UNSW
SYDNEY

TOWARDS A SELF-CONTAINED THEORY FOR STOCHASTICS IN THE COMPLEX PLANE

Robert Cantwell

Supervisors: A/Prof Pierre Lafaye de Micheaux and Dr Gery Geenens

School of Mathematics and Statistics
UNSW Sydney

November 2022

Submitted in partial fulfillment of the requirements of the degree of
Bachelor of Science (Advanced Mathematics) with Honours

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: _____ Date: _____

Acknowledgements

First and foremost, my greatest thanks must go to my supervisors Pierre and Gery for their support, guidance and teachings throughout this project. Incredible.

Thanks is also due to my family, whom I have lived with this past year, more closely than usual due to the pandemic restrictions. Their support and encouragement, despite not always understanding what I was banging on about, helped me go further than I otherwise may have and solve many problems by talking about them.

To Coti Manassero, Andrea Morello, Kate Selway and Vince Todd in applied disciplines, a big thank you for your patient explanations on why you use complex numbers in the first place. Also to Daniel Chan and Jan Zika for taking the time to assist an honours student emailing them somewhat out of the blue, thank you.

Thank you Yudhi for your fresh eyes and insightful feedback, both when proofreading parts of this thesis and throughout the whole year.

Finally, to the reader. I hope that you are able to take away some of the fascination of complex numbers in a stochastic setting. Thank you very much for the time invested to read this — without you, the words on these pages mean nothing.

Robert Cantwell, 19 November 2021.

Notation

To avoid confusion with various notation in the literature for handling complex numbers, the notation used throughout this thesis is defined here.

i	The imaginary unit, $i^2 = -1$.
$\Re(z), \Im(z)$	The real and imaginary components x and y of complex number $z = x + iy$.
z^*	The conjugate of z, $z^* = \Re(z) - i\Im(z)$.
$\lvert z\rvert$	The modulus of z, $\lvert z\rvert = \sqrt{\Re(z)^2 + \Im(z)^2}$.
$\mathrm{Arg}(z)$	The principal argument of z, $\mathrm{Arg}(z) = \tan^{-1}\frac{\Im(z)}{\Re(z)} \pm \pi \in (-\pi, \pi]$.
$\mathbb{R}, \mathbb{Q}, \mathbb{Z}$	The field of real numbers, rationals and integers.
$\mathbb{C}$	The field of complex numbers.
$M_{mn}(\mathbb{R})$	The field of $m \times n$ matrices with real entries.
$M_{mn}(\mathbb{C})$	The field of $m \times n$ matrices with complex entries.
$\mathbb{P}(\cdot)$	Probability of event $\cdot$.
$\mathbb{E}(\cdot)$	The expectation of $\cdot$.
$\mathbf{A} = [a_{kl}]$	The matrix $\mathbf{A}$ where the kl^{th} entry is a_{kl}.
$\mathbf{A}^\top,\ \mathbf{A}^\mathcal{H}$	The transpose and Hermitian (conjugate transpose) of matrix $\mathbf{A}$.
$\boldsymbol{z}$	A vector $(z_1, z_2, ..., z_n)^\top$ in $\mathbb{C}^n$ for some $n \in \mathbb{N}$.
$\lVert\boldsymbol{z}\rVert$	The L_2-norm $\sqrt{\sum_{k=1}^{n} \lvert z_k\rvert^2}$ of $\boldsymbol{z}$.
$\overline{Z}_n$	The arithmetic mean $\frac{1}{n}\sum_{k=1}^{n} Z_k$ of a sample $\{Z_1, \ldots, Z_k\}$.
$\overline{Z}_n^{(g)}$	The geometric mean $\left(\prod_{k=1}^{n} Z_k\right)^{\frac{1}{n}}$ of a sample $\{Z_1, \ldots, Z_k\}$.
$Z \sim F$	The random variable Z has distribution F.
$\xrightarrow{\text{d}}, \xrightarrow{\mathbb{P}}, \xrightarrow{\text{a.s.}}$	Convergence in distribution, in probability and almost surely.
$o(\cdot)$	Small o notation where $\lim_{\cdot \to 0} \frac{o(\cdot)}{\cdot} = 0$.
$\mathrm{mod}(m)$	Modulo m, i.e. the remainder after dividing by m.

Specific points of interest are that:

- the imaginary unit is given by i (as opposed to j, $\imath$, and $\jmath$ which are also often encountered in the literature).
- to avoid any confusion, neither i nor j will be used as index counters in sets, sums or products.
- the conjugate of complex number z is given by z^* (not $\bar{z}$ to avoid confusion with a sample average) and the Hermitian by $\boldsymbol{z}^{\mathcal{H}}$.
- where possible, z and w will be used to represent complex numbers, whilst x, y, u and v will denote real numbers;
- boldface is used to indicate a vector rather than a scalar;
- capital letters will be used for matrices or random variables, with matrices upright and the context should be clear.

Abstract

This thesis is an original approach to complex statistics from within the complex field. Data represented as complex numbers arise in numerous disciplines such as electrical engineering, neuroscience, geophysics and meteorology. Proper statistical analysis of such data is conditional on a fully-fledged theory of probability and random variables in $\mathbb{C}$. Yet, to date, most of the available theory has been derived by simply translating notation from $\mathbb{R}^2$ to $\mathbb{C}$. Such theory based on $\mathbb{R}^2$ inherently overlooks the complex multiplication, the availability of which often being the motivation to work in $\mathbb{C}$ in the first place. Our self-contained approach uncovers new results to augment existing theory, which are essential to establish a complete theory of stochastics in the complex plane, different to that of real bivariate random variables.

Contents

Chapter 1

Introduction

Complex numbers are a convenient mathematical construct, much like real numbers, with origins in roots of polynomials where real numbers were insufficient (Descartes (1637); Merzbach and Boyer (2011)). The field of complex analysis provides results that exhibit solutions to otherwise cumbersome or unsolvable real-valued problems. One such example in probability is using Cauchy's residue theorem to evaluate the integral $\int_{-\infty}^{\infty} \frac{e^{itx}}{\pi(1+x^2)}\,\mathrm{d}x$, which yields the characteristic function of a Cauchy distribution. However, from a statistical perspective complex numbers can also be used to represent data. Whilst seemingly counter-inuitive, complex numbers — and specifically the complex multiplication — provide a convenient mathematical representation of real phenomena in fields such as electrical engineering, neuroscience, geophysics and meteorology. Furthermore, wherever random polynomials occur, such as in Bayesian time series methods or random matrix analysis, random complex numbers will also occur as the polynomial roots. In quantum physics, recent work by Renou et al. (2021) even suggests that complex numbers are not completely "imaginary" as coined by Descartes (1637), but in fact essential in the study of our universe from a quantum perspective.

