating function.
MKE	Mellin kind Edgeworth.
MKGK	Mellin kind gamma kernel.
MKLK	Mellin kind log-normal kernel.
MKS	Mellin kind statistics.
MT	Mellin transform.
PDF	Probability density function.
RV	Random variable.
SAR	Synthetic aperture radar.

The authors are with the Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway (e-mail: torgeir.brenn@gmail.com, stian.normann.anfinsen@uit.no).

I. INTRODUCTION

ESTIMATING the probability density function (PDF) is a central part of many data analysis applications. This includes various model based image analysis tasks using parametric PDFs. The choice of model is a trade-off: Advanced models can be highly accurate for a relatively wide range of data, but are usually mathematically and computationally demanding. Parameter estimation may also pose challenges. Simple models are implemented easily and run fast, but are less flexible and may not provide a good fit to the data.

In the case of near-Gaussian data, the Gram-Charlier [8] and Edgeworth [9] series expansions provide attractive alternatives. They combine the simplicity of a fitted Gaussian distribution with the flexibility and accuracy of accounting for higher order moments of the data, i.e. skewness, excess kurtosis, etc. However, these methods have not proven as effective for non-negative random variables (RVs), that is, RVs which maps to zero on the entire negative part of the real line (support $\subseteq \mathbb{R}_{\geq 0} = [0, \infty)$). Radar intensity data naturally fall into this category, and for the purpose of synthetic aperture radar (SAR) image analysis, several distributions have been suggested as data models. These distributions are also relevant for other coherent imaging modalities, including ultrasound, sonar and laser images. They are commonly based upon a doubly stochastic product model [10], [11], which means that the observed RV is modelled as the product of two unobservable RVs, and its PDF is consequentially found through a multiplicative convolution. There are numerous other fields in economics, science and engineering that also make use of heavy-tailed distribution models for non-negative RVs. [12], [13]

Mellin kind statistics (MKS) were introduced by Nicolas in [14] and has proven to be a powerful framework designed to deal with the product model and non-negative RVs. In MKS, the Fourier transform (FT) is replaced by the Mellin transform (MT), giving a Mellin kind characteristic function (MKCF) in place of the classical characteristic function (CF). The MKCF of a product $X \cdot Y$ of independent RVs is the product of the constituent MKCFs, matching the property the CF has with respect to the sum $X + Y$. Logarithmic moments and cumulants are statistics with natural inherent qualities in MKS, and can be retrieved in an analogous way to their classical linear counterparts. The framework has since been expanded to the matrix-variate case [15]. Furthermore, it has been utilized extensively for estimation problems through the method of logarithmic cumulants (MoLC) [6], used as a tool to understand the physical process underlying the acquisition of coherent images [11], and it has also been used to produce asymptotic expansions for PDFs [16].

The paper is organized as follows. In Section II, we briefly summarize the Gram-Charlier and Edgeworth series, the Bell polynomials, the MT and its properties, and MKS. In Section III, we introduce a complete framework for the Mellin kind equivalents of the classical series expansions, including a new series based on a gamma distribution kernel. These series, along with other classical and modern methods, are used to approximate known PDFs in Section IV and to estimate unknown PDFs in Section V. We give our conclusions in Section VI.

II. Theory

A. Classical Series Expansion Methods

For a RV X with unknown PDF $f_X(x)$, its CF $\Phi_X(\omega)$ is the FT of $f_X(x)$ [17], that is

$$\Phi_X(\omega) \equiv \mathcal{F}[f_X(x)](\omega) = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x)\,\mathrm{d}x = \mathrm{E}\{e^{j\omega X}\}, \quad (1)$$

where j is the imaginary unit, the expectation operation $\mathrm{E}\{\cdot\}$ is performed with respect to X, and $\omega \in \mathbb{R}$ is the transform variable. The linear moments of order ν are defined as $m_\nu \equiv \mathrm{E}\{X^\nu\}$. The natural logarithm of the CF is called the cumulant generating function (CGF), since the cumulants $c_{X,\nu},\ \nu \in \mathbb{Z}_{>0}$ can (if they all exist) be retrieved from

$$\log \Phi_X(\omega) = \sum_{\nu=1}^{\infty} c_{X,\nu} \frac{(j\omega)^\nu}{\nu!}. \quad (2)$$

Let $\alpha(x) = (2\pi)^{-1/2} e^{-x^2/2}$ denote the standardized (zero mean, unit variance) Gaussian kernel with CF $\Phi_\alpha(\omega)$ [18]. Unique to the Gaussian distribution is the property of its cumulants: $c_{\alpha,\nu} = 0\,\forall\,\nu \geq 3$ [19]. Combining (1) and (2) with the CGF of $\alpha(x)$ gives the CF of X as

$$\Phi_X(\omega) = \exp\left\{\sum_{\nu=1}^{\infty} [c_{X,\nu} - c_{\alpha,\nu}] \frac{(j\omega)^\nu}{\nu!}\right\} \Phi_\alpha(\omega). \quad (3)$$

where we see for standardized X

$$c_{X,\nu} - c_{\alpha,\nu} = \begin{cases} 0 & \nu = 1,2 \\ c_{X,\nu} & \nu \geq 3 \end{cases}. \quad (4)$$

The PDF of X can now be retrieved [17] from (3) via the inverse FT

$$f_X(x) = \exp\left\{\sum_{\nu=3}^{\infty} c_{X,\nu} \frac{(-\mathrm{D}_x)^\nu}{\nu!}\right\} \alpha(x), \quad (5)$$

where $\exp\{\cdot\}$ is the exponential function and $\mathrm{D}_x = \mathrm{d}/\mathrm{d}x$ is the derivative operator. To get a more tractable expression, the exponential function is reduced via its power series $\exp\{x\} = \sum_{k=0}^{\infty} x^k/k!$ to give an infinite double sum. Now we can collect the terms according to the power of $(-\mathrm{D}_x)$ and recollect the definition of the Hermite polynomials, $H_\nu(x)\alpha(x) = (-\mathrm{D}_x)^\nu \alpha(x)$ [17], to get the classical Gram-Charlier series

$$f_X(x) = \left[1 + \frac{c_{X,3}}{6} H_3(x) + \frac{c_{X,4}}{24} H_4(x) + \cdots\right]. \quad (6)$$

Edgeworth's idea was to assume that the nearly-Gaussian RV X was a standardized sum

$$X = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} \frac{Z_i - m}{\varsigma}, \quad (7)$$

where the RVs $Z_1, Z_2, \ldots, Z_r$ are independent and identically distributed (IID), each with mean m, variance ς^2 and higher order cumulants $c_{Z,\nu} = \varsigma^\nu \lambda_\nu$. The dimensionless λ_ν will simplify the following derivation, and the properties of the cumulants give [8]

$$c_{X,\nu} = \frac{\lambda_\nu}{r^{\frac{\nu}{2}-1}}\ ,\ \ \nu \geq 3. \quad (8)$$

Collecting the terms in (5) based on their power of $r^{-1/2}$ instead gives the Edgeworth series [9]

$$\begin{aligned} f_X(x) =& \alpha(x) + r^{-\frac{1}{2}} \left[\frac{\lambda_3}{6} H_3(x)\right] \alpha(x) \quad (9) \\ &+ r^{-1} \left[\frac{\lambda_4}{24} H_4(x) + \frac{\lambda_3^2}{72} H_6(x)\right] \alpha(x) + O\left(r^{-\frac{3}{2}}\right). \end{aligned}$$

Its convergence is found to be superior to the Gram-Charlier series both with few terms and asymptotically [9].

B. The Bell Polynomials

Named in honor of Eric T. Bell who introduced what he called partition polynomials in [20], the partial Bell polynomials are defined as [21]

$$B_{n,r}(x_1, x_2, \ldots, x_{n-r+1}) = \sum_{\Xi_r} n! \prod_{i=1}^{n-r+1} \frac{1}{j_i!} \left(\frac{x_i}{i!}\right)^{j_i}, \quad (10)$$

where the sum is the over the set Ξ_r of all combinations of non-negative integers $j_1, \ldots, j_n$ which satisfy $j_1 + 2j_2 + \cdots + (n-r+1)j_{n-r+1} = n-r+1$ and $r = j_1 + k_2 + \cdots + j_{n-r+1}$. The nth complete Bell polynomial is the sum

$$B_n(x_1, \ldots, x_n) = \sum_{r=1}^{n} B_{n,r}(x_1, x_2, \ldots, x_{n-r+1}). \quad (11)$$

The Bell polynomials satisfy [21]

$$\exp\left\{\sum_{\nu=1}^{\infty} x_\nu \frac{t^\nu}{\nu!}\right\} = \sum_{n=0}^{\infty} B_n(x_1, \ldots, x_n) \frac{t^n}{n!}, \quad (12)$$

and a well-known use of this result is to retrieve the νth order moment from the cumulants of order $\leq \nu$ [22] as

$$m_\nu = B_\nu(c_1, \ldots, c_\nu). \quad (13)$$

Another use of (12) is to sort the terms in (5) to retrieve the Gram-Charlier series in a simple manner [23]. In [24], we demonstrate how the Bell polynomials can also be used to sort the terms in the classical Edgeworth series, providing a simpler way of evaluating (9) up to arbitrary power of $r^{-1/2}$.

