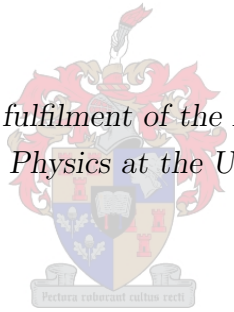
Gaussian and non-Gaussian-based Gram-Charlier and Edgeworth expansions for correlations of identical particles in HBT interferometry

by

Michiel Burger De Kock



Thesis presented in partial fulfilment of the requirements for the degree of Master of Theoretical Physics at the University of Stellenbosch

Department of Physics
University of Stellenbosch
Private Bag X1, 7602 Matieland, South Africa

Supervisor: Prof H.C. Eggers

March 2009

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 02-03-2009

Abstract

Gaussian and non-Gaussian-based Gram-Charlier and Edgeworth expansions for correlations of identical particles in HBT interferometry

M.B. De Kock

Department of Physics
University of Stellenbosch
Private Bag X1, 7602 Matieland, South Africa

Thesis: MSc (Theoretical Physics)

March 2009

Hanbury Brown–Twiss interferometry is a correlation technique by which the size and shape of the emission function of identical particles created during collisions of high-energy leptons, hadrons or nuclei can be determined. Accurate experimental datasets of three-dimensional correlation functions in momentum space now exist; these are sometimes almost Gaussian in form, but may also show strong deviations from Gaussian shapes. We investigate the suitability of expressing these correlation functions in terms of statistical quantities beyond the normal Gaussian description. Beyond means and the covariance matrix, higher-order moments and cumulants describe the form and difference between the measured correlation function and a Gaussian distribution. The corresponding series expansion is the Gram-Charlier series and in particular the Gram-Charlier Type A expansion found in the literature, which is based on a Gaussian reference distribution. We investigate both the Gram-Charlier Type A series as well as generalised forms based on non-Gaussian reference distributions, as well as the related Edgeworth expansion. For testing purposes, experimental data is initially represented by a suite of one-dimensional analytic non-Gaussian distributions. We conclude that the accuracy of these expansions can be improved dramatically through a better choice of reference distribution, suggested by the sign and size of the kurtosis of the experimental distribution. We further extend our investigation to simulated samples

of such test distributions and simplify the theoretical expressions for unbiased estimators (k-statistics) for the case of symmetric distributions.

Abstrak

Gaussian and non-Gaussian-based Gram-Charlier and Edgeworth expansions for correlations of identical particles in HBT interferometry

M.B. De Kock

Departement Fisika
Universiteit van Stellenbosch
Privaatsak X1, 7602 Matieland, Suid Afrika

Tesis: MSc (Teoretiese Fisika)

Maart 2009

Hanbury Brown–Twiss interferometrie is 'n korrelasietegniek waarmee die grootte en vorm van die emissiefunksie van identiese deeltjies wat tydens botsings van hoë-energie leptone, barione en kerne geskep is, bepaal kan word. Daar bestaan nou akkurate eksperimentele driedimensionele korrelasiefunksies in momentum-ruimte wat somtyds amper gaussiese vorm het, maar soms ook sterk afwykings van gaussiese vorms toon. Ons ondersoek die toepaslikheid van nie-gaussiese statistiese hoeveelhede in die uitdrukking van sulke korrelasiefunksies. Benewens die gemiddeldes en kovariansiematriks beskryf die hoër-orde momente en kumulante die vorm en die verskil tussen die korrelasiefunksie en 'n Gaussverdeling. Die algemene uitbreiding wat hiermee gepaard gaan is die Gram-Charlier reeks, terwyl die Gram-Charlier Tipe A reeks wat in die literatuur verskyn gebaseer is op die Gauss verdeling. Ons ondersoek sowel die Tipe A reeks asook die meer algemene Gram-Charlier reeks gebaseer op verskillende nie-gaussiese verwysingsverdelings, asook die verwante Edgeworth-reeks. Eksperimentele data word vir toetsdoeleindes aanvanklik voorgestel deur eendimensionele analitiese nie-gaussiese verdelings. Ons kom tot die gevolgtrekking dat die akkuraatheid van reeksuitbreidings dramaties verbeter kan word deur 'n beter keuse van 'n verwysingsverdeling te maak, ingelig deur die teken en grootte van die kurtosis van die eksperimentele verdeling. Ons ondersoek verder gesimuleerde steekproewe van die toetsverdelings en vereenvoudig die teoretiese uitdrukkings vir onsydige beramer (k-statistiek) vir die geval van identiese deeltjies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to femtoscopy

"Femtoscopy" is the latest name for what is also variously known in experimental high energy physics as "Bose-Einstein Correlations" and "HBT" or "intensity interferometry". The underlying physics is that of two-particle quantum mechanical wavefunctions, which for identical bosons exhibit the well-known "Bose-Einstein effect", while the names "HBT" and "intensity interferometry" are more accurately associated with the experimental techniques required.

1.1 Intensity interferometry and the Bose-Einstein effect

In the 1940s and 1950s, astronomers made use of amplitude interferometry, in which wave amplitudes are measured simultaneously by two spatially separated detectors, to measure the diameters of stars. The angular diameters of smaller stars was not resolvable due to the phase fluctuations of the light in the detector, caused by the atmosphere.

In 1956, radio astronomers Robert Hanbury Brown and Richard Q. Twiss found a way to circumvent this problem with the introduction of a new kind of interferometry now known as intensity interferometry[1]. In their "HBT" experiment, they were able to measure the spatial size of both laboratory mercury light sources and very distant stars by aiming two photomultiplier tubes at the source and measuring the correlations in intensities of light received. The positive correlation observed was correctly interpreted as a consequence of the bosonic nature of photons, which like to bunch together because of Bose-Einstein statistics.

In subsequent decades, colliding energies where two or more identical pions could be produced where reached in various elementary particle accelerators. Pions are also bosons, and in quantum mechanics the square of the (wave function) amplitude is the probability of the pion with an eigenmomentum $\mathbf{p}$ being at a point in space $\mathbf{x}$, thus the HBT intensity formalism for electromagnetic waves could be directly translated into the language and experimental

measurement of the high-energy physics laboratory.

In a nutshell, the intensity correlations used for electromagnetic waves could be directly translated into differential cross sections:

$$C(\mathbf{E}_1, \mathbf{E}_2) = \frac{\langle I_1 I_2 \rangle}{\langle I_1 \rangle \langle I_2 \rangle} \qquad \longrightarrow \qquad C(\mathbf{p}_1, \mathbf{p}_2) = \frac{(d^6\sigma/d^3\mathbf{p}_1\, d^3\mathbf{p}_2)}{(d^3\sigma/d^3\mathbf{p}_1)\,(d^3\sigma/d^3\mathbf{p}_2)} \tag{1.1.1}$$

where as usual σ is the binned particle count divided by the luminosity and $I = |\mathbf{E}|^2$. In both cases, C is called the "correlation function".

In 1959 Goldhaber et al. [2] were looking for the rho-meson in antiproton-proton annihilations and instead of finding the rho meson they discovered that identically charged pion piars are more likely emitted to the same hemisphere than the opposite hemisphere. They first simply reported this observation. Then this like-sign pion correlation was explained in terms of Bose-Einstein statistics in 1960 by G. Goldhaber, S. Goldhaber, W. Y. Lee and A. Pais [3] hence the name GGLP effect.

While in astronomy, intensity interferometry or the HBT method has been supplanted by improved amplitude interferometry, it has in high energy collisions grown into a precision tool to probe the size, shape, time development and freezeout conditions of the particle emitting particles.

Usually, experiments begin with the simplest case of identical charged pions, basically for two reasons: firstly, because pions are the lightest mesons and are thus produced in large quantities which improves the statistics of the experiment and secondly, because charged pions are relatively easy to identify. Particle mis-identification is known to reduce the observable Bose-Einstein correlations due to admixture of non-interfering particle pairs, that are not subject to Bose-Einstein symmetrization, to the data sample. Since the advent of good particle identification and copious production, HBT of identical bosons both of higher mass (charged and neutral kaons, etc) and even photons, which has zero mass and other inherent experimental difficulties, as well as fermionic identical pairs (proton-proton etc) are being studied systematically.

A major advance occurred when the formalism was extended to relate measured correlations to the theoretical "kernel", the squared two-particle wave function, of non-identical particle pairs. Improved particle identification has made it possible to measure such non-identical pairs accurately also. In this thesis, we shall restrict our focus to pairs of identical particles only.

While the differential cross section of Eq. 1.1.1 implies measurement of correlations in the six-dimensional space spanned by the momenta $(\mathbf{p}_1, \mathbf{p}_2)$, statistics forced experiments in the early days to measure correlations in lower-dimensional variables, for example the invariant

four-momentum difference

$$Q = \sqrt{-(p_1 - p_2)^2} = \sqrt{(\mathbf{p}_1 - \mathbf{p}_2)^2 - (E_1 - E_2)^2}. \tag{1.1.2}$$

With the advent of higher collision energies and larger colliding systems such as heavy ions, it also became possible to measure correlations in three-dimensional and sometimes even the full six-dimensional space. The topic for this thesis, systematic quantification of deviations from simple Gaussian shape of the correlation, was inspired by the experimental measurements in these higher-dimensional spaces as well as previous theoretical frameworks to describe such deviations [4]. This will be considered in more detail in Section 1.5. Before doing so, however, we briefly review the basic theoretical notions in Sections 1.2 and 1.3. Section 1.4, setting out the situation for particle emission regions that are exactly Gaussian, is a necessary prelude to considering non-Gaussian cases.

1.2 Correlations from Bose-Einstein statistics

Knowing that the correlation function is dependent on the spatial distribution of the source, the question we have to answer is how to connect the known experimental correlation function $C(\mathbf{p}_1, \mathbf{p}_2)$ to the unknown "emission function", the distribution of distances between emitted pions. To illustrate the basic principles, we consider here an oversimplified model and then introduce the more general case in the next section.

The essence of the effect is based on the Bose-Einstein symmetrised wave function for two identical noninteracting bosons emitted at points[1] $\mathbf{x}_a$ and $\mathbf{x}_b$ and propagating into the detectors, where two momenta $(\mathbf{p}_1, \mathbf{p}_2)$ are measured and which jointly form the eigenvalue of the two-particle wavefunction in six-dimensional momentum space,

$$\langle \mathbf{x}_a, \mathbf{x}_b \,|\, \mathbf{p}_1, \mathbf{p}_2 \rangle = \frac{1}{\sqrt{2}} \left[\langle \mathbf{x}_a \,|\, \mathbf{p}_1 \rangle \langle \mathbf{x}_b \,|\, \mathbf{p}_2 \rangle \;\pm\; \langle \mathbf{x}_a \,|\, \mathbf{p}_2 \rangle \langle \mathbf{x}_b \,|\, \mathbf{p}_1 \rangle \right], \tag{1.2.1}$$

where $\psi_{\mathbf{p}}(\mathbf{x}) = \langle \mathbf{x} \,|\, \mathbf{p} \rangle$ is the coordinate space representation of the one-particle eigenket with eigenvalue $\mathbf{p}$. The wave function is symmetrical because we are dealing with identical bosons; replacing the upper $+$ sign with the lower $-$ deals with the fermion case. We will be focusing on non-interacting identical bosons and ignore Coulomb and final-state interactions in this thesis. This leads us to the plane wave approximation for two identical free bosons,

$$\Psi_{12}(\mathbf{x}_a, \mathbf{x}_b) = \langle \mathbf{x}_a, \mathbf{x}_b \,|\, \mathbf{p}_1, \mathbf{p}_2 \rangle = \frac{1}{\sqrt{2}} \left(e^{i\mathbf{p}_1 \cdot \mathbf{x}_a + i\mathbf{p}_2 \cdot \mathbf{x}_b} + e^{i\mathbf{p}_1 \cdot \mathbf{x}_b + i\mathbf{p}_2 \cdot \mathbf{x}_a} \right). \tag{1.2.2}$$

Interpreting $|\Psi_{12}|^2$ as the probability that particles with measured momenta $(\mathbf{p}_1, \mathbf{p}_2)$ were emitted at points $(\mathbf{x}_a, \mathbf{x}_b)$, the number of particles counted in a given $(\mathbf{p}_1, \mathbf{p}_2)$ bin would be

[1] Note that we write three-dimensional quantities in **bold**.

the joint probability of emission $\rho_2(\mathbf{x}_a, \mathbf{x}_b)$ times $|\Psi_{12}|^2$, integrated over all space. Assuming that the emission of particles is statistically independent,

$$\rho_2(\mathbf{x}_a, \mathbf{x}_b) = \rho(\mathbf{x}_a)\,\rho(\mathbf{x}_b), \tag{1.2.3}$$

these considerations lead us to an expression connecting the measured correlation function $C(\mathbf{p}_1, \mathbf{p}_2)$ to the symmetrised wave function and emission probabilities,

$$C(\mathbf{p}_1, \mathbf{p}_2) = \int d^3\mathbf{x}_a\, d^3\mathbf{x}_b\, \rho(\mathbf{x}_a)\,\rho(\mathbf{x}_b)\,|\Psi_{12}|^2, \tag{1.2.4}$$

which for the noninteracting boson wave function Eq. 1.2.2 takes the simple form

$$C(\mathbf{q}, \mathbf{K}) = \int d^3\mathbf{x}_a\, d^3\mathbf{x}_b\, \rho(\mathbf{x}_a)\,\rho(\mathbf{x}_b)\;(1 + \cos(\mathbf{q}\cdot\mathbf{r})) \tag{1.2.5}$$

where $\mathbf{q} = \mathbf{p}_1 - \mathbf{p}_2$ and $\mathbf{r} = \mathbf{x}_a - \mathbf{x}_b$ are the momentum difference and spatial separation of the emission points respectively. These can clearly be seen as part of the coordinate transformations to sum and difference coordinates,

$$\mathbf{q} = \mathbf{p}_1 - \mathbf{p}_2, \tag{1.2.6}$$

$$\mathbf{K} = \tfrac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2), \tag{1.2.7}$$

$$\mathbf{r} = \mathbf{x}_a - \mathbf{x}_b, \tag{1.2.8}$$

$$\mathbf{s} = \tfrac{1}{2}(\mathbf{x}_a + \mathbf{x}_b). \tag{1.2.9}$$

We note that, while for static, non-expanding sources as in Eq. 1.2.3, $C(\mathbf{q}, \mathbf{K})$ does not depend on the sum variable $\mathbf{K}$, this is not true for expanding sources which naturally leads to a momentum dependent effective source size which is also seen in the experimental data.

Because both $C(\mathbf{q}, \mathbf{K})$ and $\rho(\mathbf{r})$ are real, the cosine can generally be written as $\exp(i\mathbf{q}\cdot\mathbf{r})$. The fact that $\cos(\mathbf{q}\cdot\mathbf{r})$ is even in $\mathbf{q}$ illustrates the general property that the correlation for identical particles must of necessity be even in $\mathbf{q}$ always,

$$C(-\mathbf{q}, \mathbf{K}) = C(\mathbf{q}, \mathbf{K}). \tag{1.2.10}$$

The property that identical bosons will lead to a symmetric transform and even functions will play a prominent role in the subsequent parts of this Thesis.

Defining the *relative distance distribution*

$$S(\mathbf{r}) = \int d^3\mathbf{s}\,\rho\left(\mathbf{s} + \tfrac{1}{2}\mathbf{r}\right)\rho\left(\mathbf{s} - \tfrac{1}{2}\mathbf{r}\right), \tag{1.2.11}$$

i.e. the probability that two particles are emitted at a spatial separation of $\mathbf{r}$, the correlation function becomes

$$C(\mathbf{q},\mathbf{K}) = \int d^3\mathbf{r}\left(1+e^{i\mathbf{q}\cdot\mathbf{r}}\right)\int d^3\mathbf{s}\,\rho\left(\mathbf{s}+\tfrac{1}{2}\mathbf{r}\right)\,\rho\left(\mathbf{s}-\tfrac{1}{2}\mathbf{r}\right) \tag{1.2.12}$$

$$= 1+\int d^3\mathbf{r}\,S(\mathbf{r})\,e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.2.13}$$

because $\int d^3\mathbf{r}S(\mathbf{r}) = 1$. For noninteracting bosons, the "true" correlator

$$R(\mathbf{q},\mathbf{K}) = C(\mathbf{q},\mathbf{K}) - 1 \tag{1.2.14}$$

is therefore the Fourier transform of the relative distance distribution,

$$R(\mathbf{q},\mathbf{K}) = C(\mathbf{q},\mathbf{K}) - 1 = \int d^3\mathbf{r}\,S(\mathbf{r})\,e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.2.15}$$

where again the $\mathbf{K}$ dependence happens to be absent for the noninteracting kernel $e^{i\mathbf{q}\cdot\mathbf{r}}$ but will be present in the general case. We can alternatively write Eq. 1.2.5 in the form

$$R(\mathbf{q},\mathbf{K}) = \int d^3\mathbf{x}_a\,d^3\mathbf{x}_b\,\rho(\mathbf{x}_a)\,\rho(\mathbf{x}_b)\,e^{i\mathbf{q}\cdot\mathbf{x}_a}e^{-i\mathbf{q}\cdot\mathbf{x}_b} \;=\; |\tilde{\rho}(\mathbf{q})|^2, \tag{1.2.16}$$

where

$$\tilde{\rho}(\mathbf{q}) = \int d^3\mathbf{x}\rho(\mathbf{x})e^{i\mathbf{q}\cdot\mathbf{x}} \tag{1.2.17}$$

is the Fourier transform of the individual source distribution.

Eqs. 1.2.15 and 1.2.16 both illustrate how the unknown source function $\rho(\mathbf{x})$ can be connected to a known measurement in the form of $C(\mathbf{q},\mathbf{K})$. Ideally, one would like to invert these integral equations to find explicit answers for $S(\mathbf{r})$ as a function of $C(\mathbf{q},\mathbf{K})$: this is the ambition of the "imaging" programme of Brown et al. [4]. Note that, due to the squaring of the amplitude $|\Psi_{12}|$, it will never be possible to measure distributions in absolute space coordinates $\mathbf{x}_a, \mathbf{x}_b$ or $\rho(\mathbf{x})$ directly: all we can get are quantities related to the square $|\rho(\mathbf{x})|^2$.

Staying within our simple model, we illustrate the basic physics by examining the correlation function resulting from a one-dimensional spatial distribution of emission points: if we assume that pions are emitted according to a Gaussian probability distribution (in coordinate space) with width σ_x,

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}}\exp(-\frac{x^2}{2\sigma_x^2}), \tag{1.2.18}$$

the correlator (in momentum space) is also Gaussian,

$$C(q) = 1+\exp(-\sigma_x^2 q^2). \tag{1.2.19}$$

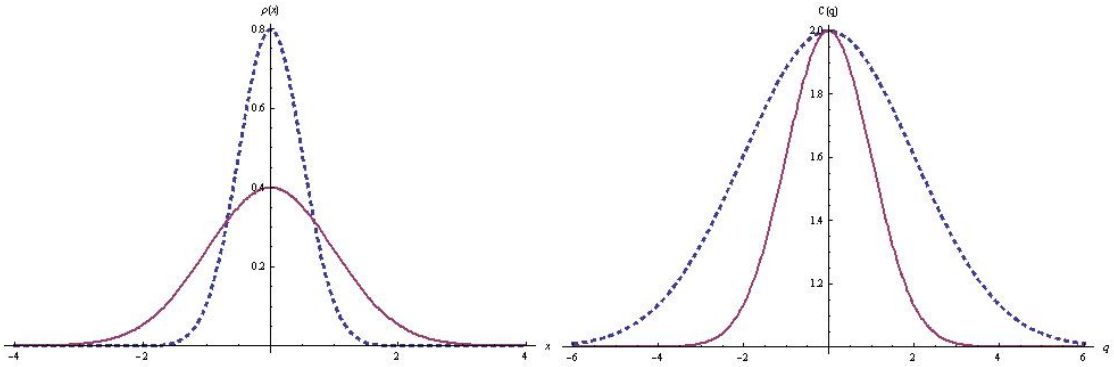


Figure 1.1: The Fourier transforms of narrow and wide Gaussians, showing how a narrow distribution in coordinate space (dotted line, left panel) translates into a wide distribution in momentum space (dotted line, right panel) while a wide distribution in coordinate space translates into a narrow one in momentum space (solid lines).

Writing this in terms of the width σ_q of the measured distribution,

$$C(q) = 1 + \exp(-\frac{q^2}{2\sigma_q^2}). \tag{1.2.20}$$

we see that

$$\sigma_x = \sqrt{2}/\sigma_q \tag{1.2.21}$$

whenever Gaussians are involved. (Normally $\sigma_x = 1/\sigma_q$, but the squares of quantities, which are inherent to the theory, lead to the $\sqrt{2}$ factor.)

Plotting two Gaussians for widths, $\sigma_x = \frac{1}{2}$ and $\sigma_x = 1$, together in Figure 1.1 illustrates the well-known connection of dual function space of Fourier transform pairs: Large-scale structures in q transform into small-scale structures in x. Thus information about small x is contained in the behaviour of $C(q)$ at large q and vice versa. The figure also shows that it is impossible to "squeeze" both a function and its Fourier transform to a delta function: the narrower the one, the wider the other becomes.

We finally note that the correlation function (Eq. 1.2.19) has a maximum value $C(0, \mathbf{K}) = 2$. Experimentally, $C(0, \mathbf{K})$ usually deviates from 2 due to various effects, among them the fact that sources are often not statistically independent as assumed above. Experiments hence often parametrise their Gaussian measurements by including a "intercept parameter" λ

$$C(q) = 1 + \lambda \exp(-\frac{q^2}{2\sigma_q^2}). \tag{1.2.22}$$

A third interference term has also been left out of Eq. 1.2.22, see [5], but the equation is sufficient for our discussion.

1.3 Wigner-function formulation

The toy model of the previous section is, as stated, an oversimplified view. For a start, there is an inherent contradiction in the interpretation of $\mathbf{x}_a$ and $\mathbf{x}_b$ as emission points of pions, thereby fixing their location, even while using the same quantities in the plane waves, which assume that the pions are entirely delocalised. Interactions, relativistic effects, frame dependences, and flow are but a few of the important effects that have to be considered in a proper description.

Of several more sophisticated approaches more in line with basic quantum mechanics, we briefly outline one based on Wigner phase-space densities $S(x, p)$ below. In developing the following formalism we follow Ref. [6]. A natural starting point is the Lorentz-invariant probability density function of each particle type. In the theory, the unnormalised distribution of a single particle is written in terms of a creation operator $\hat{a}^\dagger_{\mathbf{p}}$ and annihilation operator $\hat{a}_{\mathbf{p}}$ acting upon a "vacuum" $\langle\cdot\rangle$,

$$N_1(\mathbf{p}) = E\langle\hat{a}^\dagger_{\mathbf{p}}\hat{a}_{\mathbf{p}}\rangle. \tag{1.3.1}$$

The theoretical distribution is assumed to be equal to the experimentally measured Lorentz-invariant cross section, which counts the number of particles end up in a particular momentum bin times their energy,

$$N_1(\mathbf{p}) = E\frac{dN}{d^3\mathbf{p}} = E\langle\hat{a}^\dagger_{\mathbf{p}}\hat{a}_{\mathbf{p}}\rangle; \tag{1.3.2}$$

similarly for the two particle distribution one equates field theory and experimental cross section

$$N_2(\mathbf{p}_1, \mathbf{p}_2) = E_1E_2\frac{dN}{d^3\mathbf{p}_1 d^3\mathbf{p}_2} = E_1E_2\langle\hat{a}^\dagger_{\mathbf{p}_1}\hat{a}^\dagger_{\mathbf{p}_2}\hat{a}_{\mathbf{p}_2}\hat{a}_{\mathbf{p}_1}\rangle. \tag{1.3.3}$$

The experimental cross sections are normalized to the total number of particles,

$$\int\frac{d^3\mathbf{p}}{E}N_1(\mathbf{p}) = \langle N\rangle \tag{1.3.4}$$

and the two particle density distribution to the total number of particle pairs,

$$\int\frac{d^3\mathbf{p}}{E}N_2(\mathbf{p}_1, \mathbf{p}_2) = \langle N(N-1)\rangle, \tag{1.3.5}$$

where $\langle\ldots\rangle$ now indicates an experimental sample average. We then construct the two-particle correlator as the ratio of one- and two- particle spectra,

$$C(\mathbf{p}_1, \mathbf{p}_2) = \frac{N_2(\mathbf{p}_1, \mathbf{p}_2)}{N_1(\mathbf{p}_1)N_1(\mathbf{p}_2)}, \tag{1.3.6}$$

which apart from possible prefactors is also the experimental definition of the two-particle correlator.

Defining the relative four-momentum q^μ and the average four-momentum K^μ,

$$q^\mu = (E_1 - E_2, \mathbf{p}_1 - \mathbf{p}_2), \tag{1.3.7}$$

$$K^\mu = (\tfrac{1}{2}(E_1 + E_2), \tfrac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2)). \tag{1.3.8}$$

In analogy with our simplified model, the Wigner-function description of the two-particle correlator is built out of the emission function $S(x, p)$, in terms of four-vectors x and p. Using the wave function $|\Psi_{12}|^2$ as a weight, the correlator is

$$C(\mathbf{p}_1, \mathbf{p}_2) = \frac{\int d^4x_a\, d^4x_b\, S(x_a, p_1)\, S(x_b, p_2)\, |\Psi_{12}|^2}{\int d^4x_a\, S(x_a, K) \int d^4x_b\, S(x_b, K)}. \tag{1.3.9}$$

This relation can be derived in a number of ways, see [6]. The Wigner function $S(x, p)$ is the quantum mechanical equivalent of a classical phase-space distribution. It is real but not positive definite and when one of the variables x or p are integrated out, it gives a distribution in the other coordinate. Assuming identical, non-interacting particles and substituting plane waves into the equation, the result is

$$C(\mathbf{q}, \mathbf{K}) = 1 + \frac{\left|\int dx^4\, S(x, K)\, e^{iq\cdot x}\right|^2}{\left|\int dx^4\, S(x, K)\right|^2}. \tag{1.3.10}$$

There is a subtle difference between Eq. 1.3.10 and Eq. 1.3.9, Eq. 1.3.10 is always equal to or greater than one while this is not necessarily true for Eq. 1.3.9. A number of approximations are also needed to derive Eq. 1.3.10 from Eq. 1.3.9, most notably the smoothness approximation and the on-shell approximation:

- The *smoothness approximation* is the assumption that our emission function has a smooth enough dependence on the momentum so that one may approximate in the denominator
$$S(x_a, K - \tfrac{1}{2}q)\, S(x_b, K + \tfrac{1}{2}q) \approx S(x_a, K)\, S(x_b, K). \tag{1.3.11}$$
- The *on-shell approximation* is used because in principle the emission function depends on the off-shell momentum K, i.e. $K^\mu K_\mu = \frac{1}{2}(m^2 + E_1E_2 - \mathbf{p}_1 \cdot \mathbf{p}_2) \neq m^2$. This is a problem because no one knows the off-shell behaviour of S. Fortunately, it is possible to approximate the off-shell $K^0 = \frac{1}{2}(E_1 + E_2)$ by a "on-shell energy" $E_K = \sqrt{m^2 + \mathbf{K}^2}$ which ensures that $K^\mu K_\mu \simeq m^2$.

These approximations are necessary to simplify the four-dimensional theoretical Wigner function to the measured three-dimensional relative coordinate measurements.

It is important to notice that the emission function $S(x, K)$ is a scalar function over an eight-dimensional space spanned by the four-vectors x^μ and K^μ, while $C(\mathbf{q}, \mathbf{K})$ is a scalar

over the six-dimensional space spanned by $\mathbf{q}$ and $\mathbf{K}$. The necessary reduction to the accessible six dimensions comes in two steps.