Given data arise as random complex numbers, statistical analysis is dependent upon a fully-fledged theory of probability and random variables on $\mathbb{C}$. However, to date, most of the available theory simply translates concepts and notation from real bivariate statistics. This thesis explores the groundwork to set up a self-contained theory of stochastics in $\mathbb{C}$ that takes full advantage of the field structure — in particular the multiplication. Existing theory for complex random variables is now motivated from within $\mathbb{C}$, rather than $\mathbb{R}^2$. We also extend the theory with new results which give precisely how real univariate statistics is a subset of complex statistics, and highlight the differences between complex and real bivariate statistics.

The next section of this introductory chapter further motivates the consideration of complex-valued data (hereafter just 'complex data') and statistical methods thereupon by outlining applications in various fields. Finally the second section provides the structure for the rest of the thesis.

1.1 Motivation

This section motivates the investigation of probability and statistics on $\mathbb{C}$ by providing an overview of how and where complex data arise. We also answer here why complex data should be statistically manipulated within $\mathbb{C}$ as opposed to splitting it into its real components.

1.1.1 Where do complex data arise

Complex data arise in contexts wherever the complex multiplication makes sense algebraically or geometrically.

Firstly, in the historical use of complex numbers to solve polynomials (Descartes (1637); Merzbach and Boyer (2011)), wherever random polynomials occur, so too will random complex roots of the polynomial. Random polynomials have been studied since at least 1943 by Mark Kac (Kac (1943)) and the study of the distribution of the roots followed with Erdös and Turán (1950) and Hammersley (1956). Random polynomials remain an open area of research today with applications in random matrix theory (Tao (2012); Edelman and Wang (2013)) or Bayesian time series analysis (Chib and Greenberg (1994); Edelman and Wang (2013); Can et al. (2019); Amry (2020); Suparman (2020)).

Secondly, complex numbers are a convenient representation of a waveform by Euler's formula $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. Where real instruments detect behaviour of an underlying (usually electromagnetic) wave, if one takes the wave as the real component of a complex number, then physical behaviour such as phase shifts and amplitude scaling are represented by linear transformations of complex numbers. Mathematical details are elaborated in Section 2.3. Importantly, it is the polar components of complex data which represent quantities of physical interest in most of the following applications. Fields which analyse waveforms as complex data include: signal processing (Picinbono and Chevalier (1995); Fuhrmann (1999); Eriksson et al. (2010); Schreier and Scharf (2010); Adali and Schreier (2014); Yan et al. (2019)), magnetotelluric methods in geophysics (Cagniard (1953); Caldwell et al. (2004); Varentsov et al. (2003); personal correspondence with Dr Kate Selway and Dr Constanza Manassero) and functional Magnetic Resonance Imaging (fMRI) in neuroscience (Ogawa et al. (1990); Bandettini et al. (1993); Rowe and Logan (2004); Adrian et al. (2018)). Complex numbers are also used in quantum physics to describes states in terms of wave functions. However, in this application complex numbers represent probabilities rather than real outcomes (Schrödinger (1926); Wigner (1932); Aaronson (2013); Weingarten (2002); Goyal et al. (2010); Jansson (2019); Renou et al. (2021); personal correspondence with Dr Andrea Morello). A separate discussion of complex probability and quantum theory will take place in Chapter 3.

Thirdly, complex numbers may also be used to model real bivariate data where the geometry of the complex multiplication as a rotation is meaningful. Again, the physical quantities of interest are represented by the polar components of a complex number such as wind speed and direction, or ocean current strength and direction (Gonella (1972); Mooers (1973); Burt et al. (1979); Foreman and Henry (1989); Goh et al. (2006)). Whilst rotation matrices may achieve the same rotational geometry in $\mathbb{R}^2$, complex data also arise in these fields as the discrete Fourier transform of time series observations (details on the transform provided in Section 7.3). Complex random variables have even been used to represent real bivariate random variables in other fields, such as economics (Svetunkov (2012)). Yet, although Svetunkov (2012) remark on different conclusions from methods such as linear modelling — which is to be expected as the complex multiplication interacts real and imaginary components in a particular fashion — no justification is given as to why the complex

multiplication makes sense. Thus, complex data exist in certain fields where the complex multiplication has a physical interpretation, and as such complex random variables are not directly analogous to real bivariate random variables, which cannot naturally be multiplied.

1.1.2 Probability and statistics over complex numbers

Just as data in the aforementioned fields are naturally represented as complex numbers, complex random variables constructed to represent randomness in observations should also be manipulated within $\mathbb{C}$. Almost all probabilistic and statistical results to date for complex random variables have been developed by treating the random variable $Z = X + iY$ as a real bivariate random variable $(X, Y)^\top$ (primary references are: Wooding (1956); Eriksson et al. (2010); Schreier and Scharf (2010); Ducharme et al. (2016)). We identify two main contributing factors which have led to this bivariate real approach to complex random variables. Firstly, by $\mathbb{R}^2$ and $\mathbb{C}$ being isomorphic as additive groups, statistical objects pertaining to sums, including the characteristic function and widely-used normal distribution (Wooding (1956)) coincide. Secondly, to perform calculus on real-valued functions of complex random variables, the differential structure of $\mathbb{C}$ is relaxed to imitate that of $\mathbb{R}^2$ by identifying $\mathbb{C}$ as a two-dimensional real Riemannian manifold (further details in Chapter 2), which also means one can identify the distribution of Z with the bivariate distribution of X, Y. However, whilst working in $\mathbb{R}^2$ has established a partial theory of complex random variables, objects pertaining to complex multiplication that do not have analogues in $\mathbb{R}^2$, for example the Mellin transform, are largely ignored.

Furthermore, recent applications of machine learning to complex data have remarked that setting computations up directly in $\mathbb{C}$ has both computational and algorithmic advantages over deconstructing complex numbers and manipulating their real Cartesian components (Trabelsi et al. (2018); Chakraborty and Yu (2019); Sob et al. (2020); Wang et al. (2020)). This concurs with recent publications on the handling of complex numbers in floating point arithmetic (Jeannerod et al. (2017); Lefèvre and Muller (2019); Muller et al. (2018)), whereby from a computational perspective there are efficient methods to encode complex multiplication in a computer which avoid an accumulation of rounding errors in intensive computations.

Overall, both from an analytic and computational perspective, a theory of complex random variables should be self-contained within $\mathbb{C}$ and not constantly having to reference bivariate real statistics to obtain new results. This thesis takes the first steps towards establishing such a self contained theory by motivating existing results within $\mathbb{C}$ which then facilitates novel extensions of the theory. It should not be surprising that results pertaining to sums of complex random variables (e.g. characteristic function, normal distribution as an additive error distribution) coincide with their bivariate counterparts as $\mathbb{C}$ and $\mathbb{R}^2$ are isomorphic as additive groups. However, at times the motivation within $\mathbb{C}$ will also lead to different preferences of notation (such as how we write degrees of freedom for a t-distribution), which align more intuitively with familiar results in terms of statistical modelling.

