



UIT

THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Science and Technology

Modeling Probability Density Functions of Non-Negative Random Variables Using Novel Series Expansions Based On Mellin Kind Statistics

—
Torgeir Brenn

FYS-3941, Master's Thesis in Applied Physics and Mathematics, June 2017



Contents

Abstract	i
Acknowledgements	ii
Mathematical Nomenclature	ii
Abbreviations	iii
1 Introduction	1
1.1 Motivation	1
1.2 Subjects and Goals	2
1.3 Method	2
1.4 Organization	2
1.5 Contributions of This Thesis	3
2 Theory	5
2.1 Terminology	5
2.2 Classical Statistics	6
2.2.1 The Characteristic Function	6
2.2.2 Moments	6
2.2.3 Cumulants	7
2.2.4 The Sum of Independent RVs	8
2.2.5 The Gaussian Distribution	9
2.3 Mellin Kind Statistics	10
2.3.1 The Mellin Transform	10
2.3.2 The Inverse Mellin Transform	10
2.3.3 A MT Example: Euler's Gamma Function	10
2.3.4 Some General Properties of the Mellin Transform	10
2.3.5 The Mellin Kind Characteristic Function	11
2.3.6 Log-Moments	11
2.3.7 Log-Cumulants	12
2.3.8 The Product of Independent RVs	14
2.3.9 The Gamma Distribution	14
2.3.10 The Beta Prime Distribution	16
2.3.11 The Log-Normal Distribution	17
2.4 The Classical Statistics Series Expansions	18
2.4.1 The Chebyshev-Hermite Polynomials	19
2.4.2 The Gram-Charlier Gaussian Kernel Series	20
2.4.3 The Laguerre Polynomials	21
2.4.4 The Gram-Charlier Gamma Kernel Series	22
2.4.5 The Edgeworth Series	23
2.5 The Bell Polynomials and the Stirling Numbers	25
2.5.1 Defining the Bell Polynomials	25
2.5.2 A Property of the Bell Polynomials	26
2.5.3 The Bell Polynomials and Faà di Bruno's Formula	26
2.5.4 Defining the Stirling Numbers	27
2.5.5 A Property of the Stirling Numbers	27
2.6 Dissimilarity Measures of PDFs	28
2.6.1 Terminology: Divergence, Distance and Metric	28
2.6.2 The Kullback-Leibler Distance	29

2.6.3	The Bhattacharyya Distance	29
2.6.4	Caveats Related to the Series Expansion Methods	30
3	Mellin Kind Series Expansion Framework	31
3.1	Fundamentals	31
3.1.1	The Analogy With the Classical Series Expansion Methods	31
3.1.2	A Mellin Derivative	32
3.1.3	The Mellin Kind Gram-Charlier Series With Arbitrary Kernel	32
3.1.4	The Mellin Derivative and the Arbitrary Kernel	33
3.2	The Mellin Kind Gamma Kernel Series	34
3.2.1	The Laguerre Polynomials and the Mellin Transform	34
3.2.2	The $M_n(x)$ Polynomials	34
3.2.3	Defining the Mellin Kind Gamma Kernel Series	35
3.2.4	Choosing the Kernel Parameters	36
3.2.5	The MKGK Series With Realistic Numbers of Terms	37
3.3	The Mellin Kind Log-Normal Kernel Series	38
3.3.1	The Logarithmic Hermite Polynomials	38
3.3.2	Non-Standardized Log-Normal Data	39
3.3.3	Defining the Mellin Kind Log-Normal Kernel Series	40
3.3.4	The Obvious Choices of the Kernel Parameters	40
3.3.5	The MKLK Series With Realistic Numbers of Terms	40
3.4	The Mellin Kind Edgeworth Series	41
3.4.1	The Log-Cumulant Differences	41
3.4.2	Deriving the Mellin Kind Edgeworth Series	42
3.4.3	The Work of Pastor et al.	44
3.5	The Mellin Kind Beta Prime Kernel Series	46
3.5.1	The $M'_n(\cdot)$ Polynomials	46
3.5.2	Defining the Mellin Kind Beta Prime Kernel Series	48
3.5.3	Choosing the Kernel Parameters	48
3.5.4	The MKBK Series With Realistic Numbers of Terms	49
4	The Classical Series Expansions Revisited	51
4.1	Bell Polynomials, Moments and Cumulants	51
4.2	The Gram-Charlier Series with Arbitrary Kernel	53
4.2.1	The Differential Operator and the Arbitrary Kernel	53
4.2.2	The Gaussian Kernel	54
4.2.3	The Gamma Kernel	55
4.3	The Edgeworth Series	56
4.4	Author's Comment	57
5	Results	59
5.1	Preliminaries	59
5.1.1	Method – Approximation and Estimation	59
5.1.2	The Target PDFs	60
5.1.3	Results Included in the Appendix	62
5.2	Approximation	62
5.2.1	Two Introductory Examples	62
5.2.2	Approximations to Numerous Target PDFs Tabulated	64
5.2.3	Convergence, Two Examples	66
5.2.4	Convergence, Numerous Target PDFs	67
5.2.5	Alternative Choices of the Kernel Parameters	70
5.2.6	Alternative Kernel Parameter Values in a SAR Scenario	79

5.3	Estimation	80
5.3.1	Revisiting the Introductory Examples	81
5.3.2	Estimations of Numerous Target PDFs Tabulated	82
5.3.3	Convergence and the Number of Data Points	83
5.3.4	Alternative to the MoLC Parameter Estimates	84
5.3.5	The MKBK Series Used in a SAR Change Detection Scenario	89
5.4	Experiments on Real SAR Data	92
5.4.1	Modeling SAR Image Data	92
5.4.2	Modeling in a SAR Change Detection Scenario	97
6	Conclusion	101
6.1	Conclusions about the Theoretical Contributions	101
6.1.1	The MK Series Expansion Framework	101
6.1.2	Use of the Bell Polynomials	102
6.1.3	Other Contributions to the Classical Series Expansions	103
6.2	Conclusions from the Experiments	103
6.2.1	The MKLK and MKE Series	103
6.2.2	The MKGK Series	104
6.2.3	The MKBK Series	104
6.2.4	Modeling Real Data	104
6.3	Future Work	105
A	Supplementary Theoretical Results	107
A.1	The Laguerre Polynomials Scaled	107
A.1.1	Orthogonality	107
A.1.2	Rodrigues Formula	107
A.2	Observations on $M_n(x)$	108
A.2.1	$M_n(x)$ as Linear Combinations of Laguerre Polynomials	108
A.2.2	The Leading Coefficient of $M_n(x)$	109
A.2.3	$M_n(x)$ and a Confluent Hypergeometric Function	109
A.3	The Logarithmic Hermite Polynomials	110
A.4	Program Code	111

List of Figures

5.1	Approximating a K Distribution	63
5.2	Approximating a GFD	64
5.3	Convergence Visualized	67
5.4	Convergence When Approximating Known Distributions, Four Target PDFs	69
5.5	Convergence When Approximating Known Distributions, Other Target PDFs	70
5.6	Alternative Parameter Choices, Approximating a Gamma Distribution	72
5.7	Alternative Parameter Choices, Approximating an Inverse Gamma Distribution	73
5.8	Alternative Parameter Choices, Approximating a GFD	74
5.9	Alternative Parameter Choices, Approximating a K Distribution	75
5.10	Alternative Parameter Choices, Approximating a G^0 Distribution	76
5.11	Convergence Visualized, Non-Tailored Parameter Values	79
5.12	Estimating a K Distribution	81
5.13	Estimating a GFD	82
5.14	Convergence When Estimating Unknown Distributions, Five Target PDFs	85
5.15	The MKGK Series, Parameter Estimates When Modeling K Distributions	88
5.16	The MKBK Series Used in SAR Change Detection	90
5.17	SAR Water Data	93
5.18	SAR Park Data	95

5.19	SAR Urban Data	96
5.20	SAR Change Detection Histograms	98
5.21	SAR Change Detection Convergence	99

List of Tables

2.1	Mellin Transform Properties	11
2.2	Stirling Numbers	28
5.1	Broad Comparison of the MK Series Expansions as Approximations	65
5.2	Broad Comparison of the MK Series Expansions as Estimations	83
5.3	Biased and Unbiased Log-Variance Estimators in the MKLK/MKE Series	86
5.4	MK Series Expansions Fitted to SAR Water Data	92
5.5	MK Series Expansions Fitted to SAR Park Data	94
5.6	MK Series Expansions Fitted to SAR Urban Data	97

Abstract

This master's thesis introduces a framework for series expansions based on *Mellin kind* (MK) statistics, which was introduced in [Nicolas, 2002].¹ That is, we derive the analogies to the classical Gram-Charlier and Edgeworth series, based on the log-moments and log-cumulants, which are the natural sets of descriptors in MK statistics.

We introduce the MK Gram-Charlier series with arbitrary kernel $\rho(x)$,

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \mathcal{P}_n(x) \right] \rho(x), \quad (1)$$

where $B_n(\cdot)$ is the n th complete Bell polynomial and $f_X(x)$ is the *probability density function* (PDF) of a non-negative random variable, which the series expansions seeks to approximate. $\Delta\kappa_1$ is the difference in log-cumulants between $f_X(x)$ and the kernel $\rho(x)$, which is a known PDF with certain constraints, including $\rho(x) = 0$ for $x < 0$. The functions $\mathcal{P}_n(x)$ depend on the choice of kernel. In this thesis, we present the following choices of kernel.

The Mellin Kind Gamma Kernel Series Substituting $\rho(x)$ with the gamma PDF $\gamma(x; a, b) = b^a x^{a-1} e^{-bx} / \Gamma(a)$ gives

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) M_n(x) \right] \gamma(x; a, b), \quad (2)$$

where the functions $M_n(x) = x^{-a+1} e^x (-Dx)^n [x^{a-1} e^{-x}]$ are n th degree polynomials in x , defined using the derivative operator $D \equiv d/dx$. We derive several results regarding $M_n(x)$.

The Mellin Kind Beta Prime Kernel Series Substituting $\rho(x)$ with the beta prime PDF $\beta'(x; a_1, a_2, b) = b(bx)^{a_1-1} / (B(a_1, a_2)(1+bx)^{a_1+a_2})$, where $B(a_1, a_2)$ is the beta function, gives

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) M'_n\left(\frac{bx}{1+bx}\right) \right] \beta'(x; a_1, a_2, b), \quad (3)$$

where the functions $M'_n(\cdot)$ are n th degree polynomials in $bx/(1+bx)$, given by

$$M'_n\left(\frac{bx}{1+bx}\right) = \frac{(1+bx)^{a_1+a_2}}{(bx)^{a_1-1}} (-D_x x)^n \left[\frac{(bx)^{a_1-1}}{(1+bx)^{a_1+a_2}} \right]. \quad (4)$$

The Mellin Kind Log-Normal Kernel Series Substituting $\rho(x)$ with the log-normal PDF $\Lambda(x; \mu, \sigma) = \exp\{-(\log x - \mu)^2 / (2\sigma^2)\} / (x\sigma\sqrt{2\pi})$ gives

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) H_n(\log x) \right] \Lambda(x; a, b), \quad (5)$$

where $H_n(\log x)$ is the n th Hermite polynomial. Unlike the expansions of the gamma and beta prime PDFs, this series is not new, as it was implicitly given in [Pastor, 2016]. We derived the series independently from that work.

Under certain assumptions, the terms of the series expansion of $\Lambda(x; \mu, \sigma)$ can be rearranged to give the MK Edgeworth series,

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n\left(\frac{\kappa_3(-Dx)^3}{6}, \dots, \frac{\kappa_{n+2}(-Dx)^{n+2}}{(n+1)(n+2)}\right) \right] \Lambda(x; \mu, \sigma), \quad (6)$$

¹This reference is translated to English in [Nicolas and Anfinson, 2012].

where κ_n is the n th log-cumulant of the PDF $f_X(x)$ we seek to approximate. This series was first derived in [Pastor et al., 2014] and we did *not* derive it independently of that work. However, this thesis contains a different approach, leading to important differences in both the derivation and the expression of the series.

We also applied the Bell polynomials to the classical series expansions of the standardized Gaussian PDF $\alpha(x)$. That is, we expressed the Gram-Charlier as

$$f_X(x) = \left[1 + \sum_{n=3}^{\infty} \frac{1}{n!} B_n(0, 0, c_{X,3}, \dots, c_{X,n}) H_n(x) \right] \alpha(x), \quad (7)$$

where $c_{X,n}$ is the n th order cumulant of X , and the Edgeworth series as

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n \left(\frac{c_{X,3}(-D)^3}{6}, \dots, \frac{c_{X,n+2}(-D)^{n+2}}{(n+1)(n+2)} \right) \right] \alpha(x), \quad (8)$$

where $(-D)^n \alpha(x) = H_n(x) \alpha(x)$ allows for a computable result. These are, to our knowledge, the first explicit expressions of the classical Gram-Charlier and Edgeworth series, which are more than a century old.

We conducted a broad numerical investigation as to the performance of the MK series expansions in approximating and estimating several target distributions, and fitting real-world data. We focused on distributions which are relevant in radar imagery, but their general nature allows us to draw conclusions which apply to all non-negative random phenomena.

Acknowledgements

I would like to thank my supervisor, Stian N. Anfinsen, for providing the subject for this thesis, his invaluable feedback, and his assistance throughout the writing process.

There are several others at UiT who deserve my gratitude, across multiple institutes at the Faculty of Science and Technology. For example, I have taken classes with only one or two other students, and I have been the only student to show up to three-hour lectures, but I have never experienced cancellations because of this. This dedication has been paralleled by the admirable helpfulness of the student advisors. As the nordic proverb goes: *none mentioned, none forgotten*.

Last, but never least, I would like to thank my family for their unceasing support of all my endeavors.

Mathematical Nomenclature

\equiv "Defined as", "by definition"

$f(x) \xleftrightarrow{\mathcal{F}} F_{\mathcal{F}}(\omega)$ $F_{\mathcal{F}}(\omega) = \mathcal{F}[f(x)](\omega)$. That is, $f(x)$ and $F_{\mathcal{F}}(\omega)$ is a Fourier transform pair.

$f(x) \xleftrightarrow{\mathcal{M}} F_{\mathcal{M}}(s)$ $F_{\mathcal{M}}(s) = \mathcal{M}[f(x)](s)$. That is, $f(x)$ and $F_{\mathcal{M}}(s)$ is a Mellin transform pair.

$\log x$ Natural logarithm of x (base e).

$\exp\{x\}$ The (natural) exponential function. e^x is also sometimes used.

$\langle \theta \rangle$ Empirical (or sample) value of the parameter θ . For example, if the RVs X_1, \dots, X_n all have mean m , then the sample x_1, \dots, x_n has empirical mean $\langle m \rangle \equiv \sum_{i=1}^n x_i/n$.

$\hat{\theta}$	Estimate of the parameter θ . For example, $\hat{m}_{\text{ML}} = \langle m \rangle$, which is to say that the maximum likelihood estimate of the mean is the empirical mean.
D_x, D	The derivative operator $D_x \equiv \frac{d}{dx}$. We use D (without subscript) when it is obvious which variable the expression is being differentiated with respect to.
$E\{X\}$	Expectation (expected value) of X . The concept extends to $E\{g(X)\}$. See e.g. [Kendall et al., 1994] for more details.
$\text{Re}(s)$	Real part of s .
Domain	The domain of a function $f(x)$ is the set of all values of the argument x for which the function f is defined.
Support	The support of a function $f(x)$ is the subset of the domain containing exactly the values of the argument x for which $f(x) \neq 0$.
$\mathbb{R}_{\geq 0}$	The set of all non-negative real numbers $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$. That is, $x \in [0, \infty)$.
$\mathbb{R}_{\neq 0}$	The set of all non-zero real numbers $\mathbb{R}_{\neq 0} = \{x \in \mathbb{R} \mid x \neq 0\}$. That is, $x \in (-\infty, 0) \cup (0, \infty)$.
$\mathbb{Z}_{\geq 0}$	The set of all non-negative integers, $\mathbb{Z}_{\geq 0} = \{x \in \mathbb{Z} \mid x \geq 0\}$. That is, $x \in \{0, 1, 2, 3, \dots\}$.

Abbreviations

CF	<i>Characteristic function</i>
CGF	<i>Cumulant generating function</i>
e.g.	<i>Exempli gratia (latin for "for example")</i>
eq.	<i>Equation</i>
FT	<i>Fourier transform</i>
GFD	<i>Generalized gamma distribution</i>
i.e.	<i>Id est (latin for "that is")</i>
IID	<i>Independently identically distributed</i>
MGF	<i>Moment generating function</i>
MK	<i>Mellin kind, e.g. MK series expansions, MK statistics, MK CF etc.</i>
MK BK series	<i>Mellin kind beta prime kernel series</i>
MKE series	<i>Mellin kind Edgeworth series</i>
MKGK series	<i>Mellin kind gamma kernel series</i>
MK LK series	<i>Mellin kind log-normal kernel series</i>
ML	<i>Maximum likelihood</i>
MoLC	<i>Method of log-cumulants</i>
MT	<i>Mellin transform</i>

PDF	<i>Probability density function</i>
RV	<i>Random variable</i>
SAR	<i>Synthetic aperture radar</i>
s.t.	<i>Such that</i>
w.r.t.	<i>With respect to</i>

Chapter 1

Introduction

1.1 Motivation

In the broadest sense, the purpose of this thesis is to advance the understanding of a modern approach within the realm of theoretical statistics, which is used to approximate certain probability densities. An immediate application is towards the analysis of satellite *synthetic aperture radar* (SAR) images. For instance, we may wish to make an inference about certain surface properties, like tracking sea ice or assessing if an agricultural area is at risk of experiencing drought. Another example is monitoring deforestation in tropical forest areas where the Norwegian government has paid for its preservation. These tasks can be accomplished by using techniques such as classification, see [Doulgeris et al., 2008]; clustering, see [Doulgeris et al., 2011]; and segmentation, see [Doulgeris, 2015]. All of these methods require a model for the data in the image. Due to the stochastic nature of the pixel measurements, the model will be a probability density. Thus, the modeling of probability densities is a central problem in these applications, and there are currently several models being used. Some are very simple and fast but lacking in accuracy, while others are complicated but more accurately describes the underlying natural phenomena.

The purpose of this thesis was to examine whether a new statistical framework gives rise to novel approximations to the complicated models, an endeavor in which we were successful. During the course of this work, we have also made significant contributions to the understanding of this framework, regarding both the theoretical foundations and the performance of the methods. We have conducted comprehensive testing which indicates what types of scenarios in which the methods are faster and/or more accurate than current methods. Our findings suggest that these methods can improve speed and accuracy in some real-world applications. This can potentially enable operational services to produce more accurate classification maps, by using the improved tools at their disposal.

These methods would naturally be applicable to other fields as well. Ultrasound and laser speckle imaging are other examples of coherent imaging techniques which experience the same interference phenomenon as SAR images, and therefore have the same statistical characteristics. In fact, the methods can potentially find use in entirely separate fields, as long as the data is non-negative. In [Pastor et al., 2014], the authors used one of the models discussed in this thesis to model wireless networks. Those measurements arise from a radically different process than in coherent imaging, but both are associated with heavy-tailed distributions.

Getting more technical, the classical Gram-Charlier and Edgeworth series have been used to model certain types of data for a long time, see e.g. [Blinnikov and Moessner, 1998]. Unrelated

to these series, *Mellin kind* (MK) statistics was introduced in [Nicolas, 2002]. The supervisor for this thesis, Stian N. Anfinsen, has applied MK statistics to *Earth Observation* (EO) problems, e.g. in [Anfinsen and Eltoft, 2011]. During his work in the EO field, he identified the possibility to improve on current methods, while also seeing a clear analogy between the Edgeworth series in classical statistics and properties in MK statistics. The intuition that there could and should exist such a series gave rise to this thesis.

1.2 Subjects and Goals

This thesis was preceded by a project paper, in which we aimed to derive a specific new series expansion of the gamma distribution, using MK statistics. We successfully derived that expansion, and during the course of that work we identified several related topics which warranted further research, thus justifying this master's thesis. The original goals for this thesis were:

- Expand and generalize the work from the project paper, in order to create a framework for MK series expansions.
- Investigate this framework and identify other MK series expansions.
- Thoroughly examine the performance of the discovered MK series expansions.

During the course of this work, we also made interesting discoveries which were unexpected in the sense that they were not covered by the original goals. This included a significant contribution to the classical Edgeworth series, and tools which can prove useful in other areas of MK statistics. These, along with the other contributions of this thesis, are listed in Section 1.5.

Based on this work, we are working on two scientific papers. They are currently draft versions, intended for imminent submission. One of them is appended to this thesis, as it includes experiments which supplement those we conduct here.

1.3 Method

The starting point for this thesis was the aforementioned project paper. This also dictated the working method, which in broad strokes were

1. Expand and refine the theoretical aspects of the MK series expansions framework.
2. Review the literature in search of new ideas.
3. Repeat steps 1 and 2 until the thesis presents a clear, concise, and comprehensive review of the MK series expansions framework.
4. Perform numerical experiments to document strengths and weaknesses of the MK series expansions.

Naturally, the work did not progress nearly as smoothly as the above list might suggest, with dead ends, surprises, frustration, and satisfaction, usually when least expected.

1.4 Organization

This thesis is organized as follows. In Chapter 2, we review the theoretical background needed for the rest of this thesis. In Chapter 3, we present the framework for the MK series expansions, including the series themselves. In Chapter 4, we take some of the new ideas which were used

in Chapter 3, and apply them to the classical series expansions. In Chapter 5, we perform numerical experiments on the MK series expansions from Chapter 3. We conclude in Chapter 6.

The decision was made to limit Chapter 2 to only a review of the literature.¹ That is, everything in that chapter has been derived and described previously. Most of the content in the other chapters is new material (introduced in this thesis), with all exceptions explicitly stated. As a consequence of this, we introduce the classical Gram-Charlier and Edgeworth series in Section 2.4, but revisit them in Chapter 4 to propose our changes and improvements there.

The most notable departure from the above guidelines, is the inclusion of SAR specific theory in Chapter 5. While the modeling of SAR data was a key motivation for this thesis, the derived series expansion methods can just as easily be applied to similar models from entirely different fields. Thus, SAR theory does not permeate this thesis, and is only mentioned when necessary, e.g. to motivate an experiment.

Footnotes are primarily used for supplementary information, i.e. information that is not strictly necessary to understand the material. In practice, this includes a lot of the formalism, which cannot and should not be omitted entirely, but is not of interest to all readers and could potentially be disruptive to the flow of the text. Also, a lot of the SAR specific information has been relegated to footnotes, in recognition of the fact that some readers may find it not interesting.

1.5 Contributions of This Thesis

To round off the introduction, we discuss and list the contributions of this thesis. This is for the benefit of readers which are familiar with the theoretical background and are only interested in the new findings we present.

The biggest contributions of this thesis are:²

- The derivation of the MK series expansion with arbitrary kernel.
- The demonstration of how the Mellin derivatives can be used in the context of MK statistics.³
- The recursive definition in eq. (3.16) and lemma 1, which concern the relation between a Mellin derivative and the arbitrary kernel.
- The MK gamma kernel series and MK beta prime kernel series, and subsequent results about the polynomials which arise during their derivation, including lemmas 1, 2, 4, and 5.
- The novel use of the Bell polynomials, which gives explicit expressions for the MK series expansions.
- The derivation of the classical Gram-Charlier series expansion with arbitrary kernel. This mirrors our derivation of the MK series expansion, and includes lemma 3.
- The ideas of a series expansion with arbitrary kernel was carried over to the classical Gram-Charlier series, including lemma 3.

¹There are two exceptions to this rule, i.e. Chapter 2 includes two minor results which were derived during the course of writing this thesis. They are listed at the end of Section 1.5.

²Several of these results are connected to the work in [Pastor et al., 2014], [Pastor et al., 2016], and [Pastor, 2016]. In Section 3.4.3 we discuss in detail how we derived the results to clarify what was done independently of Pastor et al.

³Most sources do not name the Mellin derivatives, i.e. they are just differential operators with no clear link to MK statistics. This can explain why they have not been used before in this context.

- The use of the Bell polynomials was also applied to the classical Gram-Charlier and Edgeworth series, providing explicit expressions for these.

In addition to this, a few minor results were derived during the course of this work, some of which can possibly be called small contributions. Examples of this are⁴

- A simpler proof of how the power function affects the log-cumulants, see Section 2.3.7.
- An alternate proof of how Faà di Bruno's formula and the Bell polynomials represent the relationship between (log-)moments and (log-)cumulants, see Section 2.5.3.
- A new (and much simpler) formula for the coefficients of the Gram-Charlier series around the gamma kernel, see Section 4.2.3.
- A proof of the scalability of the generalized Laguerre polynomials in terms of their Rodrigues formula and orthogonality property, see Appendix A.1.

⁴This list is not exhaustive, but it should be clear from the text which results were derived for this thesis, and which were found in the literature.

Chapter 2

Theory

2.1 Terminology

The key foundation of this master's thesis is the idea of Jean-Marie Nicolas to replace the Fourier transform with the Mellin transform in the definition of the characteristic function. It was originally presented in French in [Nicolas, 2002] and later translated to English in [Nicolas and Anfinzen, 2012]. There, the terms *first-kind statistics* and *second-kind statistics* were assigned to the classical (Fourier kind) and MK framework, respectively. However, this thesis will instead adhere to the naming convention coined in [Anfinzen and Eltoft, 2011]. That is, *classical statistics* refers to the framework where the characteristic function is the Fourier transform of the probability density function, while MK statistics refers to the framework where the Mellin transform is used instead. Also, the terms *characteristic function* and *cumulant generating function* are used instead of *first/second characteristic function*. This choice of terminology is deliberate and preempts any possible confusion about e.g. the *second characteristic function of the second kind*, which will here be referred to as the *Mellin kind (log-)cumulant generating function*, abbreviated to MK CGF. This more precisely conveys the nature of the functions, instead of just numerically labeling them.

When referring to functions in classical statistics, the symbols used for the characteristic and cumulant generating functions are upper case (Φ and Ψ respectively). The lower case symbols ϕ and φ are used when referring to MK statistics. This is also in accordance with [Anfinzen and Eltoft, 2011]. The moment and log-moment generating functions are not assigned their own greek letters because they are seldom used in this thesis.

The classical linear moments and cumulants themselves are assigned the roman letters m and c , while the greek letters μ and κ denote the logarithmic moments (log-moments) and logarithmic cumulants (log-cumulants) respectively. Again, this mirrors [Anfinzen and Eltoft, 2011]. Historically, μ has often been assigned to the classical (linear) moments, but this deviation from the norm can be justified by the limited presence of the classical descriptors in this thesis and the benefit of clean and simple symbols for the MK descriptors, which are absolutely central to this work. We let ς denote the classical variance, and σ denote its logarithmic counterpart.

The well-known probability density functions that are central to the thesis are assigned their own greek letters, e.g. the Gaussian $\alpha(\cdot)$. By convention $\alpha(x; m, \varsigma)$ denotes the Gaussian distribution with parameters m, ς , while $\alpha(x)$ denotes the standardized version, that is $m = 0, \varsigma = 1$. The same goes for the other distributions and will be specified as they are formally introduced later in this chapter.

2.2 Classical Statistics

2.2.1 The Characteristic Function

Let X be a *random variable* (RV) with *probability density function* (PDF) $f_X(x)$.¹ The *characteristic function* (CF) $\Phi_X(\omega)$ of X is defined as the *Fourier transform* (FT)² of its PDF:

$$\Phi_X(\omega) = \mathcal{F}[f_X(x)](\omega) = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx = \mathbb{E}\{e^{j\omega X}\}, \quad (2.1)$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator w.r.t. X , $j \equiv \sqrt{-1}$ is the imaginary unit, and ω is a real-valued transform variable, often interpreted as a frequency.

2.2.2 Moments

The *moment generating function* (MGF) is defined as

$$\text{MGF}_X(\omega) \equiv \mathbb{E}\{e^{\omega X}\}. \quad (2.2)$$

I.e., the MGF and the CF are related by $\text{MGF}_X(\omega) = \Phi_X(-j\omega)$ and this close relationship means that for the purpose of this thesis, only the CF is needed to develop the theory and avoiding use of the MGF will reduce confusion.³ The moments $m_n\{X\}$ can be defined in two ways, in terms of the expectation of a power of X , and in terms of the derivative of the CF;

$$m_n\{X\} = \mathbb{E}\{X^n\} = \int_{-\infty}^{\infty} x^n f_X(x) dx, \quad (2.3)$$

$$m_n\{X\} = (-j)^n \frac{d^n}{d\omega^n} \Phi_X(\omega)|_{\omega=0}. \quad (2.4)$$

If all of the moments exist⁴, then we can define them as derivatives of the CF. This is based on the Taylor series expansion of the exponential function $e^{j\omega x}$ around $\omega = 0$, which can be used to rewrite the integral in eq. (2.1) as

$$\Phi_X(\omega) = \sum_{n=0}^{\infty} m_n \frac{(j\omega)^n}{n!}. \quad (2.5)$$

Note that m_0 is included in this sum, and can be viewed either as the integral of the PDF $\mathbb{E}\{X^0\} = \int_{-\infty}^{\infty} f_X(x) dx = 1$ or $\Phi_X(0)$, hence $m_0 = 1$ trivially for any PDF. One use of this "zereth" moment is that any non-negative function $f(x)$ divided by its integral (m_0) can by

¹A detailed review of basic concepts like the conditions for the existence of the PDF will not be included here. The vast majority of books on the fundamentals of statistics and probability cover this in detail, e.g. [Kendall et al., 1994].

²There are several ways to define the FT with regards to the sign in the exponential, the scaling and the use of ordinary or angular frequency. The choices in the transform also defines the inverse, s.t. the inverse $\mathcal{F}^{-1}[\mathcal{F}[f(x)]] = f(x)$. The definition used here is common for this purpose (see [Kendall et al., 1994]), as the CF is exactly the FT of the PDF with these choices.

³The MGF is designed to have the convenient property that $m_n = \text{MGF}^{(n)}(0)$, but as shown in this section the CF has the same property with only an extra factor $(-j)^n$.

⁴The existence of m_n requires that the integral in eq. (2.3) converges for the corresponding x^n [Kendall et al., 1994]. There are several well-known distributions where some moments do not exist or are ∞ , but the distributions encountered in this thesis are mostly well behaved. When referring to m_n in the following, it is implied that it exists for that particular distribution unless otherwise stated.

definition be interpreted as a PDF. The moments m_n are often called the *raw moments* to separate them from the *centralized moments* (also known as *moments about the mean*), defined as

$$\tilde{m}_n \equiv E\{(X - m_1)^n\}. \quad (2.6)$$

That is, the centralized moments compensate for the mean of the RV, giving $\tilde{m}_1 = 0$. With respect to naming, the first raw moment m_1 is the mean, while the second centralized moment $\tilde{m}_2 = \zeta^2$ is the variance. Additionally, the *standardized moments* are the dimensionless quantities defined as \tilde{m}_n/ζ^n , where \tilde{m}_3/ζ^3 is the skewness and \tilde{m}_4/ζ^4 is the kurtosis. To conclude the discussion on moments, note the property that the set of all moments of an RV, if they exist, almost always uniquely describes the PDF of the RV.⁵

2.2.3 Cumulants

In the same way as the moments can be found from the CF, the cumulants⁶ $c_n\{X\}$ are found via the *cumulant generating function* (CGF)

$$\Psi_X(\omega) \equiv \log \Phi_X(\omega). \quad (2.7)$$

If the moments m_n exist, then the cumulants also exist [Sundt et al., 1998] and the CGF can be written as⁷

$$\Psi_X(\omega) = \sum_{n=1}^{\infty} c_n \frac{(j\omega)^n}{n!}, \quad (2.8)$$

which also gives a method for finding the cumulants when the CGF is known:

$$c_n = (-j)^n \frac{d^n}{d\omega^n} \Psi_X(\omega)|_{\omega=0}. \quad (2.9)$$

The analogy to calculating moments from the CF is immediately clear. Another analogy with the moments is that the set of all cumulants also describes a distribution uniquely. The relationship between moments and cumulants is given in e.g. [Pitman, 2002] and [Rota and Shen, 2000] as

$$c_n = \sum_{i=1}^n (-1)^{i-1} (i-1)! B_{n,i}(m_1, \dots, m_{n-i+1}), \quad (2.10)$$

$$m_n = B_n(c_1, \dots, c_n). \quad (2.11)$$

Here, $B_{n,i}(m_1, \dots, m_{n-i+1})$ is the partial Bell polynomial and $B_n(\cdot)$ is the n th complete Bell polynomial, which we will define in Section 2.5. Eq. (2.10) is a special case of Faà di Bruno's formula, which is described in [Johnson, 2002]⁸ and will be examined in Section 2.5.3. Note that in order to compute c_n , only m_1, \dots, m_n up to order n is required and, similarly, to find m_n one only needs to know c_1, \dots, c_n .

⁵[Heyde, 1963] provides a rare counterexample where the set of all moments can belong to distinct distributions, but the statement is still useful in order to provide insight to what the descriptive constants (e.g. the moments) actually are.

⁶Both moments and cumulants are functions of the RV X through its PDF, as indicated by the notation $m_n\{X\}$ and $c_n\{X\}$. However, they are usually referred to just as m_n and c_n .

⁷Unlike the CF, the power series of the CGF does not include the term $n = 0$. Usually, the zeroth cumulant is not defined even though the zeroth moment is. If c_0 were to be defined, then $c_0 = 0$ since $\log \Phi_X(0) = 1$ for all true PDFs, and the term $n = 0$ in the sum definition of the CGF would be zero anyway.

⁸Using the same notation as Johnson, $f(t)$ is the CGF and $g(t) = e^t$, s.t. all derivatives of $g(t)$ are also e^t (greatly simplifying the expression) and $g(f(t))$ is the CF.

Shift Invariance The cumulants of order $n \geq 2$ are *shift-invariant*. That is to say that if the RV X has cumulants $c_{X,n}$ and $Y = a + X$ where a is some constant, then Y has cumulants $c_{Y,1} = c_{X,1} + a$ and $c_{Y,n} = c_{X,n}$ for $n \geq 2$. To see this, state the CF of Y in terms of the CF of X

$$\Phi_Y(\omega) = \mathbb{E}\{e^{j\omega Y}\} = \mathbb{E}\{e^{j\omega(a+X)}\} = e^{j\omega a}\Phi_X(\omega), \quad (2.12)$$

and the cumulants are found via the CGF $\Psi_Y(\omega) = \log \Phi_Y(\omega) = j\omega a + \Psi_X(\omega)$ using eq. (2.9),

$$c_{Y,n} = (-j)^n \frac{d^n}{d\omega^n} [j\omega a + \Psi_X(\omega)]|_{\omega=0} = \begin{cases} a + c_{X,1} & , \quad n = 1, \\ c_{X,n} & , \quad n \geq 2. \end{cases} \quad (2.13)$$

Scaling Another property of the cumulants is that for constant $a \in \mathbb{R}_{\neq 0}$ the RV aX has cumulants $a^n c_{X,n}$, i.e. scaling X with a results in the n th-order cumulant being scaled with a^n . To see this, start with

$$\Phi_{aX}(\omega) = \mathbb{E}\{e^{j\omega aX}\} = \mathbb{E}\{e^{j(a\omega)X}\} = \Phi_X(a\omega), \quad (2.14)$$

and use the power series definition of the cumulants in eq. (2.8) to see that

$$\Psi_{aX}(\omega) = \Psi_X(a\omega) = \sum_{n=1}^{\infty} \frac{(ja\omega)^n}{n!} c_{X,n} = \sum_{n=1}^{\infty} \frac{(j\omega)^n}{n!} (a^n c_{X,n}), \quad (2.15)$$

i.e. the cumulants of aX are $a^n c_{X,n}$.

Upon encountering the cumulants for the first time, it is often questioned what they *are*, and what their purpose is. The moments have a mathematical definition and physical and statistical interpretations which are simple and well-established. The cumulants on the other hand, are defined through the CGF (via the CF or the MGF) and in practice they are usually computed as combinations of the empirical moments. However, both the moments and the cumulants are sets of descriptive constants of a distribution which can be useful in describing or specifying it. From this point of view, their value must be based on their usefulness, and as [Kendall et al., 1994] notes, the cumulants are more useful than the moments. This will become apparent in the following sections. The cumulants were first defined explicitly in [Thiele, 1889], and [Hald, 2000] recounts their early history.

2.2.4 The Sum of Independent RVs

Let X and Y be two RVs with PDFs $f_X(x)$ and $f_Y(y)$ and *joint* PDF $f_{X,Y}(x,y)$. X and Y are said to be *independent* if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \forall x,y. \quad (2.16)$$

From e.g. [Stark and Woods, 2012] we know that their sum $Z = X + Y$ is then also an RV and has PDF

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx. \quad (2.17)$$

This is the *additive convolution*⁹ of $f_X(x)$ with $f_Y(y)$, usually written

$$f_Z(z) = f_X(x) * f_Y(y). \quad (2.18)$$

⁹Usually this is referred to as just *convolution*. In this thesis however, we encounter another type of convolution, so it is necessary to use the prefix here.

Let X model the signal and Y model the noise. If the noise is additive and independent of the signal, then $Z = X + Y$ is the measurable system output. This *additive noise* model is very much used, and Z is commonly analyzed via the FT. The convolution property of the FT, which can be found in e.g. [McClellan et al., 2003], says that

$$\mathcal{F}[f(x) * g(x)](\omega) = \mathcal{F}[f(x)](\omega)\mathcal{F}[g(x)](\omega), \quad (2.19)$$

i.e. the FT of a convolution of two functions is equal to the product of the FTs of each function. Inserting the PDFs $f_X(x)$ and $f_Y(y)$ for $f(x)$ and $g(x)$, we see that the CF of the sum of independent RVs is the product of the CF of each of these RVs. That is, if X and Y have CFs $\Phi_X(\omega)$ and $\Phi_Y(\omega)$, then

$$\Phi_Z(\omega) = \Phi_X(\omega)\Phi_Y(\omega). \quad (2.20)$$

Taking the logarithm on both sides, we get the CGF of Z :

$$\Psi_Z(\omega) = \Psi_X(\omega) + \Psi_Y(\omega), \quad (2.21)$$

from which we can see that the cumulants of a sum $Z = X + Y$ are the sums of the corresponding cumulants of X and Y , i.e. $c_{Z,n} = c_{X,n} + c_{Y,n}$.

2.2.5 The Gaussian Distribution

Perhaps the most ubiquitous distribution is the Gaussian (normal), which has PDF

$$\alpha(x; m, \varsigma) = \frac{1}{\sqrt{2\pi\varsigma}} \exp\left\{-\frac{(x-m)^2}{2\varsigma^2}\right\}. \quad (2.22)$$

At this point, we recall that omitting the parameters from $\alpha(\cdot)$ denotes the standardized version of the PDF, which for the Gaussian is $m = 0, \varsigma = 1$, i.e.

$$\alpha(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \quad (2.23)$$

To convert between the two, see that

$$\alpha(x; m, \varsigma) = \frac{1}{\varsigma} \alpha\left(\frac{x-m}{\varsigma}\right). \quad (2.24)$$

[Bryc, 2012] gives the CF and CGF of the Gaussian distribution as

$$\Phi_X(\omega) = \exp\left\{j\omega m - \frac{(\varsigma\omega)^2}{2}\right\}, \quad (2.25)$$

$$\Psi_X(\omega) = j\omega m - \frac{(\varsigma\omega)^2}{2}. \quad (2.26)$$

The cumulants are found from this using eq. (2.9): $c_1 = m, c_2 = \varsigma^2$ and $c_n = 0 \forall n > 2$. This is a unique property of the Gaussian distribution – there are no distributions where $c_1, c_2, \dots, c_n \neq 0$ and $c_k = 0 \forall k > n$, for any $n > 2$. In other words, the CGF cannot be a polynomial of finite degree $n > 2$, see [Lukacs, 1970]. Conversely, the even moments of the Gaussian distribution are all non-zero (functions of ς), i.e. the cumulants describe this distribution in the simplest and most elegant way possible.

2.3 Mellin Kind Statistics

MK statistics was introduced in [Nicolas, 2002]. It is the foundation for this thesis, along with the work of others like [Anfinson and Eltoft, 2011] and [Pastor et al., 2014], who have expanded and applied Nicolas' findings. The elements of MK statistics which are key to this thesis will now be presented. For a recent and very comprehensive guide to all aspects of MK statistics, see [Nicolas, 2016].

2.3.1 The Mellin Transform

The *Mellin transform* (MT) is named after Finnish mathematician Hjalmar Mellin. The MT of a function $f(x)$ is

$$\mathcal{M}[f(x)](s) \equiv \int_0^{\infty} x^{s-1} f(x) dx = F_{\mathcal{M}}(s) \Leftrightarrow f(x) \xrightarrow{\mathcal{M}} F_{\mathcal{M}}(s), \quad (2.27)$$

where $s \in \mathbb{C}$ is a complex transform variable.

2.3.2 The Inverse Mellin Transform

As explained in [Flajolet et al., 1995], the *fundamental strip* is the largest open strip $\langle a, b \rangle$ for $\text{Re}(s)$ in which the integral in (2.27) converges. It is a vertical strip in the complex plane defined by its boundaries on the real line, hence the name. The inverse MT is defined in the same reference as

$$f(x) = \mathcal{M}^{-1}[F_{\mathcal{M}}(s)](x) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} x^{-s} F_{\mathcal{M}}(s) ds. \quad (2.28)$$

This integral is along a vertical line in the complex plane, given by a constant real value $c \in \mathbb{R}$. The inverse MT exists when $\{c \in \mathbb{R} : a < c < b\}$, i.e. when $s \in S_f$, where S_f is the fundamental strip.

2.3.3 A MT Example: Euler's Gamma Function

When examining the MT, a simple first example is Euler's gamma function $\Gamma(s)$, usually referred to just as the gamma function. It is defined as

$$\Gamma(s) \equiv \int_0^{\infty} x^{s-1} e^{-x} dx, \quad (2.29)$$

for s complex with positive real part, see [Davis, 1959] for more information.¹⁰ Combining eqs. (2.27) and (2.29), we see that the gamma function is simply the MT of e^{-x} , i.e.

$$e^{-x} \xleftrightarrow{\mathcal{M}} \Gamma(s). \quad (2.30)$$

2.3.4 Some General Properties of the Mellin Transform

There are several good publications concerning the MT. [Bertrand et al., 2000] include some derivations of the general properties of the MT. From its definition, the MT is trivially linear,

Table 2.1: Some basic Mellin transform properties.

	$f(x)$	$F_{\mathcal{M}}(s) = \mathcal{M}[f(x)](s)$	Constraint
Scaling	$f(bx)$	$b^{-s}F_{\mathcal{M}}(s)$	$b > 0, s \in S_f$
Multiplication	$x^n f(x)$	$F_{\mathcal{M}}(s+n)$	$s+n \in S_f$
Differentiation	$D^n f(x)$	$(-1)^n \prod_{k=1}^n (s-k)F_{\mathcal{M}}(s-n)$	$n \in \mathbb{Z}_{\geq 0}, s-n \in S_f$

i.e. if $f(x) \xleftrightarrow{\mathcal{M}} F_{\mathcal{M}}(s)$ and $g(x) \xleftrightarrow{\mathcal{M}} G_{\mathcal{M}}(s)$, then $af(x) + bg(x) \xleftrightarrow{\mathcal{M}} aF_{\mathcal{M}}(s) + bG_{\mathcal{M}}(s)$ for constant a, b . Other properties central to this thesis are listed in Table 2.1.

These properties can be combined to give

$$D^n x^n f(x) \xleftrightarrow{\mathcal{M}} (-1)^n \prod_{k=1}^n (s-k)F_{\mathcal{M}}(s), \quad n \in \mathbb{Z}_{\geq 0}, \quad (2.31)$$

under the constraint that $s \in S_f$, which is just the original constraint for $f(x) \xleftrightarrow{\mathcal{M}} F_{\mathcal{M}}(s)$.

2.3.5 The Mellin Kind Characteristic Function

The *Mellin kind characteristic function* (MK CF) $\phi_X(s)$ of a non-negative RV¹¹ X was defined in [Nicolas, 2002] as the MT of its PDF $f_X(x)$:

$$\phi_X(s) = \mathcal{M}[f_X(x)](s) = \int_0^{\infty} x^{s-1} f_X(x) dx = E\{X^{s-1}\}, \quad (2.32)$$

where the expectation is again w.r.t. X . Also, the PDF $f_X(x)$ is the inverse MT of the MK CF,

$$f_X(x) = \mathcal{M}^{-1}[\phi(s)](x). \quad (2.33)$$

From its definition in eq. (2.27) it is clear that the MT only makes sense for non-negative RVs. There are several important PDFs designed for such RVs, including the log-normal and gamma distributions, both of which will be discussed later in this thesis. Indeed, as demonstrated in [Nicolas, 2002], the MT seems to be fundamentally better suited to the distributions of non-negative RVs, compared to the FT. The intuition behind this is that since the MT works on exactly the values of x where $f_X(x)$ is non-zero, it is tailored to these distributions in a sense. On the other hand, we see from the FT definition in eq. (2.1), that it is better suited to PDFs supported on the whole real line.

2.3.6 Log-Moments

The *log-moments* μ_n are defined in [Nicolas, 2002] as

$$\mu_n\{X\} = E\{(\log X)^n\} = \int_0^{\infty} (\log x)^n f_X(x) dx. \quad (2.34)$$

¹⁰It is named after the famous mathematician Leonhard Euler who presented the integral form in [Euler, 1738].

¹¹Formally, X is said to be non-negative if $f_X(x) = 0 \forall x < 0$, i.e. if its PDF is supported on (a subset of) $\mathbb{R}_{\geq 0}$.

They can be retrieved from the MK CF by rewriting the transform kernel $x^{s-1} = e^{(\log x)(s-1)}$ in eq. (2.32) to get

$$\phi_X(s) = \int_0^{\infty} e^{(\log x)(s-1)} f_X(x) dx, \quad (2.35)$$

then inserting the power series expansion for the exponential function as a function of s at $s = 1$, which gives

$$\phi_X(s) = \int_0^{\infty} \sum_{n=0}^{\infty} \frac{[(\log x)(s-1)]^n}{n!} f_X(x) dx. \quad (2.36)$$

Finally, we reverse the order of integration and summation¹² to recognize the integral definition of the log-moments from eq. (2.34), and see that

$$\phi_X(s) = \sum_{n=0}^{\infty} \frac{(s-1)^n}{n!} \int_0^{\infty} (\log x)^n f_X(x) dx = \sum_{n=0}^{\infty} \mu_n \frac{(s-1)^n}{n!}. \quad (2.37)$$

As in the classical case, this requires of the existence of all μ_n , and implies that

$$\mu_n = \left. \frac{d^n}{ds^n} \phi_X(s) \right|_{s=1}. \quad (2.38)$$

This illustrates that the log-moments are the MK statistics equivalent of the (classical) moments m_n .

The log-moments were used before the introduction of the MK CF. For example, it is well known that for data samples $\{x_1, \dots, x_n\}$ from a gamma distribution with known scale but unknown shape, the sample *log-mean*

$$\langle \mu \rangle \equiv \frac{1}{n} \sum_{i=1}^n \log x_i \quad (2.39)$$

is a sufficient statistic for the shape parameter, see [Pitman, 1937]. Note that by definition, the log-mean¹³ is the first-order log-moment, i.e. $\mu = \mu_1 = E\{\log X\}$.

2.3.7 Log-Cumulants

The *log-cumulant generating function* (or MK CGF) is defined as $\varphi_X(s) = \log \phi_X(s)$. If all log-cumulants κ_n exist, we have

$$\varphi_X(s) = \sum_{n=1}^{\infty} \kappa_n \frac{(s-1)^n}{n!}, \quad (2.40)$$

$$\kappa_n = \left. \frac{d^n}{ds^n} \varphi_X(s) \right|_{s=1}. \quad (2.41)$$

For a detailed description of the conditions for the existence of log-moments and log-cumulants, it is again referred to [Nicolas, 2002]. In the same article, Nicolas argues that since the log-cumulants are constructed the same way as the classical cumulants, the relationships between

¹²This is required in the classical case as well, and [Kendall et al., 1994] discuss when this reversal is mathematically valid. It is beyond the scope of this thesis to discuss this at length; the representations used are valid unless otherwise stated.

¹³This deviates from [Nicolas, 2002], where the term log-mean referred to a different entity. This naming convention was used (seemingly for the first time) in [Pastor et al., 2016].

μ_1, \dots, μ_n and $\kappa_1, \dots, \kappa_n$ must be the same as in the classical case, i.e.

$$\kappa_n = \sum_{i=1}^n (-1)^{i-1} (i-1)! B_{n,i}(\mu_1, \dots, \mu_{n-i+1}), \quad (2.42)$$

$$\mu_n = B_n(\kappa_1, \dots, \kappa_n). \quad (2.43)$$

Scale Invariance The log-cumulants of order $n \geq 2$ are *scale invariant*. That is, if X has log-cumulants κ_n , then for some positive¹⁴ constant a , the log-cumulants of aX are $\kappa_{aX,1} = \kappa_{X,1} + \log a$ and $\kappa_{aX,n} = \kappa_{X,n}$ for $n \geq 2$. This is analogous to the shift invariance of the classical cumulants. The straightforward proof was provided in [Pastor et al., 2014], and starts with the MK CF definition from eq. (2.32) for aX ,

$$\phi_{aX}(s) = \mathbb{E}\{(aX)^{s-1}\} = a^{s-1} \mathbb{E}\{X^{s-1}\} = a^{s-1} \phi_X(s), \quad (2.44)$$

which gives the MK CGF

$$\varphi_{aX}(s) = \log(a^{s-1}) + \log \phi_X(s) = (s-1) \log a + \varphi_X(s). \quad (2.45)$$

Using the definition in eq. (2.41), the log-cumulants of aX are

$$\kappa_{aX,n} = \begin{cases} \log a + \kappa_{X,1} & , \quad n = 1, \\ \kappa_{X,n} & , \quad n \geq 2, \end{cases} \quad (2.46)$$

i.e. the log-cumulants of order $n \geq 2$ are scale-invariant, while the first order log-cumulant is shifted by $\log a$.

Power Transformation For the power transformation RV $Y = X^a$, where $a \in \mathbb{R}_{\neq 0}$ is a constant, the log-cumulants of order $n \geq 2$ are

$$\kappa_{Y,n} = a^n \kappa_{X,n}. \quad (2.47)$$

A rigorous proof can be found in [Nicolas, 2016], but mimicking the approach of the scaling property of the classical cumulants in Section 2.2.3 is also possible. We see that $\phi_Y(s) = \mathbb{E}\{X^{a(s-1)}\}$ implies that $(s-1)$ is scaled by a in the MK CGF and using the power series definition from eq. (2.40) gives

$$\varphi_Y(s) = \sum_{n=1}^{\infty} \kappa_n \frac{[a(s-1)]^n}{n!} = \sum_{n=1}^{\infty} (a^n \kappa_n) \frac{(s-1)^n}{n!}, \quad (2.48)$$

where the log-cumulants of Y are recognized as $a^n \kappa_{X,n}$.

To summarize, shifting the RV X affects the cumulants like scaling X affects the log-cumulants, and scaling X affects the cumulants like applying the power transformation to X affects the log-cumulants. Even the proofs are similar – this demonstrates the logarithmic nature of MK statistics.

After introducing the cumulants in Section 2.2.3, it was argued that their merit should be based on their usefulness as sets of descriptive constants. This also applies to the log-moments and log-cumulants, which are even further removed from a simple and intuitive interpretation of what they *are*. When introducing them, [Nicolas, 2002] demonstrated their usefulness in some cases, while others like [Anfinsen and Eltoft, 2011], [Krylov et al., 2013] and [Pastor et al., 2016] have recently extended their use to a wider range of applications. This thesis seeks to add to this growing body of research.

¹⁴For $a = 0$ the RV aX is fixed to 0, and for $a < 0$ it is a negative RV and hence unsuited to MK statistics. Thus, the requirement $a > 0$ does not reduce the generality of the result.

2.3.8 The Product of Independent RVs

Revisiting the model with independent RVs X, Y , what if the noise is *multiplicative*? That is, what if the measurable output is the RV $Z = X \cdot Y$? The usual approach has been to work with $\log Z = \log X + \log Y$, which reduces the problem to the well-known additive case. Using MK statistics, a new approach was presented in [Nicolas, 2002]. The PDF of Z is stated in [Nicolas, 2002] as

$$f_Z(z) = \int_0^{\infty} f_X(x) f_Y\left(\frac{z}{x}\right) \frac{dx}{x}, \quad (2.49)$$

which is the *multiplicative convolution*, also known as the *Mellin convolution*, denoted

$$f_Z(z) = f_X(x) \hat{*} f_Y(y). \quad (2.50)$$

Again referring to [Nicolas, 2002], a fundamental property of the MT is that

$$\mathcal{M}[f_X(x) \hat{*} f_Y(y)](s) = \mathcal{M}[f_X(x)](s) \mathcal{M}[f_Y(y)](s), \quad (2.51)$$

which is very similar to the FT additive convolution property in eq. (2.19). The implication of this is that the standard transform domain analysis of the additive model is possible for the multiplicative model, if the MT is used instead of the FT. In particular, if the MK CFs of X, Y are $\phi_X(s), \phi_Y(s)$ then the MK CF and CGF of Z are

$$\phi_Z(s) = \phi_X(s) \phi_Y(s), \quad (2.52)$$

$$\varphi_Z(s) = \varphi_X(s) + \varphi_Y(s). \quad (2.53)$$

This implies that in the case of multiplicative noise, the log-cumulants are additive, i.e.

$$\kappa_{Z,n} = \kappa_{X,n} + \kappa_{Y,n}. \quad (2.54)$$

The quotient of $Z = X/Y$ of two independent RVs was also discussed in [Nicolas, 2002]. We will not recite the entire review of that case, but one property is of interest later in this thesis: The log-cumulants of the quotient are given by

$$\kappa_{Z,n} = \kappa_{X,n} + (-1)^n \kappa_{Y,n}, \quad (2.55)$$

that is, the odd log-cumulants are the differences and the even log-cumulants the sums of the corresponding log-cumulants of the constituent RVs.

2.3.9 The Gamma Distribution

The gamma distribution has several parametrizations, and this thesis uses two of the most common ones. Let X be a gamma distributed RV. Its PDF with *shape* parameter $L > 0$ and *location* parameter $m = E\{X\}$ is then

$$\gamma(x; L, m) = \begin{cases} \left(\frac{L}{m}\right)^L \frac{x^{L-1}}{\Gamma(L)} \exp\left\{-\frac{Lx}{m}\right\} & x \geq 0, \\ 0 & x < 0, \end{cases} \quad (2.56)$$

where $\Gamma(\cdot)$ is the gamma function from Section 2.3.3. The alternative parametrization uses *shape* $a > 0$ and *scale* $b > 0$, giving the PDF

$$\gamma(x; a, b) \equiv \frac{b^a x^{a-1}}{\Gamma(a)} e^{-bx}, \quad x \geq 0, \quad (2.57)$$

where $\gamma(x; a, b) = 0 \forall x < 0$ is again the case. To convert between the two, see that $a \leftrightarrow L$, $b \leftrightarrow L/m$. The shape/location parametrization in $\gamma(x; L, m)$ is common for modeling SAR data, where m is the mean radar intensity and L has a physical interpretation as "(equivalent) number of looks", see [Anfinen et al., 2009] for more information. It represents the averaging done in the image formation, thus it is a *global* parameter in the sense that it is constant for the whole image and independent of the surface properties. This is in contrast to the *local* location parameter, which is assumed uniform only for segments of the image corresponding to exactly one class in the scene (forest, farmland etc.). The practical implication is that the number of looks can be estimated based on the entire image, and for modern sensors and methods this allows us to treat L as a known parameter.

The gamma distribution has classically been fitted to data using the *maximum likelihood* (ML) estimates as discussed in [Choi and Wette, 1969]. In general, this thesis adheres to a convention where $f_X(x)$ is a PDF model based on assumptions and observations of a physical process. Conversely, the distributions which are assigned greek letters, e.g. $\gamma(x; a, b)$, $\alpha(x; m, \varsigma)$ etc. are *kernels*, i.e. they are tools which can be tailored arbitrarily for specific purposes, in this case to facilitate the series expansions. As with the Gaussian kernel, omitting the scale by writing $\gamma(x; a)$ refers to the standardized case $b = 1$, which we often use in this thesis for the sake of brevity. The PDF is then simplified to

$$\gamma(x; a) = \frac{x^{a-1}}{\Gamma(a)} e^{-x}, \quad x \geq 0, \quad (2.58)$$

and since the scaling is simply a multiplication bx of the argument, subsequent generalizations to arbitrary scale are usually trivial mathematical exercises. In this thesis we do not explicitly use versions with standardized shape, but it can be noted that $a = 1$ reduces $\gamma(\cdot)$ to the exponential PDF, which is simply be^{-bx} .