Firstly, the four-vectors K^μ and q^μ are through their definition constrained by the so-called "mass-shell constraint",

$$K \cdot q = p_1^2 - p_2^2 = m_1^2 - m_2^2 = 0. \tag{1.3.12}$$

so that we can eliminate q^0 and write it in terms of a three-velocity $\beta = \mathbf{K}/K^0$

$$q^0 = \mathbf{q} \cdot \mathbf{K}/K^0 = \sum_{i=1}^{3} q_i \beta_i \tag{1.3.13}$$

The immediate result of this is that measurements of spacetime quantities depend only on the combination $(x_i - \beta_i t)$, i.e. we cannot separate information on the spatial point of emission from the time of emission [7]. This necessarily introduces a model-dependence if we want to reconstruct an emission function depending on the original eight-dimensional space.

The second reduction of a variable is effected by an integral over time. Eq. 1.3.9 can be written in a different form by working with the four-vector relative spacetime coordinate

$$r^\mu = (t_a - t_b, \mathbf{x}_a - \mathbf{x}_b), \tag{1.3.14}$$

in terms of which we define a relative source function, an integral over relative time $r^0 = t_a - t_b$,

$$S(\mathbf{r}, \mathbf{K}) = \frac{\int d^4x_a \, d^4x_b \, dr^0 \, S_1(x_a, p_1) \, S_2(x_b, p_2) \, \delta^4(r - x_a + x_b)}{\int d^4x_a \, S(x_a, p_1) \int d^4x_b \, S(x_b, p_2)} \tag{1.3.15}$$

in terms of which the correlator becomes

$$\boxed{C(\mathbf{q}, \mathbf{K}) - 1 = \int d^3\mathbf{r} \, S(\mathbf{r}, \mathbf{K}) \, \left[|\Psi_{12}(\mathbf{q}, \mathbf{r})|^2 - 1 \right]} \tag{1.3.16}$$

As in the toy model of the previous section, we therefore again obtain a direct relation between the measured correlator and the relative distance distributions. Eq. 1.3.16 therefore captures in a nutshell what we can and cannot learn about the structure of the source from the correlator.

The term $\left[|\Psi_{12}(q, r)|^2 - 1 \right]$ can be viewed as a kernel transformation, transforming between a coordinate-space basis and a relative-momentum space basis,

$$C(\mathbf{q}, \mathbf{K}) - 1 = \int d^3\mathbf{r} \, S(\mathbf{r}, \mathbf{K}) \, \mathcal{K}(\mathbf{r}, \mathbf{q}) \tag{1.3.17}$$

where in most theoretical cases, $\mathcal{K}$ turns out to be independent of the average momentum $\mathbf{K}$.

Substituting the free particle wave function and ignoring final-state Coulomb and strong interactions, we can reduce the correlator to

$$R(\mathbf{q},\mathbf{K}) = C(\mathbf{q},\mathbf{K}) - 1 = \int d^3\mathbf{r} S(\mathbf{r},\mathbf{K})\, \cos(\mathbf{q}\cdot\mathbf{r}). \tag{1.3.18}$$

Again like in the simple model we return to the Fourier transform, but not of the spatial distribution $\rho(x)$ but the relative source emission function $S(\mathbf{r},\mathbf{K})$. More complex interactions will lead to different kernels, but they should all reduce to the Fourier transform in the non-interacting case and thus exhibit some of the same behaviour.

1.4 Gaussian parametrizations of the correlator

While the dependence on the average momentum $\mathbf{K}$ is important in many ways, the transformation coupling $\mathbf{r}$ with $\mathbf{q}$ in the kernel transformations (Eqs. 1.3.16–1.3.18) implies that the formalism to be developed below is the same for any value of $\mathbf{K}$. For this reason we may drop $\mathbf{K}$ from $C(\mathbf{q},\mathbf{K})$ and $S(\mathbf{r},\mathbf{K})$ while always keeping in mind the implicit dependence on it.

The simplest descriptions of correlators, and the ones which were used traditionally, involve their parametrisation in terms of one or more known analytic probability density functions. As $C(\mathbf{q})$ is a scalar, its mathematical parametrisation must be a scalar function of $\mathbf{q}$.

Most common among such parametrisations is the three-dimensional Gaussian. Even though many experimental correlators are not Gaussian, it is still the starting point for a description of the correlation function. Also, much of this thesis will be concerned with systematic expansions, and the Gaussian will be needed later as the basis for one of the simplest expansions available.

1.4.1 The cartesian coordinate system

The question immediately arises which coordinate system to use for decomposing $\mathbf{q}$ into components. Bertsch [8] first introduced the so-called "cartesian" system, in which the direction of the three orthogonal axes are determined by physical considerations. In this *out-side-longitudinal (osl)* coordinate system, as shown in Figure 1.4.1, the longitudinal "*long*" or z-axis remains the same for all particle pairs. It is the most important direction in collisions as it corresponds to the beam axis of the colliding particles or nuclei (or, in the case of electron-positron collisions, the thrust axis).

The other two "*out*" and "*long*" axes, are oriented differently for each pion pair: the *out*- or x-direction is chosen parallel to the transverse component of the pair momentum $\mathbf{K}_\perp$ and the *side* or y-direction is the remaining axis orthogonal to the other two.

Because the *out* axis is always chosen to be parallel to $\mathbf{K}_\perp$, the transverse component of $\mathbf{K}$, the *side* component is by definition always zero. In the cartesian system, we therefore have

$$\mathbf{q} = q_o\hat{x} + q_s\hat{y} + q_l\hat{z}, \tag{1.4.1}$$

$$\mathbf{K} = K_\perp\hat{x} + K_l\hat{z}. \tag{1.4.2}$$

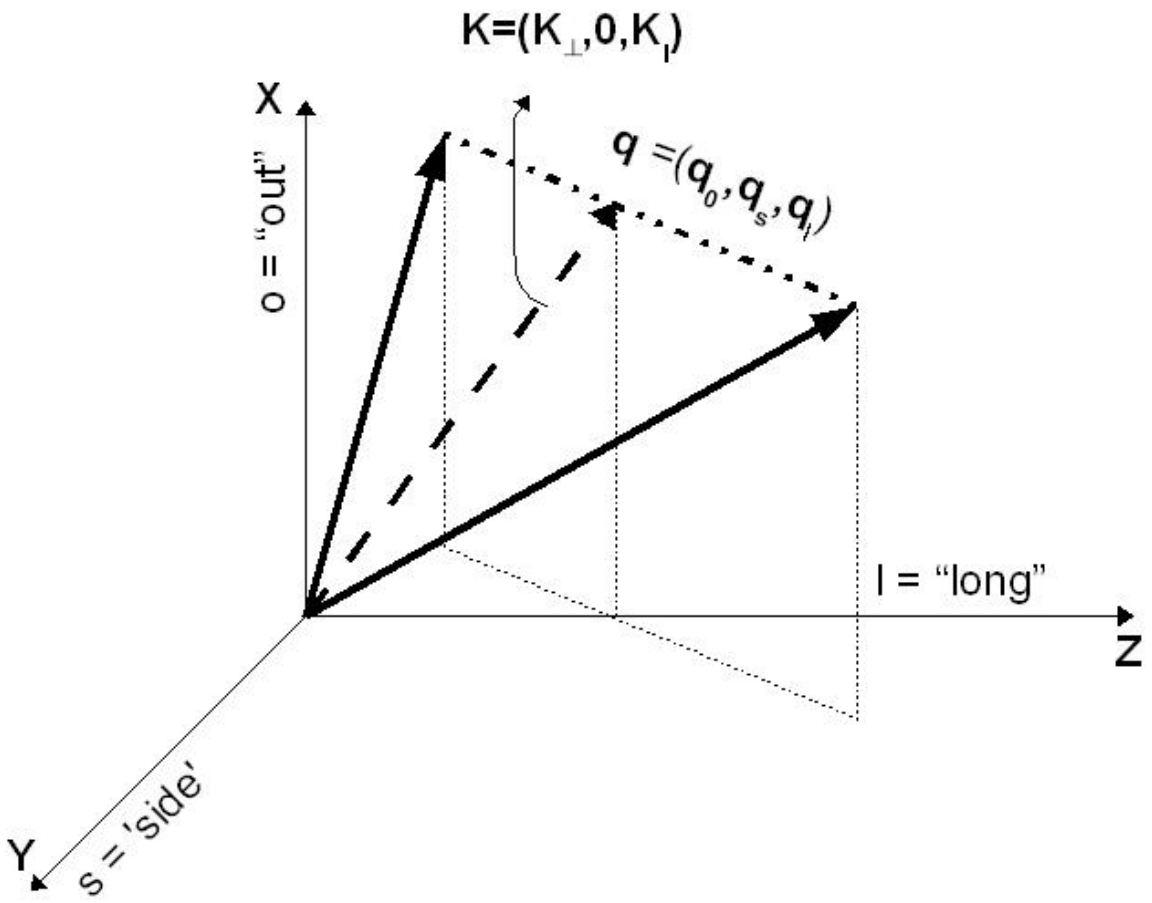


Figure 1.2: The cartesian coordinate system: the *long* axes is in the direction of the beam axis, the *out* direction is chosen parallel to the transverse component of the pair momentum $\mathbf{K}_\perp$ and the *side* direction is chosen perpendicular to both.

1.4.2 Gaussian parametrization

In the cartesian coordinate system, a Gaussian correlator can be written as

$$C(\mathbf{q}) - 1 = \lambda \exp\Big(-\sum_{i,j=o,s,l} R_{ij}^2 q_i q_j\Big), \tag{1.4.3}$$

where the R_{ij}^2 are free parameters. The more general form for a three-dimensional Gaussian is

$$C(\mathbf{q}) - 1 = \frac{\exp\left[-\frac{1}{2}\sum_{i,j}(q_i - \lambda_i)(\lambda^{-1})_{ij}(q_j - \lambda_j)\right]}{(2\pi)^{3/2}(\det\lambda)^{1/2}}, \tag{1.4.4}$$

where $\lambda_{i,j}$ are elements of the covariance matrix of the correlations between two components of $\mathbf{q}$, λ_i are the three components of the "offset", i.e. the displacement of the peak of the Gaussian from the origin, and λ^{-1} is the inverse to the covariance matrix.

The inverse matrix λ^{-1} is directly related to the expectation value of the emission points in coordinate space [7],

$$R_{ij}^2 = (\lambda^{-1})_{ij} = E[(x_i - \beta_i t)(x_j - \beta_j t)] \tag{1.4.5}$$

where $E[\cdot]$ is defined in terms of the normalised Wigner function

$$E[f] = \frac{\int dx^4\, f(x)\, S(x,K)}{\int dx^4 S(x,K)} \tag{1.4.6}$$

For Gaussian S and C, this establishes a direct relationship between the experimentally measurable "radii" $R_{i,j}$ and variances of the emission function S.

For central collisions or when the reaction plane of the collision is not determined, the data sample of $\mathbf{q}$ vectors will be azimuthally symmetric about the z-axis. As a result, the off-diagonal cross-terms R_{os} and R_{sl} must be identically zero. We already know that $\mathbf{K}$ is zero in the y-direction. Together, these simplifications lead to the relations

$$R_s^2 = \langle\tilde{y}^2\rangle, \tag{1.4.7}$$
$$R_o^2 = \langle(\tilde{x} - \beta_\perp\tilde{t})^2\rangle, \tag{1.4.8}$$
$$R_l^2 = \langle(\tilde{z} - \beta_l\tilde{t})^2\rangle, \tag{1.4.9}$$
$$R_{ol}^2 = \langle(\tilde{x} - \beta_\perp\tilde{t})(\tilde{z} - \beta_l\tilde{t})\rangle, \tag{1.4.10}$$
$$R_{os}^2 = 0, \tag{1.4.11}$$
$$R_{sl}^2 = 0. \tag{1.4.12}$$

where $\tilde{x} = x - E[x]$ etc, $E[x]$ being termed "offsets" in the literature.

In order to simplify matters even further, we can boost every pair along the longitudial axis to the frame where $(\mathbf{p}_1 + \mathbf{p}_2) \cdot \hat{z} = 0$, this is called the longitudinally comoving system (LCMS). The utility of the LCMS frame lies in the fact that $\beta_l = 0$, which leads to the simplification

$$R_l^2 = \langle \tilde{z}^2 \rangle, \tag{1.4.13}$$

Theoretical calculations are mostly carried out in the pair centre-of-mass system (PCMS), $\mathbf{p}_1 + \mathbf{p}_2 = 0$, in which $\beta_l = \beta_\perp = 0$ and hence

$$R_o^2 = \langle \tilde{x}^2 \rangle, \tag{1.4.14}$$

$$R_{ol}^2 = 0. \tag{1.4.15}$$

In all these coordinate system the symmetry of identical particles implies that the "offsets" are zero. In general the more symmetries the frame or the emission function, the fewer parameters are needed to describe the data.

1.4.3 q-moments

Traditionally, variances for the relative distance distribution were determined in terms of the R_{ij} by performing least-square fits to experimental data of $C(\mathbf{q})$. In their papers Heinz et al.[9], the authors suggested that they could also be found directly from $C(\mathbf{q})$ by direct measurement of the covariance matrix λ in Eq. 1.4.4 in terms of a normalised correlator

$$\lambda_{ij} = \langle q_i \, q_j \rangle = \frac{\int d^3\mathbf{q} \, q_i \, q_j \, [C(\mathbf{q}) - 1]}{\int d^3\mathbf{q} \, [C(\mathbf{q}) - 1]} \tag{1.4.16}$$

and then inverting this covariance matrix to obtain the matrix of $\lambda^{-1} = R_{ij}^2$ values. While this was implemented only in a few cases, it does relate to issues which will be raised later in this thesis.

Deviation of the correlator from a Gaussian shape can be characterized by the higher order "q-moments" such as $\langle q_i q_j q_k q_l \rangle$, which should also be directly measurable. The authors did notice that these moments can be determined in terms of derivatives of the relative source distribution, because for noninteracting pions, we can invert to find S in terms of C,

$$S(\mathbf{r}) = \int d^3\mathbf{q} e^{i\mathbf{r}\cdot\mathbf{q}} \, [C(\mathbf{q}, \mathbf{K}) - 1] \tag{1.4.17}$$

following from which the q-moments could be found from

$$\langle q_{i1} \cdots q_{in} \rangle = \frac{(-\partial)^n}{\partial_{ri1} \cdots \partial_{rin}} \ln S_K(\mathbf{r}) \Big|_{\mathbf{r}=0} \tag{1.4.18}$$

It is not hard to see that the relative source function, Eq. 1.3.15, and the generating function are one and the same. The q-moments does not necessarily exist and consequently the derivatives of Eq. 1.4.18 will also not exist. Another drawback is that if the correlator is not strictly greater than one the relations in this section does not hold, because of neglected correlation effects like long-lived resonances.

1.5 Beyond Gaussian

Realistic two-particle correlation functions deviate from a Gaussain slightly too significantly in some cases [10]. This implies that the space-time variance does not agree with the fitted Gaussian radius parameters and that these parameters do not contain all the information in $C(\mathbf{q}, \mathbf{K})$. Thus we must either augment the Gaussian radius parameters with more characterising parameters or we may proceed to reconstruct the correlation function directly from the information contained therein without using a specific parameterisation.

1.5.1 Imaging Methods

Imaging methods start from the equation:

$$C(\mathbf{q}, \mathbf{K}) - 1 = \int d^3\mathbf{r} K(\mathbf{q}, \mathbf{r}) S_{\mathbf{K}}(\mathbf{r}) \tag{1.5.1}$$

The kernel is $K(\mathbf{q}, \mathbf{r}) = |\Psi_{\mathbf{q}}(\mathbf{r})|^2 - 1$ where $\Psi_{\mathbf{q}}$ describes the propagation of a pair, which is created with a center of mass seperation r and detected with relative momentum $\mathbf{q}$. In the case of free particle propagation $K(\mathbf{q}, \mathbf{r}) = \cos(\mathbf{q} \cdot \mathbf{r})$. In practice inversion of Eq. 1.5.1 is complicated by finite measurement statistics and the higly oscillatory nature of the wave function. The function $S_K(\mathbf{r})$ is thus decomposed in a finite number of basis functions usually cartesian or spherical harmonics [11]. The coefficients are then computed directly from the data.

1.5.2 The Lévy index of stability

The lévy index of stability, $0 < \alpha \leq 2$ represents a family of two- and three-particle Bose Einstein correlation functions, [12]. If the underlying process consists of many random elementary steps and these steps have a finite mean and variance the Central Limit theorem states that Gaussian distribution will arise in the limit of the steps approaching infinity and $\alpha = 2$, if these steps do not have a finite variance or mean the result is one of the Lévy stable distributions in the limiting case and $\alpha < 2$. Thus the stable laws arise as a consequence of a

generalised Central Limit theorems and the correlator is then described in terms of these stable laws. This is a model independent characterization of the correlation function by special functions.

1.5.3 Gram-Charlier series

The Gram-Charlier series takes the idea of the q-moments further by using higher moments to characterize to correlator's deviation from Gaussian and reconstruct the correlator without using fitting [13]. The expansion is model independent, but does require a basis distribution. Normally the Gaussian distribution is used as a reference function with the same covariance and offsets as the density we are trying to approximate. The correction terms then involve cumulants and cumulant products as coefficients of the Hermite polynomials. We will discuss the relation between cumulants and moments in Section 2.3.1. More generally, different reference functions than the Gaussian could also be used. The correction terms will then involve derivatives of the approximating density and are not necessarily polynomials.

Chapter 2

Mathematical and statistical basics

While the ultimate aim is to apply our results to the full three-dimensional imaging of pion source distributions, the issues appearing in systematic expansions must first be understood and controlled in their one-dimensional forms. We therefore concentrate on the simplest one-dimensional forms of the relevant mathematical and statistical structures.

For a start, we abandon the nomenclature of momentum and coordinate space and revert to the symbols used in the mathematical and statistical literature. In this and the next chapters, the variable x is the one-dimensional equivalent of the three-momentum difference $\mathbf{q}$ of Chapter 1, while t is equivalent to the relative distance of emission $\mathbf{r}$.

In this chapter we will discuss the mathematical basics of the thesis, while in the next chapter we will discuss Gram-Charlier series and finally look at some numerical studies, on which we will base our conclusions.

2.1 Basic Probability and Measure

The equivalent to the experimental measurement of a three-momentum difference $\mathbf{q}$ for a pion pair is a single realisation x of an outcome of a random variable X, distributed according to some parent distribution. Assuming that the experimental correlation function $R(\mathbf{q})$ is continous on $\mathbb{R}$, the equivalent normalised **probability density function (PDF)**, which we will for the moment give the generic name $f(x)$, will also be assumed continuous. The set of possible outcomes for X or "support" of $f(x)$, will usually range anywhere from $-\infty$ to $+\infty$, and $f(x)$ is normalised accordingly,

$$\int_{-\infty}^{+\infty} dx\, f(x) = 1. \tag{2.1.1}$$

We shall usually omit the limits of integration except where they differ from $(-\infty, +\infty)$.

For any PDF, the following definitions will be needed:

Definition 1 *Given a function $h(x)$ bounded on $\mathbb{R}$ and a PDF $f(x)$, the expectation value of $h(x)$ with respect to f is defined as*

$$E_f[h(x)] \quad = \quad \int h(x)\, f(x)\, dx. \tag{2.1.2}$$

Ref. [14]

This definition can easily be extended to the multivariate case by replacing the integral over $\mathbb{R}^1$ by an integral over $\mathbb{R}^p$. We are interested in describing PDFs in terms of moments, so we define the *moments* of $f(x)$ as

Definition 2

$$\mu'_j \quad = \quad E_f[x^j] = \int_{\mathbb{R}} x^j\, f(x)\, dx. \tag{2.1.3}$$

*X has a **finite jth moment** if the integral converges. It can be shown that if X has a finite jth moment, then it has a finite moment of each order less than j. Ref. [14]*

In statistics, the moments μ'_j about a constant a are usually defined by $\int_\infty^\infty (x-a)^r\, f(x)\, dx$. If a is equal to the mean μ'_1, they are called the *central moments* or moments about the mean,

$$\mu_j = \int_{-\infty}^{\infty} (x - \mu'_1)^j\, f(x)\, dx$$

and the prime is dropped [15]. In our case, the mean of our densities will almost always be zero ($a = \mu_1 = 0$) and so the two definitions coincide, i.e. all moments will be central moments. Note that because we are interested in probability densities, the moment of order zero is $\mu_0 = \int f(x)\, dx = 1$. When we mention a moment of order j, we assume the integral (Eq. 2.1.3) converges.

We will also use the usual norms to investigate the difference between approximations of functions and the functions themselves:

Definition 3 *Norms*

$$\left\|f(x)\right\|_1 = \left(\int_a^b |f(x)|dx\right) \tag{2.1.4}$$

$$\left\|f(x)\right\|_2 = \left(\int_a^b |f(x)|^2 dx\right)^{1/2} \tag{2.1.5}$$

$$\left\|f(x)\right\|_\infty = \sup_{a<x<b} |f(x)| \tag{2.1.6}$$

define three different norms on the linear spaces $\mathcal{L}^1(a,b)$, $\mathcal{L}^2(a,b)$ *and* $\mathcal{L}^\infty(a,b)$ *respectively, these spaces consist of all complex-valued measurable functions such that* $\int_b^a |f|^1 dx < \infty$, $\int_b^a |f|^2 dx < \infty$ *and* $f < \infty$. *Ref. [14].*

Writing $\mathcal{L}^n(\mathbb{R})$ we imply $\mathcal{L}^n(-\infty, \infty)$.

2.2 Generating Functions and the Characteristic function

Consider the infinite series of moments

$$M_X(t) = 1 + t\mu_1 + \frac{1}{2!}t^2\mu_2 + \frac{1}{3!}t^3\mu_3 + \frac{1}{4!}t^4\mu_4 + \ldots = \sum_{j=0}^{\infty} \frac{t^j}{j!}\mu_j, \tag{2.2.1}$$

assuming that moments of all orders exist, and that the series converges. As $\mu_j = E[x^j]$, this can be compactly described as an expectation value:

Definition 4 *The moment generating function is defined as*

$$M_X(t) = E[e^{tx}] = \int e^{tx}\, f(x)\, dx. \tag{2.2.2}$$

From either Eq. 2.2.1 or Eq. 2.2.2, it is clear that the moments are also derivatives of $M_X(t)$ evaluated at $t = 0$,

$$\mu_j = \frac{d^j}{dt^j} M_X(t)\Big|_{t=0}.$$

The infinite series (Eq. 2.2.1) may of course be divergent for some real t, either because a higher-order moment could be infinite or because the moments increase so rapidly that the integral (Eq. 2.2.2) diverges even though the moments are finite. To overcome this problem, we define the characteristic function:

Definition 5 *The Characteristic function is defined as*

$$\overline{F_X(t)} = E[e^{itx}] = \int e^{itx}\, f(x)\, dx. \tag{2.2.3}$$

We have assumed that the density function exists but even if it did not we could have viewed the characteristic function as the Fourier transform of a probability measure $f(x)dx$. Our notation explicitly shows that the characteristic function is the complex conjugate of the Fourier transform of the density $\overline{F_X(t)}$, see Section 2.6. Since the integral of the density function is $\int_{\mathbb{R}} f(x)dx = 1$, we have

$$|\overline{F_x(t)}| \leq \int_{\mathbb{R}} |e^{itx}| f(x)dx = \int_{\mathbb{R}} f(x)dx = 1 \tag{2.2.4}$$

and the integral converges thus absolutely and uniformly in t. It may therefore be integrated and differentiated under the integral sign with respect to t, so that

$$\frac{1}{i^j}\,\frac{d^j}{dt^j}\overline{F_X(t)} = \int_{\mathbb{R}} e^{itx}\, x^j\, f(x)\, dx \tag{2.2.5}$$

and evaluating the expression at zero gives

$$(-i)^j\,\frac{d^j}{dt^j}\overline{F_X(t)}\bigg|_{t=0} = \int^{\mathbb{R}} x^j f(x)dx = \mu_j \tag{2.2.6}$$

provided that μ_j exists. Inserting this into the Taylor expansion for $\overline{F_X(t)}$, this can also be written as a moment expansion,

$$\overline{F_X(t)} = 1 + it\mu_1 + \frac{1}{2!}(it)^2\mu_2 + \frac{1}{3!}(it)^3\mu_3 + \ldots \;=\; \sum_{j=0}^{\infty} \frac{(it)^j}{j!}\mu_j, \tag{2.2.7}$$

If only the first j moments exist, the first j derivatives of $\overline{F_X(t)}$ will exist at $t = 0$.

Clearly the formalism for $M_X(t)$ and $\overline{F(t)}$ is identical apart from the factor i,

$$\overline{F_X(t)} = M_X(it),$$

and so we may use one or the other interchangeably, reverting to the characteristic function whenever convergence is an issue.

2.3 Cumulants

Some of the series expansions explored in this thesis make use of *cumulants* rather than moments. Cumulants are most easily defined through the "cumulant generating function" which is the logarithm of the moment generating function,

$$K_X(t) = \log M_X(t). \tag{2.3.1}$$

In analogy with Eq. 2.2.7, the expansion of the cumulant generating function has the set of cumulants as coefficients,

$$K_X(t) = \kappa t + \frac{1}{2!}t^2\kappa_2 + \frac{1}{3!}t^3\kappa_3 + \frac{1}{4!}t^4\kappa_4 + \cdots = \sum_{j=1}^{\infty} \frac{t^j}{j!}\kappa_j. \tag{2.3.2}$$

Just as the characteristic function of any $f(x)$ is determined by the complete set of moments $\mu_j, j = 0, 1, 2, \ldots$, it may also be specified in terms of the set of cumulants $\kappa_j, j = 0, 1, 2, \ldots$. The question then naturally arises why they should be preferred over the moments. The following desirable properties motivate to some extent their usage:

- Most statistical calculations using cumulants are simpler than the corresponding calculations using moments.
- The cumulant of the sum of n independent random variables y_k,

$$X = \sum_{k=1}^{n} Y_k, \tag{2.3.3}$$

  is equal to the sum of cumulants of each variable,

$$\kappa_j(X) = \kappa_j(\textstyle\sum_{k=1}^{n} Y_k) = \displaystyle\sum_{k=1}^{n} \kappa_j(Y_k). \tag{2.3.4}$$

- Where two variables in the case of multivariate distribution are statistically independent, the cross cumulants involving those variables are zero. A simple example is given by the well-known covariance (termed $\kappa_{1,1}$ in the language of cumulants) which is zero if the two variables are independent,

$$\mathbf{cov}(X, Y) = E(XY) - E(X)\,E(Y) = 0 \tag{2.3.5}$$

- Gram-Charlier Type A series used for approximations to distributions are most conveniently expressed using cumulants.

Unlike moments, cumulants also behave very simply under affine transformations from X to a new random variable Y

$$Y = a + bX. \tag{2.3.6}$$

This is easily shown through the generating function

$$\begin{aligned} M_Y(t) &= E[\exp\{t(a + bX)\}] && (2.3.7)\\ &= \exp(at)M_X(bt), && (2.3.8) \end{aligned}$$

so the cumulant generating functions are related by

$$K_Y(t) = at + K_X(bt), \tag{2.3.9}$$

giving us the relations between cumulants

$$\kappa_1(Y) = a + b\kappa_1(X), \quad \kappa_2(Y) = b^2\kappa_2(X), \quad \kappa_3(Y) = b^3\kappa_3(X), \tag{2.3.10}$$

and so on. The change of origin affects only the mean or first cumulant. For this reason, cumulants are sometimes called semi-invariants.