TABLE I
MELLIN TRANSFORM PROPERTIES

	$f(x)$	$F(s) = \mathcal{M}[f(x)](s)$	Condition
Linearity	$af(x) + bg(x)$	$aF(s) + bG(s)$	a, b constant, $s \in S_f$
Scaling	$f(ax)$	$a^{-s}F(s)$	$a > 0$, $s \in S_f$
Multiplication	$x^n f(x)$	$F(s+n)$	$s + n \in S_f$
Differentiation	$D_x^n f(x)$	$(-1)^n (s-1)_n F(s-n)$	$n \in \mathbb{Z}_{\geq 0}$, $s - n \in S_f$
A combination	$D_x^n x^n f(x)$	$(-1)^n (s-1)_n F(s)$	$s \in S_f$
A Mellin derivative	$(D_x x)^n f(x)$	$(-1)^n (s-1)^n F(s)$	$s \in S_f$

C. The Mellin Transform

The MT of a function $f(x)$ is

$$\mathcal{M}[f(x)](s) \equiv \int_0^\infty x^{s-1} f(x) \mathrm{d}x = F(s) \Leftrightarrow f(x) \xrightarrow{\mathcal{M}} F(s), \quad (14)$$

where $s \in \mathbb{C}$ is the transform variable. The MT is limited to functions which satisfy $f(x) = 0 \, \forall \, x < 0$, i.e. f(x) has support $\subseteq \mathbb{R}_{\geq 0}$. The fundamental strip S_f is the largest open strip $\langle a, b \rangle$ of $\mathrm{Re}(s)$ for which the integral in (14) converges. If $s \in S_f$, then $f(x)$ is retrievable via the inverse MT [25],

$$f(x) = \mathcal{M}^{-1}[F(s)](x) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} x^{-s} F(s) \mathrm{d}s, \quad (15)$$

where the integral is taken along a vertical line in the complex plane, with the fundamental strip defined by its real (vertical) boundaries.

Some general properties of the MT are listed in Table I [26], [27]. The differentiation properties introduces the falling factorial, defined as $(s-1)_n = (s-1)(s-2)\cdots(s-n)$, and the Mellin derivative[1] operator, defined as $D_x x$. Note that the final two properties differ because differentiation and multiplication operations are non-commutative. E.g., for $n = 2, f(x) = 1$ we have $(D_x x)^2 = D_x x D_x x = 1$, whereas $D_x^2 x^2 = D_x 2x = 2$.

D. Mellin Kind Statistics

While the idea of using the Mellin transform (MT) as a tool for statistical analysis had been proposed earlier [28], it did not receive much attention until the introduction of MKS in [14]. The MKCF $\phi_X(s)$ is defined as the MT of the PDF

$$\phi_X(s) \equiv \mathcal{M}[f_X(x)](s) = \int_0^\infty x^{s-1} f_X(x) \mathrm{d}x = \mathrm{E}\{X^{s-1}\}. \quad (16)$$

The log-moments are defined as $\mu_\nu \equiv \mathrm{E}\{(\log X)^\nu\}$, where $\nu \in \mathbb{Z}_{\geq 0}$. The MKCF can be expressed in terms of the log-moments by rewriting the transform kernel $x^{s-1} = e^{(\log x)(s-1)}$ in (16), inserting the power series expansion for the exponential function, and finally changing the order of integration and summation to recognize the log-moments from their definition, i.e.

$$\phi_X(s) = \sum_{\nu=0}^\infty \mu_\nu \frac{(s-1)^\nu}{\nu!}. \quad (17)$$

As in the classical case, this depends on the existence of all μ_ν, and under this condition it is also possible to retrieve the log-moments from

$$\mu_\nu = D_s^\nu \phi_X(s)\big|_{s=1}. \quad (18)$$

The log-cumulant generating function (MKCGF) is defined $\varphi_X(s) = \log \phi_X(s)$. Provided all log-cumulants κ_ν exist, we then have

$$\varphi_X(s) = \sum_{\nu=1}^\infty \kappa_\nu \frac{(s-1)^\nu}{\nu!}, \quad (19)$$

$$\kappa_\nu = D_s^\nu \varphi_X(s)\big|_{s=1}. \quad (20)$$

The equivalent result as (13) trivially holds for the log-moments and log-cumulants, since their relations are identical [14], [29].

For a more detailed review of the fundamental properties of MKS, see e.g. [14] (English translation: [30]), while [15] emphasizes the analogy to classical statistics and expands the framework to the matrix-variate case. A comprehensive list of MKCFs and MKCGFs for several distributions can be found in [31].

III. A FRAMEWORK FOR THE MELLIN KIND SERIES EXPANSION METHODS

A. The Mellin Kind Gram-Charlier Series Expansion with Arbitrary Kernel

For a non-negative RV X and an arbitrary continuous PDF kernel $\theta(x)$ with support $\mathbb{R}_{\geq 0}$, whose log-moments and log-cumulants all exist, it is possible to mirror the approach in Section II leading up to (3), giving the MKCF of X as

$$\phi_X(s) = \exp\left\{ \sum_{\nu=1}^\infty [\kappa_{X,\nu} - \kappa_{\theta,\nu}] \frac{(s-1)^\nu}{\nu!} \right\} \phi_\theta(s). \quad (21)$$

In the same way [23] applied the Bell polynomials to the classical Gram-Charlier series, applying the result in (12) gives us

$$\phi_X(s) = \left[\sum_{n=0}^\infty B_n(\Delta\kappa_1, \Delta\kappa_2, \ldots, \Delta\kappa_n) \frac{(s-1)^n}{n!} \right] \phi_\theta(s), \quad (22)$$

[1] $D_x x$ and $x D_x$ are called "Mellin derivatives" in [2].

where $\Delta\kappa_n = \kappa_{X,n} - \kappa_{\theta,n}$ is used for brevity. Table I contains the Mellin derivative property, which has not been used in the context of MKS before now. It provides an inverse MT of (22), leading up to

$$f_X(x) = \left[\sum_{n=0}^{\infty} B_n(\Delta\kappa_1, \Delta\kappa_2, \ldots, \Delta\kappa_n)\frac{(-\mathrm{D}_x x)^n}{n!}\right]\theta(x), \quad (23)$$

which is the Mellin kind Gram-Charlier series expansion with arbitrary kernel.

B. The Mellin Kind Gamma Kernel Series

Let

$$\gamma(x; a, b) = \begin{cases} \frac{b^a x^{a-1}}{\Gamma(a)} e^{-bx} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases} \quad (24)$$

denote the gamma distribution PDF with shape a and scale b. We will now substitute $\theta(x) \rightarrow \gamma(x; a, b)$ in (23) to get the Mellin kind gamma kernel (MKGK) series. To get an applicable expression, it is necessary to evaluate $(-\mathrm{D}_x x)^n \gamma(x; a, b)$. By letting $b = 1$, it is possible to define the polynomials $M_n(x)$ implicitly as

$$M_n(x)\gamma(x; a) = (-\mathrm{D}_x x)^n \gamma(x; a). \quad (25)$$

or[2]

$$M_n(x) = x^{-(a-1)} e^x (-\mathrm{D}_x x)^n [x^{a-1} e^{-x}], \quad (26)$$

and since $\mathrm{D}_x x$ is scale invariant, the generalization to arbitrary b is simply to replace x with bx in the polynomials, i.e.

$$M_n(bx)\gamma(x; a, b) = (-\mathrm{D}_x x)^n \gamma(x; a, b). \quad (27)$$

In Appendix A we prove that $M_n(x)$ is a linear combination of the well-known generalized Laguerre polynomials [33] and give explicit polynomials for $n = 0, 1, 2, 3$. The MKGK series can now be completed by substituting $\gamma(x; a, b)$ for $\theta(x)$ and (27) into (23) to yield

$$f_X(x) \approx \left[\sum_{n=0}^{N} \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \ldots, \Delta\kappa_n) M_n(bx)\right] \gamma(x; a, b), \quad (28)$$

where the sum was truncated to finite N.

The parameters a and b of the kernel can be chosen such that the log-cumulants $\kappa_{\gamma,1}$ and $\kappa_{\gamma,2}$ match the respective population log-cumulants $\kappa_{X,1}$ and $\kappa_{X,2}$. This way we can approximate any given PDF model for X. If the model is unknown, then $\kappa_{\gamma,1}$ and $\kappa_{\gamma,2}$ can be matched with the corresponding sample log-cumulants which amounts to producing MoLC estimates of a and b [6]. This simplification is considerable, as the Bell polynomials of degree 0 through 6 consist of 30 terms in total, only 6 of which are non-zero if $\Delta\kappa_1 = \Delta\kappa_2 = 0$, and (28) is reduced to

$$f_X(x) \approx \left[1 + \sum_{n=3}^{N} \frac{1}{n!} B_n(0, 0, \Delta\kappa_3, \ldots, \Delta\kappa_n) M_n(bx)\right] \gamma(x; a, b). \quad (29)$$

The first few terms in the MKGK are presented in (40).

[2]This definition mirrors the Rodrigues type formulae [32], with the Mellin derivative replacing the standard differentiation operator.

C. The Mellin Kind Log-normal Kernel Series

Now insert for $\theta(x)$ the log-normal PDF kernel

$$\Lambda(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} \quad (30)$$

with log-mean $\mu = \mathrm{E}\{\log X\}$ and log-variance $\sigma^2 = \mathrm{E}\{(\log X - \mu)^2\}$ [16] to obtain the Mellin kind log-normal kernel (MKLK) series. To evaluate $(-\mathrm{D}_x x)^n \Lambda(x; \mu, \sigma)$, we use the proof (59) of Lemma 3 from Appendix B to see that

$$f_X(x) = \left[\sum_{n=0}^{\infty} \frac{1}{n!\sigma^n} B_n(\Delta\kappa_1, \ldots, \Delta\kappa_n) H_n\left(\frac{\log x - \mu}{\sigma}\right)\right] \Lambda(x; \mu, \sigma). \quad (31)$$

Matching of the log-mean μ and log-variance σ^2 to $\kappa_{X,1}$ and $\kappa_{X,2}$ not only results in that most of the terms vanish, as with the MKGK series, but since [31]

$$\kappa_{\Lambda,\nu} = \begin{cases} \mu & \nu = 1 \\ \sigma^2 & \nu = 2 \\ 0 & \nu \geq 3 \end{cases}, \quad (32)$$

we have that $\Delta\kappa_n = \kappa_{X,n}$ for $n > 2$. That is, for $\Lambda(x; \mu, \sigma)$ with tailored parameters,

$$f_X(x) \approx \left[1 + \sum_{n=3}^{N} \frac{B_n(0, 0, \kappa_{X,3}, \ldots, \kappa_{X,n})}{n!\sigma^n} \cdot H_n\left(\frac{\log x - \mu}{\sigma}\right)\right] \Lambda(x; \mu, \sigma), \quad (33)$$

with the first few terms presented in (41).