1.2 Thesis Overview

The rest of the thesis is organised as follows. Chapter 2 provides a background into complex numbers and functions, in particular how complex numbers represent real phenomena and also the differential tools that emerge from the identification of $\mathbb{C}$ as a two-dimensional real Riemannian manifold. Chapter 3 addresses whether we really need to change the differential structure in $\mathbb{C}$ by discussing tangential ideas of complex probability with a novel unification of measure theory and quantum mechanics. Chapter 4 examines the probabilistic description of complex random variables and their distributions. Chapter 5 details specific complex distributions including the complex uniform, normal, t- and log-normal distributions, as well as a novel complex Bernoulli distribution. Chapter 6 then considers statistical results involving complex random samples including asymptotic results (which are now proven in $\mathbb{C}$), estimation methods and hypothesis testing. Finally, Chapter 7 discusses dependence measures for complex random variables and sets up a key result for inference on linear model estimation in the complex case. Due to the extent of groundwork required to establish this theory, theoretical examples are given, with simulated data where required, however no application to a real-world data set is provided here.

Chapter 2

Complex Numbers

This chapter gives a brief overview of complex numbers, how they arise as data and important tools to handle functions of complex variables for use in later chapters.

2.1 Complex Numbers and Real Representations

Complex numbers were historically developed in order to solve seemingly unsolvable polynomials, such as $z^2 + 1 = 0$.[1] The origins of the term imaginary to describe parts of complex numbers can be found in Descartes (1637, Book 3 pp. 40–66), where René Descartes came across polynomials apparently without any positive ("true") or negative ("false") roots. There are two dominant forms of denoting complex numbers and two convenient isomorphisms of $\mathbb{C}$ to spaces over the field of real numbers.

2.1.1 Cartesian form

A complex number is simply any number z of the form $z = x + iy$, where x and y are real numbers, and i is the imaginary unit with the property $i^2 = -1$. Complex addition and multiplication are defined as may be expected in $\mathbb{R}$ somewhat treating i as a pronumeral, but employing the equality $i^2 = -1$. That is, for $z = x + iy$ and $w = u + iv$,

$$\begin{aligned} z + w &= (x + u) + i(y + v), \\ z \times w &= (xu - yv) + i(xv + yu). \end{aligned}$$

2.1.2 $\mathbb{R}^2$ representation

A single complex number can be pictured as the ordered pair of points $(x, y)^\top \in \mathbb{R}^2$, where the first entry is the real component and the second the imaginary (Hamilton (1837)). This representation can be easily extended to the multivariate case by considering a complex vector $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y} \in \mathbb{C}^n$ as the real vector $(\boldsymbol{x}^\top, \boldsymbol{y}^\top)^\top \in \mathbb{R}^{2n}$.

Linear conversions of real bivariate statistics to complex notation particularly take advantage of the augmented vector structure through the following notation and relationship (Eriksson et al. (2010); Schreier and Scharf (2010); Adali and Schreier (2014)). Let $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ and $\boldsymbol{z}^* = \boldsymbol{x} - i\boldsymbol{y}$ be vectors in $\mathbb{C}^n$, and define the augmented vector of $\boldsymbol{z}$ to be

$$\underline{\boldsymbol{z}} \coloneqq \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{z}^* \end{pmatrix} \in \mathbb{C}^{2n}.$$

[1] More accurately for solving depressed cubics in 1500s Italy (Merzbach and Boyer (2011)).

Then define the augmented conversion matrix $\mathbf{J}_n \in M_{2n,2n}(\mathbb{C})$ to be

$$\mathbf{J}_n \coloneqq \frac{1}{2}\begin{pmatrix} \mathbf{I}_n & \mathbf{I}_n \\ -i\,\mathbf{I}_n & i\,\mathbf{I}_n \end{pmatrix},$$

where $\mathbf{I_n}$ is the $n \times n$ identity matrix (van den Bos (1995)). The conversion matrix $\mathbf{J}_n$ then satisfies the following three properties:

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \mathbf{J}_n \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{z}^* \end{pmatrix}; \tag{2.1.1}$$

$$\det(\mathbf{J}_n) = \left(\frac{i}{2}\right)^n; \tag{2.1.2}$$

$$\mathbf{J}_n^{-1} = 2\mathbf{J}^{\mathcal{H}}. \tag{2.1.3}$$

This conversion has been the foundation of much of complex statistics to date (Eriksson et al. (2010); Schreier and Scharf (2010); Adali and Schreier (2014)), as it allows simple translation of well-studied $2n$-variate real methods and coincides nicely with Wirtinger calculus (Wirtinger (1927)) which is explored in Section 2.4.

However, the definition of a single complex number as an ordered pair of points in $\mathbb{R}^2$ is not complete in the sense that it does not establish a field isomorphism between $\mathbb{C}$ and $\mathbb{R}^2$ due to the fact that we require the extension of $\mathbb{R}^2$ into a field by defining (usual) addition $+$ as

$$\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x+u \\ y+v \end{pmatrix},$$

and (*unusual*) multiplication $\times$ as

$$\begin{pmatrix} x \\ y \end{pmatrix} \times \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} xu - yv \\ xv + yu \end{pmatrix}.$$

Observing in $\mathbb{R}^2$, complex multiplication may seem somewhat arbitrary, however it has a useful geometric interpretation when considering the polar form of a complex number.

2.1.3 Polar form

The polar form of a complex number describes it in terms of the modulus (Euclidean length) and argument (angle of rotation) of the equivalent vector in $\mathbb{R}^2$. Specifically, the complex number $z = x + iy$ can also be written as

$$z = re^{i\theta} = r\cos(\theta) + i\,r\sin(\theta), \tag{2.1.4}$$

where $r = |z| = \sqrt{x^2 + y^2}$ and θ is the angle z makes with the positive real axis. The size of the corresponding acute angle for the argument may be found by taking $\tan^{-1}\left(\frac{y}{x}\right)$, however adjustments of adding or subtracting π may be required based off the signs of x and y. Furthermore, the angles are only unique up to a 2π ($360°$) rotation, hence one can rather think of the argument in terms of the equivalence

class of all equivalent angles modulo 2π. The second equality in (2.1.4) makes use of Euler's formula (Ahlfors (1979, p. 44)). Unless otherwise stated, we will take the principal argument $\text{Arg}(z)$ to be the corresponding angle in the interval $(-\pi, \pi]$. Figure 2.1 illustrates the different complex number representations.