Using this property we can **standardize** our distribution, by dividing our random variable with the square root of the variance, $\sqrt{\kappa_2}$, and substracting the mean κ_1,

$$Y = \frac{\sum_{k=1}^{n} X_k - \mu_1(X)}{\sqrt{\kappa_2(X)}}, \tag{2.3.11}$$

giving the cumulants of the standardized variable Y as:

$$\kappa_1(Y) = 0, \quad \kappa_2(Y) = 1, \quad \kappa_3(Y) = \frac{\kappa_3(X)}{\sqrt{\kappa_2(X)}^3}, \text{ etc.} \tag{2.3.12}$$

2.3.1 Relationship between cumulants and moments

The relations between moments and cumulants can be derived through their respective expansions and Eq. 2.3.1.

$$\exp\left(\sum_{j=1}^{\infty} \frac{t^j}{j!}\kappa_j\right) = \sum_{j=0}^{\infty} \frac{t^j}{j!}\mu_j. \tag{2.3.13}$$

Expanding the exponential

$$1 + t\kappa^i + t^2(\kappa_2/2! + \kappa_1^2/2!) + t^3\kappa_3/3! + t^4\kappa_4/4! + t^3\kappa_1\kappa_2/2! + t^4\{\kappa_1\kappa_3/6 + \kappa_2\kappa_2/8\} \tag{2.3.14}$$

$$+ t^3\kappa_1^3/3! + t^4\kappa_1^2\kappa_2/4 + t^4\kappa_1^4/4 + \ldots \tag{2.3.15}$$

and equating coefficients of t^j, we find

$$\mu_1 = \kappa_1, \tag{2.3.16}$$

$$\mu_2 = \kappa_2 + \kappa_1^2, \tag{2.3.17}$$

$$\mu_3 = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3, \tag{2.3.18}$$

$$\mu_4 = \kappa_4 + 4\kappa_1\kappa_3 + 3\kappa_2^2 + 6\kappa_1^2\kappa_2 + \kappa_1^4, \tag{2.3.19}$$

$$\mu_5 = \kappa_5 + 5\kappa_1\kappa_4 + 10\kappa_2\kappa_3 + 10\kappa_1^2\kappa_3 + 15\kappa_1\kappa_2^2 + 10\kappa_1^3\kappa_2 + \kappa_1^5 \qquad \text{etc.} \tag{2.3.20}$$

If $\kappa_1 = 0$, the formulae simplify to

$$\mu_1 = \kappa_1, \tag{2.3.21}$$
$$\mu_2 = \kappa_2, \tag{2.3.22}$$
$$\mu_3 = \kappa_3, \tag{2.3.23}$$
$$\mu_4 = \kappa_4 + 3\kappa_2^2, \tag{2.3.24}$$
$$\mu_5 = \kappa_5 + 10\kappa_2\kappa_3, \tag{2.3.25}$$
$$\mu_6 = \kappa_6 + 15\kappa_2\kappa_4 + 10\kappa_3^2 + 15\kappa_2^3. \tag{2.3.26}$$

Formulae giving cumulants in terms of moments may be found by formal inversion of Eq. 2.3.14 or by expansion of $\log[M_X(\xi)]$ and combining terms. The expressions obtained for the first four cumulants are

$$\kappa_1 = \mu_1, \tag{2.3.27}$$
$$\kappa_2 = \mu_2 - \mu_1^2, \tag{2.3.28}$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3, \tag{2.3.29}$$
$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4. \tag{2.3.30}$$

2.3.2 Intuitive interpretation of cumulants

In the univariate case, the first four cumulants can be directly identified with the concepts location, variance, skewness and kurtosis, all of which have a natural interpretation in terms of the shape of the marginal distributions. See Fig. 2.1 and for a simplified illustration.

Due to symmetry, the skewness $\kappa_3/\kappa_2^{3/2}$ of the distributions we shall be considering will always be zero. Kurtosis, defined as

$$\gamma_4 = \frac{\kappa_4}{\kappa_2^2} \tag{2.3.31}$$

will, however, be especially important later on. Distributions with kurtosis equal to zero are called mesokurtic. Those for which it is positive are called leptokurtic and those with negative kurtosis are platykurtic. All the cumulants of order higher than two also gives a sense of the non-Gaussian nature of the function as all higher cumulants $\kappa_{j\geq 3}$ are zero for the Gaussian distribution.

2.4 Orthogonality

The Gram-Charlier series treated in the next chapter reduces to an orthogonal system in certain cases, and we will later contrast it with a system of orthogonal polynomials. We

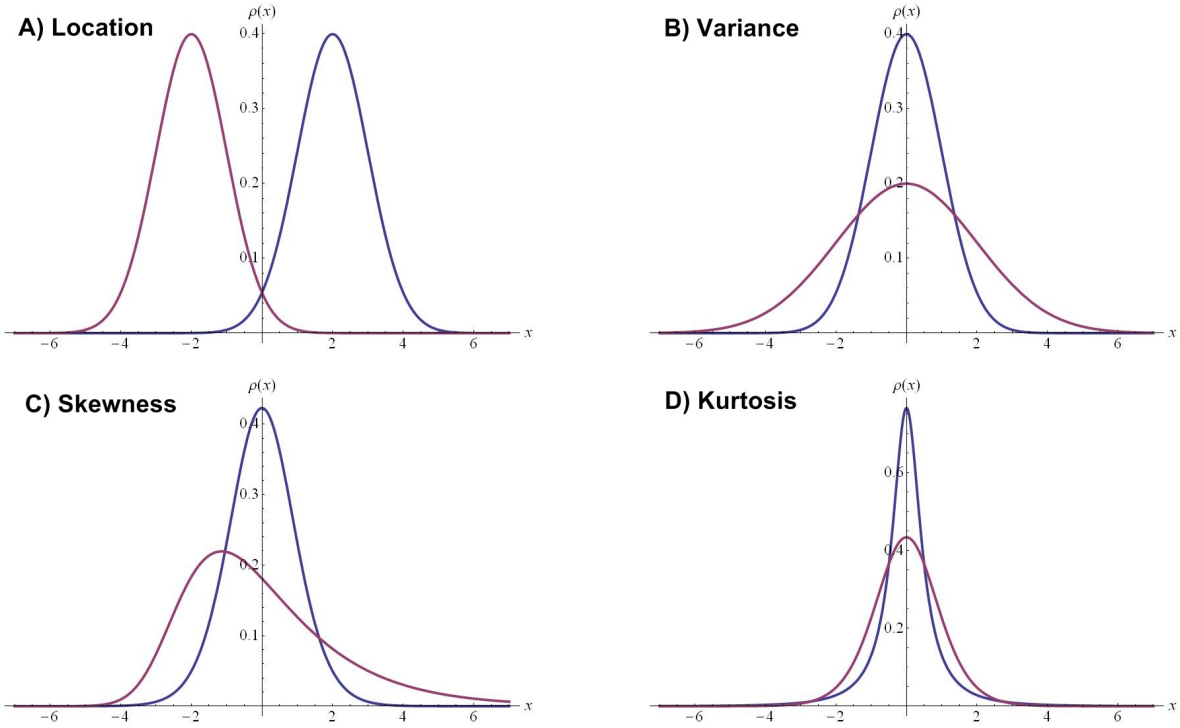


Figure 2.1: Location, Variance, Skewness and Kurtosis: κ_1 (Top left) determines the location of the distribution, while κ_2 (Top right) is a measure of its width,κ_3 (Bottom left) is a measure of skewness, i.e. how asymmetric a distribution is, while κ_4 (Bottom right) is related to kurtosis, which is directly related to the excess or deficit in the "tails" of the distribution at large $|x|$ compared to a Gaussian.

therefore need to introduce the concept of orthogonality. First, though, we have to clarify the meaning of linear independence and orthonormality:

Let $f_1(x), f_2(x), \ldots, f_\ell(x)$ be a set of linearly independent functions. Linear independence is defined by the following property: if

$$\sum_{j=1}^{\ell} c_j f_j(x) = 0 \qquad \Rightarrow \qquad c_j = 0 \,\forall j. \tag{2.4.1}$$

Linear independence can be checked by using the Wronskian

$$W(f_1, \ldots, f_n) = \begin{vmatrix} f_1 & f_2 & \cdots & f_n \\ f_1' & f_2' & \cdots & f_n' \\ \vdots & \vdots & \cdots & \vdots \\ f_1^{(n)} & f_2^{(n)} & \cdots & f_n^{(n)} \end{vmatrix} \quad \text{for every } x \text{ in the support.} \tag{2.4.2}$$

If the Wronskian is non-zero somewhere in the domain of the functions, they are linearly independent.

Definition 6 *A set of functions* $\{\phi_0(x), \phi_1(x), \ldots, \phi_\ell(x)\}$, ℓ *finite or infinite, is orthonormal with respect to a weight function* $w(x)$ *if*

$$(\phi_n, \phi_m) = \int_a^b \phi_n(x)\phi_m(x)\, w(x)\, dx = \delta_{nm}, \qquad n, m = 0, 1, 2, \ldots, \ell \tag{2.4.3}$$

Here $\phi_n(x)$ *is real-valued and belongs to the class* $\mathcal{L}^2(a, b)$. *Ref. [16].*

Quoting Szegő [16], we then have the following theorem:

Theorem 1 *Let the real-valued functions*

$$f_0(x), f_1(x), \ldots, f_\ell(x), \qquad \ell \text{ finite or infinite,} \tag{2.4.4}$$

be of the class $\mathcal{L}^2(a, b)$ *and linearly independent. Then an orthonormal set of functions*

$$\phi_0(x), \phi_1(x), \phi_2(x), \ldots, \phi_\ell(x) \tag{2.4.5}$$

exists such that, for $j = 0, 1, 2, \ldots, \ell$

$$\phi_j(x) = c_{j0}f_0(x) + c_{j1}f_1(x) + \cdots + c_{jj}f_j(x), \qquad c_{jj} > 0. \tag{2.4.6}$$

The orthonormal set (Eq. 2.4.5) is uniquely determined. Ref. [16].

These functions are necessarily linearly independent.

The procedure for constructing these functions is called **orthogonalization**. They can be constructed from inner products and determinants by the Gram-Schmidt like procedure, which can be summarised in analytic form:

$$\phi_\ell(x) = (D_{\ell-1}D_\ell)^{-1/2} \begin{vmatrix} (f_0, f_0) & (f_0, f_1) & \ldots & (f_0, f_\ell) \\ (f_1, f_0) & (f_1, f_1) & \ldots & (f_1, f_\ell) \\ \vdots & \vdots & \ldots & \vdots \\ (f_{\ell-1}, f_0) & (f_{\ell-1}, f_1) & \ldots & (f_{\ell-1}, f_\ell) \\ f_0(x) & f_1(x) & \ldots & f_\ell(x) \end{vmatrix}, \qquad n = 0, 1, 2, \ldots \tag{2.4.7}$$

where, for $\ell \geq 1$,

$$D_\ell = \begin{vmatrix} (f_0, f_0) & (f_0, f_1) & \ldots & (f_0, f_\ell) \\ (f_1, f_0) & (f_1, f_1) & \ldots & (f_1, f_\ell) \\ \vdots & \vdots & \ldots & \vdots \\ (f_\ell, f_0) & (f_\ell, f_1) & \ldots & (f_\ell, f_\ell) \end{vmatrix} \tag{2.4.8}$$

and we define $D_{-1} = 1$ and $D_0(x) = f_0(x)$.

We can now with our set $\phi_n(x)$ of orthogonal functions express an arbitrary function in terms of Fourier coefficients:

Definition 7 *Let $\{\phi_n(x)\}$ be a given orthonormal set, finite or infinite. To an arbitrary real-valued function $f(x)$ let there correspond the formal Fourier expansion*

$$f(x) \approx c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_\ell\phi_\ell(x) + \ldots. \tag{2.4.9}$$

The coefficients c_n, called the Fourier coefficients of $f(x)$ with respect to the given system, are defined by

$$c_\ell = (f, \phi_\ell) = \int_a^b f(x)\phi_\ell(x)\, w(x)\, dx, \qquad \ell = 0, 1, 2, \ldots. \tag{2.4.10}$$

Ref. [16].

This series has the following very useful minimum property:

Theorem 2 *Let $\phi_j(x), f(x), c_j$ have the same meaning as in the previous definition. Let $l \geq 0$ be a fixed integer, and $a_0, a_1, \ldots, a_l$ arbitrary real constants. If we write*

$$g(x) = a_0\phi_0(x) + a_1\phi_1(x) + \cdots + a_l\phi_l(x), \tag{2.4.11}$$

and the coefficients a_j are variable, the integral

$$\int_a^b \{f(x) - g(x)\}^2\, w(x)\, dx \tag{2.4.12}$$

becomes a minimum if and only if $a_j = c_j, v = 0, 1, 2, \ldots, l$. Ref. [16]

2.5 Orthogonal Polynomials

If we choose our set of linear independent functions as the powers of x, all inner products in equation Eqs. 2.4.7 and 2.4.8 become the moments of the weight function $w(x)$ and the determinant D_n,

$$D_n = \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \vdots & \vdots & \ldots & \vdots \\ \mu_n & \mu_{n+1} & \ldots & \mu_{2n} \end{vmatrix}, \tag{2.5.1}$$

is called a Hankel determinant. It can be proven positive [17]. The set of orthogonal functions $\phi_j(x)$ then reduces to a set of polynomials, which we shall denote by $P_j(x)$.

Theorem 3 *Given a positive, continous weight function $w(x)$ on $\mathbb{R}$ with infinite support and finite moments, there exists a unique sequence of polynomials $\{P_j(x)\}_{j=0}^{\infty}$,*

$$P_j = x^j + \textit{lower order terms}, \qquad j = 0, 1, \ldots, \tag{2.5.2}$$

and a sequence of positive numbers $\{\xi_n\}_0^{\infty}$, with $\xi_0 = 1$ such that

$$\int_{\mathbb{R}} P_m(x) P_n(x)\, w(x)\, dx = \xi_n \delta_{m,n} \tag{2.5.3}$$

Ref. [17].

This can be proven by using induction and applying orthogonalization. The polynomials can be found from the expression

$$P_n(x) = \frac{1}{D_{n-1}} \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \vdots & \vdots & \ldots & \vdots \\ \mu_{n-1} & \mu_n & \ldots & \mu_{2n-1} \\ 1 & x & \ldots & x^n \end{vmatrix} \tag{2.5.4}$$

and the constants ξ_n are given by

$$\xi_n = \int_{\mathbb{R}} P_n^2(x)\, w(x)\, dx = \int_{\mathbb{R}} x^n P_n(x) dx = \frac{1}{D_{n-1}} \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_n \\ \mu_1 & \mu_2 & \ldots & \mu_{n+1} \\ \vdots & \vdots & \ldots & \vdots \\ \mu_n & \mu_{n+1} & \ldots & \mu_{2n} \end{vmatrix} \tag{2.5.5}$$

which is just

$$\xi_n = \frac{D_n}{D_{n-1}}. \tag{2.5.6}$$

2.6 Fourier transforms

The Fourier transform is an essential tool for the Gram-Charlier series and for the basic physics of the experiment as set out in Chapter 1. We will first discuss the more general class of integral transforms to which the Fourier transform belongs, connect the transform with the characteristic function and then state a few properties which we will need later.

While in Chapter 1 the primary space is the momentum space spanned by $(\mathbf{q}, \mathbf{K})$, we shall in this and later chapters use the notation x to denote the equivalent variable. The dual space in one dimension will be spanned by the one-dimensional variable t, equivalent to "coordinate space" $\mathbf{x}$ in Chapter 1. For more details, see the introduction to Chapter 1.

2.6.1 Integral Transforms

An integral transform Tf of a function f takes the following general form:

$$(Tf)(t) = \int_a^b K(t,x)\, f(x)\, dx. \tag{2.6.1}$$

This maps a function f in x-space onto another function Tf in t-space. This interpretation has physical significance in the time-frequency relation of Fourier transforms and in the momentum-space vs coordinate space relations in the Bose-Einstein effect. There is a infinite number of possible transforms each specified by a choice of the function $K(t,x)$ called the kernel or nucleus of the transform. The integral transform is a linear operator since the integral is a linear operator.

2.6.2 The characteristic function as a Fourier transform

We define the Fourier transform $F(t)$ of $f(t)$ as

$$F(t) = \int_{\mathbb{R}} f(x) e^{-itx} dx \tag{2.6.2}$$

Some comments: Firstly, the factor 2π will appear in the inverse transform, Eq. 2.6.6; this is mostly a choice of aesthetics in general and specifically we want the characteristic function and transform to be as close as possible. Secondly, we use a capital letter to denote the transform of the equivalent lower-case letter.

We also write $\overline{\mu_j}$ for the complex conjugate of μ_j. It is now easy to see that the characteristic function is the complex conjugate of the Fourier transform of the probability density, given that $f(x)$ and its moments are real, $\overline{\mu_j} = \mu_j$,

$$\overline{F_X(t)} = \overline{\int_{\mathbb{R}} e^{-itx} f(x)\, dx} = \int_{\mathbb{R}} e^{itx} f(x)\, dx \tag{2.6.3}$$

The complex conjugation can be dropped in any case when working with even functions.

Two simplifications of the Fourier transform for even and uneven functions are the Fourier cosine and the Fourier sine transforms:

$$F_c(x) = \int_0^\infty f(x)\, \cos(tx)\, dx \qquad \text{if} \quad f(x) = f(-x), \tag{2.6.4}$$

$$F_s(x) = \int_0^\infty f(x)\, \sin(tx)\, dx \qquad \text{if} \quad f(x) = -f(-x). \tag{2.6.5}$$

Due to the symmetry of identical particles, we will primarily be interested in the cosine transform and we can use e^{itx} and $\cos(tx)$ interchangeably.

2.6.3 Inverse Fourier transforms

The inverse Fourier transform is

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(t) e^{itx} dt. \tag{2.6.6}$$

A few examples of symmetrical distributions and their Fourier transform are given in figure 2.2.

The Fourier transform pair in three-dimensional space are integrals over $\mathbb{R}^3$,

$$F(\mathbf{t}) = \int f(\mathbf{x} e^{-i\mathbf{t}\cdot\mathbf{x}} d^3x, \tag{2.6.7}$$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^3} \int F(\mathbf{t}) e^{i\mathbf{t}\cdot\mathbf{x}} d^3t. \tag{2.6.8}$$

2.6.4 Properties of the Fourier transforms

2.6.4.1 Convolution Theorem

Definition 8 *We define the operation*

$$f \star g = \int_{\mathbb{R}} f(x)\, g(x-y)\, dy \tag{2.6.9}$$

as the convolution of the two functions f and g over the interval $(-\infty, \infty)$. This is also the probability density of the sum of two random variables.

Introducing the inverse Fourier transform, we find

$$\begin{aligned} \int_{\mathbb{R}} g(y) f(x-y) dy &= \frac{1}{2\pi} \int_{\mathbb{R}} g(y) \int_{\mathbb{R}} F(t) e^{-it(x-y)} dt dy = \frac{1}{2\pi} \int_{\mathbb{R}} F(t) \left[\int_{\mathbb{R}} g(y) e^{ity} dy \right] e^{-itx} dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} F(t) \overline{G(t)} e^{-itx} dt, \end{aligned}$$

and so the Fourier transform of $f \star g$ is $F\,\overline{G}$.

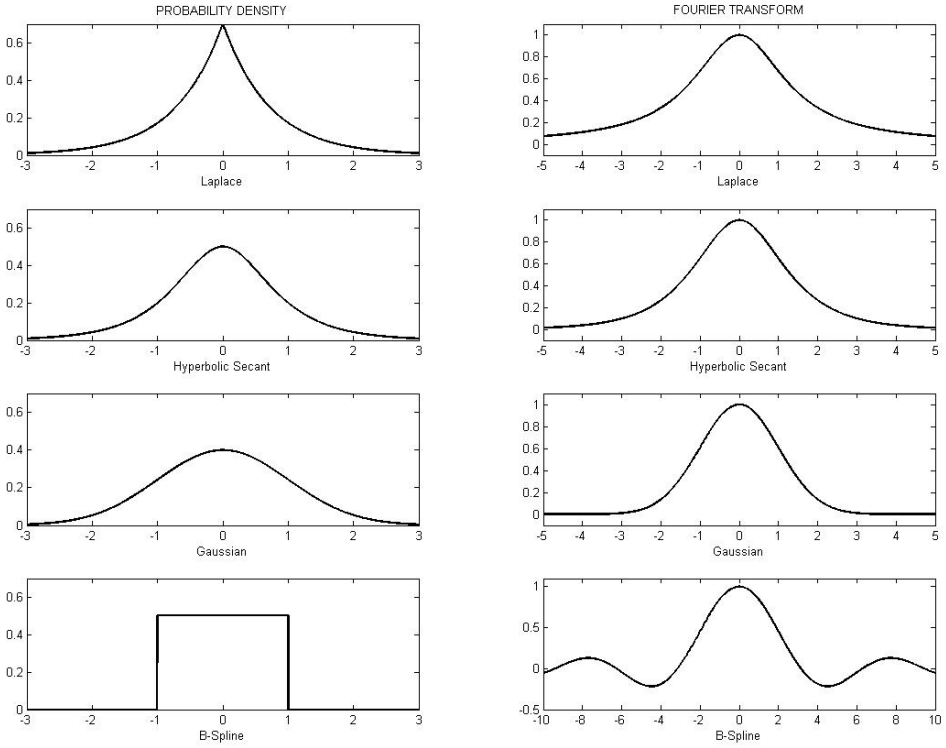


Figure 2.2: Graphical examples of some distributions and their Fourier transforms. The formulas for the distributions can be found in Table 2.1 and the free parameters have been chosen to give standardised distributions i.e. $\mu_1 = 0, \mu_2 = 1$.

2.6.4.2 Fourier transform of derivatives

Extending the Fourier transform $F(t) = \int_{-\infty}^{\infty} f(x)e^{-itx}dx$ to derivatives, we define

$$F_1(t) = \int_{-\infty}^{\infty} \frac{df(x)}{dx} e^{-itx} dx. \tag{2.6.10}$$

Integrating by parts, we obtain

$$F_1(t) = e^{-itx} f(x)|_{-\infty}^{\infty} - it \int_{-\infty}^{\infty} f(x)e^{-itx}dx. \tag{2.6.11}$$

If $f(x)$ vanishes as $x \to \pm\infty$, which apart for some special cases has to occur for the Fourier transform of $f(x)$ to exist, we have

$$F_1(t) = -itF(t), \tag{2.6.12}$$

that is, the transform of the derivative $f'(x)$ is $(-it)$ times the transform of the original function. This may be generalized to

$$F_j(t) = \int e^{-itx} \frac{d^j}{dx^j} f(x)\, dx \;=\; (-it)^j F(t) \qquad j = 0, 1, 2, \cdots . \tag{2.6.13}$$

2.6.4.3 Hankel transform

If the function $f(x, y) = f(r)$ depends only on the radial component $r = \sqrt{x^2 + y^2}$, the two-dimensional Fourier-transform reduces to the Hankel transform. Let $x + iy = re^{i\theta}$ and $u + iv = qe^{i\phi}$ so that

$$x = r\cos\theta, \qquad y = r\sin\theta, \tag{2.6.14}$$

$$r = \sqrt{x^2 + y^2}, \qquad q = \sqrt{u^2 + v^2}, \tag{2.6.15}$$

$$u = q\cos\phi, \qquad v = q\sin\phi. \tag{2.6.16}$$

Then the Fourier transform

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r) e^{-i(ux+vy)} dx\, dy \tag{2.6.17}$$

reduces to

$$\begin{aligned} F(q) &= \int_0^{\infty} \int_0^{2\pi} f(r) e^{-irq(\cos(\theta-\phi))} r dr d\theta = \int_0^{\infty} \int_0^{2\pi} f(r) e^{-irq\cos\theta} r dr d\theta \\ &= 2\pi \int_0^{\infty} f(r) J_0(qr) r dr, \end{aligned} \tag{2.6.18}$$

where J_0 is the Bessel function of the first kind.

2.7 Table of symmetrical probability density functions

For later use and comparision, we tabulate on the next page some common and some useful, if less well known, probability density functions. All are by design symmetric about x=0. The numbers in Column 1 refer to the relevant exercises of [15]. Note that the formula for the Beta distribution is wrong in the Sixth Edition of [15] but correct in the Fifth Edition. For the Student's t distribution, moment of order r only exists if $r < 2m - 1$, where m is a free parameter in the Student's t distribution, see Table 2.1.

In Figure 2.3, the positive half of some of these is shown in graphical form on a linear as well as a logarithmic scale. As cumulants are sensitive to the tails of distributions, the log scale is a useful reference to illustrate the sign of kurtosis for the different distributions.

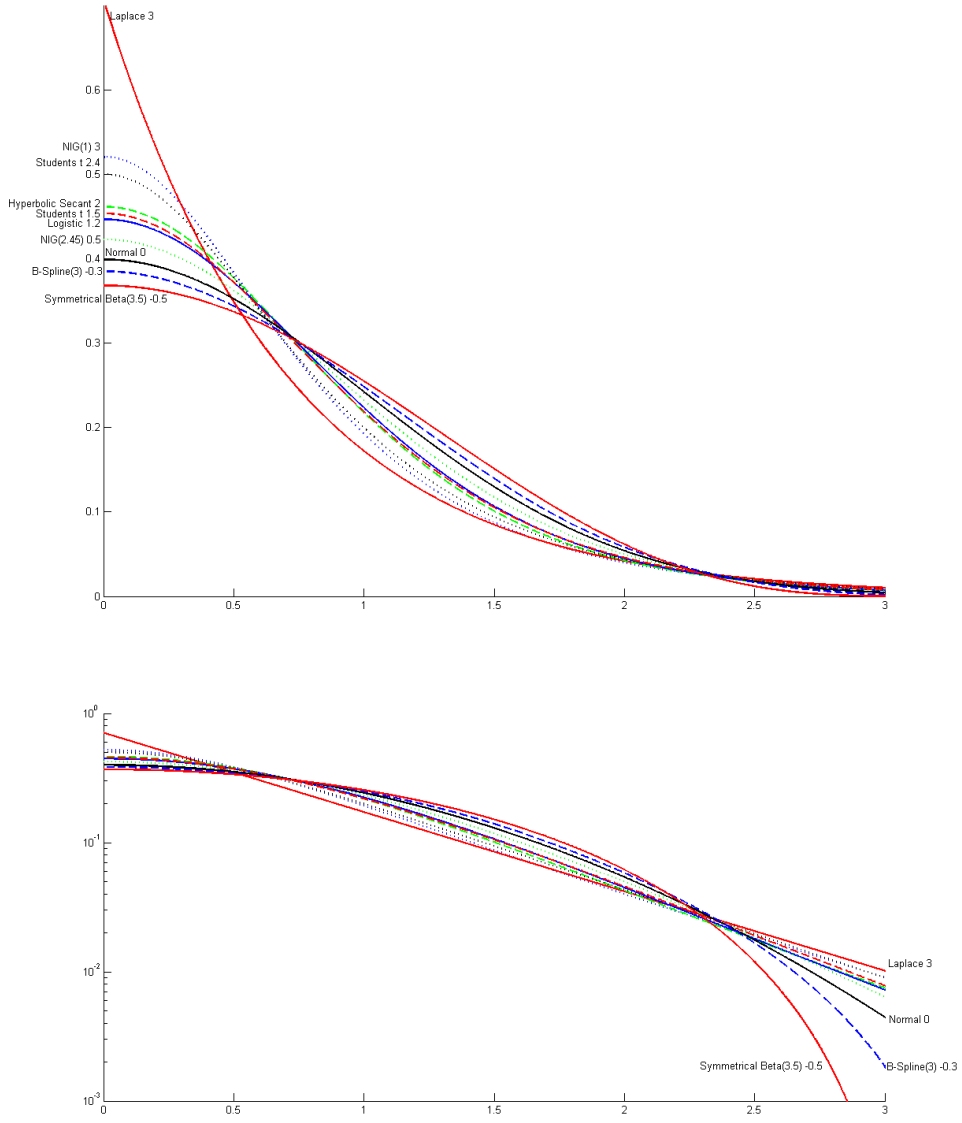


Figure 2.3: Examples of some probability density functions and their Fourier transforms. The values of the parameters in this figure can be found in Table 4.1. We also list the kurtosis(κ_4) of the distributions after there name: Laplace 3, Normal Inverse Gaussian(high) 3, Hypersecant 2, Student's t 1.5, Logistic 1.2, Normal Inverse Gaussian(low), Gaussian 0, B-Spline -0.3, Symmetric Beta -0.5. Note that we plotted two Normal Inverse Gaussians; one with high positive kurtosis and one with low poistive kurtosis.