D. Mellin Kind Edgeworth Series

Recall that the classical Edgeworth series is based on the assumption in (7) which leads to (8). In Appendix C we prove that replacing X with $\log X$ in (7) gives the log-cumulant differences

$$\Delta\kappa_\nu = \kappa_{X,\nu} - \kappa_{\Lambda,\nu} = \begin{cases} 0 & \nu = 1, 2 \\ \frac{\lambda_\nu}{r^{\frac{\nu}{2}-1}} & \nu \geq 3 \end{cases}, \quad (34)$$

where (32) was also used. Using this result and inserting the MKCF $\phi_\Lambda(s)$ of the log-normal kernel into (21) gives

$$\phi_X(s) = \exp\left\{\sum_{\nu=3}^{\infty} \frac{\lambda_\nu}{r^{\frac{\nu}{2}-1}} \frac{(s-1)^\nu}{\nu!}\right\} \phi_\Lambda(s). \quad (35)$$

Shifting the index $\nu \rightarrow \nu + 2$, this can instead be viewed as a power series in $r^{-1/2}$

$$\phi_X(s) = \exp\left\{\sum_{\nu=1}^{\infty} \zeta_\nu(s-1) \frac{r^{-\nu/2}}{\nu!}\right\} \phi_\Lambda(s), \quad (36)$$

where

$$\zeta_\nu(s) = \frac{\lambda_{\nu+2}}{(\nu+1)(\nu+2)} s^{\nu+2}. \quad (37)$$

Since the function $\zeta_\nu(s-1)$ is independent of r, property (12) gives

$$\phi_X(s) = \left[\sum_{n=0}^{\infty} B_n\Big(\zeta_1(s-1), \ldots, \zeta_n(s-1)\Big) \frac{r^{-n/2}}{n!}\right] \phi_\Lambda(s). \quad (38)$$

This is a polynomial in $(s-1)$, so the inverse MT can be applied as for the MKGK and MKLK series to yield

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} B_n\Big(\zeta_1(-D_x x),\ \ldots,\ \zeta_n(-D_x x)\Big)\frac{r^{-n/2}}{n!}\right]\Lambda(x;\mu,\sigma), \tag{39}$$

where the ζ function from (37) is now used with index n and the operator $-D_x x$ as input.

As in (33), the term corresponding to $n = 0$ is unity, indicating that this is a series around the tailored log-normal kernel. Lemma 3 is again used to replace the Mellin derivative with the Hermite polynomials, and the first few terms of the Mellin kind Edgeworth (MKE) series are presented in (42). Note that the first correction term of the MKLK and MKE series are equal.

The MKE series is identical to the series presented in [1], but it is derived independently as the result of a different approach. The authors of [1] used a change of variable, $s-1 \to j\omega$, the fact that if $Y = \log X$, and then [14] showed that $\phi_X(s) = \Phi_Y(\omega)$. This allowed for the inverse FT to be used, while the present approach involves the Mellin derivative and the inverse MT. Secondly, the role of the Bell polynomials in the series expansion methods is illuminated to reveal an alternative representation to the one in [1], which mirrored the classical Edgeworth series in [9].[3]. Thirdly, use of $B_n(\cdot)$ allows generalization into a framework of approximations with arbitrary kernel function.

IV. Approximating Known Distributions

In this section we use the MKGK, MKLK and MKE series and other methods to approximate known PDFs. That is, we assume that the distribution parameters are known and do not need to estimate the log-cumulants $\kappa_{X,n}$ or in fact any quantity in (40), (41) and (42) from data.

For all simulations in this paper, we compute the Bhattacharyya distance $d_B(f_X(x), \hat{f}_X(x))$ [35], the Kullback-Leibler distance[4] $d_{KL}(f_X(x), \hat{f}_X(x))$, and the maximum error (i.e. the Kolmogorov-Smirnov test or L^∞ norm distance) to ensure that our conclusions are not colored by our choice of dissimilarity measure. In most cases the results were highly consistent, allowing us to present only one or two of the measures for brevity.

Since the series expansions are not in general true probability measures (they do not always integrate to unity and exhibit negative values of $\hat{f}_X(x)$) [9], we needed to slightly modify the estimates to ensure that the ratios and logarithms in $d_B(\cdot)$ and $d_{KL}(\cdot)$ do not fail, but also gives fair results.[5]

A. Broad Comparison of the Methods

We start with a general comparison of 7 methods based on log-cumulants. In addition to the MKGK, MKLK and MKE series, the methods tested are the gamma, log-normal, K [6], and generalized gamma (GΓD) [7] distributions with parameters fitted by the MoLC. Note that the MoLC gamma method corresponds to the kernel of the MKGK series, i.e. $N \le 2$ in (29), while the MoLC log-normal method corresponds to the kernels in both the MKLK and MKE series, i.e. $N \le 2$ in (33) and truncating the entire sum in (39). The series are here corrected only for $\kappa_{X,3}$ and $\kappa_{X,4}$, specifically $N = 4$ in the MKGK and MKLK series and truncating terms of $O\big(r^{-3/2}\big)$ in the MKE series. Note that fewer terms render the MKLK and MKE series identical. In Section IV-B we examine how the series depend on the number of terms.

In Fig. 1, we approximate the K distribution with PDF [14]

$$\mathcal{K}(x;\mu,L,M) = \frac{2LM}{\mu\Gamma(L)\Gamma(M)}\left(\frac{LMx}{\mu}\right)^{\frac{M+L}{2}-1} K_{M-L}\left(2\sqrt{\frac{LMx}{\mu}}\right), \tag{43}$$

where $x, \mu, L, M > 0$, and the G^0 distribution with PDF

$$G^0(x;g,L,M) = \frac{L^L\Gamma(L-M)x^{L-1}}{g^M\Gamma(L)\Gamma(-M)(g+Lx)^{L-M}}, \tag{44}$$

where $x, g, L > 0$, $M < 0$. These two distributions are given in [36] as two distributions with physical foundation which arise when modelling observed SAR intensity of a heterogeneous scene.

Fig. 1 (a) shows the relative success of all methods in modelling the K distribution with a high shape parameter $L = 16$, known in the SAR literature as the number of looks [37]. The MKLK, MKE, and MKGK series visibly improve

[3] The Bell polynomials can be used in the classical (linear) case as well [23], [34], [24].

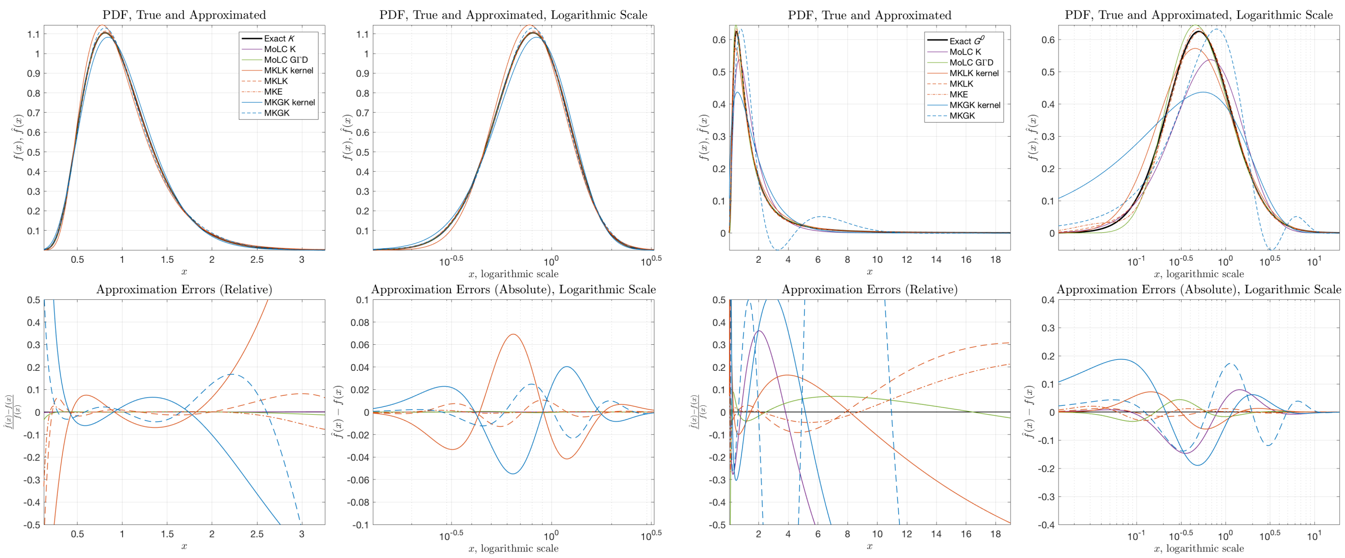
[4] We take distance to mean a non-negative definite symmetric function, and $d_{KL}(f_X(x), \hat{f}_X(x))$ is the symmetrized version of the Kullback-Leibler divergence. This follows the nomenclature of [35].