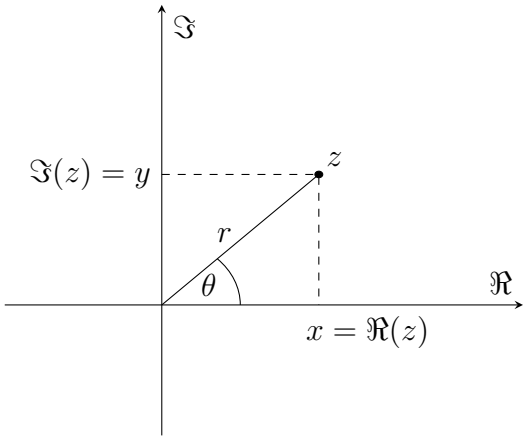

Figure 2.1: A plot of $z = x + iy = re^{i\theta}$

Now, considering complex operations, it is clear that for $z = re^{i\theta}$ and $w = se^{i\alpha}$,

$$\begin{aligned} z + w &= (r\cos(\theta) + s\cos(\alpha)) + i(r\sin(\theta) + s\sin(\alpha)) \\ &= \sqrt{r^2 + 2rs\cos(\theta - \alpha) + s^2}\, e^{i\tan^{-1}\left(\frac{r\sin(\theta)+s\sin(\alpha)}{r\cos(\theta)+s\cos(\alpha)}\right)} \\ &= \sqrt{r^2 + 2\Re(zw^*) + s^2}\, e^{i\tan^{-1}\left(\frac{r\sin(\theta)+s\sin(\alpha)}{r\cos(\theta)+s\cos(\alpha)}\right)}, \\ z \times w &= rse^{i(\theta+\alpha)}. \end{aligned}$$

Whilst trying to express the addition again in polar form can be cumbersome, we see that geometrically the complex multiplication represents a scaling of the two moduli and addition of the arguments. That is, scale the first number's length by that of the second and rotate the first anticlockwise by the second's argument. In $\mathbb{R}^2$, this is similar to how a rotation matrix may be used, and in fact there exists a *field* isomorphism between the complex numbers and a 2×2 real matrices.

2.1.4 2×2 *matrix representation*

The field, $\mathbb{C}$, of complex numbers is isomorphic to a subfield of $M_{22}(\mathbb{R})$ (Copson (1935)). Specifically, it is the field of all 2×2 real-valued matrices of the form

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \tag{2.1.5}$$

for $x, y \in \mathbb{R}$, which represents the complex number $z = x + iy$. We shall denote this subfield by $M_{\mathbb{C}}$.

Within $M_{\mathbb{C}}$ the usual matrix addition and multiplication (which commutes for such matrices) is equivalent to complex addition and multiplication. Furthermore, these matrices are invertible (except in the case of the zero matrix) and the inverse corresponds to the inverse of the equivalent complex number, the determinant of

such a matrix yields the square of the complex modulus, and the transpose gives the conjugate. Hence, it may be more intuitive to consider the isomorphism between $\mathbb{C}$ and $M_{\mathbb{C}}$ than that between $\mathbb{C}$ and $\mathbb{R}^2$, as the natural matrix algebra maps directly to that of $\mathbb{C}$.

Understanding complex numbers in $\mathbb{R}^2$ or $M_{\mathbb{C}}$ is often how computers store complex numbers. As will be explored in Section 2.5, however, floating point complex arithmetic may be improved in terms of accuracy and rounding errors beyond naïve real implementations (Jeannerod et al. (2017); Muller et al. (2018)).

2.2 On Powers and Principality

It is important here to address the non-uniqueness of arguments for complex numbers. Indeed, restricting the principal argument to the interval $(-\pi, \pi]$ results in a single argument for each complex number, however this should not be interpreted to induce an ordering of angles. That is, the statement $\pi - \varepsilon > -\pi + \varepsilon$ for some small $\varepsilon > 0$ is not valid for these two angles, which happen to be very close, either side of the negative real axis. In the language of complex analysis, this ray in the complex plane which is the boundary of the arbitrary interval chosen to restrict the argument to is called the branch cut (Ahlfors (1979); Brown and Churchill (2009)). Branch cuts cause discontinuities in functions which involve the argument in a form other than Euler's exponential or within trigonometric functions. However, in the context of complex statistics we will only explicitly restrict the argument in the context of integrals, so branch cuts will not affect our calculations.

2.2.1 Arg *or* arg . . . *Argh!*

In terms of calculating arguments, the notation throughout this thesis will be very precise. Should we wish to consider in general any angle, or rather, all possible values in the equivalence class of a given angle modulo 2π, we will use the function arg. On the other hand, the principal argument function Arg gives only the corresponding angle in the interval $(-\pi, \pi]$. There is, thus, only a single principal argument per complex number. If we wish to take a unique argument within a different interval of length 2π, we will denote the lower bound of that interval by a subscript, for example Arg_0 takes the corresponding angle in $(0, 2\pi]$. One can consider the more general function arg as representing the equivalence class of Arg modulo 2π, that is[2]

$$\arg(z) = \text{Arg}(z) + 2k\pi\,,\ \forall\, k \in \mathbb{Z}. \tag{2.2.1}$$

Operating on products, we have that for $z, w \in \mathbb{C}$,

$$\arg(zw) = \arg(z) + \arg(w), \tag{2.2.2}$$

in the sense that both sides give the same (infinitely many) values (Ahlfors (1979)). However, it may be that $\text{Arg}(zw) \neq \text{Arg}(z) + \text{Arg}(w)$, despite both sides here still

[2]Where we wish to denote a multivalued object, we will use $\forall$ in the expression, rather than saying 'for all' before or after, which may be confused for an identity.

representing the same angle modulo 2π. To convert an arbitrary angle $\alpha \in \mathbb{R}$ to its corresponding principal argument, one can take

$$(\alpha + \pi) \mathrm{mod}(2\pi) - \pi \in (-\pi, \pi]. \tag{2.2.3}$$

The main occurrence of the argument where branch cuts need to be handled with care is when considering the complex logarithm, and by extension, powers of complex numbers. This is important in the context of moments for complex random variables and the log-normal distribution. Importantly, the complex number 0 does not have a defined argument, and hence the argument and logarithm are not defined for zero, however we will employ the convention regarding powers that $0^s = 0$ for all $s \in \mathbb{C}$ such that $0 < |s| < \infty$.

2.2.2 The complex logarithm

The complex exponential is defined by the power series, or equivalently Euler's formula such that for $z = x + iy \in \mathbb{C}$, (Ahlfors (1979); Brown and Churchill (2009))

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^x e^{iy} = e^x(\cos(y) + i\sin(y)). \tag{2.2.4}$$

Similar to the real exponential, we have that $e^z e^w = e^{z+w}$. However this results in periodicity of the complex exponential in the imaginary component of z as $e^{i2\pi k} = 1$ for any integer k. So, for all $k \in \mathbb{Z}$,

$$e^z = e^{z+i2\pi k}. \tag{2.2.5}$$

Now turning to the logarithm as the inverse of the exponential, because the complex exponential is a many-to-one function, the complex logarithm must necessarily be one-to-many, or multivalued. That is, if $z = re^{i\theta}$ for some $r > 0$ and $\theta \in \mathbb{R}$,

$$\log(z) = \ln(r) + i\theta + i2\pi k\,,\ \forall\, k \in \mathbb{Z} \tag{2.2.6}$$

$$= \ln(|z|) + i\arg(z). \tag{2.2.7}$$

Some texts in fact define the arg function as above after first introducing the complex logarithm (e.g. Ahlfors (1979)). Importantly, we make a distinction between the multivalued complex logarithm, log, and the single-valued real logarithm ln.