Table 2.1: Table of some Symmetric PDFs

PDF Name and reference	**Probability density (PDF)** $\mathbf{f(x)}$	**Support** (Outcome space) $\mathbf{S(x)}$	**Parameter domain(s)**	**Characteristic function** $\Phi(t)$	**Cumulants** of order $2r$ κ_{2r}
Laplace Ex 4.3 pp. 159	$\frac{\alpha}{2}e^{-\alpha\lvert x\rvert}$	$-\infty < x < \infty$	$\alpha > 0$	$\frac{\alpha^2}{\alpha^2+t^2}$	$\frac{2(2r-1)!}{a^{2r}}$
Hypersecant [18]	$\frac{1}{2\alpha}\mathrm{sech}(\frac{\pi}{2\alpha}x)$	$-\infty < x < \infty$	$\alpha > 0$	$\mathrm{sech}(\alpha t)$	$(-1)^{r+1}(2^{2r}-1)2^{2r-1}\alpha^{2r}\frac{\mathrm{B}_{2r}}{r}$
Logistic Ex 4.21 pp 161	$\frac{\exp(-x/\alpha)}{\alpha(1+\exp(-x/\alpha))^2}$	$-\infty < x < \infty$	$\alpha > 0$	$\frac{\alpha\pi t}{\sinh(\alpha\pi t)}$	$(-1)^{r-1}\frac{(2\alpha\pi)^{2r}}{2r}\mathrm{B}_{2r}$
Normal Exa 3.4 pp 79	$\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{x^2}{2\sigma^2})$	$-\infty < x < \infty$	$\sigma > 0$	$\exp(\frac{-\sigma^2t^2}{2})$	$\begin{cases} \sigma^2 & : \ r = 2 \\ 0 & : \ r \neq 2 \end{cases}$
B-Spline Ex 3.23 pp 122	$g_1(x) \star g_2(x) \star \ldots g_N(x)$ $g(x) = 1/h$	$-\frac{Nh}{2} \leq x \leq \frac{Nh}{2}$ $-\frac{h}{2} \leq x \leq \frac{h}{2}$	$h > 0$ $N = 1, 2, 3 \ldots$	$(\frac{\sin(ht/2)}{ht/2})^N$	$N\mathrm{B}_{2r}h^{2r}/(2r)$
Family of Curves					
NIG [19]	$\frac{\delta\alpha\exp(\delta\alpha)}{\pi\sqrt{\delta^2+x^2}}K_1(\alpha\sqrt{\delta^2+x^2})$	$-\infty < x < \infty$	$\alpha > 0, \delta > 0$	$\exp[\delta(\lvert\alpha\rvert - \sqrt{\alpha^2-t^2})]$	$\kappa_2 = \frac{\delta}{\alpha}, \kappa_4 = \frac{3\delta}{\alpha^3}, \kappa_6 = \frac{45\delta}{\alpha^5}$
Student's t [20] Exa 3.13 pp 93	$\frac{1}{aB(\frac{1}{2},m)}(1+\frac{x^2}{a^2})^{-m-1/2}$	$-\infty < x < \infty$	$a > 0, m > \frac{9}{2}$	$\frac{K_m(a\lvert t\rvert)(a\lvert t\rvert)^m}{\Gamma(m)2^{m-1}}$	$\mu_2 = \frac{a^2}{2m-3}$ $\mu_4 = \frac{6a^4}{(2m-3)(2m-5)}$ $\mu_6 = \frac{240a^6}{(2m-3)^2(2m-5)(2m-7)}$
Beta Ex 6.1	$\frac{1}{aB(\frac{1}{2},m+1)}(1-\frac{x^2}{a^2})^m$	$-a < x < a$	$a > 0, m > 0$	${}_0F_1(\frac{3}{2}+m, -\frac{a^2t^2}{4})$	$\mu_2 = \frac{a^2}{2m+3}$ $\mu_4 = -\frac{6a^4}{(2m+3)(2m+5)}$ $\mu_6 = \frac{240a^6}{(2m+3)^2(2m+5)(2m+7)}$

Chapter 3

Series Approximations

3.1 The physical motivation

As set out in Chapter 1, we have from a physics point of view the following challenge: We have a non-Gaussian correlation function $R(\mathbf{q}) = C(\mathbf{q}) - 1$ whose non-Gaussian character we wish to describe systematically in terms of expansions. A long-term goal would be to use the same expansion both on $R(\mathbf{q})$ in momentum space as well as on the relative distance distribution of sources $S(\mathbf{r})$, but for the moment we are looking only for robust systematic expansions which describe the shape of $R(\mathbf{q})$.

The imaging programme of Brown, Pratt, Danielewicz and others [4] made use of orthogonal polynomial expansions, which are generally well-known in physics as they are common in quantum mechanics. In this chapter, we pursue the alternative route of expansions arising from statistics, notably the "Gram-Charlier" and the "Edgeworth" series which these authors conceived early in the 20th century. We also consider some extensions which have been proposed. This line of research in high energy physics was started in Ref. [21] and Ref. [22].

From the statistics viewpoint, we have a non-Gaussian "experimental distribution" which we shall call $g(x)$ with moments μ_j and cumulants κ_j. As $g(x)$ is equivalent to the experimental $R(\mathbf{q})$, its value $g(x)$ at any point x is assumed known. There is no requirement that $g(x)$ is exactly equal or close to some known analytical PDF; however, for testing purposes, we shall often assume that $g(x)$ has the form of some non-Gaussian PDF such as listed in Table 2.1.

In addition to the known "experimental" $g(x)$, the Gram-Charlier and Edgeworth expansions assume the existence of a *reference* PDF, which we shall call $f(x \,|\, \theta)$. This PDF almost always has a known analytical form; the first and most obvious being the Gaussian itself. While x is the variable, θ is the set of parameters which govern a given analytical

PDF. As Column 4 of Table 2.1 shows, the analytical PDFs we are using have one or at most two parameters. These parameters are free, i.e. we are free to choose or otherwise determine them to give the best results for a given type of expansion.

The moments of the reference PDF will be denoted by ξ_j and the corresponding cumulants λ_j. All the formulae of Chapter 2 apply to both $g(x)$ and $f(x)$ with the appropriate substitution. For example, we will have characteristic functions $\overline{F_X(t)}$ for $f(x)$ and $\overline{G_X(t)}$ for $g(x)$ and correspondingly expressions for the moments,

$$\mu_j = \frac{1}{i^j} D_t^j \, \overline{G_X(t)} \bigg|_{t=0} = \int x^j \, g(x) \, dx, \tag{3.1.1}$$

$$\xi_j(\theta) = \frac{1}{i^j} D_t^j \, \overline{F_X(t)} \bigg|_{t=0} = \int x^j \, f(x \,|\, \theta) \, dx, \tag{3.1.2}$$

and so on for cumulants and all other quantities and expressions. For compactness, we have introduced the shorthand $D_t = (d/dt)$. Note that the moments and cumulants of f depend on the free parameters θ.

As set out in Chapter 1, the correlation function for identical bosons is symmetric,

$$R(-\mathbf{q}) = R(\mathbf{q}), \tag{3.1.3}$$

and so we assume our one-dimensional PDFs $f(x)$ and $g(x)$ to be symmetric about zero. This will lead to many simplifications in the series expansions because all moments and cumulants of odd order are identically zero,

$$\mu_{2j+1} = 0, \qquad \kappa_{2j+1} = 0, \tag{3.1.4}$$

$$\xi_{2j+1} = 0, \qquad \lambda_{2j+1} = 0. \tag{3.1.5}$$

We shall, however, retain the odd-ordered terms in many derivations below to keep them as general as possible. We must also mention that the one-dimensional version of Eq. 3.1.3 is not necessarily true, but we assume in this thesis that the relation will also hold in one dimension.

3.2 The mathematical goal

In order to introduce series expansions in general, we start with the specific example of the Gram-Charlier series: Given a non-Gaussian probability density $g(x)$, we formally expand it in a series of derivatives of some reference PDF,

$$g(x) = c_0 f(x) - \frac{c_1}{1!} f'(x) + \frac{c_2}{2!} f''(x) - \frac{c_3}{3!} f^{(3)}(x) + \cdots = \sum_{j=0}^{\infty} (-1)^j \frac{c_j}{j!} f^{(j)}(x \,|\, \theta), \tag{3.2.1}$$

where $f^{(j)}(x \mid \theta) = \left(\frac{d}{dx}\right)^j f(x) = D_x^j f(x \mid \theta)$.

A number of other series will be proposed below, but before going into detail we must first consider whether such series converge. It is, however, not truly important whether the infinite expansion converges to $g(x)$ or not, because in practice one will always only measure the first few terms anyway. The important question is rather to what extent a *partial sum* of k-th order

$$S_k(x \mid \theta) = \sum_{j=0}^{k} (-1)^j \frac{c_j}{j!} f^{(j)}(x \mid \theta) \tag{3.2.2}$$

gives a good approximation for $g(x)$, and for which k. In mathematical terms, we are therefore looking at questions relating to *asymptotic series*. While convergence ($\lim_{k\to\infty} S_k(x \mid \theta) = g(x)$) would be nice, there are two more practical questions to be answered:

1. The first question is whether an "optimal order" k can be found for which S_k closely approaches $g(x)$; mathematically speaking, we are looking both for a k for which some norm of the difference is small, and for a suitable norm which best measures such differences,
$$\min_k \left\| S_k(x \mid \theta) - g(x) \right\| . \tag{3.2.3}$$

2. Secondly, such a minimum order should be realistic in the sense that it must be experimentally accessible. A series which converges only slowly, i.e. for which a minimum difference is attained only for (say) $k = 26$, will be of little use to the experimentalist.

3.2.1 Asymptotic Series

An asymptotic series is a series expansion of a function in a variable x which may or may not converge, but whose partial sum can be made an arbitrarily good approximation to a given function for large enough x. An asymptotic expansion has the following properties [23],

$$\lim_{x\to\infty} x^k S_k(x|\theta) - g(x) = 0 \qquad \text{for fixed } k, \tag{3.2.4}$$

$$\lim_{k\to\infty} x^k S_k(x|\theta) - g(x) = \infty \qquad \text{for fixed } x. \tag{3.2.5}$$

In other words the deviation from the expanded function can be made arbitrary small if the argument of the expanded function is sufficiently large, although at any fixed value of the argument x the series may not be convergent.

3.3 Series expansions: general relations

3.3.1 Gram-Charlier series

Returning to the Gram-Charlier ansatz Eq. 3.2.1, we must for the further derivation restrict the choice of $f(x)$ to functions with the following properties:

- Moments ξ_j of all orders of $f(x)$ must be finite.
- Derivatives of any required order must exist.
- The support of $g(x)$ and $f(x)$ is assumed compact. There is no need for the supports of the two PDFs to be the same, but it can be useful.
- If $x = a$ or $x = b$ and the support of $f(x) \in (a, b)$, there exists high order "contact", meaning that

$$\lim_{x\to a} x^j f^{(j)}(x) = \lim_{x\to b} x^j f^{(j)}(x) = 0. \tag{3.3.1}$$

These assumptions are very strong and can be lessened if the need arises to do so.

The coefficients in the series will be determined by forming moments on both sides. In view of this we need to look at the moments of the derivatives of $f(x)$, defined as [24]

$$L_j(f^{(k)}) = \frac{1}{j!}\int x^j\, f^{(k)}(x)\, dx. \tag{3.3.2}$$

Note the $j!$ factor which is different from the definition of the normal moments and will become important when we try to generalize the equation. Using the "contact" property and by repeated integration by parts, we find that

$$L_j(f^{(k)}) = \begin{cases} (-1)^k\, L_{j-k}(f) \;=\; (-1)^k\, \xi_{j-k} & \text{for } j \geq k \\ 0 & \text{for } j < k. \end{cases} \tag{3.3.3}$$

Multiplying each side of (Eq. 3.2.1) with x^ℓ and integrating, using the previous identity, we obtain an expression for the moment μ_ℓ of $g(x)$ in terms of the moments ξ_j of $f(x)$,

$$\mu_\ell = c_0\xi_\ell(\theta) + c_1\xi_{\ell-1}(\theta) + c_2\xi_{\ell-2}(\theta) + \cdots + c_\ell\xi_0(\theta) \qquad \ell = 0, 1, 2, \cdots. \tag{3.3.4}$$

This is a triangular system of linear equations in the unknown c's. For measured moments μ_j and non-analytic $g(x)$, it may be written in matrix terms

$$\begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} \xi_0 & 0 & 0 & 0 & \cdots \\ \xi_1 & \xi_0 & 0 & 0 & \cdots \\ \xi_2 & \xi_1 & \xi_0 & 0 & \cdots \\ \xi_3 & \xi_2 & \xi_1 & \xi_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ \vdots \end{bmatrix} \tag{3.3.5}$$

and solved iteratively for the coefficients $c_j, j = 0, 1, \ldots$. If $g(x)$ is known in analytic form, we can also alternatively find the c_j in terms of the Fourier transforms

$$\overline{F(t \mid \theta)} = \int e^{itx} f(x \mid \theta)\, dx \qquad \overline{G(t)} = \int e^{itx} g(x)\, dx. \tag{3.3.6}$$

Using the derivatives property of the Fourier Transform, Section 2.6.4.2 which is also based on integration by parts and infinite contact at the edges of the $f(x)$, the Fourier transform of the series (Eq. 3.2.1) is

$$G(t) = F(t \mid \theta) \left[c_0 + c_1 it + c_2 \frac{(it)^2}{2!} \cdots + c_k \frac{(it)^k}{k!} + \ldots \right]. \tag{3.3.7}$$

Looking at Eq. 3.3.7 we can see that the $i^k c_k$ equals the coefficient of t^k and that hence

$$c_k = (-i)^k D_t^k \frac{G(t)}{F(t \mid \theta)} \bigg|_{t=0}. \tag{3.3.8}$$

A third expression involving cumulant differences can also be derived from Eq. 3.3.7. Exponentiating the cumulant generating functions (Eq. 2.3.2) and writing it for t,

$$G(t) = \exp \left[\sum_{j=1}^{\infty} \frac{(it)^j}{j!} \kappa_j \right], \tag{3.3.9}$$

$$F(t \mid \theta) = \exp \left[\sum_{j=1}^{\infty} \frac{(it)^j}{j!} \lambda_j \right], \tag{3.3.10}$$

and defining cumulant differences,

$$\Delta_j = \kappa_j - \lambda_j(\theta), \tag{3.3.11}$$

we have

$$\sum_{j=0}^{\infty} \frac{(it)^j}{j!} c_j = \frac{G(t)}{F(t \,|\, \theta)} = \exp\left(\sum_{j=1}^{\infty} \frac{(it)^j}{j!} \Delta_j \right) = \sum_{k=0}^{\infty} \frac{1}{k!} \left[\sum_{j=1}^{\infty} \frac{(it)^j}{j!} \Delta_j \right]^k . \tag{3.3.12}$$

The relationship between the c_j and the Δ_j is the same as the one between moments μ_j and cumulants κ_j in Eq. 2.3.1, implying that we can identify the coefficients of the Gram-Charlier series as the "moments" of the cumulant differences, see Eq. 2.3.1. Doing the calculation explicitly for the first few coefficients gives,

$$c_0 + ic_1 t - c_2 \frac{t^2}{2!} - ic_3 \frac{t^3}{3!} + \cdots = 1 + \left(i\Delta_1 t - \Delta_2 \frac{t^2}{2!} - i\frac{\Delta_3 t^3}{3!} \cdots \right) \tag{3.3.13}$$

$$+ \frac{1}{2!} \left(i\Delta_1 t - \Delta_2 t^2 + \cdots \right)^2 + \frac{1}{3!} \left(i\Delta_1 t + \cdots \right)^3 + \cdots . \tag{3.3.14}$$

Equating orders in t gives,

$$c_0 = 1, \tag{3.3.15}$$

$$c_1 = \Delta_1, \tag{3.3.16}$$

$$c_2 = \Delta_2 + \Delta_1^2, \tag{3.3.17}$$

$$c_3 = \Delta_3 + 3\Delta_2\Delta_1 + \Delta_1^3. \tag{3.3.18}$$

The Gram-Charlier expansion does not necessarily lead to an orthogonal expansion, which implies that there is no minimum property such as the one that exists in Fourier series, see Theorem 2. It does have one important property in common with orthogonal systems in the fact that determination of later coefficients does not affect earlier ones; in other words computation of c_{j_1} relies on c_{j_2}, with $j_2 < j_1$.

There are three systems namely Hermite, Laguerre and Jacobi orthogonal polynomial systems that are both Gram-Charlier systems and orthogonal polynomials. It can also be proven that they are the only systems with this property [25]. We turn first to the Hermite system which seems to be ideal for our purposes.

The general Gram-Charlier expansion will converge if the function to be approximated $g(x)$ approaches zero faster than $\sqrt{f(x)}$ at the end points of the distribution [26].

3.3.2 Edgeworth series

In this section, we follow Cramér [27]. Assume that the variable x is the sum of n uncorrelated random variables as in Eq. 2.3.12 and that it has been standardised,

$$x = \frac{(\sum_{k=1}^{n} y_k) - \mu_1(x)}{\sqrt{\kappa_2(x)}} = \frac{(\sum_{k=1}^{n} y_k) - n\mu_1(y)}{\sigma\sqrt{n}} \tag{3.3.19}$$

where $\mu_1(y)$ and $\sigma^2 = \kappa_2(y)$ are the moment and second cumulant of the underlying y variables. In terms of the Fourier transform $G(t)$ of the distribution of y, the Fourier transform of $g(x \mid n)$ using the convolution property from Section 2.6.4.1 is

$$G(t \mid n) = \left[G\left(\frac{t}{\sigma\sqrt{n}} \right) \right]^n = \exp\left[n \sum_{j=1}^{\infty} \frac{\lambda_j}{j!} \left(\frac{it}{\sigma\sqrt{n}} \right)^j \right]. \tag{3.3.20}$$

where the λ_j are the cumulants of y. If the Gram-Charlier expansion for y is given by Eq. 3.3.12

$$\sum_{j=0}^{\infty} c_j \frac{(it)^j}{j!} = \frac{G(t)}{F(t)} = \sum_{k=0}^{\infty} \frac{1}{k!} \left[\sum_{j=1}^{\infty} \frac{(it)^j}{j!} \Delta_j \right]^k \tag{3.3.21}$$

then, replacing $G(t)$ with $G(t \mid n)$, the relation for x will be

$$\sum_{j=0}^{\infty} c_j(n) \frac{(it)^j}{j!} = \sum_{k=0}^{\infty} \frac{1}{k!} \left[n \sum_{j=1}^{\infty} \frac{\Delta_j}{j!} \left(\frac{it}{\sigma\sqrt{n}} \right)^j \right]^k \tag{3.3.22}$$

Where we also assumed that the refernce function is also a sum of variables and has been standardised. We also use the notation to show that the coefficients are dependant on n. Instead of expanding in powers of t which gives rise to the Gram-Charlier series, we now expand in the powers of n. This is just a reordering of the series; both series are equivalent in the sense that if one converges so does the other one. The difference in the series arises when we truncate the infinite sum to a partial sum, the Edgeworth series will only contain coefficients up to a certain order of n and Gram-Charlier A-series will be a mixture of orders. This makes the Edgeworth expansion preferable for nearly Gaussian Distributions, because we can estimate the error and thus have a true asymptotic expansion.

In the following table we compare the magnitude of c_j for the first few values of j when $\Delta_2 = 0$ and $\Delta_1 = 0$.

Subscript j	Order of c_j
3	$n^{-1/2}$
$4, 6$	n^{-1}
$5, 7, 9$	$n^{-3/2}$
$8, 10, 12$	n^{-2}
$11, 13, 15$	$n^{-5/2}$

3.4 Expansions with Gaussian reference PDF

3.4.1 Hermite polynomials

In the Gram-Charlier A series, the Gaussian density is used for the reference PDF $f(x\,|\,\theta)$ with the single parameter $\theta = \sigma$,

$$f(x\,|\,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{3.4.1}$$

$$\overline{F(t\,|\,\sigma)} = \exp\left(-\frac{\sigma^2 t^2}{2}\right). \tag{3.4.2}$$

Closely associated with the Gaussian are the Hermite polynomials, defined in terms of its Rodrigues formula as

$$H_j(x) = \frac{(-1)^j}{f(x|\sigma=1)} D_x^j\, f(x|\sigma=1) \tag{3.4.3}$$

from which follows

$$H_2(x) = x^2 - 1, \tag{3.4.4}$$

$$H_4(x) = x^4 - 6x^2 + 3, \tag{3.4.5}$$

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15, \tag{3.4.6}$$

$$H_8(x) = x^8 - 28x^6 + 210x^4 - 420x^2 + 105, \text{ etc.} \tag{3.4.7}$$

Note that the above are the statistical version of the Hermite polynomials, while in physics an alternative choice with $\sigma = \frac{1}{\sqrt{2}}$ and e^{-x^2} is made for f. The equivalent physics forms are

$$H_2'(x) = 4x^2 - 2, \tag{3.4.8}$$

$$H_4'(x) = 16x^4 - 48x^2 + 12, \tag{3.4.9}$$

$$H_6'(x) = 64x^6 - 480x^4 + 720x^2 - 120, \tag{3.4.10}$$

$$H_8'(x) = 256x^8 - 3584x^6 + 13440x^4 - 13440x^2 + 1680, \text{ etc.} \tag{3.4.11}$$

As mentioned, the system of Hermite polynomials is one of few whose elements are both derivatives of the reference function and mutually orthogonal. This can be shown also by repeated integration by parts and use of the Rodrigues formula. Assuming $m \le \ell$

$$\int_{-\infty}^{\infty} H_m(x) H_\ell(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \;=\; (-1)^\ell \int_{-\infty}^{\infty} H_m(x) D_x^\ell \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

$$= (-1)^\ell \left[H_m(x) D_x^{\ell-1} \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx \right]_{-\infty}^{\infty} + (-1)^{\ell-1} \int_{-\infty}^{\infty} \left[D_x H_m(x)\right] D_x^{\ell-1} \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx$$

The term in square brackets is zero and from the definition of the Hermite polynomials, the integral becomes

$$m(-1)^{\ell-1} \int_{-\infty}^{\infty} H_{m-1} D_x^{\ell-1} \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx.$$

Continuing this process we get

$$\int_{-\infty}^{\infty} H_m(x) H_\ell(x) \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx = m!\, \delta_{m\ell} \tag{3.4.12}$$

In particular, $m = 0$ gives $\int_{\mathbb{R}} H_\ell(x) \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} dx = 0$ for all $\ell > 0$. This argument can easily be extended to show that the Hermite polynomials with the Gaussian distribution are orthogonal for there chosen value of σ.

3.4.2 Expressions for the coefficients

For $\sigma \neq 1$, substitution of the Rodrigues formula into Eq. 3.2.1 yields a series expansion in Hermite polynomials

$$g(x) = \sum_{j=0}^{\infty} (-1)^j \frac{c_j}{j!} D_x^j f(x \,|\, \sigma) \tag{3.4.13}$$

$$= \sum_{j=0}^{\infty} \frac{c_j}{\sigma^j j!} H_j\left(\frac{x}{\sigma}\right) f(x \,|\, \sigma). \tag{3.4.14}$$

A direct consequence of the orthogonality of the system is that coefficients in terms of moments are exactly the Hermite polynomials integrated over our density of interest,

$$c_\ell = \frac{1}{\ell! \sigma^l} \int_{-\infty}^{\infty} H_\ell\left(\frac{x}{\sigma}\right) g(x) dx. \tag{3.4.15}$$

A different approach to compute the expansion coefficients is using Eq. 3.3.12.

$$\frac{G(t)}{F(t)} = \exp\left(-\frac{\sigma^2 t^2}{2}\right) \sum_{j=0}^{\infty} \mu_j \frac{(it)^j}{j!} \tag{3.4.16}$$

$$= \sum_{j=1}^{\infty} c_j \frac{(it)^j}{j!} \tag{3.4.17}$$

The coefficients of the expansion can now easily be read off and they are

$$c_0 = 1 \tag{3.4.18}$$
$$c_1 = \mu_1, \tag{3.4.19}$$
$$c_2 = \mu_2 - \sigma^2, \tag{3.4.20}$$
$$c_3 = \mu_3 - 3\mu_1\sigma^2, \tag{3.4.21}$$
$$c_4 = \mu_4 - 6\mu_2\sigma^2 + 3\sigma^4, \tag{3.4.22}$$
$$c_5 = \mu_5 - 10\mu_3\sigma^2 + 15\mu_1\sigma^4, \tag{3.4.23}$$
$$c_6 = \mu_6 - 15\mu_4\sigma^2 + 45\mu_2\sigma^4 - 15\sigma^6, \tag{3.4.24}$$

which agrees with Eq. 3.4.15. The same can be done for cumulants by exponentiating the sum giving, Eq. 3.3.12,

$$\frac{G(t)}{F(t)} = \exp\left(-\frac{\sigma^2 t^2}{2} + \sum_{j=1}^{\infty} \kappa_j \frac{(it)^j}{j!}\right) \tag{3.4.25}$$
$$= \sum_{j=1}^{\infty} c_j \frac{(it)^j}{j!}. \tag{3.4.26}$$

Remembering that all higher-order cumulants for Gaussians are zero $\lambda_{j\geq 3} = 0$ while $\lambda_2 = \sigma^2$, the first coefficients in are, in terms of cumulants,

$$c_1 = \kappa_1, \tag{3.4.27}$$
$$c_2 = \kappa_2 + \kappa_1^2 - \lambda_2, \tag{3.4.28}$$
$$c_3 = \kappa_3 + \kappa_1^3 + 3\kappa_1(\kappa_2 - \lambda_2), \tag{3.4.29}$$
$$c_4 = \kappa_4 + 4\kappa_1\kappa_3 + \kappa_1^4 + 6\kappa_1^2(\kappa_2 - \sigma^2) + 3(\kappa_2 - \lambda_2)^2. \tag{3.4.30}$$

This is a specific example of the more general form in terms of $\Delta_j = \kappa_j - \lambda_j$,

$$c_1 = \Delta_1, \tag{3.4.31}$$
$$c_2 = \Delta_2 + \Delta_1^2, \tag{3.4.32}$$
$$c_3 = \Delta_3 + 3\Delta_1\Delta_2 + \Delta_1^3, \tag{3.4.33}$$
$$c_4 = \Delta_4 + 4\Delta_1\Delta_3 + \Delta_1^4 + 6\Delta_1^2\Delta_2 + 3\Delta_2^2. \tag{3.4.34}$$

3.4.3 Gram-Charlier Type A series

Normally the first and obvious simplification is to assume central moments $\mu_1 = 0$ and choose the free parameter σ^2 of $f(x)$ to be equal to the measured cumulant κ_2 of $g(x)$. This gives

$\Delta_1 = \Delta_2 = 0$ and the simplified coefficients

$$c_1 = c_2 = 0, \tag{3.4.35}$$

$$c_3 = \mu_3 = \kappa_3, \tag{3.4.36}$$

$$c_4 = \mu_4 - 3\mu_2^2 = \kappa_4, \tag{3.4.37}$$

$$c_5 = \mu_5 - 10\mu_3\mu_2 = \kappa_5, \tag{3.4.38}$$

$$c_6 = \mu_6 - 15\mu_4\mu_2 + 30\mu_2^3 = \kappa_6 + 10\kappa_3^2, \tag{3.4.39}$$

$$c_8 = \mu_8 - 28\mu_6\mu_2 + 210\mu_4\mu_2^2 - 315\mu_2^4 = \kappa_8 + 56\kappa_5\kappa_3 + 35\kappa_4^2. \tag{3.4.40}$$

Inserting these coefficients into Eq. 3.2.1, we obtain the "Gram-Charlier Type A Series"

$$\begin{aligned} g(x) &= f(x)\left[1 + \frac{1}{3!}\mu_3\, H_3(x) + \frac{1}{4!}(\mu_4 - 3\mu_2^2)\, H_4(x) + \frac{1}{5!}(\mu_5 - 10\mu_3\mu_2)\, H_5(x)\right. \\ &\qquad \left. + \frac{1}{6!}(\mu_6 - 15\mu_4\mu_2 + 30\mu_2^3)\, H_6(x) + \ldots\right] \qquad (3.4.41) \\ &= f(x)\left[1 + \frac{1}{3!}\kappa_3\, H_3(x) + \frac{1}{4!}\kappa_4\, H_4(x) + \frac{1}{5!}\kappa_5\, H_5(x + \frac{1}{6!}(\kappa_6 + 10\kappa_3^2)\, H_6(x) + \ldots\right]. \qquad (3.4.42) \end{aligned}$$

3.4.4 Nonstandardised Gram-Charlier series

Matching the variance of the experimental and reference PDF's might not be the best strategy for the partial sum ($\sigma^2 \neq \kappa_2$) when approximating distributions. This question we will investigate further in Chapter 4, but for future use the coefficients Eq. 3.4.34 with this strategy and uneven moments zero become,

$$c_1 = c_3 = c_5 = 0, \tag{3.4.43}$$

$$c_2 = \frac{1}{2!}(\Delta_2), \tag{3.4.44}$$

$$c_4 = \frac{1}{4!}(\Delta_4 + 3(\Delta_2)^2), \tag{3.4.45}$$

$$c_6 = \frac{1}{6!}(\Delta_6 + 15\Delta_2(\Delta_4 + \Delta_2^2)). \tag{3.4.46}$$

3.4.5 Gauss Edgeworth series

In terms of Section 3.3.2, the re-grouped series formed by collecting terms that are of equal order in n yields,

$$g(x) = f(x)\Bigg[1 + \kappa_3\frac{H_3(x)}{2!} + \left(\frac{\kappa_4 H_4(x)}{4!} + 10\frac{\kappa_3^2 H_6(x)}{6!}\right) + \left(\frac{\kappa_5 H_5(x)}{5!} + 35\frac{\kappa_3\kappa_4 H_7(x)}{7!} + 280\frac{\kappa_3^3 H_9(x)}{9!}\right) + \left(\frac{\kappa_6 H_6(x)}{6!} + 56\frac{\kappa_3\kappa_5 H_7(x)}{8!} + 35\frac{\kappa_4^2 H_8(x)}{8!} + 2100\frac{\kappa_4\kappa_3^2 H_{10}(x)}{10!} + 15400\frac{\kappa_3^4 H_{12}(x)}{12!}\right) + \cdots\Bigg], \quad (3.4.47)$$

where the expression in each line is associated with a higher power of $n^{-1/2}$, see McCullagh [28].