The gamma distribution is an example of a PDF eligible for the MT. An analysis like the one conducted in [Nicolas, 2002] reveals several nice and interesting properties of the gamma distribution that can be derived via the MT. First, the classical and the MK CFs are

$$\Phi_X(\omega) = \left(\frac{L}{m}\right)^L \frac{\exp\left\{jL \arctan\left(\frac{m\omega}{L}\right)\right\}}{\left(\omega^2 + \left(\frac{L}{m}\right)^2\right)^{\frac{L}{2}}}, \quad (2.59)$$

$$\phi_X(s) = \left(\frac{m}{L}\right)^{s-1} \frac{\Gamma(s+L-1)}{\Gamma(L)}. \quad (2.60)$$

Nicolas even used the MK CF to derive the *classical* moments in a much simpler way than using the classical CF, namely

$$m_n = \phi_X(n+1) = \left(\frac{m}{L}\right)^n \frac{\Gamma(L+n)}{\Gamma(L)}. \quad (2.61)$$

The log-cumulants are of particular interest in this thesis, and are given by

$$\kappa_{\gamma, n} = \begin{cases} \psi^{(0)}(L) + \log\left(\frac{m}{L}\right) & n = 1, \\ \psi^{(n-1)}(L) & n \geq 2, \end{cases} \quad (2.62)$$

where $\psi^{(n)}(L)$ is the polygamma function, defined in e.g. [Abramowitz and Stegun, 1964] as

$$\psi^{(n)}(x) = D_x^{n+1} \log \Gamma(x). \quad (2.63)$$

In some applications, e.g. SAR, multiplicative models¹⁵ including gamma distributions are used, often resulting in distributions which are very complicated to evaluate and analyze. For instance,

¹⁵For a full overview of the log-cumulants in the multiplicative models, see [Bombrun et al., 2011] and [Anfinen and Eltoft, 2011]. [Deng et al., 2016] provides a physical interpretation of the models.

the product of two independent gamma distributed RVs are by definition K distributed, see [Nicolas, 2002]. The K PDF includes a modified Bessel function of the second kind, which can be both analytically and numerically prohibiting. By the properties of MK statistics, the log-cumulants associated with the K distribution are the sums of log-cumulants of its constituent RVs. Due to the simplicity of the expression for the log-cumulants of a gamma distributed RV, the expression for the log-cumulants of a K distributed RV is also simple.

Finding the MT of the gamma distribution is straightforward when combining eq. (2.29) with the scaling and multiplication properties from Table 2.1:

$$\phi_\gamma(s) = \mathcal{M}[\gamma(x; a, b)](s) = \frac{b^a}{\Gamma(a)} \int_0^\infty x^{s-1} x^{a-1} e^{-bx} dx = b^{a-s} \frac{\Gamma(s+a-1)}{\Gamma(a)}. \quad (2.64)$$

That is, the constant multiplier is preserved and the gamma function is shifted due to x^{a-1} , while the scaling b results in a factor b^{-s} .

2.3.10 The Beta Prime Distribution

If X_1 and X_2 are independent with PDFs $\gamma(x; a_1, b_1)$ and $\gamma(x; a_2, b_2)$, their ratio X_2/X_1 is beta prime¹⁶ distributed with PDF

$$f_{X_2/X_1}(x) = \beta' \left(x; a_1, a_2, \frac{b_2}{b_1} \right) \equiv \frac{b_2/b_1}{B(a_1, a_2)} \frac{\left(\frac{b_2}{b_1} x \right)^{a_1-1}}{\left(1 + \frac{b_2}{b_1} x \right)^{a_1+a_2}}, \quad x, a_1, a_2, b_1, b_2 \geq 0. \quad (2.65)$$

where $\beta'(x; a_1, a_2, b) = 0 \forall x < 0$ and $B(\cdot)$ is the beta function

$$B(a_1, a_2) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1 + a_2)}. \quad (2.66)$$

We define $b \equiv b_2/b_1$ for brevity, i.e.

$$\beta'(x; a_1, a_2, b) \equiv \frac{b}{B(a_1, a_2)} \frac{(bx)^{a_1-1}}{(1+bx)^{a_1+a_2}}, \quad x, a_1, a_2, b > 0. \quad (2.67)$$

The beta prime distribution has numerous parametrizations, perhaps since there are several ways to interpret it. One of the most well-known is the F distribution, which is a reparametrization of the unit scale beta prime distribution. In [Frery et al., 1997], the authors argue that the generalized beta prime distribution describes the intensity SAR distribution of extremely heterogeneous (e.g. urban) areas, but they called it the G^0 distribution and used yet another parametrization. In [Dubey, 1970], the author derived the beta prime distribution as a special case of the compound gamma distribution. That is, if X is gamma distributed for fixed scale (conditional PDF $\gamma(\cdot)$), but the scale itself is a RV which also follows the gamma distribution. Dubey found that X is then beta prime distributed, but used a parametrization which was different still. These alternative interpretations are not included merely to confuse the reader,

¹⁶Regarding naming conventions, this distribution is also known as the beta distribution of the second kind. However, the literature usually reserves these terms for the unscaled versions ($b = 1$), with scaling and/or the power transformation reserved for distributions with the prefix *generalized*. Thus, we deviate slightly from these conventions, but if X has the unscaled PDF $\beta'(x; a_1, a_2)$, then bX for constant $b > 0$ simply has PDF $\beta'(x; a_1, a_2, b)$, i.e. the conversion is trivial in practice, but essential for our use. For more details, see e.g. [Nicolas, 2016], where $\beta'(x; a_1, a_2, b)$ is the special case (unit power parameter) of what he called the generalized F distribution.

but to illustrate the many uses of $\beta'(x; a_1, a_2, b)$. Our quotient definition in eq. (2.67) emphasizes how it can be used in multitemporal imaging (change detection), where two independent images of the same scene are analyzed with the null hypothesis that the scene has not undergone change. If each image is assumed gamma distributed, the test statistic will follow the beta prime distribution, and values very different from unity suggest a rejection of the null hypothesis, indicating that something has changed in the scene.

The classical moments were derived in [Goovaerts and De Pril, 1980], where they found that only the moments of order $n < a_2$ exist. Deriving the log-cumulants is comparatively much simpler, as [Nicolas, 2002] included a general expression for the quotient of two independent RVs. If X follows the beta prime distribution, we have from [Nicolas, 2016] that

$$\kappa_{\beta',n} = \begin{cases} \psi^{(0)}(a_1) - \psi^{(0)}(a_2) - \log b & n = 1, \\ \psi^{(n-1)}(a_1) + (-1)^n \psi^{(n-1)}(a_2) & n \geq 2. \end{cases} \quad (2.68)$$

Recall from Section 2.3.9 that the shape parameter is interpreted as the number of looks in SAR imagery, and that it is global and independent of the scene. This has even greater implications for the beta prime distribution, as we can assume that $a_1 = a_2 = L$, i.e. the number of looks is equal for both images. This is the case regardless of whether the scene has undergone change, and its global nature allows us to treat it as known. We can then use the parametrization

$$\beta'(x; a, b) \equiv \frac{b\Gamma(2a)}{[\Gamma(a)]^2} \frac{(bx)^{a-1}}{(1+bx)^{2a}}, \quad x, a, b > 0, \quad (2.69)$$

where we used a instead of L since this PDF will be used as a kernel later in this thesis. The log-cumulants from eq. (2.68) are simplified to

$$\kappa_{\beta',n} = \begin{cases} -\log b & n = 1, \\ 2\psi^{(n-1)}(a) & n = 2, 4, 6, \dots, \\ 0 & n = 3, 5, 7, \dots, \end{cases} \quad (2.70)$$

i.e. the odd log-cumulants, except for the first, are all zero.

2.3.11 The Log-Normal Distribution

The non-negative RV X follows a log-normal distribution if $\log X$ follows a Gaussian distribution, that is

$$\alpha(\log x; m = \mu, \varsigma = \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad (2.71)$$

where $\mu \equiv E\{\log X\} = \mu_1$ is the log-mean and $\sigma^2 \equiv E\{(\log X - \mu)^2\} = \kappa_2$ is called the *log-variance*. Let $\Lambda(x; \mu, \sigma)$ denote the PDF of X , commonly called the log-normal distribution. It becomes

$$\Lambda(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad x > 0, \quad (2.72)$$

and obviously $\Lambda(x; \mu, \sigma) = 0 \forall x \leq 0$. The factor x^{-1} is explained in [Kendall et al., 1994] as arising from the conservation of differential probability

$$\Lambda(x; \mu, \sigma) = \alpha(\log x; \mu, \sigma) \left| \frac{d \log x}{dx} \right| = \frac{\alpha(\log x; \mu, \sigma)}{x}. \quad (2.73)$$

The standardized version has zero log-mean and unit log-variance, i.e.

$$\Lambda(u) = \frac{1}{u\sqrt{2\pi}} \exp\left\{-\frac{(\log u)^2}{2}\right\}. \quad (2.74)$$

In order to convert between the two, use

$$\log u = \frac{\log x - \mu}{\sigma} \Rightarrow u = (xe^{-\mu})^{1/\sigma}, \quad (2.75)$$

which can be interpreted as scaling being the logarithmic equivalent of shifting (accounting for the log-mean of the data), and the power transformation being the logarithmic equivalent of scaling (accounting for the log-variance of the data). The relationship between the standardized and non-standardized log-normal PDFs is

$$\Lambda((xe^{-\mu})^{1/\sigma}) = \frac{1}{\sqrt{2\pi}}(xe^{-\mu})^{-1/\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad (2.76)$$

$$\frac{1}{x\sigma}(xe^{-\mu})^{1/\sigma} \Lambda((xe^{-\mu})^{1/\sigma}) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad (2.77)$$

as also [Pastor et al., 2016] states. In terms of x and u , the conversion is simply

$$\frac{u}{x\sigma} \Lambda(u) = \Lambda(x; \mu, \sigma). \quad (2.78)$$

The log-normal distribution is another example of a distribution well suited to the MT, in that it has both problematic behavior w.r.t. the classical framework and simple and elegant results using MK statistics. [Heyde, 1963] demonstrated how the set of all (classical) moments of the log-normal distribution exists, but does not uniquely define a distribution. That is, other distributions than the log-normal may have the same moments. This is an unusual property, as mentioned in Section 2.2.2. In [Tellambura and Senaratne, 2010], the authors discuss the CF and moment generating function of the log-normal distribution, which have serious limitations w.r.t. their existence and lack exact closed-form expressions. Using MK statistics however, many of the elegant properties of the Gaussian distribution from Section 2.2.5 are mirrored: The MK CF of a log-normally distributed RV is given in [Nicolas, 2016] as

$$\phi_{\Lambda}(s) = e^{\mu(s-1)} e^{\sigma^2 \frac{(s-1)^2}{2}}, \quad (2.79)$$

giving the log-cumulants

$$\kappa_{\Lambda,n} = \begin{cases} \mu & n = 1, \\ \sigma^2 & n = 2, \\ 0 & n \geq 3. \end{cases} \quad (2.80)$$

2.4 The Classical Statistics Series Expansions

In this section, the series expansions using the (classical) moments and cumulants will be presented. An important reference is [Wallace, 1958], where the author summarizes the historical aspects back to the original works [Chebyshev, 1860], [Gram, 1883], [Chebyshev, 1890], [Edgeworth, 1905], [Charlier, 1905], and [Charlier, 1906]. [Hald, 2000] also recounts the history of the Gram-Charlier series, with an emphasis on the cumulants. The derivation and even the notation has not changed significantly since [Wallace, 1958], as seen in the modern and more comprehensive reference [Kendall et al., 1994]. This section presents the approach and results of these references, while we suggest some changes and improvements to these classical methods in Chapter 4.

2.4.1 The Chebyshev-Hermite Polynomials

The series expansions were originally derived using an orthogonality property of the Chebyshev-Hermite (or just Hermite) polynomials $H_n(x)$. This property allows us to write the PDF $f_X(x)$ as

$$f_X(x) = \sum_{n=0}^{\infty} \xi_n H_n(x) \alpha(x), \quad (2.81)$$

and easily evaluate as many of the coefficients ξ_0, ξ_1, \dots as we want. Before proceeding with the derivation of the series, we investigate the necessary aspects of these polynomials.

The n th degree Hermite polynomial $H_n(x)$ is defined in terms of the derivatives of the standardized Gaussian PDF in eq. (2.23), given in [Kendall et al., 1994] as

$$(-D_x)^n \alpha(x) = H_n(x) \alpha(x), \quad (2.82)$$

where the factor $(-1)^n$ ensures that the leading coefficient is 1.¹⁷ The Rodrigues formula¹⁸ for $H_n(x)$, which connects the polynomials to the derivatives of the Gaussian kernel, can be found in e.g. [Rottmann, 2003] or easily derived from eq. (2.82) as

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} D^n e^{-\frac{x^2}{2}}. \quad (2.83)$$

The first few Hermite polynomials are

$$H_0(x) = 1, \quad (2.84)$$

$$H_1(x) = x, \quad (2.85)$$

$$H_2(x) = x^2 - 1, \quad (2.86)$$

$$H_3(x) = x^3 - 3x, \quad (2.87)$$

$$H_4(x) = x^4 - 6x^2 + 3, \quad (2.88)$$

$$H_5(x) = x^5 - 10x^3 + 15x, \quad (2.89)$$

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15, \quad (2.90)$$

$$H_7(x) = x^7 - 21x^5 + 105x^3 - 105x. \quad (2.91)$$

As [Rottmann, 2003] remarks, $H_n(x)$ is a solution to the differential equation

$$D_x^2 y - x D_x y + n y = 0, \quad (2.92)$$

for $n \in \mathbb{Z}_{\geq 0}$. Also, the Hermite polynomials are orthogonal w.r.t. $\alpha(x)$ in the sense that

$$\int_{-\infty}^{\infty} H_k(x) H_n(x) \alpha(x) dx = n! \delta_{kn}, \quad (2.93)$$

where

$$\delta_{kn} = \begin{cases} 1 & k = n, \\ 0 & k \neq n, \end{cases} \quad (2.94)$$

is the Kronecker delta function. This property will be used when deriving the Gram-Charlier series.

¹⁷The *leading coefficient* is the coefficient of the term of the highest power in the polynomial, i.e. in the n th degree polynomial $P_n(x) = cx^n + P_{n-1}(x)$, the leading coefficient is c .

¹⁸The term *Rodrigues formula* is used for any polynomial definition of this form, as explained in [Askey, 2005].

2.4.2 The Gram-Charlier Gaussian Kernel Series

As before, let $f_X(x)$ denote the PDF of X . If we assume that the PDF can be written in terms of the derivatives of the Gaussian kernel, then eq. (2.82) allows us to replace the differential operator with the known $H_n(x)$ polynomials and get eq. (2.81). To find the coefficients ξ_n , we multiply both sides with $H_k(x)$, integrate from $-\infty$ to $+\infty$ and swap the order of integration and summation. Using also the orthogonality property, we get

$$\int_{-\infty}^{\infty} f_X(x) H_k(x) dx = \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \xi_n H_k(x) H_n(x) \alpha(x) = \sum_{n=0}^{\infty} \xi_n k! \delta_{kn} = \xi_n n!, \quad (2.95)$$

$$\xi_n = \frac{1}{n!} \int_{-\infty}^{\infty} f_X(x) H_n(x) dx. \quad (2.96)$$

Since $H_n(x)$ is a known polynomial of x , each coefficient ξ_n is a linear combination of the moments m_n . For standardized X (shifted to $m_1 = 0$, scaled to $\zeta^2 = 1$), the first few coefficients are found using eqs. (2.84) through (2.91)

$$\xi_0 = 1, \quad (2.97)$$

$$\xi_1 = \xi_2 = 0, \quad (2.98)$$

$$\xi_3 = \frac{1}{6} m_3, \quad (2.99)$$

$$\xi_4 = \frac{1}{24} (m_4 - 3), \quad (2.100)$$

$$\xi_5 = \frac{1}{120} (m_5 - 10m_3), \quad (2.101)$$

$$\xi_6 = \frac{1}{720} (m_6 - 15m_4 + 30), \quad (2.102)$$

$$\xi_7 = \frac{1}{5040} (m_7 - 21m_5 + 105m_3). \quad (2.103)$$

In [Kendall et al., 1994] the authors represented the coefficients in terms of the cumulants c_n :

$$\xi_0 = 1, \quad (2.104)$$

$$\xi_1 = \xi_2 = 0, \quad (2.105)$$

$$\xi_3 = \frac{1}{6} c_3, \quad (2.106)$$

$$\xi_4 = \frac{1}{24} c_4, \quad (2.107)$$

$$\xi_5 = \frac{1}{120} c_5, \quad (2.108)$$

$$\xi_6 = \frac{1}{720} (c_6 + 10c_3^2), \quad (2.109)$$

$$\xi_7 = \frac{1}{5040} (c_7 + 35c_3c_4), \quad (2.110)$$

where eqs. (2.10) and (2.11) can be used to convert between the two representations. Naturally, the true moments and cumulants of the distribution we seek to model are usually not known to us, and we must use the empirical values instead. Inserting the coefficients into eq. (2.81) gives the Gram-Charlier series around the Gaussian kernel

$$f_X(x) = \alpha(x) \left[1 + \frac{1}{6} c_3 H_3(x) + \frac{1}{24} c_4 H_4(x) + \dots \right], \quad (2.111)$$

where the cumulant-based representation of the coefficients was used. Each term in the series compensates for the n th cumulant c_n of X , since the Gaussian kernel has $c_{\alpha,n} = 0$ for $n \geq 3$. The Gram-Charlier series can be adjusted to account for non-standardized RVs, but here X is assumed standardized for brevity.¹⁹ Using the cumulant representation simplifies the Gram-Charlier series, since the coefficients ξ_n have simpler expressions when stated in terms of cumulants for $n \geq 4$.

2.4.3 The Laguerre Polynomials

The generalized (or associated) Laguerre polynomials are denoted $L_n^{(a)}(x)$, where $n \in \mathbb{Z}_{\geq 0}$. They are solutions to Laguerre's equation

$$xy'' + (a + 1 - x)y' + ny = 0, \quad (2.112)$$

which can be found in [Szeg, 1939]. The same reference also gives the Rodrigues formula as

$$L_n^{(a)}(x) = \frac{x^{-a}e^x}{n!} D^n [x^{a+n}e^{-x}]. \quad (2.113)$$

Shifting $a \rightarrow a - 1$, we can rewrite this as

$$L_n^{(a-1)}(x) \frac{x^{a-1}}{\Gamma(a)} e^{-x} = \frac{1}{n! \Gamma(a)} D^n [x^{a-1+n}e^{-x}], \quad (2.114)$$

where $1/\Gamma(a)$ can be moved inside the argument of $D^n[\cdot]$ since it is a constant factor w.r.t. x . Then, we recognize the gamma kernel $\gamma(x; a)$ with scale $b = 1$ from eq. (2.58) to see that

$$L_n^{(a-1)}(x) \gamma(x; a) = \frac{1}{n!} D^n [x^n \gamma(x; a)] \quad \text{for } x \geq 0. \quad (2.115)$$

This is analogous to the how the Hermite polynomials were defined in eq. (2.82), with the difference being a factor $(-1)^n/n!$ and the multiplication of the kernel with x^n .

The first few generalized Laguerre polynomials are

$$L_0^{(a)}(x) = 1, \quad (2.116)$$

$$L_1^{(a)}(x) = -x + a + 1, \quad (2.117)$$

$$L_2^{(a)}(x) = \frac{x^2}{2} - (a + 2)x + \frac{(a + 1)(a + 2)}{2}, \quad (2.118)$$

$$L_3^{(a)}(x) = -\frac{x^3}{6} + \frac{(a + 3)x^2}{2} - \frac{(a + 2)(a + 3)x}{2} + \frac{(a + 1)(a + 2)(a + 3)}{6}, \quad (2.119)$$

$$L_4^{(a)}(x) = \frac{x^4}{24} - \frac{(a + 4)x^3}{6} + \frac{(a + 3)(a + 4)x^2}{4} - \frac{(a + 2)(a + 3)(a + 4)x}{6} + \frac{(a + 1)(a + 2)(a + 3)(a + 4)}{24}. \quad (2.120)$$

In the same way that the Hermite polynomials are orthogonal w.r.t. the Gaussian kernel in eq. (2.93), the Laguerre polynomials are orthogonal²⁰ w.r.t. $x^a e^{-x}$:

$$\int_0^{\infty} L_n^{(a-1)}(x) L_k^{(a-1)}(x) x^{a-1} e^{-x} dx = \Gamma(a) \binom{n+a-1}{n} \delta_{nk} = \Gamma(a) \frac{\Gamma(n+a)}{\Gamma(n+1)\Gamma(a)} \delta_{nk}, \quad (2.121)$$

$$\int_0^{\infty} L_n^{(a-1)}(x) L_k^{(a-1)}(x) \gamma(x; a) dx = \frac{\Gamma(n+a)}{\Gamma(n+1)\Gamma(a)} \delta_{nk}. \quad (2.122)$$

¹⁹See Section 4.2.2 for the necessary corrections for non-standardized X .

²⁰This result can be found in [Szeg, 1939], while e.g. [Fowler, 1996] generalizes the binomial coefficient to allow for non-integer a .

Recall that $\gamma(x; a)$ assumes unit scale, but in order to tailor the kernel to data, arbitrary scale b is necessary. Appendix A.1 contains some straightforward proofs that the orthogonality property of eq. (2.122) and the Rodrigues formula of eq. (2.113) are extended to the scaled kernel by replacing $L_n^{(a-1)}(x)$ with $L_n^{(a-1)}(bx)$. This allows us to use the two-parameter $\gamma(x; a, b)$ in the following.

2.4.4 The Gram-Charlier Gamma Kernel Series

Historically, the Gaussian distribution has been almost exclusively the kernel of choice for the Gram-Charlier series. After developing the Gram-Charlier series with a Gaussian kernel, [Kendall et al., 1994] mentions that the same approach is possible with two other pairs of orthogonal polynomials and kernels. The most important pair for this thesis is the Laguerre polynomials and gamma kernel pair,²¹ with their orthogonal relationship stated in eq. (2.122). The Gram-Charlier series with a gamma kernel was also derived in [Gaztanaga et al., 2000], where it was given a different name. [Inglada and Mercier, 2007] suggested (but did not present) a series expansion around the gamma kernel, if the data were more heavy-tailed than the SAR data they used.

Letting $f_X(x)$ be the PDF of a non-negative RV X , assume that the PDF can be written in terms of (scaled) Laguerre polynomials and the gamma kernel, i.e.

$$f_X(x) = \sum_{n=0}^{\infty} \xi_n L_n^{(a-1)}(bx) \gamma(x; a, b). \quad (2.123)$$

Repeating the steps in Section 2.4.2 and using the orthogonality property of eq. (2.122), gives an expression for the coefficients ξ_n

$$\int_0^{\infty} f_X(x) L_k^{(a-1)}(bx) dx = \int_0^{\infty} \sum_{n=0}^{\infty} \xi_n L_k^{(a-1)}(bx) L_n^{(a-1)}(bx) \gamma(x; a, b) dx = \sum_{n=0}^{\infty} \xi_n \frac{\Gamma(n+a)}{\Gamma(n+1)\Gamma(a)} \delta_{kn}, \quad (2.124)$$

$$\xi_n = \frac{n!}{\prod_{i=0}^{n-1} (a+i)} \int_0^{\infty} f_X(x) L_n^{(a-1)}(bx) dx, \quad (2.125)$$

where the integration limits now reflect the non-negative nature of X .²² Like $H_n(x)$, the Laguerre polynomials are known polynomials in x , reducing the integral to a linear combination of the moments m_n of X . The first few coefficients are

$$\xi_0 = 1, \quad (2.126)$$

$$\xi_1 = -\frac{b}{a} m_1 + 1, \quad (2.127)$$

$$\xi_2 = \frac{b^2}{a(a+1)} m_2 - \frac{2b}{a} m_1 + 1. \quad (2.128)$$

²¹The other pair is the Jacobi polynomials and the beta distribution, see [Durbin and Watson, 1951]. Note that the beta distribution is closely related to the beta prime distribution, which we discussed in Section 2.3.10, although Durbin and Watson used the unit scale version.

²²Also, since n is a non-negative integer we have $\Gamma(n+1) = n!$ and $\Gamma(n+a)/\Gamma(a) = \prod_{i=0}^{n-1} (a+i)$.

In practice, the Laguerre polynomials and gamma kernel are scaled and shaped to match the data by solving the equations

$$\begin{aligned} m_1 &= \frac{a}{b} \\ m_2 &= \frac{a(a+1)}{b^2} \end{aligned} \quad \Rightarrow \quad \begin{aligned} a &= \frac{\langle m_1 \rangle^2}{\langle m_2 \rangle - \langle m_1 \rangle^2} = \frac{\langle m_1 \rangle^2}{\langle \varsigma \rangle^2} \\ b &= \frac{\langle m_1 \rangle}{\langle m_2 \rangle - \langle m_1 \rangle^2} = \frac{\langle m_2 \rangle}{\langle \varsigma \rangle^2} \end{aligned}, \quad (2.129)$$

for a and b by replacing m_1 and m_2 with the empirical moments $\langle m_1 \rangle$ and $\langle m_2 \rangle$. Then, the coefficients are simplified to

$$\xi_0 = 1, \quad (2.130)$$

$$\xi_1 = \xi_2 = 0, \quad (2.131)$$

$$\xi_3 = -\frac{b^3}{a(a+1)(a+2)}m_3 + 1, \quad (2.132)$$

$$\xi_4 = \frac{b^4}{a(a+1)(a+2)(a+3)}m_4 - \frac{4b^3}{a(a+1)(a+2)}m_3 + 3, \quad (2.133)$$

where we usually have to insert the empirical moments, as we discussed in Section 2.4.2. The Gram-Charlier gamma kernel series is

$$f_X(x) = \left[1 + \sum_{n=3}^{\infty} \xi_n L_n^{(a-1)}(bx) \right] \gamma(x; a, b). \quad (2.134)$$

In Section 4.2.3, we present some observations and results which simplify this classical method and provides some additional insight.

2.4.5 The Edgeworth Series

Inherent to all series expansion methods is the fact that the truncated series seldom represents a true PDF. That is, the approximation can take negative values, it may not integrate to 1, or both. The Edgeworth series does *not* attempt to correct this, but as [Blinnikov and Moessner, 1998] showed, it does significantly improve on the Gram-Charlier series, both with few terms and asymptotically.

There are several ways to derive the Edgeworth series – [Blinnikov and Moessner, 1998] had one approach, [Kendall et al., 1994] had another. In this thesis, we derive the Edgeworth series the same way as [Hall, 2013]. Edgeworth sought to modify the Gaussian kernel so that the asymptotic behavior described by the central limit theorem was faster. Let Z_1, Z_2, \dots, Z_r be *independently identically distributed* (IID) RVs with mean m , variance ς^2 and higher-order cumulants $c_n = \varsigma^n \lambda_n$. In the same way that the moments were standardized in Section 2.2.2, $\lambda_n = c_n/\varsigma^n$ are the cumulants scaled by the corresponding inverse power of the standard deviation. The use of λ_n simplifies the Edgeworth series derivation and expression. Let

$$X = \frac{1}{\sqrt{r}} \sum_{i=1}^r \frac{Z_i - m}{\varsigma} \quad (2.135)$$

denote the standardized sum (i.e. zero mean, unit variance) of Z_1, Z_2, \dots, Z_r .²³ Now, combining

²³The central limit theorem states that as $r \rightarrow \infty$ then $f_X(x) \rightarrow \alpha(x)$, where $\alpha(x)$ is the standardized Gaussian kernel.

equations (2.7) and (2.8), the CFs of X and the kernel $\alpha(x)$ are

$$\Phi_X(\omega) = \exp \left\{ \sum_{n=1}^{\infty} c_{X,n} \frac{(j\omega)^n}{n!} \right\}, \quad (2.136)$$

$$\Phi_\alpha(\omega) = \exp \left\{ \sum_{n=1}^{\infty} c_{\alpha,n} \frac{(j\omega)^n}{n!} \right\}, \quad (2.137)$$

where $c_{X,n}$ is the n th-order cumulant of the distribution $f_X(x)$. Combining these, it is possible to express the CF of the unknown PDF $f_X(x)$ as a function of the known CF of the Gaussian kernel, i.e.

$$\Phi_X(\omega) = \exp \left\{ \sum_{n=1}^{\infty} [c_{X,n} - c_{\alpha,n}] \frac{(j\omega)^n}{n!} \right\} \Phi_\alpha(\omega). \quad (2.138)$$

Now the elegant properties of the cumulants can be exploited:

1. The cumulants of order $n \geq 2$ are shift-invariant (Section 2.2.3).
2. Scaling a RV with ς^{-1} like in eq. (2.135) implies that c_n is scaled by ς^{-n} (Section 2.2.3).
3. The cumulants of a sum of independent RVs are the sums of the cumulants of each RV (Section 2.2.4).
4. The cumulants of $\alpha(x; m, \varsigma)$ are $c_{\alpha,n} = 0$ for $n \geq 3$ (Section 2.2.5).
5. Since X by definition has zero mean and unit variance, its first and second order cumulants are the same as those of a RV which follows the standardized $\alpha(x)$.

Using all of these properties gives the cumulant differences

$$c_{X,n} - c_{\alpha,n} = \begin{cases} 0 & n = 1, 2, \\ \frac{\lambda_n}{r^{\frac{n}{2}-1}} & n \geq 3. \end{cases} \quad (2.139)$$

The first two cumulant differences follow directly from property 5. Disregarding order $n = 1$, property 1 says that $Z_i - m$ has the same cumulants as Z_i (i.e. $\varsigma^n \lambda_n$). Property 3 then says that the sum of all Z_i has cumulants $r\varsigma^n \lambda_n$, as they are r IID RVs. Since X is this sum, scaled with $(\sqrt{r}\varsigma)^{-1}$, property 2 gives the cumulants of X as $\lambda_n/r^{\frac{n}{2}-1}$. Finally, property 4 says that for $n \geq 3$, the cumulant differences are just the cumulants of X .

The FT property $(-D)^n f(x) \xleftrightarrow{\mathcal{F}} (j\omega)^n F_{\mathcal{F}}(\omega)$ now allows for an inverse FT on the CF, giving an expression for the PDF

$$f_X(x) = \exp \left\{ \sum_{n=3}^{\infty} \frac{\lambda_n}{r^{\frac{n}{2}-1}} \frac{(-D)^n}{n!} \right\} \alpha(x). \quad (2.140)$$

Using the power series representation of e^x , this can be thought of as an infinite sum of infinite sums,

$$f_X(x) = \alpha(x) \sum_{k=0}^{\infty} \frac{\left[\sum_{n=3}^{\infty} \frac{\lambda_n}{r^{\frac{n}{2}-1}} \frac{(-D)^n}{n!} \right]^k}{k!}. \quad (2.141)$$

The question becomes how to prioritize which terms are included in the finite truncation. It is possible to recover the Gram-Charlier series by sorting the terms by their order of D . Edgeworth's idea was to sort the terms by their dependency on r instead:

$$f_X(x) = \alpha(x) + r^{-\frac{1}{2}} \left[\frac{1}{6} \lambda_3 H_3(x) \right] \alpha(x) + r^{-1} \left[\frac{1}{24} \lambda_4 H_4(x) + \frac{1}{72} \lambda_3^2 H_6(x) \right] \alpha(x) + O\left(r^{-\frac{3}{2}}\right), \quad (2.142)$$

where the r -factors disappear when inserting the empirical cumulants for $\lambda_n/r^{\frac{n}{2}-1}$.

Compared to the Gram-Charlier series in eq. (2.111), the Edgeworth series approximation when disregarding terms of $O(r^{-3/2})$ differs only by the term $\lambda_3^2 H_6(x)/72$, corresponding to $k = 2, n = 3$ in eq. (2.141). X is by definition standardized, but allowing for non-standardized data is done the same way as for the Gram-Charlier series, see Section 4.2.2.

[Blinnikov and Moessner, 1998] provided a formula for finding the terms associated with any given power of $r^{-\frac{1}{2}}$, but it is omitted from this thesis for the sake of brevity. Chapter 4 provides a novel and perhaps simpler way of representing both the Gram-Charlier and Edgeworth series and also contains some observations on this classical approach, including a more thorough explanation of the derivation leading up to eq. (2.138).

2.5 The Bell Polynomials and the Stirling Numbers

This section introduces the Bell polynomials and Stirling numbers, but will be limited to the properties which are required later in this thesis. For a more thorough and in-depth review, see e.g. [Johnson, 2002].

2.5.1 Defining the Bell Polynomials

The Bell polynomials are named in honor of Eric Temple Bell, who introduced them in [Bell, 1927] under the name partition polynomials. The *partial Bell polynomials* $B_{n,r}(\cdot)$ are defined in e.g. [Mihoubi, 2008] as

$$B_{n,r}(x_1, x_2, \dots, x_{n-r+1}) = \sum_{\mathcal{K}_r} n! \prod_{i=1}^{n-r+1} \frac{1}{j_i!} \left(\frac{x_i}{i!}\right)^{j_i}, \quad (2.143)$$

where the sum is over the set \mathcal{K}_r of all combinations of non-negative integers j_1, j_2, j_3, \dots which satisfy

$$j_1 + j_2 + j_3 + \dots = r, \quad (2.144)$$

$$j_1 + 2j_2 + 3j_3 + \dots = n. \quad (2.145)$$

For given values of n and r , the integers j_i with index $i > n - r + 1$ must all be zero. Thus, the sums above can be truncated at j_{n-r+1} , which implies that the partial Bell polynomial $B_{n,r}(\cdot)$ has argument $x_1, x_2, \dots, x_{n-r+1}$, as indicated in eq. (2.143).

The n th *complete Bell polynomial* $B_n(x_1, \dots, x_n)$ is defined in [Mihoubi, 2008] as

$$B_n(x_1, \dots, x_n) = \sum_{r=1}^n B_{n,r}(x_1, x_2, \dots, x_{n-r+1}). \quad (2.146)$$

In this thesis, the complete Bell polynomials are the most used, so the prefix is usually omitted. In any case the notation clearly distinguishes between the two versions, since the subscripts (n

and n, r) differ. The first complete Bell polynomials are

$$B_0 = 1, \quad (2.147)$$

$$B_1(x_1) = x_1, \quad (2.148)$$

$$B_2(x_1, x_2) = x_1^2 + x_2, \quad (2.149)$$

$$B_3(x_1, \dots, x_3) = x_1^3 + 3x_1x_2 + x_3, \quad (2.150)$$

$$B_4(x_1, \dots, x_4) = x_1^4 + 6x_1^2x_2 + 4x_1x_3 + 3x_2^2 + x_4, \quad (2.151)$$

$$B_5(x_1, \dots, x_5) = x_1^5 + 10x_2x_1^3 + 15x_2^2x_1 + 10x_3x_1^2 + 10x_3x_2 + 5x_4x_1 + x_5, \quad (2.152)$$

$$B_6(x_1, \dots, x_6) = x_1^6 + 15x_2x_1^4 + 20x_3x_1^3 + 45x_2^2x_1^2 + 15x_2^3 + 60x_3x_2x_1 + 15x_4x_1^2 + 10x_3^2 \quad (2.153) \\ + 15x_4x_2 + 6x_5x_1 + x_6.$$

Finally, recall that the Bell polynomials were used in eqs. (2.10), (2.11), (2.42) and (2.43) for converting between (log-)moments and (log-)cumulants.

2.5.2 A Property of the Bell Polynomials

Using an expression from [Mihoubi, 2008] and recognizing the power series of e^x gives

$$\exp \left\{ \sum_{k=1}^{\infty} x_k \frac{s^k}{k!} \right\} = \sum_{n=0}^{\infty} B_n(x_1, \dots, x_n) \frac{s^n}{n!}. \quad (2.154)$$

This is the key property of the Bell polynomials w.r.t. this thesis, and it will be used on several occasions.

2.5.3 The Bell Polynomials and Faà di Bruno's Formula

The use of Bell polynomials in this thesis is in fact just an alternative representation of a special case of Faà di Bruno's formula. In its most general form, Faà di Bruno's formula extends the chain rule of differentiation to higher derivatives. It is presented in [Johnson, 2002] as

$$D_s^n g(f(s)) = \sum_{\mathcal{K}_n} n! D_t^r g(t = f(s)) \prod_{i=1}^n \frac{1}{j_i!} \left(\frac{D_s^i f(s)}{i!} \right)^{j_i}, \quad (2.155)$$

where $D_t^r g(t = f(s))$ means that $g(\cdot)$ is differentiated $r = j_1 + \dots + j_n$ times w.r.t. to $f(s)$, not s .²⁴ The sum is over the set \mathcal{K}_n , which is similar to the set \mathcal{K}_r used in eq. (2.143), but r is allowed to take any integer value.²⁵ This allows us to combine eqs. (2.143) and (2.146) to get

$$B_n(x_1, \dots, x_n) = \sum_{\mathcal{K}_n} n! \prod_{i=1}^n \frac{1}{j_i!} \left(\frac{x_i}{i!} \right)^{j_i}, \quad (2.156)$$

i.e. using \mathcal{K}_n fulfills the same purpose as a summation over every possible \mathcal{K}_r . The connection between log-moments and log-cumulants²⁶ is obtained as a special case of eq. (2.155), where

²⁴A simple example is $g(s) = e^s$, $f(s) = as$ with constant a . Then $D_s^r g(f(s)) = D_s^r e^{as} = a^r e^{as}$, but the exponential function is its own derivative, i.e. $D_t^r g(t) = g(t)$, so $D_t^r g(t = f(s)) = g(f(s)) = e^{as}$.

²⁵Eqs. (2.144) and (2.145) only have solutions for $r \in \{1, 2, \dots, n\}$, which implies that other values of r do not contribute to \mathcal{K}_n .

²⁶As this relationship is exactly the same for the classical moments and cumulants with identical results as the MK case, only the latter is reviewed here.

$f(s)$ is the MK CGF $\varphi(s)$ and $g(t) = e^t$, i.e. $g(f(s))$ is the MK CF $\phi(s)$. Then, all derivatives of $g(t)$ w.r.t. t also equal $\phi(s)$, which greatly simplifies the expression to

$$D_s^n \phi(s) = \sum_{\kappa_n} n! \phi(s) \prod_{i=1}^n \frac{1}{j_i!} \left(\frac{D_s^i \varphi(s)}{i!} \right)^{j_i}. \quad (2.157)$$

Eqs. (2.38) and (2.41) state that when evaluated at $s = 1$, the n th derivative of $\phi(s)$ is μ_n and the n th derivative of $\varphi(s)$ is κ_n . Also, $\phi(1) = 1$ trivially, so eq. (2.156) gives

$$\mu_n = \sum_{\kappa_n} n! \prod_{i=1}^n \frac{1}{j_i!} \left(\frac{\kappa_i}{i!} \right)^{j_i} = B_n(\kappa_1, \dots, \kappa_n), \quad (2.158)$$

which completes a proof of eq. (2.43). While this relationship is previously known (in terms of the classical descriptors), other sources like [Pitman, 2002] proves it in a different way. We performed the above derivation²⁷ because the result is so central to this thesis (if such an obvious mathematical traversal can be called a derivation). Also, in doing so we introduced notation which is used in the important references [Blinnikov and Moessner, 1998] and [Pastor et al., 2014].

2.5.4 Defining the Stirling Numbers

Named after James Stirling, the Stirling numbers of the second kind are denoted by either $S(n, k)$ or $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$, with the latter form adopted by this thesis. These are the only kind of Stirling numbers that will be used in this thesis, allowing us to simply refer to them as the Stirling numbers. [Lengyel, 1994] interprets them as the number of possible ways to partition n labelled objects into k subsets which are nonempty and unlabelled, with the explicit definition

$$\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} = \frac{1}{k!} \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} i^n, \quad (2.159)$$

where $\binom{k}{i}$ is the binomial coefficient. Some general observations can be made:

- $\left\{ \begin{smallmatrix} n \\ n \end{smallmatrix} \right\} = 1 \forall n \in \{0, 1, 2, \dots\}$, since there is exactly one way to partition n objects into n unlabelled and non-empty subsets (one object into each subset).
- $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} = 0 \forall k > n$, since there can be no such partition without any empty subsets.
- $\left\{ \begin{smallmatrix} n \\ 0 \end{smallmatrix} \right\} = 0 \forall n > 0$, since there is no way to distribute a positive number of objects into zero subsets.
- $\left\{ \begin{smallmatrix} n \\ 1 \end{smallmatrix} \right\} = 1$, since all objects must then go into the only subset.

The first few Stirling numbers are presented in Table 2.2.

2.5.5 A Property of the Stirling Numbers

According to [Wagner, 1996], the Stirling numbers satisfy the equation

$$\sum_{k=0}^n \left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} \prod_{i=0}^{k-1} (x - i) = x^n, \quad (2.160)$$

which we will use in Appendix A.2.1.

²⁷This result was listed in Section 1.5 as a minor contribution of this thesis.

Table 2.2: The first few Stirling numbers. The terms where $n = 0$ or $k = 0$ are excluded since they are all 0 with the exception of $\{0_0\} = 1$. These special cases are discussed in the text.

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$n = 1$	1	0	0	0	0
$n = 2$	1	1	0	0	0
$n = 3$	1	3	1	0	0
$n = 4$	1	7	6	1	0
$n = 5$	1	15	25	10	1

2.6 Dissimilarity Measures of PDFs

It is essential to qualitatively and quantitatively assess the methods presented in this thesis, as their merit cannot be based on mathematical elegance alone. [Blinnikov and Moessner, 1998] show how the classical series expansion methods are prone to severe divergence in several cases. In those instances it is enough to simply say that the methods have qualitatively failed, but otherwise, quantitative studies are necessary to rank competing methods. Specifically, this involves assigning numerical values to the dissimilarity of two PDFs.

2.6.1 Terminology: Divergence, Distance and Metric

Before presenting the dissimilarity measures, a quick review of the relevant terminology is appropriate. Consider a function $d(f(x), g(x))$ which maps the dissimilarity of $f(x)$ and $g(x)$ to \mathbb{R} , and may satisfy any number of the following properties from [Theodoridis and Koutroumbas, 2009]:

1. $d(f(x), g(x)) \geq 0$, non-negativity.
2. $d(f(x), g(x)) = 0 \Leftrightarrow f(x) = g(x)$, definiteness.
3. $d(f(x), g(x)) = d(g(x), f(x))$, symmetry.
4. $d(f(x), g(x)) \leq d(f(x), h(x)) + d(h(x), g(x))$, triangular inequality.

This thesis adopts the terminology of [Frery et al., 2014] where

- $d(f(x), g(x))$ is a *metric* only if all four properties hold.
- $d(f(x), g(x))$ is a *distance* if only the triangular inequality is relaxed.
- $d(f(x), g(x))$ is a *divergence* measure if it is also non-symmetric (i.e. $d(\cdot)$ is only non-negative definite).

These definitions are not ubiquitous, with e.g. [Theodoridis and Koutroumbas, 2009] having a slightly different naming convention. However, the chosen nomenclature is more than sufficient for this thesis, facilitating simple and precise discussions about the dissimilarity measures encountered.

Regarding the choice of dissimilarity measures, [Frery et al., 2014] discussed this in the context of SAR data. That article is concerned with matrix-variate data, but it can be assumed relevant to this thesis since the gamma distribution is a special case of the matrix-variate Wishart distribution considered by Frery et al. Also, if the methods of this thesis are later extended to

the matrix-variate case (building on the theory in [Anfinson and Eltoft, 2011]), the observations and choices of measures made in this sections are easily generalized to match that situation.

2.6.2 The Kullback-Leibler Distance

The Kullback-Leibler distance is a natural choice when measuring distances between distributions. It is based on the Kullback-Leibler divergence, which itself is a popular dissimilarity measure, and is discussed in e.g. [Theodoridis and Koutroumbas, 2009] and [Frery et al., 2014]. The Kullback-Leibler divergence $\tilde{d}_{\text{KL}}(f(x), g(x))$ between two PDFs $f(x), g(x)$ is defined as

$$\tilde{d}_{\text{KL}}(f(x), g(x)) \equiv \int f(x) \log \frac{f(x)}{g(x)} dx, \quad (2.161)$$

where the integral is over the whole domain, in practice \mathbb{R}^1 in the univariate case, with a straightforward generalization to matrix-variate RVs. The notation \tilde{d} is used to clarify that it is (just) a divergence measure, i.e. that it is non-symmetric. This thesis will instead use the symmetrized Kullback-Leibler *distance*

$$d_{\text{KL}}(f(x), g(x)) \equiv \frac{1}{2} \left[\tilde{d}_{\text{KL}}(f(x), g(x)) + \tilde{d}_{\text{KL}}(g(x), f(x)) \right], \quad (2.162)$$

which can also be found in [Frery et al., 2014].

2.6.3 The Bhattacharyya Distance

In order to ensure that the results are assessed comprehensively, an alternative measure is needed. We will use the Bhattacharyya distance, as we assume that it is good measure for our experiments, based on its success in relatively similar scenarios in [Frery et al., 2014]. The Bhattacharyya distance is given in [Kailath, 1967] as

$$d_{\text{B}}(f(x), g(x)) \equiv -\log \int \sqrt{f(x)g(x)} dx, \quad (2.163)$$

where the integral is again over the whole domain. It is clear from the definition that the Bhattacharyya distance satisfies all the properties of a distance measure and that $d_{\text{B}}(f(x), g(x)) \leq \infty$. [Kailath, 1967] also points out that $d_{\text{B}}(\cdot, \cdot)$ does not in general satisfy the triangular inequality, but that

$$\sqrt{1 - \exp\{-d_{\text{B}}(\cdot, \cdot)\}} = \sqrt{1 - \int \sqrt{f(x)g(x)} dx} \quad (2.164)$$

does, making it a metric by the conventions followed here.²⁸

When comparing two particular cases of the same distribution (e.g. two gamma distributions with different shapes and scales), it is often advisable to rewrite eqs. (2.162) and (2.163) in terms of the distribution parameters, as done in [Frery et al., 2014]. However, this thesis is mostly concerned with comparing *different* distributions, several of which have analytically very complicated PDFs, like the K distribution. The remedy is to simply evaluate the PDFs (both true and estimated) at each value of the discretized x , and use eqs. (2.162) and (2.163) to compute the distances directly by numerical integration.

²⁸The closely related Hellinger distance is also assessed in [Frery et al., 2014]. Like the metric in eq. (2.164), it is easily retrieved from the Bhattacharyya distance. That is, the results in this thesis can trivially be converted to other well-known measures, including metrics.

2.6.4 Caveats Related to the Series Expansion Methods

Most reviews of the classical Gram-Charlier and Edgeworth methods of approximating and estimating PDFs, like [Blinnikov and Moessner, 1998], comment on the following qualitative properties of the series expansions:

- They are not true PDFs, i.e. they can take negative values and generally do not integrate to unity.
- They are susceptible to divergence, especially in the tail(s).
- Increasing the number of terms does not necessarily decrease the error.

These problems are inherent to all series expansion methods, including those presented in this thesis, and require special attention when applying dissimilarity measures.

The most prominent issue is the possibility of non-positive estimates in the series expansion methods, i.e. $g(x) = \hat{f}(x) \leq 0$ for some x . This can potentially cause extreme values or failures in eqs. (2.163) and (2.162), which uses ratios and logarithms. This is a problem inherent to the methodology, causing problems also in applications like likelihood ratio testing or change detection, where reciprocals of $\hat{f}(x)$ appear.

A more subtle issue stems from the fact that with the series approximation methods, $\hat{f}(x)$ does not necessarily integrate to 1. If $\int_{\mathbb{R}} \hat{f}(x) dx > 1$, the Bhattacharyya distance in eq. (2.163) is artificially lowered, even yielding negative distance values in some situations.

Both issues are corrected for by altering the PDF estimates. We replacing non-positive values with machine epsilon²⁹ and subsequently standardize the estimates by their integrals.

²⁹Approximately $2.22 \cdot 10^{-16}$ for the double precision data type used here.

Chapter 3

Mellin Kind Series Expansion Framework

Recall the close analogy between classical and MK statistics, which we examined in Sections 2.2 and 2.3. In the classical case we have the Gram-Charlier and Edgeworth series, as presented in Section 2.4, and the motivation behind this thesis can be summarized as wanting to explore if there exists analogies to these series expansions, within MK statistics.

3.1 Fundamentals

We will now introduce a framework for the MK series expansion methods, starting with a MK Gram-Charlier series with arbitrary kernel distribution $\rho(x)$. The derivation closely follows the classical case outlined in Section 2.4, using results from Chapter 2.

3.1.1 The Analogy With the Classical Series Expansion Methods

Let X be a non-negative RV with PDF $f_X(x)$, i.e. $f_X(x) = 0 \forall x < 0$. In addition, all log-cumulants of X must exist. Then, the MK CF $\phi_X(s)$ is given by eq. (2.40) as

$$\phi_X(s) = \exp\{\varphi_X(s)\} = \exp\left\{\sum_{k=1}^{\infty} \kappa_{X,k} \frac{(s-1)^k}{k!}\right\}, \quad (3.1)$$

where $\kappa_{X,k}$ is the k th-order log-cumulant of X . In the same way, let $\rho(x)$ denote an arbitrary kernel¹ with MK CF

$$\phi_\rho(s) = \exp\{\varphi_\rho(s)\} = \exp\left\{\sum_{k=1}^{\infty} \kappa_{\rho,k} \frac{(s-1)^k}{k!}\right\}, \quad (3.2)$$

where $\kappa_{\rho,k}$ is the k th-order log-cumulant of an RV which follows the kernel PDF $\rho(x)$. Analytical expressions for the log-cumulants of a great many distributions are found in [Nicolas, 2016], while [Nicolas and Anfinson, 2012] is a translated version of Nicolas' original work [Nicolas, 2002] which discusses fewer distributions, but is available online in English.

In the same way the CFs were combined in eq. (2.138) when deriving the classical series series, equations (3.1) and (3.2) can be used to give the MK CF of the unknown PDF $f_X(x)$ as

$$\phi_X(s) = \exp\left\{\sum_{k=1}^{\infty} [\kappa_{X,k} - \kappa_{\rho,k}] \frac{(s-1)^k}{k!}\right\} \phi_\rho(s), \quad (3.3)$$

¹Naturally, the kernel has to have support $\mathbb{R}_{>0}$ and all its log-moments (and thus log-cumulants) must exist.

i.e. it is stated in terms of the known MK CF for the kernel, and the log-cumulants of X . This expression can be simplified by defining the *log-cumulant differences* $\Delta\kappa_k \equiv \kappa_{X,k} - \kappa_{\rho,k}$. Using the result in eq. (2.154) gives²

$$\phi_X(s) = \left[\sum_{n=0}^{\infty} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \frac{(s-1)^n}{n!} \right] \phi_\rho(s). \quad (3.4)$$

The factors $B_n(\cdot)/n!$ are constants w.r.t. s and are, along with the summation, not affected by the linear inverse MT. However, a MT property for inverse transformations of $(s-1)^n$ is necessary in order to proceed.

3.1.2 A Mellin Derivative

At this point in the classical case in Section 2.4.5, the convenient FT property $(-D)^n f(x) \xleftrightarrow{\mathcal{F}} (j\omega)^n F_{\mathcal{F}}(\omega)$ allowed for an inverse transform resulting in an expression for $f_X(x)$. Table 2.1 does *not* include a result which is directly applicable to eq. (3.4), but [Bateman et al., 1954] has the following MT property on page 308:

$$(Dx)^n f(x) \xleftrightarrow{\mathcal{M}} (-1)^n (s-1)^n F_{\mathcal{M}}(s), \quad (3.5)$$

where $(Dx)^n f(x)$ must be continuous from 0 to ∞ for any non-negative integer n . This MT property apparently contradicts eq. (2.31), if not for the fact that $(Dx)^n$ is non-commutative, i.e. $(Dx)^n f(x) \neq D^n x^n f(x)$. These are operators based on the derivative operator D , and the two operators we will use the most in this thesis are xD and Dx . Higher powers of these operators are, in general, not commutative. To demonstrate how to evaluate these operators, we expand the case $n = 2$:

$$(Dx)^2 f(x) = (Dx)(Dx)f(x) = (Dx)[f(x) + xDf(x)] = f(x) + 3xDf(x) + x^2D^2f(x) \quad (3.6)$$

$$D^2x^2f(x) = D^2[x^2f(x)] = D[2xf(x) + x^2Df(x)] = 2f(x) + 4xDf(x) + x^2D^2f(x) \quad (3.7)$$

and the difference of $f(x) + xDf(x)$ is clear.

In [Boyadzhiev, 2009], $(Dx)^n$ is actually referred to as a *Mellin derivative*, presumably due to its MT properties. This terminology will be used here as well, and the subject is reviewed in more detail in Section 3.3.1 and Appendix A.2.1.

3.1.3 The Mellin Kind Gram-Charlier Series With Arbitrary Kernel

Using the MT property of the Mellin derivative, the MK Gram-Charlier series with arbitrary kernel $\rho(x)$ can now be completed. Assuming that $(Dx)^n \rho(x)$ is continuous on $\mathbb{R}_{\geq 0}$ for all $n \in \mathbb{Z}_{\geq 0}$, the result in eq. (3.5) can be applied to eq. (3.4) to give

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \frac{(-Dx)^n}{n!} \right] \rho(x), \quad (3.8)$$

where $(-Dx)^n = (-1)^n (Dx)^n$. The term corresponding $n = 0$ always equals 1 (recall that $B_0 = 1$), so that term was separated from the sum. This also emphasizes that it is a series around the kernel $\rho(x)$, based on Mellin derivatives of that kernel and the differences between

²This use of the Bell polynomials can also be applied to the classical case, as is demonstrated in Section 4.1.

the log-cumulants of X and the kernel log-cumulants. Finally, the series is truncated to provide an approximation or estimate

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \frac{(-Dx)^n}{n!} \right] \rho(x) \approx f_X(x). \quad (3.9)$$

This arbitrary kernel approach is applied to the classical case in Chapter 4.

3.1.4 The Mellin Derivative and the Arbitrary Kernel

In practice it is necessary to rewrite $(-Dx)^n \rho(x)$, in the same way eq. (2.82) allowed $H_n(x)\alpha(x)$ to replace $(-D)^n \alpha(x)$ in the classical Gram-Charlier series. Inserting the arbitrary kernel and its MK CF into eq. (3.5) gives

$$(-Dx)^n \rho(x) \xleftrightarrow{\mathcal{M}} (s-1)^n \phi_\rho(s). \quad (3.10)$$

This can be used to indirectly define the functions $\mathcal{P}_n(x)$ as

$$\mathcal{P}_n(x)\rho(x) = (-Dx)^n \rho(x), \quad (3.11)$$

that is, they are designed to satisfy

$$\mathcal{P}_n(x)\rho(x) \xleftrightarrow{\mathcal{M}} (s-1)^n \phi_\rho(s). \quad (3.12)$$

As the notation indicates, $\mathcal{P}_n(\cdot)$ is an n th degree polynomial in (some function of) x , at least for all of the PDF kernels we explore in this thesis. Eq. (3.11) can easily be manipulated to isolate $\mathcal{P}_n(\cdot)$,³

$$\mathcal{P}_n(x) = \frac{1}{\rho(x)} (-Dx)^n \rho(x), \quad (3.13)$$

where the factors in the kernel that are independent of x cancel each other out.

We now derive a recursive definition of $\mathcal{P}_{n+1}(x)$, starting with

$$\mathcal{P}_{n+1}(x) = \frac{1}{\rho(x)} (-Dx)(-Dx)^n \rho(x) = \frac{1}{\rho(x)} (-Dx)\mathcal{P}_n(x)\rho(x). \quad (3.14)$$

In order to simplify this expression, we view this as the differentiation (w.r.t. x) of the product of $\mathcal{P}_n(x)$ and $x\rho(x)$, i.e.

$$\mathcal{P}_{n+1}(x) = \frac{1}{\rho(x)} [-\mathcal{P}_n(x)Dx\rho(x) - x\rho(x)D\mathcal{P}_n(x)], \quad (3.15)$$

where $-Dx\rho(x) = \mathcal{P}_1(x)\rho(x)$ is recognized to give

$$\mathcal{P}_{n+1}(x) = \mathcal{P}_n(x)\mathcal{P}_1(x) - xD\mathcal{P}_n(x). \quad (3.16)$$

This allows us to compute subsequent $\mathcal{P}_n(x)$ if we have determined $\mathcal{P}_1(x)$ using eq. (3.13).

³This mirrors the Rodrigues formulas in eqs. (2.83) and (2.113), with the Mellin derivative Dx replacing the differential operator D . We use the term Rodrigues formula for these expressions in the following.

3.2 The Mellin Kind Gamma Kernel Series

Now, the gamma kernel $\gamma(x; a, b)$ defined in eq. (2.57) will be used to give a first example of a MK series expansion. It was termed the *Mellin kind gamma kernel* (MKGK) series in the project paper which preceded this thesis. We insert $\gamma(x; a, b)$ for the arbitrary kernel $\rho(x)$ in eq. (3.9) to get

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \frac{(-Dx)^n}{n!} \right] \gamma(x; a, b), \quad (3.17)$$

and the last hurdle is to replace $(-Dx)^n \gamma(x; a, b)$ with a more practical expression.