[5] Specifically, to remedy the common feature that the series expansion methods produce estimates $\hat{f}_X(x)$ which integrate to < 1 and result in $d_B(\cdot) < 0$, we divided $\hat{f}_X(x)$ by its integral.

$$f_X(x) = \left[1 + \frac{\Delta\kappa_3}{6}M_3(bx) + \frac{\Delta\kappa_4}{24}M_4(bx) + \frac{\Delta\kappa_5}{120}M_5(bx) + \frac{\Delta\kappa_6+10\Delta\kappa_3^2}{720}M_6(bx) + \frac{\Delta\kappa_7+35\Delta\kappa_3\Delta\kappa_4}{5040}M_7(bx) + \cdots\right]\gamma(x;a,b) \tag{40}$$

$$f_X(x) = \left[1+\frac{\kappa_3}{6\sigma^3}H_3\left(\frac{\log x-\mu}{\sigma}\right)+\frac{\kappa_4}{24\sigma^4}H_4\left(\frac{\log x-\mu}{\sigma}\right)+\frac{\kappa_5}{120\sigma^5}H_5\left(\frac{\log x-\mu}{\sigma}\right)+\frac{\kappa_6+10\kappa_3^2}{720\sigma^6}H_6\left(\frac{\log x-\mu}{\sigma}\right)+\cdots\right]\Lambda(x;\mu,\sigma) \tag{41}$$

$$f_X(x) = \Lambda(x;\mu,\sigma) + \frac{1}{r^{1/2}}\left[\frac{\lambda_3}{6\sigma^3}H_3\left(\frac{\log x-\mu}{\sigma}\right)\right]\Lambda(x;\mu,\sigma) + \frac{1}{r}\left[\frac{\lambda_4}{24\sigma^4}H_4\left(\frac{\log x-\mu}{\sigma}\right) + \frac{\lambda_3^2}{72\sigma^6}H_6\left(\frac{\log x-\mu}{\sigma}\right)\right]\Lambda(x;\mu,\sigma) + \mathrm{O}\left(\frac{1}{r^{3/2}}\right) \tag{42}$$



(a) (b)

Fig. 1. PDF approximations of known distributions produced by seven methods, with the series expansion methods corrected for $\kappa_{X,3}$ and $\kappa_{X,4}$. (a) K distribution with parameters $\mu = 1$, $L = 16$, $M = 10$, (b) G^0 distribution with parameters $g = 2$, $L = 4$, $M = -2$.

on their kernels, but they are not able to attain the accuracy of the three-parameter GΓD when only corrected for $\kappa_{X,3}$ and $\kappa_{X,4}$. The K distribution is of course exact in this case, in the sense that any deviation is solely the result of computational inaccuracies, e.g. numerical iterative solutions terminated after achieving some predefined accuracy. In the following, we will disregard such technicalities, instead stating the solutions as exact.

Fig. 1 (b) represents a more challenging case with a heavier tail. The MKGK series is ill-suited to this case, but the MKLK and MKE series outperform their kernel and even the MoLC K and GΓD methods.

In Table II we show the results of experiments where we have again used the K and G^0 distributions as targets, but also included the gamma distribution from (24), the inverse gamma distribution [14]

$$\gamma^{-1}(x; a, b) = \frac{b^a x^{-a-1}}{\Gamma(a)} e^{-\frac{b}{x}}, \tag{45}$$

where $x, b > 0,\ a > 0$, and the generalized gamma distribution [38], [39], [7]

$$\mathrm{G\Gamma D}(x; a, b, d) = \frac{|d|b}{\Gamma(a)} (bx)^{ad-1} \exp\{-(bx)^d\}, \tag{46}$$

where $x, b > 0$, $a > 0$, $d \neq 0$. We only present the Kullback-Leibler distance in this case as the three dissimilarity measures are consistent, with $d_{KL}(\cdot)$ having the best contrast when the approximations were highly accurate.

From Table II, we can see that the series expansions improve on their kernel, and with only two correcting terms they are competing with the flexible and accurate MoLC K and GΓD methods. Disregarding distributions which are exact matches, the series expansions prove the most accurate for 5 of the 10 distributions tested, with the standout performers being the MoLC GΓD method and the MKE series. These two methods exhibit flexibility in the sense that they successfully model all our target distribution. This property is remarkable for a method as fast as the MKE series.

B. *Convergence of the Novel Series Expansion Methods*

We will now examine if and how the MKLK, MKE and MKGK series converge to the true PDF as we correct for successively higher order log-cumulants.

We found it necessary to present both $d_B(\cdot)$ and $d_{KL}(\cdot)$ in Fig. 2, as the measures are in discord in cases (b) and (c). Still, we clearly see that the MKE series converges while the MKLK and MKGK series are less well-behaved. This closely resembles the convergence properties of the classical Gram-Charlier and Edgeworth series, which were examined in [9].

Compared to Table II, we see that the MKE series provides a better approximation to $\mathcal{K}(x; \mu = 1; L = 16, M = 10)$ than the top performer MoLC GΓD when corrected for $\kappa_{X,5}$ and beyond. The same is the case for $\gamma(x; a = 4, b = 2)$ and when correcting for at least $\kappa_{X,7}$ in $\gamma^{-1}(x; a = 16, b = 2)$ (these cases were not included in Fig. 2). For $\mathcal{K}(x; \mu = 2; L = 4, M = 2)$ and $\gamma^{-1}(x; a = 4, b = 2)$, none of the series exceeded the MoLC GΓD when limited to corrections up to $\kappa_{X,8}$.

It is clear that the MKE series in particular is a very attractive alternative when approximating the known PDFs of non-negative RVs.

V. Synthetic Data Experiments

In this section, the scenario is that the true distributions are *not* known, that is, we must estimate the distribution cumulants and log-cumulants by replacing them with the corresponding empirical entities. The parameters are estimated using the MoLC.

TABLE II
COMPARISON OF PDF APPROXIMATION METHODS FOR DIFFERENT DISTRIBUTIONS, KULLBACK-LEIBLER DISTANCE TO THE TRUE PDF, SERIES EXPANSIONS CORRECTED FOR $\kappa_{X,3}$ AND $\kappa_{X,4}$, BEST METHOD IN BOLD

	$\mathcal{K}(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[μ, L, M]		$G^0(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[g, L, M]		$\gamma(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[a, b]		$\gamma^{-1}(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[a, b]		GΓD$(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[a, b, d]	
$\boldsymbol{\rho}$	[1, 16, 10]	[2, 4, 2]	[1, 16, −10]	[2, 4, −2]	[4, 2]	[16, 8]	[4, 2]	[16, 8]	[4, 2, 2]	[16, 8, 2]
MoLC K [6]	Exact	Exact	$6.67{\cdot}10^{-3}$	$6.78{\cdot}10^{-2}$	$\mathbf{1.02{\cdot}10^{-3}}$	$5.40{\cdot}10^{-4}$	$4.20{\cdot}10^{-2}$	$1.06{\cdot}10^{-2}$	$7.88{\cdot}10^{-3}$	Failure
MoLC GΓD [7]	$\mathbf{3.51{\cdot}10^{-7}}$	$\mathbf{2.55{\cdot}10^{-5}}$	$9.29{\cdot}10^{-5}$	$1.87{\cdot}10^{-3}$	Exact	Exact	$\mathbf{4.11{\cdot}10^{-6}}$	$\mathbf{1.05{\cdot}10^{-7}}$	Exact	Exact
MKLK, kernel only	$3.38{\cdot}10^{-3}$	$1.68{\cdot}10^{-2}$	$6.84{\cdot}10^{-4}$	$5.92{\cdot}10^{-3}$	$1.83{\cdot}10^{-2}$	$4.69{\cdot}10^{-3}$	$1.85{\cdot}10^{-2}$	$4.74{\cdot}10^{-3}$	$1.91{\cdot}10^{-2}$	$4.75{\cdot}10^{-3}$
MKLK series	$8.44{\cdot}10^{-5}$	$3.00{\cdot}10^{-3}$	$6.52{\cdot}10^{-6}$	$1.11{\cdot}10^{-3}$	$2.61{\cdot}10^{-3}$	$1.45{\cdot}10^{-4}$	$2.29{\cdot}10^{-3}$	$1.30{\cdot}10^{-4}$	$2.56{\cdot}10^{-3}$	$1.40{\cdot}10^{-4}$
MKE series	$6.26{\cdot}10^{-6}$	$6.79{\cdot}10^{-4}$	$\mathbf{3.96{\cdot}10^{-6}}$	$\mathbf{9.33{\cdot}10^{-4}}$	$3.94{\cdot}10^{-3}$	$\mathbf{1.31{\cdot}10^{-5}}$	$7.82{\cdot}10^{-3}$	$1.32{\cdot}10^{-5}$	$6.40{\cdot}10^{-3}$	$\mathbf{1.30{\cdot}10^{-5}}$
MKGK, kernel only	$2.74{\cdot}10^{-3}$	$1.00{\cdot}10^{-2}$	$1.88{\cdot}10^{-2}$	$9.48{\cdot}10^{-2}$	Exact	Exact	$7.29{\cdot}10^{-2}$	$1.89{\cdot}10^{-2}$	$4.81{\cdot}10^{-3}$	$1.75{\cdot}10^{-3}$
MKGK series	$8.57{\cdot}10^{-4}$	$1.46{\cdot}10^{-1}$	$3.50{\cdot}10^{-3}$	$1.99{\cdot}10^{0}$	Exact	Exact	$5.68{\cdot}10^{-2}$	$2.26{\cdot}10^{-3}$	$\mathbf{7.20{\cdot}10^{-4}}$	$4.72{\cdot}10^{-4}$