3.4.6 Convergence

For a simple case such as the Gaussian, formal convergence criteria do exist. It can be shown [27] that whenever the integral

$$\int_{-\infty}^{\infty} e^{-x^2/4}\, g(x)\, dx \quad (3.4.48)$$

is convergent and if $g(x)$ is of bounded variation in $(-\infty, \infty)$, the series will converge to $g(x)$. A quick investigation of our set of test distribution shows that the Gram-Charlier Type A series does not converge for any of the positive kurtosis distributions, which is not surprising because positive kurtosis generally imply heavier tails than the (zero-kurtosis) Gaussian. For the negative kurtosis distributions on the other hand it does converge by the virtue of their finite support. Using a series approximation outside its radius of convergence usually leads to a asymptotic approximation, but in this case Gram-Charlier Type A series is not an true asymptotic expansion as we cannot estimate the error. To get a true asymptotic expansion we have to introduce the Edgeworth expansion. There also exist experimental criteria for convergence, see Ref. [29].

3.4.7 Comparing partial sums

We compare the first three terms in order of the Gram-Charlier partial sum with the first three terms in $n^{-1/2}$ of the Edgeworth partial sum while standardising $\kappa_2 = \sigma^2$,

$$\text{Gram-Charlier Type A} \qquad f(x)\left(1 + \kappa_3 \frac{H_3(x)}{2!} + \frac{\kappa_4 H_4(x)}{4!}\right), \tag{3.4.49}$$

$$\text{Edgeworth} \qquad f(x)\left(1 + \kappa_3 \frac{H_3(x)}{2!} + (\frac{\kappa_4 H_4(x)}{4!} + 10\frac{\kappa_3^2 H_6(x)}{6!})\right). \tag{3.4.50}$$

For symmetrical densities the four-term Edgeworth and Gram-Charlier partial sums are identical. Now compare 6th order for Gram-Charlier and Edgeworth up to n^{-2} in Edgeworth with uneven moments zero (Symmetrical densities) and again standardising $\kappa_2 = \sigma^2$,

$$\text{Gram-Charlier Type A} \qquad f(x)\left(1 + \frac{\kappa_4 H_4(x)}{4!} + \frac{\kappa_6 H_6(x)}{6!}\right), \tag{3.4.51}$$

$$\text{Edgeworth} \qquad f(x)\left(1 + \frac{\kappa_4 H_4(x)}{4!} + (\frac{\kappa_6 H_6(x)}{6!} + 35\frac{\kappa_4^2 H_8(x)}{8!})\right). \tag{3.4.52}$$

Note that both these partial sums only require the first three even cumulants to be known and the difference between the two partial sums is only one term.

We will also later investigate if using a unstandardised PDF for the Gram-Charlier can improve the partial sum, but first we refer to Figure 3.1 as an example of the previous partial sums. We plotted two test distributions to illustrate the difference between the partial sums, first the logistic distribution with variance 1 and kurtosis 1.2 and secondly a Symmetrical Normal Inverse Gaussian curve with variance 1 and with a low kurtosis of $\kappa_4 = 1/2$,

$$g_1(x) = \frac{\pi}{2\sqrt{3}(1 + \cosh(\pi x/\sqrt{3}))}, \tag{3.4.53}$$

$$g_2(x) = 6e^6 \frac{K_1(\sqrt{36 + 6x^2})}{\pi\sqrt{6 + x^2}}. \tag{3.4.54}$$

The free parameter σ was chosen to minimise the difference at $x = 0$ between the two densities. $\sigma = 1.13357$ for the Logistic density and $\sigma = 0.878375$ for the SNIG curve.

Complete results and discussion are in Chapter 4, but to aid the discussion we state some preliminary results between the series of which the figure is an example. We deduce the following:

- For the $\kappa_2 = \sigma^2$ Gram-Charlier Type A series as done in the textbooks, the 4th order in general improves the approximation for all test PDFs, while adding more terms worsens the approximation overall. As the tails are improved the centre is worsened, see properties of a asymptotic series, Section 3.2.1.

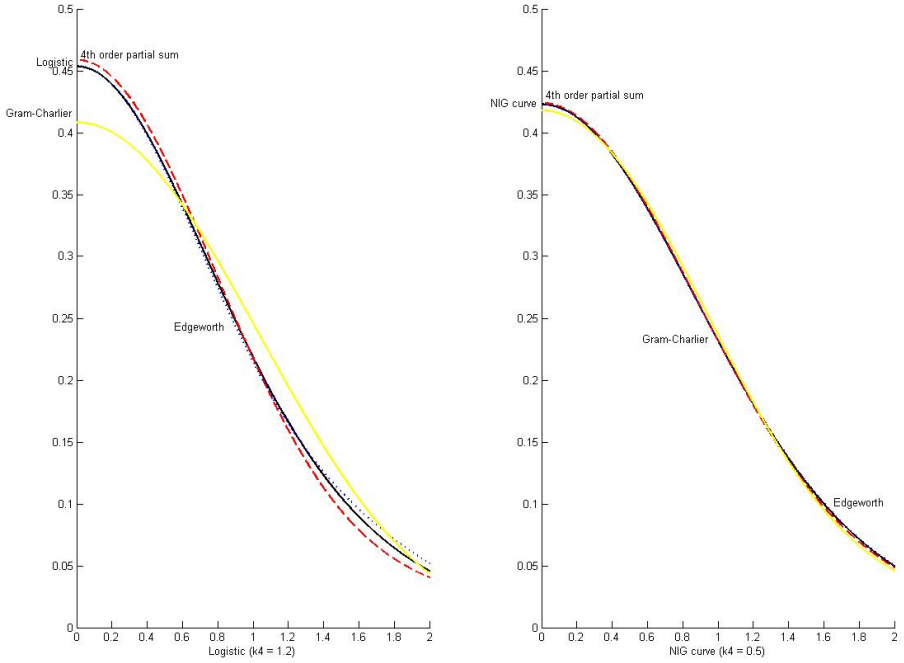


Figure 3.1: Comparing Gram-Charlier with Edgeworth partial sums for a logistic test PDF (left panel) and SNIG test PDF (right panel). The solid line is the test PDF; dashed line is the 4th order partial sum; dotted line is the Edgeworth expansion and the faded line is the 6th order Gram-Charlier expansion.

- For the minimization strategies adding more terms should worsen the behavior near $x = 0$ but the minimization tries to counteract it, which makes the behaviour erratic. The error stays approximately the same for the first few terms.
- The Edgeworth expansion, Eq. 3.3.22, outperforms the Gram-Charlier dramatically in all cases where the sum of variables assumption can be justified i.e. all examples except the Student's t distribution, see [30].
- Only if the kurtosis is very low can we safely add the sixth term to the textbook Gram-Charlier Type A series.
- We did not plot it but in the case of negative kurtosis the Gram-Charlier Type A and Edgeworth expansions are convergent, but because the coefficients are negative the tails of the distribution are also negative which makes the Gram-Charlier Type A and Edgeworth useless for approximating negative kurtosis distributions.

In summary if we use the textbook Gram-Charlier Type A series ($\kappa_2 = \sigma^2$) we can in general only approximate up to the 4th moment, while if the kurtosis is small we can add more orders. If we use a minimization policy we can add more orders, but the approximation near $x = 0$ stays poor. The Edgeworth expansion is much more successful, but gives strong non-unimodal behaviour for large kurtosis. The non-unimodal may even be an advantage in some cases, see Ref. [31] and Ref. [29]. Both the Gram-Charlier Type A and Edgeworth expansions are unusable for negative kurtosis test PDFs.

3.5 Generalized Gram-Charlier Series

The Gram-Charlier Type A series is only successful if the distribution we are trying to approximate is close to normal. For "experimental" PDFs $g(x)$ with negative kurtosis, the partial sums may be negative for some intervals in x, while for PDFs with positive kurtosis, the Type A series does not converge and the partial sum is not a unimodal function. This leads to the question whether there are better reference PDFs, beyond Gaussians, which can solve these problems.

Removing the Gaussian reference as foundation of series expansions immediately leads both to bigger problems and to bigger opportunities. Clearly, there will not be a single PDF which can be expected to replace Gaussians once and for all. Which PDF to use as reference will of course be guided by the desire to have the reference PDF model the available $g(x)$ as closely as possible. This is reminiscent of the situation in fitting, where one also chooses parametrisations to be as close to the data as possible.

In 1924, Romanovsky took the first step along this route [32] followed by Hildebrandt [33]. Both used solutions to Pearson differential equation as a reference function, because Pearson's distributions lead to orthogonal polynomial systems when taking derivatives.

$$\frac{df(x)}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} f(x) = \frac{a_0 + a_1 x}{Q(x)} f(x), \tag{3.5.1}$$

with a_0, a_1, a_2, b_0, b_2 real constants and $Q(x) = b_0 + b_1 x + b_2 x^2$. Solutions of this equation have two useful properties: Firstly, for each $j \geq 1$, the function obtained from the Rodrigues formula

$$P_j(x) = \frac{1}{f(x)} D_x^j \left[f(x) Q(x)^j \right] \tag{3.5.2}$$

is a polynomial of at most degree j, and secondly, for each $j \geq 1$, these polynomials $P_j(x)$ solve the self-adjoint second-order Sturm-Liouville differential equation [34; 35],

$$D_x \left[f(x) Q(x)\, D_x P_j(x) \right] - \Lambda_j\, P_j(x) = 0 \tag{3.5.3}$$

on the same interval for which $f(x)$ is a solution to the differential equation, with $\Lambda_j = j[a_1 + (n+1)b_2]$. It is known [25] that, in terms of polynomial solutions, only the Hermite, Jacobi and Laguerre polynomials are solutions to the Rodrigues formula (Eq. 3.5.2). As we have seen, a Gaussian $f(x)$ yields the Hermite polynomials, while the beta distribution leads to the Jacobi polynomials system and the gamma distribution to the generalised Laguerre polynomials.

Series which use the beta and gamma densities as reference PDFs are also considered Gram-Charlier series [15]. The major difference when we move away from the Gaussian reference PDF is that the infinite "contact" assumption, (the property of zero surface terms in the integration by parts) may be lost, meaning that the whole derivation would become invalid. To fix this problem, we need to introduce the "*correction function*" $Q(x)$ enters the general Rodrigues formula (Eq. 3.5.2) also into our series ansatz. We therefore generalise Eq. 3.2.1 to the *Generalised Gram-Charlier Series* [33]

$$g(x) = c_0 f(x) - \frac{c_1}{1!} D_x \left[f(x)Q(x)\right] + \frac{c_2}{2!}\frac{d^2}{dx^2}\left[f(x)Q(x)^2\right] + \cdots + \frac{(-1)^k c_k}{k!}\frac{d^k}{dx^k}\left[f(x)Q(x)^k\right] + \cdots \tag{3.5.4}$$

and proceed, in the sections below, to explore the consequences for three resulting series expansions: one for the negative-kurtosis beta distribution, and two with positive kurtosis, namely the Student's t distribution and the Symmetrical Normal Inverse Gaussian.

3.6 Negative-kurtosis reference PDF: the Symmetrical Beta distribution

We now consider a Generalised Gram-Charlier series using the symmetrical beta distribution (Pearson Type II) as reference PDF $f(x)$; see Table 2.1 and Ref. [32] for some details. It has finite support on the real line, $x \in (-a, +a)$, and its use will solve the negative tail problem which the (Gaussian) Gram-Charlier Type A expansion encountered. Not surprisingly, it improves the approximation for negative-kurtosis experimental PDFs $g(x)$ by virtue of having negative kurtosis itself and thus reducing the coefficients in the expansion. Using the symmetrical beta-function and renaming the parameter $m = v - \frac{1}{2}$ to simplify the algebra, we have the reference PDF and its characteristic function,

$$f(x|a,v) = \frac{(1-\frac{x^2}{a^2})^{v-1/2}}{aB(1/2, v+1/2)}, \tag{3.6.1}$$

$$\overline{F(t|a,v)} = {}_0F_1(v+1, -\frac{1}{4}a^2t^2), \tag{3.6.2}$$

where ${}_0F_1$ is the Confluent Hypergeometric Function [36]. As noted before, the symmetric beta PDF does not have infinite contact at the extremities $(-a, a)$, and, being a polynomial, the beta PDF will be identically zero after $2m$ derivatives and will thus have only a finite number of moments. However, the Rodrigues formala with a correction function

$$Q(x) = \left(1 - \frac{x^2}{a^2}\right) \tag{3.6.3}$$

increases the v parameter by an integer j before taking the j-th derivative so that polynomials of all orders are defined. Note that the $[f(x \,|\, v)Q(x)^j]$ in the Rodrigues formula can also be written as $f(x \,|\, v+j)$ up to a normalization constant, so that the polynomials can be written as

$$C_j(x|a, v) = \frac{(-1)^j}{f(x|a, v)} D_x^j \, f(x|a, v + j) \tag{3.6.4}$$

The $(-1)^j$ factor as with the Hermite polynomials ensure we have positive polynomials and we also normalize the function after multiplying it with the correction function, which simplifies the Fourier transform. We state the first few polynomials with $a = 1$; they are the set of Gegenbauer or Ultraspherical polynomials [36] times a constant $\frac{j! 2^j \Gamma(v+1)\Gamma(2v+j)}{\Gamma(2v)\Gamma(v+1+j)}$,

$$\begin{aligned}
C_2(x|v) &= \frac{4(v+1)(v+2)}{v(1+2v)} \left[2(v+1)x^2 - v\right], \qquad (3.6.5)\\
C_4(x|v) &= \frac{64v\Gamma(2v+4)}{\Gamma(2v)\Gamma(v+5)} \left[4\Gamma(v+4)x^4 - 12\Gamma(v+3)x^2 + 3\Gamma(v+2)\right],\\
C_6(x|v) &= \frac{6! 2^6 v\Gamma(2v+6)}{90\Gamma(2v)\Gamma(2v+7)} \left[\Gamma(v+6)x^6 - 60\Gamma(v+5)x^4 + 90\Gamma(v+4)x^2 - 15\Gamma(v+3)\right].
\end{aligned}$$

The Ultraspherical polynomials are orthogonal,

$$\int_{-a}^{a} C_m(x|v)\, C_j(x|v)\, f(x|v)\, dx = \frac{\Gamma(v)\Gamma(v+1)\Gamma(2v)}{(2a)^{2j} j! \Gamma(v+j+1)\Gamma(v+j)\Gamma(2v+j)} \, \delta_{mj}. \tag{3.6.6}$$

In particular, $m = 0$ gives $\int_{-a}^{a} C_j(x) f(x|v) dx = 0$ for all $j > 0$. The partial sum is,

$$\begin{aligned}
g(x) &= \sum_{j=0}^{\infty} \frac{c_j}{j!} D_x^j f(x|a, v + j), \qquad (3.6.7)\\
&= \sum_{j=0}^{\infty} \frac{c_j}{a^j j!} (-1)^j C_j \left(\frac{x}{a}, v\right) f(x|a, v). \qquad (3.6.8)
\end{aligned}$$

The Fourier transform of the partial sum as seen below, does not factorize anymore; rather, we get

$$G(t) = {}_0F_1(v+1, -\frac{a^2t^2}{4}) - \frac{c_2t^2}{2!}{}_0F_1(v+3, -\frac{a^2t^2}{4}) + \frac{c_4t^4}{4!}{}_0F_1(v+5, -\frac{a^2t^2}{4}) - \frac{c_6t^6}{6!}{}_0F_1(v+7, -\frac{a^2t^2}{4}) + \cdots . \quad (3.6.9)$$

The nonfactorization means that the coefficients cannot be expressed simply in terms of cumulants. Comparing terms and then inverting them, we get the expansion coefficients

$$\begin{aligned} c_2 &= \mu_2 - \frac{a^2}{2(v+1)}, \\ c_4 &= \mu_4 - \frac{3a^2c_2}{v+3} - \frac{3a^4}{4(v+1)(v+2)}, \\ c_6 &= \mu_6 - \frac{15a^2c_4}{2(v+5)} - \frac{45a^4c_2}{4(v+3)(v+4)} - \frac{15a^6}{8(v+1)(v+2)(v+3)}. \end{aligned} \quad (3.6.10)$$

These coefficients are still dependent on the moments of the reference distribution but the correction function adds a factor to each term.

Convergence

It can be shown [16] that, whenever the integral

$$\int_{-a}^{a} (1 - \frac{x^2}{a^2})^{\frac{1-2v}{4}} g(x)dx \quad (3.6.11)$$

is convergent and if $g(x)$ is of bounded variation in $(-a, a)$, the series will converge to $g(x)$ in every continuity point of $g(x)$. This theorem implies that the Fourier-Jacobi series converges for any distribution with limited support, the same as the Gram-Charlier Type A series.

3.7 Expansions with positive-kurtosis reference PDF.

3.7.1 Student's t distribution

We now turn to a reference PDF which improves the series expansion for positive kurtosis distributions. As in the previous two expansions, we would like an orthogonal polynomial system. The Pearson weight which corresponds to the symmetrical positive-kurtosis distributions is the "Student's t" or Pearson Type VII distribution, which has a PDF and

characteristic function of

$$f(x|m,a) = \frac{1}{aB(\frac{1}{2},m)}(1+\frac{x^2}{a^2})^{-(m+1/2)}, \tag{3.7.1}$$

$$\overline{F(t|m,a)} = \frac{a^m|t|^m K_m(a|t|)}{\Gamma(m)2^{m-1}}, \tag{3.7.2}$$

where $K_m(x)$ is the modified Bessel function of the second kind. The problem with this distribution is that it has only $2m-1$ finite moments which implies that we only have a finite basis (fewer than m orthogonal polynomials). If we add the correction function $Q(x) = (1+\frac{x^2}{a^2})$ we get the *Hildebrandt polynomials* [33]

$$S_j(x|a,m) = \frac{(-1)^j}{f(x|a,m)} D_x^j f(x|a,m+j). \tag{3.7.3}$$

The way we defined the polynomials, we also normalize the function after multiplying it with the correction function, which simplifies the Fourier transform. The first few even polynomials with $a=1$ are

$$S_2(x|m) = \frac{4(m-1)(m-2)}{2m-1}\left[2(m-1)x^2-1\right], \tag{3.7.4}$$

$$S_4(x|m) = \frac{\Gamma(m)\Gamma(m-1)}{2^4\Gamma(m-4)(2m-1)(2m-3)}\left[\frac{4}{\Gamma(m-3)}x^4 - \frac{12}{\Gamma(m-2)}x^2 + \frac{3}{\Gamma(m-1)}\right],$$

$$S_6(x|m) = \frac{2^8\Gamma(2m-5)\Gamma(m)^2}{\Gamma(m-6)\Gamma(2m)}\left[\frac{8}{\Gamma(m-5)}x^6 - \frac{60}{\Gamma(m-4)}x^4 + \frac{90}{\Gamma(m-3)}x^2 - \frac{15}{\Gamma(m-2)}\right].$$

As with the set of hermite polynomials, the Hildebrandt polynomials can also be proven orthogonal. The Truncated Fourier Transform of the partial sum S_6 is, assuming $t>0$ and $m>6$,

$$G(t) = \frac{K_m(at)(at)^m}{\Gamma(m)2^{m-1}} + \frac{c_2}{2!}\frac{K_{m-2}(at)(at)^m}{a^2\Gamma(m-2)2^{m-3}} + \frac{c_4}{4!}\frac{K_{m-4}(at)(at)^m}{a^4\Gamma(m-4)2^{m-5}} + \frac{c_6}{6!}\frac{K_{m-6}(at)(at)^m}{a^6\Gamma(m-6)2^{m-7}}. \tag{3.7.5}$$

Comparing terms and inverting, we get the expansion coefficients

$$\begin{aligned} c_2 &= \mu_2 - \frac{a^2}{2(m-1)}, \\ c_4 &= \mu_4 - \frac{3a^2c_2}{m-3} - \frac{3a^4}{4(m-1)(m-2)}, \\ c_6 &= \mu_6 - \frac{15a^2c_4}{2(m-5)} - \frac{45a^4c_2}{4(m-3)(m-4)} - \frac{15a^6}{8(m-1)(m-2)(m-3)}. \end{aligned} \tag{3.7.6}$$

These coefficients are the same as the Pearson Type II coefficients only with $(v+j)$ being replaced with $(m-j)$.

3.7.2 Symmetric Normal Inverse Gauss

The Student's t case gives us a finite basis, so we conclude that to get a complete basis we will have to choose between the Gram-Charlier series and orthogonality. We can keep orthogonality and construct an polynomial system from the Symmetric Normal Inverse Gauss distribution, SNIG polynomials in short. This is not an Gram-Charlier series, as at no point do we take derivatives of the SNIG PDF, or other option is that we take the SNIG PDF as a reference function for the Gram-Charlier series. This expansion will not be orthogonal but we will be able to compute coefficients by matching moments.

$$\begin{aligned} f(x|a,\delta) &= \frac{a\delta \exp(a\delta) K_1\left(a\sqrt{x^2+\delta^2}\right)}{\pi\sqrt{x^2+\delta^2}} \\ \overline{F(t|a,\delta)} &= \exp\left[a\delta - \delta\sqrt{a^2+t^2}\right]. \end{aligned} \tag{3.7.7}$$

Proving that these two expressions are Fourier transform pairs is not that straightforward, so we make a brief detour using Ref. [19] to do so:

Fourier transform of SNIG

We start with a function which resembles the above $F(t|a,\delta)$ but which we expand into a *two-dimensional* dual space (t, A) corresponding to two-dimensional space variables (x, y),

$$F(t, A) = \exp\left[a\delta - \delta\sqrt{A^2+t^2}\right]. \tag{3.7.8}$$

An inverse Fourier transform in both of these variables

$$f(x,y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(t,A)\, e^{-i(tx+yA)}\, dt\, dA, \tag{3.7.9}$$

reduces to a Hankel transform because of the radial dependence of F

$$\begin{aligned} f(q) &= \frac{1}{2\pi}\int_0^{\infty} F(r) J_0(qr) r dr \\ &= \frac{e^{a\delta}\delta}{2\pi(q^2+\delta^2)^{3/2}}. \end{aligned} \tag{3.7.10}$$

Note that $q = \sqrt{x^2+y^2}$ and $r = \sqrt{A^2+t^2}$. Changing back from radial coordinates to cartesian coordinates gives,

$$f(x,y) = \frac{e^{a\delta}\delta}{2\pi(x^2+y^2+\delta^2)^{3/2}}. \tag{3.7.11}$$

Fourier transforming y back to A and then choosing $A = a$ results in,

$$f(x) = \frac{a\delta \exp(a\delta) K_1\left(a\sqrt{x^2+\delta^2}\right)}{\pi\sqrt{x^2+\delta^2}}. \tag{3.7.12}$$

We return to the construction of SNIG-based expansions. Differentiating our density function does not lead to polynomials and we cannot factorize out our density. The functions resulting from differentiatin are, however, reasonably simple. Defining an auxiliary function $\phi(x)$ and the derivatives of the SNIG distribution $h_k(x|a,\delta)$,

$$\begin{aligned} \phi_k(x|\delta,) &= \frac{a^{k+1}\delta \exp(a\delta) K_{k+1}\left(a\sqrt{x^2+\delta^2}\right)}{\pi(x^2+\delta^2)^{k/2}}, \\ h_k(x|a,\delta) &= (-1)^k D_x^k \left(\frac{a\delta \exp(a\delta) K_1\left(a\sqrt{x^2+\delta^2}\right)}{\pi\sqrt{x^2+\delta^2}} \right), \end{aligned} \tag{3.7.13}$$

simplifies the results. The auxiliary function has the nice property of $D_x\phi_k(x) = -x\phi_{k+1}(x)$. In terms of the ϕ_k, the derivatives are

$$\begin{aligned} h_0(x|a,\delta) &= \phi_0, \\ h_2(x|a,\delta) &= x^2\phi_2 - \phi_1, \\ h_4(x|a,\delta) &= x^4\phi_4 - 6x^2\phi_3 + 3\phi_2, \\ h_6(x|a,\delta) &= x^6\phi_6 - 15x^4\phi_5 + 45x^4\phi_4 - 15x^4\phi_3; \end{aligned} \tag{3.7.14}$$

note the striking resemblance to Hermite polynomials. The expansion, Eq. 3.2.1, now becomes

$$\begin{aligned} g(x) &= c_0 h_0 + c_2\frac{h_2}{2!} + c_4\frac{h_4}{4!} + c_6\frac{h_6}{6!} + \cdots \\ \overline{G(t)} &= \exp\left[a\delta - a\sqrt{\delta^2+t^2}\right](1 - c_2\frac{t^2}{2!} + c_4\frac{t^4}{4!} - c_6\frac{t^6}{6!} + \cdots) \end{aligned} \tag{3.7.15}$$

The probability density function cannot be factorised like the previous expansion, but the Fourier transform remains the same and the coefficient can thus easily be computed. The first few even-order coefficients are, in terms of moments of g,

$$\begin{aligned} c_0 &= 1, \\ c_2 &= \mu_2 - \frac{\delta}{a}, \\ c_4 &= \mu_4 - 6\frac{\mu_2\delta}{a} + \frac{3\delta(a\delta-1)}{a^3}, \\ c_6 &= \mu_6 - 15\frac{\mu_4\delta}{a} + 45\frac{\mu_2(1+a\delta)}{a^3} - 15\frac{\delta(3-3a\delta-a^2\delta^2)}{a^5}. \end{aligned} \tag{3.7.16}$$

or in terms of cumulants (assuming odd-ordered moments and cumulants zero),

$$\begin{aligned} c_2 &= \kappa - \frac{\delta}{a}, \\ c_4 &= \kappa_4 - 3\frac{\delta}{a^3} + 3c_2^2, \\ c_6 &= \kappa_6 - 45\frac{\delta}{a^5} - 30c_2^3 + 15c_2c_4. \end{aligned} \tag{3.7.17}$$

3.7.3 Orthogonal polynomials from the SNIG distribution

Using again the SNIG as reference PDF, we can construct orthogonal polynomials using the process of orthogonalization. Using the determinants as set out in Chapter 2,

$$P_k(x) = \frac{1}{D_k}\begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & & \vdots \\ \mu_{k-1} & \mu_k & \cdots & \mu_{2k+1} \\ 1 & x & \cdots & x^k \end{vmatrix} \tag{3.7.18}$$

The first few polynomials are

$$D_2 = \frac{a}{\delta}, \tag{3.7.19}$$

$$P_2(x) = \frac{1}{D_2}\begin{vmatrix} 1 & 0 & \frac{\delta}{a} \\ 0 & \frac{\delta}{a} & 0 \\ 1 & x & x^2 \end{vmatrix} = x^2 - \frac{\delta}{a}, \tag{3.7.20}$$

$$D_4 = \frac{3\delta^3(3 + 2a\delta(12 + a\delta(9 + 2a\delta)))}{a^9}, \tag{3.7.21}$$

$$P_4(x) = \frac{1}{D_4}\begin{vmatrix} 1 & 0 & \frac{\delta}{a} & 0 & \frac{3\delta(1+a\delta)}{a^3} \\ 0 & \frac{\delta}{a} & 0 & \frac{3\delta(1+a\delta)}{a^3} & 0 \\ \frac{\delta}{a} & 0 & \frac{3\delta(1+a\delta)}{a^3} & 0 & \frac{15\delta(3+3a\delta+a^2\delta^2)}{a^5} \\ 0 & \frac{3\delta(1+a\delta)}{a^3} & 0 & \frac{15\delta(3+3a\delta+a^2\delta^2)}{a^5} & 0 \\ 1 & x & x^2 & x^3 & x^4 \end{vmatrix} \tag{3.7.22}$$

$$= x^4 - x^2\frac{3(15 + 2a\delta(7 + 2a\delta))}{a^2(3 + 2a\delta)} + \frac{3\delta(12 + a\delta(9 + 2a\delta))}{a^3(3 + 2a\delta)}. \tag{3.7.23}$$

In the next section we compare these three expansions with each other.