3.2.1 The Laguerre Polynomials and the Mellin Transform

In the classical case in Section 2.4.4, the Laguerre polynomials were used. We now review a negative result in order to demonstrate that the same polynomials are not directly applicable in the MK framework.

In order to find the MT of $L_n^{(a-1)}(x)\gamma(x; a)$, we combine eqs. (2.31), (2.64), and (2.115) to get

$$L_n^{(a-1)}(x)\gamma(x; a) = \frac{1}{n!} D^n [x^n \gamma(x; a)] \xleftrightarrow{\mathcal{M}} \frac{(-1)^n}{n!} \prod_{k=1}^n (s-k) \phi_\gamma(s), \quad (3.18)$$

where $\phi_\gamma(s)$ is the MK CF of the gamma distribution, found in eq. (2.64).⁴ However, it is observed that the right-hand side of eq. (3.18) is not on the same form as the right-hand side of eq. (3.10), which indicates that we cannot use $L_n^{(a-1)}(x)$ here.

Also, it is easy to verify that the Laguerre polynomials satisfy eq. (3.16) for $n = 0$, but not for $n = 1$. That is, $L_n^{(a-1)}(x)$ is not the right replacement for $\mathcal{P}_n(x)$ for *any* choice of kernel.

3.2.2 The $M_n(x)$ Polynomials

Instead following the instructions outlined in Section 3.1.4, we introduce $M_n(x)$, indirectly defined in terms of the unit scale gamma kernel from eq. (2.58) via

$$M_n(x)\gamma(x; a) = (-Dx)^n \gamma(x; a). \quad (3.19)$$

$M_n(x)$ is an n th degree polynomial in x , a fact we prove later in this section. Like in eq. (3.13), we can easily find the Rodrigues formula for $M_n(x)$ as

$$M_n(x) = \frac{\Gamma(a)}{x^{a-1}e^{-x}} (-Dx)^n \left[\frac{x^{a-1}e^{-x}}{\Gamma(a)} \right] = x^{-a+1} e^x (-Dx)^n [x^{a-1}e^{-x}]. \quad (3.20)$$

We used the unscaled version $\gamma(x; a)$ for brevity, but it is straightforward to show that we get the scaled version of $\gamma(x; a, b)$ by substituting $x \rightarrow bx$ in eq. (3.20). This is done the same way as for the Rodrigues formula of the generalized Laguerre polynomials in Section 2.4.3, but is even easier since $(Dx)^n$ is scale invariant. That is, for $u = bx$ we have

$$(D_u u)^n = \left(\frac{d}{dbx} bx \right)^n = \left(\frac{1}{b} \frac{d}{dx} bx \right)^n = \left(\frac{d}{dx} x \right)^n = (D_x x)^n. \quad (3.21)$$

⁴The project paper was solely focused on the MKGK series and had a slightly different approach to its derivation. There, eq. (3.18) was the starting point and we subsequently found a linear combination of the Laguerre polynomials which had MT $(s-1)^n \phi_\gamma(s)$. This result is reproduced in this thesis in Appendix A.2.1.

Replacing x with bx in eq. (3.19) gives

$$M_n(bx) \frac{b^{a-1} x^{a-1} e^{-bx}}{\Gamma(a)} = (-D_x x)^n \left[\frac{b^{a-1} x^{a-1} e^{-bx}}{\Gamma(a)} \right], \quad (3.22)$$

and since b is unaffected by $(D_x x)^n$, it is possible to multiply both sides with b and move the scaling factor inside the square brackets on the right-hand side to get

$$M_n(bx) \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)} = (-D_x x)^n \left[\frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)} \right], \quad (3.23)$$

$$M_n(bx) \gamma(x; a, b) = (-D_x x)^n \gamma(x; a, b), \quad (3.24)$$

showing the scalability of eq. (3.19).

In order to find the explicit expressions for the $M_n(\cdot)$ polynomials, both the result in eq. (3.16) and the Rodrigues formula eq. (3.20) are valid approaches, leading to

$$M_0(x) = 1, \quad (3.25)$$

$$M_1(x) = x - a, \quad (3.26)$$

$$M_2(x) = x^2 - (2a + 1)x + a^2, \quad (3.27)$$

$$M_3(x) = x^3 - 3(a + 1)x^2 + (3a^2 + 3a + 1)x - a^3, \quad (3.28)$$

$$M_4(x) = x^4 - (4a + 6)x^3 + (6a^2 + 12a + 7)x^2 - (4a^3 + 6a^2 + 4a + 1)x + a^4, \quad (3.29)$$

where the scaling was again fixed to $b = 1$ for brevity.

Lemma 1 $M_n(x)$ is an n th degree polynomial in x .

Proof We can insert $M_n(x)$ into the recursive definition in eq. (3.16) to get

$$M_{n+1}(x) = M_n(x)M_1(x) - xDM_n(x), \quad (3.30)$$

and from eq. (3.26) we know that $M_1(x)$ is a 1st degree polynomial in x . Assuming that $M_n(x)$ is an n th degree polynomial in x , $M_n(x)M_1(x)$ is trivially an $(n + 1)$ th degree polynomial. Also, $DM_n(x)$ must be an $(n - 1)$ th degree polynomial, i.e. $xDM_n(x)$ is of degree n . Thus, if $M_n(x)$ is an n th degree polynomial, then $M_{n+1}(x)$ is an $(n + 1)$ th degree polynomial, but $M_1(x)$ is a 1st degree polynomial, so we have now proved by induction that $M_n(x)$ is an n th degree polynomial in x . This result is general in the sense that it applies to any kernel in which $\mathcal{P}_1(x)$ in eq. (3.16) is a 1st degree polynomial in x .

Like the Hermite polynomials, the leading coefficient of $M_n(x)$ is $1 \forall n$. This property is proven in Appendix A.2, where the connection with the Laguerre polynomials is also examined.

Finally, observe that $M_n(x)\gamma(x; a)$ and therefore also $(Dx)^n\gamma(x; a)$ is continuous for any $n \in \mathbb{Z}_{\geq 0}$.⁵ This ensures that the series expansion will also be continuous.

The discovery and formulation of $\mathcal{P}_n(x)$ in general and the $M_n(x)$ polynomials specifically is a contribution of this thesis.

3.2.3 Defining the Mellin Kind Gamma Kernel Series

Now, eq. (3.24) inserted into eq. (3.17) gives the MKGK series

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) M_n(bx) \right] \gamma(x; a, b). \quad (3.31)$$

⁵In the classical case, $D^n\gamma(x; a)$ is discontinuous at $x = 0$ for some values of n , which is why the "kernel" in Section 2.4.4 is $x^n\gamma(x; a)$. We revisit this issue in Section 4.2.3.

3.2.4 Choosing the Kernel Parameters

In the classical case in Section 2.4.2, the parameters of the Gaussian kernel $\alpha(x)$ were chosen to match $f_X(x)$ by setting the kernel mean and variance equal to the empirical (observed) values. Thus, $c_{X,n} - c_{\alpha,n} = 0$ for $n = 1, 2$, thereby removing some terms from eq. (2.111).

For the MKGK series, it is proposed to choose a and b s.t. $\Delta\kappa_n = 0$ for $n = 1, 2$. The resulting estimates are not optimal, in the sense that they differ from the ML estimators of a and b .⁶ However, a significant advantage of this choice is the resulting simplification of the expressions, due to the nature of the Bell polynomials. In fact, if all arguments of the Bell polynomials are non-zero, then

$$\sum_{n=0}^6 B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \quad (3.32)$$

has 30 terms. If $\Delta\kappa_1 = \Delta\kappa_2 = 0$, this is reduced to only 6 terms since 24 of the terms have a dependency on at least one of the first two arguments. We will in the following refer to kernels who satisfy $\Delta\kappa_1 = \Delta\kappa_2 = 0$ as *tailored*, since they match the first two log-cumulants of the data.

Finding the parameters of the tailored $\gamma(x; a, b)$ can be done using the estimation algorithm known as the *method of log-cumulants* (MoLC). The MoLC will be briefly presented here – see [Krylov et al., 2013] for an in-depth review including application to the gamma and log-normal distributions and numerous other PDFs of non-negative RVs. The expressions for the first two log-cumulants of $\gamma(x; a, b)$ are found by reparametrizing eq. (2.62):

$$\kappa_{\gamma,1} = \psi^{(0)}(a) - \log b, \quad (3.33)$$

$$\kappa_{\gamma,2} = \psi^{(1)}(a). \quad (3.34)$$

We know that $\Delta\kappa_n = 0 \Rightarrow \kappa_{\gamma,n} = \kappa_{X,n}$, and that we must estimate the unknown $\kappa_{X,n}$. For $\kappa_{X,1}$, the standard estimator based on the data $\{x_1, \dots, x_n\}$, is the empirical or sample log-mean⁷

$$\langle \kappa_{X,1} \rangle = \langle \mu_{X,1} \rangle = \frac{1}{n} \sum_{i=1}^n \log x_i. \quad (3.35)$$

This is also the minimum variance unbiased estimator. For $\kappa_{X,2}$, the sample log-variance

$$\langle \kappa_{X,2} \rangle = \langle \sigma^2 \rangle = \frac{1}{n} \sum_{i=1}^n (\log x_i - \langle \mu_{X,1} \rangle)^2, \quad (3.36)$$

is the globally minimum variance estimate. However, it is slightly biased, so the scaled version $\frac{n}{n-1} \langle \sigma^2 \rangle$, which is the minimum variance unbiased estimator, is also sometimes used. Using the empirical log-cumulants $\langle \kappa_{X,1} \rangle$ and $\langle \kappa_{X,2} \rangle$ gives us

$$\langle \kappa_{X,1} \rangle = \psi^{(0)}(\hat{a}) - \log \hat{b}, \quad (3.37)$$

$$\langle \kappa_{X,2} \rangle = \psi^{(1)}(\hat{a}). \quad (3.38)$$

These two equations are solved for the estimates \hat{a} and \hat{b} . In practice, \hat{a} is computed numerically, since the polygamma function is not invertible. When we have a value for \hat{a} , the final step is to directly compute $\hat{b} = \exp\{\psi^{(0)}(\hat{a}) - \langle \kappa_{X,1} \rangle\}$.

⁶The ML estimators are not trivial for the gamma distribution, as discussed in [Choi and Wette, 1969], and for other distributions they may be very difficult to ascertain. [Nicolas and Anfinson, 2012] discusses the strengths and weaknesses of different estimators for the gamma distribution parameters.

⁷From eq. (2.42) it is easy to see that $\kappa_1 = \mu_1$ and $\kappa_2 = \sigma^2$.

The MKGK series with tailored kernel is

$$f_{X,N}(x) = \left[1 + \sum_{n=3}^N \frac{1}{n!} B_n(0, 0, \Delta\kappa_3, \dots, \Delta\kappa_n) M_n(x) \right] \gamma(x; a, b), \quad (3.39)$$

where $\Delta\kappa_n = \langle \kappa_{X,n} \rangle - \kappa_{\gamma,n}$.

3.2.5 The MKGK Series With Realistic Numbers of Terms

From a practical point of view, we must ask how many terms in the MKGK series should usually be included. More precisely: What is the highest order log-cumulants difference that should be accounted for, or equivalently, what is a "good" value of N in eq. (3.39)? A few precautions are necessary: The series can diverge, like [Blinnikov and Moessner, 1998] demonstrated for the classical methods. Moreover, the empirical log-cumulants are usually estimates of the true log-cumulants and therefore associated with bias and variance, depending on the estimator and the sample size.⁸ [Kendall et al., 1994] notes that for the classical cumulants, sampling fluctuations renders estimates of c_n unreliable for $n > 4$. In [Krylov et al., 2013], the authors show how the log-cumulants are less sensitive to outliers in small datasets, due to their logarithmic nature. They are however not concerned with higher order log-cumulants like those encountered in the MKGK series, so the disadvantage associated with high values of N is still an open question, and one we revisit in Chapter 5. For now we will merely demonstrate how certain properties of the Bell polynomials affect the MKGK series as N increases.

$N = 0, 1, 2$: The correction arising from $\Delta\kappa_1$ and $\Delta\kappa_2$ is embedded in the tailored gamma kernel, giving

$$f_{X,0}(x) = f_{X,1}(x) = f_{X,2}(x) = \gamma(x; a, b). \quad (3.40)$$

This happens since the gamma kernel has two parameters, and it is analogous to adjusting the mean and variance of a Gaussian kernel to fit the data.

$N = 3$:

$$f_{X,3}(x) = \left[1 + \frac{\Delta\kappa_3}{6} M_3(bx) \right] \gamma(x; a, b) \quad (3.41)$$

This is analogous to correcting for the skewness of the data in the classical case, as $\kappa_{X,3}$ represents *logarithmic skewness*.

$N = 4$:

$$f_{X,4}(x) = \left[1 + \frac{\Delta\kappa_3}{6} M_3(bx) + \frac{\Delta\kappa_4}{24} M_4(bx) \right] \gamma(x; a, b) \quad (3.42)$$

This is analogous to additionally correcting for the kurtosis of the data in the classical case, since $\kappa_{X,4}$ is the *logarithmic kurtosis*.

$N = 6$: Omitting $N = 5$ for the sake of brevity, the sixth term is interesting because it introduces the first non-linear term:

$$f_{X,6}(x) = \left[1 + \frac{\Delta\kappa_3}{6} M_3(bx) + \frac{\Delta\kappa_4}{24} M_4(bx) + \frac{\Delta\kappa_5}{120} M_5(bx) + \frac{\Delta\kappa_6 + 10\Delta\kappa_3^2}{720} M_6(bx) \right] \gamma(x; a, b).$$

From eqs. (2.147) through (2.153) we see that $10\Delta\kappa_3^2$ is the first non-linear term in the sequence of Bell polynomials which does not depend on either $\Delta\kappa_1$ or $\Delta\kappa_2$.

⁸See [Kendall et al., 1994] for a review of different estimators, like the standard sample mean or the unbiased k statistic.

$N = 7$: The final term which we present explicitly, $N = 7$ introduces the first cross-term, $\Delta\kappa_3\Delta\kappa_4$:

$$f_{X,7}(x) = \left[1 + \frac{\Delta\kappa_3}{6}M_3(bx) + \frac{\Delta\kappa_4}{24}M_4(bx) + \frac{\Delta\kappa_5}{120}M_5(bx) + \frac{\Delta\kappa_6 + 10\Delta\kappa_3^2}{720}M_6(bx) + \frac{\Delta\kappa_7 + 35\Delta\kappa_3\Delta\kappa_4}{5040}M_7(bx) \right] \gamma(x; a, b). \quad (3.43)$$

From the series presented here, the analogy with the classical case in Section 2.4.4 is very clear, as is the eventual introduction of non-linear and cross terms in the log-cumulant differences. For a tailored kernel, the Bell polynomials of order $n \leq 5$ are actually reduced to linear terms in $\Delta\kappa_n$, resulting in this relatively simple representation of the MKGK series.

3.3 The Mellin Kind Log-Normal Kernel Series

Inserting the log-normal kernel PDF $\Lambda(x; \mu, \sigma)$ from eq. (2.72) into eq. (3.9) gives

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \frac{(-Dx)^n}{n!} \right] \Lambda(x; \mu, \sigma). \quad (3.44)$$

We know that $\Lambda(x; \mu, \sigma)$ has support $\mathbb{R}_{>0}$ and in the next section it will become clear that $(-Dx)^n\Lambda(x; \mu, \sigma)$ is continuous for all $n \in \mathbb{Z}_{\geq 0}$. That is, $\Lambda(\cdot)$ is a valid choice of kernel, and as we will see in the following, it is a highly interesting choice as well.

3.3.1 The Logarithmic Hermite Polynomials

It is now necessary to evaluate $(-Dx)^n\Lambda(x; \mu, \sigma)$ in eq. (3.44). However, we can in this case find an expression in terms of *known* polynomials and avoid the need to use the recursive definition in eq. (3.11). To keep the derivation simple, we start with the standardized $\Lambda(x)$, and extend the result to arbitrary log-mean and log-variance afterwards.

See first that the log-normal kernel in eq. (2.74) is related to the Gaussian kernel in eq. (2.23) by

$$\Lambda(x) = \frac{1}{x}\alpha(\log x). \quad (3.45)$$

Applying $(-Dx)^n$ to both sides gives us

$$\begin{aligned} (-Dx)^n\Lambda(x) &= (-Dx)^n \frac{1}{x}\alpha(\log x) = (-1)^n (Dx)^n \frac{1}{x}\alpha(\log x) = (-1)^n (Dx)^{n-1} D\alpha(\log x) \\ &= (-1)^n \frac{1}{x} (Dx)^{n-1} D\alpha(\log x) = \frac{1}{x} (-xD)^{n-1} \alpha(\log x), \end{aligned} \quad (3.46)$$

where the final step was to group the terms differently. That is, we moved the parenthesis in $x(Dx) \cdots (Dx)D$ to write $(xD) \cdots (xD)$ instead. To evaluate the effects of applying $(xD)^n$ to $\alpha(\log x)$, note that $d \log x / dx = x^{-1}$, i.e. $dx / d \log x = x$, giving

$$\begin{aligned} (-xD)^n \alpha(\log x) &= (-xD)^{n-1} (-xD) \alpha(\log x) = (-xD)^{n-1} (-x) \frac{d\alpha(\log x)}{dx} \\ &= (-xD)^{n-1} \left(-\frac{dx}{d \log x} \right) \frac{d\alpha(\log x)}{dx} = (-xD)^{n-1} \left(-\frac{d}{d \log x} \right) \alpha(\log x), \end{aligned} \quad (3.47)$$

where the chain rule for calculating the derivate of the composition of two functions was used. We can repeat this n times and use the result in eq. (3.46) to reveal the key relation

$$(-Dx)^n \Lambda(x) = \frac{1}{x} \left(-\frac{d}{d \log x} \right)^n \alpha(\log x). \quad (3.48)$$

Substituting x with $\log x$ in the definition of the Hermite polynomials in eq. (2.82) gives

$$\left(-\frac{d}{d \log x} \right)^n \alpha(\log x) = H_n(\log x) \alpha(\log x), \quad (3.49)$$

and from the relationship of the log-normal and Gaussian PDFs in eq. (3.45) we clearly see that

$$(-Dx)^n \Lambda(x) = \frac{1}{x} H_n(\log x) \alpha(\log x) = H_n(\log x) \Lambda(x). \quad (3.50)$$

This can be used in a standardized version of eq. (3.44).

Finally, note that two of the results in this section are valid for any function $f(x)$, namely that

$$(Dx)^n \frac{1}{x} f(x) = \frac{1}{x} (xD)^n f(x), \quad (3.51)$$

and

$$(xD)^n f(x) = \left(\frac{d}{d \log x} \right)^n f(x). \quad (3.52)$$

While it is certainly possible that other authors have derived these results, it should be pointed out that the present derivation was done independently. One of the most significant contributions of this thesis is to illuminate the roles of the Mellin derivatives in MK statistics, since they have not been used in this context before. The above results are a part of this contribution.

3.3.2 Non-Standardized Log-Normal Data

Let $\log u = (\log x - \mu)/\sigma$ as in Section 2.3.11, and apply the result in eq. (2.78) to get

$$(-D_{xx})^n \Lambda(x; \mu, \sigma) = (-D_{xx})^n \frac{u}{x\sigma} \Lambda(u). \quad (3.53)$$

Since $\frac{d \log u}{d \log x} = 1/\sigma$, eq. (3.52) implies that

$$(-xD_x)^n = (-uD_u)^n \frac{1}{\sigma^n}, \quad (3.54)$$

and the result in eq. (3.51) gives

$$(-D_{xx})^n \Lambda(x; \mu, \sigma) = \frac{1}{x} (-xD_x)^n \frac{u}{\sigma} \Lambda(u) \quad (3.55)$$

$$(-D_{xx})^n \Lambda(x; \mu, \sigma) = \frac{1}{x} (-uD_u)^n \frac{1}{\sigma^n} \frac{u}{\sigma} \Lambda(u) \quad (3.56)$$

$$(-D_{xx})^n \Lambda(x; \mu, \sigma) = \frac{u}{x\sigma} \frac{1}{\sigma^n} (-D_u)^n \Lambda(u). \quad (3.57)$$

We can now use eq. (3.50) to insert $H_n(\log u)$ and see that

$$(-D_{xx})^n \Lambda(x; \mu, \sigma) = \frac{u}{x\sigma} \frac{1}{\sigma^n} H_n(\log u) \Lambda(u), \quad (3.58)$$

and the derivation is completed by re-inserting for $\log u$ in the Hermite polynomials and recognizing the non-standardized $\Lambda(x; \mu, \sigma)$ from eq. (2.78) to yield

$$(-D_{xx})^n \Lambda(x; \mu, \sigma) = \frac{1}{\sigma^n} H_n \left(\frac{\log x - \mu}{\sigma} \right) \frac{u}{x\sigma} \Lambda(u) = \frac{1}{\sigma^n} H_n \left(\frac{\log x - \mu}{\sigma} \right) \Lambda(x; \mu, \sigma). \quad (3.59)$$

In Appendix A.3, we derive this generalized result directly, i.e. without examining the standardized case first.

3.3.3 Defining the Mellin Kind Log-Normal Kernel Series

Using the result from eq. (3.59) to replace $(-Dx)^n$ with the logarithmic Hermite polynomials in eq. (3.44) gives the *Mellin kind log-normal kernel* (MKLK) series

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N \frac{1}{n! \sigma^n} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) H_n\left(\frac{\log x - \mu}{\sigma}\right) \right] \Lambda(x; \mu, \sigma). \quad (3.60)$$

Upon inspection of the coefficients in eq. (2.111), it is clear that this is perfectly analogous to the classical Gram-Charlier series with $\Lambda(x)$ replacing $\alpha(x)$ and $H_n(\log x)$ replacing $H_n(x)$. In Chapter 4, the approach used here is applied to the classical case, shedding further light on the connection between classical and MK statistics.

3.3.4 The Obvious Choices of the Kernel Parameters

As with the MKGK series, we can use MoLC estimates of the kernel parameters to achieve $\Delta\kappa_1 = \Delta\kappa_2 = 0$. Eq. (2.80) reveals that this amounts to equating the log-mean and log-variance in $\Lambda(x; \mu, \sigma)$ to the first and second order empirical log-cumulants, i.e.

$$\hat{\mu} = \langle \kappa_{X,1} \rangle, \quad (3.61)$$

$$\hat{\sigma}^2 = \langle \kappa_{X,2} \rangle. \quad (3.62)$$

This is conceptually the same as computing the tailored shape and scale parameters of the MKGK series using eqs. (3.37) and (3.38), but we see that for the log-normal kernel, this exercise is trivial.⁹ This matches the classical case, where the mean and variance of the Gaussian kernel are set equal to the first and second empirical cumulants.

Eq. (2.80) also tells us $\Delta\kappa_n = \kappa_{X,n} \forall n \geq 3$, since $\kappa_{\Lambda,n} = 0$ for $n \geq 3$. In other words, the first two log-cumulant differences are both zero, and all higher order log-cumulant differences are reduced to only the log-cumulants of X . This simplifies the MKLK series to

$$f_{X,N}(x) = \left[1 + \sum_{n=3}^N \frac{1}{n! \sigma^n} B_n(0, 0, \kappa_{X,3}, \dots, \kappa_{X,n}) H_n\left(\frac{\log x - \mu}{\sigma}\right) \right] \Lambda(x; \mu, \sigma), \quad (3.63)$$

where we again have to substitute $\kappa_{X,n}$ with $\langle \kappa_{X,n} \rangle$, unless we know the true distribution of X .

3.3.5 The MKLK Series With Realistic Numbers of Terms

After the thorough review concerning the terms in the MKGK series in Section 3.2.5, this section will be brief and merely state the MKLK series with $N = 7$:

$$f_{X,7}(x) = \left[1 + \frac{\kappa_3}{6\sigma^3} H_3\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_4}{24\sigma^4} H_4\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_5}{120\sigma^5} H_5\left(\frac{\log x - \mu}{\sigma}\right) \right. \\ \left. + \frac{\kappa_6 + 10\kappa_3^2}{720\sigma^6} H_6\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_7 + 35\kappa_3\kappa_4}{5040\sigma^7} H_7\left(\frac{\log x - \mu}{\sigma}\right) \right] \Lambda(x; \mu, \sigma). \quad (3.64)$$

Since the log-cumulants of $\Lambda(x; \mu, \sigma)$ of order $n \geq 3$ are all zero, there is no ambiguity in writing κ_n instead of $\kappa_{X,n}$. Note the striking resemblance with eq. (2.111).

⁹As we discuss later in this thesis, there is also much less ambiguity in the choice of estimates due to the fact the first two empirical log-cumulants are also the ML estimates for the log-mean and log-variance.

3.4 The Mellin Kind Edgeworth Series

When reviewing the classical series expansion methods in Section 2.4, both the Gram-Charlier series (Gaussian kernel) and the Edgeworth series could be derived from the double infinite sum in eq. (2.141). In this section, we apply a similar approach using MK statistics. This leads us to the same series expansion as the one first derived in [Pastor et al., 2014] and [Pastor et al., 2016], but our derivation and expression of the series differs from those works.

3.4.1 The Log-Cumulant Differences

Section 2.4.5 lists the five properties of the classical cumulants used to justify rewriting the cumulant differences as $\lambda_n/r^{n/2-1}$. In this section, we conduct the same investigation w.r.t. the log-cumulants, and use the result to derive the *Mellin kind Edgeworth* (MKE) series.

First, the underlying assumption must be stated. Saying that X is the standardized sum of IID RVs Z_1, \dots, Z_r , like in eq. (2.135), does not make directly sense in a MK statistics situation. Even if we impose the restriction that Z_i must be non-negative (to ensure a non-negative X), summation is not a concept which is suited for the logarithm operation, in the sense that $\log(a+b)$ is not associated with any useful mathematical relationship. Physically, the additive model is better suited to the classical methods in any case. Instead, assume that the *logarithm* of X is the standardized sum of Z_1, \dots, Z_r , that is

$$\log X = \frac{1}{\sqrt{r}} \sum_{i=1}^r \frac{Z_i - m}{\zeta}, \quad (3.65)$$

where the only assumptions on Z_1, \dots, Z_r are that they are IID with classical mean m , classical variance ζ^2 and higher order cumulants $c_n = \zeta^n \lambda_n$, just like in Section 2.4.5. Thus, the only real restriction on the constituent RVs is that they are IID, they are even allowed to take negative values.

The easiest way of finding the log-cumulants of X is to acknowledge that since the moments of $\log X$ and the log-moments of X are both defined as $E\{(\log X)^n\}$, they must obviously be equal. We also know from eqs. (2.10) and (2.42) that the relationship between the log-moments and log-cumulants matches that of their classical counterparts, so the log-cumulants of X must equal the cumulants of $\log X$, which in turn is the same standardized sum as in Section 2.4.5. Still, deriving this result from the relevant log-cumulant properties can be instructive, and is done in the following.

Starting with a list which is equivalent to the one in Section 2.4.5, we summarize the properties needed to proceed:

1. The log-cumulants of order $n \geq 2$ are scale-invariant (Section 2.3.7).
2. Applying the power transformation $X^{1/\zeta}$ to the RV X results in a scaling of the log-cumulants $\kappa_{X,n}$ with ζ^{-n} (Section 2.3.7).
3. The log-cumulants of a product of independent RVs are the sums of the log-cumulants of each RV (Section 2.3.8).
4. The log-cumulants of $\Lambda(x; \mu, \sigma)$ of order $n \geq 3$ are all zero. (Section 2.3.11).
5. Since $\log X$ by definition has zero mean and unit variance, X has zero log-mean and unit log-variance, i.e. its first and second order log-cumulants are the same as those of a RV which follows the standardized $\Lambda(x)$. For non-standardized $\log X$, this property is mirrored by $\Lambda(x; \mu, \sigma)$ when μ and σ are tailored to the data (in practice set to the first and second empirical log-cumulant).

Combining properties 4 and 5 trivially gives

$$\Delta\kappa_n = \kappa_{X,n} - \kappa_{\Lambda,n} = \begin{cases} 0 & n = 1, 2, \\ \kappa_{X,n} & n \geq 3. \end{cases} \quad (3.66)$$

To examine $\kappa_{X,n}$, we start with eq. (3.65), multiply both sides with $\varsigma\sqrt{r}$, and then apply the exponential function to isolate X as

$$X^{\varsigma\sqrt{r}} = \prod_{i=1}^r e^{Z_i - m} = e^{-rm} \prod_{i=1}^r e^{Z_i} \Rightarrow X = e^{-\frac{\sqrt{r}m}{\varsigma}} \left(\prod_{i=1}^r e^{Z_i} \right)^{\frac{1}{\varsigma\sqrt{r}}}, \quad (3.67)$$

and we can recognize the multiplicative situation from Section 2.3.8. Before applying the remaining properties, the log-cumulants of e^{Z_i} are required. The log-moments are simply

$$\mu_{\{e^{Z_i}, n\}} = E \{ (\log e^{Z_i})^n \} = E \{ Z_i^n \} = m_{\{Z_i, n\}}. \quad (3.68)$$

That is, by the definitions of the moments and log-moments in eqs. (2.3) and (2.34), the log-moments of e^{Z_i} equal the moments of Z_i . Since the combinatoric relationship between the two sets of logarithmic descriptors is the same as for the classical descriptors, the log-cumulants of e^{Z_i} must also equal the cumulants of Z_i . In other words, we can see that

$$\kappa_{\{e^{Z_i}, n\}} = c_{Z_i, n} = \varsigma^n \lambda_n \quad (3.69)$$

by combining eq. (3.68) with eqs. (2.10) and (2.42). Thus, property 3 implies that the log-cumulants of $\prod_{i=1}^r e^{Z_i}$ are $r\kappa_{\{e^{Z_i}, n\}} = r\varsigma^n \lambda_n$, and since $\kappa_{X,1}$ is not of interest, property 1 states that the scaling factor $e^{-\frac{\sqrt{r}m}{\varsigma}}$ can be disregarded. Finally, property 2 is applied to give for $n \geq 2$:

$$\kappa_{X,n} = \frac{1}{(\varsigma\sqrt{r})^n} r\varsigma^n \lambda_n = \frac{\lambda_n}{r^{\frac{n}{2}-1}}, \quad (3.70)$$

and the final result is practically identical to the classical result in eq. (2.139), namely

$$\Delta\kappa_n = \kappa_{X,n} - \kappa_{\Lambda,n} = \begin{cases} 0 & n = 1, 2, \\ \frac{\lambda_n}{r^{\frac{n}{2}-1}} & n \geq 3. \end{cases} \quad (3.71)$$

3.4.2 Deriving the Mellin Kind Edgeworth Series

With the given assumptions on the underlying nature of the data, it is now possible to complete the derivation of the MKE series. Starting from eq. (3.3), the arbitrary kernel is replaced with the log-normal kernel,

$$\phi_X(s) = \exp \left\{ \sum_{k=1}^{\infty} \Delta\kappa_k \frac{(s-1)^k}{k!} \right\} \phi_{\Lambda}(s), \quad (3.72)$$

and the result in eq. (3.71) gives

$$\phi_X(s) = \exp \left\{ \sum_{k=3}^{\infty} \frac{\lambda_k}{r^{\frac{k}{2}-1}} \frac{(s-1)^k}{k!} \right\} \phi_{\Lambda}(s). \quad (3.73)$$

As in the classical case, the Edgeworth series distinguishes itself from the Gram-Charlier by arranging the terms by their power of $r^{-1/2}$. Since $1/r^{\frac{k}{2}-1} = (r^{-1/2})^{k-2}$, an index shift $k \rightarrow k-2$ is necessary, i.e.

$$\phi_X(s) = \exp \left\{ \sum_{k=1}^{\infty} \frac{\lambda_{k+2}}{(r^{1/2})^k} \frac{(s-1)^{k+2}}{(k+2)!} \right\} \phi_{\Lambda}(s). \quad (3.74)$$

By rewriting $(k+2)! = k!(k+1)(k+2)$, the power series representation in terms of $r^{-1/2}$ is now recognized as

$$\phi_X(s) = \exp \left\{ \sum_{k=1}^{\infty} \frac{\lambda_{k+2}(s-1)^{k+2}}{(k+1)(k+2)} \frac{(r^{-1/2})^k}{k!} \right\} \phi_{\Lambda}(s) = \exp \left\{ \sum_{k=1}^{\infty} \xi_k(s-1) \frac{(r^{-1/2})^k}{k!} \right\} \phi_{\Lambda}(s), \quad (3.75)$$

where we define the coefficients ξ_k for the sake of brevity as

$$\xi_k(s-1) \equiv \frac{\lambda_{k+2}(s-1)^{k+2}}{(k+1)(k+2)}, \quad (3.76)$$

and we note that they are independent of r .

Up to this point, the approach has not been significantly different from that of [Pastor et al., 2014] and [Pastor et al., 2016], except w.r.t. notation and the fact that the format of those references required a much more condensed derivation. Now, however, one of the key contributions of this thesis is to use the Bell polynomials via the property in eq. (2.154) to write the MKE series in a new and more compact way:

$$\phi_X(s) = \left[\sum_{n=0}^{\infty} B_n(\xi_1(s-1), \xi_2(s-1), \dots, \xi_n(s-1)) \frac{(r^{-1/2})^n}{n!} \right] \phi_{\Lambda}(s). \quad (3.77)$$

Since all coefficients $\xi_n(s-1)$ are scaled powers of $(s-1)$, all $B_n(\cdot)$ are polynomials of $(s-1)$, and the linear property of the MT ensures that an inverse transform using eq. (3.5) is permitted like in Section 3.3. That is,

$$f_X(x) = \left[\sum_{n=0}^{\infty} B_n(\xi_1(-Dx), \xi_2(-Dx), \dots, \xi_n(-Dx)) \frac{(r^{-1/2})^n}{n!} \right] \Lambda(x; \mu, \sigma), \quad (3.78)$$

where

$$\xi_n(-Dx) = \frac{\lambda_{n+2}(-Dx)^{n+2}}{(n+1)(n+2)}. \quad (3.79)$$

Explicitly, the truncated MKE series is

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} B_n \left(\frac{\lambda_3(-Dx)^3}{6}, \dots, \frac{\lambda_{n+2}(-Dx)^{n+2}}{(n+1)(n+2)} \right) \frac{(r^{-1/2})^n}{n!} \right] \Lambda(x; \mu, \sigma), \quad (3.80)$$

where the use of the non-standardized $\Lambda(x; \mu, \sigma)$ again requires scaling with σ^{-n} when inserting the Hermite polynomials. Also, the powers of r can be re-matched to their corresponding powers of λ , as they were just a temporary tool. Then, finally

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n \left(\frac{\kappa_3(-Dx)^3}{6}, \dots, \frac{\kappa_{n+2}(-Dx)^{n+2}}{(n+1)(n+2)} \right) \right] \Lambda(x; \mu, \sigma), \quad (3.81)$$

which is the most compact form of the MKE series. Eq. (3.59) gives the relationship needed to replace $(-Dx)^{n+2}$ with the Hermite polynomials in order to implement the series. The first few

terms are

$$\begin{aligned}
f_X(x) &= \Lambda(x; \mu, \sigma) + \frac{1}{r^{1/2}} \left[\frac{\lambda_3}{6\sigma^3} H_3 \left(\frac{\log x - \mu}{\sigma} \right) \right] \Lambda(x; \mu, \sigma) \\
&+ \frac{1}{r} \left[\frac{\lambda_3^2}{72\sigma^6} H_6 \left(\frac{\log x - \mu}{\sigma} \right) + \frac{\lambda_4}{24\sigma^4} H_4 \left(\frac{\log x - \mu}{\sigma} \right) \right] \Lambda(x; \mu, \sigma) \\
&+ \frac{1}{r^{3/2}} \left[\frac{\lambda_3^3}{1296\sigma^9} H_9 \left(\frac{\log x - \mu}{\sigma} \right) + \frac{\lambda_3\lambda_4}{144\sigma^7} H_7 \left(\frac{\log x - \mu}{\sigma} \right) + \frac{\lambda_5}{120\sigma^5} H_5 \left(\frac{\log x - \mu}{\sigma} \right) \right] \Lambda(x; \mu, \sigma) \\
&+ \frac{1}{r^2} \left[\frac{\lambda_3^4}{31104\sigma^{12}} H_{12} \left(\frac{\log x - \mu}{\sigma} \right) + \frac{\lambda_3^2\lambda_4}{1728\sigma^{10}} H_{10} \left(\frac{\log x - \mu}{\sigma} \right) \right. \\
&+ \left. \left(\frac{\lambda_3\lambda_5}{720} + \frac{\lambda_4^2}{1152} \right) \frac{1}{\sigma^8} H_8 \left(\frac{\log x - \mu}{\sigma} \right) + \frac{\lambda_6}{720\sigma^6} H_6 \left(\frac{\log x - \mu}{\sigma} \right) \right] \Lambda(x; \mu, \sigma) \\
&+ O\left(\frac{1}{r^{5/2}}\right).
\end{aligned} \tag{3.82}$$

Here, r was retained to illustrate how each successive term is associated with an incremented power of $r^{-1/2}$. Combining the factors λ_n with their corresponding power of r allows the (empirical) log-cumulants to be inserted when fitting the MKE series to data. Note the similarity with the classical result in eq. (2.142). Also, we see that to correct for only the log-cumulants of order $\leq N$, we need to discard terms of order $r^{-\frac{N-1}{2}}$. This will later allow us to compare the MKE series with the other series expansions.

3.4.3 The Work of Pastor et al.

Eq. (3.81) is exactly the same result as in [Pastor et al., 2014], but the derivation is different in several key areas. Where this thesis introduced¹⁰ Bell polynomials as an alternative way to represent the combinatoric aspect of the series, other authors have used eq. (2.155) directly. Examples of this are [Blinnikov and Moessner, 1998], who used it on the classical Edgeworth series, and [Pastor et al., 2014], [Pastor, 2016] who used it on what is here called the MKE series. Pastor presented the series again in [Pastor et al., 2016] in a compact but easily readable format under the name *EEL*. That series is reproduced below, but we have slightly modified the notation to match the rest of this thesis:

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \sum_{\mathcal{K}_n} \frac{1}{\sigma^{n+2r}} H_{n+2r} \left(\frac{\log x - \mu}{\sigma} \right) \prod_{i=1}^n \frac{1}{j_i} \left(\frac{\kappa_{i+2}}{(i+2)!} \right)^{j_i} \right] \Lambda(x; \mu, \sigma) \tag{3.83}$$

$$\mathcal{K}_n = \left\{ (j_1, \dots, j_n; r) : j_i \in \mathbb{Z}_{\geq 0}, \sum_{i=1}^n i j_i = n, r = \sum_{i=1}^n j_i \right\}. \tag{3.84}$$

We see that \mathcal{K}_n is the same set as in Section 2.5.3. That is, the representation in eq. (3.81) avoids the use of the set \mathcal{K}_n , while eq. (3.83) avoids the use of Bell polynomials. The latter also permits the insertion of the Hermite polynomials. This cannot be done before evaluating the Bell polynomials in eq. (3.81), as $[(-Dx)^n]^2$ must be replaced with $H_{2n}(\cdot)$, not $[H_n(\cdot)]^2$.

Another and perhaps more significant difference between the two approaches is how $f_X(x)$ is retrieved from the MK CF. In [Pastor et al., 2014], [Pastor et al., 2016], and [Pastor, 2016], the

¹⁰Author's comment: While it is impossible to be absolutely certain that it has never been done, personally I have not found the Bell polynomials used this way before, even in the classical Gram-Charlier or Edgeworth series.

author(s) used a change of transform variables $s - 1 = j\omega$. For the two RVs X and Y , where $Y = \log X$, they used the fact that the MK CF of X equals the classical CF of Y , that is

$$\phi_X(s) = E\{X^{s-1}\} = E\{X^{j\omega}\} = E\{e^{j\omega Y}\} = \Phi_Y(\omega), \quad (3.85)$$

where the MK CF and CF are recognized from their definitions in eqs. (2.32) and (2.1). This is a general result, first proven in [Nicolas, 2002]. A key insight on behalf of Pastor et al. was that this could also be applied to the MK CF of the standardized kernels $\Lambda(x)$ and $\alpha(x)$. From there, they used the inverse FT was used to recover $f_Y(y)$, and going from $f_Y(y)$ to $f_X(x)$ was just a matter of converting from $\alpha(\log x)$ to $\Lambda(x)$.

This is in contrast to our approach, where the Mellin derivatives allowed us to use the inverse MT to derive the (same) MKE series. That is to say, two of the contributions of this thesis is to highlight the use of the Mellin derivatives, and to provide a derivation of the MKE series which is wholly within the framework of MK statistics. Hopefully, this can contribute to the popularity of Dx in the context of MK statistics, but also the MKE series, which our experiments in Chapter 5 promote as a formidable method.

Authors Comment At this point it is necessary to be completely forthright about the process behind this thesis and the preceding project paper, in order to clarify what was developed independently of Pastor et al.

Before assigning this subject to me, the supervisor of this thesis, Stian N. Anfinsen, provided copies of [Pastor et al., 2014], [Pastor et al., 2016], and [Pastor, 2016] among a great many publications (he was an examiner of the latter). The primary task was to develop a MK series expansion of the gamma kernel, based on an intuition that it was the analogy to the classical series expansions. The MKGK series was derived successfully and independently. We can claim this, since we frankly believed that the different approach we took necessarily would lead to a different result. The original derivation of the MKGK series, as presented in the project paper, was significantly different from the version presented here.

During the early stages of work on this thesis, we tried to expand the MKGK series to a full framework for MK series expansions, still under the conviction that our approach would lead to fundamentally different results than the work of Pastor et al. Shortly after deriving the MKLK series,¹¹ I realized that the result was identical to what Pastor et al. called the *Gram-Charlier series using log-cumulants*,¹² which they stated using $Y = \log X$, the inverse FT, and D_y^n as explained above. From there, the dots were connected, so to speak, i.e. I recognized the importance of the relationship in eq. (3.85) and subsequently understood how the two derivations was related.

To summarize, the MKLK series was developed independently of Pastor et al., but it is identical to the series they called the Gram-Charlier series using log-cumulants. The MKE series however, was decided *not* derived independently. Our use of the Bell-polynomials in both the classical and MK series expansions on the other hand, is original work, presented for the first time in this thesis. The MK Gram-Charlier series with arbitrary kernel, the MKGK series, and the series expansion in Section 3.5 have not been discussed before in any capacity, to the best of my knowledge.

See also Chapter 1 for more information about the contributions of this thesis.

¹¹The MKGK series in its present form took a few months to derive, with the inverse MT of $(s - 1)^n \phi_\gamma(s)$ using the little known Mellin derivative proving the most significant hurdle. With the knowledge of $(Dx)^n$ etc., the MKLK series was derived in a matter of hours, with the biggest challenge being the rigorous proof of eq. (3.59), which was completed in a few days.

¹²It appears that, [Pastor et al., 2014], [Pastor et al., 2016], and [Pastor, 2016] uses the log-normal kernel exclusively.

3.5 The Mellin Kind Beta Prime Kernel Series

Another candidate for replacing $\rho(x)$ in eq. (3.9) is the beta prime distribution $\beta'(x; a_1, a_2, b)$ from eq. (2.67). Compared to the gamma and log-normal kernels, the beta prime PDF has an additional parameter, serving as an example of a series expansion around a more complex kernel. We define the *Mellin kind beta prime kernel* (MKBK) series as

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) \frac{(-Dx)^n}{n!} \right] \beta'(x; a_1, a_2, b). \quad (3.86)$$

3.5.1 The $M'_n(\cdot)$ Polynomials

We will now examine the functions which can replace $(-Dx)^n$ in eq. (3.86). That is, we will define $M'_n(\cdot)$ (read "M-prime") as the functions arising from replacing $\rho(x)$ with $\beta'(x; a_1, a_2, b)$ in eq. (3.11). As we will show shortly, $M'_n(\cdot)$ is an n th degree polynomial in $\frac{bx}{1+bx}$, hence we refer to them as $M'_n\left(\frac{bx}{1+bx}\right)$.

Let $M'_n\left(\frac{bx}{1+bx}\right)$ be implicitly defined by

$$M'_n\left(\frac{bx}{1+bx}\right) \beta'(x; a_1, a_2, b) = (-Dx)^n \beta'(x; a_1, a_2, b). \quad (3.87)$$

The Rodrigues formula is again found by isolating the polynomials and letting the constant factors cancel each other out

$$M'_n\left(\frac{bx}{1+bx}\right) = \frac{(1+bx)^{a_1+a_2}}{(bx)^{a_1-1}} (-D_x x)^n \left[\frac{(bx)^{a_1-1}}{(1+bx)^{a_1+a_2}} \right], \quad (3.88)$$

and we know from eq. (3.21) that $D_x x$ is scale invariant, i.e. we can substitute $u = bx$ and $D_u u = D_x x$. The recursive formula in eq. (3.16) is applicable, and inserting $M'_n\left(\frac{u}{1+u}\right)$ gives

$$M'_{n+1}\left(\frac{u}{1+u}\right) = M'_n\left(\frac{u}{1+u}\right) M'_1\left(\frac{u}{1+u}\right) - u D_u M'_n\left(\frac{u}{1+u}\right), \quad (3.89)$$

and we note that $x D_x$ is easily found to be scale invariant by the same reasoning as in eq. (3.21). To use the result in eq. (3.89), we must first find $M'_1(\cdot)$ from eq. (3.88), that is,

$$M'_1\left(\frac{u}{1+u}\right) = \frac{(1+u)^{a_1+a_2}}{u^{a_1-1}} (-D_u u) \left[\frac{u^{a_1-1}}{(1+u)^{a_1+a_2}} \right] = \left[(a_1 + a_2) \frac{u}{1+u} - a_1 \right]. \quad (3.90)$$

This is in fact all that is required to find $M'_n(\cdot)$ for arbitrary n using eq. (3.89). For example, to find $M'_2(\cdot)$ we need

$$\begin{aligned} -u D_u M'_1\left(\frac{u}{1+u}\right) &= -(a_1 + a_2) \frac{u}{(1+u)^2} = -(a_1 + a_2) \frac{u + u^2 - u^2}{(1+u)^2} \\ &= (a_1 + a_2) \left[\left(\frac{u}{1+u}\right)^2 + \frac{u}{1+u} \right], \end{aligned} \quad (3.91)$$

which gives

$$\begin{aligned} M'_2\left(\frac{u}{1+u}\right) &= \left[M'_1\left(\frac{u}{1+u}\right) \right]^2 - u D_u M'_1\left(\frac{u}{1+u}\right) \\ &= (a_1 + a_2)^2 \left(\frac{u}{1+u}\right)^2 - 2(a_1 + a_2) a_1 \frac{u}{1+u} + a_1^2 (a_1 + a_2) \left[\left(\frac{u}{1+u}\right)^2 + \frac{u}{1+u} \right] \\ &= (a_1 + a_2)(a_1 + a_2 + 1) \left(\frac{u}{1+u}\right)^2 - (2a_1 + 1)(a_1 + a_2) \frac{u}{1+u} + a_1^2. \end{aligned} \quad (3.92)$$

The next two $M'_n(\cdot)$ polynomials are

$$\begin{aligned} M'_3\left(\frac{u}{1+u}\right) &= (a_1 + a_2)(a_1 + a_2 + 1)(a_1 + a_2 + 2)\left(\frac{u}{1+u}\right)^3 \\ &\quad - 3(a_1 + 1)(a_1 + a_2)(a_1 + a_2 + 1)\left(\frac{u}{1+u}\right)^2 \\ &\quad + a_1(a_1 + a_2)(3a_1 + 1)\frac{u}{1+u} - a_1^3, \end{aligned} \quad (3.93)$$

and

$$\begin{aligned} M'_4\left(\frac{u}{1+u}\right) &= (a_1 + a_2)(a_1 + a_2 + 1)(a_1 + a_2 + 2)(a_1 + a_2 + 3)\left(\frac{u}{1+u}\right)^4 \\ &\quad - (a_1 + a_2)(a_1 + a_2 + 1)[(a_1 + a_2 + 2)(3a_1 + 1) + 3(a_1 + 1)]\left(\frac{u}{1+u}\right)^3 \\ &\quad + (a_1 + a_2)[3(a_1 + 1)(a_1 + 2)(a_1 + a_2 + 1) + a_1(3a_1 + 1)(a_1 + a_2)]\left(\frac{u}{1+u}\right)^2 \\ &\quad - a_1(a_1 + a_2)[4a_1(a_1 + 1) + 1]\frac{u}{1+u} + a_1^4. \end{aligned} \quad (3.94)$$

Recall from Section 2.3.10 that in SAR imagery, we can assume that the shape parameters of multitemporal images are equal, with modern estimates so good that they are considered equal to the true values, i.e. $\hat{a}_1 = \hat{a}_2 = L$. The recursive definition in eq. (3.89) is still applicable, but the polynomials are greatly simplified to

$$M'_1\left(\frac{u}{1+u}\right) = \left[2L\frac{u}{1+u} - L\right], \quad (3.95)$$

$$M'_2\left(\frac{u}{1+u}\right) = 2L(2L+1)\left(\frac{u}{1+u}\right)^2 - 2L(2L+1)\frac{u}{1+u} + L^2, \quad (3.96)$$

$$\begin{aligned} M'_3\left(\frac{u}{1+u}\right) &= 2L(2L+1)(2L+2)\left(\frac{u}{1+u}\right)^3 - (12L^3 + 18L^2 + 6L)\left(\frac{u}{1+u}\right)^2 \\ &\quad + (6L^3 + 2L^2)\frac{u}{1+u} - L^3, \end{aligned} \quad (3.97)$$

$$\begin{aligned} M'_4\left(\frac{u}{1+u}\right) &= 2L(2L+1)(2L+2)(2L+3)\left(\frac{u}{1+u}\right)^4 - (24L^4 + 56L^3 + 42L^2 + 10L)\left(\frac{u}{1+u}\right)^3 \\ &\quad + (24L^4 + 46L^3 + 42L^2 + 12L)\left(\frac{u}{1+u}\right)^2 - (8L^4 + 8L^3 + 2L^2)\frac{u}{1+u} + L^4. \end{aligned} \quad (3.98)$$

Lemma 2 $M'_n\left(\frac{u}{1+u}\right)$ is an n th degree polynomial in $\frac{u}{1+u}$.

Proof This is not as straightforward as for the $M_n(x)$ polynomials in Section 3.2.2, but the proofs are quite similar. Assuming that $M'_n\left(\frac{u}{1+u}\right)$ is an n th degree polynomial in $\frac{u}{1+u}$, we start with the recursive definition of $M'_{n+1}(\cdot)$ in eq. (3.89) and immediately see that since $M'_1\left(\frac{u}{1+u}\right)$ is a first degree polynomial in $\frac{u}{1+u}$, then

$$M'_n\left(\frac{u}{1+u}\right)M'_1\left(\frac{u}{1+u}\right) \quad (3.99)$$

is of degree $n + 1$. For $uD_u M'_n\left(\frac{u}{1+u}\right)$, the chain rule of differentiation states that

$$D_u \left(\frac{u}{1+u} \right)^n = \frac{1}{u} \frac{n}{1+u} \left(\frac{u}{1+u} \right)^n, \quad (3.100)$$

and multiplying with u we can rewrite this as

$$uD_u \left(\frac{u}{1+u} \right)^n = \frac{n + nu - nu}{1+u} \left(\frac{u}{1+u} \right)^n = \left(\frac{u}{1+u} \right)^n \left[n - n \frac{u}{1+u} \right]. \quad (3.101)$$

This means that if $M'_n\left(\frac{u}{1+u}\right)$ is of degree n , then $uD_u M'_n\left(\frac{u}{1+u}\right)$ is of degree $n + 1$, and $M'_{n+1}\left(\frac{u}{1+u}\right)$ is also of degree $n + 1$. Since we have already evaluated $M'_1\left(\frac{u}{1+u}\right)$ in eq. (3.90), our proof by induction is complete.

3.5.2 Defining the Mellin Kind Beta Prime Kernel Series

As with the previously presented series, we insert the result from eq. (3.87) into eq. (3.86) to get a workable expression for the MKBK series:

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N \frac{1}{n!} B_n(\Delta\kappa_1, \Delta\kappa_2, \dots, \Delta\kappa_n) M'_n\left(\frac{bx}{1+bx}\right) \right] \beta'(x; a_1, a_2, b). \quad (3.102)$$

While also being another tool with which to approximate PDFs, the MKBK series demonstrates the flexibility of the MK series expansion framework – expanding a more complicated and three-parameter kernel is possible with only minor issues.

If we can assume $a_1 = a_2 = L$, $\beta'(x; a_1, a_2, b)$ in eq. (3.102) is replaced with $\beta'(x; L, b)$ from eq. (2.69), as discussed in Section 2.3.10. Recall from Section 3.5.1 that the $M'_n(\cdot)$ polynomials are found in the same way as for the three-parameter case, but are significantly simplified.

3.5.3 Choosing the Kernel Parameters

Compared to the MKGK, MKLK and MKE series, the kernel in the MKBK series has an additional parameter. This allows us to choose the parameters s.t. the tailored kernel satisfies $\Delta\kappa_1 = \Delta\kappa_2 = \Delta\kappa_3 = 0$. This is achieved by choosing the MoLC parameter estimates. That is, the first three log-cumulants of $\beta'(x; a_1, a_2, b)$ are found from eq. (2.68) via

$$\kappa_{\beta',1} = \psi^{(0)}(a_1) - \psi^{(0)}(a_2) - \log b, \quad (3.103)$$

$$\kappa_{\beta',2} = \psi^{(1)}(a_1) + \psi^{(1)}(a_2), \quad (3.104)$$

$$\kappa_{\beta',3} = \psi^{(2)}(a_1) - \psi^{(2)}(a_2). \quad (3.105)$$

By the same reasoning as in Section 3.2.4, we replace $\kappa_{\beta',n}$ with the empirical log-cumulants $\langle \kappa_{X,n} \rangle$. Then, we have a set of three equations which must be solved for the three unknowns \hat{a}_1 , \hat{a}_2 , and \hat{b} . In practice, the shape parameter estimates \hat{a}_1 and \hat{a}_2 have to be computed simultaneously with an iterative numerical procedure,¹³ with \hat{b} solved trivially based on \hat{a}_1 , \hat{a}_2 , and $\langle \kappa_{X,1} \rangle$.¹⁴

¹³In [Li et al., 2011] the authors faced the same problem with a separate three-parameter distribution. They used the asymptotic formula for the polygamma function to get approximate, but explicit expressions for the MoLC estimates, resulting in a much faster computation. There is nothing that suggest such an approach is impossible for the beta prime distribution, but it is beyond the scope of this thesis to attempt it.

¹⁴The built-in MATLAB functions for solving sets of equations were prohibitively slow, but a specialized program written in C for finding the MoLC shape estimates of the K distribution was provided by the supervisor for this thesis, Stian N. Anfinssen. He received it from Olivier Harant and Lionel Bombrun, who at the time worked at Grenoble Institute of Technology. The program was relatively easily modified to account for the slightly different expression for the log-cumulants of the beta prime distribution.

With the kernel tailored, the MKBK series is simplified to

$$f_{X,N}(x) = \left[1 + \sum_{n=4}^N \frac{1}{n!} B_n(0, 0, 0, \Delta\kappa_4, \Delta\kappa_5, \dots, \Delta\kappa_n) M'_n \left(\frac{bx}{1+bx} \right) \right] \beta'(x; a_1, a_2, b). \quad (3.106)$$

If $a_1 = a_2 = L$, the beta prime kernel has only two parameters, and eq. (2.70) gives the MoLC estimates

$$\langle \kappa_{X,1} \rangle = -\log \hat{b}, \quad (3.107)$$

$$\langle \kappa_{X,2} \rangle = \psi^{(1)}(\hat{L}), \quad (3.108)$$

but also that

$$\kappa_{\beta',3} = \kappa_{\beta',5} = \dots = 0. \quad (3.109)$$

This resembles the equations for the gamma kernel parameters in Section 3.2.4, but is slightly simpler. Finding L still involves numerical approaches since there is no exact way of inverting the polygamma function, but it does not involve solving for two shape parameters simultaneously as in the general case.¹⁵ The MKBK series with equal shape parameters and tailored kernel is

$$f_{X,N}(x) = \left[1 + \sum_{n=4}^N \frac{1}{n!} B_n(0, 0, \Delta\kappa_3, \Delta\kappa_4, \dots, \Delta\kappa_n) M'_n \left(\frac{bx}{1+bx} \right) \right] \beta'(x; L, b). \quad (3.110)$$

3.5.4 The MKBK Series With Realistic Numbers of Terms

The MKBK series expansion of the tailored three-parameter kernel, with $N = 8$ is

$$\begin{aligned} f_{X,8}(x) = & \left[1 + \frac{\Delta\kappa_4}{24} M'_4 \left(\frac{bx}{1+bx} \right) + \frac{\Delta\kappa_5}{120} M'_5 \left(\frac{bx}{1+bx} \right) + \frac{\Delta\kappa_6}{720} M'_6 \left(\frac{bx}{1+bx} \right) \right. \\ & \left. + \frac{\Delta\kappa_7}{5040} M'_7 \left(\frac{bx}{1+bx} \right) + \frac{\Delta\kappa_8 + 35\Delta\kappa_4^2}{40320} M'_8 \left(\frac{bx}{1+bx} \right) \right] \beta'(x; a_1, a_2, b). \end{aligned} \quad (3.111)$$

Compared to the MKGK and MKLK series in Sections 3.2.5 and 3.3.5, we see that $\Delta\kappa_3 = 0$ means that $f_{X,N}(x) = \beta'(x; a_1, a_2, b)$ also for $N = 3$ (i.e. the first correction is for $\Delta\kappa_4$), and that the first non-linear term which is non-zero is delayed until $N = 8$.