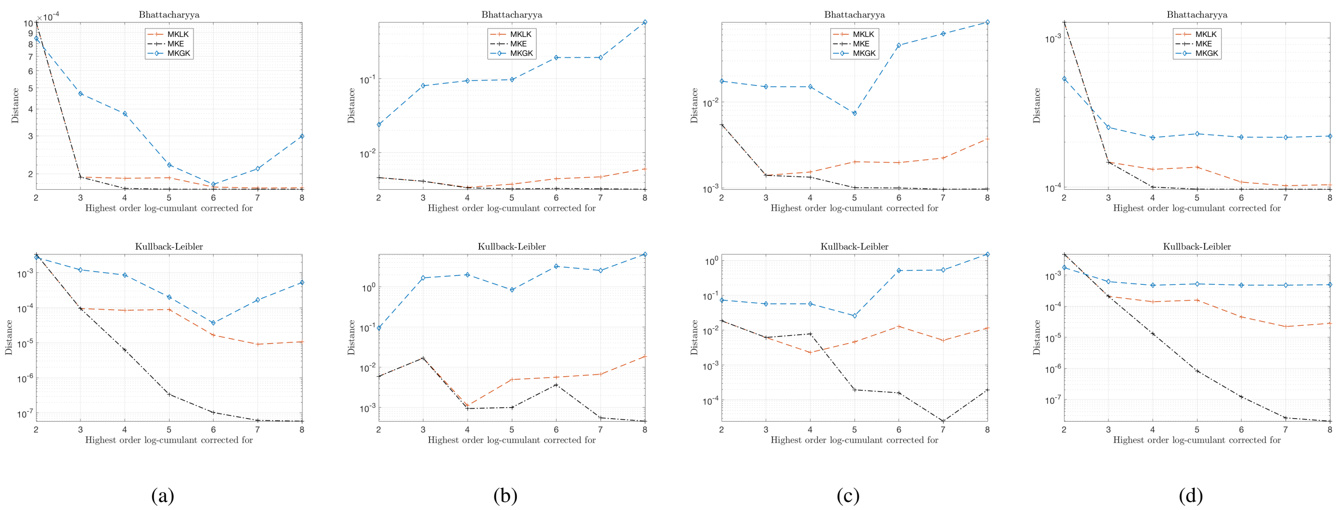


Fig. 2. Bhattacharyya distance (top) and Kullback-Leibler distance (bottom) to the true distribution as a function of the number of terms in the series expansion approximations. True distributions (a) $\mathcal{K}(x;\mu=1;L=16,M=10)$, (b) $G^0(x;g=2,L=4,M=-2)$, (c) $\gamma^{-1}(x;a=4,b=2)$, (d) GΓD$(x;a=16,b=8,d=2)$.

A. Broad Comparison Based on Data

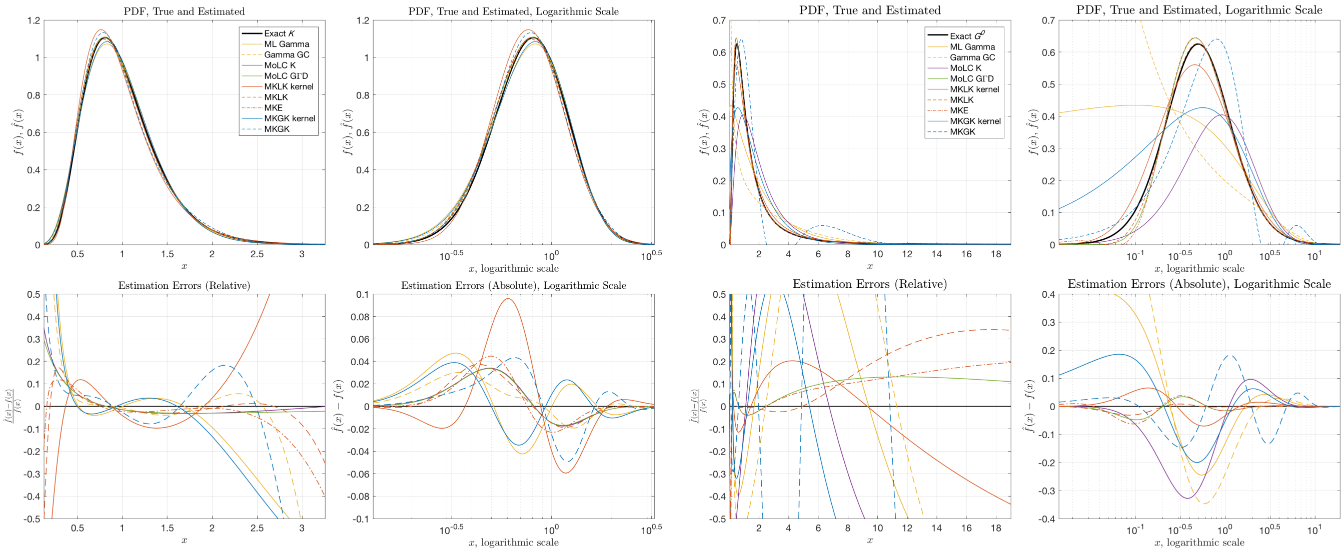
We start with a broad comparison as in Section IV-A. In addition to the seven methods used there, we also compare the gamma distribution fitted with the maximum likelihood estimates of the parameters [4] and the classical Gram-Charlier series with a gamma kernel, which is a series expansion using the classical (empirical) cumulants with the gamma kernel tailored using the method of moments [5].

In Fig. 3 (a), all methods proved reasonably successful, with the MKLK and MKE series outperforming their log-normal kernel to compete with the more advanced and computationally demanding MoLC K and GΓD estimates. Fig. 3 (b) demonstrates a more challenging scenario, due to its heavier tail. The MKGK series diverged, while the MKLK and MKE series were again among the best. In a direct comparison with Fig. 1, we see significantly higher errors when the distribution is unknown.

It should be noted that the MoLC GΓD method occasionally failed in testing: The closed form expressions used to estimate the parameters sometimes failed due to arguments of a square root being negative. This is inherent to the novel approach presented in [7], which provides a quick and relatively simple way of estimating the parameters in the flexible GΓD, that is to say that the issues are not nearly severe enough to dismiss the method. Here, we simply check the arguments to the square root and set them to zero when necessary to get an estimate. The MoLC K model had similar problems, leading to up to 5 of the 1000 estimates being discarded.

In Table III, the results from Fig. 3 are tabulated, along with several other experiments corresponding to other underlying distributions. The series expansion methods were only corrected for κ_3, κ_4 (two terms). We present the Kullback-Leibler distance since all dissimilarity measures were for the most part in accordance in this situation. An exception is that in some of the cases where the MKLK series exhibited slightly lower $d_{KL}(\cdot)$ than the MKE series, the latter had the lowest $d_B(\cdot)$ of the two, essentially implying that the MKLK and MKE series performed evenly in this scenario.

From Table III we can draw the conclusion that, not surprisingly, the best results are achieved when making accurate assumptions about the data model. E.g., for K distributed data, the fitted $\mathcal{K}(\cdot)$ outperforms the other methods. Likewise, for gamma-distributed data, the simple gamma models outperformed the more complex models, and we recall that the



(a) (b)

Fig. 3. PDF estimates produced by 9 methods, based on 1000 synthesized data points, with the series expansion methods corrected for κ_3, κ_4. (a) K distribution with parameters $\mu = 1$, $L = 16$, $M = 10$, (b) G^0 distribution with parameters $g = 2$, $L = 4$, $M = -2$.

TABLE III
COMPARISON OF PDF ESTIMATION METHODS FOR DIFFERENT TYPES OF DATA, KULLBACK-LEIBLER DISTANCES TO THE TRUE PDF, SERIES EXPANSIONS CORRECTED FOR $\kappa_{X,3}$ AND $\kappa_{X,4}$,[A] BEST METHOD FOR EACH TYPE OF DATA IN BOLD

	$\mathcal{K}(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[μ, L, M]		$G^0(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[g, L, M]		$\gamma(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[a, b]		$\gamma^{-1}(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[a, b]		GΓD$(x;\boldsymbol{\rho})$, $\boldsymbol{\rho}$=[a, b, d]	
$\boldsymbol{\rho}$	[1, 16, 10]	[2, 4, 2]	[1, 16, −10]	[2, 4, −2]	[4, 2]	[16, 8]	[4, 2]	[16, 8]	[4, 2, 2]	[16, 8, 2]
Fitted $\gamma(\cdot)$ [4]	$3.72{\cdot}10^{-3}$	$1.02{\cdot}10^{-2}$	$2.04{\cdot}10^{-2}$	$1.35{\cdot}10^{-1}$	$\mathbf{9.55{\cdot}10^{-4}}$	$\mathbf{9.85{\cdot}10^{-4}}$	$7.90{\cdot}10^{-2}$	$2.03{\cdot}10^{-2}$	$5.60{\cdot}10^{-3}$	$2.17{\cdot}10^{-3}$
$\gamma(\cdot)$ expansion [5]	$3.78{\cdot}10^{-3}$	$2.06{\cdot}10^{-2}$	$7.69{\cdot}10^{-2}$	$5.72{\cdot}10^{-1}$	$1.71{\cdot}10^{-3}$	$1.95{\cdot}10^{-3}$	$1.30{\cdot}10^{0}$	$5.54{\cdot}10^{-2}$	$3.95{\cdot}10^{-3}$	$2.77{\cdot}10^{-3}$
MoLC K [6]	$\mathbf{1.13{\cdot}10^{-3}}$	$\mathbf{1.37{\cdot}10^{-3}}$	$7.76{\cdot}10^{-3}$	$1.19{\cdot}10^{-1}$	$2.23{\cdot}10^{-3}$	$1.64{\cdot}10^{-3}$	$4.60{\cdot}10^{-2}$	$1.17{\cdot}10^{-2}$	$8.77{\cdot}10^{-3}$	Failure
MoLC GΓD [7]	$1.46{\cdot}10^{-3}$	$1.64{\cdot}10^{-3}$	$\mathbf{1.56{\cdot}10^{-3}}$	$\mathbf{3.48{\cdot}10^{-3}}$	$1.59{\cdot}10^{-3}$	$1.49{\cdot}10^{-3}$	$\mathbf{1.51{\cdot}10^{-3}}$	$\mathbf{1.43{\cdot}10^{-3}}$	$\mathbf{1.67{\cdot}10^{-3}}$	$\mathbf{1.48{\cdot}10^{-3}}$
MKLK, kernel only	$4.39{\cdot}10^{-3}$	$1.78{\cdot}10^{-2}$	$1.65{\cdot}10^{-3}$	$6.90{\cdot}10^{-3}$	$1.93{\cdot}10^{-2}$	$5.68{\cdot}10^{-3}$	$1.95{\cdot}10^{-2}$	$5.74{\cdot}10^{-3}$	$2.01{\cdot}10^{-2}$	$5.74{\cdot}10^{-3}$
MKLK series	$2.16{\cdot}10^{-3}$	$8.15{\cdot}10^{-3}$	$2.69{\cdot}10^{-3}$	$4.73{\cdot}10^{-3}$	$5.43{\cdot}10^{-3}$	$2.36{\cdot}10^{-3}$	$5.45{\cdot}10^{-3}$	$3.11{\cdot}10^{-3}$	$6.69{\cdot}10^{-3}$	$2.63{\cdot}10^{-3}$
MKE series	$2.34{\cdot}10^{-3}$	$7.17{\cdot}10^{-3}$	$2.62{\cdot}10^{-3}$	$6.81{\cdot}10^{-3}$	$9.75{\cdot}10^{-3}$	$2.89{\cdot}10^{-3}$	$1.34{\cdot}10^{-2}$	$3.72{\cdot}10^{-3}$	$1.23{\cdot}10^{-2}$	$3.18{\cdot}10^{-3}$
MKGK, kernel only	$3.78{\cdot}10^{-3}$	$1.12{\cdot}10^{-2}$	$1.98{\cdot}10^{-2}$	$9.63{\cdot}10^{-2}$	$1.01{\cdot}10^{-3}$	$1.01{\cdot}10^{-3}$	$7.41{\cdot}10^{-2}$	$1.99{\cdot}10^{-2}$	$5.84{\cdot}10^{-3}$	$2.55{\cdot}10^{-3}$
MKGK series	$8.03{\cdot}10^{-3}$	$2.23{\cdot}10^{-1}$	$9.34{\cdot}10^{-3}$	$2.02{\cdot}10^{0}$	$2.28{\cdot}10^{-2}$	$3.41{\cdot}10^{-3}$	$1.64{\cdot}10^{-1}$	$5.03{\cdot}10^{-3}$	$1.04{\cdot}10^{-2}$	$2.47{\cdot}10^{-3}$