Chapter 4

Results for analytical test distributions

4.1 Overview of questions and strategies

In this and the next chapter, we make quantitative comparisons between the non-Gaussian distribution $g(x)$ and the various expansions which try to approximate it. If real data was available, a normalised histogram would be constructed which would represent $g(x)$. The purpose of this thesis is, however, more limited: before applying an expansion or strategy to real data, we wish to test it in a "controlled environment" where we know exactly what $g(x)$ should be, allowing quantitative comparisons regarding the quality of a given expansion or strategy.

In carrying out this programme, we shall employ two methodologies. In the first methodology, applied in this chapter, we shall pretend that the non-Gaussian data is perfectly reproduced by some analytic non-Gaussian probability density function (PDF), which we shall call the "test PDF". This avoids issues related to sampling and allows us to concentrate on the various ways in which the expansions approach (or do not approach) the desired form of the analytic test PDF.

In the second methodology, we shall in Chapter 5 go one step further: while still assuming that $g(x)$ itself is some analytical PDF, we will generate artificial data samples which are distributed according to it and do our analysis with these samples. This will allow us to explore the impact and significance of sampling fluctuations, for example how stable a given quantity or method is with regard to normal statistical fluctuations in data.

Note that the reference PDF $f(x)$ will always be analytic in form. In Section 4.2.1, we consider the simplest case where $f(x) = \exp(-x^2/2\sigma^2)/\sqrt{2\pi\sigma^2}$ is a Gaussian with one free

parameter σ, while in Section 4.3, three different non-Gaussian strategies will be compared.

Questions which we want to address with numerical simulation:

1. Which strategy works best to fix the free parameter or parameters of the reference PDF?
2. Are there advantages in using non-Gaussian reference PDFs as compared to the usual Gaussian reference?
3. How does a given expansion change as the order k of the partial sum S_k is increased?

These questions will be addressed in various ways in the remainder of this chapter and in the next one.

4.2 Strategies to fix the free parameters

The following strategies for fixing the free parameters of $f(x|\theta)$ (one or two, as the case may be) will be investigated:

a) **Zeroing second-order difference:**
This textbook strategy chooses the second cumulant of the reference PDF equal to that of the test PDF, $\lambda_2 = \kappa_2$ where for the Gaussian reference $\lambda_2 = \sigma^2$. This sets the second coefficient identically to zero, $\Delta_2 = 0$, and thereby reduces the algebra of the expansions significantly.

Arbitrarily choosing $\kappa_2 = 1$ for our test PDFs, this strategy immediately also implies a standardized reference PDF ($\lambda_2 = 1$). Parameter values to bring our suite of test PDFs into their standardized form are set out in Table (-). For one-parameter distributions, setting $\kappa_2 = 1$ fixes the parameter and thereby all higher-order cumulants. In the case of the two-parameter distributions (the SNIG and Student distributions in our case), standardization reduces the two free parameters to one. For illustrative purposes, we have therefore chosen two sets of parameters for these distributions which both yield $\kappa_2 = 1$.

b) **Minimization of cumulant differences:**
Another approach to fixing the reference distribution parameter(s) is to determine it from not only from the second-order Δ_2, but to take into account Δ_4 and Δ_6 as well. With this in mind, we minimize the sum of coefficients (which are functions of the Δs) to determine an appropriate θ,

$$\min_{\theta}\left(c_2^2 + c_4^2 + c_6^2\right) \tag{4.2.1}$$

c) **Minimization of differences at x=0:**
A third approach is based on the dual function spaces of the Fourier pairs t and x. We augment the first few cumulants $\kappa_2, \kappa_4, \kappa_6$ with the value and derivatives of $g(x)$ at zero, $g(x{=}0), g''(0), g^{(4)}(0)$. For a single free parameter this is expressed mathematically as

$$\min_{\theta} \left[g(0) - S_k(0|\theta) \right]^2 . \tag{4.2.2}$$

The first and obvious criticism of this approach is that we have thereby singled out the specific point $x = 0$ in the distribution. We point out, however, that $g(0)$ was already singled out when we started talking about moments and cumulants about zero, because $g(0)$ contributes zero to all moments and we will not be able to reconstruct it if we only have a finite number of cumulants. The present strategy will improve our asymptotic approximations as $g(0)$ is also by definition the worst-behaved point of the approximation.

The second criticism is that we may not be able to measure $g(0)$ because of experimental limitations or bad statistics. This may be so. Possibly, however, we may solve this problem by using the empirical characteristic function which is a point fourier transform of the data to explore the behaviour of the characteristic function and the data set at zero.

d) **Edgeworth:**
The strategy here is also to set $\lambda_2 = \kappa_2$ as in Case a) above. The difference arises because, as set out in Section 3.3.2, the inclusion of terms in an expansion is determined not by order, but rather by the Edgeworth ordering in terms of a sum variable n. For the purposes of the test below, this means that the strategy is identical with that of a) above but with an extra term $\frac{35}{8!}\kappa_4^2 H_8(x)$ included in $S_6(x)$.

4.2.1 Results for Gaussian reference PDF

Table 4.1 shows the numerical values of the cumulants of the PDFs we are interested in. Table 4.2 contains numerical results illustrating minimization strategies with the same Gaussian reference PDF but different test PDFs. The first column indicates the test PDF being approximated, while in column 2 the value of σ which optimised a given strategy is listed. The values for c_2, c_4 and c_6 in the next three columns are fixed by the value of σ. The last three columns set out three different norms used to illustrate the differences between the test PDF and the partial sum: an integral over the *square of the difference* between the partial sum and the reference PDF (L_2 norm), an integral over the *absolute value of the difference*

Test PDF	Standardising Relation	κ_2	κ_4	κ_6	κ_8
SNIG	$\alpha = \delta = 1$	1	3	45	1575
Laplace	$\alpha = \frac{1}{2}$	1	3	30	630
Student's t	$a = \sqrt{2m-3}, m = 3.75$	1	2.4	192	∞
Hypersecant	$\alpha = 1$	1	2	16	272
Student's t	$a = \sqrt{2m-3}, m = 4.5$	1	1.5	30	∞
Logistic	$s = \sqrt{3}/\pi$	1	1.2	$\frac{48}{7} \approx 6.86$	86.5
SNIG	$\alpha = \delta = \sqrt{6}$	1	0.5	1.25	$\frac{175}{24} \approx 7.29$
Normal	$\sigma = 1$	1	0	0	0
3-Spline	$x \to x/\sqrt{3}$	1	-0.3	$\frac{3}{7} \approx 0.429$	-1.35
Beta	$a = \sqrt{2m+3}, m = 3.5$	1	-0.5	$\frac{10}{7} \approx 1.429$	$-\frac{155}{6} \approx -9.69$

Table 4.1: Values of cumulants for test PDFs. Values of parameters in Column 1 are chosen to make $\kappa_2 = 1$, which then fixes all higher-order cumulants. The PDFs are orderer from high kurtosis to low kurtosis.

(L_1 norm), and the *maximum difference* between $S_6(x)$ and $g(x)$ over some range of x. The purpose of using different norms is to ensure that comparision between strategies is not overly dependent on the specifics of the norm used. In all cases, the sixth-order partial sum S_6 was used.

Results for different test PDFs are ordered from high kurtosis (c_4) to low kurtosis (top of table to bottom).

For each test PDF, there are four lines in the table, with each line corresponding to one of the four strategies set out above: Line 1 is the textbook $\Delta_2 = 0$ case, Line 2 the minimization of $\sum_j c_{2j}^2$, Line 3 the minimization of $[S_6(0) - g(0)]$, and Line 4 the $\Delta_2 = 0$ case but with the additional Edgeworth term added in.

From Table 4.2, we can infer the following:

- The Edgeworth expansion gives better results than the other strategies except for the Student's t distribution and high-kurtosis SNIG where it is better only with respect to the L_2 norm but worst in the L_1 case. This behaviour can be explained by the assumptions that went into the derivation of the Edgeworth expansion: we assumed that the distribution we are approximating is a sum of random variables, in the case of the Student's t distribution with its divergent higher moments this assumption is not very good.

- For test PDFs with high kurtosis, the Edgeworth expansion gives better L_2 approxi-

mation but oscillates more which leads to greater negative densities: see Figure 4.2 top row (SNIG with large kurtosis).

- In the case of nonstandardised Gram-Charlier i.e. not choosing $\kappa_2 = \sigma^2$, the accuracy of the partial sum does improve as expected, especially in the case of Student's t distribution. The difference between the two different minimization strategies are minimal.

In Figure 4.1 and Figure 4.2, the partial sums are plotted against the test PDF. Figure 4.1 shows the general performance for distributions with low kurtosis and Figure 4.2, gives the performance for high kurtosis distributions. The Edgeworth expansion gives a much better approximation at $x = 0$ even when we tried to minimize the Gram-Charlier expansion at that point.

In the second and third rows, Figure 4.2 also shows two test PDFs (Laplace and Student) which are highly non-Gaussian and therefore not likely to be well approximated with expansions based on a Gaussian reference PDF. In the case of the Laplace PDF, the cusp and nonzero first derivative at $x = 0$ cannot possibly be approximated by a Gaussian which has a zero derivative there. In the case of the Student's t distribution, the problem lies with its divergent higher moments which cause the Edgeworth expansion to fail completely.

Test PDF	σ	$\mathbf{c_2}$	$\mathbf{c_4}$	$\mathbf{c_6}$	$\int(\mathbf{S_6}-\mathbf{g})^2$	$\int\lvert\mathbf{S_6}-\mathbf{g}\rvert$	$\mathbf{max}(\mathbf{S_6}-\mathbf{g})$
SNIG	1	0	3	45	0.143	0.771	0.240
$(\alpha=\delta=1)$	1.346	-0.813	4.984	0.336	0.010	0.198	0.056
	1.348	-0.818	5.006	0	0.010	0.198	0.056
	1	0	3	45	0.009	0.229	0.070
Laplace	1	0	3	30	0.089	0.568	0.408
	1.262	-0.592	4.051	0.247	0.023	0.250	0.269
	1.263	-0.596	4.066	0	0.023	0.250	0.269
	1	0	3	30	0.012	0.244	0.081
Student's t	1	0	2.4	192	3.031	3.559	1.096
$(m=3.75)$	1.731	-1.996	14.36	0.808	0.015	0.279	0.108
$a=\sqrt{2m-3}$	1.732	$-2.$	14.4	0.	0.015	0.279	0.108
	1	0	2.4	192	2.316	3.146	1.329
Hypersecant	1	0	2	16	0.020	0.282	0.134
	1.214	-0.474	2.673	0.201	0.004	0.127	0.068
	1.216	-0.479	2.687	0.	0.004	0.127	0.068
	1	0	2	16	0.001	0.084	0.024
Student's t	1	0	1.5	30	0.063	0.515	0.222
$(m=4.5)$	1.369	-0.873	3.788	0.365	0.003	0.115	0.048
$a=\sqrt{2m-3}$	1.371	-0.880	3.821	0.	0.003	0.115	0.048
	1	0	1.5	30	0.027	0.353	0.140
Logistic	1	0	1.2	6.857	0.003	0.114	0.052
	1.158	-0.340	1.547	0.151	0.0009	0.060	0.028
	1.160	-0.346	1.560	0.	0.0009	0.059	0.028
	1	0	1.2	6.857	0.0001	0.023	0.007
SNIG	1	0	0.5	1.25	0.0001	0.021	0.009
$(\alpha=\delta=\sqrt{6})$	1.072	-0.150	0.567	0.077	6×10^{-5}	0.015	0.007
	1	-0.159	0.576	0.	6×10^{-5}	0.015	0.007
	1	0	0.5	1.25	1×10^{-6}	0.002	0.0007

Table 4.2: Gaussian reference PDF: Numerical results for various minimization strategies. Results for different test PDFs are ordered from high kurtosis (c_4) to low kurtosis (top of table to bottom). The four lines for each test PDF refer respectively to the four strategies set out in the text, namely a) $c_2=0$, b) cumulant minimization, c) minimization at $x=0$, and d) Edgeworth term added.

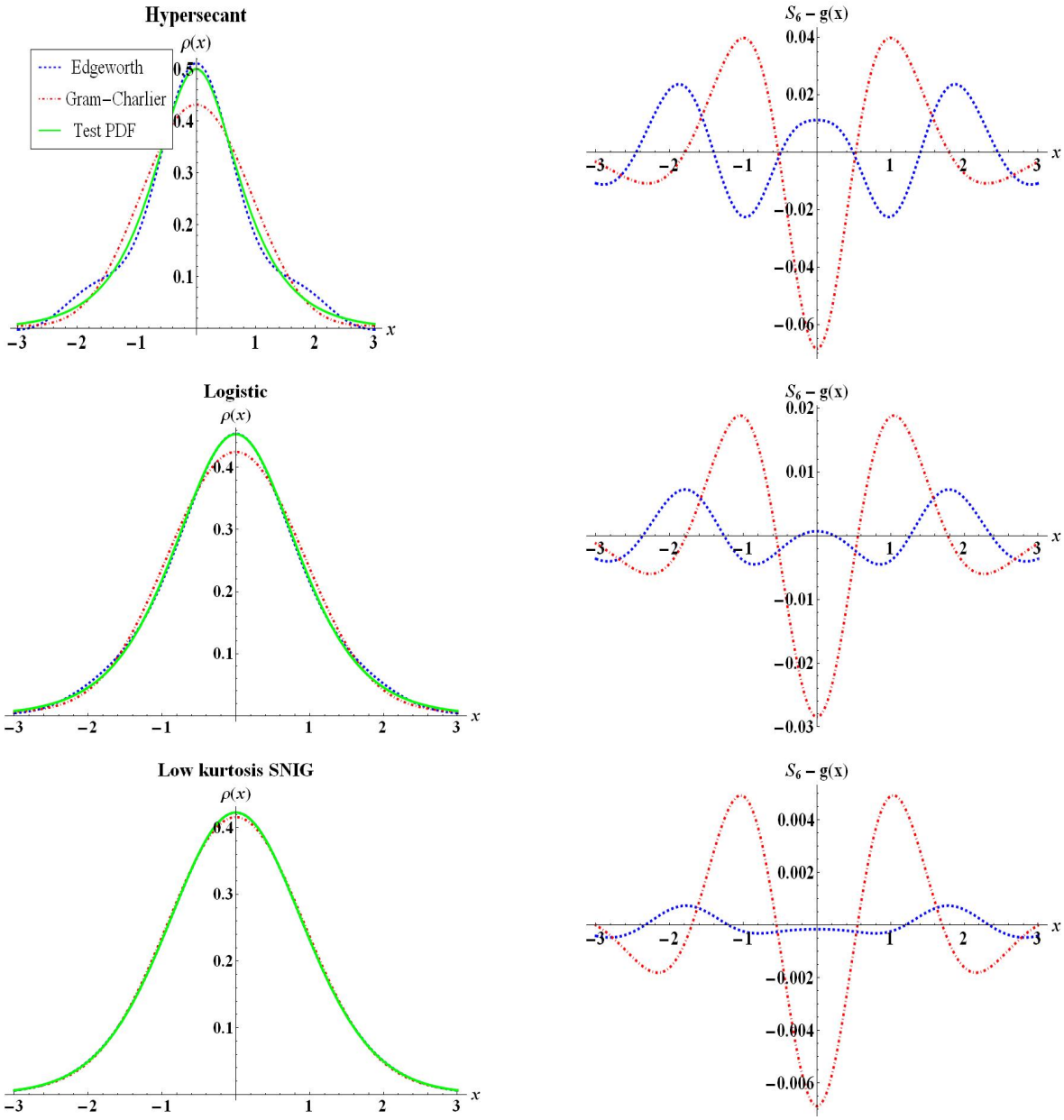


Figure 4.1: Results for Gaussian reference PDF I: minimization strategies on PDFs with low positive kurtosis. Top row: Hypersecant test PDF; centre row: Logistic test PDF; bottom row: Symmetrical Normal Inverse Gaussian test PDF with a low kurtosis of 0.5, see Table 4.1. The left panels show the functions in the x-space, while the right panels show the difference between the test PDFs and the partial sums (note the very different scales). The solid lines represent the test PDF, dashed lines the Edgeworth partial sum and the dot-dashed line the minimization schemes. We show only one of the minimizations as both look the same to the naked eye. All parameters and coefficients for these figures are listed in Table 4.2.

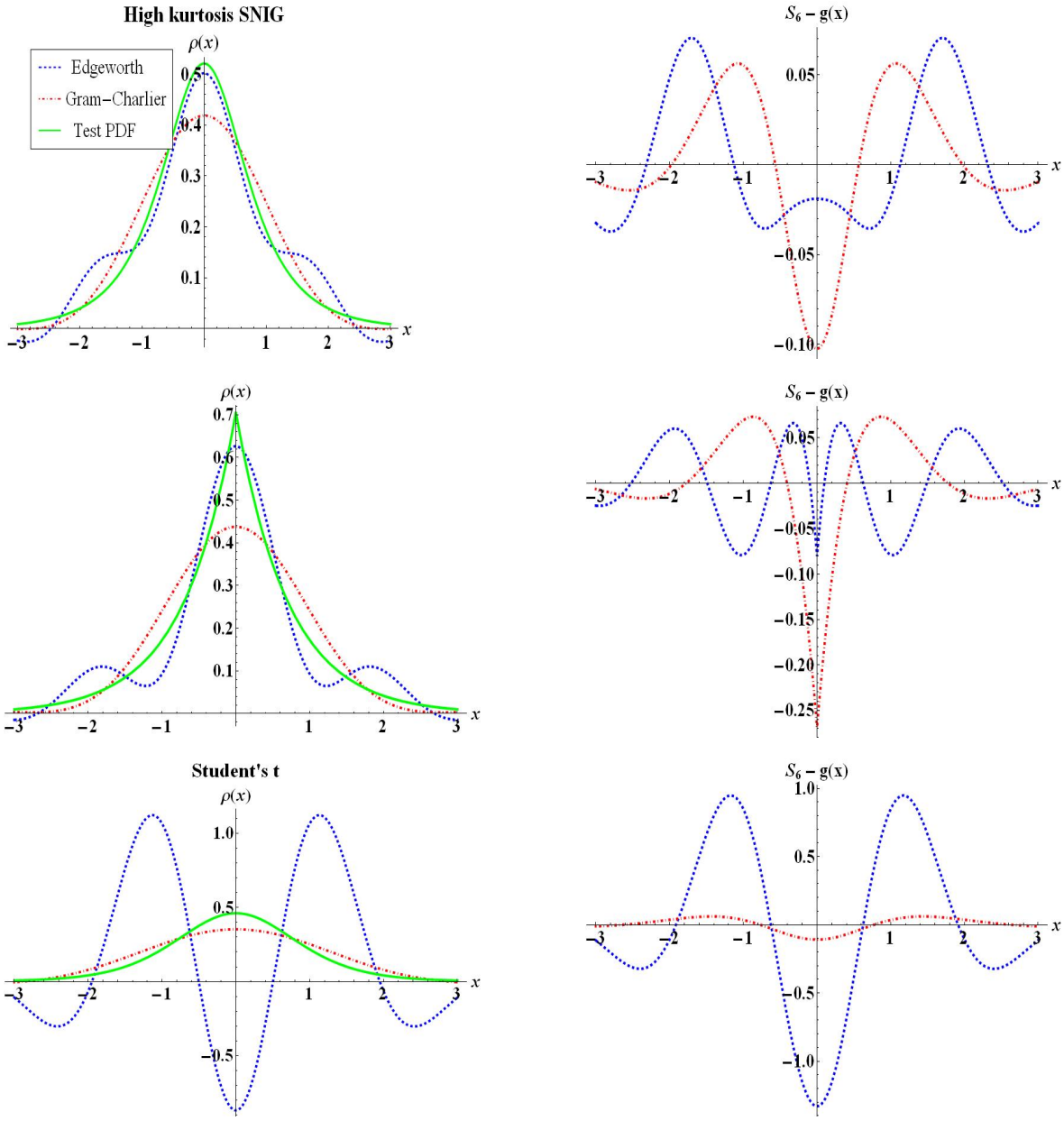


Figure 4.2: Results for Gaussian reference PDF II: minimization strategies on PDFs with high kurtosis. Top row: Symmetrical Normal Inverse Gaussian test PDF with a high kurtosis of 3; centre row: Laplace test PDF; bottom row: Student test PDF. The left panel shows the functions in the x-space, while the right panel shows the difference between the test PDF and the partial sum. The solid line is the test PDF, dashed line is the Edgeworth partial sum and the dot-dashed line is the $c_2 = 0$ minimization. Neither Laplace nor Student can be approximated well because of their highly non-Gaussian character. A Laguerre expansion should be more appropriate for the Laplace distribution. All parameters and coefficients for these figures are listed in Table 4.2.

4.3 Non-Gaussian reference functions

As set out in Chapter 3, it is also possible to use non-Gaussian PDFs as reference. The complications arising from this choice are possibly worth it if the resulting expansion is more accurate or converges more quickly than the equivalent Gaussian reference.

The immediate question is of course which non-Gaussian PDF to choose as a reference. This cannot be answered in general, because one will want to choose the reference PDF to be as close to the given test PDF or data as possible.

4.3.1 Negative kurtosis: Beta distribution

To investigate the suitability of an ultraspherical expansion for probability densities, we plot a B-spline test PDF against a corresponding sixth-order partial sum. The second B-spline is the convolution of three standardized uniform distributions, which is itself then standardized. The probability density is

$$g(x) = \begin{cases} \frac{1}{16}(x+3)^2 & \text{if } -3 \le x \le -1, \\ \frac{1}{8}(3-x^2) & \text{if } -1 \le x \le 1, \\ \frac{1}{16}(x-3)^2 & \text{if } 1 \le x \le 3. \end{cases} \tag{4.3.1}$$

- The characteristic function is $(\sin t)^3/t^3$ with the first three even moments $\mu_2 = 1, \mu_4 = 2.6$ and $\mu_6 = \frac{205}{21} = 9.7619$.

- The partial sum is

$$S_6 = \left[1 + \frac{c_2}{a^2 2!} C_2\left(\frac{x}{a}\right) + \frac{c_4}{a^4 4!} C_4\left(\frac{x}{a}\right) + \frac{c_6}{a^6 6!} C_6\left(\frac{x}{a}\right)\right] \frac{(1-\frac{x^2}{a^2})^{v-1/2}}{aB(1/2, v+1/2)}. \tag{4.3.2}$$

- Choosing a value for the parameters a and m of the beta distribution is the next step. We use a to ensure $c_2 = 0$, and v is used to minimise $|g(0) - S_6(0)|^2$. We have

$$S_6(0) = \frac{1}{5040a^7\sqrt{\pi}\Gamma(7/2+v)}\Big[(11025a^6 - 22050a^4(4+v) + 22932a^2(4+v)(5+v) - 8200(4+v)(5+v)(6+v))\Gamma(4+v)\Big], \tag{4.3.3}$$

$$g(0) = \frac{3}{8}. \tag{4.3.4}$$

- Solving these equations gives $v = 3.34761$, $a = 2.94877$, $c_2 = 0$, $c_4 = 0.160998$ and $c_6 = 0.151507$.

- Comparing the Gegenbauer (Fourier-Jacobi) expansion with the Gram-Charlier type A partial sum, we see that both are a pretty good approximations: see top picture in Figure 4.3. The middle picture is the difference between the partial sums and the B-spline, while the bottom picture is a loglinear plot of the top picture.

- We will not look more explicitly at the Beta distribution as it is widely documented in the computational polynomial field to approximate limited supported functions. It is very succesful, but as the Fourier transform of a distribution with bounded support usually oscillates it cannot represent the relative distance distribution which is necessarily positive definite. Our interest in negative kurtosis distributions is therefore limited.

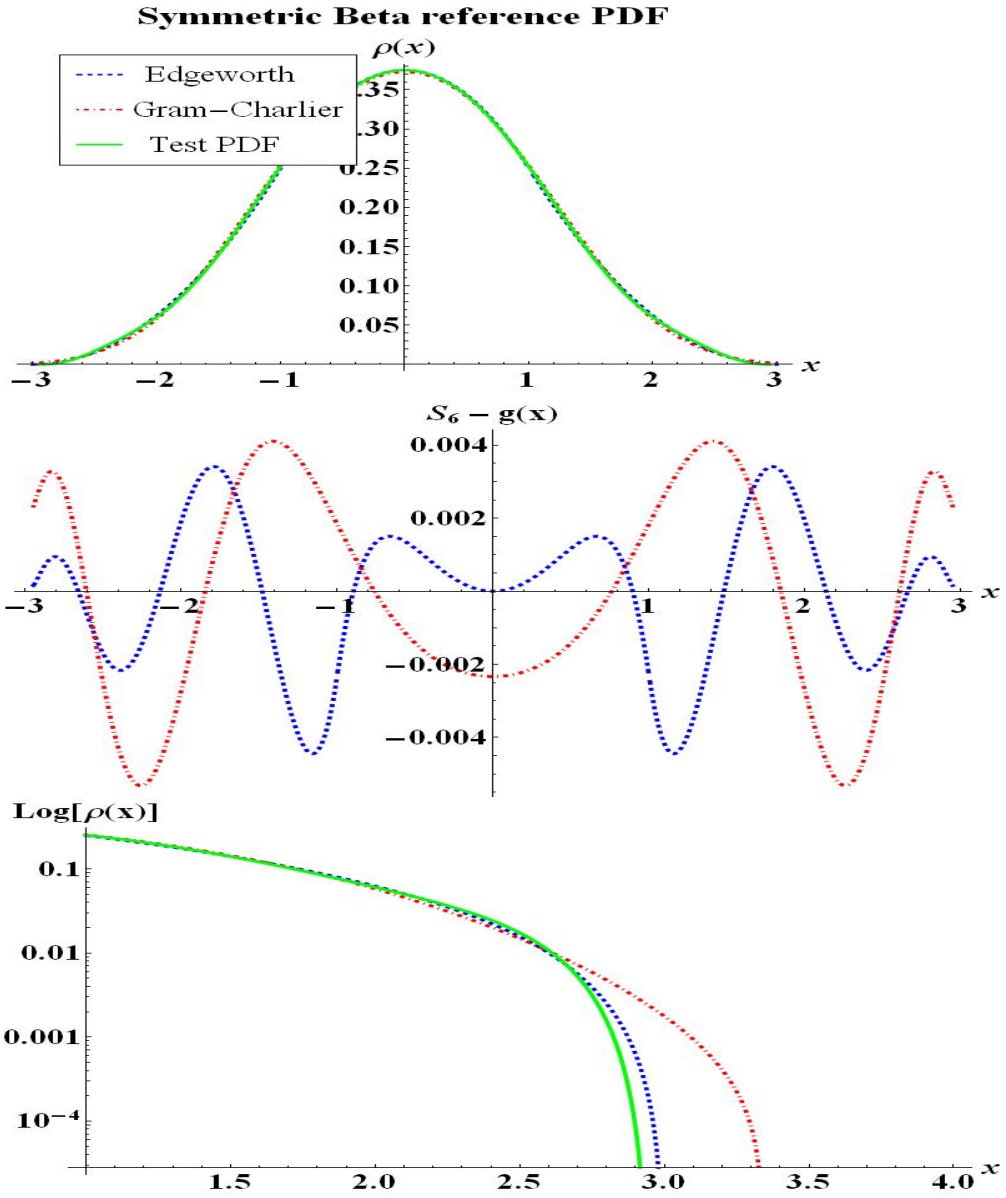


Figure 4.3: Example of symmetrical beta reference PDF: Top panel shows the B-spline test PDF (solid line) and the corresponding partial sums, Jacobi-Fourier (Dotted line) and the Gram-Charlier Type A (Dot-dashed line). Middle panel is the difference between the partial sums and test PDF. Bottom panel is a log plot of the top panel.