In the case of $a_1 = a_2 = L$, we have $\Delta\kappa_3 \neq 0$ and then

$$\begin{aligned} f_{X,7}(x) = & \left[1 + \frac{\Delta\kappa_3}{6} M'_3 \left(\frac{bx}{1+bx} \right) + \frac{\Delta\kappa_4}{24} M'_4 \left(\frac{bx}{1+bx} \right) + \frac{\Delta\kappa_5}{120} M'_5 \left(\frac{bx}{1+bx} \right) \right. \\ & \left. + \frac{\Delta\kappa_6 + 10\Delta\kappa_3^2}{720} M'_6 \left(\frac{bx}{1+bx} \right) + \frac{\Delta\kappa_7 + 35\Delta\kappa_3\Delta\kappa_4}{5040} M'_7 \left(\frac{bx}{1+bx} \right) \right] \beta'(x; L, b), \end{aligned} \quad (3.112)$$

where we found in Section 2.3.10 that $\kappa_{\beta',n} = 0 \Rightarrow \Delta\kappa_n = \kappa_{X,n}$ for $n = 3, 5, 7, \dots$

¹⁵Also, from Section 2.3.10 we know that this assumption leads to all the kernel log-cumulants of odd order (except the first, naturally) being zero.

Chapter 4

The Classical Series Expansions Revisited

Comparing the classical methods in Section 2.4 and the novel approach in Section 3.1 raises the question: Can some of the techniques be applied to the classical case? E.g.: Is there a classical equivalent to the MK Gram-Charlier series with arbitrary kernel in eq. (3.9)? This section demonstrates which results are applicable to the classical case, and which are not.

4.1 Bell Polynomials, Moments and Cumulants

Recall from Section 2.4 that the classical methods were introduced in the late 19th and early 20th century. The Bell polynomials were introduced in [Bell, 1927], and were naturally not available to Gram, Charlier, Edgeworth, and their peers. Now it is time to examine how the Bell polynomials can convey the combinatoric relationships in the classical series expansion methods.

Combining eqs. (2.7) and (2.8) gives the CF of the RV X in terms of the cumulants $c_{X,k}$ of X , i.e.

$$\Phi_X(\omega) = \exp \left\{ \sum_{k=1}^{\infty} c_{X,k} \frac{(j\omega)^k}{k!} \right\}. \quad (4.1)$$

Applying eq. (2.154) gives

$$\Phi_X(\omega) = \sum_{n=1}^{\infty} B_n(c_{X,1}, c_{X,2}, \dots, c_{X,n}) \frac{(j\omega)^n}{n!}, \quad (4.2)$$

where the moments $m_n = B_n(c_{X,1}, \dots, c_{X,n})$ are recognized from either eq. (2.5) or (2.11). As mentioned above, this relationship was known *before* the Bell polynomials were defined. According to [Hald, 2000], it was known already when the cumulants¹ were first given an explicit definition in [Thiele, 1889]. Eq. (4.2) is also not particularly useful in itself, as the Bell polynomials can simply be replaced with the moments. However, when deriving the Gram-Charlier and Edgeworth series, $B_n(\cdot)$ proves to be a very useful tool.

To see why it is useful to apply the Bell polynomials, we need to thoroughly examine the derivation of the Gram-Charlier and Edgeworth series. Let $\rho(x)$ be a PDF whose cumulants all exist and is continuously differentiable an arbitrary number of times.² Letting $c_{\rho,n}$ denote the

¹Then called *semi-invariants*.

²That is, in addition to the existence of all its cumulants, $\rho(x)$ must be non-negative everywhere, integrate to unity and its derivative $D^n \rho(x)$ of order n must exist and be continuous for all $n \in \mathbb{Z}_{\geq 0}$.

cumulants of $\rho(x)$, its CF can be given as

$$\Phi_\rho(\omega) = \exp \left\{ \sum_{n=1}^{\infty} c_{\rho,n} \frac{(j\omega)^n}{n!} \right\}. \quad (4.3)$$

Then the CF of the unknown PDF $f_X(x)$ is expressed in terms of the known CF and cumulants of the kernel and the easily estimated cumulants of X . We have done these operations before, to arrive at eqs. (2.138) and (3.3), but we will now take care to show every intermediate step, in order to afterwards show why this requires that we use the cumulants, and not the moments. The cumulants allow us to manipulate

$$\Phi_X(\omega) = \frac{\exp \left\{ \sum_{n=1}^{\infty} c_{X,n} \frac{(j\omega)^n}{n!} \right\}}{\exp \left\{ \sum_{n=1}^{\infty} c_{\rho,n} \frac{(j\omega)^n}{n!} \right\}} \Phi_\rho(\omega) \quad (4.4)$$

$$\Phi_X(\omega) = \exp \left\{ \sum_{n=1}^{\infty} c_{X,n} \frac{(j\omega)^n}{n!} - \sum_{n=1}^{\infty} c_{\rho,n} \frac{(j\omega)^n}{n!} \right\} \Phi_\rho(\omega) \quad (4.5)$$

$$\Phi_X(\omega) = \exp \left\{ \sum_{n=1}^{\infty} [c_{X,n} - c_{\rho,n}] \frac{(j\omega)^n}{n!} \right\} \Phi_\rho(\omega). \quad (4.6)$$

This approach is not possible with moments, as eq. (4.4) would be replaced with the ratio of sums from eq. (2.5), i.e.

$$\Phi_X(\omega) = \frac{\sum_{n=0}^{\infty} m_{X,n} \frac{(j\omega)^n}{n!}}{\sum_{n=0}^{\infty} m_{\rho,n} \frac{(j\omega)^n}{n!}} \Phi_\rho(\omega), \quad (4.7)$$

which does not permit the same steps as the cumulant representation between eqs. (4.4) and (4.6). In order to recover an expression for the unknown $f_X(x)$, the FT property $(-D)^n f(x) \xrightarrow{\mathcal{F}} (j\omega)^n F_{\mathcal{F}}(\omega)$ must be applied. This requires that the exponential function in eq. (4.6) is reduced to a polynomial in $(j\omega)^n$. The most general form is to use the power series definition

$$\exp\{x\} \equiv \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad (4.8)$$

which gives

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} \frac{\left[\sum_{n=1}^{\infty} [c_{X,n} - c_{\rho,n}] \frac{(j\omega)^n}{n!} \right]^k}{k!} \Phi_\rho(\omega), \quad (4.9)$$

and afterward the inverse FT yields

$$f_X(x) = \sum_{k=0}^{\infty} \frac{\left[\sum_{n=1}^{\infty} [c_{X,n} - c_{\rho,n}] \frac{(-D)^n}{n!} \right]^k}{k!} \rho(x). \quad (4.10)$$

The Gram-Charlier series is then completed by sorting these terms in order of their power of D .³

³Recall that the Edgeworth series instead sorts by r , which was introduced into the expression for the cumulants.

We now propose to use the Bell polynomials instead. It is well known that $B_n(c_1, \dots, c_n) = m_n$, but the idea is to instead apply it to the cumulant *differences* $c_{X,n} - c_{\rho,n}$ in eq. (4.6). This operation is not linear, i.e. $B_n(c_{X,1} - c_{\rho,1}, \dots, c_{X,n} - c_{\rho,n})$ and $m_{X,n} - m_{\rho,n}$ are not generally equal. As it turns out, the Bell polynomials represent the simplest way of sorting the terms in eq. (4.10) by their order of D. This will be demonstrated in the next section.

4.2 The Gram-Charlier Series with Arbitrary Kernel

Using the Bell polynomial of eq. (2.154) in eq. (4.6) gives

$$\Phi_X(\omega) = \left[\sum_{n=0}^{\infty} B_n(c_{X,1} - c_{\rho,1}, \dots, c_{X,n} - c_{\rho,n}) \frac{(j\omega)^n}{n!} \right] \Phi_{\rho}(\omega). \quad (4.11)$$

The part in the square brackets is a formal power series in $(j\omega)^n$, so $f_X(x)$ can be retrieved via the inverse FT as

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} B_n(c_{X,1} - c_{\rho,1}, \dots, c_{X,n} - c_{\rho,n}) \frac{(-D)^n}{n!} \right] \rho(x), \quad (4.12)$$

where the term corresponding to $n = 0$ is always 1 and was put outside the sum to highlight the fact that it is an expansion around the kernel. This is then truncated to a finite number of terms N to give the Gram-Charlier series with arbitrary kernel

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N B_n(\Delta c_1, \dots, \Delta c_n) \frac{(-D)^n}{n!} \right] \rho(x) \approx f_X(x), \quad (4.13)$$

where $\Delta c_n = c_{X,n} - c_{\rho,n}$. This is analogous to eq. (3.9).

4.2.1 The Differential Operator and the Arbitrary Kernel

We know from Section 2.4.1 that $(-D)^n \alpha(x)$ can be replaced with $H_n(x) \alpha(x)$ in the Gaussian case, but what can we say about $(-D)^n \rho(x)$ in general? It turns out that it is possible to derive a general recursion relation, like with the MK framework in Section 3.1.4. Mirroring that approach, we indirectly define $P_n(x)$ by

$$P_n(x) \rho(x) = (-D)^n \rho(x), \quad (4.14)$$

which gives the Rodrigues type formula

$$P_n(x) = \frac{1}{\rho(x)} (-D)^n \rho(x), \quad (4.15)$$

where the typical case is that there are constant factors in $\rho(x)$ which cancel. We see that

$$P_{n+1}(x) = \frac{1}{\rho(x)} (-D)(-D)^n \rho(x) = \frac{1}{\rho(x)} (-D) P_n(x) \rho(x), \quad (4.16)$$

and evaluate this as using the product rule of differentiation to get

$$P_{n+1}(x) = \frac{1}{\rho(x)} [-P_n(x) D \rho(x) - \rho(x) D P_n(x)]. \quad (4.17)$$

Finally, we recognize $-D \rho(x) = P_1(x) \rho(x)$ to get the recursive definition

$$P_{n+1}(x) = P_n(x) P_1(x) - D P_n(x). \quad (4.18)$$

Lemma 3 *If $P_1(x)$ is linear (a first degree polynomial) in x , then $P_n(x)$ is an n th degree polynomial in x .*

Proof Assuming that $P_n(x)$ is an n th degree polynomial in x , we see that $P_n(x)P_1(x)$ is of degree $n + 1$ trivially. Naturally, the derivative operator reduces the degree of a polynomial by one, i.e. $DP_n(x)$ is of degree $n - 1$. Thus, by eq. (4.18) we have that $P_{n+1}(x)$ is an $(n + 1)$ th degree polynomial in x if $P_n(x)$ is of degree n . The initial condition was that $P_1(x)$ is a first degree polynomial, and assuming that $P_n(x)$ is of degree n for arbitrary n implies that it is true for $n + 1$. Thus the proof by induction is complete.

4.2.2 The Gaussian Kernel

Selecting the Gaussian PDF $\alpha(x)$ as the kernel is by far the most common example of the Gram-Charlier series. It results in several significant simplifications in eq. (4.13), and the Hermite polynomials $H_n(x)$ are easily identified from their definition in eq. (2.82). For X standardized (zero mean, unit variance),

$$f_{X,N}(x) = \left[1 + \sum_{n=1}^N \frac{1}{n!} B_n(\Delta c_1, \dots, \Delta c_n) H_n(x) \right] \alpha(x). \quad (4.19)$$

Choosing the kernel parameters s.t. $\Delta c_1 = \Delta c_2 = 0$ has been standard procedure since the Gram-Charlier series was first derived. Also, the Gaussian kernel property that $c_{\alpha,n} = 0 \forall n \geq 3$ further simplifies the expression. Thus, the Gram-Charlier Gaussian kernel series is

$$f_{X,N}(x) = \left[1 + \sum_{n=3}^N \frac{1}{n!} B_n(0, 0, c_{X,3}, \dots, c_{X,n}) H_n(x) \right] \alpha(x), \quad (4.20)$$

since the terms corresponding to $n = 1, 2$ are zero when the kernel is tailored to the data. This can trivially be verified to be the same series as in Section 2.4.2.

For X non-standardized, i.e. arbitrary mean and variance, the Gram-Charlier series must be corrected. We recall that for $u = \frac{x-m}{\varsigma}$, eq. (2.24) gives the relationship between the non-standardized and standardized Gaussian kernels as

$$\alpha(u) = \frac{1}{\varsigma} \alpha(x; m, \varsigma). \quad (4.21)$$

The definition of the Hermite polynomials in eq. (2.82) in terms of the non-standardized argument is

$$(-D_u)^n \frac{1}{\varsigma} \alpha\left(\frac{x-m}{\varsigma}\right) = H_n\left(\frac{x-m}{\varsigma}\right) \frac{1}{\varsigma} \alpha\left(\frac{x-m}{\varsigma}\right), \quad (4.22)$$

where the factors ς^{-1} cancel. In order to substitute this into the Gram-Charlier series, we need to replace D_u with D_x , but $D_x = D_u/\varsigma$, so

$$(-D_x)^n \alpha(x; m, \varsigma) = \frac{1}{\varsigma^n} (-D_u)^n \alpha(x; m, \varsigma) = \frac{1}{\varsigma^n} H_n\left(\frac{x-m}{\varsigma}\right) \alpha(x; m, \varsigma), \quad (4.23)$$

and the Gram-Charlier Gaussian kernel series for non-standardized data is

$$f_{X,N}(x) = \left[1 + \sum_{n=3}^N \frac{1}{n! \varsigma^n} B_n(0, 0, c_{X,3}, \dots, c_{X,n}) H_n\left(\frac{x-m}{\varsigma}\right) \right] \alpha(x; m, \varsigma). \quad (4.24)$$

The Edgeworth series is corrected the same way, i.e. by replacing $H_n(x)$ with $\varsigma^{-n} H_n\left(\frac{x-m}{\varsigma}\right)$, and $\alpha(x)$ with $\alpha(x; m, \varsigma)$.

4.2.3 The Gamma Kernel

The gamma kernel $\gamma(x; a, b)$ is not directly applicable to eq. (4.13), as $D^n\gamma(x)$ is discontinuous at $x = 0$ for high enough n .⁴ Thus, the manipulation from eq. (4.11) to eq. (4.12) is not permitted. In the MK statistics situation in Section 3.2.3, $\gamma(x; a, b)$ does not have this problem, as $x = 0$ is the lower bound in the MT in eq. (2.27).

However, this thesis can provide some new insight on this method as well, namely a simplification using the confluent hypergeometric function of the first kind.⁵ Its definition is given in e.g. [Daalhuis, 2010] as

$${}_1F_1(a; b; x) = \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+k-1)x^k}{b(b+1)\cdots(b+k-1)k!}, \quad (4.25)$$

with the convention that empty products ($k = 0$) equal 1. [Daalhuis, 2010] also notes that ${}_1F_1(\cdot)$ is related to the Laguerre polynomials by

$${}_1F_1(-n; a; x) = \frac{n!}{a(a+1)(a+2)\cdots(a+n-1)} L_n^{(a-1)}(x). \quad (4.26)$$

As n is a non-negative integer, ${}_1F_1(-n; a; x)$ is a polynomial in x of degree n . This can be seen from its definition as

$${}_1F_1(-n; a; x) = \sum_{k=0}^n \frac{n(n-1)\cdots(n-k+1)(-1)^k x^k}{a(a+1)(a+2)\cdots(a+k-1)k!}, \quad (4.27)$$

that is, since $n(n-1)\cdots(n-k+1) = 0 \forall k > n$, the sum can be truncated to the terms $k = 0, 1, \dots, n$.

In the current literature the confluent hypergeometric function has not been applied to the Gram-Charlier series expansion before now.⁶ Replacing x with bx , the property from eq. (4.26) can be inserted into the result for the coefficients of the Gram-Charlier gamma kernel series in eq. (2.125) to give

$$\xi_n = \frac{n!}{\prod_{i=0}^{n-1} (a+i)} \int_0^{\infty} f_X(x) L_n^{(a-1)}(bx) dx = \int_0^{\infty} f_X(x) {}_1F_1(-n; a; bx) dx = E\{{}_1F_1(-n; a; bx)\}, \quad (4.28)$$

where the expectation is w.r.t. x . This permits a compact explicit formula for the Gram-Charlier gamma kernel series:

$$f_X(x) = \left[1 + \sum_{n=3}^{\infty} E\{{}_1F_1(-n; a; bx)\} L_n^{(a-1)}(bx) \right] \gamma(x; a, b). \quad (4.29)$$

To evaluate each ξ_n in practice, we simply replace the powers x^k in the definition of ${}_1F_1(-n; a+1; x)$ with the moments m_k to get an explicit formula for the coefficients,

$$\xi_n = \sum_{k=0}^n m_k \frac{n(n-1)\cdots(n-k+1)(-1)^k b^k}{a(a+1)(a+2)\cdots(a+k-1)k!}, \quad (4.30)$$

⁴Formally, applying the inverse FT to $\Phi_\gamma(\omega)$ is not permitted because the $(j\omega)^n \Phi_\gamma(\omega)$ is not integrable for high enough n , as $\int_{\mathbb{R}} |(j\omega)^n \Phi_\gamma(\omega)| d\omega < \infty$ is not satisfied.

⁵Also known as Kummer's function, after the man who introduced them in [Kummer, 1837].

⁶This is obviously a statement with the caveat that it is not possible to know everything about the current state of the literature. However, the Gram-Charlier series with a gamma kernel is very little used, and [Gaztanaga et al., 2000], which does discuss the series, contains no mention of the hypergeometric functions at all.

where the term $k = 0$ is 1 for any n . Tailoring the gamma kernel by using eq. (2.129) simplifies the terms $k = 1, 2$ significantly, giving

$$\xi_n = 1 - n + \frac{n(n-1)}{2} + \sum_{k=3}^n m_k \frac{n(n-1)\cdots(n-k+1)(-1)^k b^k}{a(a+1)(a+2)\cdots(a+k-1)k!} \quad (4.31)$$

$$\xi_n = \frac{(n-1)(n-2)}{2} + \sum_{k=3}^n m_k \frac{n(n-1)\cdots(n-k+1)(-1)^k b^k}{a(a+1)(a+2)\cdots(a+k-1)k!} \quad (4.32)$$

Alternatively, the hypergeometric functions are readily available in textbooks, mathematics software and online.

4.3 The Edgeworth Series

As mentioned in Section 2.4.5, [Blinnikov and Moessner, 1998] provided a formula for finding the terms associated with each power of $r^{-1/2}$ in eq. (2.142). In this section, an alternative formula based on the Bell polynomials is presented.

The premise for the Edgeworth series is that the cumulant differences can be formulated as in eq. (2.139), allowing us to go from eq. (4.6) to

$$\Phi_X(\omega) = \exp \left\{ \sum_{k=3}^{\infty} \frac{\lambda_k}{r^{\frac{k}{2}-1}} \frac{(j\omega)^k}{k!} \right\} \Phi_{\alpha}(\omega). \quad (4.33)$$

As in [Blinnikov and Moessner, 1998], the power series in the expression for the CF $\Phi_X(\omega)$ must be viewed not in terms of $(j\omega)$, but rather in terms of $r^{-1/2}$. Mirroring our approach in Section 3.4.2, this is achieved with an index shift $k \rightarrow k-2$, and coincidentally this also implies that the summation starts from $k=1$. Then,

$$\Phi_X(\omega) = \exp \left\{ \sum_{k=1}^{\infty} \frac{\lambda_{k+2}}{(r^{1/2})^k} \frac{(j\omega)^{k+2}}{(k+2)!} \right\} \Phi_{\alpha}(\omega), \quad (4.34)$$

and using $(k+2)! = k!(k+1)(k+2)$ gives

$$\Phi_X(\omega) = \exp \left\{ \sum_{k=1}^{\infty} \frac{\lambda_{k+2}(j\omega)^{k+2}}{(k+1)(k+2)} \frac{(r^{-1/2})^k}{k!} \right\} \Phi_{\alpha}(\omega) = \exp \left\{ \sum_{k=1}^{\infty} \xi_k(j\omega) \frac{(r^{-1/2})^k}{k!} \right\} \Phi_{\alpha}(\omega), \quad (4.35)$$

where the coefficients ξ_k are defined as

$$\xi_k(j\omega) = \frac{\lambda_{k+2}(j\omega)^{k+2}}{(k+1)(k+2)}, \quad (4.36)$$

i.e. they are independent of r . Up to this point, the derivation has loosely followed [Blinnikov and Moessner, 1998], albeit with significant differences in notation. Now, however, the Bell polynomials are proposed as a simpler way of achieving a polynomial representation in terms of $(j\omega)$. Using eq. (2.154) gives

$$\Phi_X(\omega) = \left[\sum_{n=0}^{\infty} B_n(\xi_1(j\omega), \dots, \xi_n(j\omega)) \frac{(r^{-1/2})^n}{n!} \right] \Phi_{\alpha}(\omega). \quad (4.37)$$

Upon closer inspection, all $B_n(\cdot)$ are polynomials of $(j\omega)$ since all $\xi_n(j\omega)$ are monomials of $(j\omega)$. Like in Section 2.4.5, the linearity of the FT now permits an inverse transform to recover the PDF, using the property that $(-D)^n f(x) \xleftrightarrow{\mathcal{F}} (j\omega)^n F_{\mathcal{F}}(\omega)$ to get

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} B_n(\xi_1(-D), \dots, \xi_n(-D)) \frac{(r^{-1/2})^n}{n!} \right] \alpha(x), \quad (4.38)$$

where

$$\xi_n(-D) = \frac{\lambda_{n+2}(-D)^{n+2}}{(n+1)(n+2)}. \quad (4.39)$$

Explicitly, for X standardized,

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} B_n \left(\frac{\lambda_3(-D)^3}{6}, \dots, \frac{\lambda_{n+2}(-D)^{n+2}}{(n+1)(n+2)} \right) \frac{(r^{-1/2})^n}{n!} \right] \alpha(x), \quad (4.40)$$

and the result eq. (4.23) can be used to easily extend this to non-standardized data the usual way. The first few terms were stated in eq. (2.142). Again it is possible to recombine the powers of r with their corresponding λ_n to retrieve the cumulants, i.e.

$$f_X(x) = \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} B_n \left(\frac{c_3(-D)^3}{6}, \dots, \frac{c_{n+2}(-D)^{n+2}}{(n+1)(n+2)} \right) \right] \alpha(x). \quad (4.41)$$

As with the MKE series, $(-D)^n$ cannot be replaced with $H_n(x)$ in this representation, as the non-linear terms of $B_n(\cdot)$ would then result in e.g. $[H_n(x)]^2$ instead of the correct $H_{2n}(x)$.

4.4 Author's Comment

I have not seen the Bell polynomials used when deriving the Gram-Charlier or Edgeworth series before. In fact, my general impression is that the derivation of these series expansions is usually presented in ways which are more complicated than necessary. The derivation should highlight their similarity by emphasizing the fact that they differ only in the order that they include the terms in the double infinite sum in eq. (2.141).

For someone not intimately familiar with the theory, it is easy to misinterpret the Gram-Charlier series in eq. (2.111) as

$$f_X(x) = \left[1 + \sum_{n=3}^{\infty} \frac{c_n}{n!} H_n(x) \right] \alpha(x). \quad (4.42)$$

This arises from the fact that eq. (4.20) is simplified because $B_n(0, 0, c_{X,3}, \dots, c_{X,n}) = c_{X,n}$ for $n \leq 5$. It is then important to also emphasize that this is just the result of the properties of the tailored Gaussian kernel. The use of the Bell polynomials seeks to convey this non-trivial combinatoric relationship in a simple and concise way.

The Bell polynomials were also used in the Edgeworth series in eq. (4.40) in a representation which to my knowledge has not been presented or discussed before. This is proposed as an alternative to [Blinnikov and Moessner, 1998], where the authors presented the same Edgeworth series without using the Bell polynomials. Recall from Section 2.5 that the complete Bell polynomials are *sums of a sum of products of powers of functions of the cumulants*, with the inner sum being over the set \mathcal{K} of all combinations of integers which solve two equations. This complicated mathematical relationship is compressed into just $B_n(\cdot)$, which is readily available for any reasonable n for anyone wanting to implement the Edgeworth series. To put things

bluntly, eq. (4.40) is much more suited to a world where the Bell polynomials can be found on the Internet in a matter of seconds.

It can be discussed whether the use of Bell polynomials helps the reader to actually understand the series, but it is my belief that this novel way of presenting the Gram-Charlier and Edgeworth series has some distinct advantages. Namely, it is very clear what the differences between the two methods are, and the resulting expressions are explicit and compact. In fact, the Gram-Charlier series has traditionally been presented by merely stating that the terms are collected according to their power of D , while $B_n(\cdot)$ actually does this.

Another topic is the failure of eq. (4.13) to extend to the gamma kernel. Fundamentally, it is an example of the limitations of the classical methods when it comes to non-negative RVs. The problem comes directly from attempting to apply the FT to a function with support on $\mathbb{R}_{\geq 0}$, and it is no longer surprising that the MT seems better suited to these functions. If no other kernels than the Gaussian are used for the Gram-Charlier series, it can perhaps be just as instructive to start directly with $\alpha(x)$.

The positive result that was presented w.r.t. the gamma kernel in Section 4.2.3 is a novel approach to computing the coefficients ξ_n . The result in eq. (4.29) is the most compact explicit representation of the series that I am aware of, and it came about from having to actually implement the series. I needed to hard-code the coefficients to save time on the simulations themselves, but using the currently available methods was cumbersome, to say the least. In my experience, the fastest way of computing ξ_n was to find ${}_1F_1(-n; a + 1; x)$ online. At the time of writing, the function can be evaluated using the web interface of Wolfram Alpha, e.g. [http://www.wolframalpha.com/input/?i=Hypergeometric1F1\(-7,a%2B1,x\)](http://www.wolframalpha.com/input/?i=Hypergeometric1F1(-7,a%2B1,x)) for $n = 7$, and replacing the powers of x with the corresponding moments from there. A trick is the fact that the contribution of the terms $k = 0, 1, 2$ combined is always $(n - 1)(n - 2)/2$ for the tailored kernel, as shown in Section 4.2.3.

Chapter 5

Results

5.1 Preliminaries

Before presenting any results, we will review the methods and PDFs used, and comment on the relevant results included in a paper we have written based on the work in this thesis.

5.1.1 Method – Approximation and Estimation

The first part of this chapter is concerned with approximating a known *target* PDF, i.e. both the distribution family and its parameter values are available to the series expansions. This means that the kernel parameters and log-cumulant differences are exact, with the latter computed from the theoretical log-cumulants $\kappa_{X,n}$ of the target PDF in each case.

In the second part of this chapter, the series expansions only have access to random data generated/synthesized from the target PDFs. That is, the distributions are unknown,¹ and thus the kernel parameters and log-cumulants have to be estimated, e.g. by $\langle \kappa_{X,n} \rangle$. For the latter, this is usually done by computing the empirical log-moments of the data sample $\{x_1, x_2, \dots, x_n\}$, namely

$$\langle \mu_N \rangle \equiv \frac{1}{n} \sum_{i=1}^n (\log x_i)^n, \quad (5.1)$$

and converting these to the empirical log-cumulants using eq. (2.42), but we assess alternative approaches as well.

On a slightly technical note, when discretizing x we chose grid resolutions which ensured that there would be around 10,000 x -values, and it has been verified that this is sufficient to ensure that the results are not affected by the discretization. We limited x to the values where the target PDF was at least 1/1000 of its maximum. This means that we have not focused on the ability of the methods to model the tails of the target PDFs. This was a deliberate choice, as it quickly became clear to us that the MK series expansions generally perform poorly in the tails, due to perturbations caused by the polynomial corrections.² This mirrors the classical series expansions, as demonstrated in [Blinnikov and Moessner, 1998]. Thus, we chose to assess the performance for the values of x where the methods can feasibly be applied.³

¹Naturally, *we* know the target PDFs and can evaluate the success of the series with the distances from Section 2.6.

²This arbitrary choice affects the quantitative results, but we performed some additional testing to verify that it did not have too much impact on our conclusions. If anything, the threshold could reasonably have been more restrictive, and this would presumably have cast a more favorable light on the MK series expansions.

³The Cornish-Fisher expansion is better suited to approximation of the tails, see [Hill and Davis, 1968] for

5.1.2 The Target PDFs

Throughout this chapter we will approximate and estimate several different distributions, with multiple sets of parameters for each. This is to ensure that the conclusions are not biased due to certain series expansions performing exceptionally for particular target PDFs. We chose to use a handful of PDFs which are based on transformations and/or generalizations of the gamma distribution. These are commonly encountered in coherent imaging, e.g. SAR. We use the shape and location parametrization from eq. (2.56), as we are simulating data. We recall from Section 2.3.9 that the shape is interpreted as the number of looks, a global parameter which is estimated based on the entire image.⁴ Finally, we remark that for some of the transformations and generalizations of $\gamma(x; L, m)$, the location m does not equal the mean $E\{x\}$. This fact will be noted for each distribution it applies to.

Gamma Distribution The most basic target PDF used is the gamma distribution with shape (number of looks) and location parametrization from eq. (2.56).

Inverse Gamma Distribution If X is gamma distributed with shape L and location m , we say that $Y = 1/X$ follows the *inverse gamma* distribution with PDF

$$\gamma^{-1}(y; L, m) = \left(\frac{L}{m}\right)^L \frac{1}{\Gamma(L)} y^{-L-1} \exp\left\{-\frac{L}{my}\right\}, \quad (5.2)$$

where $\gamma^{-1}(y; L, m) = 0$ for $y \leq 0$. The present notation is a combination of [Nicolas, 2016] and [Li et al., 2011], adapted to be consistent with the rest of this thesis. The mean $E\{Y\}$ of an inverse gamma distributed RV does *not* equal its location parameter m , but, as both references point out,

$$E\{Y\} = \begin{cases} \frac{1}{m} \frac{L}{L-1} & L > 1, \\ \infty & \text{otherwise.} \end{cases} \quad (5.3)$$

Naturally, the mean of $1/Y$ is m , since $1/Y$ is by definition gamma distributed with location m . The log-cumulants of an inverse gamma distributed RV were already in [Nicolas, 2002] found to be

$$\kappa_{\gamma^{-1}, n} = \begin{cases} -\psi^{(0)}(L) + \log\left(\frac{L}{m}\right) & n = 1, \\ (-1)^n \psi^{(n-1)}(L) & n \geq 2. \end{cases} \quad (5.4)$$

Generalized Gamma Distribution With X still gamma distributed, we say that $Z = X^d$ follows the *generalized gamma distribution* (GFD) with PDF

$$\text{GFD}(z; L, m, \nu) = \left(\frac{L}{m}\right)^{L\nu} \frac{|\nu|}{\Gamma(L)} z^{L\nu-1} \exp\left\{-\left(\frac{Lz}{m}\right)^\nu\right\}, \quad (5.5)$$

where $\text{GFD}(z; L, m, \nu) = 0$ for $z \leq 0$ and the *power* parameter ν is non-zero (but can be negative). The notation used here is again an adaptation of several references to maintain consistency throughout this thesis. For more information on the GFD, see [Stacy, 1962], where it was originally introduced; [Tadikamalla, 1979], for a random sampling procedure; [Nicolas, 2016], for a summary of its classical statistics and MK statistics; and [Li et al., 2011], where the authors present a clever and fast method of fitting $\text{GFD}(\cdot)$ using the MoLC while avoiding

the classical version and [Pastor et al., 2016] for the MK statistics version.

⁴In real-world applications, the current estimators for L have such good accuracy that it can be realistic to consider it a known quantity when estimating the target PDFs based on the data. Hence, some of our testing will be conducted with the true value of L available to the series expansions in estimation scenarios.

the cumbersome iterative numerical approach. These two most recent sources include the distribution log-cumulants

$$\kappa_{\text{GFD},n} = \begin{cases} \frac{\psi^{(0)}(L)}{\nu} + \log\left(\frac{m}{L}\right) & n = 1, \\ \frac{\psi^{(n-1)}(L)}{\nu^n} & n \geq 2. \end{cases} \quad (5.6)$$

The last two distributions we use as targets arise from the doubly stochastic product model, namely the K and G^0 distributions. They were proposed in [Jakeman and Pusey, 1978] and [Frery et al., 1997] as the models for SAR intensity of heterogeneous and extremely heterogeneous scenes, respectively. This indicates that they are good candidates for our analysis, as the goal is to forecast the performance of the methods in real-world scenarios.

K Distribution Recall from Section 2.3.9, that the product of two gamma distributed RVs is K distributed. That is, it has PDF

$$K(x; L, m, M) = \frac{2LM}{m\Gamma(L)\Gamma(M)} \left(\frac{LMx}{m}\right)^{\frac{M+L}{2}-1} K_{M-L}\left(2\sqrt{\frac{LMx}{m}}\right), \quad x, L, m, M > 0. \quad (5.7)$$

where $K_{M-L}(\cdot)$ is the modified Bessel function of the second kind with order $M - L$ and $K(x; L, m, M) = 0$ for $x \leq 0$. As discussed in [Nicolas, 2002], the K distribution is most easily analyzed using MK statistics. From the review of the product model in Section 2.3.8, we see that $K(x; L, m, M)$ is the Mellin convolution (from eq. (2.50)) of two $\gamma(\cdot)$ PDFs.⁵ The log-cumulants of a K distributed RV are then readily retrieved using the additive nature of the MK CGF in the product model, specifically eq. (2.53), yielding

$$\kappa_{K,n} = \begin{cases} \psi^{(0)}(L) + \psi^{(0)}(M) + \log\left(\frac{m}{LM}\right) & n = 1, \\ \psi^{(n-1)}(L) + \psi^{(n-1)}(M) & n \geq 2. \end{cases} \quad (5.8)$$

G^0 Distribution If the depicted scene is extremely heterogeneous, e.g. an urban area, [Frery et al., 1997] argues that the K distribution is not applicable, instead proposing the G^0 distribution with PDF

$$G^0(x; g, L, M) = \frac{L^L \Gamma(L - M) x^{L-1}}{g^M \Gamma(L) \Gamma(-M) (g + Lx)^{L-M}}, \quad x, g > 0, M < 0, \quad (5.9)$$

with $G^0(x; g, L, M) = 0$ for $x \leq 0$. As mentioned in Section 2.3.10, the beta prime distribution in eq. (2.67) is actually an alternative parametrization of the G^0 distribution. However, our inclusion of the beta prime distribution in this thesis is not redundant, as it is used exclusively as a kernel. The G^0 distribution, on the other hand, is used as a target PDF, e.g. to simulate real-world data.⁶

The log-cumulants are naturally just a re-parametrized version of the beta prime log-cumulants in eq. (2.68), i.e.

$$\kappa_{G^0,n} = \begin{cases} \psi^{(0)}(L) - \psi^{(0)}(-M) - \log\left(\frac{g}{L}\right) & n = 1, \\ \psi^{(n-1)}(L) + (-1)^n \psi^{(n-1)}(-M) & n \geq 2. \end{cases} \quad (5.10)$$

⁵These PDFs have shapes L and M , and location 1 and m , respectively. The former has unit mean by convention, since it is not possible to discern the contribution to the mean from the constituent RVs.

⁶This is same situation as with the two parametrizations of the gamma distribution – one is reserved for use as a kernel, the other for data.

5.1.3 Results Included in the Appendix

Based on the findings in this thesis, we have written drafts for two papers, with the intention of submitting them to scientific journals. One of these contains an extensive analysis of the MKGK, MKLK and MKE series. That analysis compares these series expansions to other methods of approximating and estimating PDFs. These other methods include a wide array of classical and state-of-the-art methods, ranging from simple to complicated, and from fast to computationally demanding. This comparison allowed us to assess the merits of the MK series framework under the premise that it must provide some advantage over the competing methods.

Instead of repeating the same experiments, or doing something very similar, the current version of that paper is included in the appendix of this thesis. Thus, we can here devote full attention to only the MK series expansions.

5.2 Approximation

In this section, we assess the mathematical properties of the MK series expansions, by approximating known target PDFs. In order to get a basic understanding of the methods, we first present two cases which are analyzed in-depth. Then, we perform a broader (but shallower) analysis with several target PDFs. This will hopefully shed some light on whether some methods are better suited to model certain types of distributions. Finally, we vary the number of correcting terms in the series to assess the convergence of the methods.

5.2.1 Two Introductory Examples

We start by plotting two target PDFs, their approximations, and the corresponding errors. This will hopefully give us a sense of what the MK series expansions *are*, what they look like and how they modify their kernel, before we move on to more substantial testing.

The kernel parameters were tailored by requiring $\Delta\kappa_1 = \Delta\kappa_2 = 0$, and for the beta prime kernel $\Delta\kappa_3 = 0$ also. The series are corrected for log-cumulant differences up to order 4.⁷ This choice is largely arbitrary, however it is not unfounded. Correcting only up to order 3 renders the MKLK and MKE series identical, and does not allow for any terms beyond the kernel in the MKBK series. Conversely, [Kendall et al., 1994] comments that estimates of the classical cumulants beyond the fourth are unreliable due to sample fluctuations, so our initial assumption is that the most realistic comparison is limited to the first four log-cumulants.⁸

Figure 5.1 is the first case we examine, and the immediate impression from the top two plots is that all methods are fairly successful. In fact, the methods are so successful that it is hard to discern each individual approximation. In order to compare the methods, we can examine the approximation error in the bottom plots instead. The K distribution has positive linear skewness which is manifested in the top left plot as a heavy right tail. As this trait is shared among virtually all distributions associated with non-negative RVs, it is not a very useful observation. Hence, we also plotted the PDF on the logarithmic scale to evaluate its *logarithmic skewness*,

⁷Specifically, this implies $N = 4$ in the MKGK series, like in eq. (3.42), $N = 4$ also in the MKLK series in eq. (3.63), discarding terms of order $r^{-3/2}$ and beyond in the MKE series in eq. (3.82), and $N = 4$ again in the MKBK series in eq. (3.106).

⁸We will test this assumption thoroughly later in this chapter, but we initially expect that when estimating unknown distributions, the error resulting from the sample fluctuations outweigh the benefit of the correcting terms beyond $\Delta\kappa_4$. Even though we now approximate *known* distributions, we use the same, realistic truncation.

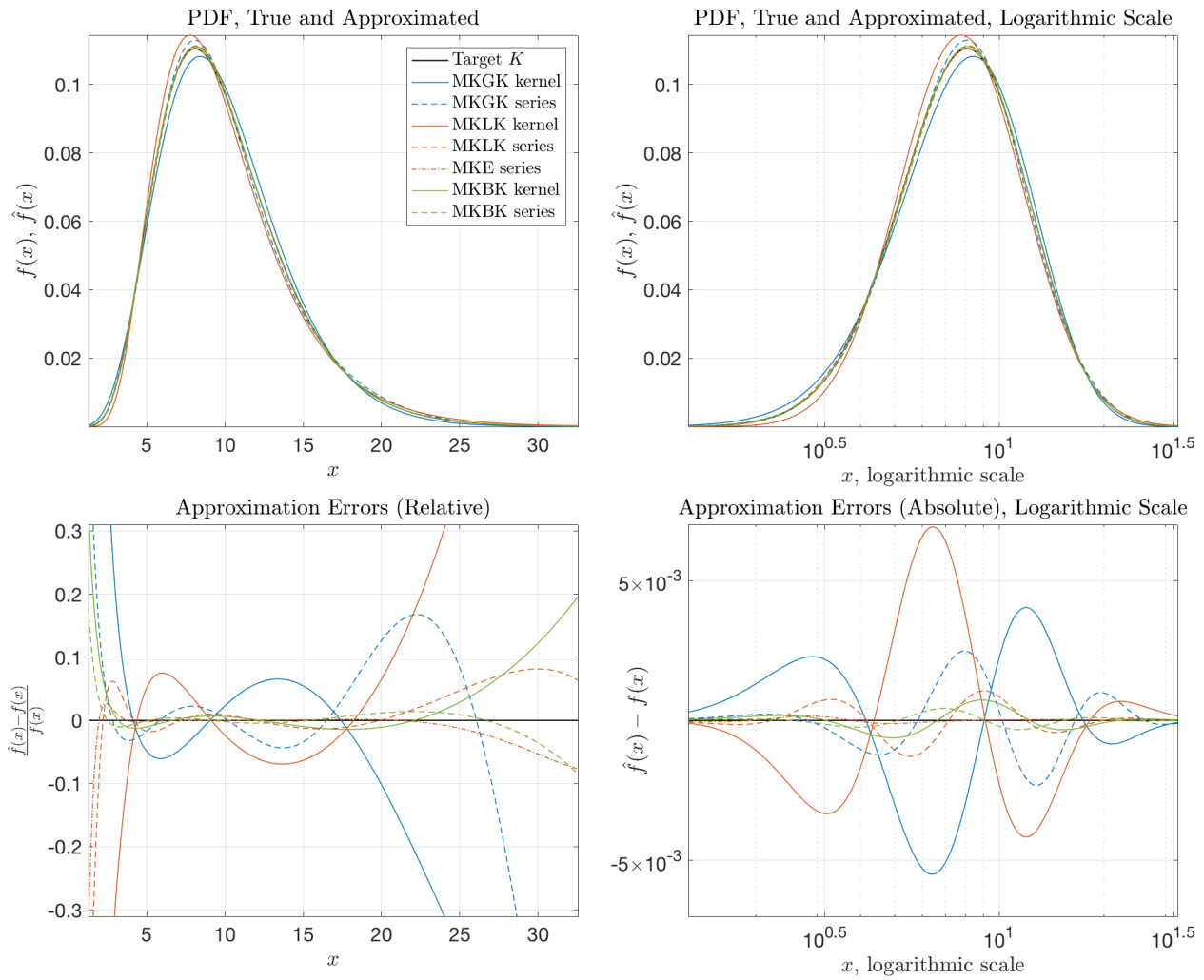


Figure 5.1: The four MK series expansions used to approximate the K distribution with parameters $L = 16$, $m = 10$, and $M = 10$. The solid lines are the kernels, and the dashed lines are the series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$.

quantitatively κ_3 . Alternatively, we can compute $\kappa_{K,3}$ using eq. (5.8) to see that it is negative.⁹ Visually, this corresponds to seeing that the left tail is heavier than the right, when x is plotted on a logarithmic scale.

The relative approximation errors (lower left) are naturally quite low for $x \in [4, 18]$, i.e. when the target PDF is at its highest. When the target PDF approaches zero, the relative errors increase in magnitude, but the absolute errors (lower right) tend to zero. We can easily see that, in terms of absolute errors, all series improve on their kernels, with the MKE series seemingly outperforming the others.

The second case examined is the GFD, presented in Figure 5.2. We see that the approximations struggle to fit the target PDF, i.e. it is a far bigger challenge than the first case. In terms of the absolute approximation errors, they are of the same magnitude as in Figure 5.1, but that target distribution had much larger probability density values. The result is naturally that the relative errors in the lower left plot tells the true story of methods struggling to model the target PDF, in several instances failing completely.

⁹By properties of $\psi^{(2)}(\cdot)$, both gamma and K distributed RVs always have negative κ_3 . The literature on MK statistics and SAR sometimes refers to logarithmic skewness relative to the gamma distribution, e.g. [Anfinssen et al., 2011]. A K distributed RV always has negative κ_3 relative to its constituent gamma RVs.

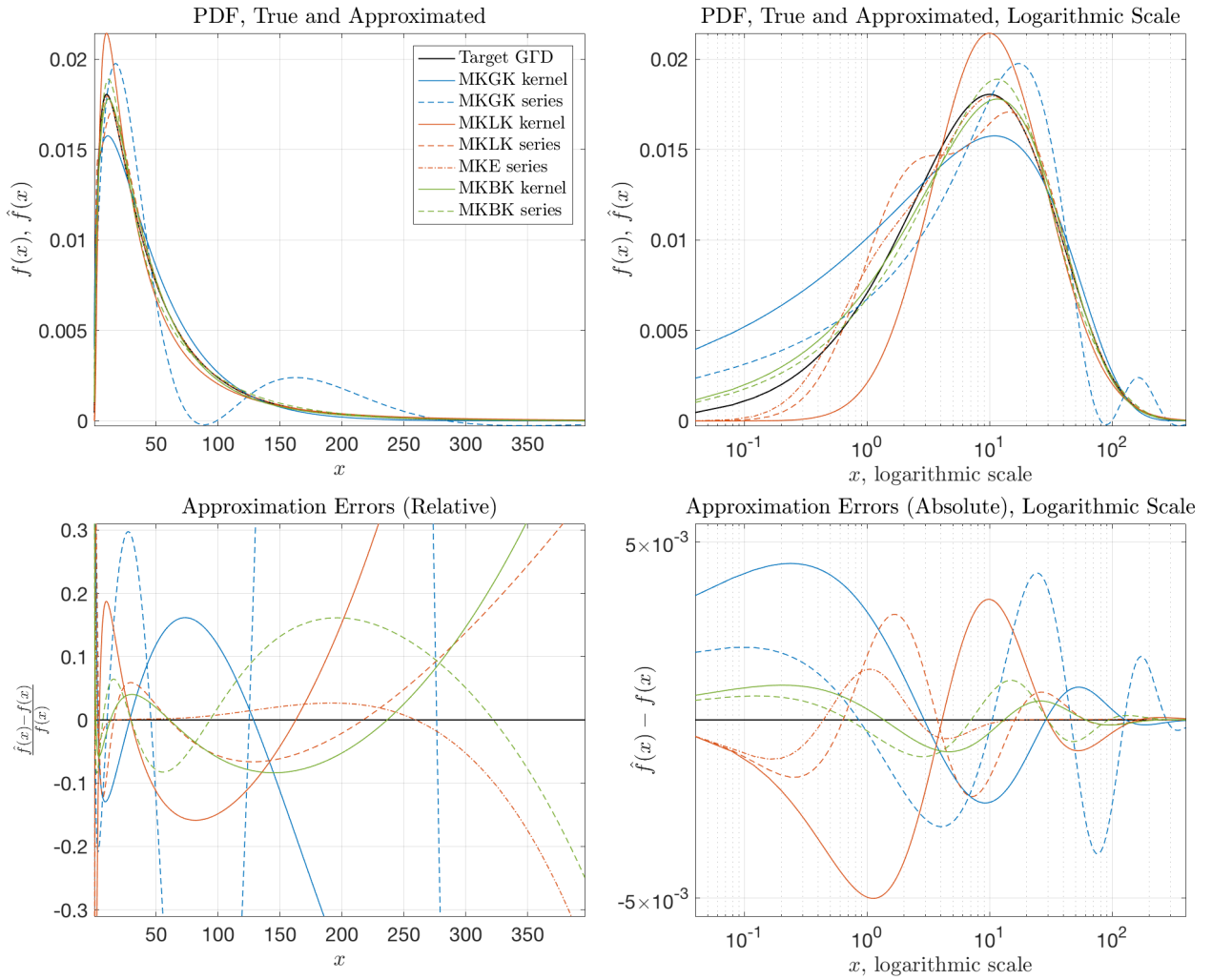


Figure 5.2: The four MK series expansions used to approximate the GFD with parameters $L = 4$, $m = 10$, and $\nu = 0.5$. The solid lines are the kernels, and the dashed lines are the series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$.

From the top left plot we see that the target PDF has a very heavy tail and that the MKGGK series diverges completely. From the top right plot we see that the logarithmic skewness is again negative (this can be verified with eq. (5.6)), but it also gives us an interesting perspective as to what is happening with the MKGGK series. Specifically, the correcting terms reduces the large approximation error of the gamma kernel below 10^0 , but the resulting perturbations cause a divergence around 10^2 , and even negative PDF approximation values on some parts of the x axis. This trade-off can seemingly be justified by a visual inspection of the top right plot, but on the linear scale the improvements all but disappear and the perturbations are enhanced. The same happens with the MKBK series, but on a much smaller scale. As we will see in the next section, the numerical dissimilarity measures also indicate that only the MKLK and MKE series are better than their kernel in this case.¹⁰

5.2.2 Approximations to Numerous Target PDFs Tabulated

In this section, we perform a broad comparison with the series expansions applied to four distinct sets of known parameters for each of the five target PDFs discussed in Section 5.1.2. In several

¹⁰However, changing the power parameter to $\nu = 2$ but leaving the others unchanged, results in a target PDF for which the MKGGK series is the top performer. This could indicate that its emphasis on the low x values is sometimes justified.

cases, one of the kernels coincide with the target PDF, i.e. the MKGK series expansion of the gamma kernel is used to model a gamma distribution, or the MKBK series expansion of the beta prime kernel is used to model distributions that are special cases of $\beta'(\cdot)$. Note also that the respective series have no correcting terms, as all log-cumulant differences are zero. In these cases, the only information in $d_{\text{KL}}(\cdot)$ is the inaccuracy stemming from our computations, e.g. errors of magnitude 10^{-16} resulting in $d_{\text{KL}}(\cdot) \sim 10^{-8}$. As this is not of interest to us here, we simply state the methods as "exact" in these cases.

We present only the Kullback-Leibler distance $d_{\text{KL}}(\cdot)$ between the approximations and the target PDFs. The other dissimilarity measures discussed in Section 2.6 were largely in agreement with $d_{\text{KL}}(\cdot)$, but did not have as good contrast, especially when several methods performed well. In the cases where the Bhattacharyya distance disagreed w.r.t. which approximation was closest to the true PDF, we will explicitly state this.

Table 5.1: Approximating a broad range of target distributions with MK series expansions, Kullback-Leibler distance to the true PDF. The kernels are tailored and the series expansions correct for $\Delta\kappa_3$ and $\Delta\kappa_4$, best method in bold.

Method \ θ	$\gamma(x; \theta), \theta=[L, m]$		$\gamma^{-1}(x; \theta), \theta=[L, m]$		GFD($x; \theta$), $\theta=[L, m, d]$		$K(x; \theta), \theta=[L, m, M]$		$G^0(x; \theta), \theta=[L, g, M]$	
	[4, 1]	[4, 10]	[4, 1]	[4, 10]	[4, 10, 0.5]	[4, 10, 2]	[4, 10, 1]	[4, 10, 10]	[4, 1, -10]	[4, 10, -10]
MKGK kernel	Exact	Exact	$7.34 \cdot 10^{-2}$	$7.34 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$	$4.81 \cdot 10^{-3}$	$6.61 \cdot 10^{-3}$	$4.22 \cdot 10^{-3}$	$8.54 \cdot 10^{-3}$	$8.54 \cdot 10^{-3}$
MKGK series	Exact	Exact	$5.69 \cdot 10^{-2}$	$5.69 \cdot 10^{-2}$	$1.10 \cdot 10^0$	$7.20 \cdot 10^{-4}$	$1.54 \cdot 10^0$	$2.14 \cdot 10^{-2}$	$2.89 \cdot 10^{-2}$	$2.72 \cdot 10^{-2}$
MKLK kernel	$1.83 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	$1.87 \cdot 10^{-2}$	$1.87 \cdot 10^{-2}$	$1.59 \cdot 10^{-2}$	$1.91 \cdot 10^{-2}$	$3.87 \cdot 10^{-2}$	$9.26 \cdot 10^{-3}$	$5.66 \cdot 10^{-3}$	$5.66 \cdot 10^{-3}$
MKLK series	$2.61 \cdot 10^{-3}$	$2.61 \cdot 10^{-3}$	$2.29 \cdot 10^{-3}$	$2.29 \cdot 10^{-3}$	$2.52 \cdot 10^{-3}$	$2.56 \cdot 10^{-3}$	$2.27 \cdot 10^{-2}$	$7.70 \cdot 10^{-4}$	$3.87 \cdot 10^{-4}$	$3.87 \cdot 10^{-4}$
MKE series	$4.02 \cdot 10^{-3}$	$3.76 \cdot 10^{-3}$	$8.90 \cdot 10^{-3}$	$8.34 \cdot 10^{-3}$	$4.06 \cdot 10^{-4}$	$6.09 \cdot 10^{-3}$	$3.91 \cdot 10^{-3}$	$1.62 \cdot 10^{-4}$	$1.59 \cdot 10^{-4}$	$1.59 \cdot 10^{-4}$
MKBK kernel	Exact	Exact	Exact	Exact	$1.00 \cdot 10^{-3}$	$6.91 \cdot 10^{-3}$	$2.50 \cdot 10^{-4}$	$9.69 \cdot 10^{-5}$	Exact	Exact
MKBK series	Exact	Exact	Exact	Exact	$1.98 \cdot 10^{-3}$	$6.69 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$	$6.21 \cdot 10^{-5}$	Exact	Exact

Method \ θ	$\gamma(x; \theta), \theta=[L, m]$		$\gamma^{-1}(x; \theta), \theta=[L, m]$		GFD($x; \theta$), $\theta=[L, m, d]$		$K(x; \theta), \theta=[L, m, M]$		$G^0(x; \theta), \theta=[L, g, M]$	
	[16, 1]	[16, 10]	[16, 1]	[16, 10]	[16, 10, 0.5]	[16, 10, 2]	[16, 10, 1]	[16, 10, 10]	[16, 1, -10]	[16, 10, -10]
MKGK kernel	Exact	Exact	$1.89 \cdot 10^{-2}$	$1.89 \cdot 10^{-2}$	$4.55 \cdot 10^{-3}$	$1.75 \cdot 10^{-3}$	$6.83 \cdot 10^{-4}$	$2.74 \cdot 10^{-3}$	$1.88 \cdot 10^{-2}$	$1.88 \cdot 10^{-2}$
MKGK series	Exact	Exact	$2.26 \cdot 10^{-3}$	$2.26 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$4.72 \cdot 10^{-4}$	$2.36 \cdot 10^{-2}$	$8.57 \cdot 10^{-4}$	$3.50 \cdot 10^{-3}$	$3.50 \cdot 10^{-3}$
MKLK kernel	$4.69 \cdot 10^{-3}$	$4.69 \cdot 10^{-3}$	$4.74 \cdot 10^{-3}$	$4.74 \cdot 10^{-3}$	$4.50 \cdot 10^{-3}$	$4.75 \cdot 10^{-3}$	$5.36 \cdot 10^{-2}$	$3.38 \cdot 10^{-3}$	$6.83 \cdot 10^{-4}$	$6.83 \cdot 10^{-4}$
MKLK series	$1.45 \cdot 10^{-4}$	$1.45 \cdot 10^{-4}$	$1.30 \cdot 10^{-4}$	$1.30 \cdot 10^{-4}$	$1.54 \cdot 10^{-4}$	$1.39 \cdot 10^{-4}$	$4.00 \cdot 10^{-2}$	$8.43 \cdot 10^{-5}$	$6.53 \cdot 10^{-6}$	$6.53 \cdot 10^{-6}$
MKE series	$1.31 \cdot 10^{-5}$	$1.31 \cdot 10^{-5}$	$1.32 \cdot 10^{-5}$	$1.32 \cdot 10^{-5}$	$1.34 \cdot 10^{-5}$	$1.30 \cdot 10^{-5}$	$6.93 \cdot 10^{-3}$	$6.23 \cdot 10^{-6}$	$4.03 \cdot 10^{-6}$	$4.03 \cdot 10^{-6}$
MKBK kernel	Exact	Exact	Exact	Exact	$1.15 \cdot 10^{-4}$	$7.79 \cdot 10^{-3}$	$4.62 \cdot 10^{-6}$	$4.31 \cdot 10^{-5}$	Exact	Exact
MKBK series	Exact	Exact	Exact	Exact	$4.09 \cdot 10^{-5}$	$7.35 \cdot 10^{-3}$	$2.92 \cdot 10^{-5}$	$1.04 \cdot 10^{-5}$	Exact	Exact

The results are presented in Table 5.1, and from these we can get several insights about the series expansions. Immediately, the MKE series stands out as the top performer in most cases. In fact, when using the Bhattacharyya distance, the MKE series provides the closest approximation also for the four cases where $L = 16$ with the $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$ distributions, and GFD(\cdot) with $\theta = [4, 10, 2]$. On the other hand, the MKE series must share its title of best performer with the MKLK series in the case of the G^0 distribution with $\theta = [16, 10, -10]$ – the two were not separable at the present uncertainty levels in terms of $d_{\text{B}}(\cdot)$. Otherwise, the dissimilarity measures were in agreement.