Just as we can restrict the principal argument to obtain a single-valued function, the principal logarithm can also be developed and denoted

$$\mathrm{Log}(z) = \ln(|z|) + i\mathrm{Arg}(z), \tag{2.2.8}$$

which has image in a horizontal strip in the complex plane with imaginary part in $(-\pi, \pi]$ (Ahlfors (1979); Brown and Churchill (2009)). Similarly, if we wish to use a different unique argument e.g. $\mathrm{Arg}_0 \in (0, 2\pi]$, we will denote the corresponding logarithm with the same subscript e.g. Log_0.

As with the argument, whilst $\log(zw) = \log(z) + \log(w)$, we do not necessarily have that $\text{Log}(zw) = \text{Log}(z) + \text{Log}(w)$ as the imaginary parts may differ by 2π. However, it still holds that

$$e^{\text{Log}(zw)} = e^{\text{Log}(z)+\text{Log}(w)}. \tag{2.2.9}$$

The complex logarithm is important in the definition of powers of complex numbers, which are not as straightforward as for real numbers.

2.2.3 Complex powers

For some $x \in \mathbb{R}$, we can define powers of the form x^s for any $s \in \mathbb{C}$. Where s is an integer, x^s is interpreted as multiplying x by itself s times. Where s is not an integer, x^s can be understood by

$$x^s = e^{s \ln(x)}, \tag{2.2.10}$$

noting that the real logarithm, ln, is a single-valued function (coinciding with $\text{Log}(z)$ for $z = x + i0$). However, if we now consider powers of the form z^s for $z \in \mathbb{C}$, then this may be similarly defined by

$$z^s := e^{s \log(z)} \tag{2.2.11}$$

which is now a multivalued object by nature of the complex logarithm. Specifically,

$$z^s = e^{s\text{Log}(z)} e^{is2\pi k}, \ \forall\, k \in \mathbb{Z}. \tag{2.2.12}$$

We observe that, for s an integer, $e^{is2\pi k} = 1$ for all integers k and thus z^s is single-valued and interpreted as multiplying z by itself s times. Furthermore, if s is rational with reduced form $\frac{p}{q}$ for integers p and q, then z^s takes q distinct complex values (Ahlfors (1979)). For all other values of $s \in \mathbb{C}$, the complex power z^s may take countably many values. We will call the principal value that given by

$$e^{s\text{Log}(s)}. \tag{2.2.13}$$

2.3 Representing Real Data with Complex Numbers

Given that any single complex number can be thought of as a vector in $\mathbb{R}^2$, a natural question when statistically handling complex data is: why not treat complex data as bivariate real data? In short, the answer is that the complex multiplication, which is not natural in $\mathbb{R}^2$, is useful to develop linear functions that correspond to physically meaningful phenomena in terms of modulus scaling and argument shifts. Complex notation is also more interpretable and compact than handling random 2×2 matrices, despite the field isomorphism to $M_{\mathbb{C}}$. The rest of this section expands further on why the complex multiplication should be retained by manipulating complex data in $\mathbb{C}$.

We do not address the area of random polynomials, which rather obviously require complex numbers to describe all roots by the fundamental theorem of algebra. We note however that to factorise an n-degree polynomial $P_n(z)$, the factorisation relies upon the complex multiplication. In particular, $P_n(z) = \prod_{k=1}^{n}(z - \alpha_k)$ for

roots $\alpha_k \in \mathbb{C}$, $k = 1, \ldots, n$. Hence, it is not possible to replace complex numbers here by equivalent vectors in $\mathbb{R}^2$ as we cannot multiply them. Likewise, we cannot use 2×2 matrices and keep $z \in \mathbb{R}$ or $\mathbb{C}$ as the addition or multiplication no longer makes sense.

In other applications, it is the geometry of the complex multiplication that makes complex numbers useful representations of real-world observations. For example, when handling a signal such as a cosinusoidal waveform $x(t) = A\cos(\omega t)$, common features of interest to encode or understand information within the signal are amplitude changes ($BA\cos(\omega t)$) and phase shifts ($A\cos(\omega t + \phi)$), where $A, B, \omega, \phi \in \mathbb{R}$ and t can be thought of as a time parameter (usually positive real). Within $\mathbb{R}$, such amplitude and phase changes could be represented by a function $f_{B,\phi} : \mathbb{R} \to \mathbb{R}$,

$$f_{B,\phi}(x(t)) = Bx(t + \phi).$$

Of course, this is not a convenient function to work with, being non-linear in x due to the time-shift. On the other hand, consider the (artificially) complex signal $z(t) = e^{i\omega t}$, understanding that the true signal is the real part of this complex signal. Subsequently, by taking advantage of the polar form of complex numbers and the multiplication in the complex field, any arbitrary amplitude change or phase shift is easily represented by the linear function $g_{B,\phi} : \mathbb{C} \to \mathbb{C}$,

$$g_{B,\phi}(z(t)) = Be^{i\phi}z(t).$$

The real part of this complex linear function is the desired signal with the appropriate amplitude change and phase shift described by the real-valued function $f_{B,\phi}$. The linearity of $g_{B,\phi}$ then easily passes through the linear combination of a complex Fourier series to act upon any desired signal (Schreier and Scharf (2010, p. 12)). Similarly, when dealing with bivariate data mapped to complex values through the Cartesian representation of complex numbers, if rotation and scaling is physically meaningful e.g. for wind data, then again the geometry of the complex algebra facilitates representing these operations by a linear function in the complex plane, rather than left multiplication by a rotation matrix in $\mathbb{R}^2$.