4.3.2 Positive kurtosis reference functions

As mentioned, reference PDFs $f(x)$ and the resulting set of expansion functions have different properties. Each of the choices places the focus on a different aspect and has their own advantages and disadvantages. In the case of positive-kurtosis PDFs, we observe the following:

- The Hildebrandt polynomials are derivatives of the reference PDF (in this case the Student's t), and they are orthogonal. This makes their coefficients easily computable, although not simple functions of cumulant differences due to the nontrivial correction function, $Q(x) \neq 1$. The set of Hildebrandt polynomials is, however, finite, which might seem to disqualify them for expansions. For low-kurtosis test PDFs, this is not really a problem because we can choose the Student's parameter m to be large, which correspondingly gives us moments up to order $2m - 1$ and polynomials up to order m, so that the divergent behaviour does not affect the partial sum. It does, of course, restrict the choice of m.

- The situation is different for a SNIG reference PDF. As in the case of the Hildebrandt polynomials, the coefficients and higher derivatives are easy to compute. The coefficients are also expressible as cumulant differences, which aids statistical interpretation of the expansion. The price we pay for these properties is that the expansion functions are not orthogonal. The expansion is still, however, an convergent series for all of the test PDFs, except Laplace.

  Lack of orthogonality implies that we cannot rely on the Theorem 2 proving that the set of coefficients obtained minimises the difference between partial sum and $g(x)$. We must therefore separately investigate the impact of the minimum property.

- The SNIG polynomials are not derivatives of the SNIG reference PDF but rather orthogonal with respect to the SNIG distribution by construction. The orthogonality property gives us the minimum property of Fourier Series, but the coefficients and polynomials are computationally more complex.

4.3.3 Results for non-Gaussian reference PDFs

In Figure 4.4 and Figure 4.5, the partial sums are plotted against the test PDF. Figure 4.4 shows the general performance for distributions with low positive kurtosis and Figure 4.5, gives the performance for high kurtosis distributions. In the left-hand panels of the plots, the test PDFs are given by a solid line, the Hildebrandt polynomials dashed, the SNIG Gram-Charlier dot-dashed and the SNIG polynomials by a dotted line.

Clearly, non-Gaussian reference PDFs with low kurtosis do an excellent job. As the differences between the partial sums and $g(x)$ are nearly indistinguishable in the left-hand panels, we plot in the right-hand panels the differences between the partial sums and the test PDF. Again note the differences in scale between plots.

Table 4.3 contains the results of the non-Gaussian reference PDFs. The first column indicates the test PDF being approximated, while in column 2 and 3 the value of free parameters which has been optimised in the $(S_6(0) - g(0))^2$ sense are listed. The values for c_2, c_4 and c_6 in the next three columns are fixed by the value of free parameters and the moments. The last three columns set out three different norms as in the previous section.

For each test PDF, there are four lines, with each line corresponding to one of the four partial sums set out above: Line 1 is the textbook $c_2 = 0$ Gram-Charlier from Table 4.2, Line 2 the Hildebrandt polynomials, Line 3 the SNIG Gram-Charlier series, and Line 4 the SNIG polynomials. The first test PDF is the SNIG with large kurtosis, but two of the reference functions also use the SNIG. We expect that the distribution will coincide and reduce all the coefficient to zero, but because we are using numerical minimization to solve for the free parameters, they end up in a local minimum.

From Table 4.3, we can infer the following:

- For test PDFs with large kurtosis, the SNIG polynomials outperforms the other options. Even for the Laplace distribution which is the worst case the approximation are not too bad, see Figure 4.3.[1]

- For test PDFs with lower kurtosis, the SNIG Gram-Charlier and the Hildebrandt polynomials perform slightly better than the SNIG polynomials.

- The Hildebrandt polynomials gives the best approximation to Student's t distribution as expected.

[1] We again note that Laplace has a cusp at zero and will hence never look good in expansions. The Laguerre expansion would be a better choice for the Laplace distribution, because it also has a cusp at zero.

a) Gram-Charlier: $\sigma^2 = \kappa_2$								
b) Hildebrandt polynomials:								
c) SNIG Gram-Charlier:								
d) SNIG polynomials:								
Test PDF	a	m or δ	$\mathbf{c_2}$	$\mathbf{c_4}$	$\mathbf{c_6}$	$\int(\mathbf{S_6}-\mathbf{g})^2$	$\int\lvert\mathbf{S_6}-\mathbf{g}\rvert$	$\mathbf{max}(\mathbf{S_6}-\mathbf{g})$
SNIG	1	n/a	0	3	45	0.143	0.771	0.239
($\alpha=\delta=1$)	2.387	6.110	0.443	2.409	-17.69	$6.\times10^{-5}$	0.0154	0.0057
	1.028	2.616	-1.544	2.932	-14.924	8.7×10^{-5}	0.0189	0.00694
	0.893	0.983	0.031	0.004	-0.0008	7.1×10^{-7}	0.0020	0.0067
Laplace	1	n/a	0	3	30	0.0895	0.568	0.408
	2.046	6.455	0.616	3.220	4.600	0.0110	0.183	0.102
	0.901	2.347	-1.605	1.097	-50.41	0.0150	0.226	0.111
	0.613	0.550	0.0232	-0.01	0.0020	0.0056	0.137	0.086
Student's t	1	n/a	0	2.4	192	3.0311	3.559	1.096
($m=3.75$)	2.805	6.004	0.2139	1.403	120.7	5.9×10^{-6}	0.0055	0.0017
	0.549	3.778	-5.886	37.71	-439.2	2.1×10^{-4}	0.0090	0.0366
	1.019	1.327	-0.085	0.025	0.0048	6.8×10^{-5}	0.0059	0.0201
Hypersecant	1	n/a	0	2	16	0.020	0.282	0.134
	2.687	7.143	0.4125	1.607	0.0043	5.0×10^{-5}	0.0142	0.0053
	1.319	1.387	-0.052	0.194	0.209	9.8×10^{-5}	0.0173	0.0083
	0.892	1.062	-0.052	0.005	-0.0005	2.8×10^{-6}	0.0037	0.0010
Student's t	1	n/a	0	1.5	30	0.0631	0.515	0.222
($m=4.5$)	2.485	6.000	0.3826	0.708	0.0046	3.3×10^{-13}	1.25×10^{-6}	3.7×10^{-7}
	0.859	3.186	-2.709	8.435	-22.96	1.2×10^{-6}	0.0027	6.2×10^{-4}
	1.133	1.545	-0.105	0.031	-0.0088	2.2×10^{-5}	0.0113	0.0031
Logistic	1	n/a	0	1.2	6.857	0.0032	0.114	0.051658
	3.460	9.666	0.3092	0.916	1.769	3.117×10^{-6}	0.0038	0.0013
	1.609	2.861	-0.778	0.955	-2.105	7.2×10^{-7}	0.0018	6.5×10^{-4}
	1.048	1.417	-0.096	0.020	-0.0054	8.3×10^{-6}	0.0070	0.0020
SNIG	1	n/a	0	0.5	1.25	0.00010	0.0206	0.0092
($\alpha=\delta=\sqrt{6}$)	4.032	11.39	0.2176	0.203	-0.404	5.6×10^{-9}	1.6×10^{-4}	6.0×10^{-5}
	2.494	3.182	-0.276	0.113	-0.072	8.4×10^{-11}	1.9×10^{-5}	7.5×10^{-6}
	1.905	2.289	-0.104	0.031	-0.014	7.7×10^{-8}	6.7×10^{-4}	1.6×10^{-4}

Table 4.3: Numerical results for various non-Gaussian reference PDFs. Results for different test PDFs are ordered from high kurtosis (c_4) to low kurtosis (top of table to bottom). The four lines for each test PDF refer respectively to: 1) Gaussian reference Gram-Charlier as before, 2) Hildebrandt polynomials from the Student's t reference PDF, 3) Gram-Charlier series using SNIG reference PDF, 4) constructed orthogonal polynomials from SNIG.

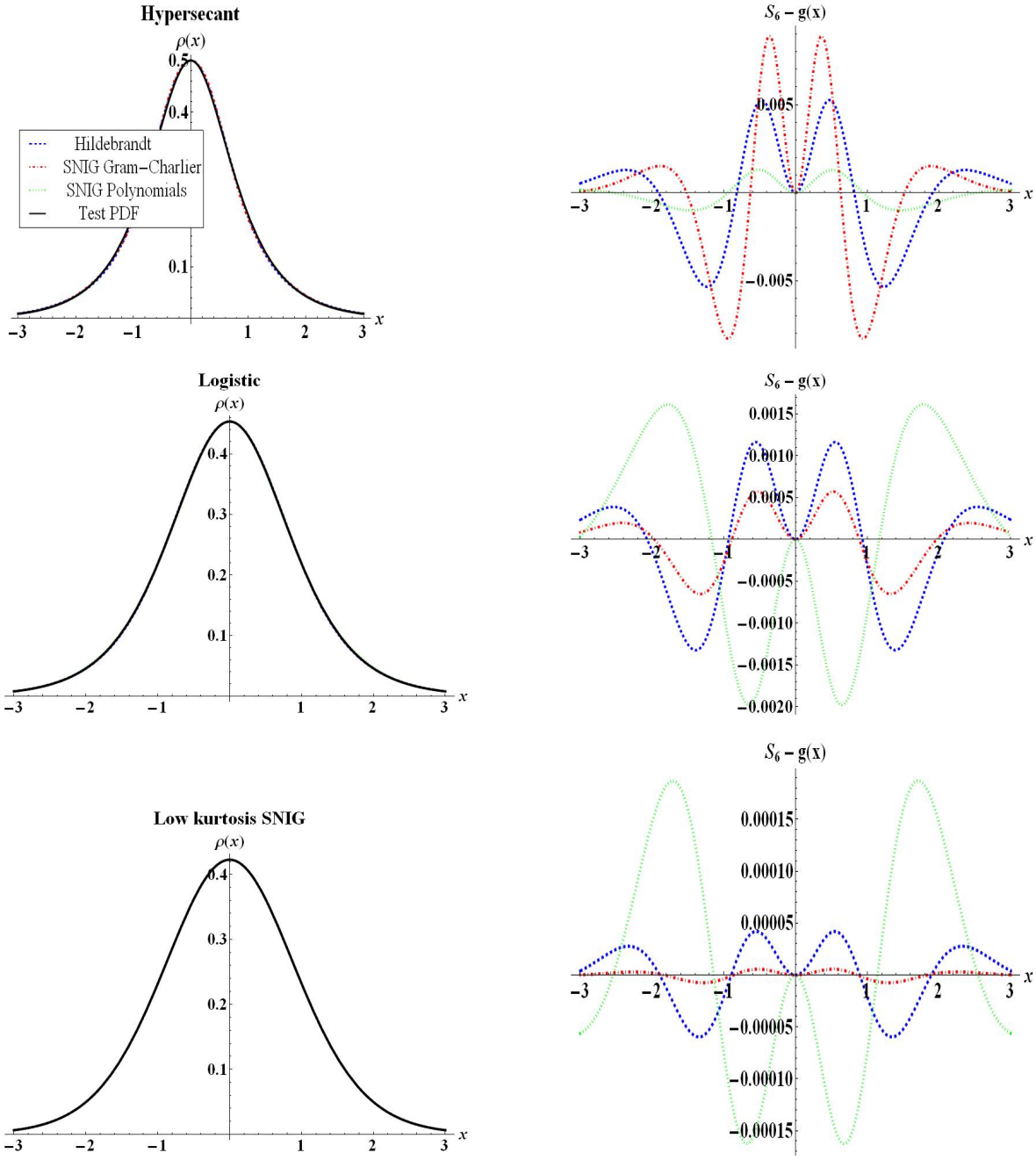


Figure 4.4: Non-Gaussian reference PDFs with small kurtosis. The dashed line is the Hildebrandt polynomials, the dot-dashed line is the SNIG Gram-Charlier and the SNIG polynomials is the dotted line. In the left panels, the partial sums are plotted and in the right panels the differences between partial sum and test PDF. The top row is for the Hypersecant test PDF, middle for the Logistic PDF and bottom for the SNIG with small kurtosis. The coefficients and parameters of these plots are contained in Table 4.3.

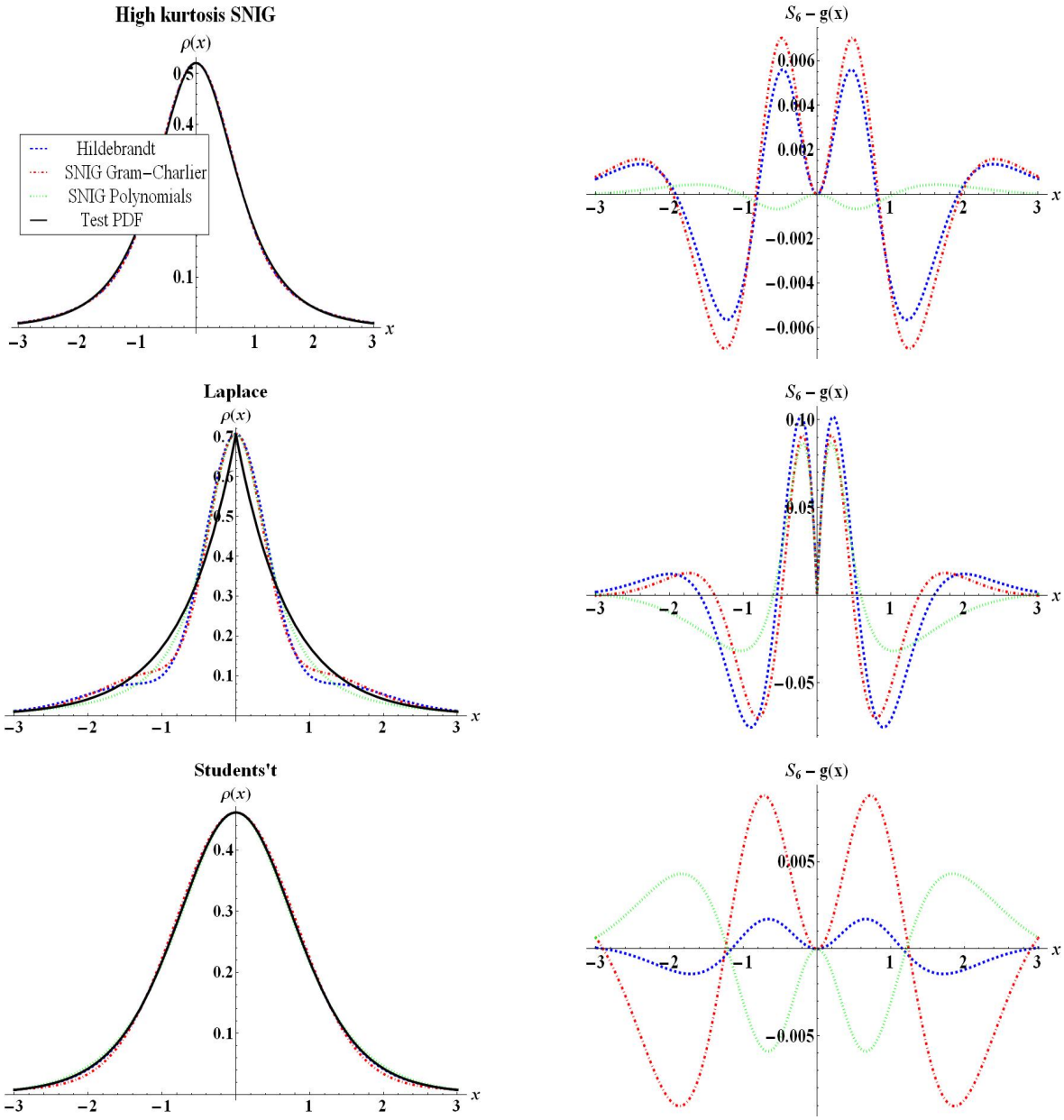


Figure 4.5: Non-Gaussian reference PDFs with high kurtosis. The dashed lines represent the Hildebrandt polynomials, the dot-dashed line the SNIG Gram-Charlier and the dotted line the SNIG polynomials. In the left panels, the partial sums are plotted and in the right panels the differences between partial sum and test PDF. The top row is for the NIG PDF with high kurtosis, middle row for the Laplace PDF and bottom for the Student's t PDF with high kurtosis. The coefficients and parameters of these plots are contained in Table 4.3.

4.4 Non-Gaussian reference: dependence on the order of the partial sum

Table 4.4 contains the results of the differences $S_k(x) - f(x)$ for various orders of the partial sum. The first column indicates the method used, while in column 2, 3 and 4 the order of the partial sum is indicated. For each method, there are three lines, with each line corresponding to one of the three norms as set out in the previous tables. The test PDF throughout this exercise was chosen to be the hypersecant and again we have applied the minimization strategy $((S_6(0) - g(0))^2)$ throughout.

Table 4.4 clearly illustrates the asymptotic nature of the approximations:

- The Gram-Charlier Type A series and SNIG polynomials worsens with the 6th order term and improves with the 8th order term.
- The Hildebrandt polynomials improves with the 6th order term and worsens with the 8th order term.
- The SNIG Gram-Charlier partial sums improves in the L_1 and L_2 norms, but worsens in the 8th term in the L_∞ norm.
- Figure 4.6 and Figure 4.7 illustrates the information in the Table graphically.

What we can deduce from Table 4.4 is that we require more research in this case, probably in an Edgeworth kind of stratagem to reorder these series to have a better error convergence as some of these series are convergent. Note that differences are not large, so this approach is promising.

Partial Sum	S_4	S_6	S_8	**Partial Sum**	S_4	S_6	S_8
Gram-Charlier	0.001	0.004	0.001	Hildebrandt	0.0005	4.9×10^{-5}	0.0002
	0.077	0.127	0.077		0.0459	0.0142	0.0310
	0.024	0.068	0.024		0.0148	0.0053	0.0114
SNIG	0.0002	0.0001	1.7×10^{-6}	SNIG	1.6×10^{-6}	2.8×10^{-6}	1.7×10^{-6}
Gram-Charlier	0.0459	0.0173	0.0029	polynomials	0.0031	0.0037	0.0029
	0.0093	0.0083	0.0096		0.00087	0.00099	0.00098

Table 4.4: Comparison in different norms, between order of the same partial sum. The first row is the L_2 norm, the second row L_1 and the third row the L_∞ norm

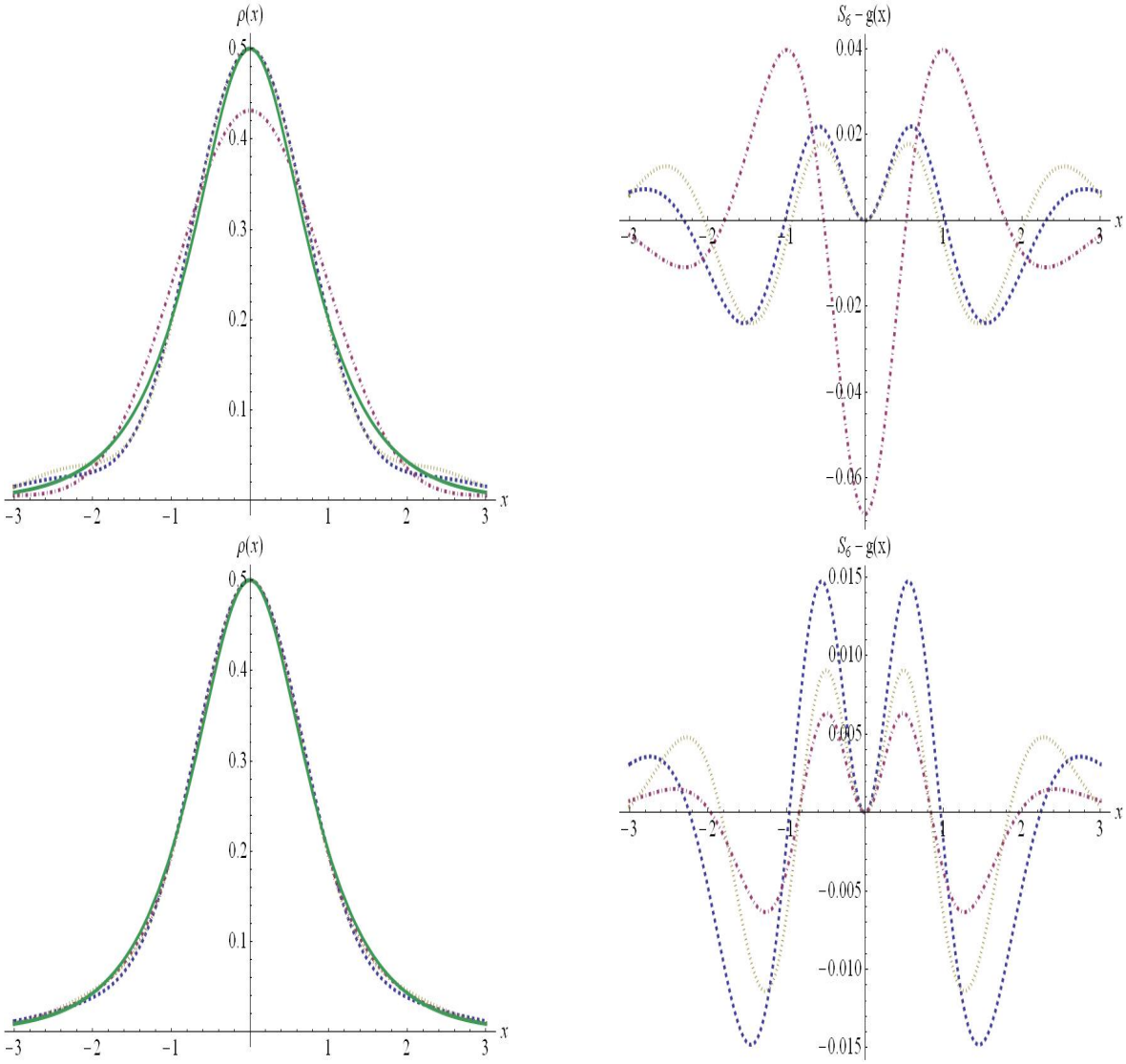


Figure 4.6: Comparison between different orders in partial sum: The top panel is the textbook Gram-Charlier of the Hypersecant test PDF; Bottom panel is the Hildebrandt polynomials of the Hypersecant test PDF; On the left hand is the distributions and on the right the difference between the partial sum and the test PDF. Dashed line is the 4th order partial sum; Dot-dashed is the 6th order; the dotted line is the 8th order and the solid line is the test PDF. We also applied the minimization strategy ($(S_6(0) - g(0))^2$) throughout.

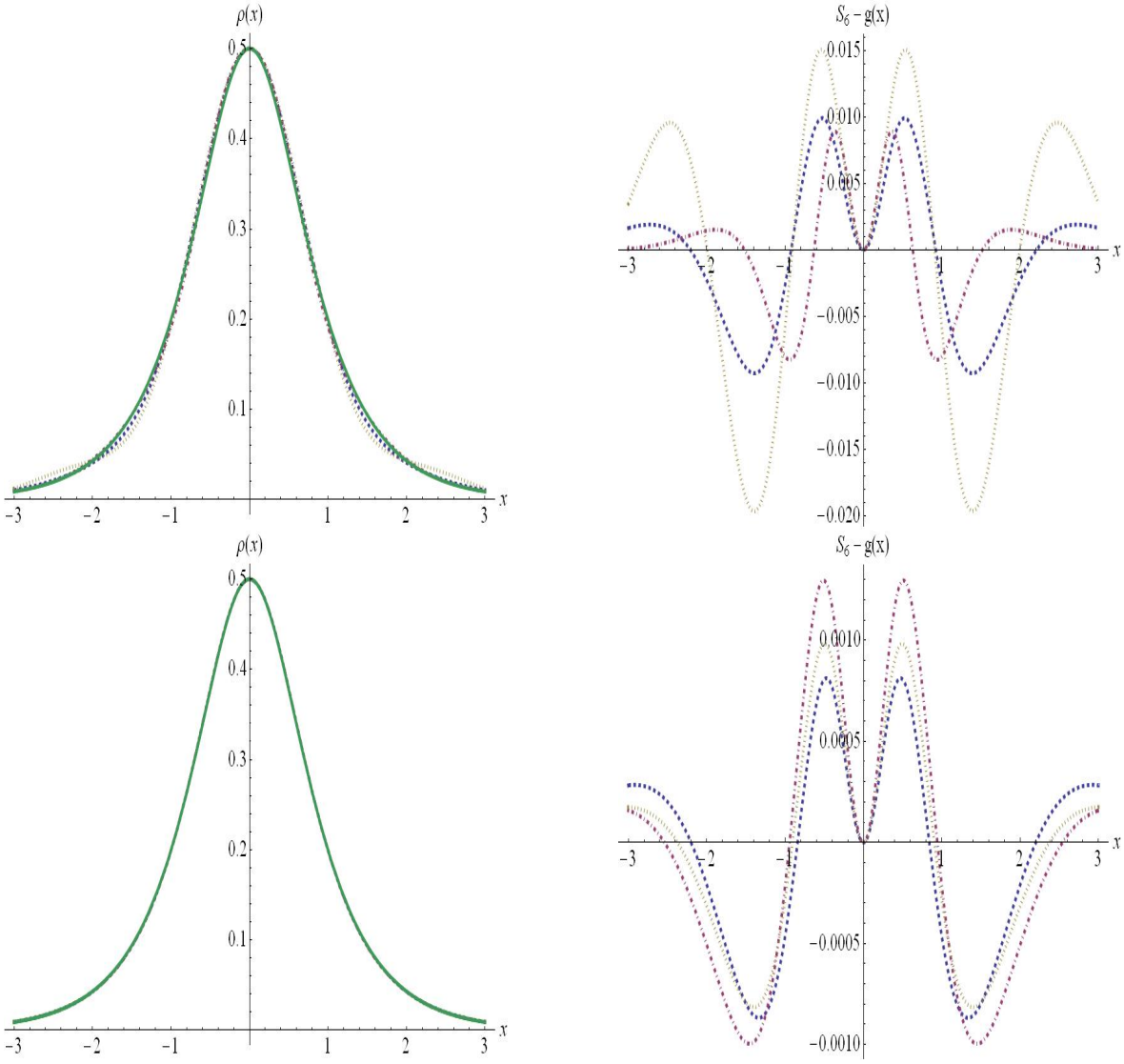


Figure 4.7: Comparison between different orders in partial sum: The top panel is the SNIG Gram-Charlier of the Hypersecant test PDF; Bottom panel is the SNIG polynomials of the Hypersecant test PDF; On the left hand is the distributions and on the right the difference between the partial sum and the test PDF. Dashed line is the 4th order partial sum; Dot-dashed is the 6th order; the dotted line is the 8th order and the solid line is the test PDF. We also applied the minimization strategy $((S_6(0) - g(0))^2)$ throughout.