Fitting a beta prime PDF kernel to the $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$ distributions is an interesting problem. To see why, we combine eqs. (2.62) and (2.68) to see that tailoring the beta prime kernel to the gamma target PDF implies that

$$\Delta\kappa_2 = 0 \Rightarrow \psi^{(1)}(a_1) + \psi^{(1)}(a_2) = \psi^{(1)}(L), \quad (5.11)$$

$$\Delta\kappa_3 = 0 \Rightarrow \psi^{(2)}(a_1) - \psi^{(2)}(a_2) = \psi^{(2)}(L). \quad (5.12)$$

Clearly, $a_1 = L$ and $\psi^{(1)}(a_2) = \psi^{(2)}(a_2) = 0$. According to [Abramowitz and Stegun, 1964], $\psi^{(1)}(a_2)$ has only one positive real root, at $a_2 \approx 1.46$. It is easily verified that this value is *not* a root of $\psi^{(2)}(a_2)$. However, [Abramowitz and Stegun, 1964] also contains asymptotic

formulas which show that $\psi^{(1)}(a_2) \rightarrow 0$ and $\psi^{(2)}(a_2) \rightarrow 0$ as $a_2 \rightarrow \infty$. Similarly, we can use the log-cumulants of the inverse gamma distribution from eq. (5.4) to see that

$$\Delta\kappa_2 = 0 \Rightarrow \psi^{(1)}(a_1) + \psi^{(1)}(a_2) = \psi^{(1)}(L) \quad (5.13)$$

$$\Delta\kappa_3 = 0 \Rightarrow \psi^{(2)}(a_1) - \psi^{(2)}(a_2) = -\psi^{(2)}(L), \quad (5.14)$$

and by the same argument as before, $a_2 = L$, while $a_1 \rightarrow \infty$. We tried a few numerical methods, and all of them correctly found $a_1 = L$ for the gamma target PDF and $a_2 = L$ for the inverse gamma target PDF. All of the methods produced very high values for the other shape parameter (as expected, since the true value was ∞), with the level of accuracy determining exactly how high. These values invariably caused problems within the gamma functions in the beta prime kernel PDF in eq. (2.66). Specifically, $\Gamma(a_1 + a_2)$ was evaluated as infinity, but this was remedied with the usual tricks.¹¹ Then, the MKBK kernel and series actually modeled the gamma and inverse gamma distributions exactly, theoretically explained in e.g. [McDonald, 1995] by the fact that $\beta'(\cdot)$ is reduced to $\gamma(\cdot)$ or $\gamma^{-1}(\cdot)$ when $a_2 \rightarrow \infty$ or $a_1 \rightarrow \infty$, respectively. To summarize, the MKBK kernel and series exactly modeled gamma and inverse gamma distributions, but required a careful implementation to do so.

From the $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$ distributions we can isolate the effect of varying only the location m or the shape L . If the latter is kept constant while m is varied, the results are unchanged except for the MKE series with $L = 4$. All other parameters equal, we see that the approximations are generally more accurate for $L = 16$ than for $L = 4$ – in many cases the difference is several orders of magnitude.

The K distributions with $\boldsymbol{\theta} = [4, 10, 1]$ and $\boldsymbol{\theta} = [16, 10, 1]$ are different from the rest as they are nonzero at $x = 0$. As we can see, these cases are well modeled by the beta prime distribution in particular, but the MKBK series failed to improve on its kernel. In general, the MKBK series seems like a good candidate to model the K distribution, perhaps since both distributions have two shape parameters.

For the G^0 distribution, the parameter g acts in the same way as the location, i.e. the methods are, with a single exception, identical for $g = 1$ and $g = 10$.

The MKGK series improved on its kernel in 9 of the 16 cases where it did not exactly model the target PDF, the MKBK series improved on its kernel in 5 of 8 cases, but the MKLK and MKE series improved on the log-normal kernel in all 20 cases. It is still too early to draw any conclusions on the matter, but this might be due to the fact that the log-cumulants of order $n \geq 3$ of the log-normal distribution are all zero. Recall that the resulting simplification is that the log-cumulant differences equal the target PDF's log-cumulants.

5.2.3 Convergence, Two Examples

We will now examine if and how the MK series expansions converge as the number of correcting terms are increased. In [Blinnikov and Moessner, 1998], the authors performed a similar analysis of the classical Edgeworth and Gram-Charlier series expansion of the Gaussian kernel.

We start with two visual examples, that is, we plot the PDFs as we correct for each successive log-cumulant difference. We limit ourselves to a single method applied to two different target PDFs, simply to demonstrate how convergence and divergence *looks*, with a much more thorough and broader approach in the next section.

¹¹We had to evaluate the logarithm of the beta prime PDF without the scaling beta function, apply the exponential function and finally normalize the PDF by its numerical integral.

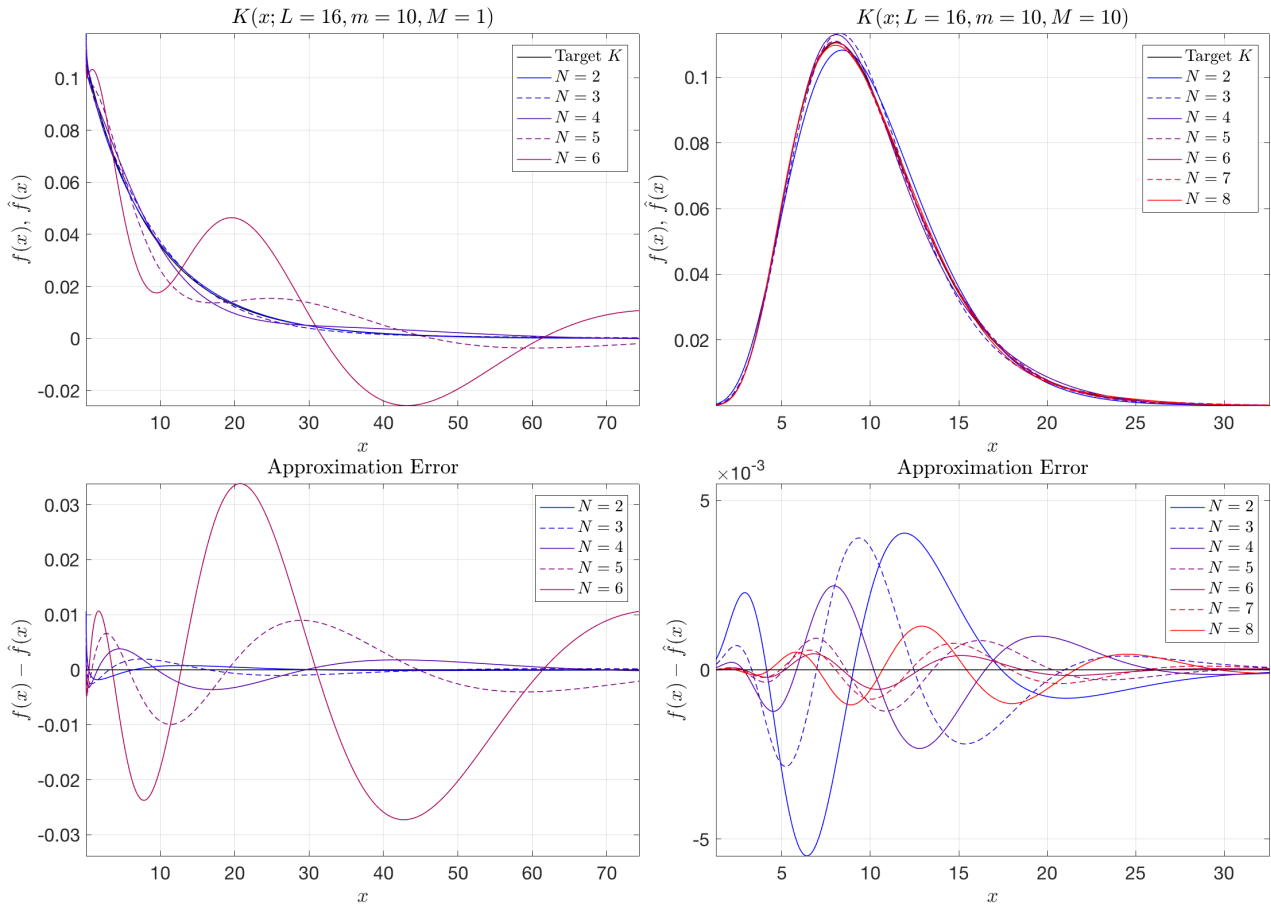


Figure 5.3: The MKGK series expansions used to approximate two K target distributions, correcting the tailored kernel with up to $N = 6$ (left) and $N = 8$ (right) in eq. (3.39). Target PDF $f(x)$ and MKGK series approximations $\hat{f}(x)$ (top), errors $f(x) - \hat{f}(x)$ (bottom).

The cases are presented in Figure 5.3, where the plots are colored on a scale from blue (for the kernel) to red ($N = 8$), with even N values given solid lines and odd N given dashed lines for clarity.

When looking at all the PDF approximations at once, it is hard to draw conclusions beyond the fact that the MKGK series expansion diverges in the case where $M = 1$ and that it models the K distribution with $M = 10$ very well. Therefore, we also included plots of the approximation errors, which gives us some more insight. The errors oscillate around 0 in a periodic manner, and an interesting observation is that the extrema of these oscillations are shifted on the x axis as the number of correcting terms is increased. Naturally, when the MKGK series expansion converges to the target PDF, the oscillations are also damped, and when it diverges they are amplified. As we will see in the next section, the MKGK diverges from the first correcting term when approximating $K(x; L = 16, m = 1, M = 1)$ and starting at $N = 7$ when approximating $K(x; L = 16, m = 1, M = 10)$. Upon close inspection of Figure 5.3, it is also possible to see this visually.

5.2.4 Convergence, Numerous Target PDFs

Now it is time to conduct a thorough investigation of the convergence of all the series expansions, and for a wide range of target distributions. As the targets, we use the same 20 PDFs as in Section 5.1, with a special emphasis on four of them, which we present first.

Recall that throughout Chapter 3 we suggested choosing the kernel parameters that satisfy

$\Delta\kappa_1 = \Delta\kappa_2 = 0$ (we assess other methods of choosing the kernel parameters later). That is to say that the baseline for our experiment is to correct for the first two log-cumulant differences. The tailored beta prime kernel even corrects for $\Delta\kappa_3$, i.e. the MKBK series starts to modify its kernel one term later than the other series.

In Figure 5.4 we present plots of the two dissimilarity measures discussed in Section 2.6, and also the maximum absolute error, for the MK series expansions used to approximate four different target PDFs. The γ^{-1} and G^0 distributions are special cases of the beta prime kernel, so the MKBK series is omitted in those instances. We clearly see that the MKE and MKLK series are identical when correcting only for $\Delta\kappa_3$, which we already knew since eq. (3.41) is identical to eq. (3.82) when discarding terms of $O(r^{-1})$.

As in the previous section, the MKE series is the stand-out performer. In Figure 5.4 it has the remarkable property that the approximation is improved with each successive term, for all targets and dissimilarity measures. The MKGK series seemingly converges for the $\gamma^{-1}(\cdot)$ case, but for the other targets it diverges beyond $\Delta\kappa_5$, $\Delta\kappa_6$, or even with the very first correcting term. The MKBK series is only presented for two of the cases, diverging in one and converging in the other. Indeed, it seems like this GFD is very challenging for all but the MKE series. In general, the MKLK series is more stable than the others, neither converging nor diverging with particular conviction.

The dissimilarity measures are largely in agreement, barring a few subtle differences. This allows us to only present one of them in the following, and we see that the Kullback-Leibler distance has the best contrast when the distance between the series expansions and the target PDF becomes very small.

In Figure 5.5 we present the convergence of rest of the cases from Table 5.1, i.e. the 16 cases which were not included in Figure 5.4.

We can make many of the same observations as in Section 5.2.2: The series expansions are seemingly invariant to the location parameter m in the $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$, and the same can be said for the g parameter in the G^0 distribution. All series generally perform better for $L = 16$ than $L = 4$, and this extends to a faster and more regular convergence, e.g. the GFDs with $m = 10$ and $\nu = 2$.

Perhaps the biggest difference from Figure 5.4 is that Figure 5.5 contains examples of the MKE series not improving with each successive correction term. Still, the general trend is that the MKE series' advantage over the other methods increases with each extra term, providing the best approximation for all but one target PDF when correcting for up to $\Delta\kappa_5$. However, it is important to note that the MKE series is more complex than the other series. That is, correcting for up to $\Delta\kappa_5$, we know from Section 3.3.5 that the MKLK series is

$$f_X(x) \approx \left[1 + \frac{\kappa_3}{6\sigma^3} H_3\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_4}{24\sigma^4} H_4\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_5}{120\sigma^5} H_5\left(\frac{\log x - \mu}{\sigma}\right) \right] \Lambda(x; \mu, \sigma), \quad (5.15)$$

and from Section 3.4.2 we know that the corresponding MKE series is

$$f_X(x) \approx \left[1 + \frac{\kappa_3}{6\sigma^3} H_3\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_3^2}{72\sigma^6} H_6\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_4}{24\sigma^4} H_4\left(\frac{\log x - \mu}{\sigma}\right) \right. \\ \left. + \frac{\kappa_3^3}{1296\sigma^9} H_9\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_3\kappa_4}{144\sigma^7} H_7\left(\frac{\log x - \mu}{\sigma}\right) + \frac{\kappa_5}{120\sigma^5} H_5\left(\frac{\log x - \mu}{\sigma}\right) \right] \Lambda(x; \mu, \sigma), \quad (5.16)$$

where we have arranged the terms in the order they arise when evaluating the successive Bell

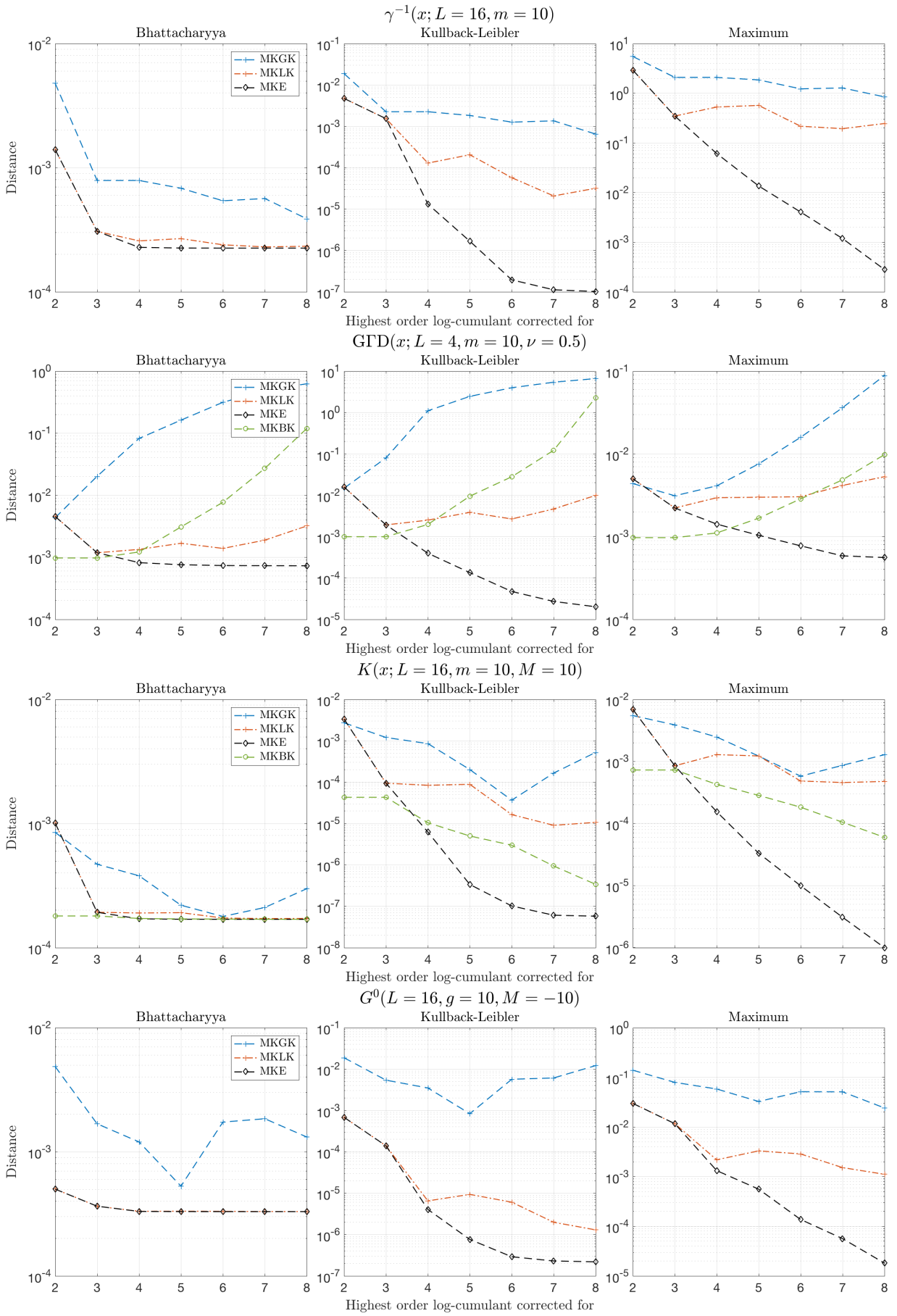


Figure 5.4: The four MK series expansions used to approximate four different target distributions, correcting for up to $\Delta\kappa_8$.

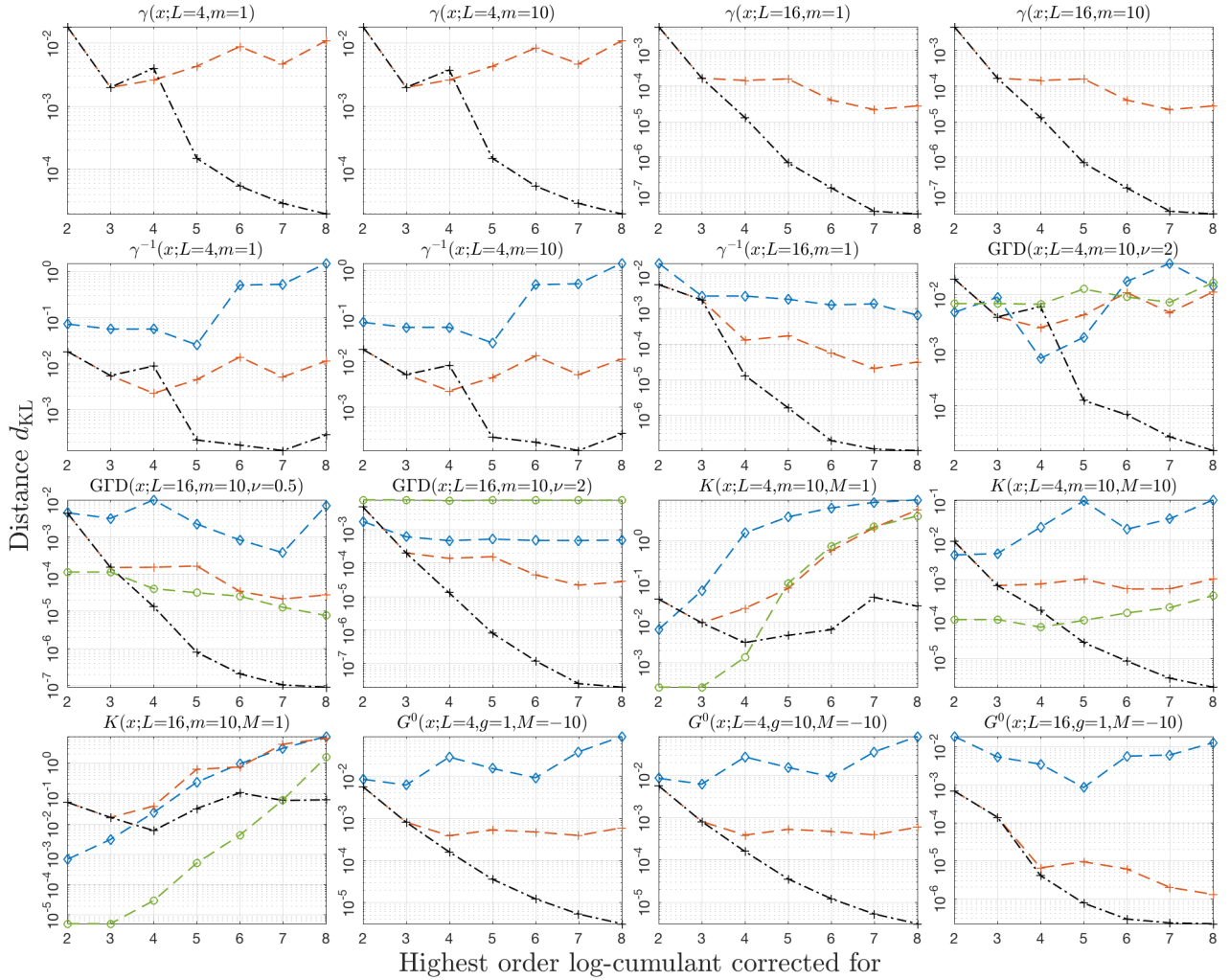


Figure 5.5: The MK series expansions used to approximate the target PDFs from Table 5.1, excluding those included in Figure 5.4. As in that figure, the MKGK series is blue, the MKLK series is orange, the MKE series is black, and the MKBK series is green. The dissimilarity measure used is the Kullback-Leibler distance.

polynomials in eq. (3.81).¹² We see that the MKE series contains all three correcting terms of the MKLK series, but also three additional terms, correcting for κ_3^2 , $\kappa_3\kappa_4$, and κ_3^3 . In this scenario the targets PDFs are known, so there is no uncertainty in the log-cumulants, and this additional complexity is not punished. Indeed, based on the performance we have seen in this section we can say that the extra terms are definitely merited.¹³

5.2.5 Alternative Choices of the Kernel Parameters

The time has come to assess the validity of the method of choosing the kernel parameters used throughout Chapter 3. In Section 3.2.4 we argued that the nature of the Bell polynomials results in significant simplifications if we choose the kernel parameters s.t. $\Delta\kappa_1 = \Delta\kappa_2 = 0$ (in

¹²Recall that $\kappa_{\Delta,n} = 0$ for all $n \geq 3$, which allowed us to simply use κ_n without ambiguity when expanding the log-normal kernel.

¹³In Section 5.3, we estimate the log-cumulants with an incurred uncertainty in each correcting term. Based on the findings in Table 5.1 and Figures 5.4 and 5.5, it was very interesting to see whether the additional complexity of the MK series was still merited. As a spoiler of sorts, we can mention that when there was enough data points to justify a correction for $\Delta\kappa_4$, the MKE series outperformed the MKLK in almost all the cases we examined.

practice using the MoLC), and we called these the *tailored* parameter values.¹⁴ This choice naturally cannot be based on convenience (in the sense of mathematical simplicity) alone, and we will now test whether our choice is associated with an advantage in accuracy as well.

In the classical case, alternative choices of the kernel parameters have not received much attention. This can in large part be explained by the fact that the Gaussian has almost exclusively been the kernel of choice, and the ML estimates of the mean and variance are in fact equal to the method of moments estimates which give $\Delta_{c_1} = \Delta_{c_2} = 0$.¹⁵ A correction for the bias of the ML variance estimate of the Gaussian distribution has been known since [Gauss, 1823], i.e. even before the development of the Gram-Charlier and Edgeworth series. However, the use of unbiased variance estimates in the series expansions is not common. [Kendall et al., 1994] do not comment on the particular case of the biased variance estimate, but the authors write (in a very general context):

“[We] have retained the terms in D and D^2 because the approximation may perhaps be slightly improved by taking m and $[\zeta^2]$ in the [kernel] distribution not quite equal to the mean and variance of [the target distribution].”

To summarize, the lack of attention to alternative parameter choices in the classical case can reasonably be attributed to the popularity of the Gaussian kernel. Its ML and method of moments parameter estimators overlap, and with the exception of a scaling factor $\frac{n}{n-1}$, they are also the minimum variance unbiased estimators. In addition to this, $\Delta_{c_1} = \Delta_{c_2} = 0$ is required in the classical Edgeworth series, just as $\Delta_{\kappa_1} = \Delta_{\kappa_2} = 0$ is required to develop the MKE series.

Regarding the classical Gram-Charlier series around the gamma kernel, the authors of [Gaztanaga et al., 2000] did not discuss alternatives to the parameter choice in eq. (2.129).

In Figures 5.6 through 5.10 we display the effects (on the Kullback-Leibler distance to the target PDF) of varying each kernel parameter separately. We display relative parameter values, where the reference is the tailored value at unity.

Regarding the cases which were "exact" in Table 5.1, we can now compare the MKGK series to the gamma target PDF in Figure 5.6, as it is only exact for the tailored parameter values.¹⁶ The MKBK series, however, is included for the G^0 distribution but omitted for the $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$ PDFs. For the latter two cases, one of the beta prime shape parameters will tend to ∞ , as we discussed in Section 5.2.2. While it could be possible to vary the finite shape parameter, the MKBK series breaks down as the beta prime kernel is reduced to the gamma or inverse gamma PDF. This is easy to see from the Rodrigues formula of the $M'(\cdot)$ polynomials in eq. (3.88), which obviously assumes that both a_1 and a_2 are finite.

For the MKGK, MKLK, and MKBK series, non-tailored parameter values imply that the correction terms start already with Δ_{κ_1} , i.e. $n = 1$ in eqs. (3.31), (3.60), and (3.102). That is,

¹⁴It is perhaps clearer why one would consider alternative choices of the kernel parameters when the target PDF is unknown. Then, we have to estimate the parameters, and the MoLC estimator is not usually the same as e.g. the ML estimator, and it might have bias which we could want to correct for. This will be discussed more in detail when we repeat this analysis with unknown target PDFs.

¹⁵The ML estimates of the Gaussian distribution is found in many (if not most) statistics textbooks, e.g. [Kendall et al., 1994]. Recall from Section 2.2.5 that $c_1 = m$ and $c_2 = \zeta^2$, which we replace with their empirical counterparts as appropriate. The sample mean is the *minimum variance unbiased* estimator for m . That is, $\langle m \rangle = \hat{m}_{\text{ML}}$. The sample variance $\langle \zeta^2 \rangle$ is actually biased, but scaling it with $\frac{n}{n-1}$, where n is the number of samples, gives the minimum variance unbiased estimator.

¹⁶As we can see, $d_{\text{KL}}(\cdot)$ is not zero as it theoretically should be, but between 10^{-8} and 10^{-7} . This is only due to numerical inaccuracies in the implementation, and thus serves as an ultimate benchmark for non-exact methods with the present precision.

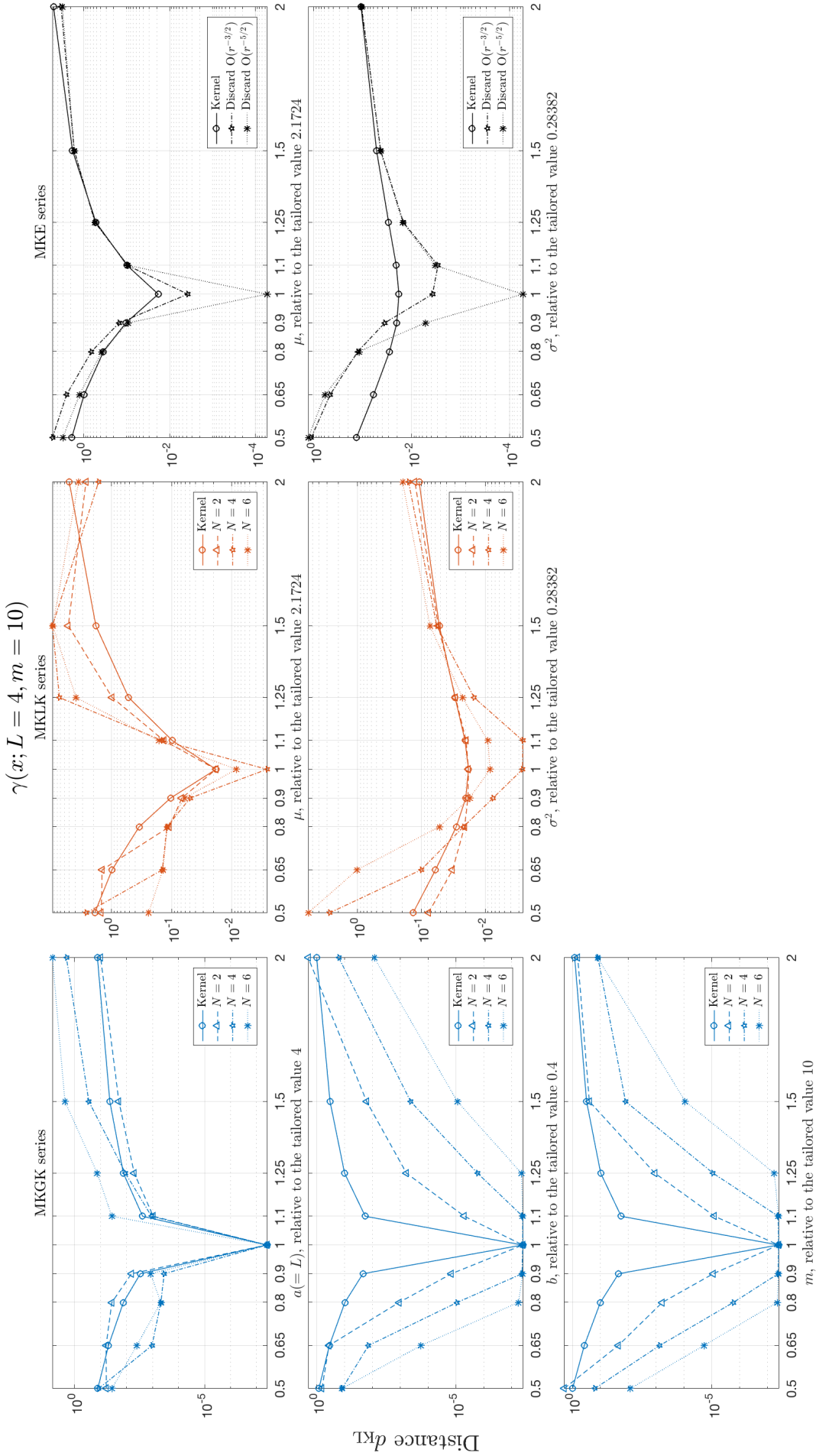


Figure 5.6: The MKGK, MKLK and MKE series expansions used to approximate a gamma distribution with parameters $L = 4$ and $m = 10$. The kernel parameters are varied between 0.5 and 2 times the value they have when the kernel is tailored (formally when $\Delta\kappa_1 = \Delta\kappa_2 = 0$). The error measure is the Kullback-Leibler distance to the target PDF.

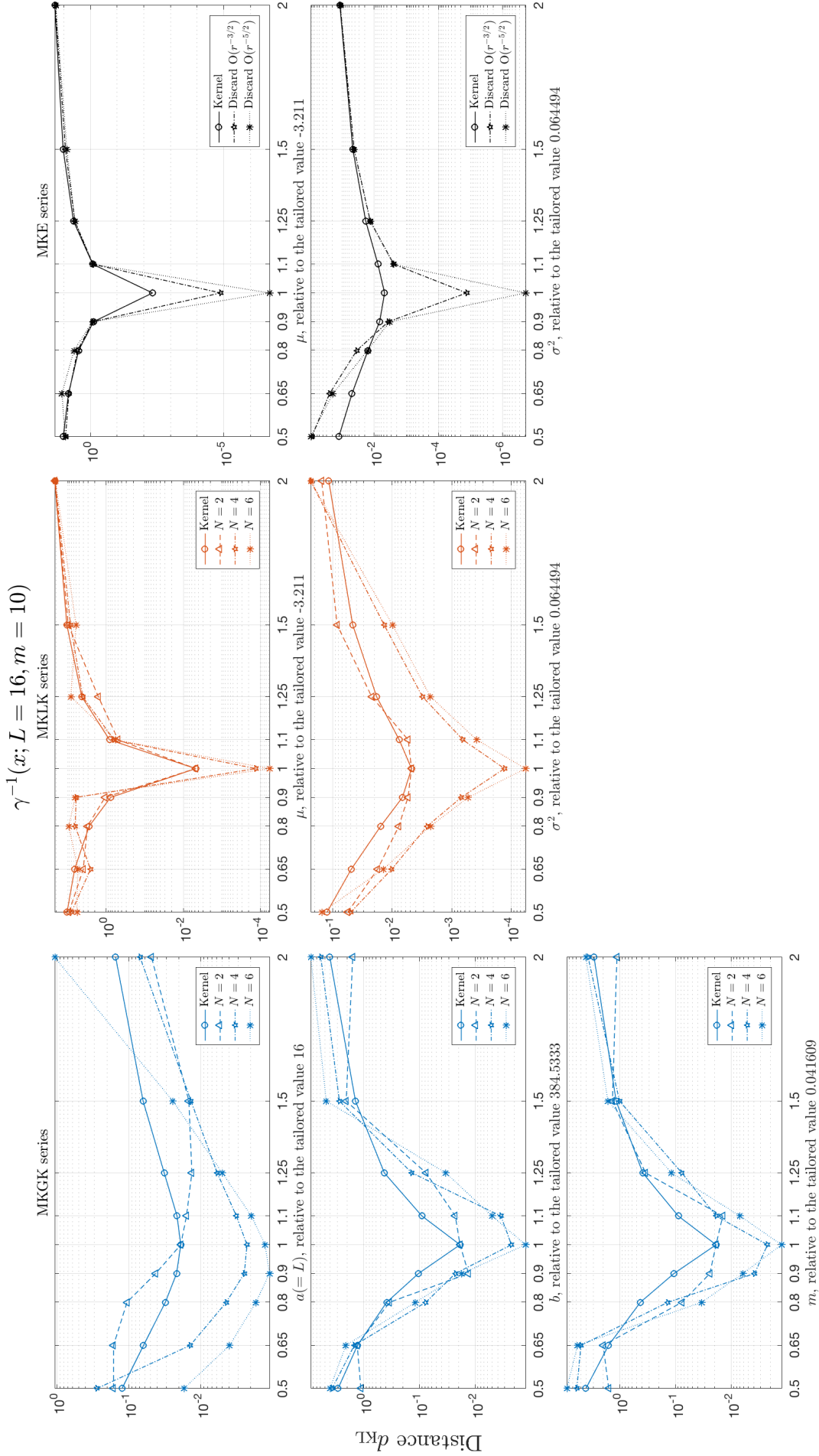


Figure 5.7: The MKGK, MKLK and MKE series expansions used to approximate a inverse gamma distribution with parameters $L = 16$ and $m = 10$. The kernel parameters are varied between 0.5 and 2 times the value they have when the kernel is tailored (formally when $\Delta\kappa_1 = \Delta\kappa_2 = 0$). The error measure is the Kullback-Leibler distance to the target PDF.

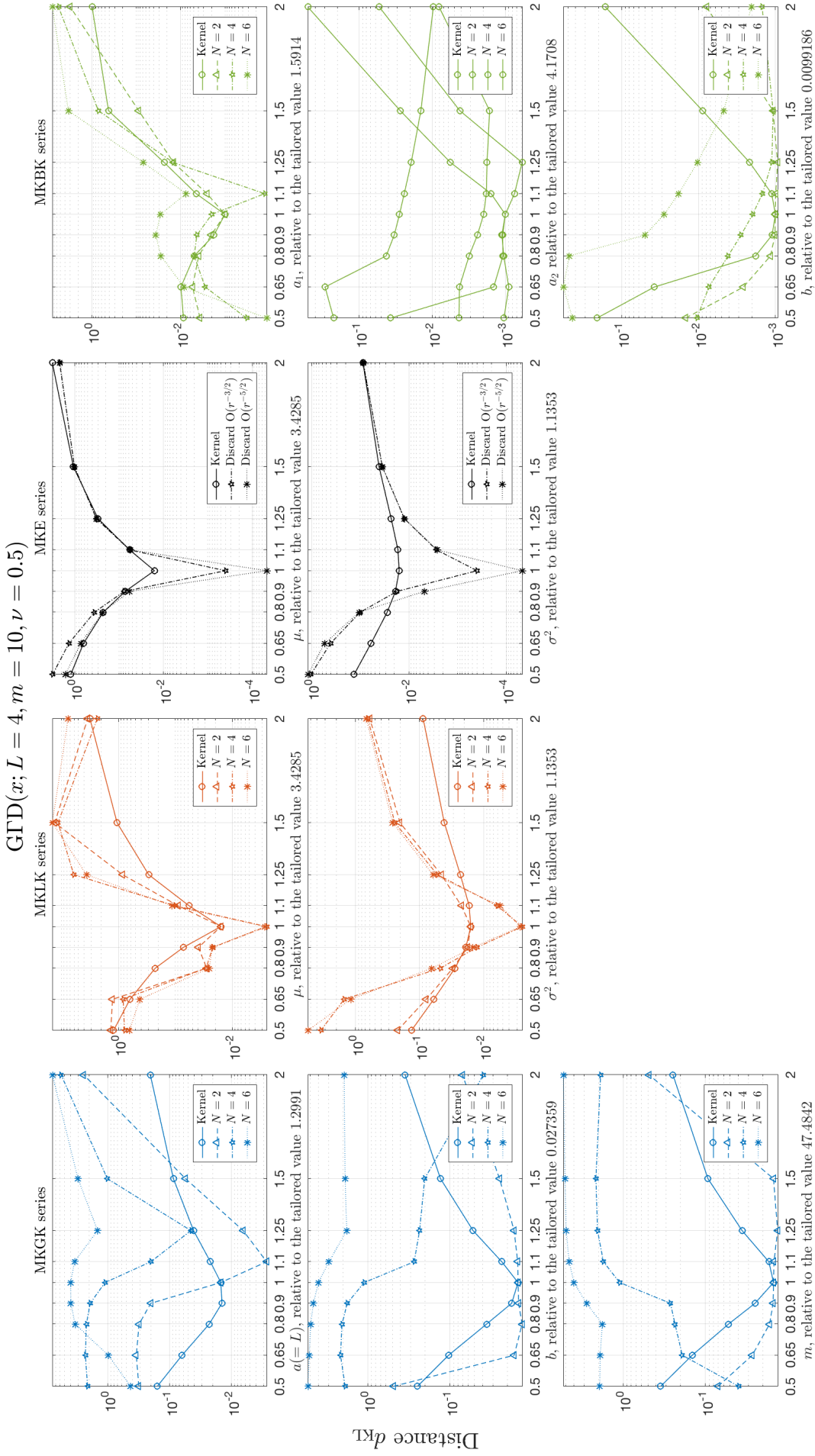


Figure 5.8: The MKGK, MKLK, MKE and MKBK series expansions used to approximate a GFD with parameters $L = 4$, $m = 10$, and $\nu = 0.5$. The kernel parameters are varied between 0.5 and 2 times the value they have when the kernel is tailored (formally when $\Delta\kappa_1 = \Delta\kappa_2 = 0$, and also $\Delta\kappa_3 = 0$ for the MKBK series). The error measure is the Kullback-Leibler distance to the target PDF.

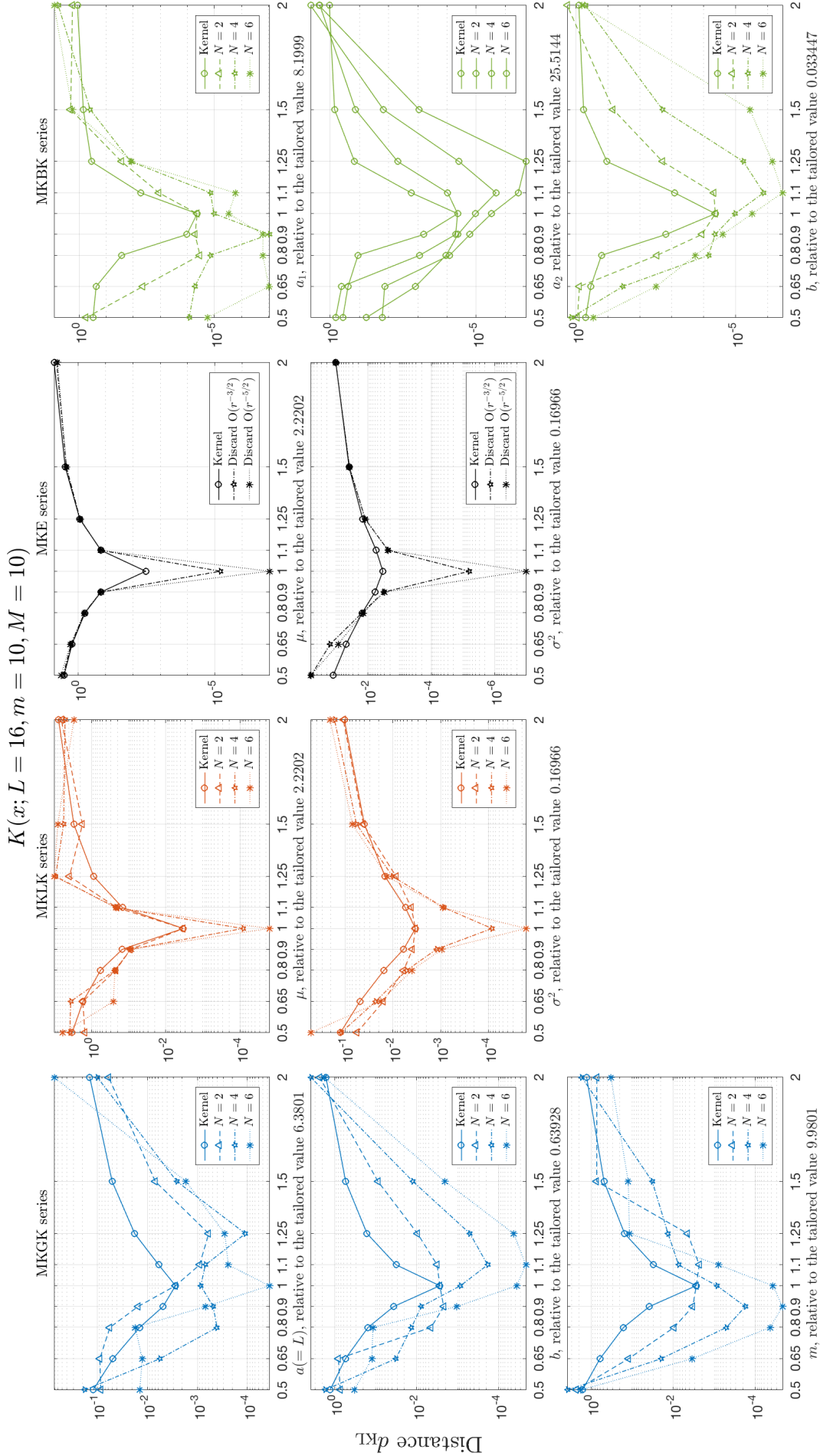


Figure 5.9: The MKGGK, MKLK, MKE and MKBK series expansions used to approximate a K distribution with parameters $L = 16$, $m = 10$, and $M = 10$. The kernel parameters are varied between 0.5 and 2 times the value they have when the kernel is tailored (formally when $\Delta\kappa_1 = \Delta\kappa_2 = 0$, and also $\Delta\kappa_3 = 0$, and also $\Delta\kappa_3 = 0$ for the MKBK series). The error measure is the Kullback-Leibler distance to the target PDF.

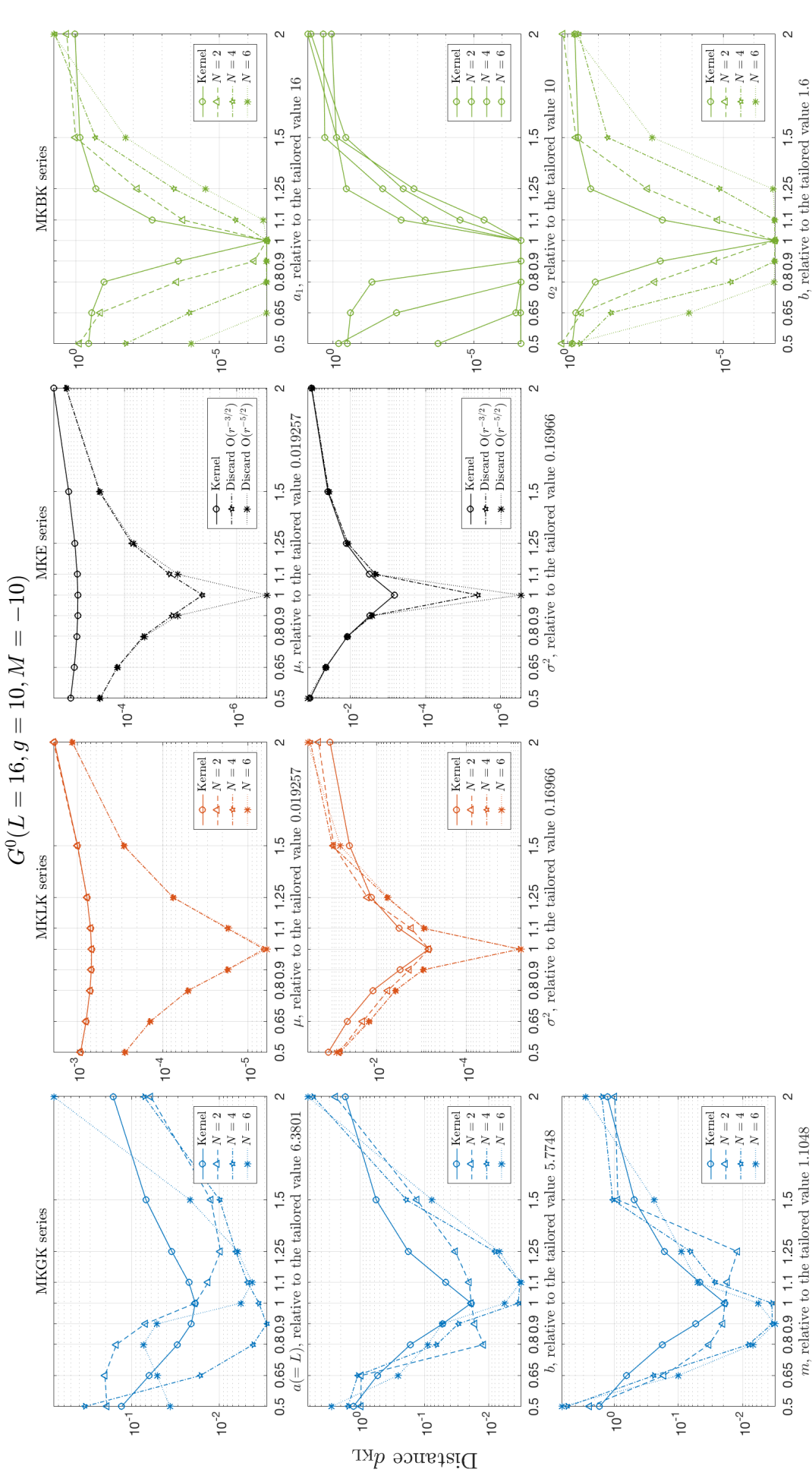


Figure 5.10: The MKGK, MKLK, MKE and MKBK series expansions used to approximate a G^0 distribution with parameters $L = 16, g = 10$, and $M = -10$. The kernel parameters are varied between 0.5 and 2 times the value they have when the kernel is tailored (formally when $\Delta\kappa_1 = \Delta\kappa_2 = 0$, and also $\Delta\kappa_3 = 0$ for the MKBK series). The error measure is the Kullback-Leibler distance to the target PDF.

$N = 2$ now corresponds to

$$f_{X,N}(x) = \left[1 + \Delta\kappa_1 M_1(bx) + \frac{\Delta\kappa_1^2 + \Delta\kappa_2}{2} M_2(bx) \right] \gamma(x; a, b), \quad (5.17)$$

for the MKGK series, and the corresponding versions when the log-normal or beta prime kernel is used instead. The MKE series however, is fundamentally based on the assumption that the first two log-cumulant differences are zero, as we discussed in Section 3.4.1. There is no easy way to amend eq. (3.81) in order to correct for non-zero $\Delta\kappa_1$ or $\Delta\kappa_2$, so we do not attempt to. That is, we simply change the kernel parameters without making further changes to the series expansions. This is a significant difference in approach to the other series, and we must keep this in mind when analyzing the results.

Gamma Distribution Starting with Figure 5.6, it is interesting to observe that the MKGK series seems very sensitive to changes in the shape a ; its accuracy at 1.1 of the tailored value is of the same magnitude (10^{-2}) as when the target PDF is inverse gamma in Figure 5.7. For the scale b , however, we observe something interesting (recall that the $\mu = L/b$, and since L is constant when b and μ are varied, they are each others inverse). When increasing the number of correcting terms N , the MKGK series becomes remarkably resilient to fluctuations in b and μ , achieving accuracies comparable to the exact solution at below 10^{-7} even when b is at 0.8 or 1.25 relative to the tailored value. This is encouraging w.r.t. modeling of SAR data: As we discussed in Section 2.3.9, the number of looks L can often be estimated based on the entire image, i.e. very precisely. Perhaps the MKGK series can then correct for inaccuracies in the location estimates – we return to this subject later.

The other methods in Figure 5.6, the MKLK and MKE series, are more sensitive to changes in the log-mean μ than the log-variance σ^2 . This is a trend we recognize with the other target PDFs also, especially the K and G^0 distributions. Another trend which is exemplified with the $\gamma(\cdot)$ target PDF is the fact that when the kernel parameters are not tailored, the benefit of correcting that kernel is lost, or even negative in the form of increased distances to the target compared to the bare kernel. This is apparent for the MKLK series with μ greater than its tailored value and the MKE series with σ^2 below its tailored value in Figure 5.6, but also in several other cases. With μ at double its tailored value, the MKLK series again seemingly outperforms its kernel, but these are errors well above $10^0 = 1$.¹⁷ When the errors are of this magnitude, we should be careful when drawing conclusions – saying that a method with an error of 4 is better than one with an error of 6 is a bit pointless when the fact of the matter is that both those methods have diverged severely and thus failed.

Inverse Gamma Distribution The story told by Figure 5.7 is not very different from the experiment with the gamma distribution, except that the MKGK series is not exact and thus behaves differently. It is less sensitive to variations in the shape parameter, even achieving slightly higher accuracies for non-tailored values, depending on the number of correcting terms N . In general, the MKGK series expansion improves on its kernel when modeling $\gamma^{-1}(\cdot)$.

Generalized Gamma Distribution Moving on to the three-parameter target PDFs, we recall from Table 5.1 and Figure 5.2 that this parametrization of the GFD was challenging to model, a fact which is reflected in Figure 5.8. The MKLK and MKE series behave similarly as for the $\gamma(\cdot)$ and $\gamma^{-1}(\cdot)$ targets, while the MKGK and MKBK series seem chaotic. There is, however, insight to be had from a closer inspection. Clearly, $N = 6$ for the MKGK series was too many correcting terms while $N = 2$ seems like the best choice. However, non-tailored parameter values

¹⁷These are divergent cases, e.g. the MKGK series in Figure 5.2 was visually far from its target, and was tabulated in Table 5.1 at $1.10 \cdot 10^0$.

greatly improve the approximations for both $N = 2$ and $N = 4$. Indeed, with L at 1.1 times its tailored value, the MKGK series with $N = 2$ is comparable to the best approximation of the MKLK series, although the latter peaks in performance with its tailored parameters and for $N = 6$. Since the MKGK is improved both for higher L and higher μ , it is interesting to test whether there are further improvements to be had if both parameters are higher than their tailored value simultaneously. In this case, further testing determined that there is a slight benefit.¹⁸ The MKBK series experiences something remarkable, for $N = 4$ and a_1 at 1.1 times its tailored value, the distance to the target is reduced by more than an order of magnitude to $1.27 \cdot 10^{-4}$, which would have made it the top performer in that case w.r.t. Table 5.1. As with the MKGK series, further testing showed that the distance can be slightly improved still by altering the parameters simultaneously.

K Distribution As we commented already in Section 5.2.1, the K distribution with the parametrization $L = 16$, $m = 10$, and $M = 10$ is seemingly easier for the series expansions to approximate than the above GFD. This observation is reinforced when examining Figure 5.9, since the general theme is that the accuracy is increased as N is incremented. As before, the MKLK and MKE series are far more sensitive to changes in the log-mean than the log-variance parameter and achieve their smallest distance to the target PDF for the tailored values. They again lose most, if not all, of the benefit of correcting the kernels when the parameters are far from their tailored values. Further testing strongly suggests that simultaneously varying both μ and σ^2 does not improve either of the series expansions of the log-normal kernel.

The MKGK again performs better without non-tailored kernels, competing with the accuracy of the MKLK series, which is certainly in contrast to what one would believe after reading Table 5.1. Allowing the kernel scale and shape to deviate from their tailored values simultaneously further improved the MKGK series again. For the MKBK series with $N = 4$, scaling the tailored a_1 with 0.9 reduced the distance from $1.04 \cdot 10^{-5}$ to $9.37 \cdot 10^{-8}$, which is an extreme improvement. This is also much better than the corresponding MKE series, which was the top performer in that case in Table 5.1. A similar distance was achieved with $N = 6$, again with a_1 lower than its tailored value. It was slightly better than the MKE series corrected for the up to the same log-cumulant order, but we must be careful about drawing conclusions, as the effects of our discretization of x and other aspects of our implementation cannot be disregarded at this level of accuracy.

G^0 Distribution Finally, we have the G^0 target distribution in Figure 5.10. In terms of the MKGK, MKLK and MKE series, there is not much to add to the already extensive analysis. The MKGK is slightly better for non-tailored parameter values, the MKLK and MKE series are more sensitive to variations in the log-variance than the log-mean, and there is no benefit to correcting the non-tailored log-normal kernels. The fact that the MKLK and MKE series so consistently peak at their tailored values can perhaps be linked to the observation that they always improved on their kernels in Table 5.1. Recall that in Section 5.2.2 we speculated that this consistency could be due to the fact that the log-cumulant differences of order $n \geq 3$ are reduced to the target log-cumulants.

The MKBK series is now exact, and is thus interesting as a comparison to when the MKGK series modeled the gamma distribution in Figure 5.6. The story is much the same, the MKGK series is extremely successful in correcting for non-exact values of any of the three parameters.

¹⁸Examining this exhaustively is beyond the scope of this thesis, but possibly a subject for future work.

5.2.6 Alternative Kernel Parameter Values in a SAR Scenario

Continuing our investigation on alternative parameter values from the previous section, we now test the mathematical feasibility of a radically different reason for having non-tailored parameters; Can the series expansions be used to correct for the effects of a physical process?

Specifically, we concentrate on a SAR scenario. Expanding on the discussion about the number of looks L and the K distribution from Sections 2.3.9, 2.3.10, and 5.1.2, say we have a SAR image of a scene containing different classes. As we described when introducing the gamma distribution, L is for SAR images a global value which depends on the processing scheme in the image formation and represents the level of averaging or smoothing. The mean radar intensity m is a local property of the surface, quantifying the mean level of backscatter. If we assume completely homogeneous classes, i.e. that every pixel corresponding to e.g. "snow" will reflect the same amount of the radar pulse on average (in the direction of the sensor), then the gamma distribution model is valid. However, this idealistic assumption is usually quite naive, as classes like "forest" and "ocean" will certainly contain variations in the local radar reflectivity. This heterogeneity leads to a multiplicative model, represented in the K distribution by the *texture* shape parameter M . For more information on this subject, see e.g. [Oliver and Quegan, 2004].

In order to account for heterogenous classes, the traditional approach has been to depart from $\gamma(x; L, m)$ and use $K(x; L, m, M)$ instead. The question now is, can we keep using the simpler gamma PDF, but employ the MKGK series expansions as a way of correcting for the physical variations usually modeled by M ? To conclude that this is possible, we would naturally have to conduct a thorough examination on a variety of real-world data. Here, we must limit our ambitions to simply seeing if it is at all feasible, by using non-tailored parameters values in a comparison similar to the one in Figure 5.3.