[A] The $\gamma(\cdot)$ expansion was naturally corrected for the corresponding linear cumulants instead.

MKGK kernel is the MoLC estimate of $\gamma(\cdot)$, while the "fitted $\gamma(\cdot)$" uses the maximum likelihood estimates. This is to say that when we synthesize data from complicated distributions in the following, one must keep in mind that if the data in reality follows a simpler model, then the simple methods should be tried first.

The MKLK and MKE series are very often among the top performers, indicating that they are good candidates when it is difficult to make assumptions on the data. However, their results are not as strong as they were in Section IV, presumably since they require the estimation of 4 quantities (the log-cumulants of order 1 through 4). In the following sections, we vary the number of terms used and observations available.

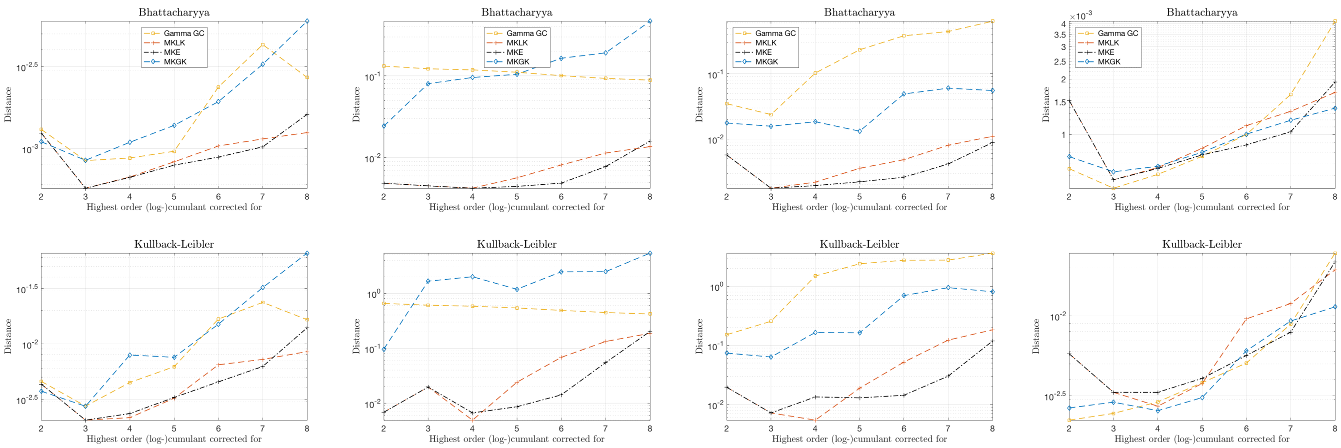
B. The Impact of the Number of Terms

Here we examine series expansion methods (the classical Gram-Charlier series with a gamma kernel [5], the MKGK, MKLK and MKE series), with respect to their performance as the number of correction terms are varied. In Fig. 4 we present the same four distributions as in Fig. 2 and correct for up to the 8th order (log-)cumulant. We performed the same computations for all 10 distributions in Table III, but the results were too similar to warrant presenting them all.

Clearly, both gamma kernel series are prone to divergence, only performing well in the case of GΓD data in (d). The MKLK and MKE fared better, with the latter outperforming the former when correcting for between $\kappa_{X,5}$ and $\kappa_{X,7}$, but not otherwise. The contrast with Fig. 2 is clear: The benefit of additional corrections does not outweigh the error introduced by having to correct for additional log-cumulants with this number of data points.

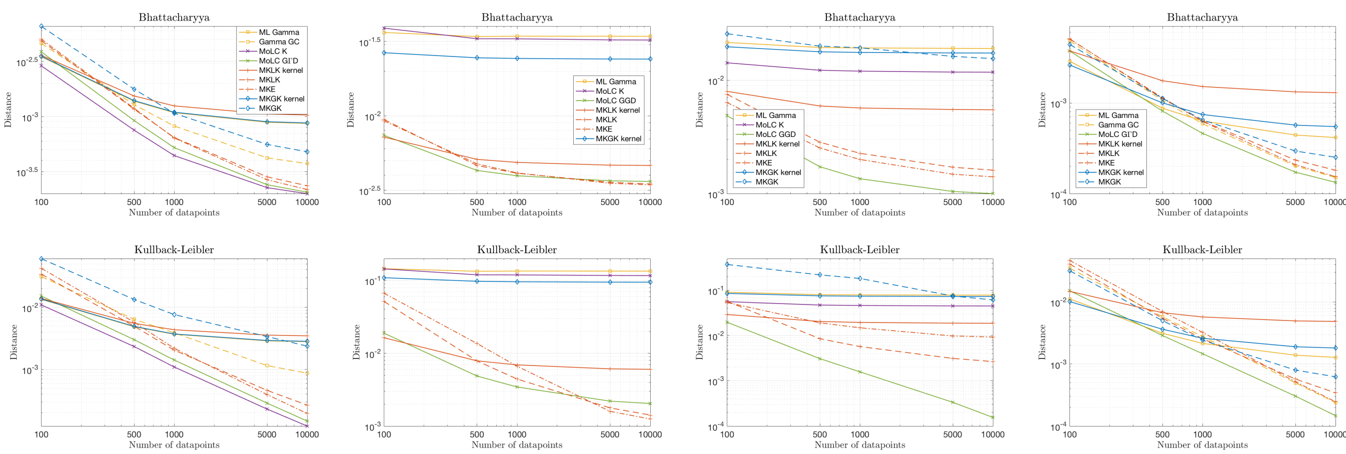
C. The Impact of the Number of Observations

This final analysis is concerned with how the performance of the methods depends on the number of data points (obser-



(a) (b) (c) (d)

Fig. 4. Bhattacharyya distance (top) and Kullback-Leibler distance (bottom) to the true distribution as a function of the number of terms in the series expansion estimates. Mean of 1000 iterations with 1000 synthesized data points. True distributions (a) $\mathcal{K}(x; \mu = 1, L = 16, M = 10)$, (b) $G^0(x; g = 2, L = 4, M = -2)$, (c) $\gamma^{-1}(x; a = 4, b = 2)$, (d) G$\Gamma$D$(x; a = 16, b = 8, d = 2)$.



(a) (b) (c) (d)

Fig. 5. Bhattacharyya distance (top) and Kullback-Leibler distance (bottom) to the true distribution as a function of the number of data points. Mean of 1000 iterations, series expansion methods corrected with two terms as in Fig. 3. True distributions (a) $\mathcal{K}(x; \mu = 1, L = 16, M = 10)$, (b) $G^0(x; g = 2, L = 4, M = -2)$, (c) $\gamma^{-1}(x; a = 4, b = 2)$, (d) G$\Gamma$D$(x; a = 16, b = 8, d = 2)$. The classical Gram-Charlier gamma kernel series was omitted from (b) and (c) for readability, as it was divergent (much higher distances than the others), and the MKGK series was also left out from (b) for the same reason. The MoLC K method failed in (d), as in Tables II and III.

vations). Specifically: Should the quantity of data impact our choice of PDF estimate? For the series expansion methods we also ask whether vast amounts of data permit more terms.

Fig. 5 presents the same distributions as Fig. 4, but fixed to two correcting terms and with the number of observations now varying from 100 to 10000. The series expansion methods benefit more from the increase in observations, which is not surprising as the (log-)cumulants used in the corrections are in fact estimated themselves. Especially the methods based on the log-normal kernel demonstrate their value as they approach the accuracy of the MoLC GΓD method. We also see that the MKE series benefits even more from an increase in the quantity of the data then the MKLK series, presumably because its second correcting term also accounts for $\kappa_{X,3}^2$, i.e. it is more complex.

Our final investigation seeks to shed light on the practical question of whether there is an ideal number of correcting terms for a given number of data points, and if this also depends on the nature of the data (true distribution) at hand.