Chapter 5

Results for sampling distributions

While the previous chapter was concerned with theoretical calculations, this chapter is devoted to the sampling of these distributions and how good the series expansions are when applied not to an analytical test PDF but to a "real" data sample.

5.1 Unbiased estimators

Before we can apply our results for series expansions to samples, we must first consider the problem of unbiased estimators. It is well known that results from a data sample only approximate the "true" answers which are unknown in a real experimental situation. An unbiased estimator, call it "$U(Y)$" for the moment, is an experimental quantity (typically some function) of the data points Y_i such that, if the experiment would be repeated many times, the mean of the distribution of all the U measurements would be as close as possible to the "true" value, also known as the expectation value $E(U)$. Mathematically, the experimental quantity $U(Y)$ is unbiased if

$$E(U) = \text{real value of } U \text{ for infinitely many data points.} \tag{5.1.1}$$

Let $Y_1, \ldots, Y_n = \{Y_i\}_{i=1}^n$ be a sample of n independent and identically distributed data points which are distributed according to some parent distribution $f(y)$. The theoretical moments and cumulants of $f(y)$ are written as greek symbols, μ_q and κ_q respectively[1].

We first note that experimental moments m_q are all unbiased estimators of their theoretical counterparts μ_q. The usual experimental mean

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{5.1.2}$$

[1] In this chapter, the indices i, j, k, l will refer to individual data points in the sample, and to avoid confusion we will use the subscript q rather than j for the order of moments and cumulants.

clearly satisfies $E(m_1) = \mu_1$. Similarly, all higher order experimental moments are unbiased estimators of the moments μ_1, μ_2 and μ_3,

$$m_q = \frac{1}{n}\sum_i Y_i^q \qquad \Longrightarrow \qquad E(m_q) = \mu_q, \qquad q = 1, 2, \ldots . \tag{5.1.3}$$

Since the first-order cumulant and moment are identical, $k_1 = m_1$, the first cumulant k_1 is clearly unbiased. However, in second order the naive guess

$$k_2(\text{wrong}) = \frac{1}{n}\sum_i (Y_i - m_1)^2 \tag{5.1.4}$$

is biased because

$$E(k_2(\text{wrong})) = \frac{n-1}{n}\,\kappa_2. \tag{5.1.5}$$

As is well known, the unbiased estimator for κ_2 is therefore

$$k_2 = \frac{1}{n-1}\sum_i (Y_i - m_1)^2. \tag{5.1.6}$$

5.2 k-statistics

The theory of finding unbiased estimators for the "true" cumulants κ_j is called "k-statistics". Below, we reproduce a simple one-dimensional version of a multivariate formalism of Ref. [28].

Amazingly, not only cumulants but also products of moments are not *a priori* unbiased estimators, because (for example) $m_1 m_1 = n^{-2}\sum_{i=1}^n \sum_{j=1}^n Y_i Y_j$ contains "diagonal" terms where $i = j$ refers to the same data point. The unbiased estimator, termed $m_{(1)}^2$, is clearly the "off-diagonal" part of the double sum,

$$m_{(1)}^2 = \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} Y_i Y_j, \tag{5.2.1}$$

normalised so that $E(m_{(1)}^2) = \mu_1^2$. These unbiased estimators are called "symmetric means" [28]; up to order 4 they are given by

$$m_{(1)}^2 = \sum Y_i Y_j / n^{(2)}, \qquad m_{(1)} m_{(2)} = \sum Y_i Y_j^2 / n^{(2)}, \tag{5.2.2}$$

$$m_{(2)}^2 = \sum Y_i^2 Y_j^2 / n^{(2)}, \qquad m_{(1)} m_{(3)} = \sum Y_i Y_j^3 / n^{(2)}, \tag{5.2.3}$$

$$m_{(1)}^3 = \sum Y_i Y_j Y_k / n^{(3)}, \qquad m_{(1)}^2 m_{(2)} = \sum Y_i Y_j Y_j Y_k / n^{(3)}, \tag{5.2.4}$$

where sums run over unequal values of the indices, $i \neq j \neq k$ and $n^{(q)} = n(n-1)\cdots(n-q+1)$. It is a straightforward exercise to verify e.g. that $E[m_{(1)}^2 m_{(2)}] = \mu_1^2 \mu_2$ and similarly for the remaining statistics listed above.

For experimental purposes, symmetric means are inconvenient because of all the unequal indices. It is much simpler to measure all the normal moments m_q and from them to calculate the symmetric means. For example, we have by direct calculation

$$m_{(1)}^2 = \frac{n}{n-1} m_1^2 - \frac{1}{n-1} m_2. \tag{5.2.5}$$

We now turn to cumulants. Corresponding to each true κ_q we look for the corresponding unbiased estimator, denoted by the Latin k_q, such that k_q is an unbiased estimator of κ_q, i.e. we require that

$$E(k_q) = \kappa_q \qquad q = 1, 2, \ldots. \tag{5.2.6}$$

The general formalism for higher orders is written in terms of auxiliary functions ϕ with q subscripts,

$$k_2 = n^{-1} \sum_{ij} \phi_{ij}\, Y_i\, Y_j, \tag{5.2.7}$$

$$k_3 = n^{-1} \sum_{ijk} \phi_{ijk}\, Y_i\, Y_j\, Y_k, \tag{5.2.8}$$

$$k_4 = n^{-1} \sum_{ijkl} \phi_{ijkl}\, Y_i\, Y_j\, Y_k\, Y_l \tag{5.2.9}$$

and the functions ϕ are determined by the demand that they must ensure that a given sum yields an unbiased estimator for κ_j,

$$E(k_2) = E(n^{-1} \sum_{ij} \phi_{ij}\, Y_i\, Y_j) = \kappa_2, \tag{5.2.10}$$

$$E(k_3) = E(n^{-1} \sum_{ijk} \phi_{ijk}\, Y_i\, Y_j\, Y_k) = \kappa_3, \tag{5.2.11}$$

etc. These are all examples of k-statistics[2]. The results of the calculation (based on equality

[2]We note in passing that the unbiased estimators m_j and k_j do not satisfy the same relations as their "true" counterparts; for example while

$$\mu_2 \equiv \kappa_2 + \kappa_1^2, \tag{5.2.12}$$

the corresponding expression is

$$m_2 = \frac{(n-1)}{n}\, k_2 + k_1^2. \tag{5.2.13}$$

or nonequality of the sum indices) give [28]

$$\phi_{ii} = \phi_{iii} = \phi_{iiii} = 1, \tag{5.2.14}$$

$$\phi_{ij} = \phi_{iij} = \phi_{iijj} = \frac{-1}{(n-1)}, \tag{5.2.15}$$

$$\phi_{ijk} = \phi_{iijk} = \frac{2}{(n-1)(n-2)}, \tag{5.2.16}$$

$$\phi_{ijkl} = \frac{-6}{(n-1)(n-2)(n-3)}, \tag{5.2.17}$$

where the use of different subscripts implies unequal data points (i.e. ϕ_{iij} implies $i \neq j$ and so on).

Sample cumulants are best approached in terms of these symmetric means; for example, the unbiased estimators for the 2nd and 4th order cumulant are

$$k_2 = m_2 - m_{(1)}^2, \tag{5.2.18}$$

$$k_4 = m_4 - 4m_{(3)}m_{(1)} - 3m_{(2)}^2 + 12m_{(1)}^2 m_{(2)} - 6m_{(1)}^4, \tag{5.2.19}$$

simply because the expectation value of each term evaluates to the corresponding product of true moments, see also Eq. 2.3.1. From the relations between symmetric means and products of sample means, experimentally measurable estimators can then be constructed.

5.3 ℓ-statistics

In the case of HBT interferometry of identical particles, the k-statistics can be further simplified because the measurements of momentum differences for identical particles are symmetric under parity transformations: for each Y_i there will be exactly one other data point $-Y_i$. An easy way to incorporate this into our algebra is to assume that we have a data sample with $2n$ points, $\{Z_i\}_{i=1}^{2n}$ such that $Z_i = +Y_i$ for $i \leq n$ and $Z_i = -Y_i$ for $n{+}1 \leq i \leq 2n$. Let us call the cumulants of this "double sample" ℓ_q. We then apply the same formalism of k-statistics to these ℓ_q and reduce all quantities to moments m_q and cumulants k_q of the original sample $\{Y_i\}_{i=1}^{n}$.

All moments and cumulants of odd order are found to be identically zero, $\ell_{2q+1} = 0$, as they should be for a symmetrical sample. For the even orders, we find that the 2nd order cumulant of the Z-sample equals the 2nd order moment of the Y-sample, while the 4th order also simplifies considerably compared to Eqs. 5.2.18–5.2.19,

$$\ell_2 = m_2, \tag{5.3.1}$$

$$\ell_4 = m_4 - 3m_{(2)}^2. \tag{5.3.2}$$

Again we can express these in terms of normal sample moments, e.g.

$$\ell_4 = m_4 + \frac{3}{n-1}m_4 - 3\frac{n}{n-1}m_2^2 \tag{5.3.3}$$

In a similar way the 6th order can also be done,

$$\begin{aligned}
\ell_6 &= m_6 - 15m_{(4)}m_{(2)} + 30m_{(2)}^3 \\
&= m_6 - \left(\frac{n}{n-1}15m_{(4)}m_2 - \frac{15}{n-1}m_6\right) \\
&\qquad + \frac{2}{n^{(3)}}\left(15n^3\, m_2^3 - 45n^2 m_2 m_4 + 30n m_6\right) \\
&= \frac{n^2+12n+32}{(n-1)(n-2)}m_6 - \frac{n^2+4n}{(n-1)(n-2)}15m_2m_4 + \frac{2n^2}{(n-1)(n-2)}15m_2^3
\end{aligned} \tag{5.3.4}$$

5.4 Standard errors

For a proper experimental analysis, we need to know not only the unbiased estimators of various moments and cumulants, but also the standard errors for them. The standard error is defined as the square root of the variance. Again we must distinguish between a *theoretical* variance which is defined in terms of expectation values; for example

$$\mathbf{var}(\ell_2, \ell_2) = E(\ell_2\ell_2) - E(\ell_2)\,E(\ell_2), \tag{5.4.1}$$

and the corresponding *unbiased estimator* of such a variance, which we shall call "ecov" (experimental covariance). From the above, we have

$$\begin{aligned}
\mathbf{var}(\ell_2, \ell_2) &= \mathbf{var}(m_2, m_2) = \frac{1}{n^2}\mathbf{var}\left(\sum_i Y_i^2, \sum_j Y_j^2\right) \\
&= \frac{1}{n}\left(\mu_4 - \mu_2^2\right),
\end{aligned} \tag{5.4.2}$$

and the corresponding unbiased estimator is

$$\mathrm{ecov}(\ell_2, \ell_2) = \frac{1}{n}(m_4 - m_{(2)}^2). \tag{5.4.3}$$

In a similiar way, the estimator for the variance of the fourth cumulant ℓ_4 is, in terms of sums $s_q = n\, m_q$

$$\begin{aligned}
\mathrm{ecov}(\ell_4, \ell_4) = \frac{1}{n^{(4)}}\Bigg[&(n^2+7n+24)s_8 - 12(n+3)s_2s_6 + \frac{12(n^2+2n-6)}{n(n-1)}s_2^2s_4 \\
&+36s_2^2s_4 - \frac{n^3+6n^2-n-24}{n(n-1)}s_4^2 - 18s_4^2 - \frac{2(2n-3)}{n(n-1)}s_2^4\Bigg].
\end{aligned} \tag{5.4.4}$$

Table 5.1 compares the sampling cumulants k_j to their known theoretical counterparts as a function of the sample size (number of data points) n. See Table 4.1 for a list of comparable theoretical values of κ_q. The first anomaly we notice in the table is that the variance of κ_6 of the Student's t distribution does not decrease with sample size. This is the numerical implication of diverging higher moments. The second thing we notice is that the higher the kurtosis, the higher the variance in our cumulants and the more data points we need to sample before we attain a certain accuracy. Using Table 5.1 we decide on $N = 10^6$ as a conservative estimate for our S_6 partial sum comparison in the next section.

Test Distributions	Cumulant	$N = 10^4$	10^6	10^8
NIG($\alpha = \delta = 1$)	k_2	0.983	0.999	1.00
	$\mathbf{var}(k_2)$	5×10^{-4}	$5. \times 10^{-6}$	$5. \times 10^{-8}$
	k_4	2.58	3.02	3.01
	$\mathbf{var}(k_4)$	0.117	0.0026	$2. \times 10^{-5}$
	k_6	26.2	45.5	45.0
	$\mathbf{var}(k_6)$	898	21.0	2.2×10^{-4}
Student's t	k_2	0.998	0.997	1.00
$(a = \sqrt{2m-3}, m = 3.75)$	$\mathbf{var}(k_2)$	1.01	4×10^{-6}	4.4×10^{-8}
	k_4	1.83	2.28	2.36
	$\mathbf{var}(k_4)$	0.091	0.011	1.9×10^{-4}
	k_6	18.7	48.9	88.5
	$\mathbf{var}(k_6)$	3050	3700	3375
Normal $\sigma = 1$	k_2	1.01	0.999	1.00
	$\mathbf{var}(k_2)$	2×10^{-4}	$2. \times 10^{-6}$	$2. \times 10^{-8}$
	k_4	0.110	0.0029	7.5×10^{-7}
	$\mathbf{var}(k_4)$	0.0029	2.4×10^{-5}	2.4×10^{-7}
	k_6	-0.114	0.0255	0.0025
	$\mathbf{var}(k_6)$	0.0678	8.6×10^{-4}	6.7×10^{-6}
Sum of 3 Uniform	κ_2	0.992	1.00	1.00
	$\mathbf{var}(\kappa_2)$	1.5×10^{-4}	$.16 \times 10^{-6}$	1.6×10^{-8}
	κ_4	-0.384	-0.401	-0.400
	$\mathbf{var}(\kappa_4)$	9.5×10^{-4}	2.4×10^{-5}	9.6×10^{-8}
	κ_6	0.700	0.762	0.764
	$\mathbf{var}(\kappa_6)$	0.041	3.2×10^{-4}	2.9×10^{-6}

Table 5.1: Examining variance in cumulants as sample size increases.

Distribution	Gram-Charlier	SNIG Gram-Charlier	Hildebrandt	SNIG polynomials
SNIG($\alpha = \delta = 1$)	0.327	1.1×10^{-4}	0.0014	9.4×10^{-5}
Laplace	0.178	0.0173	0.0146	7.8×10^{-4}
Student's t($m = 3.75$)	0.132	0.0018	2.4×10^{-4}	0.0059
Hypersecant	0.039	1.6×10^{-4}	9.4×10^{-4}	1.7×10^{-4}
Student's t($m = 4.5$)	0.084	4.5×10^{-4}	4.5×10^{-4}	2.2×10^{-4}
Logistic	0.009	1.7×10^{-4}	1.7×10^{-4}	3.1×10^{-4}
SNIG($\alpha = \delta = \sqrt{6}$)	5.9×10^{-4}	1.5×10^{-4}	1.0×10^{-4}	0.0056

Table 5.2: Results for sampling from non-Gaussian distributions. The first column specifies the test PDF from which a sample size of 10^6 data points was drawn. Subsequent columns list the χ^2 values obtained for various series approximations: as always, a small χ^2 signifies a good approximation.

5.5 Sampling results

We drew a sample ($N = 10^6$) of the specified distribution and then measured the first three even cumulants and $f(0)$ of the distribution. We then binned the data (into 125 bins) and numerically integrated the approximation which is based on the measured quantities. Finally we used the Pearson χ^2 measure, making sure there are more than 5 counts in every bin, to compare the integrated value with the binned value. The results are given in Table 5.2, from which we can observe the following:

- Each of the methods succeeds or struggles with different distribution, while on average all the approximations are of order 10^{-4} except Gram-Charlier.
- From the Table 5.2 we can also see that the approximation is also constrained by the number of data points ($N = 10^6$) as some of the partial sums that we know from the previous section are better all have the same order of fit $\approx 10^{-4}$.
- The Hildebrandt polynomials gives worse approximations for the high-kurtosis SNIG distribution and the Laplace distribution.
- The Gram-Charlier Type A series performance is directly correlated with the kurtosis of the test PDF, for the small-kurtosis SNIG it performs just as well as the other expansions. As the kurtosis is increased the approximation worsens, for the large kurtosis SNIG it performs worse by a factor of a 1000.

Chapter 6

Conclusions

Can we systematically describe a correlator function in one dimension?

The problem was posed in Chapter 1 whether we could systematically describe the experimental correlation function $R(\mathbf{q})$. Normalising the correlation function to become a probabilty density function, we looked for statistical quantities to describe the shape of the correlator. Use of the most versatile statistical quantities, the moments and cumulants, led us directly to the topic of Gram-Charlier series.

Starting out from the generic Gram-Charlier series as an expansion in the derivatives of a reference PDF, we explored different variants, mostly by using non-Gaussian "test PDFs" to see how well a given approach would work.

In doing so, we made the following observations:

a) The generic Gram-Charlier formalism is couched in terms of a reference function which we can choose to be close to the function we are trying to approximate (the correlator). As terms of higher order are added to the series, the series represents the correlator better in terms of the cumulants.

b) Gram-Charlier series converge only in some special cases (in the L_2 sense). From a practical point of view, the important question is not the convergence but the accuracy of a partial sum with only a few terms. Gram-Charlier series are generally asymptotic in character, and so can give accurate approximations even if they do not converge. Asymptotic expansions can give arbitrary precision as long as the argument of the expansion is large enough, but this is not necessarily at the measured values of the argument.

c) When using the Gaussian as the reference function, the Gram-Charlier Series becomes the Gram-Charlier Type A series, which basically expresses the ratio between the cor-

relator and a Gaussian reference in a series of Hermite polynomials. This textbook expansion is simple but it has its problems: it cannot approximate negative kurtosis and does not converge for distributions with positive kurtosis.

d) The Gram-Charlier series is based on a Rodrigues formula without necessarily resulting in an orthogonal system. It also connects derivatives with the higher moments. The Rodrigues formula is also intimately tied to the Pearson differential equation. Solutions of the Pearson differential equation give weights which lead to orthogonal polynomials, but the solutions do not necessarily lead to a complete orthogonal system (counterexample: Student's t).

e) If we assume that the random variable from which the correlator is made up is the sum of n primieval random variables, we can reorder the terms in the series according to order of n rather than order of derivative. This is called an Edgeworth series and what is not really mentioned in the literature is that this is a general procedure not necessarily limited to the Gaussian reference PDF.

f) The Edgeworth prescription improves the partial sum remarkably well for any distribution which is close to Gaussian, including distributions with low kurtosis as they can be viewed as a sum of variables, see [30].

g) The problems with the Gram-Charlier Type A (Gaussian reference PDF) series lead us to look elsewhere for other non-Gaussian reference PDFs. Obvious candidates are the beta distribution for negative-kurtosis correlators and the Student's t distribution for a correlator with positive kurtosis.

h) The beta distribution used in the generalised Gram-Charlier series leads to the Fourier-Jacobi system, which solves the negative partial sum problem the Type A series had. However, the coefficients in the series can no longer be expressed as cumulant differences, and the limited support implies that the the Fourier transform will oscillate, making the beta distribution unsuitable as a approximation for the relative distance distribution.

i) We have looked at two PDFs which have two free parameters: the Student's t distribution and the Symetrical Normal Inverse Gauss (SNIG). Not surprisingly, the two free parameter families give better approximations than the Hermite system.

j) The Student's t distribution is a weight suited for positive kurtosis correlators. It also also generates orthogonal polynomials. Unfortunately, this distribution provides only a limited basis because it only has $2m - 1$ moments while the higher ones diverge. It

might be an appropriate basis for a correlator that has rapidly increasing cumulants or for low kurtosis in which case the Student's t distribution will have a larger m value and thus more polynomials than needed in a realistic partial sum.

k) Exploring our options in terms of positive kurtosis reference functions we have two choices for properties of the coefficients: either keep orthogonality or keep their relation to cumulant differences. We chose the SNIG distribution for both procedures as it seems to be the only other distribution which is symmetric over the real line and has two free parameters. The numerical results with all three expansions provide roughly the same results with the orthogonal SNIG polynomials being the best.

l) One of the practical consideration is how many cumulants are accessible for a finite sample; in other words, whether the sampling variance is small enough for a cumulant of a given order to be meaningful. This has been studied numerically for some test cases.

m) All the series expansion we considered are asymptotic, because they are all based on the moments(or cumulants) of the sample. The amount of terms that is used in the partial sum should be investigated for every individual case, but there are still a number of factors that should be considered in general. The size of the sample i.e. for $N = 10^6$ there is no point in adding the eight term because of the uncertainty in the 8th moment of the sample. The other factor is the shape of the underlying distribution, if it has a cusp all the expansions will perfom badly and if it is close to a certain reference distribution, that partial sum will obviously perform better.

n) The cumulants are measured experimentally by the unbiased estimators within the k-statistics formalism. In Section 5.3, we have simplified the algebra of k-statistics considerably for the estimation of cumulants of symmetric functions.

o) Among the open questions that we would like to address are:

   a) Why does the Jacobi-Fourier perform so much better than the other series expansions, even if it is still an asymptotic series?

   b) It would be interesting to apply Edgeworth re-ordering to other distributions to see if there is an improvement.

   c) The idea of modified moments is to use a polynomial and a different generating function than the Fourier transform to see if we can absorb the correction function in the Rodrigues formula.

d) Investigating the empirical characteristic function. This is a point Fourier Transform and we might be able to use it to compute the x-moments directly.

e) It is important to see how well these expansions perform on measured data and not just simulated data.

All of the above is still only a wishlist.

Bibliography

[1] Brown, R.H. and Twiss, R.Q.: The Question of Correlation between Photons in Coherent Light Rays. *Nature*, vol. 178, pp. 1447–1448, 1956.

[2] Goldhaber, G., Fowler, W.B., Goldhaber, S. and Hoang, T.F.: Pion–Pion Correlations In Antiproton Annihilation Events. *Phys. Rev. Lett.*, vol. 181, no. 3, 1959.

[3] Goldhaber, G., Goldhaber, S., Lee, W.Y. and Pais, A.: Influence of Bose-Einstein Statistics on the Antiproton-Proton Annihilation Process. *Phys. Rev.*, vol. 120, no. 1, 1960.

[4] Brown, D.A. and Danielewicz, P.: Imaging of sources in heavy-ion reactions. *Phys. Lett. B*, vol. 398, no. 3-4, pp. 252–258, April 1997.

[5] Csörgő, T., Lorstad, B. and Zimanyi, J.: Bose–Einstein Correlations for Systems with Large Halo. *Z.Phys. C*, vol. 71, p. 492, 1996.

[6] Wiedemann, U.A. and Heinz, U.: Particle Interferometry for Relativistic Heavy-Ion Collisions. *Phys. Rep.*, vol. 319, pp. 145–230, 1999.

[7] Chapman, S., Nix, J.R. and Heinz, U.: Extracting source parameters from Gaussian fits to two-particle correlations. *Phys. Rev. C*, vol. 52, no. 5, pp. 2694–2703, Nov 1995.

[8] Bertsch, G., Gong, M. and Tohyama, M.: Pion interferometry in ultrarelativistic heavy-ion collisions. *Phys. Rev. C*, vol. 37, pp. 1896–1900, May 1988.

[9] Wiedeman, U.A. and Heinz, U.: Resonance contributions to Hanbury Brown–Twiss correlation radii. *Phys. Rev. C*, vol. 56, no. 6, pp. 3625–3286, 1997.

[10] STAR collobration: Pion Interferometry of $\sqrt{s_{NN}} = 130$ GeV Au+Au Coliisions at RHIC. *Phys. Rev. Lett.*, vol. 87, 2001.

[11] Danielewicz, P. and Pratt, S.: Analysis of low-momentum correlations with Cartesian harmonics. *Phys. Lett. B*, vol. 618, no. 1-4, pp. 60–67, 2005.

[12] Csörgő, T., Hegyi, S. and Zajc, W.A.: Bose-Einstein correlations for Lvy stable source distributions. *EPJ C*, vol. 36, no. 1, pp. 67–78, 2004.

[13] Eggers, H. and Lipa, P.: HBT shape analysis with q -cumulants. *Braz. J. Phys.*, vol. 37, no. 3, 2007.

[14] N.McDonald, J. and Weiss, N.A.: *A Course in Real Analysis*. Academic Press, 1999.

[15] Stuart, A. and Ord, J.K.: *Kendall's Advanced Theory of statistics*, vol. 1, Distribution Theory of *Kendall's Library of Statistics*. Sixth edn. Arnold, 2003.

[16] Szegő, G.: *Orthogonal Polynomials*, vol. XXIII of *Colloquim Publications*. 4th edn. American Mathematical Society, 1975.

[17] Ismail, M.E. and Assche, W.V.: *Classical and Quantum Orthogonal Polynomials in One Variable*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2005.

[18] Harkness, W.L. and Harkness, M.L.: Generalized Hyperbolic Secant Distributions. *J. Am. Stat. Assoc.*, vol. 63, no. 321, pp. 329–337, 1968.

[19] Barndorf-Nielsen, O.E.: Normal Inverse Gaussian and Stochastic Volatility Modelling. *Scand. J. Stat.*, vol. 24, no. 1, pp. 1–13, 1997.

[20] Hurst, S.: The Characteristic function of the Student's t distribution. Research Report SRR95-044, Mathematical Science Institute, 1995.

[21] S, H. and Csörgő, T.: Edgeworth expansion of Bose-Einstein correlation functions. In: *Budapest, Proceedings, Relativistic heavy ion collisions*, pp. 47–51. 1993.

[22] Hegyi, S. and Csörgő, T.: Model independent shape analysis of correlations in 1, 2 or 3 dimensions. *Phys. Lett. B*, vol. 489, no. 15, 2000.

[23] Arfken, G.B. and Weber, H.-J.: *Mathematical Methods for Physicists*. Elsevier Academic Press, 2005.

[24] Samuelson, P.A.: Fitting General Gram-Charlier Series. *Ann. Math. Stat.*, vol. 14, no. 2, pp. 179–187, 1943.

[25] Dunn, E.L. and Stein, F.: Families of Sturm-Liouville Systems. *SIAM Review*, vol. 3, no. 1, pp. 54–65, 1961.

[26] Wallace, D.L.: Asymptotic Approximations to Distributions. *Ann. Math. Stat.*, vol. 29, no. 3, pp. 635–654, 1958.

[27] Cramér, H.: *Mathematical Methods of Statistics*. Princeton University Press, 1945.

[28] McCullagh, P.: *Tensor Methods in Statistics*. Monographs on statistics and applied probability. Chapman and Hall, 1987.

[29] Csörgő, T., Kittel, W., Metzger, W.J. and Novák, T.: Parametrization of Bose–Einstein Correlations in e+ e- annihilation and Reconstruction of the Space-Time Evolution of Pion Production in e+e- Annihilation. *Phys. Lett. B*, vol. 663, no. 214, 2008.

[30] Blinnikov, S. and Moessner, R.: Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement Series*, vol. 130, p. 193, 1998.

[31] Metzger, W.J., Novák, T., Kittel, W. and Csörgő, T.: Parametrization of Bose–Einstein Correlations in e+ e- annihilation and reconstruction of the source function. *Int. J. Mod. Phys. E*, vol. 16, p. 3224, 2008.

[32] Romanovsky, V.: Generalisation of Some Types of the Frequency Curves of Professor Pearson. *Biometrika*, vol. 16, no. 1/2, pp. 106–117, 1924.

[33] Hildebrandt, E.H.: Systems of Polynomials Connected with the Charlier Expansions and the Pearson Differential and Difference Equations. *Ann. Math. Stat.*, vol. 2, no. 4, pp. 379–439, 1931.

[34] Diaconis, P. and Zabell, S.: Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre. *Stat. Sci.*, vol. 6, no. 3, pp. 284–302, 1991.

[35] Beale, F.S.: On a Certain Class of Orthogonal Polynomials. *Ann. Math. Stat.*, vol. 12, no. 1, pp. 97–103, 1941.

[36] Abramowitz, M. and Stegun, I.A.: *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* Dover, 1964.