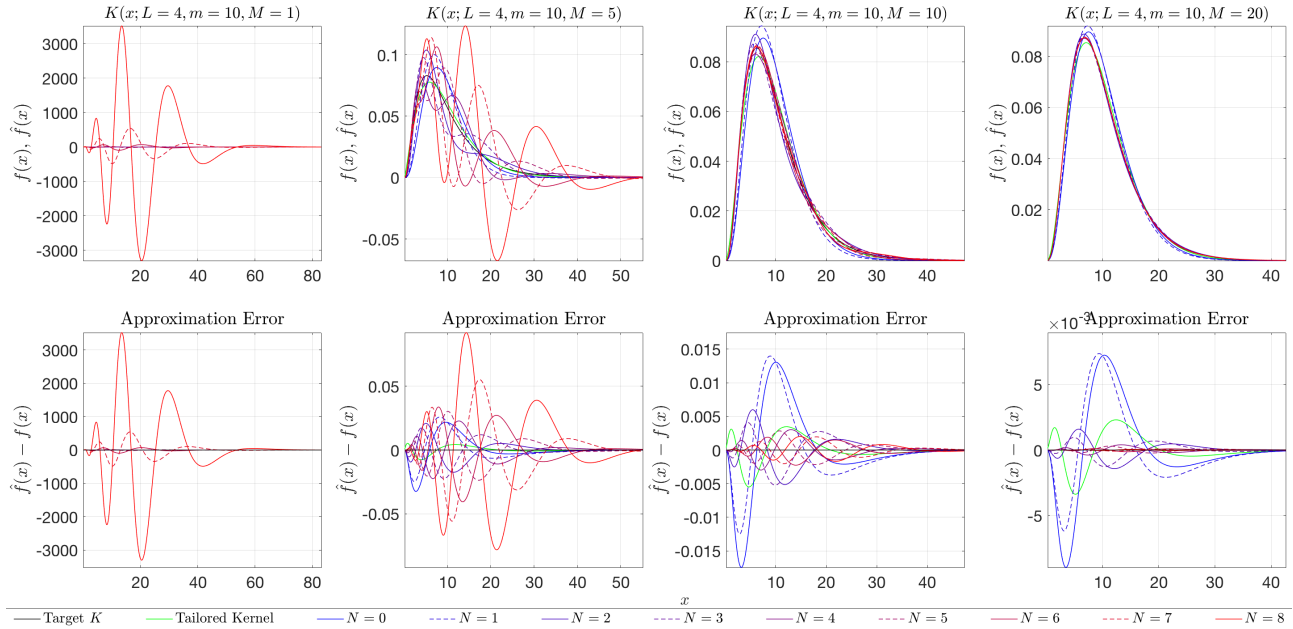


Figure 5.11: The MKGK series expansion around a non-tailored kernel with $N \leq 8$ in eq. (3.31), used to approximate four K target distributions. The kernel used was in all cases $\gamma(x; L = 4, m = 10)$, i.e. the same parameter values as the target PDF instead of those giving $\Delta\kappa_1 = \Delta\kappa_2 = 0$. Target PDF $f(x)$ and MKGK series approximations $\hat{f}(x)$ (top), errors $\hat{f}(x) - f(x)$ (bottom). The figure legend is placed below the plots.

Figure 5.11 contains the results of this mathematical experiment. From looking at the plots, we see that approach has failed for $M = 1$ and $M = 5$ but succeeded for $M = 10$ and $M = 20$. In

the latter two cases, the benefit of correcting for the non-zero $\Delta\kappa_1$ and $\Delta\kappa_2$ is great, and when $M = 20$ the MKGK series further converges as N increases further. For $M = 1$, the method is violently divergent, a scenario which is quite common also in the linear case, see [Blinnikov and Moessner, 1998]. When $M = 5$, it seems like the MKGK series is close to converging for lower values of N , but in the end fails.

Why this dependency on M ? It turns out there is both a physical and a mathematical explanation.

In [Frery et al., 1997], the authors discussed heterogeneous and homogeneous targets in depth, and they noted that $M \rightarrow \infty$ indeed corresponds to the special case of a homogeneous target, i.e. it is reduced to the gamma distribution. For the mathematical explanation, we refer to the expression of the log-cumulant differences

$$\Delta\kappa_n = \kappa_{K,n} - \kappa_{\gamma,n}, \quad (5.18)$$

with the constituent log-cumulants given by eqs. (5.8) and (2.62), we get

$$\Delta\kappa_n = \begin{cases} \psi^{(0)}(M) - \log M & n = 1, \\ \psi^{(n-1)}(M) & n \geq 2. \end{cases} \quad (5.19)$$

As mentioned when we discussed the cases $n = 2, 3$ in Section 5.2.2, the asymptotic formulas for $\psi^{(n-1)}(M)$ are found in [Abramowitz and Stegun, 1964]. We can see from these formulas that as $M \rightarrow \infty$, $\Delta\kappa_n \rightarrow 0$ for all n .¹⁹ Naturally, this is equivalent with the statement that $K(\cdot)$ is reduced to $\gamma(\cdot)$ as $M \rightarrow \infty$, as the set of all log-cumulants uniquely define a distribution, i.e. if $\Delta\kappa_n = 0 \forall n$, then the PDFs are the same.

To summarize, Figure 5.11 indicates that if the area is *slightly* heterogeneous, the MKGK series can provide a good alternative when it comes to modeling the PDF, with the greatest benefit occurring already at $N = 2$. In the rightmost plots, we can see that retaining the values of L and m from the K PDF and correcting for $\Delta\kappa_1$ and $\Delta\kappa_2$ is actually *more* accurate than the tailored kernel. This warrants additional testing and will be revisited later in this thesis, when the target PDFs are unknown.

5.3 Estimation

In this section, we remove the knowledge of the target PDFs from the series expansions. For instance, say we want to apply the MKGK series expansion to a $K(\cdot)$ target. When the target PDF was known, the log-cumulant differences in eq. (3.39) were $\Delta\kappa_n = \kappa_{K,n} - \kappa_{\gamma,n}$. Now, however, $\kappa_{K,n}$ must be estimated, i.e. replaced by the empirical log-cumulants $\langle \kappa_{X,n} \rangle$ via eq. (5.1) as described in Section 5.1.1.

When we analyzed the use of the MK series expansions as approximations to known target PDFs, we called that a *mathematical* analysis. As we now introduce uncertainties into $\kappa_{X,n}$, we can think of this as a *statistical* analysis. We can certainly expect a general reduction in accuracy as the primary effect of this introduction of uncertainty. From previous experience, we know that parameter uncertainty punishes the complex models more, so a key point of interest in this section is whether the benefit of additional correcting terms is eroded, compared to e.g. Figure 5.4.

¹⁹A detailed review of the asymptotic formulas of the polygamma function is beyond the scope of this discussion, but from [Abramowitz and Stegun, 1964] we know that as $M \rightarrow \infty$, $\psi^{(0)}(M) \rightarrow \log M$ and $\psi^{(n-1)}(M) \rightarrow 0$ for $n \geq 2$.

This section is organized as follows. We repeat the experiments of Figures 5.1 and 5.2 and Table 5.1, and briefly comment on these to get a basic understanding of using the MK series expansions methods for estimation versus approximation. We use 1,000 synthesized/generated data points in place of actual observations, and tabulate the mean of 1,000 iterations.²⁰ We then move on to an examination of convergence, repeating the experiments of Sections 5.2.3 and 5.2.4, while varying the number of data points. Finally, we examine estimates other than those we get from the MoLC, and visit a SAR-specific scenario.

5.3.1 Revisiting the Introductory Examples

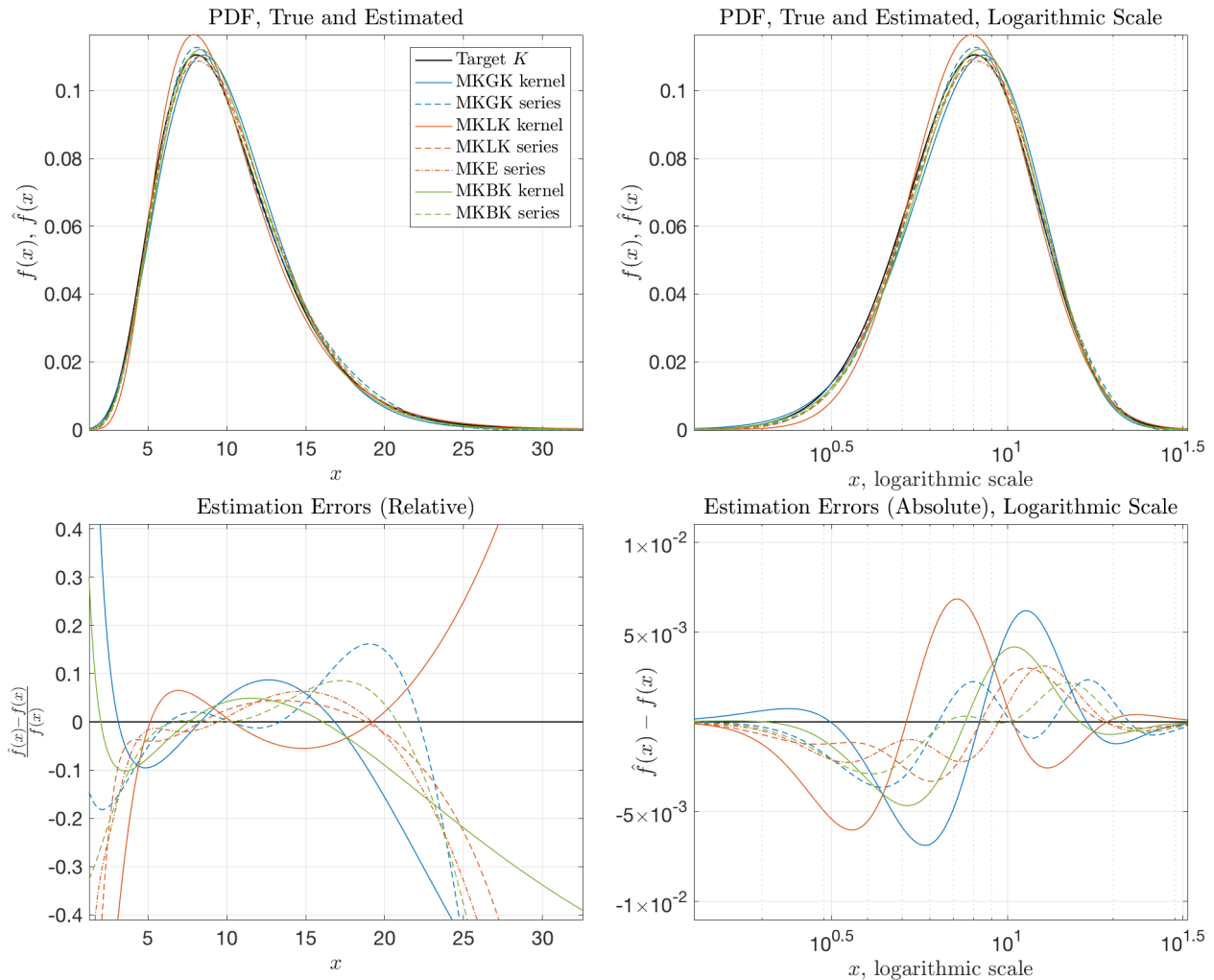


Figure 5.12: The four MK series expansions used to estimate the K distribution with parameters $L = 16$, $m = 10$, and $M = 10$, based on 1,000 synthesized data points. The solid lines are the kernels, and the dashed lines are the series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$.

Figure 5.12 shows the repetition of the experiment from Figure 5.1, with target PDF now unknown. We see that the error is increased slightly, with a bigger increase for the series expansions than their kernels. This is pretty much as expected, with the series expansions being affected more since they are more complex in the sense that they have a higher number of parameters to estimate than the bare kernels. Note that Figure 5.12 is only a single realization (with 1,000 data points), but when we later tabulate values we take the mean of the dissimilarity measures of 1,000 iterations.

²⁰It would certainly be interesting to present also the variance of the methods, but this was omitted for the sake of brevity.

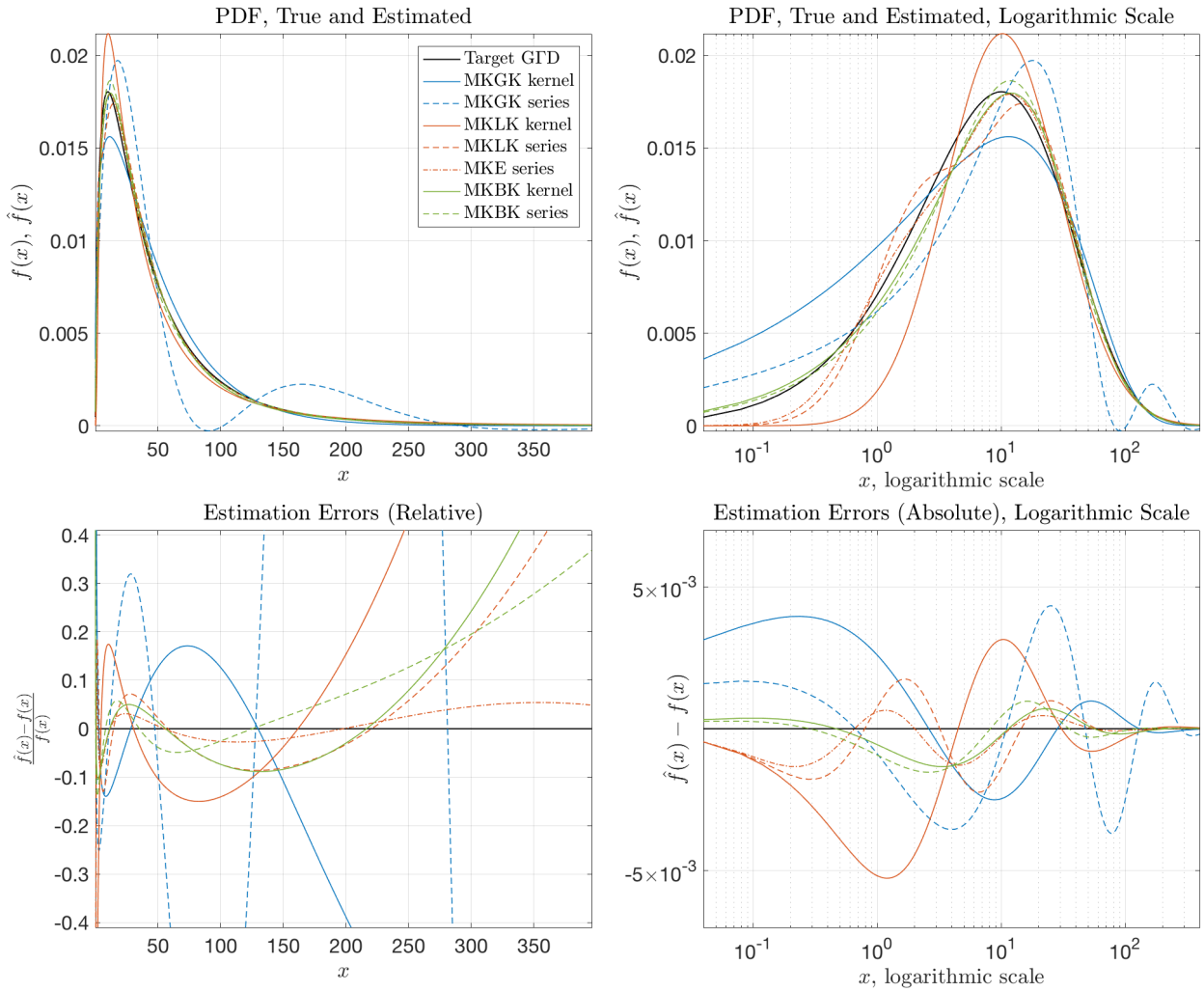


Figure 5.13: The four MK series expansions used to estimate the GFD with parameters $L = 4$, $m = 10$, and $\nu = 0.5$, based on 1,000 synthesized data points. The solid lines are the kernels, and the dashed lines are the series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$.

Figure 5.13 shows the estimation equivalent of the experiment from Figure 5.2, and the results are very similar. The differences are only apparent when carefully and directly comparing the error plots. Again we must stress that we have presented only a single realization. Repeating the experiment with different data samples gave us empirical log-cumulants which usually deviated more from their theoretical values, resulting in larger errors.

5.3.2 Estimations of Numerous Target PDFs Tabulated

In this section, we repeat the experiments from Section 5.2.2 with the target PDFs unknown this time, and we present the mean of the Kullback-Leibler distance of 1,000 iterations. One important difference is that the methods which exactly modeled targets in Table 5.1, can now be assessed as they will have some estimation error even though they coincide with the target PDF.

Table 5.2 presents our results. As in Section 5.2.2, there is some discord between the Bhattacharyya and Kullback-Leibler distances. Notably, the measures disagree on whether the MKLK or MKE series is the best expansion of the log-normal kernel. Recall that in Section 5.2.2, the MKE series in several cases had lower $d_B(\cdot)$ but higher $d_{KL}(\cdot)$. This happens here as well, but we also have some instances where $d_B(\cdot)$ favors the MKLK series when $d_{KL}(\cdot)$ favors

Table 5.2: Estimating a broad range of target distributions with MK series expansions, Kullback-Leibler distance to the true PDF. The kernels are tailored and the series expansions correct for $\Delta\kappa_3$ and $\Delta\kappa_4$, mean of 1,000 iterations, best method in bold.

Method\θ	$\gamma(x; \theta), \theta=[L, m]$		$\gamma^{-1}(x; \theta), \theta=[L, m]$		GFD($x; \theta$), $\theta=[L, m, d]$		$K(x; \theta), \theta=[L, m, M]$		$G^0(x; \theta), \theta=[L, g, M]$	
	[4, 1]	[4, 10]	[4, 1]	[4, 10]	[4, 10, 0.5]	[4, 10, 2]	[4, 10, 1]	[4, 10, 10]	[4, 1, -10]	[4, 10, -10]
MKGK kernel	1.02·10⁻³	1.05·10⁻³	7.47·10 ⁻²	7.47·10 ⁻²	1.59·10 ⁻²	5.85·10 ⁻³	7.82·10 ⁻³	5.22·10 ⁻³	9.63·10 ⁻³	9.55·10 ⁻³
MKGK series	2.15·10 ⁻²	2.40·10 ⁻²	1.63·10 ⁻¹	1.70·10 ⁻¹	1.13·10 ⁰	8.70·10 ⁻³	1.90·10 ⁰	2.98·10 ⁻²	3.85·10 ⁻²	3.65·10 ⁻²
MKLK kernel	1.93·10 ⁻²	1.93·10 ⁻²	1.96·10 ⁻²	1.97·10 ⁻²	1.68·10 ⁻²	2.01·10 ⁻²	3.71·10 ⁻²	1.02·10 ⁻²	6.68·10 ⁻³	6.62·10 ⁻³
MKLK series	5.42·10 ⁻³	6.60·10 ⁻³	5.59·10 ⁻³	7.58·10 ⁻³	6.41·10 ⁻³	5.63·10⁻³	7.88·10 ⁻²	3.25·10 ⁻³	2.64·10 ⁻³	2.64·10 ⁻³
MKE series	9.89·10 ⁻³	1.01·10 ⁻²	1.40·10 ⁻²	1.43·10 ⁻²	4.25·10 ⁻³	1.13·10 ⁻²	2.07·10 ⁻²	3.94·10 ⁻³	2.93·10 ⁻³	2.88·10 ⁻³
MKBK kernel	2.11·10 ⁻³	2.19·10 ⁻³	1.94·10⁻³	1.98·10⁻³	2.27·10⁻³	8.01·10 ⁻³	2.08·10⁻³	1.44·10⁻³	1.47·10⁻³	1.41·10⁻³
MKBK series	6.65·10 ⁻³	8.35·10 ⁻³	1.10·10 ⁻²	1.28·10 ⁻²	8.67·10 ⁻³	9.63·10 ⁻³	1.10·10 ⁻¹	4.69·10 ⁻³	3.12·10 ⁻³	3.46·10 ⁻³

Method\θ	$\gamma(x; \theta), \theta=[L, m]$		$\gamma^{-1}(x; \theta), \theta=[L, m]$		GFD($x; \theta$), $\theta=[L, m, d]$		$K(x; \theta), \theta=[L, m, M]$		$G^0(x; \theta), \theta=[L, g, M]$	
	[16, 1]	[16, 10]	[16, 1]	[16, 10]	[16, 10, 0.5]	[16, 10, 2]	[16, 10, 1]	[16, 10, 10]	[16, 1, -10]	[16, 10, -10]
MKGK kernel	1.02·10⁻³	9.78·10⁻⁴	1.99·10 ⁻²	1.93·10 ⁻²	5.57·10 ⁻³	2.61·10 ⁻³	1.92·10⁻³	3.75·10 ⁻³	1.99·10 ⁻²	1.98·10 ⁻²
MKGK series	3.43·10 ⁻³	3.53·10 ⁻³	4.93·10 ⁻³	4.97·10 ⁻³	1.58·10 ⁻²	2.49·10⁻³	8.46·10 ⁻¹	7.44·10 ⁻³	1.00·10 ⁻²	9.83·10 ⁻³
MKLK kernel	5.69·10 ⁻³	5.65·10 ⁻³	5.74·10 ⁻³	5.56·10 ⁻³	5.48·10 ⁻³	5.75·10 ⁻³	5.22·10 ⁻²	4.37·10 ⁻³	1.66·10 ⁻³	1.68·10 ⁻³
MKLK series	2.42·10 ⁻⁴	2.44·10 ⁻³	3.17·10 ⁻³	3.10·10 ⁻³	2.28·10 ⁻³	2.58·10 ⁻³	1.76·10 ⁻¹	2.18·10 ⁻³	2.86·10 ⁻³	2.69·10 ⁻³
MKE series	2.87·10 ⁻³	2.92·10 ⁻³	3.89·10 ⁻³	3.83·10 ⁻³	2.39·10 ⁻³	3.11·10 ⁻³	3.72·10 ⁻²	2.33·10 ⁻³	2.84·10 ⁻³	2.64·10 ⁻³
MKBK kernel	1.52·10 ⁻³	1.51·10 ⁻³	1.58·10⁻³	1.49·10⁻³	1.15·10⁻³	8.79·10 ⁻³	2.41·10 ⁻³	1.42·10⁻³	1.39·10⁻³	1.41·10⁻³
MKBK series	2.32·10 ⁻³	2.32·10 ⁻³	2.74·10 ⁻³	2.71·10 ⁻³	3.37·10 ⁻³	1.13·10 ⁻²	1.47·10 ⁻¹	2.56·10 ⁻³	3.12·10 ⁻³	2.88·10 ⁻³

the MKE series.

We see that for the cases in Table 5.2 where one of the kernels align with the target PDFs, that kernel is the best method (the gamma kernel for the gamma target and so on). Then, the correcting terms do not improve the accuracy, which is to be expected as the additional complexity is not warranted when the kernel coincides with the target. A comparison of Tables 5.1 and 5.2 illustrates how the kernels in general suffer less than their series expansions when the parameters/log-cumulants must be estimated. Mirroring the same trend we saw in Section 5.2.2, only the log-normal kernel is somewhat reliably improved by the series expansions correcting for both $\Delta\kappa_3$ and $\Delta\kappa_4$. We will in later sections examine whether other number of correcting terms are better, and if this depends on the quantity of data points.

The MKBK kernel must be especially discussed, as it is the top performer in so many of the cases. However interesting, comparing the MKBK kernel to the other methods is perhaps not fair. For one, it corrects for the first three log-cumulant differences, while the other methods correct for the first two (the kernels) or the first four (the series expansions). As we will see in the next section, this can actually be very significant. Also, the sheer complexity of computing the MKBK series kernel parameters, which we discussed in Section 3.5.3, is far beyond that of the other kernels. This puts the MKBK kernel in the same category as the fitted MoLC K or generalized gamma distributions.²¹ Our original goal was to see if the series expansion methods were more accurate than fitting the simple two-parameter kernels, or faster than fitting complex three-parameter PDFs, and they are definitely much faster.²² That is to say, the validity of the MK series expansion framework is not in doubt due to the performance of the fitted beta prime kernel.

5.3.3 Convergence and the Number of Data Points

Naturally, we must expect that the convergence is weaker now than when the target PDFs were known in Sections 5.2.3 and 5.2.4, due to the extra uncertainty. Additionally, we know that

²¹This is the category of three-parameter distributions, where the MoLC parameter estimates have to be found simultaneously by numerically iterative procedures, or approximated as in [Li et al., 2011].

²²Even without rigorous testing w.r.t. the time used by each method, we can safely say that even the optimized versions of fitting these three-parameter distributions are orders of magnitude slower than the series expansions of the gamma and log-normal kernels.

estimating a high number of parameters with few observations, is generally not a recipe for success. The time has come to assess how the convergence of the four MK series expansions is affected by having to estimate the log-cumulant differences.

Figure 5.14 mirrors Figure 5.4 in an estimation setting, but also includes the gamma target PDF omitted from Figure 5.4.²³ The results represent some of the most interesting findings in this chapter, as it gives valuable insight as to the best number of correcting terms w.r.t. the chosen method and the quantity and type of data.

The general impression from Figure 5.14 is that the convergence is much weaker than when the target distributions were known, but the trends from Figure 5.4 are recognizable. For instance, the MKGK series converges up to $N = 5$ or $N = 6$ for the same 3 target PDFs as in Figure 5.4, albeit only if the number of data points is sufficient. Also, recall that we used 1,000 data points in Table 5.2, and we can see that correcting only for $\Delta\kappa_3$ would have been better than also correcting for $\Delta\kappa_4$.

Regarding the log-normal kernel, corrections beyond $N = 3$ in the MKLK series rarely decrease the distance to the target PDF, and only marginally when it does. We see in Figure 5.14 that there is rarely any reason for choosing the MKLK series instead of the generally superior MKE series introduced in [Pastor et al., 2014]. Admittedly, the former was better in all cases when there was only 100 data points, but that is hardly relevant as neither series expansion improved when correcting for log-cumulants beyond the third. We recall that when correcting the tailored kernel for only $\Delta\kappa_3$, the MKLK and MKE series coincide as

$$f_X(x) \approx \left[1 + \frac{\kappa_3}{6\sigma^3} H_3 \left(\frac{\log x - \mu}{\sigma} \right) \right] \Lambda(x; \mu, \sigma). \quad (5.20)$$

We will refer to this as the *log-normal kernel corrected for logarithmic skewness*, to emphasize that it is a special case of both the MKLK and MKE series. When the quantity of data points is increased, it can be advisable to use the MKE series to correct for also $\Delta\kappa_4$ or even $\Delta\kappa_5$.

The MKBK series does not converge in any of the 5 cases, regardless of the quantity of the observations. To be fair, it can exactly model the $\gamma(\cdot)$, $\gamma^{-1}(\cdot)$ and $G^0(\cdot)$ distributions, and we have presented plenty of evidence that corrections to exact models only increases the error. Still, these findings seriously question the statistical merit of the MKBK series. That is, fitting the three-parameter beta prime PDF is an accurate method, and we have shown how to correct the kernel with a series expansion, but not that these corrections are useful in realistic scenarios.

Another observation we can make from Figure 5.14, is that correcting for $\Delta\kappa_4$ in Table 5.2 was not always justified. That is, with only 1000 data points available, the methods often performed better when corrected only for $\Delta\kappa_3$.

5.3.4 Alternative to the MoLC Parameter Estimates

As we briefly mentioned in Section 5.2.5, the MoLC estimators are not the only possible choice, and we will expand on that discussion presently. Recall from Chapter 3 that the MoLC estimates have the benefit of greatly reducing the number of correcting terms needed to correct the kernel up to given order log-cumulant differences. However, other estimators can have other benefits, as the MoLC estimators are not always unbiased nor do they necessarily have the least variance.

Perhaps the most natural alternatives are the ML estimates. The basic idea is to take the PDF, which is a function of a variable given parameters, and view it instead as the *likelihood function*

²³The experiment was repeated using the Bhattacharyya distance, with qualitatively very similar results, i.e. the results did not warrant inclusion.

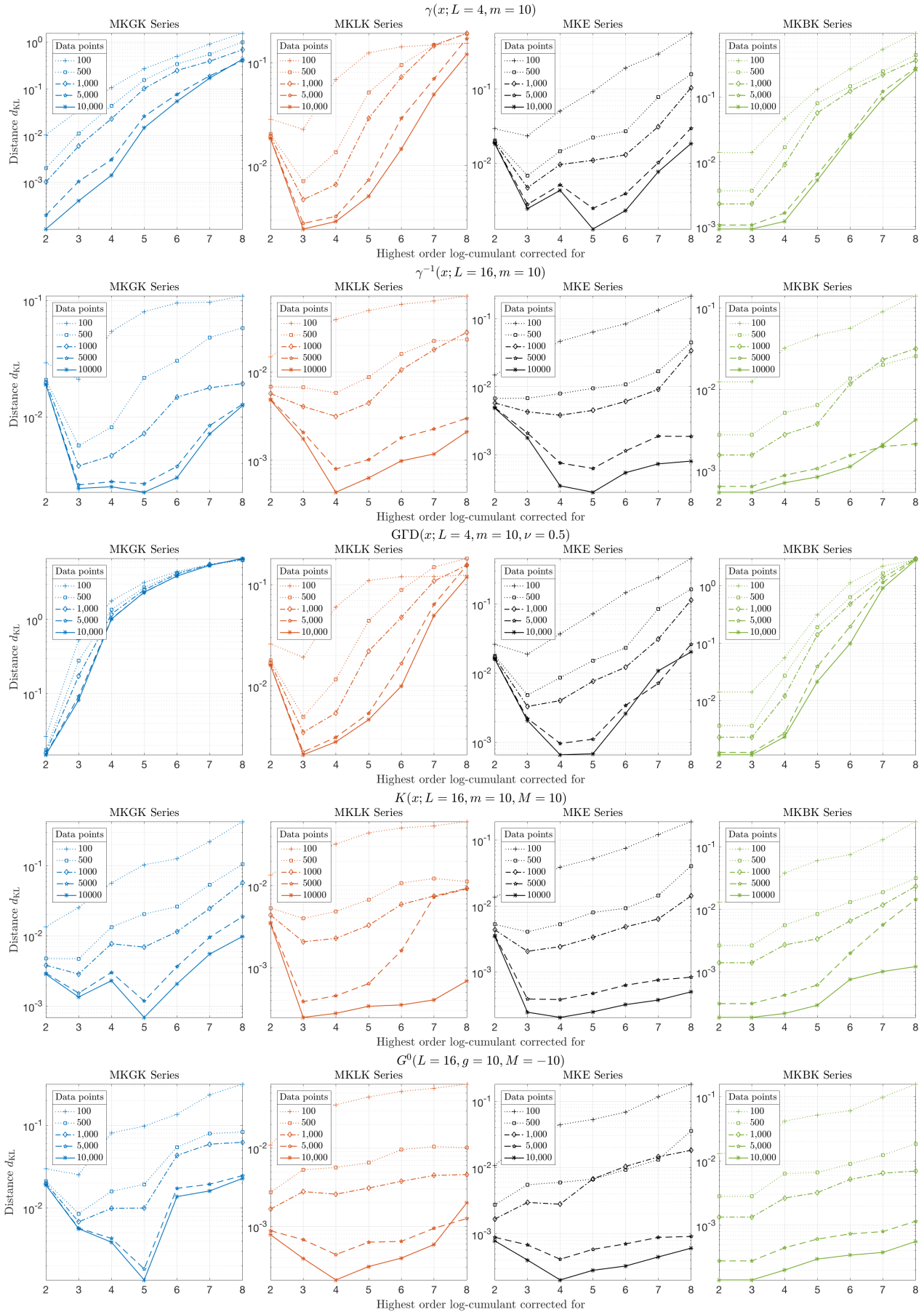


Figure 5.14: The MK series expansions estimating the same five target PDFs as in Figures 5.6 through 5.10. Plot of $d_{KL}(\cdot)$ vs. orders of log-cumulants corrected for, mean of 1,000 iterations.

of parameters given data. The ML estimates are the parameter values which maximizes the likelihood function, hence the name. This is explained thoroughly in most statistics textbooks, e.g. [Kendall et al., 1994].

The MKLK and MKE Series The log-normal distribution mirrors the Gaussian distribution in the sense that the ML estimates of the log-mean μ and log-variance σ^2 , are the same as the MoLC estimates, see [Cohen Jr., A. C., 1951] or [Krishnamoorthy, 2006]. That is, $\hat{\mu}_{\text{ML}} = \hat{\mu}_{\text{MoLC}} = \langle \mu \rangle$ and $\hat{\sigma}_{\text{ML}}^2 = \hat{\sigma}_{\text{MoLC}}^2 = \langle \sigma^2 \rangle$. In fact, $\langle \mu \rangle$ is also the minimum variance unbiased estimator for μ , and $\frac{n}{n-1} \langle \sigma^2 \rangle$, where n is the number of observations, is the minimum variance unbiased estimator for σ^2 , see [Aitchison and Brown, 1963].²⁴ We will now briefly explore whether it is beneficial to correct the bias of the empirical variance.

Table 5.3: Estimating five target distributions with the log-normal kernel corrected for logarithmic skewness, using both the biased MoLC/ML log-variance estimate $\langle \mu \rangle$, and the bias-corrected log-variance estimate $\frac{n}{n-1} \langle \mu \rangle$ (the minimum variance unbiased estimate for n data points). Kullback-Leibler distance to the true PDF, mean of 10,000 iterations.

Data Points	$\gamma(x; L=4, m=10)$		$\gamma^{-1}(x; L=16, m=10)$		GFD($x; L=4, m=10, d=0.5$)		$K(x; L=16, m=10, M=10)$		$G^0(x; L=16, g=2, M=-10)$	
	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased	Biased	Unbiased
100	$2.295 \cdot 10^{-2}$	$2.282 \cdot 10^{-2}$	$2.682 \cdot 10^{-2}$	$2.661 \cdot 10^{-2}$	$1.918 \cdot 10^{-2}$	$1.909 \cdot 10^{-2}$	$2.384 \cdot 10^{-2}$	$2.391 \cdot 10^{-2}$	$2.944 \cdot 10^{-2}$	$2.936 \cdot 10^{-2}$
1,000	$4.709 \cdot 10^{-3}$	$4.704 \cdot 10^{-3}$	$4.166 \cdot 10^{-3}$	$4.163 \cdot 10^{-3}$	$3.398 \cdot 10^{-3}$	$3.384 \cdot 10^{-3}$	$1.927 \cdot 10^{-3}$	$1.931 \cdot 10^{-3}$	$2.789 \cdot 10^{-3}$	$2.790 \cdot 10^{-3}$
10,000	$2.438 \cdot 10^{-3}$	$2.437 \cdot 10^{-3}$	$1.801 \cdot 10^{-3}$	$1.801 \cdot 10^{-3}$	$2.076 \cdot 10^{-3}$	$2.075 \cdot 10^{-3}$	$2.435 \cdot 10^{-4}$	$2.434 \cdot 10^{-4}$	$4.058 \cdot 10^{-4}$	$4.059 \cdot 10^{-4}$

Table 5.3 shows the results of our experiment. Note that we have increased the number of iterations to 10,000, which allows us to include an additional digit in our measurement to better differentiate between the methods. To better compare the methods, we applied both approaches to each of the generated samples, instead of generating 10,000 samples for the biased estimator and another 10,000 samples for the unbiased one. Also, we only corrected for the logarithmic skewness $\Delta\kappa_3$ as we learned in Section 5.3.3 that this is realistically the best choice for this number of data points.

Our findings indicate a small but consistent advantage from choosing the unbiased estimator, except for the $K(\cdot)$ target. This is interesting in its own right, and it also raises the question of whether this advantage is mirrored in the classical series. On the other hand, the difference is largest for only 100 data points, and we know from Section 5.3.3 that even correcting for only $\Delta\kappa_3$ based on just 100 data points is not always warranted. When we have larger samples, the differences between using the biased and unbiased estimates all but vanish. This is perhaps not surprising, as the estimates themselves become more and more similar as $\frac{n}{n-1} \rightarrow 1$. Thus, in terms of the sample size, it seems that the benefit of using the unbiased estimate vanishes around the same time as we get enough data points to justify corrections to the kernel at all.

The MKGK Series Estimating the gamma kernel parameters is much more intricate than in the log-normal and Gaussian distributions. Working with SAR data adds another layer of complexity, as some estimators can use data from the whole image, as we discussed in Section 5.2.6. There, we elaborated on how the K distribution model parameters are the global number of looks L , and the local mean m and texture-related M . We also performed an experiment where the values of L and m were used in the gamma kernel, and the MKGK series expansion tasked with modeling the texture. Those results were promising, so we will continue along that road here, providing the MKGK series with the true value of L in lieu of a global estimate.

²⁴These properties are matched perfectly by the classical Gaussian distribution. Recall that we speculated as to whether this lack of ambiguity in the choice of estimators is the reason why non-tailored kernel parameters have received so little attention in the classical case.

Solving for the location or mean m in eq. (2.62), we see that the MoLC estimator, given by the restraint that $\kappa_{\gamma,1} = \langle \kappa_{X,1} \rangle$, is

$$\hat{m}_{\text{MoLC}} = L \cdot \exp\{\langle \kappa_{X,1} \rangle - \psi^{(0)}(L)\}, \quad (5.21)$$

where we recall that the first order log-cumulant equals the log-mean, i.e. $\langle \kappa_{X,1} \rangle = \langle \mu_{X,1} \rangle$. On the other hand, [Oliver and Quegan, 2004] points out that if we have a sample $\{x_1, \dots, x_n\}$ of n data points, then the ML estimate of m is the sample (linear) mean

$$\hat{m}_{\text{ML}} = \langle m \rangle \equiv \frac{1}{n} \sum_{i=1}^n x_i. \quad (5.22)$$

The ML estimate for L is more complicated, see [Choi and Wette, 1969], but we assume it to be known, as discussed above.

We can also find support for using the linear sample mean in the physical SAR model. The basic idea is that the measured quantity is a RV $Y = X \cdot T$, where X and T are independent RVs representing the speckle noise and texture, respectively.²⁵ By convention, T is chosen to have unit mean. At this point we note that if X has PDF $\gamma(x; L, m)$ and T has PDF $\gamma(t; M, 1)$, Y follows the K distribution with PDF $K(y; L, m, M)$. Since X and T are (assumed) independent, the linear mean of Y is simply

$$\text{E}\{Y\} = \text{E}\{X\} \cdot \text{E}\{T\} = \text{E}\{X\} \cdot 1 = \text{E}\{X\}, \quad (5.23)$$

which is reflected in the fact that the mean of the K distribution is m . Regarding the log-mean, we have that

$$\text{E}\{\log Y\} = \text{E}\{\log X\} + \text{E}\{\log T\}, \quad (5.24)$$

but this is not especially helpful as there are no guarantees as to the numerical value of $\text{E}\{\log T\}$. That is, the linear mean is the same for the observable RV and the unobservable X , a property which the log-mean does not have. For a more thorough review of the physical SAR model, see e.g. [Oliver and Quegan, 2004] or [Deng et al., 2016].

Now we will continue our discussion from Section 5.2.6, i.e. we will examine whether the MKGK series can perform the role of modeling the texture T . Our focus now is whether the MoLC or sample mean (ML) estimates of m are best suited in this regard.

Figure 5.15 teaches us a few things. The most immediate insight is that the non-tailored kernels are more viable for lower L and higher M , which in a SAR scenario is interpreted as X explaining a bigger proportion of the variance compared to T . Additional testing with other values of L and M confirmed this trend, which we provided mathematical and physical explanations for in Section 5.2.6.

In Figure 5.11 we saw convergence for known L and m when approximating the target $K(x; L = 4, m = 10, M = 20)$. This is not the case when m must be estimated, regardless of whether we use \hat{m}_{MoLC} or \hat{m}_{ML} . Obviously, this is because the estimation uncertainty overshadows the benefit from the exact correcting terms. One thing we do recognize from the $K(x; L = 4, m = 10, M = 20)$ target in Figure 5.11, is that the non-tailored kernels with $N = 2$ outperforms the tailored kernel, at least when the number of data points is high enough. This is in agreement with the physical interpretation of the doubly stochastic product model. When simply fitting the tailored gamma kernel, the additional variance contributed by the texture variable T is falsely

²⁵The speckle phenomena is not actually noise in the usual sense, but a consequence of the measurement process. Unfortunately, delving deeper into this discussion is beyond the scope of this thesis.

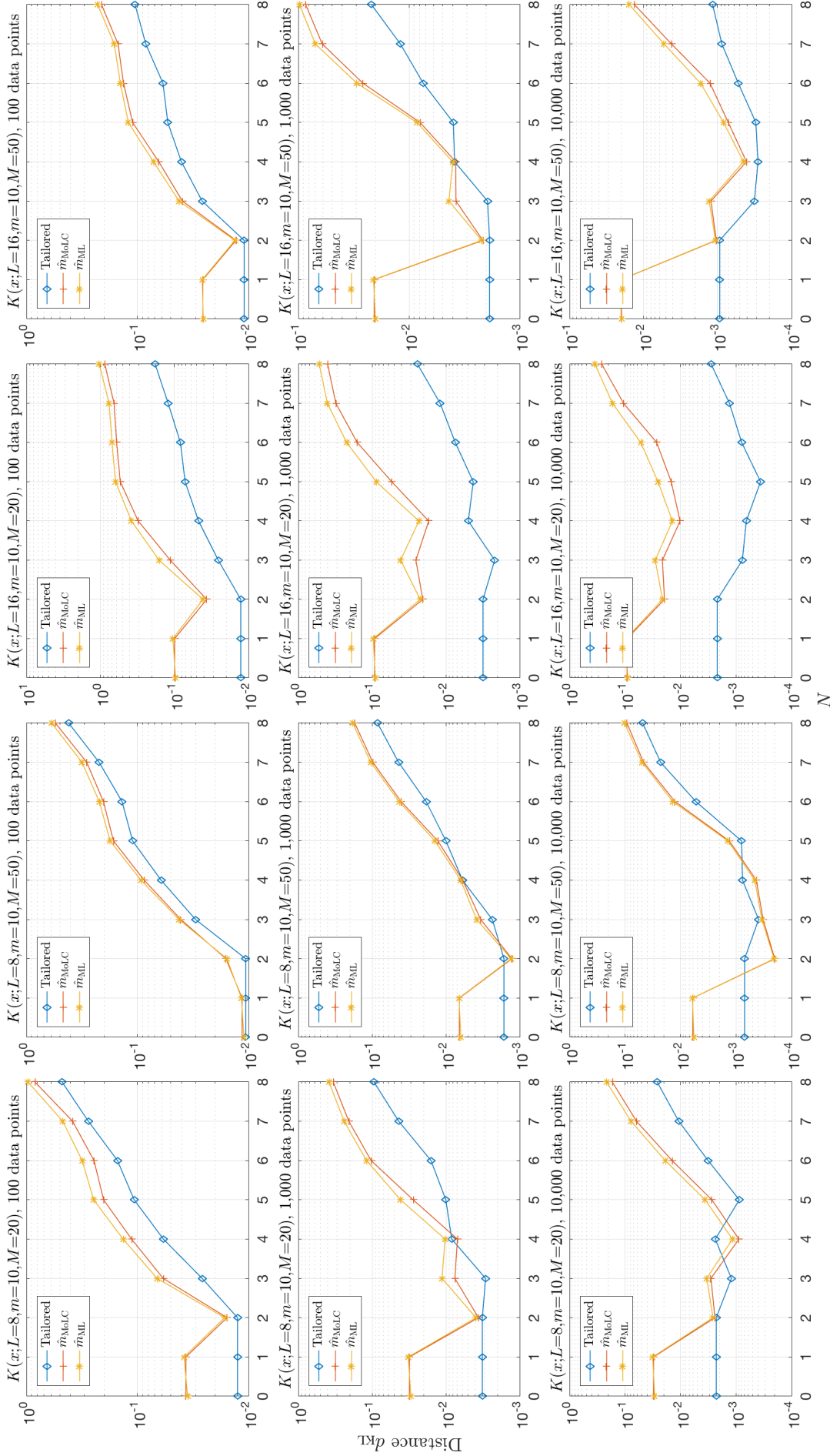


Figure 5.15: Three different gamma kernels expanded with the MKGK series and used to model several K distributions. The first kernel is the tailored version, and the other two both have shape equal to the target PDFs true value of L , and use the mean estimates \hat{m}_{MoLC} and \hat{m}_{ML} , respectively. The experiment consists of four different target PDFs, repeated for 100, 1,000, and 10,000 data points. The dissimilarity measure presented is the Kullback-Leibler distance to the target PDF, mean of 1,000 iterations. Note that using \hat{m}_{MoLC} implies that the kernel corrects for $\Delta\kappa_1$, i.e. the method does not discern between $N = 0$ and $N = 1$. Similarly, the tailored MKGK series is equal for $N = 0, 1, 2$.

attributed to the speckle noise X . This results in a fundamentally wrong shape parameter value, which is unable to fully capture to heavy-tailed nature of the physical process. However, the general message is that the tailored kernels performed better than the non-tailored ones, although the latter are more true to the physical model, probably due to the estimation uncertainty introduced.²⁶ Also, letting the shape vary for each class within the image is considerably slower than fixing it to a global value.

When the number of data points is low however, the MKGK series with tailored kernel is clearly better suited than the alternatives. This can likely be attributed to the fact that tailoring the kernel greatly reduces the number of terms in the MKGK series, as explained in Section 3.2.4. When the number of data points is low, each term is associated with a large degree of uncertainty, so fewer terms is almost always better.

The same effect is present in a much more subtle way when comparing the two non-tailored kernels and their series expansions. Close examination reveals that \hat{m}_{ML} is slightly, but consistently, better than the corresponding \hat{m}_{MoLC} for $N = 0$. However, \hat{m}_{MoLC} leads to better results for $N > 0$. We can interpret this in the following way: $\hat{m}_{\text{ML}} = \langle m \rangle$ is more accurate than \hat{m}_{MoLC} , but using \hat{m}_{ML} leads to $\Delta\kappa_1 \neq 0$. On the other hand, \hat{m}_{MoLC} ensures that $\Delta\kappa_1 = 0$, which eliminates a great many terms in the MKGK series. The fact that \hat{m}_{ML} is better for $N = 0$, but not for $N > 0$, is due to the fact that the benefit of its superior accuracy does not outweigh the inaccuracies contained in the terms that vanish when using \hat{m}_{MoLC} .

5.3.5 The MKBK Series Used in a SAR Change Detection Scenario

Recall that we discussed the use of the beta prime distribution in SAR change detection already in Section 2.3.10, and arrived at a simplified beta prime PDF with two parameters in eq. (2.69).²⁷ Based on our experiment in the previous section, we can now ask whether it is possible use the MKBK series expansion of the two-parameter beta prime kernel in eq. (3.110) to model the quotient of two SAR images of a *heterogeneous* area. The images are again represented by synthesized K distributed data, i.e. we task the MKBK series with fitting the distribution of the quotient of two IID K distributed RVs.²⁸ We also used the G^0 distribution to represent the SAR images. It was proposed on a mathematical basis in [Frery et al., 1997] for modeling extremely heterogeneous (urban) areas, with a physical explanation provided as recently as [Deng et al., 2016]. While a thorough review of SAR change detection is beyond the scope of this thesis, [Inglada and Mercier, 2007] discussed some approaches,²⁹ [Moser and Serpico, 2009] used the MK statistics in a capacity, and [Akbari. et al., 2016] is a recent example of how approximative PDF models are essential in the SAR change detection field.

If, as we claimed already in Section 2.3.9, the K distribution is analytically and numerically demanding, it is certainly well outside the realm of feasibility to work with the PDF of the

²⁶We tested this with other parameter values as well, and the results indicate that this approach could be promising, at least for certain values of L .

²⁷To recapitulate, the two-parameter beta prime PDF $\beta'(x; L, b)$ is the PDF followed by the quotient of two IID gamma RVs, each with shape L . As we discussed in Section 5.2.6, a gamma RV represents a SAR image of a homogeneous area, so two IID gamma RVs represent a pair of independent SAR images of the same area, which has not undergone change between the two acquisitions (or two indistinguishable areas).

²⁸An alternative motivation is to say that each image follows the Wishart distribution, which is the extension of the gamma distribution to multiple dimensions. Then, division is no longer possible and we must instead multiply the inverse matrix of the first image with the second. To get a scalar statistic, it is customary to take the trace of the resulting matrix product, which will itself be not exactly beta prime even though the images are exactly Wishart.

²⁹Interestingly, Inglada and Mercier found the *classical* Edgeworth series to be successful in SAR change detection.

quotient of two K distributed RVs in a simple and fast way.³⁰ Therefore, we deemed it necessary to estimate the target PDF using the non-parametric kernel density method,³¹ when computing the distances to the MKBK series. For the target PDF estimate we used 1,000,000 data points, in order to ensure a highly accurate estimate with negligible distance to the true target.

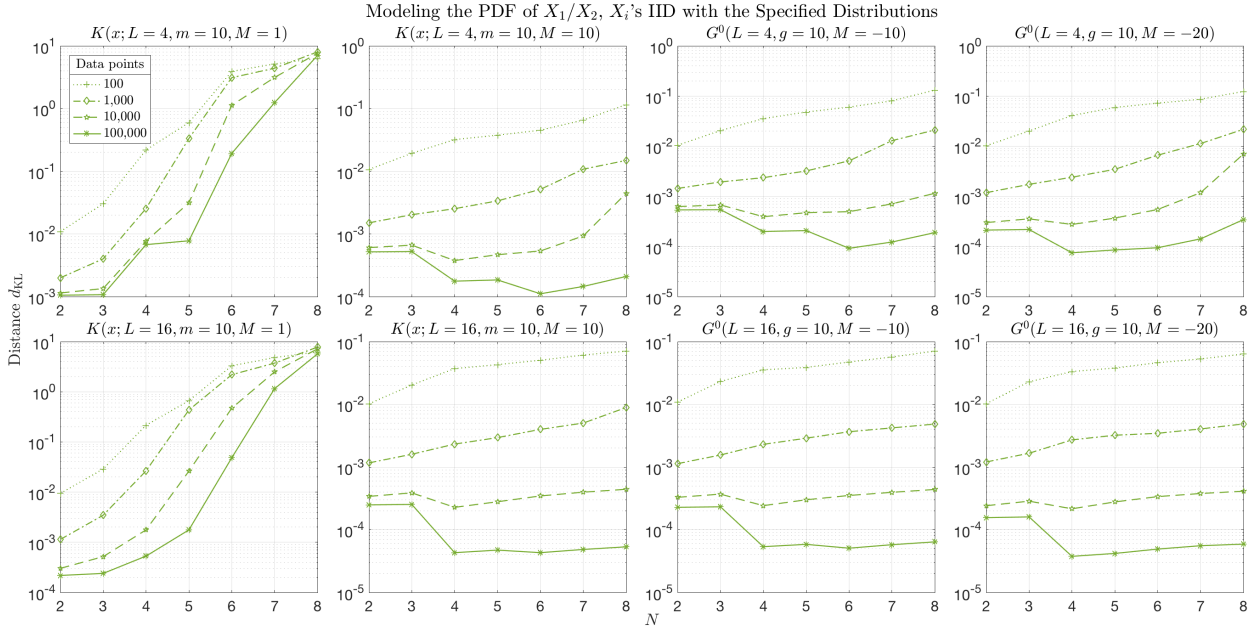


Figure 5.16: The MKBK series expansion of $\beta'(x; L, b)$ from eq. (3.110), used to model synthesized data representing SAR change images. That is, the data points are the quotients of IID RVs following K or G^0 distributions, as specified in the subfigure titles. The analytically complicated target PDF was replaced with a kernel density estimate based on 1,000,000 data points, and the measure is the Kullback-Leibler distance to this target PDF, mean of 1,000 iterations.

In Figure 5.16, we present the results of applying the MKBK series expansion of the two-parameter $\beta'(x; L, b)$ to synthesized SAR change detection scenario. Specifically, modeling the target PDFs are the distributions followed by the quotient of either two IID K distributed RVs, or two IID G^0 distributed RVs. The distribution parameters are the same as in Tables 5.1 and 5.2, except that we saw how the series expansions methods were invariant to the g parameters in $G^0(\cdot)$, so we instead present two cases where $M = -20$.

The immediate impression is that the method is fairly successful, provided the samples are large enough. The two K distribution cases with $M = 1$ are exceptions in this regards, but we recall that the $K(\cdot)$ targets with low values of M have been extremely challenging for the series expansion methods throughout this chapter, and that the dependency on M was discussed in more detail in Section 5.2.6. However, there are only slight benefits to be had from the series expansions for 10,000 data points, and no benefits if the samples are even smaller.

An interesting point can be raised if we revisit the expression for the theoretical log-cumulants of the quotient of two independent RVs in eq. (2.55). Let X denote the target RV, i.e. $X = X_1/X_2$, where X_1 and X_2 are IID, and let $\kappa_{X_i, n}$ denote the n th-order log-cumulant of X_1 and X_2 . Then

³⁰As with several other complicated product and quotient distributions, the Meijer G function is required to get an expression for the PDFs, see [Nicolas, 2011].

³¹We take the liberty of omitting an explanation of kernel density estimation, as it is not strictly relevant to this thesis. Among the original references concerning this method are [Rosenblatt, 1956] and [Parzen, 1962], with modern takes on this popular technique easily found online.

eq. (2.55) gives

$$\kappa_{X,n} = \kappa_{X_i,n} + (-1)^n \kappa_{X_i,n} = \begin{cases} 2 \cdot \kappa_{X_i,n} & n \text{ even,} \\ 0 & n \text{ odd.} \end{cases} \quad (5.25)$$

Inserting the expression for the log-cumulants of the two-parameter beta prime kernel in eq. (2.70) gives us the log-cumulant differences

$$\Delta\kappa_n = \kappa_{X,n} - \kappa_{\beta',n} = \begin{cases} \log b & n = 1, \\ 2 \cdot (\kappa_{X_i,n} - \psi^{(n-1)}(a)) & n = 2, 4, 6, \dots, \\ 0 & n = 3, 5, 7, \dots \end{cases} \quad (5.26)$$

In other words, the odd log-cumulants except the first are all zero. This effect actually manifests itself in Figure 5.16 as well. Observe how there is never any improvement when going from $N = 2$ to $N = 3$; this is presumably because $\Delta\kappa_3 = 0$, so any "correction" performed by the MKBK series at that stage only introduces estimation error. Across all subfigures, we see that this increase in error becomes smaller as the samples become larger, i.e. when $\langle \kappa_{X,n} \rangle \rightarrow \kappa_{X,n} = 0$. This effect is also noticeable when going from $N = 4$ to $N = 5$, and to a lesser degree when going from $N = 6$ to $N = 7$. Presumably, the results in Figure 5.16 would be improved by taking eq. (5.26) into account.

Based on eq. (5.26), we can also satisfy $\Delta\kappa_1 = 0$ by setting $b = 1$, regardless of the sample at hand. That is, under the assumption that we have two independent images of an area which has not undergone change, we can reduce the beta prime kernel to a single parameter by fixing $b = 1$ in eq. (2.69):

$$\beta'(x; a) \equiv \frac{\Gamma(2a)}{[\Gamma(a)]^2} \frac{x^{a-1}}{(1+x)^{2a}}, \quad x, a > 0. \quad (5.27)$$

This results in an elegant series expansion. For example, we could choose the kernel parameter a s.t. $\Delta\kappa_2 = 0$. But from eq. (5.26) we know that $\Delta\kappa_3 = \Delta\kappa_5 = 0$, and by the properties of the Bell polynomials, so then only one correcting term is required to correct for up to $\Delta\kappa_5$. That is, the MKBK series in eq. (3.102) is simplified to

$$f_X(x) = \left[1 + \frac{\Delta\kappa_4}{4!} M'_4 \left(\frac{x}{1+x} \right) + \frac{\Delta\kappa_6}{6!} M'_6 \left(\frac{x}{1+x} \right) + \frac{\Delta\kappa_4^2 + \Delta\kappa_8}{8!} M'_8 \left(\frac{x}{1+x} \right) + \dots \right] \beta'(x; a). \quad (5.28)$$

This has the remarkable property that the single-parameter $\beta'(x; a)$ can be corrected for up to $\Delta\kappa_5$ with a single correcting term.

Additionally, we can use take similar approach as in Section 5.2.6, setting the kernel shape a equal to the global estimate of the target parameter L and task the MKBK series expansion with modeling the effects of the texture. For instance, assume that X_1 and X_2 are IID K with parameters L, m, M , i.e. that $\kappa_{X_i,n} = \kappa_{K,n} \forall n$. Combining eq. (5.8) and (5.26) with $b = 1$ and $a = L$ gives

$$\Delta\kappa_n = \kappa_{X,n} - \kappa_{\beta',n} = \begin{cases} 2 \cdot \psi^{(n-1)}(M) & n \text{ even,} \\ 0 & n \text{ odd.} \end{cases} \quad (5.29)$$

This is reminiscent of eq. (5.19), and the discussion in Section 5.2.6 is highly relevant for this case as well.

To summarize, we said in previous sections that correcting three-parameter beta prime kernel with the MKBK series expansion seems unsuited at modeling the target PDFs from Section 5.1.2, by virtue of being too slow and complex. When it comes to SAR change detection on the other hand, the MKBK series not only shows promising experimental results, but also significant possibilities when it comes to simplifying the kernel and its series expansion.

5.4 Experiments on Real SAR Data

We end this chapter with a few examples of the series expansion methods applied to real SAR data. First, we fit the MK series expansions to data from different classes in a single SAR image. Then, we use the MKBK series to SAR change detection as described in Section 5.3.5.

The SAR data we use is a single polarimetric image of San Fransisco.³² This image is diverse and big enough to be the only one we need for this thesis. It is a single look image, i.e. $L = 1$, and applying a simple averaging filter gives us realistic data for the cases $L = 4$ and $L = 16$.³³

5.4.1 Modeling SAR Image Data

A Water Region The first region we look at is by far the largest, comprising over a quarter million pixels.

Figure 5.17 presents the image of San Fransisco with the "water" region framed and enlarged. It is clear that the selected region is heterogeneous in the sense that it has texture, but that it consists of only a single class.³⁴ We can perhaps recognize some swells, which are far too big to be considered texture, but must be tolerated when the tradeoff is such a large region and vast amount of pixels.