In Table IV we present the best (lowest distance) number of correction terms in the series expansion methods for the distributions in Figures 5 and 2, when the number of observations is varied. Clearly, more data points allows for more terms, as expected. In fact, it is hard to justify compensating for more than κ_3 (conceptually the logarithmic skewness), unless we have very many data points. We recall that with

TABLE IV
THE HIGHEST ORDER (LOG-)CUMULANT WHICH SHOULD BE CORRECTED FOR[A], WITH RESPECT TO THE NUMBER OF OBSERVATIONS FOR DIFFERENT TYPES OF DATA, BEST METHOD FOR EACH DISTRIBUTION IN BOLD

	$\mathcal{K}(x;\mu{=}1, L{=}16, M{=}10)$			$G^0(x;g{=}2, L{=}4, M{=}-2)$			$\gamma^{-1}(x;a{=}4, b{=}2)$			$\mathrm{G\Gamma D}(x;a{=}16, b{=}8, d{=}2)$		
Obs.	10^2	10^3	10^4	10^2	10^3	10^4	10^2	10^3	10^4	10^2	10^3	10^4
Fitted $\gamma(\cdot)$ [4]	κ_2	κ_3, κ_4	κ_4, κ_6	κ_6	κ_8	κ_3	κ_3, κ_5	κ_2, κ_3	κ_2, κ_3	κ_2	$\boldsymbol{\kappa_2, \kappa_3}$	κ_4
MKLK series	κ_2	$\boldsymbol{\kappa_3}$	κ_3, κ_4	$\boldsymbol{\kappa_2}$	$\boldsymbol{\kappa_4}$	κ_4	$\boldsymbol{\kappa_3}$	$\boldsymbol{\kappa_3, \kappa_4}$	κ_3, κ_4	κ_2	κ_3, κ_4	κ_4
MKE series	κ_2	$\boldsymbol{\kappa_3}$	$\boldsymbol{\kappa_4}$	$\boldsymbol{\kappa_2}$	κ_4	$\boldsymbol{\kappa_4, \kappa_5}$	$\boldsymbol{\kappa_3}$	$\boldsymbol{\kappa_3}$	$\boldsymbol{\kappa_5}$	κ_2	κ_3	$\boldsymbol{\kappa_4}$
MKGK series	$\boldsymbol{\kappa_2}$	κ_3	κ_5	κ_2	κ_2	κ_2	κ_2	κ_3, κ_5	κ_3, κ_5	$\boldsymbol{\kappa_2}$	κ_3, κ_4	κ_4

[A] We take this to mean the correction which results in the lower Bhattacharyya and Kullback-Leibler distance to the true distribution, based on the mean of 1000 iterations. In cases where the results are very similar or $d_B(\cdot)$ and $d_{KL}(\cdot)$ disagree, we have given both log-cumulants.

this lone correction term, the MKLK and MKE series coincide. Supporting our remarks on Fig. 5, we see that the MKE series benefits the most from an increase in observations, becoming the top performer at $10,000$ data points in all four cases. Finally, we note that [17] commented that estimation of the linear cumulants of order > 4 was unreliable due to sample fluctuations. The present findings indicate that the same can be said in the logarithmic case.

VI. CONCLUSION

We have shown how the classical Gram-Charlier and Edgeworth series have strong theoretical and practical analogies in the logarithmic domain, derived using the MT and MKS. We have introduced the Mellin derivatives in the context of MKS, providing a useful (and in this case necessary) way to retrieve the PDF via the inverse MT on the MKCF. The Bell polynomials have also been used in a new way, providing a simpler representation of the MKE series. The Mellin kind Gram-Charlier series expansion with arbitrary kernel indicates that there are undiscovered methods within the presented framework.

When approximating known distributions, we have shown how the Mellin kind series mirrors the performance of their classical counterparts [9], with the MKE series converging in a predictable manner over a range of different distributions, unlike the MKLK and MKGK series. These methods, and the MKE series in particular, are attractive alternatives which defy their simplicity to compete, often with relatively few correction terms, with state-of-the-art methods such as the GΓD and K distributions with parameter estimates computed with the MoLC. Unlike these more complicated methods, the series expansions were completely reliable in the sense that they never failed to produce an estimate throughout our testing.

In the more realistic situation where the parameters and log-cumulants of an unknown distribution must be estimated, the picture is not so clear. Again, the series around the log-normal kernel were the stand-out performers, but the cost of added complexity usually outweighed the benefit of correcting for log-cumulants beyond the logarithmic skewness κ_3. At that point, the MKLK and MKE series coincide as the log-normal PDF kernel corrected for the empirical logarithmic skewness. When the amount of data points is very high, further corrections can have merit.

APPENDIX A
OBSERVATIONS ON THE $M_n(x)$ POLYNOMIALS

A. $M_n(x)$ as a Linear Combination of Laguerre Polynomials

Lemma 1: The polynomials $M_n(x)$ defined in (26) are linear combinations of the generalized Laguerre polynomials,

$$M_n(x) = \sum_{k=0}^{n} \left\{ {n \atop k} \right\} (-1)^k k! L_k^{(a-1)}(x), \tag{47}$$

where $\left\{ {n \atop k} \right\}$ denotes the Stirling numbers of the second kind [40]

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} i^n, \tag{48}$$

which is the number of possible ways to partition n labelled objects into k non-empty and unlabelled subsets.

Proof: Starting with an identity regarding $\mathrm{D}_x^k x^k$, see that

$$\mathrm{D}_x^k x^k f(x) = \mathrm{D}_x^{k-1} x^{k-1} [(k + x\mathrm{D}_x) f(x)], \tag{49}$$

and by repetition we get

$$\mathrm{D}_x^k x^k f(x) = (x\mathrm{D}_x + k)_k f(x). \tag{50}$$

Using the fact that $\mathrm{D}_x x - x\mathrm{D}_x = 1$, we have

$$\mathrm{D}_x^k x^k f(x) = (\mathrm{D}_x x + k - 1)_k f(x). \tag{51}$$

By a property of the Stirling numbers [41], the Mellin derivative from Table I can be rewritten as

$$(-\mathrm{D}_x x)^n = \sum_{k=0}^{n} \left\{ {n \atop k} \right\} (-1)^k (\mathrm{D}_x x + k - 1)_k = \sum_{k=0}^{n} \left\{ {n \atop k} \right\} (-1)^k \mathrm{D}_x^k x^k, \tag{52}$$

and multiplying both sides with the unit scale gamma distribution $\gamma(x;a)$, the definitions of $M_n(x)$ and $L_k^{(a)}(x)$ [33] are recognized on the left and right hand sides, respectively, of

$$M_n(x) = \sum_{k=0}^{n} \left\{ {n \atop k} \right\} (-1)^k k! L_k^{(a-1)}(x). \tag{53}$$

Finally, we note that $M_n(x)$ is a nth degree polynomial.

B. The Leading Coefficient of $M_n(x)$

Lemma 2: Writing $M_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n$, the leading coefficient $a_n = 1$.

Proof: The only term containing x^n in (47) is $L_n^{(a-1)}(x)$, as the nth Laguerre polynomial is degree n. $L_n^{(a-1)}(x)$ has leading coefficient $(-1)^n/n!$, giving

$$a_n = \left\{ {n \atop n} \right\} (-1)^n n! \frac{(-1)^n}{n!} = 1, \tag{54}$$

where it was used that $\left\{ {n \atop n} \right\} = 1 \, \forall \, n$.

C. The First Few $M_n(x)$ Polynomials

$$M_0(x) = 1 \tag{55}$$
$$M_1(x) = x - a \tag{56}$$
$$M_2(x) = x^2 - (2a+1)x + a^2 \tag{57}$$
$$M_3(x) = x^3 - 3(a+1)x^2 + (3a^2 + 3a + 1)x - a^3 \tag{58}$$

Appendix B
Logarithmic Hermite Polynomials

Lemma 3:

$$(-\mathrm{D}_x x)^n \Lambda(x; \mu, \sigma) = \frac{1}{\sigma^n} H_n \left(\frac{\log x - \mu}{\sigma} \right) \Lambda(x; \mu, \sigma), \tag{59}$$

where $H_n(\cdot)$ is the nth probabilists' Hermite polynomials defined in terms of the standardized (zero mean, unit variance) Gaussian kernel $\alpha(x) = (2\pi)^{-1/2} e^{-x^2/2}$ as [17]

$$(-\mathrm{D}_x)^n \alpha(x) = H_n(x) \alpha(x). \tag{60}$$

Proof: We can use the chain rule to see that

$$\mathrm{D}_{\log x} = \frac{\mathrm{d}}{\mathrm{d} \log x} = \frac{\mathrm{d}x}{\mathrm{d} \log x} \frac{\mathrm{d}}{\mathrm{d}x} = x \frac{\mathrm{d}}{\mathrm{d}x} = x \mathrm{D}_x, \tag{61}$$

where x is to the left of D_x since it is should be multiplied with the differentiated function. The standardized log-normal and Gaussian PDFs are related by

$$\Lambda(x) = \frac{1}{\sqrt{2\pi} x} e^{-\frac{(\log x)^2}{2}} = \frac{1}{x} \alpha(\log x), \tag{62}$$

giving

$$(-\mathrm{D}_x x)^n \Lambda(x) = (-1)^n \mathrm{D}_x x \cdots \mathrm{D}_x x \frac{1}{x} \alpha(\log x) \tag{63}$$
$$= (-1)^n \frac{1}{x} x \mathrm{D}_x x \cdots \mathrm{D}_x \alpha(\log x) \tag{64}$$
$$(-\mathrm{D}_x x)^n \Lambda(x) = \frac{1}{x} (-x \mathrm{D}_x)^n \alpha(\log x). \tag{65}$$

Now we can complete the proof of Lemma 3 by replacing x with $\log x$ in (60) and using (61) and (65) to get

$$\left(-\frac{\mathrm{d}}{\mathrm{d} \log x} \right)^n \alpha(\log x) = H_n(\log x) \alpha(\log x) \tag{66}$$
$$(-\mathrm{D}_x x)^n \Lambda(x) = \frac{1}{x} H_n(\log x) \alpha(\log x) \tag{67}$$
$$(-\mathrm{D}_x x)^n \Lambda(x) = H_n(\log x) \Lambda(x). \tag{68}$$