Example 2.3.1 (QPSK)**.** *Consider playing a simple song with four notes (e.g. Mary had a Little Lamb) over the radio on a fixed frequency ω. Then, we need to be able to encode four bits of information, each corresponding to a single musical note (C,D,E,G). For transmission, if the frequency of the wave must remain fixed and the amplitude controls the volume, then the remaining method to encode information is in the phase of the wave. As such, we can choose to transmit a combination of four different signals with respect to time t, corresponding to $\cos(\omega t), -\sin(\omega t), -\cos(\omega t), \sin(\omega t)$. Mathematically, where the frequency f is understood from the radio transmission frequency, then these four signals can be represented by four equispaced complex numbers $1, i, -1, -i$, where the actual signal is given by $\Re(ze^{\omega t})$ for $z \in \{1, i, -1, -i\}$. This is called Quadrature Phase Shift Keying (QPSK) (Schreier and Scharf (2010); Adali and Schreier (2014)). The transmitted signals may be represented by four equispaced transmission nodes in the complex plane as in Figure 2.2 below.*

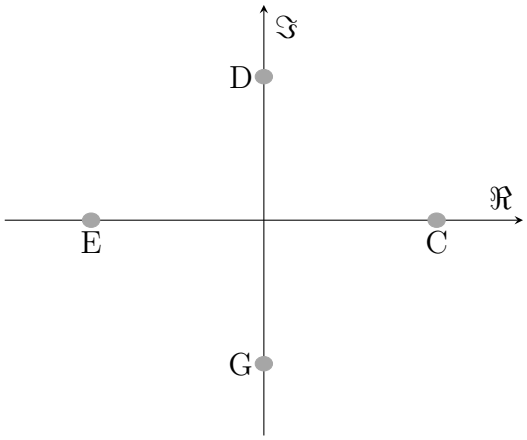

Figure 2.2: The four QPSK nodes.

Thus, a transmitter transmits Mary had a Little Lamb via a signal which over time looks like Figure 2.3.

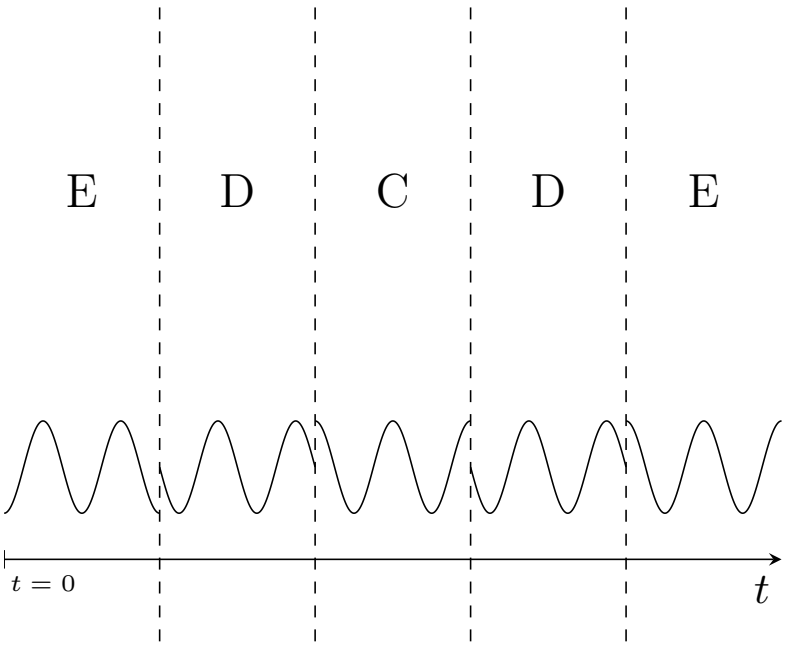

Figure 2.3: A QPSK signal of Mary had a Little Lamb (first five beats).

However, what the receiver actually receives is a signal which may be distorted by noise during transmission. This distorted signal can be represented by a point in the complex plane giving the modulus (amplitude) and argument (phase) of the received signal at a regular time interval (e.g. each beat of the song). Thus plotting the received data, the transmitted song can be represented on the complex plane in Figure 2.4.

The song data may therefore be analysed as a complex time series. In practice, the receiver may just take the nearest node to the received data point to decode which musical note was transmitted at that time in order to play the song. However, what

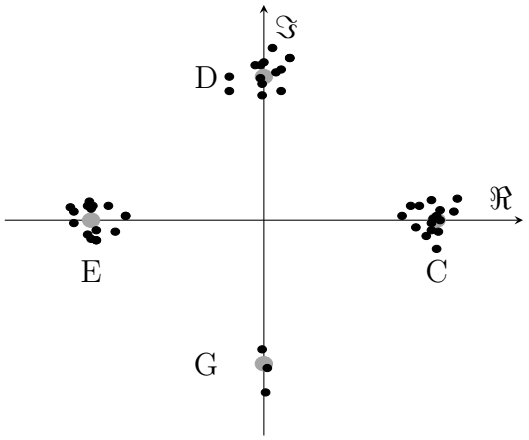

Figure 2.4: Mary had a Little Lamb, played on the complex plane.

if more than four notes needed to be transmitted? Then there needs to be more nodes, and necessarily less space between each node, thus increasing the need for reliable modelling and statistical estimation to ensure coherent signal transmission when random noise is present.

The above example is a rather simple example to convey the usefulness of representing waveform signals as complex numbers. In the present day, signal processing is further advanced to encode information not only in the phase shift, but also in the modulus (allowing the transmission nodes to be distributed on a grid in the complex plane, rather than crammed around the unit circle). A reader further interested in the exact applications in signal processing is referred to the very comprehensive book by Schreier and Scharf (2010) (further motivation on this topic may be found in their Section 1.4) which treats complex random variables in the particular application of signal processing. In other applications, such as with fMRI data, a complex time series is recorded over a cubic centimetre grid (voxels) of the brain, providing both spatial and temporal complex data. Here we will stick to a general theoretical approach without working on a specific data set.

In application where experiments record real time series data and wish to model an underlying oscillating phenomenon (e.g. electromagnetic wave), complex data often arises in the Fourier (or frequency) domain via the discrete time Fourier transform. In this form, the modulus and argument of the complex data describe the relative prominence of particular frequencies and phases of the underlying oscillating wave that generated the real observations. The discrete Fourier transform of a set of real values $\{X_1, \ldots, X_n\}$ is given by the values $\{Y_1, \ldots, Y_n\}$ where

$$Y_k = \sum_{l=1}^{n} e^{i\frac{2\pi}{n}(k-1)(l-1)} X_l \,, \; k = 1, \ldots, n.$$

Application specific context then lends further meaning to the complex values. For complex numbers returned after a brain scan by an fMRI machine, the modulus squared indicates the firing rate of a neuron and the argument gives the relative phase of its activation, i.e. not only how regularly it fires, but whether it is in

sync with other neurons (Rowe and Logan (2004); Adrian et al. (2018)). In magnetotelluric methods, the modulus represents the apparent resistivity of the Earth's crust and the phase indicates the relative phases of the orthogonal components of the electric and magnetic fields (Cagniard (1953); Caldwell et al. (2004)). Again, the polar components of the complex data $\{Y_1, \dots Y_n\}$ are of more physical interest than the Cartesian form of the data. Yet, current statistical theory tends to revolve around analysing data via their Cartesian components in additive models or simply extracting the moduli and ignoring phase data (Rowe and Logan (2004); Adrian et al. (2018)), with no investigation to our knowledge of complex multiplicative models. This provides a further motivation to establish a self-contained theory in $\mathbb{C}$, because complex multiplicative models — which act directly upon the modulus and argument — do not have analogues in $\mathbb{R}^2$. We will not reach theory of multiplicative models in this thesis, as too much groundwork is required, however we acknowledge this as an unexplored area of future research.