Qualitatively, we can see from the bottom plots in Figure 5.17 that the log-normal kernels and its two series expansions struggle heavily to fit the data when $x \rightarrow 0$, since $\Lambda(x; \mu, \sigma) \rightarrow 0$ as $x \rightarrow 0$ regardless of the parameter choices. For $L = 4$ we can see both the log-normal and gamma kernels struggling to fit the data, but while the series expansions of the former offer significant improvement, the MKGK series overcompensates around the mode and reverberates in the tail.³⁵ As we have come to expect, the series expansion methods perform well for $L = 16$ with this many data points. However, we see the same tendency to overcompensate in the MKGK series and its kernel, albeit to a much smaller degree.

Table 5.4: The MK series expansions fitted to the blue channel intensities in the "water" region of our SAR image, Kullback-Leibler distance to a kernel density estimate based on the entire region. The kernels are tailored and the series expansions correct for $\Delta\kappa_3$ and $\Delta\kappa_4$, mean of 100 iterations where 100, 1,000 or 10,000 data points were drawn at random, best method in bold.

Method\Data Points	$L = 1$			$L = 4$			$L = 16$		
	100	1,000	10,000	100	1,000	10,000	100	1,000	10,000
MKGK kernel	1.28·10⁻²	2.54·10 ⁻³	1.23·10 ⁻³	2.41·10 ⁻²	1.47·10 ⁻²	1.36·10 ⁻²	1.69·10 ⁻²	7.46·10 ⁻³	6.46·10 ⁻³
MKGK series	2.76·10 ⁰	8.67·10 ⁻¹	1.09·10 ⁻¹	1.26·10 ⁻¹	8.11·10 ⁻²	7.59·10 ⁻²	4.99·10 ⁻²	1.23·10 ⁻²	4.93·10 ⁻³
MKLK kernel	6.18·10 ⁻²	5.13·10 ⁻²	5.02·10 ⁻²	1.43·10 ⁻²	5.81·10 ⁻³	4.71·10 ⁻³	1.16·10⁻²	2.18·10 ⁻³	1.30·10 ⁻³
MKLK series	3.68·10 ⁻¹	9.36·10 ⁻²	3.79·10 ⁻²	2.88·10 ⁻²	2.41·10 ⁻³	4.81·10 ⁻⁴	3.07·10 ⁻²	2.25·10 ⁻³	2.07·10⁻⁴
MKE series	1.13·10 ⁻¹	1.45·10 ⁻²	7.91·10 ⁻³	3.47·10 ⁻²	2.19·10 ⁻³	2.77·10⁻⁴	3.60·10 ⁻²	2.49·10 ⁻³	2.07·10⁻⁴
MKBK kernel	1.71·10 ⁻²	2.03·10⁻³	2.87·10⁻⁴	1.23·10⁻²	2.00·10⁻³	7.05·10 ⁻⁴	1.29·10 ⁻²	1.58·10⁻³	2.88·10 ⁻⁴
MKBK series	2.05·10 ⁻¹	1.18·10 ⁻¹	5.17·10 ⁻²	3.44·10 ⁻²	2.62·10 ⁻³	3.30·10 ⁻⁴	3.37·10 ⁻²	2.97·10 ⁻³	2.25·10 ⁻⁴

³²Polarimetric SAR data consists of multiple channels, corresponding to different combinations of the emitted and incident radiation polarizations. These channels can be combined linearly in the Pauli basis, to better match physical interpretations of the surface, see [Cloude, 2010] for more information. As the methods presented in this thesis are limited to information from a single channel, we choose the best one w.r.t. each class we examine here.

³³Due to the inevitable spatial correlation between neighboring pixels, averaging four pixels lead to non-integer *equivalent* number of looks $L < 4$, and similarly, averaging 16 pixels leads to $L < 16$, see [Anfinsen et al., 2009].

³⁴Very shallow ocean and oil slicks are examples of ocean areas with fundamentally different properties w.r.t. radar reflectivity.

³⁵We saw the same thing happen already in Figure 5.2, where the MKGK series corrected for the gamma kernel on the left side of the mode, causing perturbations in the tail. This behavior motivates further exploration of non-tailored kernel parameter estimates, which might result in a more balanced approach.

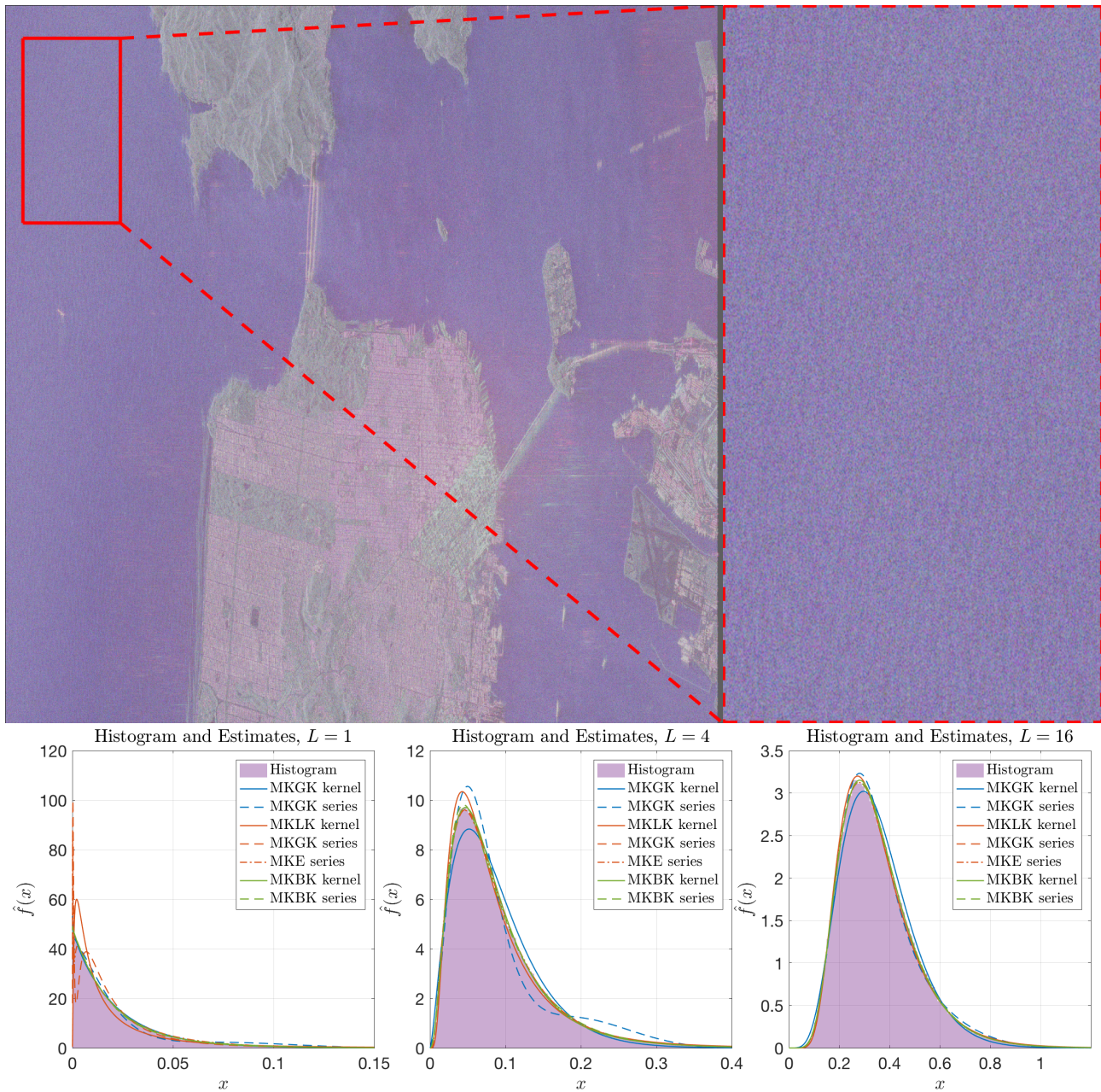


Figure 5.17: The SAR image with the chosen "water" region framed (top left), and an enlarged version of the "water" region (top right). The entire image is 2,800 by 2,800 pixels, with the "water" region consisting of over a quarter million pixels. Bottom: Histograms of all blue channel intensities within the "water" region, and the MK series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$. We included the raw single-look data (left), and averaged data corresponding to $L = 4$ (center) and $L = 16$ (right).

Table 5.4 presents the quantitative results when drawing samples at random from the "water" region. Each sample consisted of 100, 1,000, or 10,000 *different* data points, but each data point could be part of several samples. The choice was made to draw each data point in the sample independently.³⁶ As there was just over a quarter million pixels in the selected region, we can expect there to be some correlation arising from re-use of data points with a sample size of 1,000. For sample size 10,000 we expect significant correlation, since each data point was included in almost four samples on average. We measured the Kullback-Leibler distance to a

³⁶The alternative was to draw random subregions, and use all pixels in those regions to get geographically adjacent data points to test on.

kernel density estimate based on all pixels in the region.³⁷

The immediate impression from Table 5.4 is dominated by the strength of the bare kernels, which outperform the series expansions in 7 of the 9 cases. However, as we have argued before, the beta prime kernel is not really comparable to the other kernels due to its much more complicated and computationally demanding nature. Disregarding the MKBK kernel and series, the results are very much in line with our findings for synthesized data. That is, we need more than 100 or even 1,000 data points to warrant correcting for up to $\Delta\kappa_4$,³⁸ and the series expansions of the log-normal kernel is clearly the most reliable when it comes to improving on the kernel.

One key observation from this section is how the log-normal kernel and its series expansions struggle to fit PDFs which are non-zero at $x = 0$, here corresponding to single look data.. This is the result of the fact we observed in Figure 5.17, i.e. $x \rightarrow 0 \Rightarrow \Lambda(x; \mu, \sigma) \rightarrow 0 \forall \mu, \sigma$, which we discussed above.

Although it is not surprising that the three cases with 100 data points in Table 5.4 were best estimated by kernels, we should note that the three cases were best modeled by three *different* kernels, supporting the notion that SAR data analysts must master a variety of methods.

A Park Region The second region we look at is much smaller than the first, and is perhaps more realistic in that regard with its almost 18,000 pixels.

Figure 5.18 presents our "park" region, which is in fact a segment of the Golden Gate Park. We can see that the entire region belongs to a single, heterogeneous class, but there are some visible features, presumably clearings and groves. From the histograms and fitted series expansions we see largely the same trends as in Figure 5.17, even though we can see from the histograms that we have much fewer data points in this instance.

Table 5.5: The MK series expansions fitted to the green channel intensities in the "park" region of our SAR image, Kullback-Leibler distance to a kernel density estimate based on the entire region. The kernels are tailored and the series expansions correct for $\Delta\kappa_3$ and $\Delta\kappa_4$ (with an explicit exception), mean of 100 iterations where 100, 500 or 1,000 data points were drawn at random, best method in bold.

Method\Data Points	$L = 1$			$L = 4$			$L = 16$		
	100	500	1,000	100	500	1,000	100	500	1,000
MK GK kernel	$3.05 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	$1.64 \cdot 10^{-2}$	$3.30 \cdot 10^{-2}$	$2.22 \cdot 10^{-2}$	$2.11 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	$7.77 \cdot 10^{-3}$	$7.09 \cdot 10^{-3}$
MK GK series	$3.79 \cdot 10^0$	$3.53 \cdot 10^0$	$3.22 \cdot 10^0$	$1.29 \cdot 10^0$	$7.86 \cdot 10^{-1}$	$6.05 \cdot 10^{-1}$	$3.34 \cdot 10^{-1}$	$8.21 \cdot 10^{-2}$	$4.27 \cdot 10^{-2}$
MK LK kernel	$3.94 \cdot 10^{-2}$	$2.80 \cdot 10^{-2}$	$2.65 \cdot 10^{-2}$	$1.98 \cdot 10^{-2}$	$1.04 \cdot 10^{-2}$	$9.53 \cdot 10^{-3}$	$2.52 \cdot 10^{-2}$	$1.58 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$
MK LK series, $N = 3$	$2.15 \cdot 10^{-2}$	$8.68 \cdot 10^{-3}$	$8.09 \cdot 10^{-3}$	$2.77 \cdot 10^{-2}$	$3.50 \cdot 10^{-3}$	$1.92 \cdot 10^{-3}$	$4.92 \cdot 10^{-2}$	$9.89 \cdot 10^{-3}$	$6.78 \cdot 10^{-3}$
MK LK series	$1.36 \cdot 10^{-1}$	$1.35 \cdot 10^{-1}$	$2.45 \cdot 10^{-1}$	$3.34 \cdot 10^{-2}$	$4.21 \cdot 10^{-3}$	$2.84 \cdot 10^{-3}$	$1.32 \cdot 10^{-1}$	$1.02 \cdot 10^{-2}$	$7.46 \cdot 10^{-3}$
MK E series	$5.13 \cdot 10^{-2}$	$4.02 \cdot 10^{-2}$	$1.07 \cdot 10^{-1}$	$5.02 \cdot 10^{-2}$	$5.21 \cdot 10^{-3}$	$2.59 \cdot 10^{-3}$	$1.15 \cdot 10^{-1}$	$1.49 \cdot 10^{-2}$	$8.05 \cdot 10^{-3}$

Table 5.5 presents the numerical results of our experiment on the "park" region. Compared to the experiment on the much larger "water" region, we reduced the number of samples drawn to 100, and limited the maximum number of observations in each sample to 1,000.

We omitted the beta prime kernel and the MKBK series expansion, to get a comparison between the methods whose complexity and computational demand are similar. For the record, the fitted beta prime kernel was in most cases better than all other methods, and sometimes quite clearly. The MKBK series expansion failed to improve on the estimates of its kernel. Instead,

³⁷We also measured the Bhattacharyya distance, but omitted those findings as they were practically identical to $d_{KL}(\cdot)$.

³⁸Repeating the experiment while correcting only for $\Delta\kappa_3$ gave better estimates in many cases, especially for 100 and 1,000 data points. This is also in line with our previous findings. As before, we still correct for $\Delta\kappa_4$ in Table 5.4 to distinguish between the MKLK and MKE series, and between the MKBK series and its kernel.

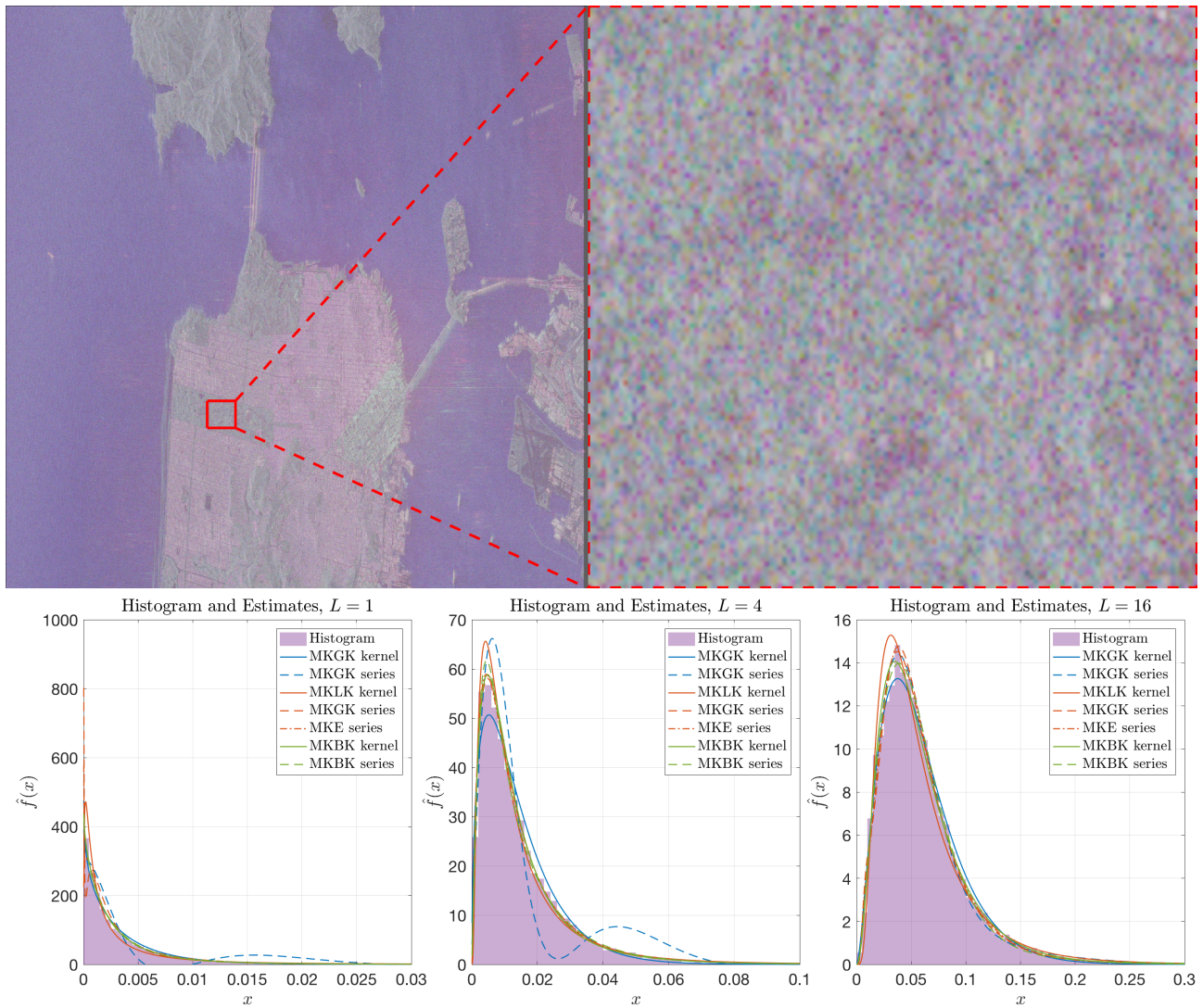


Figure 5.18: The SAR image with the chosen "park" region framed (top left), and an enlarged version of the "park" region (top right). The entire image is 2,800 by 2,800 pixels, with the "park" region consisting of almost 18,000 pixels. Bottom: Histograms of all green channel intensities within the "park" region, and the MK series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$. We included the raw single-look data (left), and averaged data corresponding to $L = 4$ (center) and $L = 16$ (right).

we included the log-normal kernel corrected for logarithmic skewness, which we recall from Section 5.3.3 is a special case of both the MKLK and MKE series, occurring at $N = 3$ in the former. Surprisingly, this series was the top performer for the single look data, improving on the weakness the expansions of $\Lambda(x; \mu, \sigma)$ had for $L = 1$ in Table 5.4.

Other than that, we can make many of the same observations as with Table 5.4, e.g. higher numbers of data points justifies more correcting terms, the MKGG series struggles to improve on its kernel, and so on. Finally, we note that repeating the experiment using the Bhattacharyya distance gave results which were more favorable towards the MKE series. Specifically, the latter outperformed the MKLK series in all but one case, although the differences were not large. This is in line with previous experiments.

An Urban Region The last region we investigate is an urban area, which in terms of size lies between the first two regions, at just above 80,000 pixels.

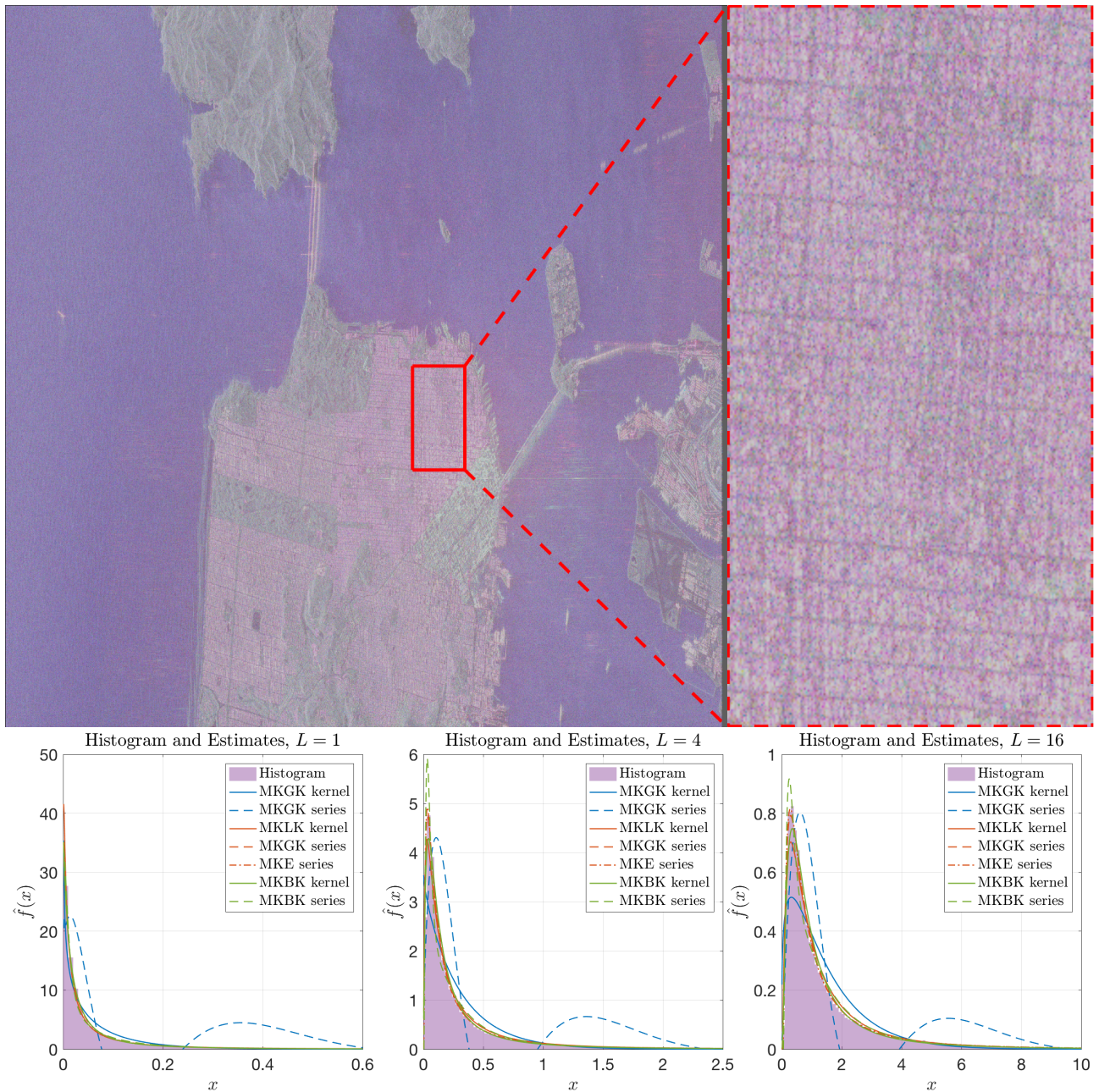


Figure 5.19: The SAR image with the chosen "urban" region framed (top left), and an enlarged version of the "urban" region (top right). The entire image is 2,800 by 2,800 pixels, with the "urban" region consisting of more than 80,000 pixels. Bottom: Histograms of all red channel intensities within the "urban" region, and the MK series expansions correcting for $\Delta\kappa_3$ and $\Delta\kappa_4$. We included the raw single-look data (left), and averaged data corresponding to $L = 4$ (center) and $L = 16$ (right).

Figure 5.19 gives an overview of the "urban" region we chose. We can see that the area is quite heterogeneous, but belonging to a single class. The only distinguishable features are the roads, which manifest as a mesh, and some pockets of green, presumably from small local parks. The histograms indicate that this is a more challenging case than the two previous ones. We could perhaps have expected this, as [Frery et al., 1997] associated urban areas with extremely heterogeneous data, which is modeled with low values of the parameter M . Throughout this chapter we have seen that, all else equal, lower values of M reduced the accuracies of the MK series expansion, e.g. in Table 5.2.

Table 5.6: The MK series expansions fitted to the red channel intensities in the "urban" region of our SAR image, Kullback-Leibler distance to a kernel density estimate based on the entire region. The kernels are tailored and the series expansions correct for $\Delta\kappa_3$ and $\Delta\kappa_4$ (with an explicit exception), mean of 100 iterations where 100, 1,000 or 5,000 data points were drawn at random, best method in bold.

Method\Data Points	$L = 1$			$L = 4$			$L = 16$		
	100	1,000	5,000	100	1,000	5,000	100	1,000	5,000
MKGK kernel	$8.32 \cdot 10^{-2}$	$6.96 \cdot 10^{-2}$	$6.83 \cdot 10^{-2}$	$1.79 \cdot 10^{-1}$	$1.62 \cdot 10^{-1}$	$1.61 \cdot 10^{-1}$	$1.44 \cdot 10^{-1}$	$1.31 \cdot 10^{-1}$	$1.30 \cdot 10^{-1}$
MKGK series	$3.83 \cdot 10^0$	$3.73 \cdot 10^0$	$3.73 \cdot 10^0$	$3.39 \cdot 10^0$	$3.32 \cdot 10^0$	$3.32 \cdot 10^0$	$2.91 \cdot 10^0$	$2.88 \cdot 10^0$	$2.88 \cdot 10^0$
MKLK kernel	$1.47 \cdot 10^{-2}$	$5.08 \cdot 10^{-3}$	$4.58 \cdot 10^{-3}$	$2.19 \cdot 10^{-2}$	$1.32 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$2.30 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$
MKLK series, $N = 3$	$2.65 \cdot 10^{-2}$	$4.71 \cdot 10^{-3}$	$4.02 \cdot 10^{-3}$	$2.32 \cdot 10^{-2}$	$6.75 \cdot 10^{-3}$	$5.79 \cdot 10^{-3}$	$1.73 \cdot 10^{-2}$	$3.16 \cdot 10^{-3}$	$2.15 \cdot 10^{-3}$
MKLK series	$7.85 \cdot 10^{-2}$	$5.57 \cdot 10^{-3}$	$3.13 \cdot 10^{-3}$	$2.40 \cdot 10^{-2}$	$6.44 \cdot 10^{-3}$	$5.37 \cdot 10^{-3}$	$1.90 \cdot 10^{-2}$	$3.19 \cdot 10^{-3}$	$2.18 \cdot 10^{-3}$
MKE series	$4.98 \cdot 10^{-2}$	$4.54 \cdot 10^{-3}$	$2.62 \cdot 10^{-3}$	$3.82 \cdot 10^{-2}$	$1.12 \cdot 10^{-2}$	$9.75 \cdot 10^{-3}$	$2.83 \cdot 10^{-2}$	$5.16 \cdot 10^{-3}$	$3.58 \cdot 10^{-3}$

Table 5.6 contains the results of our experiment on the "urban" region. The number of data points and iterations was again dictated by the total number of pixels in the region. We chose to omit the MKBK kernel and series again, in order to compare only the simpler, faster methods. It should be noted that that beta prime kernel provided the best PDF estimate in only four cases (two when using the Bhattacharyya distance), and was then only marginally better than the second best method.³⁹

Regardless of the dissimilarity measure used, the MKLK and MKE series matched each other in terms of the results. Again we recognize the trends that more data and higher values of L justify more correcting terms.

5.4.2 Modeling in a SAR Change Detection Scenario

We now continue to explore the use of the MKBK series in SAR change detection, building on the experiment of Section 5.3.5. We use the regions we picked out in the previous section, focusing on the cases $L = 4$ and $L = 16$. We omitted the case $L = 1$ as change detection on single look data was simply too difficult to model with these methods.⁴⁰ For each region, we split the pixels into two groups, shuffled them and created quotient data points, simulating the ratio of two observations of unchanged regions.⁴¹ First we present histograms with fitted MKBK series expansions with varying numbers of correcting terms. Then we repeat those experiments 100 times, using smaller samples, and tabulate the mean distances to the target PDF. The latter is again substituted by kernel density estimates based on the entire region.

As in Section 5.3.5, we use the MKBK series under the assumptions that the shapes are equal, i.e. $a_1 = a_2 = L$. This is completely true in the present experiment, since we are examining quotients of pixels from the same region of a single image. We also take into account eqs. (5.26) and (5.29), where we found that $b = 1$ and $\Delta\kappa_n = 0$ for n odd. Recall that this resulted in the highly simplified MKBK series in eq. (5.28).

Figure 5.20 presents the histograms of the quotient data we produced for $L = 4$ and $L = 16$ in the regions from the previous section. The plots corresponding to the "water" region demonstrates

³⁹The Bhattacharyya distance agrees with $d_{KL}(\cdot)$, in that the log-normal kernel corrected for logarithmic skewness performs very well in this scenario.

⁴⁰We find support for this claim in [Rignot and van Zyl, 1993], where the authors comment that the intensity quotients works best with large L . In the same article they discuss a different technique which is suitable to single look data.

⁴¹Naturally, realistic applications would have to deal with the fact that two distinct acquisitions usually do not have perfectly identical statistical properties. That is, the images might be taken from slightly different angles, vegetation is subject to seasonal changes and so on. Accounting for these effects by comparing two distinct images is a natural extension of this experiment, but beyond the scope of this thesis.

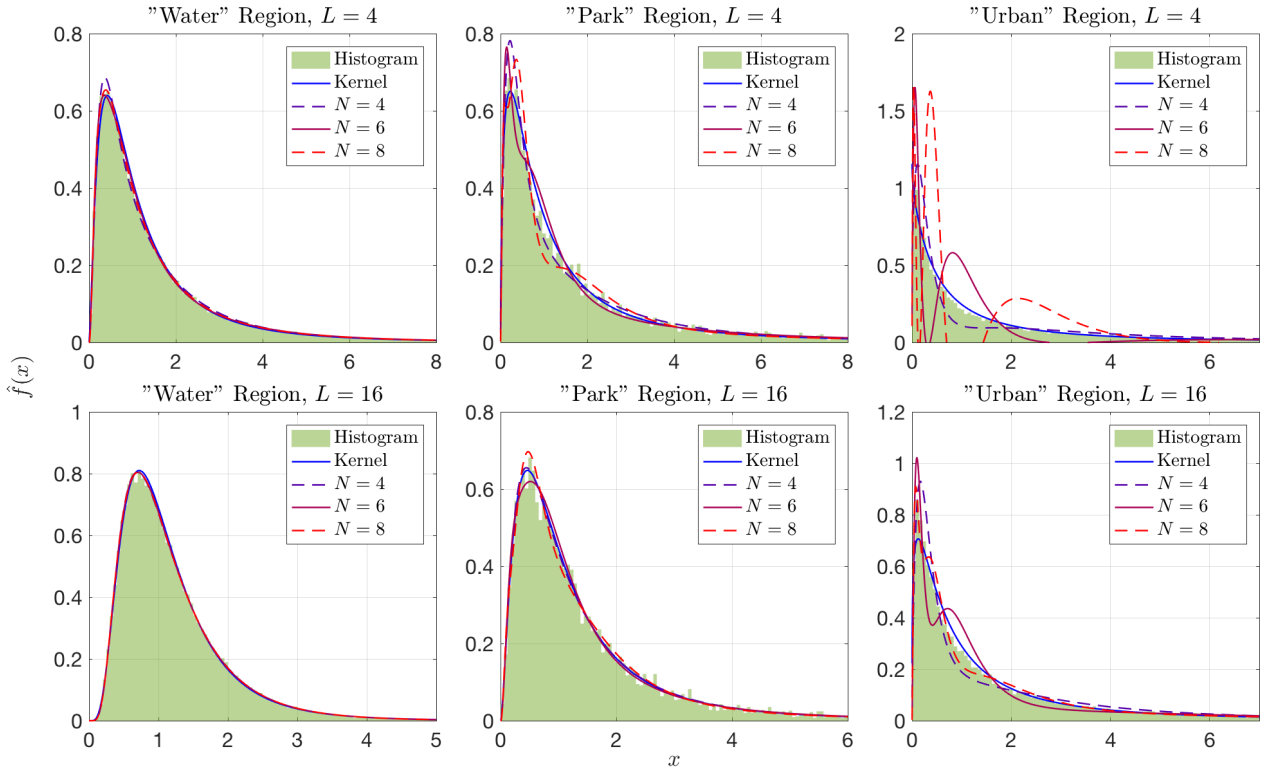


Figure 5.20: Histograms and fitted MKBK series expansions, based on data quotients produced from all pixels in each of the regions presented in Section 5.4.1. The MKBK series expansions are based on the assumptions discussed in Section 5.3.5, which culminated in eq. (5.28).

how well the beta prime kernel fits the data.⁴² In the following, we make the situation more challenging by basing the PDF estimates on smaller samples. Also, our purpose here is to (potentially) validate the approach in ideal situations, leaving it to future work to assess whether the series expansion can correct for effects such as slight changes in image angles and vegetation seasonality.

We recall from the previous section that the "park" region contained significantly fewer pixels. This is clearly apparent in the jaggedness of the histograms in Figure 5.20. This seems to have resulted in more distinct estimates, i.e. larger differences between the kernel, $N = 4$ and so on, at least for $L = 4$. When it comes to the actual PDF estimates, we see that the MKBK series with $N = 4$ and $N = 8$ tries to model the sharper peak of the "park" histogram with $L = 4$,⁴³ while the other corrections are minimal, or even detrimental.

The "urban" region was bigger than the "park" region, but significantly more heterogeneous than either of the other two. This manifested itself in data with very high variance,⁴⁴ and seemed to cause problems for the MKBK series expansion, especially for $L = 4$. Our implementation was questionable for the "urban" region, as it inevitably produced some quotients where one of the pixels corresponds to a street and the other to a building (or a small park). These effects would not be as prevalent in real urban change detection. Besides, urban areas exhibit more deterministic (less stochastic) scattering, and different methods are warranted, namely coherent change detection.

In Figure 5.21, we present the results of an experiment designed to assess the viability of the

⁴²Recall that $\beta(x; L)$ is the distribution followed by the quotient of two IID gamma RVs with shape L .

⁴³The same effect is present in the "water" data with $L = 4$, but to a much smaller degree.

⁴⁴We observed values of $x > 1,000$ for $L = 4$ and $x > 200$ for $L = 16$, but we had to truncate the histograms at $x = 7$ for the sake of clarity.

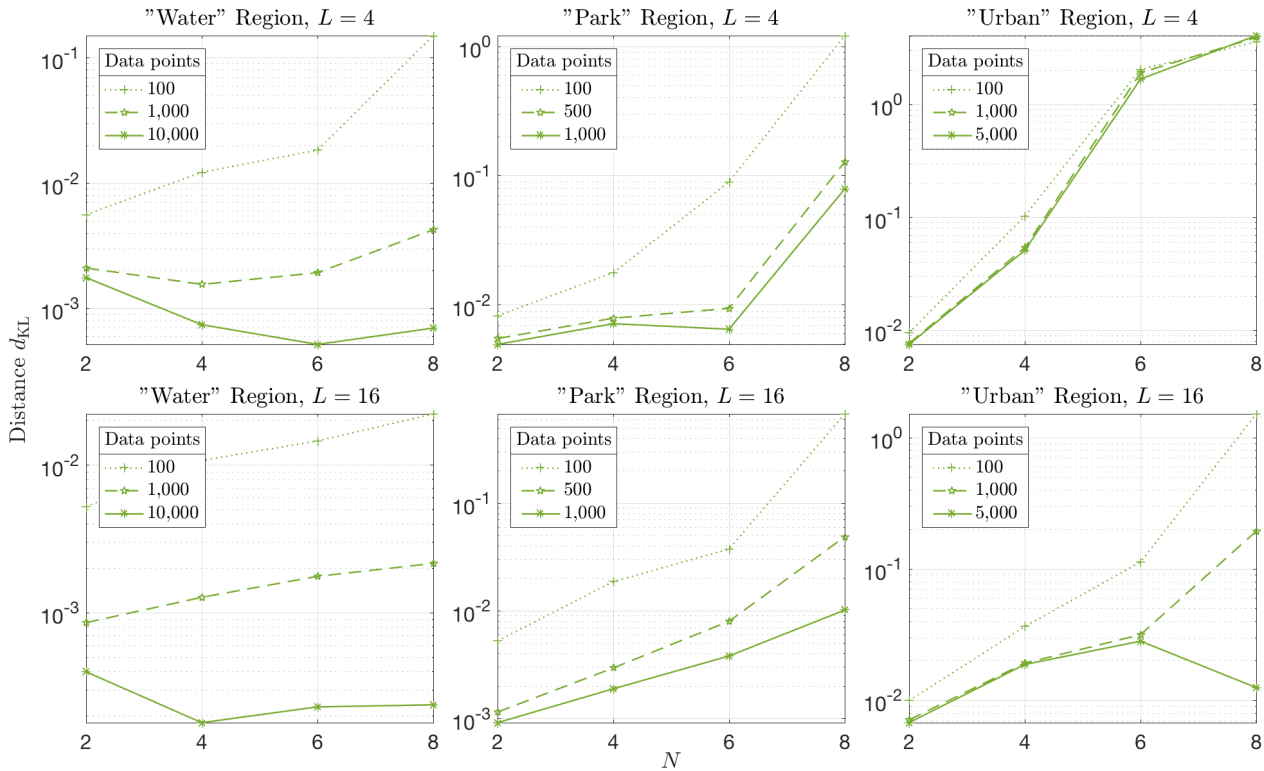


Figure 5.21: The convergence of the MKBK series expansion from eq. (5.28), applied to samples of different sizes drawn from the data sets presented in Figure 5.20. Kullback-Leibler distances, mean of 100 iterations.

one-parameter MKBK series expansions from eq. (5.28). Thus, this is an extension of the experiment from Figure 5.16 to real data. As the number of pixels was essentially halved in the process, we must expect significant correlation in the samples, since the data points are heavily reused.

We see in Figure 5.21 that our approach proved successful for the "water" region, when using the largest sample size of 10,000. Coincidentally, this region also was the least heterogeneous, as discussed in the previous section. Applying the same method to the "park" region was not successful, and comparing the plots between the two regions might indicate that the relatively smaller sample sizes do not justify using any correcting terms. We also experience failure for the "urban" region, in spite of the slightly larger sample sizes. Presumably, this is due to the undesirable effect of our implementation and the extremely heterogeneous nature of urban SAR data, which we discussed above. We base this on our experience throughout this chapter, as we have seen a link between the degree of heterogeneity and the ability of the series expansions to accurately model the data. We saw signs of this already in Section 5.2.6, and the same tendency was present in Sections 5.3.5 and 5.4.1. In the "urban" histograms in Figure 5.20 we were warned that correcting the beta prime kernel would probably be counterproductive.

To summarize, our findings were not unexpected based on our experiments on synthesized data and our impressions from the histograms in Figure 5.20. The single-parameter MKBK series does appear to have merit if the regions are large enough and not too heterogeneous. We have seen three examples which demonstrate both success and failure. 10,000 data points appears to be the point where we can be reasonably hopeful when applying the MKBK series. As the 18,000 pixel "park" region in Figure 5.18 demonstrates, the spatial resolutions of modern sensors are already at a point where this threshold is not unrealistic, with future improvements only working in favor of the series expansions.

Chapter 6

Conclusion

The time has come to conclude. The goals we listed in Section 1.2, have all been accomplished. However, it turned out that the MKE series, presented in [Pastor et al., 2014] outperformed all our novel series expansions. Thus, we could say that the main contributions of this thesis is the MK series expansion framework and the experiments, which strongly supports the theoretical and practical merit of the MKE series in particular.

The design of the experiments in Chapter 5 was based on a conscious choice to cover a wide range of topics in this limited format. Perhaps the biggest compromise was to present only the mean of the distances between the PDF estimates and the true targets in Section 5.3. Naturally, it is relevant whether some methods have larger estimation variances than others, and to determine whether one method was *significantly* better than the others, in the strict, statistical sense of the word. On the other hand, this would have ballooned e.g. Section 5.3.2 to the point where we would have had to omit other sections. The intention was to give the reader an overview of what the MK series expansions *are*, and what they could possibly be used for. From that point of view, the somewhat shallow nature of some of the experiments was deemed necessary.

The experiments were also designed such as to demonstrate both where the MK series expansion fail and where they succeed. For example, if a series expansion always converged, we reduced the number of data points or tweaked the target distribution to the point where we could demonstrate failure also.

This chapter is organized as follows. We present our conclusions on the theoretical aspects of this thesis first, and afterwards we turn to the experimental aspects. We end this chapter with some suggestions for future work.

6.1 Conclusions about the Theoretical Contributions

6.1.1 The MK Series Expansion Framework

The central topic for this thesis was the framework for series expansions of PDFs using MK statistics, and the entirety of Chapter 3 was devoted to this subject.

The most concise way to summarize the MK series expansions framework would be to say that it is an almost perfect logarithmic scale analogy of the classical series expansions. The moments and cumulants are replaced by their logarithmic counterparts, the derivative operator D_x is replaced by $D_x x$, which is closely related to $x D_x = D_{\log x}$, the Gaussian kernel is replaced by the log-normal kernel to arrive at the MKE series, and $H_n(x)$ is replaced by $H_n(\log x)$.

By introducing a full framework for the MK series expansions, we will hopefully draw more attention to the MKE series, which was introduced in [Pastor et al., 2014]. We have also shown that the MKE can be derived using the inverse MT, i.e. wholly inside the realm of MK statistics.

We laid the foundation for the MK series expansion in Section 3.1.3, by using the arbitrary kernel $\rho(x)$. We introduced $\mathcal{P}_n(x)$ as the functions implicitly defined by applying the Mellin derivative to the arbitrary kernel, thereby giving expressions we could evaluate for the series expansions. We derived a recursive formula for $\mathcal{P}_{n+1}(x)$, which turned out to be useful when we later derived specific series expansions.

Then, we replaced $\rho(x)$ with $\gamma(x)$, $\Lambda(x)$, and $\beta'(x)$ to provide relevant examples of the series expansions. The MKGK series showed how a series expansion of the gamma kernel was made possible by MK statistics.¹ The MKLK series demonstrated a strong analogy to the classical Gram-Charlier Gaussian kernel series, and serves as a useful intermediate step when trying to understand the MKE series. The MKBK series is a proof of concept of a three-parameter kernel, since there are numerous potential kernels with three parameters (or more). The MKBK series was also reduced to two or even a single parameter in a practical scenario.

In each of the series expansions, we evaluated the corresponding special cases of $\mathcal{P}_n(x)$. For the MKGK series, we named them $M_n(x)$. We proved that $M_n(x)$ is an n th degree polynomial in x with leading coefficient 1, and how they are linear combinations of the generalized Laguerre polynomials and a type of confluent hypergeometric functions.² For the MKLK series, we showed how that specific instance of $\mathcal{P}_n(x)$ is actually the well-known Hermite polynomial, but with $\log x$ replacing x in the argument. For the MKBK series, we named them $M'_n(\cdot)$, and showed that they are n th degree polynomials in $bx/(1 + bx)$.

To derive this framework, we had to employ what we called the Mellin derivatives $D_x x$ and $x D_x = D_{\log x}$. Based on the current literature, it appears that these derivative operators have not been used in the context of MK statistics before.

6.1.2 Use of the Bell Polynomials

Throughout Chapter 3, we used the Bell polynomials $B_n(x)$ to represent combinatoric relationships in the series expansions, which arises due to the very nature of the log-moments and log-cumulants. This allowed us to present concise and explicit expressions for the series expansions. Without the use of $B_n(x)$, we would have had to resort to implicit definitions such as stating “we collect the terms of like power in $D_x x$ ”, or by using sets. The most significant simplification was seen in the MKE series, where the Bell polynomials were used to represent more complex combinatorics than in the other series expansions.

These findings are also directly applicable to the classical Gram-Charlier and Edgeworth series expansions. Using the Bell polynomials in these classical methods was the basis of Chapter 4. The result is what appears to be the first explicit definition of Gram-Charlier and Edgeworth series. As these classical methods are applicable for nearly Gaussian data, which is a much more common situation than the non-negative case, this contribution might eventually prove to be the most significant of this thesis.

¹Recall that the classical Gram-Charlier gamma kernel series of Section 2.4.4 did not actually expand $\gamma(x)$, but used $x^n \gamma(x)$ as the kernel for the n th term.

²Some of these proofs and relationships were placed in the appendix.

6.1.3 Other Contributions to the Classical Series Expansions

In addition to the introduction of the Bell polynomials, we made some smaller contributions w.r.t. the classical series expansion methods. In Chapter 4, we derived the Gram-Charlier series with an arbitrary kernel, and subsequently the Edgeworth series. This derivation was different from other approaches in the literature, not only in its use of $B_n(x)$, but also in the sense that we emphasized the most general situation first. Then, the derivation branched out to the Gram-Charlier Gaussian kernel series and the Edgeworth series. This clarifies how the two series are in fact very similar, and a direct consequence of choosing the Gaussian kernel.

For the classical series expansion of the gamma kernel, we pointed out that the confluent hypergeometric function of the first kind allows for a much simpler expression for the coefficients of the series.

6.2 Conclusions from the Experiments

In the paper which is appended to this document, we compare the MKGK, MKLK, and MKE series expansions with competing methods, in experiments which are complimentary to those included in this thesis. As the MKE series was recently discovered, the experiments in that paper and this thesis constitutes a significant contribution to understanding the performance of that series, while the other MK series expansions have not been assessed before at all.

We let the target PDFs be known at first (the approximation setting), and then unknown (the estimation setting). As we expected beforehand, the convergence was degraded when the information about the target PDFs was withheld from the methods, and more so when the number of data points was low. Recall that our quantitative performance assessments disregard the tails of the distribution, as we qualitatively concluded that the MK series expansion methods are unsuited for this purpose.³

In general, our experiments promote the MKE series from [Pastor et al., 2014] as the standout performer, not least due to its flexibility in modeling all the different target distributions we have tested the series expansions on. The fact that expansions of the log-normal kernel outperform expansions of the gamma kernel for target PDFs which are so closely related to $\gamma(\cdot)$, is perhaps counterintuitive. We feel that this is due to two attributes of $\Lambda(x; \mu, \sigma)$: The log-cumulants, listed in eq. (2.80), are the simplest possible for any two-parameter PDF, with the key property that $\kappa_{\Lambda, n} = 0 \forall n \geq 3$. This entails that the log-cumulant differences in the series expansions are reduced to the target log-cumulants, which appears to be significant. The other factor is that the MoLC estimates of μ and σ^2 coincide with the ML estimates. They are also the minimum variance unbiased estimates, barring a practically insignificant correction to the log-variance estimate. Support for this claim is found in Figures 5.6 through 5.10, where the MoLC estimates were always clearly the best parameter values in the MKLK and MKE series, with a lot more ambiguity in the MKGK and MKBK series.

6.2.1 The MKLK and MKE Series

When the number of data points was low, corrections beyond $\Delta\kappa_3$ were rarely warranted. Correcting only for $\Delta\kappa_3$, as in eq. (5.20), the MKLK and MKE series coincide at what we called the log-normal kernel corrected for the logarithmic skewness.

³This mirrors their classical counterparts, see [Blinnikov and Moessner, 1998]. To recapitulate, we chose to compare only the regions of x where $f(x)$ was above 1/1000th of its maximum value. Different threshold values would have given different results, but additional testing verified that this effect was not too significant.

When the number of data points was high enough that further corrections improved the PDF estimates, the MKE series outshone the MKLK series. This mirrors the performance of the classical Gram-Charlier and Edgeworth series expansions of the Gaussian kernel. We can postulate that also the roles of the MKLK and MKE series will come to match their classical counterparts, with the MKLK series mainly serving as an intermediate step in the derivation of, and way of better understanding, the MKE series.

We have validated the use of the MKE series expansion on real SAR data. Except when the number of data points was very low, the accuracy was usually better than the gamma kernel and MKGK series, competing with the computationally much more demanding fitted beta prime distribution.

6.2.2 The MKGK Series

The initial results of the MKGK series was quite frankly disappointing, both in terms of general accuracy and convergence properties. However, experimenting with non-tailored kernel parameters revealed that alternative parameter choices were associated with higher accuracy, often comparable to that of the MKLK series.

In Sections 5.2.6 and 5.3.4 we explored a SAR-specific scenario. The gamma kernel parameters were assigned values corresponding to physical quantities,⁴ and the MKGK series was tasked with modeling another physical process which affects the measurements. In practice, L would be a highly accurate global estimate, and for each class we wish to model, we only need to estimate the mean m . As we saw in Section 5.3.4, we can then improve the resulting estimates by correcting for $\Delta\kappa_1$ and $\Delta\kappa_2$. Based on Figure 5.15, this approach might even exceed the accuracy of the much slower method of fitting both parameters freely to the data from each class.

In other words, these results were encouraging, indicating that the MKGK series could find itself useful in specific problems. This was in stark contrast to the results of simply fitting the MKGK series with tailored kernel to SAR data, which was unsuccessful.

6.2.3 The MKBK Series

At one point in Section 5.3.3, we said that the three-parameter beta prime kernel was not suited for a series expansion in most real-world applications. This was based on the failure of the series to improve on the kernel. We suspect that this is due to the kernel itself being so complex (i.e. based on three estimated parameters), that the additional model complexity of correcting terms is overshadowed by the error resulting from estimating so many quantities.

However, the MKBK series was more successful in the application to SAR change detection in Sections 5.3.5 and 5.4.2. With a few assumptions, the beta prime kernel was reduced to two or even a single parameter, eliminating the need for a time consuming numerical solution of a set of equations to find the parameter estimates. The performance was in some cases encouraging, provided that the area was not too heterogenous and that a sufficient number of data points were available.

6.2.4 Modeling Real Data

Throughout Chapter 5, we saw how the MK series expansions required relatively large amounts of data to incur improvement. A rule of thumb could be that with approximately 10,000 data

⁴As opposed to the values imposed by the constraint $\Delta\kappa_1 = \Delta\kappa_2 = 0$

points we can begin to expect better accuracy from the series expansions than their kernels. As we saw when choosing the "park" region in Figure 5.18, modern sensors are capable of producing SAR images with large number of pixels for relatively small areas.

In more general terms: As the technological development continues, the MK series expansions will become more relevant, due to the ever increasing number of data points. We know that this increase permits more complex models, but fitting a three-parameter PDF is usually extremely demanding compared to two-parameter PDFs.⁵ The series expansions, on the other hand, can utilize the information from higher order log-cumulants in a very efficient manner. That is, the series expansion methods allows us to use more complex models, without a sudden spike in computational requirements. We saw this happening in our experiments when the MKE series approaches the accuracy of fitted three-parameter distributions which require exponentially time.

6.3 Future Work

We end this chapter with some suggestions for future work based on the subjects in this thesis.

- Expand the synthesized and real data experiments of Chapter 5. Specifically, an analysis which includes the variance of the estimation errors. This permits an assessment of whether one method is significantly better than the others for certain target PDFs, in the strict, statistical sense of the word. Another factor is the computation time associated with each method, which this thesis discussed only in very broad strokes. Also, target PDFs included in this thesis are naturally far from exhaustive, and PDFs from other fields should be evaluated, e.g. data which is nearly log-normal.
- Further explore use of the MKGK and MKBK series in SAR-specific scenarios, as in Sections 5.2.6, 5.3.4, and 5.3.5.
- Insert other kernels in both the classical and MK series expansion with arbitrary kernel. Candidates for an MK series expansion are e.g. the inverse gamma or beta PDFs.⁶
- Expand the MK series expansion framework w.r.t. log-cumulant diagrams. In the literature on MK statistics, distributions are often arranged according to their positions in a log-cumulant diagram, see [Nicolas, 2002], [Anfinsen and Eltoft, 2011], [Anfinsen et al., 2011], [Bombrun et al., 2011], [Deng et al., 2016], and [Deng and López-Martínez, 2016]. It would be interesting to see if the location of the data in the diagram might predict the performance of the various series expansion methods. Also, the method of parameter estimation in [Anfinsen et al., 2011] is an interesting alternative, at least for the MKGK series.
- Extend the MK series expansion framework to matrix-variate RVs and data. A natural starting point for this would be [Anfinsen and Eltoft, 2011].
- Conduct a mathematical analysis of the error bounds of the MK series expansions. This would presumably follow a similar path to the corresponding work on the classical series, including [Esseen, 1945] and [Hsu, 1945], which was summarized in [Wallace, 1958].
- Attempt to apply the Bell-polynomials to the classical Cornish-Fisher expansion, which was presented in [Cornish and Fisher, 1938] and generalized in [Hill and Davis, 1968]. The Cornish-Fisher expansion is well suited to modeling the tails of distributions, an attractive

⁵Estimating three-parameter PDFs, such as $K(\cdot)$ and $\beta'(\cdot)$, often requires solving two equations simultaneously with a numerical iterative procedure.

⁶This is not *not* the same as the beta prime kernel, but a closely related PDF.

property when trying to detect objects or phenomena which manifests as extreme values of X . This expansion has already been extended to MK statistics in [Pastor et al., 2016], but use of $B_n(x)$ (or other polynomials) could perhaps simplify the rather complicated derivation and expression.⁷

- Explore the relationship between $M'_n(x)$ and the Jacobi Polynomials. A classical Gram-Charlier series of the beta kernel using the Jacobi polynomials can be found in e.g. [Durbin and Watson, 1951]. Since we have proven the relationship between the $M_n(x)$ polynomials used in the MKGK series, and the Laguerre polynomials used in the classical Gram-Charlier gamma kernel series, it is natural to ask whether there exists a relationship between $M'_n(x)$ and the Jacobi polynomials. Alternatively, a MK series expansion of the beta kernel might result in $\mathcal{P}_n(x)$ which are somehow related to the Jacobi polynomials.
- Investigate the properties of the functions $\mathcal{P}_n(x)$, which were introduced in Section 3.1.4. What types of kernels $\rho(x)$ result in $\mathcal{P}_n(x)$ which are n th degree polynomials of (functions of) x ? What does this tell us about which kernels are possible or even good candidates for the MK series expansion? This can also be done for the corresponding classical case in Section 4.2.1.

⁷[Kendall et al., 1994] has an alternative derivation of the Cornish-Fisher expansion, which starts with the Edgeworth series expansion. This can provide a simpler avenue for applying the Bell polynomials, compared to using the original work in [Hill and Davis, 1968].

Appendix A

Supplementary Theoretical Results

A.1 The Laguerre Polynomials Scaled

A.1.1 Orthogonality

It will be shown here how the Laguerre polynomials' orthogonality property can be extended to the two-parameter gamma PDF $\gamma(x; a, b)$ in eq. (2.57). Substituting $x \rightarrow bx$ in eq. (2.122) shows how scaling x also in $L_n^{(a-1)}(x)$ retains the property of the unscaled functions, namely

$$\int_0^{\infty} L_n^{(a-1)}(bx) L_k^{(a-1)}(bx) b^{a-1} x^{a-1} e^{-bx} b dx = \Gamma(a) \binom{n+a-1}{n} \delta_{nk} \quad (\text{A.1})$$

$$\int_0^{\infty} L_n^{(a-1)}(bx) L_k^{(a-1)}(bx) \frac{b^a x^{a-1}}{\Gamma(a)} e^{-bx} dx = \binom{n+a-1}{n} \delta_{nk} \quad (\text{A.2})$$

$$\int_0^{\infty} L_n^{(a-1)}(bx) L_k^{(a-1)}(bx) \gamma(x; a, b) dx = \binom{n+a-1}{n} \delta_{nk}, \quad (\text{A.3})$$

where $d(bx) = b dx$ was used.

A.1.2 Rodrigues Formula

We want to scale the argument of the Rodrigues formula by again substituting $u = bx$. The effect on the differential operator becomes

$$D_u^n = \frac{d^n}{du^n} = \frac{1}{b^n} \frac{d^n}{dx^n} = b^{-n} D_x^n. \quad (\text{A.4})$$

Thus, by substitution into eq. (2.113) we obtain

$$L_n^{(a-1)}(bx) = \frac{1}{n!} b^{-a+1} x^{-a+1} e^{bx} b^{-n} D_x^n [b^{a-1+n} x^{a-1+n} e^{-bx}], \quad (\text{A.5})$$

where the powers of b cancel, giving

$$L_n^{(a-1)}(bx) x^{a-1} e^{-bx} = \frac{1}{n!} D_x^n [x^n \cdot x^{a-1} e^{-bx}]. \quad (\text{A.6})$$

By multiplying both sides with $b^a/\Gamma(a)$, which is constant w.r.t. x , this can be written as

$$L_n^{(a-1)}(bx) \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)} = \frac{1}{n!} D_x^n \left[x^n \cdot \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)} \right], \quad (\text{A.7})$$

and if we restrict ourselves to $x \geq 0$, we can recognize that

$$L_n^{(a-1)}(bx) \gamma(x; a, b) = \frac{1}{n!} D_x^n [x^n \gamma(x; a, b)]. \quad (\text{A.8})$$

A.2 Observations on $M_n(x)$

A.2.1 $M_n(x)$ as Linear Combinations of Laguerre Polynomials

Lemma 4 *The polynomials $M_n(x)$ defined in eq. (3.20) are linear combinations of the generalized Laguerre polynomials, i.e.*

$$M_n(x) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k k! L_k^{(a-1)}(x), \quad (\text{A.9})$$

where $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ denotes the Stirling numbers.