The final part of the proof is to generalize the result to arbitrary log-mean μ and log-variance σ^2. Letting $\log u = (\log x - \mu)/\sigma$, the relationship between the standardized and non-standardized log-normal PDF is

$$\Lambda(x; \mu, \sigma) = \frac{u}{x\sigma} \Lambda(u), \tag{69}$$

giving

$$(-\mathrm{D}_x x)^n \Lambda(x; \mu, \sigma) = (-\mathrm{D}_x x)^n \frac{u}{x\sigma} \Lambda(u) = \frac{1}{x} (-x \mathrm{D}_x)^n \frac{u}{\sigma} \Lambda(u), \tag{70}$$

but

$$(-x \mathrm{D}_x)^n = (-u \mathrm{D}_u)^n \frac{1}{\sigma^n}, \tag{71}$$

so

$$(-\mathrm{D}_x x)^n \Lambda(x; \mu, \sigma) = \frac{u}{x \sigma^{n+1}} (-\mathrm{D}_u u)^n \Lambda(u). \tag{72}$$

We can now use (68) and (69) to finalize the proof of Lemma 3, by reinserting for u to get

$$(-\mathrm{D}_x x)^n \Lambda(x; \mu, \sigma) = \frac{u}{x \sigma^{n+1}} H_n(\log u) \Lambda(u) \tag{73}$$
$$= \frac{1}{\sigma^n} H_n \left(\frac{\log x - \mu}{\sigma} \right) \Lambda(x; \mu, \sigma). \tag{74}$$

Appendix C
The Mellin Kind Edgeworth Assumption

Lemma 4: Assuming that the logarithm of X is the standardized sum

$$\log X = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} \frac{Z_i - m}{\varsigma}, \tag{75}$$

where $Z_1, Z_2, \ldots, Z_r$ are as in Section II-A, then the log-cumulants of X are

$$\kappa_{X,\nu} = \begin{cases} 0 & \nu = 1 \\ 1 & \nu = 2 \\ \frac{\lambda_\nu}{r^{\frac{\nu}{2}-1}} & \nu \geq 3 \end{cases}. \tag{76}$$

Proof: Since both quantities are defined as $\mathrm{E}\{(\log X)^\nu\}$, the log-moments of X are equal to the moments of $\log X$. The relations between the log-moments and log-cumulants are identical to those between their classical counterparts [14], so the log-cumulants of X must also equal the cumulants of $\log X$. We already stated the cumulants of order $\nu \geq 3$ of the standardized sum in (8), and clearly the log-cumulants of X equals these.

In general, $\kappa_1 = \mu_1 = \mu$ and $\kappa_2 = \mu_2 - \mu_1^2 = \sigma^2$ [14], but the standardized sum trivially has zero mean, unit variance, i.e. X has zero log-mean, unit log-variance and the proof is complete.

References

[1] G. Pastor, I. Mora-Jiménez, A. J. Caamano, and R. Jäntti, "Log-cumulants-based Edgeworth expansion for skew-distributed aggregate interference," *ISWCS, IEEE*, pp. 390–394, 2014.

[2] K. N. Boyadzhiev, "Exponential polynomials, Stirling numbers, and evaluation of some gamma integrals," *Abstr. Appl. Anal.*, 2009.

[3] J. Pitman, "Combinatorial stochastic processes," Technical Report 621, Dept. Statistics, UC Berkeley, 2002.

[4] S. C. Choi and R. Wette, "Maximum likelihood estimation of the parameters of the gamma distribution and their bias," *Technometrics*, vol. 11, no. 4, pp. 683–690, 1969.

[5] E. Gaztanaga, P. Fosalba, and E. Elizalde, “Gravitational evolution of the large-scale probability density distribution: The Edgeworth and gamma expansions,” *Astrophys. J.*, vol. 539, no. 2, p. 522, 2000.

[6] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, “On the method of logarithmic cumulants for parametric probability density function estimation,” *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3791–3806, 2013.

[7] H.-C. Li, W. Hong, Y.-R. Wu, and P.-Z. Fan, “On the empirical-statistical modeling of sar images with generalized gamma distribution,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 386–397, 2011.

[8] A. Hald, “The early history of the cumulants and the Gram-Charlier series,” *Int. Stat. Rev.*, vol. 68, no. 2, pp. 137–153, 2000.

[9] S. Blinnikov and R. Moessner, “Expansions for nearly Gaussian distributions,” *Astron. Astrophys. Suppl.*, vol. 130, no. 1, pp. 193–205, 1998.

[10] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. SciTech Publishing, 2004.

[11] X. Deng, C. López-Martínez, and E. M. Varona, “A physical analysis of polarimetric SAR data statistical models,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3035–3048, 2016.

[12] R. J. Adler, R. E. Feldman, and M. S. Taqqu, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser, 1998.

[13] S. Foss, D. Korshunov, and S. Zachary, *An Introduction to Heavy-Tailed and Subexponential Distributions*, ser. Springer Series in Operations Research and Financial Engineering. Springer, 2011, vol. 38.

[14] J.-M. Nicolas, “Introduction aux statistiques de deuxième espèce: applications de logs-moments et des logs-cumulats à l’analyse de lois d’images radar,” *Trait. Signal*, vol. 19, no. 3, pp. 139–167, 2002.

[15] S. N. Anfinsen and T. Eltoft, “Application of the matrix-variate Mellin transform to analysis of polarimetric radar images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, June 2011.

[16] G. Pastor, I. Mora-Jiménez, A. J. Caamaño, and R. Jäntti, “Asymptotic expansions for heavy-tailed data,” *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 444–448, 2016.

[17] M. Kendall, A. Stuart, and J. Ord, *Kendall’s Advanced Theory of Statistics*, 6th ed. John Wiley & Sons, Ltd, 1994, vol. 1: Distribution theory.

[18] W. Bryc, *The Normal Distribution: Characterizations With Applications*. Springer Science & Business Media, 2012.

[19] E. Lukacs, *Characteristic Functions*. Griffin, 1970.

[20] E. T. Bell, “Partition polynomials,” *Ann. Math.*, pp. 38–46, 1927.

[21] M. Mihoubi, “Bell polynomials and binomial type sequences,” *Discrete Mathematics*, vol. 308, no. 12, pp. 2450–2459, 2008.

[22] G.-C. Rota and J. Shen, “On the combinatorics of cumulants,” *J. comb. Theory A*, vol. 91, no. 1-2, pp. 283–304, 2000.

[23] C. S. Withers and S. Nadarajah, “Charlier and Edgeworth expansions for distributions and densities in terms of Bell polynomials,” *Probab. Math. Stat.*, vol. 29, no. 2, p. 271, 2009.

[24] T. Brenn and S. N. Anfinsen, “A revisit of the Gram-Charlier and Edgeworth series expansions,” UiT The Arctic University of Norway, Department of Physics and Technology, Tech. Rep., June 2017. [Online]. Available: munin

[25] A. Fitouhi, N. Bettaibi, and K. Brahim, “The Mellin transform in quantum calculus,” *Constr. Approx.*, vol. 23, no. 3, pp. 305–323, 2006.

[26] J. Bertrand, P. Bertrand, and J.-P. Ovarlez, *The Transforms and Applications Handbook*. CRC Press LLC, 2000, ch. 11, "The Mellin Transform".

[27] H. Bateman, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms*. McGraw-Hill, 1954, vol. I & II.

[28] B. Epstein, “Some applications of the Mellin transform in statistics,” *Ann. Math. Stat.*, pp. 370–379, 1948.

[29] G. Pastor, “Probabilistic models and algorithms for energy-efficient large and dense wireless sensor networks,” Ph.D. dissertation, Aalto University, Department of Communications and Networking, 2016.

[30] J.-M. Nicolas and S. N. Anfinsen, “Introduction to second kind statistics: Application of log-moments and log-cumulants to the analysis of radar image distributions,” *Technical note*, May 2012.

[31] J.-M. Nicolas, *Lois statistiques de l’imagerie cohérente. Kit de survie pour les traiteurs d’images RSO. V-β*, Télécom ParisTech, département TSI. *Technical note*, 2016.

[32] R. Askey, *Mathematics and Social Utopias in France: Olinde Rodrigues and His Times*, ser. History of Mathematics. American Mathematical Society, 2005, vol. 28, ch. 6, pp. 105–118.

[33] G. Szeg, *Orthogonal Polynomials*. American Mathematical Society, 1939, vol. 23.

[34] C. S. Withers and S. Nadarajah, “Edgeworth–Cornish–Fisher–Hill–Davis expansions for normal and non-normal limits via Bell polynomials,” *Stochastics*, vol. 87, no. 5, pp. 794–805, 2015.

[35] A. C. Frery, A. D. Nascimento, and R. J. Cintra, “Analytic expressions for stochastic distances between relaxed complex Wishart distributions,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1213–1226, 2014.

[36] A. C. Frery, H.-J. Muller, C. d. C. F. Yanasse, and S. J. S. Sant’Anna, “A model for extremely heterogeneous clutter,” *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, 1997.

[37] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, “Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3795–3809, 2009.

[38] E. W. Stacy, “A generalization of the gamma distribution,” *Ann. Math. Stat.*, pp. 1187–1192, 1962.

[39] P. R. Tadikamalla, “Random sampling from the generalized gamma distribution,” *Computing*, vol. 23, no. 2, pp. 199–203, 1979.

[40] T. Lengyel, “On the divisibility by 2 of the Stirling numbers of the second kind,” *Fibonacci Quart.*, vol. 32, no. 3, pp. 194–201, 1994.

[41] C. G. Wagner, “Generalized Stirling and Lah numbers,” *Discrete Math.*, vol. 160, no. 1, pp. 199–218, 1996.