Under current statistical theory for complex data, the ability to represent phase shifts and amplitude changes of waves with complex linear functions is particularly useful when dealing with normal (Gaussian) distributions. The normal distribution (the family of which is closed under linear combinations) often arises in statistical estimates due to the complex central limit theorem (see Section 6.3.4).

2.4 Differentiation and Wirtinger Calculus

When dealing with complex-valued random variables and functions thereof e.g. distribution functions and likelihoods, differentiation is an important tool for analysis. In real-valued statistics, differentiation — both analytic and numerical — forms the basis of many common methods such as maximum likelihood or least squares estimation. Differentiation is also often used to train machine learning algorithms, such as in Chapter 2 of Hirose (2013), or Sob et al. (2020). To handle differentiation with complex values, this section establishes the rudimentary notions of Wirtinger calculus (a.k.a. $\mathbb{CR}$-calculus) which is used as a relaxed framework for differentiating functions which are not necessarily holomorphic, i.e., differentiable under the standard notion of complex differentiation ($\mathbb{C}$-differentiation). In the following, it is understood that a complex function $f(z)$ can also be written in terms of real variables x and y where $z = x + iy$, and hence f may also be partially differentiated with respect to x and y as much as z.

2.4.1 Complex differentiation

For a function $f : \mathbb{C} \to \mathbb{C}$ of a complex variable, we define its limit approaching a point a as follows.

Definition 2.4.1 (Ahlfors (1979), p.22)**.** *A function $f(z)$ has limit L as z tends to a, denoted*

$$\lim_{z \to a} f(z) = L, \tag{2.4.1}$$

if and only if for every real $\varepsilon > 0$, there exists a real $\delta > 0$ such that for all complex numbers $z \neq a$ satisfying $|z - a| < \delta$, then $|f(z) - L| < \varepsilon$.

With the definition of a limit established, we can now provide the definition of the derivative of a complex function.

Definition 2.4.2 (Ahlfors (1979), p.23)**.** *The derivative of the complex function $f : \mathbb{C} \to \mathbb{C}$ is defined as the limit (if it exists) of*

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h}. \tag{2.4.2}$$

The derivative (if it exists) at the point a will be denoted $f'(a)$.

In the above definition, for the derivative to exist, the limit in (2.4.2) must exist and be the same as h approaches 0 from any direction (Adali and Schreier (2014)). A function which is $\mathbb{C}$-differentiable on a set will be called holomorphic on that set.

Defining $\mathbb{C}$-differentiation as in Definition 2.4.2 gives rise to the Cauchy-Riemann differential equations as necessary conditions for a function of a complex variable to be $\mathbb{C}$-differentiable (Ahlfors (1979, p.25)). For a function $f : \mathbb{C} \to \mathbb{C}$ which may be decomposed into $f(x + iy) = u(x, y) + iv(x, y)$ for $u, v : \mathbb{R}^2 \to \mathbb{R}$, the Cauchy-Riemann equations for f at the point $w = a + ib$ are:

$$\begin{aligned} \frac{\partial u}{\partial x}(a, b) &= \frac{\partial v}{\partial y}(a, b); \\ \frac{\partial u}{\partial y}(a, b) &= -\frac{\partial v}{\partial x}(a, b). \end{aligned} \tag{2.4.3}$$

These equations are also a sufficient condition for $\mathbb{C}$-differentiability on a set where the first-order partial derivatives of the bivariate real-valued functions u and v are continuous and satisfy (2.4.3) (Ahlfors (1979, p. 26)).

One important consequence of the necessity of the Cauchy-Riemann equations is that a real-valued function of a complex variable ($v \equiv 0$) must either be a constant or non-differentiable (Ahlfors (1979)). This imposes a strict restriction upon functions of complex random variables which one may desire to take (non-constant) real values such as densities or likelihoods, for interpretability, whilst retaining the useful calculus of differentiability e.g. for maximisation. Below we present Wirtinger (a.k.a. $\mathbb{CR}$-) calculus which provides a more flexible differentiation framework.

2.4.2 Wirtinger calculus

Wirtinger calculus establishes a framework for differentiation in an $\mathbb{R}$-linear sense, that is one treats z and z^* as independent variables when differentiating. In particular, the Wirtinger derivatives below in (2.4.4)–(2.4.8), apply to any complex function which would be real-differentiable if expressed as a function in $\mathbb{R}^2$ (Eriksson et al. (2010); Schreier and Scharf (2010, Appendix 2); Adali and Schreier (2014)). That is, the Wirtinger derivatives of $f = u + iv$ exist if and only if u and v are themselves continuously differentiable in $x = \Re(z)$ and $y = \Im(z)$. A function whose Wirtinger derivatives exist is called $\mathbb{R}$-differentiable (Eriksson et al. (2010)). The derivation from Wirtinger (1927) (similar derivations are given in Ahlfors (1979, p. 27) and Remmert (1991, p. 65)) follows by using the fact that $x = \frac{z+z^*}{2}$ and $y = \frac{z-z^*}{2i}$ and applying a change of variable "forgetting" z and z^* are conjugates of each other.

The Wirtinger derivatives of f for $z = x + iy$ are defined as the following partial derivatives:

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right); \tag{2.4.4}$$

$$\frac{\partial f}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right). \tag{2.4.5}$$

To demonstrate, we apply these derivatives to a function that is not $\mathbb{C}$-differentiable, except at 0.

Example 2.4.3. *Let $f(z) = |z|^2 = zz^* = x^2 + y^2$. Taking the Wirtinger derivatives as defined in (2.4.4) and (2.4.5) clearly yields, for all $z \in \mathbb{C}$,*

$$\frac{\partial f}{\partial z} = \frac{1}{2}(2x - 2iy) = z^*,$$
$$\frac{\partial f}{\partial z^*} = \frac{1}{2}(2x + 2iy) = z.$$

The above results are what would be expected by linearly differentiating $f(z) = zz^*$ treating z and z^* as independent variables. In fact, these derivatives employ exactly that behaviour on all functions of z and z^*, proven by Branwood in 1983 (see ref [7] in Adali and Schreier (2014)) and verified in Remmert (1991). Furthermore, $\frac{\partial f}{\partial z^*} = \left(\frac{\partial f}{\partial z}\right)^*$ if and only if f is real-valued (up to a possibly complex constant). Whilst straightforward, a formal proof is presented in Appendix A.1.