Proof First, an identity regarding $D^k x^k f(x)$ for an arbitrary function $f(x)$ is proven. By applying the derivative operator and the product rule of differentiation, we see that

$$D^k x^k f(x) = D^{k-1} [k x^{k-1} f(x) + x^k D f(x)] = D^{k-1} x^{k-1} [(k + xD) f(x)]. \quad (\text{A.10})$$

This can be repeated, most easily by evaluating $g(x) = (k + xD)f(x)$, giving

$$\begin{aligned} D^k x^k f(x) &= D^{k-2} x^{k-2} [(k-1 + xD)g(x)] \\ &= D^{k-2} x^{k-2} [(k-1 + xD)(k + xD)f(x)]. \end{aligned} \quad (\text{A.11})$$

The general formula is now apparent as

$$D^k x^k f(x) = \left[\prod_{i=1}^k (xD + i) \right] f(x). \quad (\text{A.12})$$

A useful identity is that

$$(Dx - xD)f(x) = Dx f(x) - xD f(x) = f(x) + xD f(x) - xD f(x) = f(x), \quad (\text{A.13})$$

where the product rule was used again. It can be formulated without explicitly including $f(x)$, that is

$$Dx - xD = 1. \quad (\text{A.14})$$

This identity gives an alternative version of eq. (A.12) by shifting the index

$$D^k x^k f(x) = \left[\prod_{i=0}^{k-1} ((xD + 1) + i) \right] f(x) = \left[\prod_{i=0}^{k-1} (Dx + i) \right] f(x). \quad (\text{A.15})$$

Now the proof of lemma 4 can be completed. Using eq. (2.160) on $(-Dx)^n$ gives

$$(-Dx)^n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \prod_{i=0}^{k-1} (-Dx - i). \quad (\text{A.16})$$

Note that for any sequence of coefficients $\xi_0, \xi_1, \dots, \xi_{k-1}$, we have

$$\prod_{i=0}^{k-1} (-\xi_i) = (-1)^k \prod_{i=0}^{k-1} (\xi_i), \quad (\text{A.17})$$

that is, the sign can be separated from the factors. Applying this to eq. (A.16) gives

$$(-Dx)^n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k \prod_{i=0}^{k-1} (Dx + i), \quad (\text{A.18})$$

where the expression from eq. (A.15) is recognized to give

$$(-Dx)^n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k D^k x^k. \quad (\text{A.19})$$

Multiplying both sides with the gamma kernel $\gamma(x)$ completes the proof by using the definitions of $M_n(x)$ and $L_n^{(a-1)}(x)$ from eqs. (3.19) and (2.115)

$$(-Dx)^n \gamma(x; a) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k D^k x^k \gamma(x; a) \quad (\text{A.20})$$

$$M_n(x) \gamma(x; a) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k L_k^{(a-1)}(x) \gamma(x; a) \quad (\text{A.21})$$

$$M_n(x) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k k! L_k^{(a-1)}(x) \quad \text{QED.} \quad (\text{A.22})$$

A.2.2 The Leading Coefficient of $M_n(x)$

Lemma 5 $M_n(x)$ has leading coefficient $1 \forall n$.

Proof The only term containing x^n in eq. (A.9) is $(-1)^n/n! L_n^{(a-1)}(x)$, since $L_k^{(a-1)}(x)$ is a k th degree polynomials.¹ The leading coefficient of $L_k^{(a-1)}(x)$ is $(-1)^k/k!$ giving $M_n(x)$ leading coefficient

$$\left\{ \begin{matrix} n \\ n \end{matrix} \right\} (-1)^n n! \frac{(-1)^n}{n!} = 1, \quad (\text{A.23})$$

where it was used that $\left\{ \begin{matrix} n \\ n \end{matrix} \right\} = 1 \forall n$. Note that the Hermite polynomials in the classical Edgeworth series are also defined to have leading coefficient 1.

A.2.3 $M_n(x)$ and a Confluent Hypergeometric Function

Using the notation from [Daalhuis, 2010], Tricomi's confluent hypergeometric function² $U(a, b, x)$ has a close relationship with the $M_n(x)$ polynomials, namely

$$U(-n, a, x) = (-1)^n n! L_n^{(a-1)}(x), \quad (\text{A.24})$$

¹This is an alternative way of proving that $M_n(x)$ is an n th degree polynomial, but it is less general than our proof in Section 3.2.2, which we recall extends to all cases where $\mathcal{P}_1(x)$ are first degree polynomials in x .

²This is *not* the same confluent hypergeometric function ${}_1F_1(a; b; x)$ which was used in Section 4.2.3. $U(a, b, x)$ was introduced in [Tricomi, 1947].

where n must be a non-negative integer when used in the subscript of $L_n^{(a-1)}(x)$. Introducing the function $U(-n, a, x)$ slightly simplifies much of the preceding derivation, e.g. eq. (A.9), which is reduced to

$$M_n(x) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} U(-k, a, x). \quad (\text{A.25})$$

However, this was deemed potentially confusing and is only mentioned here in case there are readers who are already familiar with the confluent hypergeometric functions or are particularly interested.

A.3 The Logarithmic Hermite Polynomials

In Section 3.3 we derived $(-Dx)^n \Lambda(x) = (-Dx)^n \Lambda(x)$ first for the standardized kernel, then generalizing that result to general log-mean and log-variance. Here we will derive the result for the non-standardized $\Lambda(x; \mu, \sigma)$ directly, and we start by letting

$$u = \frac{\log x - \mu}{\sigma}, \quad (\text{A.26})$$

which gives

$$\frac{du}{d \log x} = \frac{1}{\sigma} \Rightarrow \frac{d}{du} = \frac{d}{d \log x} \frac{d \log x}{du} = \frac{d}{d \log x} \sigma, \quad (\text{A.27})$$

or, using different notation and the result in eq. (3.52),

$$D_u = \sigma(xD_x). \quad (\text{A.28})$$

The definition of the Hermite polynomials in eq. (2.82) gives

$$(-D_u)^n \alpha(u) = H_n(u) \alpha(u), \quad (\text{A.29})$$

and as shown above, D_u can be substituted along with u to give

$$\sigma^n (-xD_x)^n \alpha\left(\frac{\log x - \mu}{\sigma}\right) = H_n\left(\frac{\log x - \mu}{\sigma}\right) \alpha\left(\frac{\log x - \mu}{\sigma}\right), \quad (\text{A.30})$$

where $u = (\log x - \mu)/\sigma$ was also used. The standardized Gaussian kernel is substituted with the non-standardized version using eq. (2.24)

$$\sigma^n (-xD_x)^n \frac{1}{\sigma} \alpha(\log x; \mu, \sigma) = H_n\left(\frac{\log x - \mu}{\sigma}\right) \frac{1}{\sigma} \alpha(\log x; \mu, \sigma), \quad (\text{A.31})$$

where the factors σ^{-1} cancel. The relationship between the non-standardized version of the log-normal and Gaussian kernels is the same as for the standardized versions, i.e. $\Lambda(x; \mu, \sigma) = x^{-1} \alpha(\log x; \mu, \sigma)$ as stated in eq. (2.73). Thus, the result in eq. (3.51) is applied in the same way as in the standardized situation, now giving

$$x \sigma^n (-D_x)^n \Lambda(x; \mu, \sigma) = H_n\left(\frac{\log x - \mu}{\sigma}\right) \alpha(\log x; \mu, \sigma) \quad (\text{A.32})$$

$$(-D_x)^n \Lambda(x; \mu, \sigma) = \frac{1}{\sigma^n} H_n\left(\frac{\log x - \mu}{\sigma}\right) \frac{1}{x} \alpha(\log x; \mu, \sigma) = \frac{1}{\sigma^n} H_n\left(\frac{\log x - \mu}{\sigma}\right) \Lambda(x; \mu, \sigma). \quad (\text{A.33})$$

This derivation was done directly in the sense that the result for the standardized case in eq. (3.50) was *not* used. Only the results in eqs. (3.51) and (3.52) were necessary to directly derive the result for non-standardized $\Lambda(x; \mu, \sigma)$.

A.4 Program Code

The program code use to produce the figures and tables in Chapter 5, has been made available online at <https://github.com/TorgeirBrenn/Mellin-Kind-Series-Expansions/>. The program used was MATLAB, with one script for each experiment, designed to create the desired figure or table cells. These script are perhaps not particularly interesting, but a lot of the functionality was separated into custom functions. These functions can prove useful for others, and they are listed with a short description below. All functions were written by me, Torgeir Brenn, except where otherwise stated.

AVGFILTER.M	A simple averaging filter.
BETAPRIMEPDF.M	The beta prime distribution PDF.
BHATTADIST.M	The Bhattacharyya distance.
BPOLY.M	The Bell polynomials.
EMPLC.M	The empirical log-cumulants.
G0PDF.M	The G^0 distribution PDF.
G0RND.M	G^0 random number generator.
GCGAMFIT.M	Fitting the Gram-Charlier gamma kernel series.
GGDPDF.M	The GFD distribution PDF.
GGDRND.M	GFD random number generator.
HERMITEFAST.M	The Hermite polynomials, optimized.
HERMITEHPROB.M	The Hermite polynomials.
HERMOUTPROD.M	The Hermitian outer product.
IGAMPDF.M	The inverse gamma distribution PDF.
IGAMRND.M	Inverse gamma random number generator.
IMAGEES.M	Stretching suitable for SAR images.
KDISTFIT.M*	Fitting the K distribution.
KDISTFITEXACT.M**	Fitting the K distribution to a known target.
KLDIST.M	The Kullback-Leibler distance.
KPDF.M	The K distribution PDF.
KRND.M	K random number generator.
LAGUERREFAST.M	The Laguerre polynomials, optimized.
MKBKFIT.M**	Fitting the MKBK series.
MKBKFITEQUALSHAPES.M	Fitting the MKBK series with equal shape parameters.
MKBKFITEXACT.M**	Fitting the MKBK series to a known target.
MKBKFITGIVENPARAM.M	Fitting the MKBK series with given parameters.
MKEFIT.M	Fitting the MKE series.

MKEFITEXACT.M	Fitting the MKE series to a known target.
MKGKFIT.M	Fitting the MKGK series.
MKGKFITEXACT.M	Fitting the MKGK series to a known target.
MKGKFITGIVENPARAM.M	Fitting the MKGK series with given parameters.
MKLKFIT.M	Fitting the MKLK series.
MKLKFITEXACT.M	Fitting the MKLK series to a known target.
MKLKFITGIVENPARAM.M	Fitting the MKLK series with given parameters.
MLGAMFIT.M	Fitting the gamma distribution using the ML estimates.
MOLCGAMFIT.M	Fitting the gamma distribution using the MoLC estimates.
MOLCGGDFIT.M	Fitting the GFD using the MoLC estimates.
MOLCGGDFITEXACT.M	Fitting the GFD to a known distribution using the MoLC estimates.
MPOLY.M	The $M_n(\cdot)$ polynomials.
MPRIMEPOLY.M	The $M'_n(\cdot)$ polynomials.
NORMALIZEPDF.M	Ensures that a PDF is positive and integrates to unity.
PAULIBASIS.M	Transforms SAR data to the Pauli basis.
STIRLING2.M	The Stirling numbers.
TIMAGE.M	Displays a polarimetric SAR coherency image.

* Made by Stian N. Anfinsen, based on functions by Olivier Harant and Lionel Bombrun.

** Made by Torgeir Brenn, heavily based on functions by Stian N. Anfinsen, Olivier Harant, and Lionel Bombrun.

Bibliography

- [Abramowitz and Stegun, 1964] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover Publications.
- [Aitchison and Brown, 1963] Aitchison, J. and Brown, J. A. C. (1963). *The Lognormal Distribution*. Cambridge University Press.
- [Akbari. et al., 2016] Akbari., V., Anfinsen, S. N., Doulgeris, A. P., Eltoft, T., Moser, G., and Serpico, S. B. (2016). Polarimetric SAR change detection with the complex Hotelling-Lawley trace statistic. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):3953–3966.
- [Anfinsen et al., 2009] Anfinsen, S. N., Doulgeris, A. P., and Eltoft, T. (2009). Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):3795–3809.
- [Anfinsen et al., 2011] Anfinsen, S. N., Doulgeris, A. P., and Eltoft, T. (2011). Goodness-of-fit tests for multilook polarimetric radar data based on the Mellin transform. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7):2764–2781.
- [Anfinsen and Eltoft, 2011] Anfinsen, S. N. and Eltoft, T. (2011). Application of the matrix-variate Mellin transform to analysis of polarimetric radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2281–2295.
- [Askey, 2005] Askey, R. (2005). *Mathematics and Social Utopias in France: Olinde Rodrigues and His Times*, volume 28 of *History of Mathematics*, chapter 6: "The 1839 paper on permutations: Its relation to the Rodrigues formula and further developments", pages 105–118. American Mathematical Society.
- [Bateman et al., 1954] Bateman, H., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (1954). *Tables of Integral Transforms: Based in Part on Notes Left by Harry Bateman and Compiled by the Staff of the Bateman Manuscript Project*, volume I & II. McGraw-Hill.
- [Bell, 1927] Bell, E. T. (1927). Partition polynomials. *Annals of Mathematics*, pages 38–46.
- [Bertrand et al., 2000] Bertrand, J., Bertrand, P., and Ovarlez, J.-P. (2000). *The Transforms and Applications Handbook*, chapter 11, "The Mellin Transform". CRC Press LLC.
- [Blinnikov and Moessner, 1998] Blinnikov, S. and Moessner, R. (1998). Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement Series*, 130(1):193–205.
- [Bombrun et al., 2011] Bombrun, L., Anfinsen, S. N., and Harant, O. (2011). A complete coverage of log-cumulant space in terms of distributions for polarimetric SAR data. In *5th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR 2011)*.
- [Boyadzhiev, 2009] Boyadzhiev, K. N. (2009). Exponential polynomials, Stirling numbers, and evaluation of some gamma integrals. *Abstract and Applied Analysis*.

- [Bryc, 2012] Bryc, W. (2012). *The Normal Distribution: Characterizations With Applications*. Springer Science & Business Media.
- [Charlier, 1905] Charlier, C. V. L. (1905). Über das fehlergesetz. *Arkiv för Matematik, Astronomi och Fysik*, Vol. 2 (1905-1906)(8):9.
- [Charlier, 1906] Charlier, C. V. L. (1906). Über die darstellung willkürlicher funktionen. *Arkiv för Matematik, Astronomi och Fysik*, Vol. 2 (1905-1906)(20):35.
- [Chebyshev, 1860] Chebyshev, P. L. (1860). Sur le developpement des fonctions à une seule variable. *Bulletin de l'Académie impériale des sciences de St.-Pétersbourg*, Series 3, Vol. 1:193–202.
- [Chebyshev, 1890] Chebyshev, P. L. (1890). Sur deux théorèmes relatifs aux probabilités. *Acta Mathematica*, 14:305–315.
- [Choi and Wette, 1969] Choi, S. C. and Wette, R. (1969). Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics*, 11(4):683–690.
- [Cloude, 2010] Cloude, S. R. (2010). *Polarisation: Applications in Remote Sensing*. Oxford University Press.
- [Cohen Jr., A. C., 1951] Cohen Jr., A. C. (1951). Estimating parameters of logarithmic-normal distributions by maximum likelihood. *Journal of the American Statistical Association*, 46(254):206–212.
- [Cornish and Fisher, 1938] Cornish, E. A. and Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique*, pages 307–320.
- [Daalhuis, 2010] Daalhuis, A. O. (2010). *Confluent Hypergeometric Functions*. NIST Handbook of Mathematical Functions, p. 321–349. Cambridge University Press.
- [Davis, 1959] Davis, P. J. (1959). Leonhard Euler's integral: A historical profile of the gamma function: In memoriam: Milton Abramowitz. *The American Mathematical Monthly*, 66(10):849–869.
- [Deng and López-Martínez, 2016] Deng, X. and López-Martínez, C. (2016). Higher order log-cumulants for texture analysis of PolSAR data. In *Geoscience and Remote Sensing Symposium (IGARSS)*, pages 7541–7544. IEEE.
- [Deng et al., 2016] Deng, X., López-Martínez, C., and Varona, E. M. (2016). A physical analysis of polarimetric SAR data statistical models. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):3035–3048.
- [Doulgeris, 2015] Doulgeris, A. P. (2015). An automatic u-distribution and markov random field segmentation algorithm for polsar images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1819–1827.
- [Doulgeris et al., 2008] Doulgeris, A. P., Anfinson, S. N., and Eltoft, T. (2008). Classification with a non-gaussian model for polsar data. *IEEE Transactions on Geoscience and Remote Sensing*, 46(10):2999–3009.
- [Doulgeris et al., 2011] Doulgeris, A. P., Anfinson, S. N., and Eltoft, T. (2011). Automated non-gaussian clustering of polarimetric synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3665–3676.

- [Dubey, 1970] Dubey, S. D. (1970). Compound gamma, beta and F distributions. *Metrika*, 16(1):27–31.
- [Durbin and Watson, 1951] Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. ii. *Biometrika*, 38(1/2):159–177.
- [Edgeworth, 1905] Edgeworth, F. Y. (1905). The law of error. *Cambridge Philos. Trans.*, 20:36–66, 113–141.
- [Esseen, 1945] Esseen, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Mathematica*, 77(1):1–125.
- [Euler, 1738] Euler, L. (1738). De progressionibus transcendentibus seu quarum termini generales algebraice dari nequeunt. *Commentarii academiae scientiarum Petropolitanae*, 5:36–57.
- [Flajolet et al., 1995] Flajolet, P., Gourdon, X., and Dumas, P. (1995). Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144(1-2):3–58.
- [Fowler, 1996] Fowler, D. (1996). The binomial coefficient function. *The American Mathematical Monthly*, 103(1):1–17.
- [Frery et al., 1997] Frery, A. C., Muller, H.-J., Yanasse, C. d. C. F., and Sant’Anna, S. J. S. (1997). A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):648–659.
- [Frery et al., 2014] Frery, A. C., Nascimento, A. D., and Cintra, R. J. (2014). Analytic expressions for stochastic distances between relaxed complex Wishart distributions. *IEEE Transactions on Geoscience and Remote Sensing*, 52(2):1213–1226.
- [Gauss, 1823] Gauss, C. F. (1823). Theoria combinationis observationum erroribus minimis obnoxiae. *Werke*, 4:1.
- [Gaztanaga et al., 2000] Gaztanaga, E., Fosalba, P., and Elizalde, E. (2000). Gravitational evolution of the large-scale probability density distribution: The Edgeworth and gamma expansions. *The Astrophysical Journal*, 539(2):522.
- [Goovaerts and De Pril, 1980] Goovaerts, M. J. and De Pril, N. (1980). Survival probabilities based on pareto claim distributions: Comment. *Astin Bulletin*, 11(2):154–157.
- [Gram, 1883] Gram, J. P. (1883). Über die entwicklung reeler funktionen in reihen mittelst der methode der kleinsten quadrate. *Journal für die reine und angewandte Mathematik*, 94:41–73.
- [Hald, 2000] Hald, A. (2000). The early history of the cumulants and the Gram-Charlier series. *International Statistical Review*, 68(2):137–153.
- [Hall, 2013] Hall, P. (2013). *The Bootstrap and Edgeworth Expansion*. Springer Science & Business Media.
- [Heyde, 1963] Heyde, C. (1963). On a property of the lognormal distribution. *Journal of the Royal Statistical Society, Series B (Methodological)*.
- [Hill and Davis, 1968] Hill, G. W. and Davis, A. W. (1968). Generalized asymptotic expansions of Cornish-Fisher type. *The Annals of Mathematical Statistics*, pages 1264–1273.
- [Hsu, 1945] Hsu, P. L. (1945). The approximate distributions of the mean and variance of a sample of independent variables. *The Annals of Mathematical Statistics*, 16(1):1–29.

- [Inglada and Mercier, 2007] Inglada, J. and Mercier, G. (2007). A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5).
- [Jakeman and Pusey, 1978] Jakeman, E. and Pusey, P. (1978). Significance of K distributions in scattering experiments. *Physical Review Letters*, 40(9):546–550.
- [Johnson, 2002] Johnson, W. P. (2002). The curious history of Faà di Bruno’s formula. *The American Mathematical Monthly*, 109(3):217–234.
- [Kailath, 1967] Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60.
- [Kendall et al., 1994] Kendall, M., Stuart, A., and Ord, J. (1994). *Kendall’s Advanced Theory of Statistics*, volume 1: Distribution theory. John Wiley & Sons, Ltd, 6th edition.
- [Krishnamoorthy, 2006] Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions With Applications*. Chapman & Hall/CRC, Taylor & Francis Group.
- [Krylov et al., 2013] Krylov, V. A., Moser, G., Serpico, S. B., and Zerubia, J. (2013). On the method of logarithmic cumulants for parametric probability density function estimation. *IEEE Transactions on Image Processing*, 22(10):3791–3806.
- [Kummer, 1837] Kummer, E. E. (1837). De integralibus quibusdam definitis et seriebus infinitis. *Journal für die reine und angewandte Mathematik*, 17:228–242.
- [Lengyel, 1994] Lengyel, T. (1994). On the divisibility by 2 of the Stirling numbers of the second kind. *Fibonacci Quarterly*, 32(3):194–201.
- [Li et al., 2011] Li, H.-C., Hong, W., Wu, Y.-R., and Fan, P.-Z. (2011). On the empirical-statistical modeling of SAR images with generalized gamma distribution. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):386–397.
- [Lukacs, 1970] Lukacs, E. (1970). *Characteristic Functions*. Griffin.
- [McClellan et al., 2003] McClellan, J. H., Schafer, R. W., and Yoder, M. A. (2003). *Signal Processing First*. Pearson/Prentice Hall.
- [McDonald, 1995] McDonald, James B. & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66:133–152.
- [Mihoubi, 2008] Mihoubi, M. (2008). Bell polynomials and binomial type sequences. *Discrete Mathematics*, 308(12):2450–2459.
- [Moser and Serpico, 2009] Moser, G. and Serpico, S. B. (2009). Unsupervised change detection from multichannel SAR data by Markovian data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2114–2128.
- [Nicolas, 2002] Nicolas, J.-M. (2002). Introduction aux statistiques de deuxième espèce: applications de logs-moments et des logs-cumulats à l’analyse de lois d’images radar. *Traitement du Signal*, 19(3):139–167.
- [Nicolas, 2011] Nicolas, J.-M. (2011). Les distributions de Meijer et leurs propriétés en statistiques de Mellin. http://perso.telecom-paristech.fr/~nicolas/jmnicolas_meijer_2011D002.pdf, retrieved 2017-05-12.
- [Nicolas, 2016] Nicolas, J.-M. (2016). *Lois statistiques de l’imagerie cohérente. Kit de survie pour les traiteurs d’images RSO. V-β (14 novembre 2016)*. Télécom ParisTech, département TSI. *Technical note*.

- [Nicolas and Anfinen, 2012] Nicolas, J.-M. and Anfinen, S. N. (2012). Introduction to second kind statistics: Application of log-moments and log-cumulants to the analysis of radar image distributions. *Technical note*.
- [Oliver and Quegan, 2004] Oliver, C. and Quegan, S. (2004). *Understanding Synthetic Aperture Radar Images*. SciTech Publishing.
- [Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.
- [Pastor, 2016] Pastor, G. (2016). *Probabilistic Models and Algorithms for Energy-Efficient Large and Dense Wireless Sensor Networks*. PhD thesis, Aalto University, Department of Communications and Networking.
- [Pastor et al., 2014] Pastor, G., Mora-Jiménez, I., Caamano, A. J., and Jäntti, R. (2014). Log-cumulants-based Edgeworth expansion for skew-distributed aggregate interference. *2014 11th International Symposium on Wireless Communications Systems (ISWCS), IEEE*, pages 390–394.
- [Pastor et al., 2016] Pastor, G., Mora-Jiménez, I., Caamaño, A. J., and Jäntti, R. (2016). Asymptotic expansions for heavy-tailed data. *IEEE Signal Processing Letters*, 23(4):444–448.
- [Pitman, 1937] Pitman, E. J. G. (1937). The “closest” estimates of statistical parameters. *Mathematical Proceedings of the Cambridge Philosophical Society*, 33(02):212–222.
- [Pitman, 2002] Pitman, J. (2002). Combinatorial stochastic processes. Technical Report 621, Dept. Statistics, UC Berkeley, 2002. Lecture notes for St. Flour course.
- [Rignot and van Zyl, 1993] Rignot, E. J. M. and van Zyl, J. J. (1993). Change detection techniques for ERS-1 SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 31(4):896–906.
- [Rosenblatt, 1956] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.
- [Rota and Shen, 2000] Rota, G.-C. and Shen, J. (2000). On the combinatorics of cumulants. *Journal of Combinatorial Theory, Series A*, 91(1-2):283–304.
- [Rottmann, 2003] Rottmann, K. (2003). *Matematisk Formelsamling*. Spektrum, 14. edition.
- [Stacy, 1962] Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, pages 1187–1192.
- [Stark and Woods, 2012] Stark, H. and Woods, J. W. (2012). *Probability, Statistics, and Random Processes for Engineers*. Pearson, 4th edition.
- [Sundt et al., 1998] Sundt, B., Dhaene, J., and De Pril, N. (1998). Some results on moments and cumulants. *Scandinavian Actuarial Journal*, 1998(1):24–40.
- [Szeg, 1939] Szeg, G. (1939). *Orthogonal Polynomials*, volume 23. American Mathematical Society.
- [Tadikamalla, 1979] Tadikamalla, P. R. (1979). Random sampling from the generalized gamma distribution. *Computing*, 23(2):199–203.
- [Tellambura and Senaratne, 2010] Tellambura, C. and Senaratne, D. (2010). Accurate computation of the MGF of the lognormal distribution and its application to sum of lognormals. *IEEE Transactions on Communications*, 58(5).

- [Theodoridis and Koutroumbas, 2009] Theodoridis, S. and Koutroumbas, K. (2009). *Pattern Recognition*. Academic Press, fourth edition.
- [Thiele, 1889] Thiele, T. (1889). *Almindelig Iagttagelseslære: Sandsynlighedsregning og Mindste Kvadraters Methode*. Kjøbenhavn: Reitzel.
- [Tricomi, 1947] Tricomi, F. (1947). Sulle funzioni ipergeometriche confluenti. *Annali di Matematica Pura ed Applicata*, 26:141–175.
- [Wagner, 1996] Wagner, C. G. (1996). Generalized Stirling and Lah numbers. *Discrete Mathematics*, 160(1):199–218.
- [Wallace, 1958] Wallace, D. L. (1958). Asymptotic approximations to distributions. *The Annals of Mathematical Statistics*, 29(3):635–654.

A Framework for Mellin Kind Series Expansion Methods

Torgeir Brenn, Stian N. Anfinssen

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 for both methods, but the Mellin kind Edgeworth series in particular. 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.
GFD	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.
MKLN	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: togeir.brenn@gmail.com, stian.normann.anfinssen@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 synthetic aperture radar (SAR) images, 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.

Mellin kind statistics (MKS) were introduced by Nicolas in [12] 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 [13]. 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 [14].

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 $f_X(x)$ [15], that is

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

where j is the imaginary unit, the expectation operation $\mathbb{E}\{\cdot\}$ is performed with respect to X , and $\omega \in \mathbb{R}$ is the transform variable. The linear moments of order ν are defined as $m_\nu \equiv \mathbb{E}\{X^\nu\}$. 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 $\Phi_\alpha(\omega)$ [16]. Unique to the Gaussian distribution is the property that $c_{\alpha,\nu} = 0 \forall \nu \geq 3$ [17]. 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 [15] from (3) via the inverse FT

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

where $\exp\{\cdot\}$ is the exponential function and $D_x = d/dx$ is the derivative operator. To get a more workable 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 $(-D_x)$ and recollect the definition of the Hermite polynomials, $H_\nu(x)\alpha(x) = (-D_x)^\nu \alpha(x)$ [15], to get the classical Gram-Charlier series

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

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

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

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

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

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

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

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 [18], the partial Bell polynomials are defined as [19]

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

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

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

The Bell polynomials satisfy [19]

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

and a well-known use of this result is to retrieve the ν th order moment from the cumulants of order $\leq \nu$ [20] as

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

C. The Mellin Transform

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

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

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 $\langle a, b \rangle$

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	$(D_x x)^n f(x)$	$(-1)^n (s-1)^n F(s)$	$s \in S_f$

of $\text{Re}(s)$ for which the integral in (14) converges. If $s \in S_f$, then $f(x)$ is retrievable via the inverse MT [21],

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

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 [22], [23]. The differentiation properties introduces the falling factorial, defined as $(s-1)_n = (s-1)(s-2)\cdots(s-n)$, and the Mellin derivative¹ 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 [24], it did not receive much attention until the introduction of MKS in [12]. The MKCF $\phi_X(s)$ is defined as the MT of the PDF

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

The log-moments are defined as $\mu_\nu \equiv E\{(\log X)^\nu\}$, 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

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

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

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 [12], [25].

For a more detailed review of the fundamental properties of MKS, see e.g. [12] (English translation: [26]), while [13] 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 [27].

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

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

Deviating from the path that lead us to the classical Gram-Charlier and Edgeworth series, the Bell polynomials and their property in (12) gives us

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

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

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

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

B. The Mellin Kind Gamma Kernel Series

Let

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

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 $(-D_x x)^n \gamma(x; a, b)$. By letting $b = 1$, it is possible to define the polynomials $M_n(x)$ implicitly as

$$M_n(x) \gamma(x; a) = (-D_x x)^n \gamma(x; a). \quad (25)$$

or²

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

and since $D_x x$ is scale invariant, the generalization to arbitrary b is simply to replace x with bx in the polynomials, i.e.

$$M_n(bx) \gamma(x; a, b) = (-D_x x)^n \gamma(x; a, b). \quad (27)$$

In Appendix A we prove that $M_n(x)$ is a linear combination of the well-known generalized Laguerre polynomials [29] 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

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

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

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

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

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 $\mu = E\{\log X\}$ and log-variance $\sigma^2 = E\{(\log X - \mu)^2\}$ [14] to obtain the Mellin kind log-normal

²This definition mirrors the Rodrigues type formulae [28], with the Mellin derivative replacing the standard differentiation operator.

kernel (MKLK) series. To evaluate $(-D_x x)^n \Lambda(x; \mu, \sigma)$, we use the proof (59) of Lemma 3 from Appendix B to see that

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

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 [27]

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

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

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

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

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

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

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

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, $s - 1 \rightarrow j\omega$, the fact that if $Y = \log X$, and then [12] 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]. This use of the Bell polynomials is not limited to the Mellin kind framework, and in [Fyll inn her], we use $B_n(\cdot)$ in the classical Gram-Charlier and Edgeworth series.

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 $d_B(f(x), \hat{f}(x))$ [30], the Kullback-Leibler distance³ $d_{KL}(f(x), \hat{f}(x))$, 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 permit negative values) [9], we needed to slightly modify the

³We take distance to mean a non-negative definite symmetric function, and $d_{KL}(f(x), \hat{f}(x))$ is the symmetrized version of the Kullback-Leibler divergence. This follows the nomenclature of [30].

estimates to ensure that the ratios and logarithms in $d_B(\cdot)$ and $d_{KL}(\cdot)$ do not fail but also gives fair results.⁴

A. Broad Comparison of the Methods

We start with a general comparison of 7 methods based on the log-cumulants. In addition to the MKGK, MKLK and MKE series, the methods tested are the gamma, log-normal, K [6], and generalized gamma (G Γ D) [7] distributions with parameters given by the MoLC. Note that the MoLC gamma method corresponds to the kernel of the MKGK series, i.e. $N \leq 2$ in (29), while the MoLC log-normal method corresponds to the kernels in both the MKLK and MKE series, i.e. $N \leq 2$ 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 $O(r^{-3/2})$ 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 [12]

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

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

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

where $x, g, L > 0$, $M < 0$. These two distributions are given in [31] as the two theoretical distribution arising 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. The MKLK, MKE, and MKGK series visibly improve on their kernels, but they are not able to attain the accuracy of the three-parameter G Γ D 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.

⁴Specifically, to remedy the common feature that the series expansion methods produce estimates $\hat{f}(x)$ which integrate to < 1 and result in $d_B(\cdot) < 0$, we divided $\hat{f}(x)$ by its integral.

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

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

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

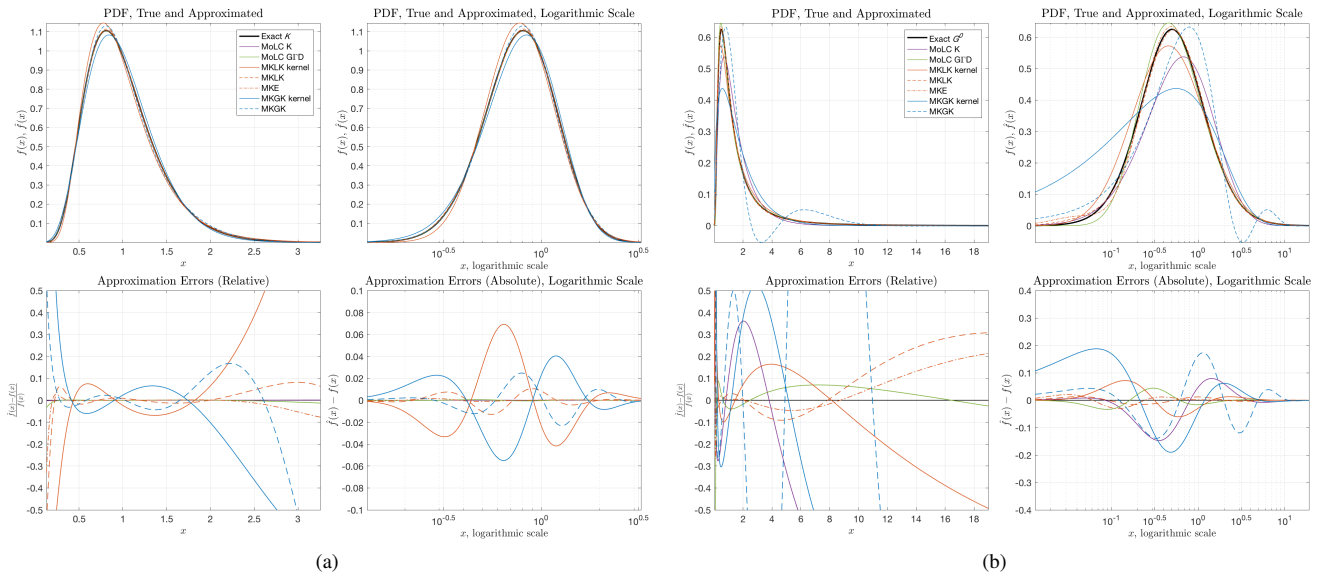


Fig. 1. PDF approximations of known distributions produced by 7 methods, with the series expansion methods corrected for $\kappa_{X,3}$ and $\kappa_{X,4}$. (a) K distribution with parameters $\mu = 1$, $L = 16$, $M = 10$, (b) G^0 distribution with parameters $g = 2$, $L = 4$, $M = -2$.

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

ρ	$\mathcal{K}(x; \rho), \rho=[\mu, L, M]$		$G^0(x; \rho), \rho=[g, L, M]$		$\gamma(x; \rho), \rho=[a, b]$		$\gamma^{-1}(x; \rho), \rho=[a, b]$		$GFD(x; \rho), \rho=[a, b, d]$	
	[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	$6.67 \cdot 10^{-3}$	$6.78 \cdot 10^{-2}$	$1.02 \cdot 10^{-3}$	$5.40 \cdot 10^{-4}$	$4.20 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$7.88 \cdot 10^{-3}$	Failure
MoLC GFD [7]	$3.51 \cdot 10^{-7}$	$2.55 \cdot 10^{-5}$	$9.29 \cdot 10^{-5}$	$1.87 \cdot 10^{-3}$	Exact	Exact	$4.11 \cdot 10^{-6}$	$1.05 \cdot 10^{-7}$	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	$6.26 \cdot 10^{-6}$	$6.79 \cdot 10^{-4}$	$3.96 \cdot 10^{-6}$	$9.33 \cdot 10^{-4}$	$3.94 \cdot 10^{-3}$	$1.31 \cdot 10^{-5}$	$7.82 \cdot 10^{-3}$	$1.32 \cdot 10^{-5}$	$6.40 \cdot 10^{-3}$	$1.30 \cdot 10^{-5}$
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	$5.68 \cdot 10^{-2}$	$2.26 \cdot 10^{-3}$	$7.20 \cdot 10^{-4}$	$4.72 \cdot 10^{-4}$

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 GFD 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 [12]

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

where $x, b > 0$, $a > 0$, and the generalized gamma distribution [32], [33], [7]

$$GFD(x; a, b, d) = \frac{|d|b}{\Gamma(a)} (bx)^{ad-1} \exp\{-(bx)^d\}, \quad (46)$$

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.

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 GFD methods. Disregarding distributions which are exact matches, the series expansions prove the most accurate in 5 of the 10 distributions tested, with the standout performers being the MoLC GFD method and 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 GFD when corrected for $\kappa_{X,5}$ and beyond. The same is the case for $\gamma(x; a = 4, b = 2)$ and when

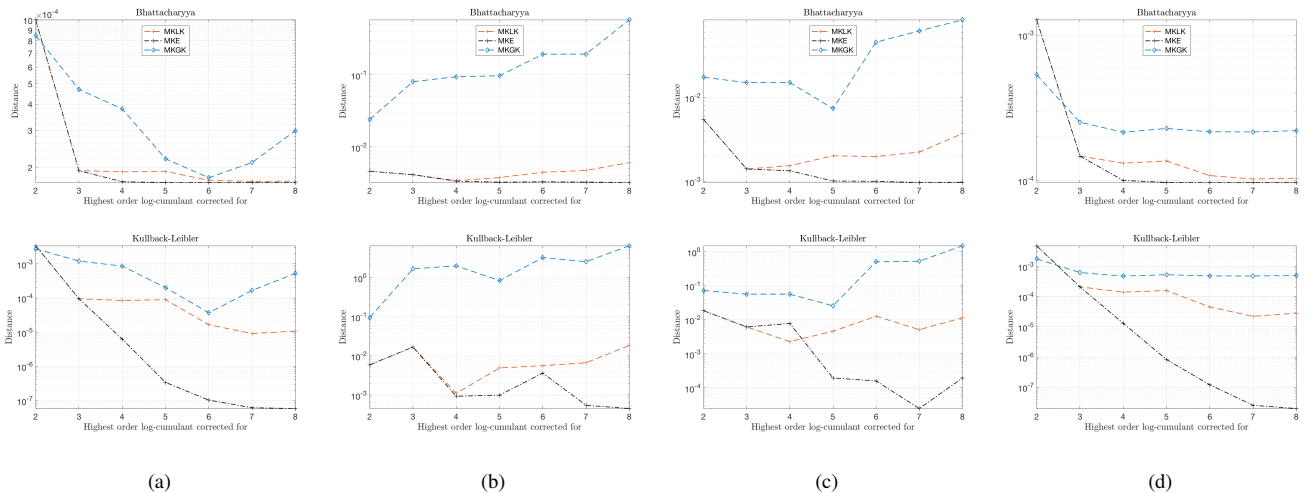


Fig. 2. Bhattacharyya distance (top) and Kullback-Leibler distance (bottom) to the true distribution as a function of the number of terms in the series expansion approximations. True distributions (a) $\mathcal{K}(x; \mu = 1; L = 16, M = 10)$, (b) $G^0(x; g = 2, L = 4, M = -2)$, (c) $\gamma^{-1}(x; a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

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 bested the MoLC GFD 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.

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 GFD 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 GFD 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 GFD, 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 method 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 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

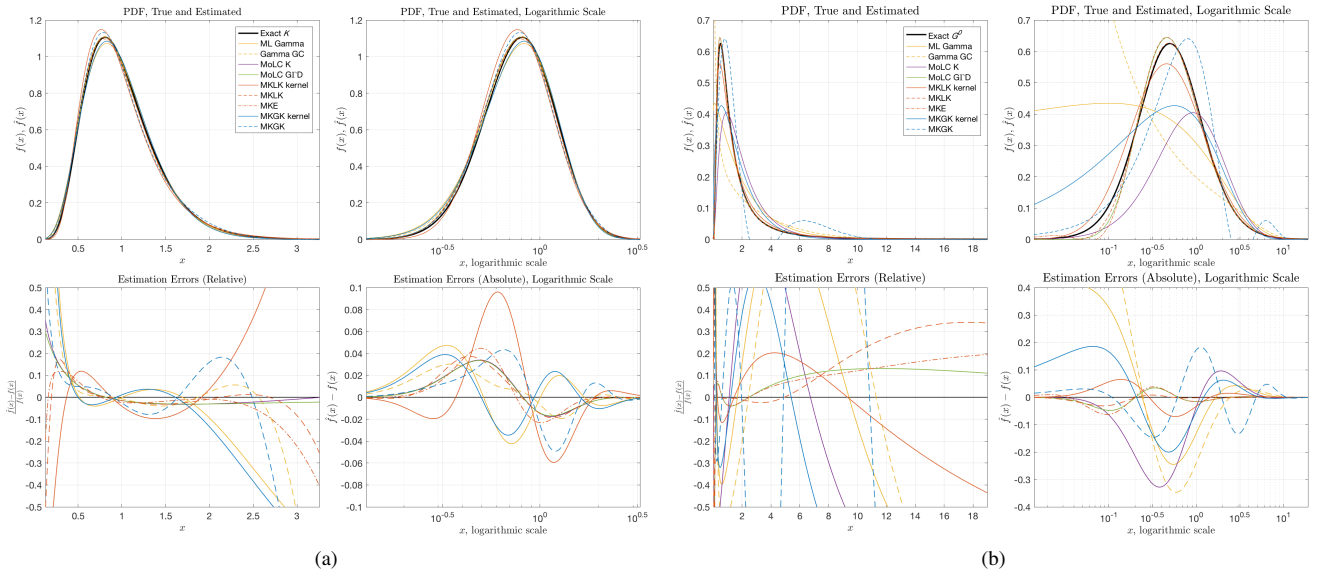


Fig. 3. PDF estimates produced by 9 methods, based on 1000 synthesized data points, with the series expansion methods corrected for κ_3, κ_4 . (a) K distribution with parameters $\mu = 1, L = 16, M = 10$, (b) G^0 distribution with parameters $g = 2, L = 4, M = -2$.

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

ρ	$\mathcal{K}(x; \rho, \rho=[\mu, L, M])$		$G^0(x; \rho, \rho=[g, L, M])$		$\gamma(x; \rho, \rho=[a, b])$		$\gamma^{-1}(x; \rho, \rho=[a, b])$		$GFD(x; \rho, \rho=[a, b, d])$	
	[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]	$3.72 \cdot 10^{-3}$	$1.02 \cdot 10^{-2}$	$2.04 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$	$9.55 \cdot 10^{-4}$	$9.85 \cdot 10^{-4}$	$7.90 \cdot 10^{-2}$	$2.03 \cdot 10^{-2}$	$5.60 \cdot 10^{-3}$	$2.17 \cdot 10^{-3}$
$\gamma(\cdot)$ 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]	$1.13 \cdot 10^{-3}$	$1.37 \cdot 10^{-3}$	$7.76 \cdot 10^{-3}$	$1.19 \cdot 10^{-1}$	$2.23 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$4.60 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	$8.77 \cdot 10^{-3}$	Failure
MoLC GFD [7]	$1.46 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$1.56 \cdot 10^{-3}$	$3.48 \cdot 10^{-3}$	$1.59 \cdot 10^{-3}$	$1.49 \cdot 10^{-3}$	$1.51 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$
MKLL, kernel only	$4.39 \cdot 10^{-3}$	$1.78 \cdot 10^{-2}$	$1.65 \cdot 10^{-3}$	$6.90 \cdot 10^{-3}$	$1.93 \cdot 10^{-2}$	$5.68 \cdot 10^{-3}$	$1.95 \cdot 10^{-2}$	$5.74 \cdot 10^{-3}$	$2.01 \cdot 10^{-2}$	$5.74 \cdot 10^{-3}$
MKLL 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.

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, MKLL and MKE series), with respect to their performance as the number of correcting terms are varied. In Fig. 4 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 GFD data in (d). The MKGK 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 (observations). 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 GFD method. We also see that the MKE series benefits even more from an increase

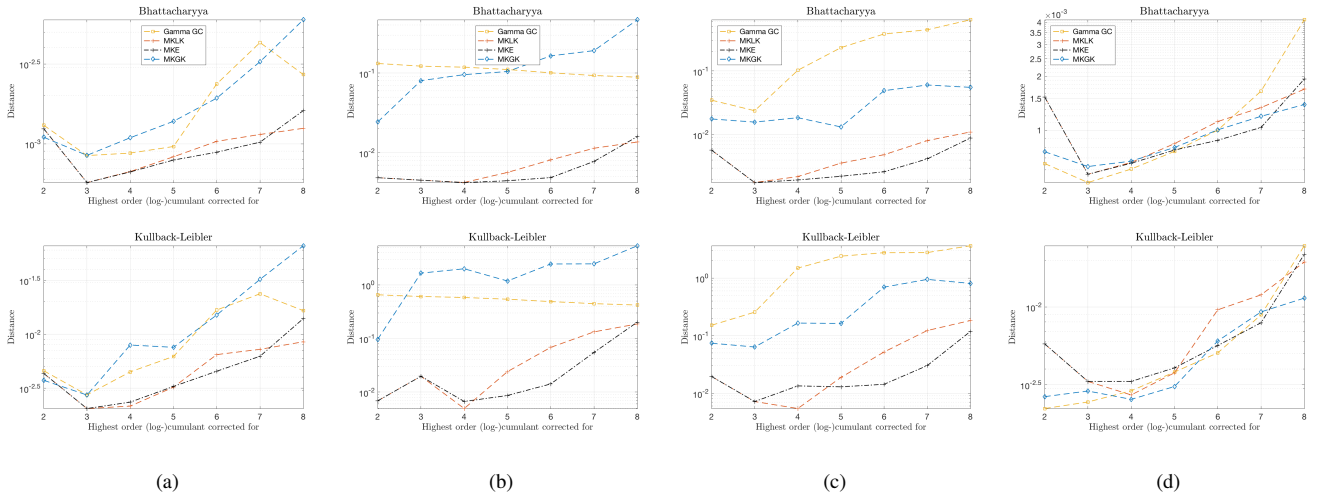


Fig. 4. Bhattacharyya distance (top) and Kullback-Leibler distance (bottom) to the true distribution as a function of the number of terms in the series expansion estimates. Mean of 1000 iterations with 1000 synthesized data points. True distributions (a) $\mathcal{K}(x; \mu = 1, L = 16, M = 10)$, (b)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

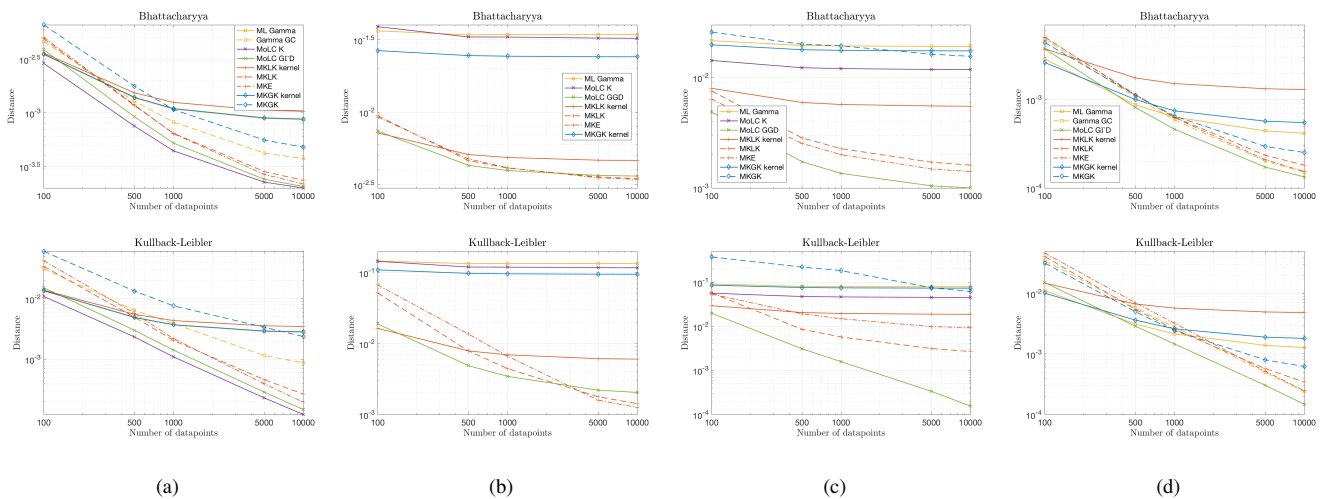


Fig. 5. Bhattacharyya distance (top) and Kullback-Leibler distance (bottom) to the true distribution as a function of the number of data points. Mean of 1000 iterations, series expansion methods corrected with two terms as in Fig. 3. True distributions (a) $\mathcal{K}(x; \mu = 1, L = 16, M = 10)$, (b) $G^0(x; g = 2, L = 4, M = -2)$, (c) $\gamma^{-1}(x; a = 4, b = 2)$, (d) $\text{GFD}(x; a = 16, b = 8, d = 2)$. 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.

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 the 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 10000 data points in all four cases. Finally, we note that [15] 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

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

Obs.	$\mathcal{K}(x; \mu=1, L=16, M=10)$			$G^0(x; g=2, L=4, M=-2)$			$\gamma^{-1}(x; a=4, b=2)$			GFD($x; a=16, b=8, d=2$)		
	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	κ_2, κ_3	κ_4
MKLG series	κ_2	κ_3	κ_3, κ_4	κ_2	κ_4	κ_4	κ_3	κ_3, κ_4	κ_3, κ_4	κ_2	κ_3, κ_4	κ_4
MKE series	κ_2	κ_3	κ_4	κ_2	κ_4	κ_4, κ_5	κ_3	κ_3	κ_5	κ_2	κ_3	κ_4
MKGK series	κ_2	κ_3	κ_5	κ_2	κ_2	κ_2	κ_2	κ_3, κ_5	κ_3, κ_5	κ_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.

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 simpler representations of the series expansion methods and especially 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 MKLG 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 GFD 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 MKLG 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,

$$M_n(x) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k k! L_k^{(a-1)}(x), \quad (47)$$

where $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ denotes the Stirling numbers of the second kind [34]

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} i^n, \quad (48)$$

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 $D_x^k x^k$, see that

$$D_x^k x^k f(x) = D_x^{k-1} x^{k-1} [(k + x D_x) f(x)], \quad (49)$$

and by repetition we get

$$D_x^k x^k f(x) = (x D_x + k)_k f(x). \quad (50)$$

Using the fact that $D_x x - x D_x = 1$, we have

$$D_x^k x^k f(x) = (D_x x + k - 1)_k f(x). \quad (51)$$

By a property of the Stirling numbers [35], the Mellin derivative from Table I can be rewritten as

$$(-D_x x)^n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k (D_x x + k - 1)_k = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k D_x^k x^k, \quad (52)$$

and multiplying both sides with the unit scale gamma distribution $\gamma(x; a)$, the definitions of $M_n(x)$ and $L_k^{(a)}(x)$ [29] are recognized on the left and right hand sides, respectively, of

$$M_n(x) = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k k! L_k^{(a-1)}(x). \quad (53)$$

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

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

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

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

$$a_n = \left\{ \begin{matrix} n \\ n \end{matrix} \right\} (-1)^n n! \frac{(-1)^n}{n!} = 1, \quad (54)$$

where it was used that $\left\{ \begin{matrix} n \\ n \end{matrix} \right\} = 1 \forall n$.

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

$$M_0(x) = 1 \quad (55)$$

$$M_1(x) = x - a \quad (56)$$

$$M_2(x) = x^2 - (2a + 1)x + a^2 \quad (57)$$

$$M_3(x) = x^3 - 3(a + 1)x^2 + (3a^2 + 3a + 1)x - a^3 \quad (58)$$

APPENDIX B
LOGARITHMIC HERMITE POLYNOMIALS

Lemma 3:

$$(-D_x x)^n \Lambda(x; \mu, \sigma) = \frac{1}{\sigma^n} H_n \left(\frac{\log x - \mu}{\sigma} \right) \Lambda(x; \mu, \sigma), \quad (59)$$

where $H_n(\cdot)$ is the n th 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 [15]

$$(-D_x)^n \alpha(x) = H_n(x) \alpha(x). \quad (60)$$

Proof: We can use the chain rule to see that

$$D_{\log x} = \frac{d}{d \log x} = \frac{dx}{d \log x} \frac{d}{dx} = x \frac{d}{dx} = x D_x, \quad (61)$$

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

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

giving

$$(-D_x x)^n \Lambda(x) = (-1)^n D_x x \cdots D_x x \frac{1}{x} \alpha(\log x) \quad (63)$$

$$= (-1)^n \frac{1}{x} x D_x x \cdots D_x \alpha(\log x) \quad (64)$$

$$(-D_x x)^n \Lambda(x) = \frac{1}{x} (-x D_x)^n \alpha(\log x). \quad (65)$$

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

$$\left(-\frac{d}{d \log x} \right)^n \alpha(\log x) = H_n(\log x) \alpha(\log x) \quad (66)$$

$$(-D_x x)^n \Lambda(x) = \frac{1}{x} H_n(\log x) \alpha(\log x) \quad (67)$$

$$(-D_x x)^n \Lambda(x) = H_n(\log x) \Lambda(x). \quad (68)$$

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

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

giving

$$(-D_x x)^n \Lambda(x; \mu, \sigma) = (-D_x x)^n \frac{u}{x\sigma} \Lambda(u) = \frac{1}{x} (-x D_x)^n \frac{u}{\sigma} \Lambda(u), \quad (70)$$

but

$$(-x D_x)^n = (-u D_u)^n \frac{1}{\sigma^n}, \quad (71)$$

so

$$(-D_x x)^n \Lambda(x; \mu, \sigma) = \frac{u}{x\sigma^{n+1}} (-D_u u)^n \Lambda(u). \quad (72)$$

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

$$(-D_x x)^n \Lambda(x; \mu, \sigma) = \frac{u}{x\sigma^{n+1}} H_n(\log u) \Lambda(u) \quad (73)$$

$$= \frac{1}{\sigma^n} H_n \left(\frac{\log x - \mu}{\sigma} \right) \Lambda(x; \mu, \sigma). \quad (74)$$

APPENDIX C
THE MELLIN KIND EDGEWORTH ASSUMPTION

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

$$\log X = \frac{1}{\sqrt{r}} \sum_{i=1}^r \frac{Z_i - m}{s}, \quad (75)$$

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

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

Proof: Since both quantities are defined as $E\{(\log X)^\nu\}$, 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 [12], 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 $\kappa_2 = \mu_2 - \mu_1^2 = \sigma^2$ [12], 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.

ACKNOWLEDGMENT

The authors would like to thank...

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] 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.
- [13] 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.

- [14] 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.
- [15] M. Kendall, A. Stuart, and J. Ord, *Kendall's Advanced Theory of Statistics*, 6th ed. John Wiley & Sons, Ltd, 1994, vol. 1: Distribution theory.
- [16] W. Bryc, *The Normal Distribution: Characterizations With Applications*. Springer Science & Business Media, 2012.
- [17] E. Lukacs, *Characteristic Functions*. Griffin, 1970.
- [18] E. T. Bell, "Partition polynomials," *Ann. Math.*, pp. 38–46, 1927.
- [19] M. Mihoubi, "Bell polynomials and binomial type sequences," *Discrete Mathematics*, vol. 308, no. 12, pp. 2450–2459, 2008.
- [20] G.-C. Rota and J. Shen, "On the combinatorics of cumulants," *J. comb. Theory A*, vol. 91, no. 1-2, pp. 283–304, 2000.
- [21] A. Fitouhi, N. Bettaibi, and K. Brahim, "The Mellin transform in quantum calculus," *Constr. Approx.*, vol. 23, no. 3, pp. 305–323, 2006.
- [22] J. Bertrand, P. Bertrand, and J.-P. Ovarlez, *The Transforms and Applications Handbook*. CRC Press LLC, 2000, ch. 11, "The Mellin Transform".
- [23] H. Bateman, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms*. McGraw-Hill, 1954, vol. I & II.
- [24] B. Epstein, "Some applications of the Mellin transform in statistics," *Ann. Math. Stat.*, pp. 370–379, 1948.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] G. Szeg, *Orthogonal Polynomials*. American Mathematical Society, 1939, vol. 23.
- [30] 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.
- [31] 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.
- [32] E. W. Stacy, "A generalization of the gamma distribution," *Ann. Math. Stat.*, pp. 1187–1192, 1962.
- [33] P. R. Tadikamalla, "Random sampling from the generalized gamma distribution," *Computing*, vol. 23, no. 2, pp. 199–203, 1979.
- [34] 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.
- [35] C. G. Wagner, "Generalized Stirling and Lah numbers," *Discrete Math.*, vol. 160, no. 1, pp. 199–218, 1996.