Important for the development of integration later, one can consider the notion of Wirtinger differentiability as the sort of differentiation that is induced by treating $\mathbb{C}$ as a two-dimensional real manifold. That is, as a topological space, we can parameterise $\mathbb{C}$ in terms of the coordinates[3] $(z, z^*)^\top$ which are mapped to the point $(x, y)^\top \in \mathbb{R}^2$. This association is what underlies the use of the augmented vector notation and conversion matrices $\mathbf{J}_n$ in most other recent texts on complex random variables (Eriksson et al. (2010); Schreier and Scharf (2010); Adali and Schreier (2014); Ducharme et al. (2016)). In the language of differential geometry, the smooth chart φ between $\mathbb{C}$ and $\mathbb{R}^2$ is given by

$$\varphi : \mathbb{C} \to \mathbb{R}^2 : \begin{pmatrix} z \\ z^* \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2.4.6}$$

Hence a function $f : \mathbb{C} \to \mathbb{C}$ is differentiable if and only if $\varphi^{-1} f \varphi$ is a differentiable function from $\mathbb{R}^2$ to $\mathbb{R}^2$, which amounts to differentiability in terms of x, y, which is exactly $\mathbb{R}$-differentiability. We remark that a similar idea is given in Chapter 2 of Hirose (2013) when setting out neural network learning in complex manifolds.

Although the Wirtinger derivatives are useful as a workaround to the stringency of $\mathbb{C}$-differentiation for real-valued functions of complex variables, one must keep in mind that we are essentially extending differentiation in $\mathbb{R}^2$ to $\mathbb{C}$ via a change

[3]The parameterisation $(z, z^*)^\top$ is obviously degenerate (in the statistical sense of the word) as if z is known, we immediately know z^*. However this is what is meant such as in Ahlfors (1979, p. 27) by "forgetting that they are in fact conjugate".

of notation — so in using this differentiation framework, the powerful tools of complex analysis developed for holomorphic functions may not apply (Picinbono (1994)). Nonetheless, these derivatives will prove useful at providing a framework to the extent that differentiation is used in much of statistics.

An aspect that has been missing from the literature which may be useful in speeding up calculations in certain contexts is a polar form of these Wirtinger derivatives. Employing a change of variables, one can obtain the explicit form of the derivatives of f as if it were a function of r and θ, where $z = re^{i\theta}$.

Proposition 2.4.4. *Where we can write $z = re^{i\theta}$ such that a function $f(z) = f(re^{i\theta})$, then the Wirtinger derivatives are given by*

$$\frac{\partial f}{\partial z} = \frac{e^{-i\theta}}{2}\left(\frac{\partial f}{\partial r} - \frac{i}{r}\frac{\partial f}{\partial \theta}\right); \tag{2.4.7}$$

$$\frac{\partial f}{\partial z^*} = \frac{e^{i\theta}}{2}\left(\frac{\partial f}{\partial r} + \frac{i}{r}\frac{\partial f}{\partial \theta}\right). \tag{2.4.8}$$

Derivation of the Cartesian and polar Wirtinger derivatives is provided in Appendix A.2.

Whilst leading to the same answer as the Cartesian Wirtinger derivatives, the polar form may present computational efficiencies, especially in circumstances involving moduli of complex numbers, which is the main non-analytic function that appears in probability distributions, particularly circular or elliptically symmetric distributions. Applying the polar Wirtinger derivatives to the function from Example 2.4.3 in Example 2.4.5 below suggests that computations would be faster, given only one partial derivative needs to be calculated either by hand, or in a computational context of automatic differentiation. Further research is required to ascertain the exact (if any) improvement in computation.

Example 2.4.5. *Let $f(z) = |z|^2 = zz^* = r^2$. Applying (2.4.7) and (2.4.8) yields, for all $z \in \mathbb{C}$,*

$$\frac{\partial f}{\partial z} = \frac{e^{-i\theta}}{2}(2r - 0) = z^*,$$
$$\frac{\partial f}{\partial z^*} = \frac{e^{i\theta}}{2}(2r - 0) = z.$$

Both forms of the Wirtinger derivatives result in treating z and z^* as independent variables and differentiating as one would a real variable. It is worth noting that a $\mathbb{C}$-differentiable function f is simply an $\mathbb{R}$-differentiable function that also satisfies the Cauchy-Riemann differential equations which can now be succinctly written as the condition $\frac{\partial f}{\partial z^*} = 0$ (Ahlfors (1979, p. 74); Remmert (1991)). If f is $\mathbb{C}$-differentiable, the Wirtinger derivative $\frac{\partial f}{\partial z}$ coincides with the $\mathbb{C}$-derivative.

Higher dimensional derivatives behave as expected in line with multidimensional real calculus (Trampitsch (2013)), where for $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y} \in \mathbb{C}^n$, one can define

$\nabla_{\boldsymbol{z}} = (\frac{\partial}{\partial z_1}, \ldots, \frac{\partial}{\partial z_n})$ and for $f : \mathbb{C}^n \to \mathbb{C}$,

$$\nabla_{\boldsymbol{z}} f = \frac{1}{2} \left(\nabla_{\boldsymbol{x}} f - i \nabla_{\boldsymbol{y}} f \right),$$
$$\nabla_{\boldsymbol{z}^*} f = \frac{1}{2} \left(\nabla_{\boldsymbol{x}} f + i \nabla_{\boldsymbol{y}} f \right).$$

For the case where $\boldsymbol{f} : \mathbb{C}^n \to \mathbb{C}^m$, the Jacobian and conjugate Jacobian matrices are constructed as in the real case, using the usual partial derivative notation but with respect to z and z^* respectively, understanding that Wirtinger derivatives are used (Trampitsch (2013, p. 9)).

Equipped with a generalised framework for differentiation, we now generalise the notion of the Taylor series expansion of a function around a point.

2.4.3 Wirtinger Taylor expansion

For use later — such as in proving the complex central limit theorem — the Taylor series expansion and approximation of an $\mathbb{R}$-differentiable function $f : \mathbb{C} \to \mathbb{C}$, which may not necessarily be $\mathbb{C}$-differentiable, can be given and is similar to what may be expected from the real bivariate Taylor expansion. In fact, it is derived by linear manipulation of the bivariate Taylor expansion, taking advantage of the additive group isomorphism between $\mathbb{C}$ and $\mathbb{R}^2$ which also have the same differential structure under Wirtinger calculus.

Univariate

Below we first present the univariate Taylor expansion which can be found without proof as Theorem 3.3 Eriksson et al. (2010, p. 5402). We will prove this formula in Appendix A.3, however first we need to establish some notation. Let $C^k_{\mathbb{R}}$ denote the set of complex functions that have continuous mixed Wirtinger derivatives of all orders up to and including k. Also, let the differential operators ∇_x and $\nabla_{\underline{z}}$ be given by (understanding that $z = x + iy$)

$$\nabla_x = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix} \quad \text{and} \quad \nabla_{\underline{z}} = \begin{pmatrix} \frac{\partial}{\partial z} \\ \frac{\partial}{\partial z^*} \end{pmatrix}.$$

Finally, whilst many sources use multinomial notation, we will persist here with matrix-vector notation, keeping in mind that for $\boldsymbol{z} = (z_1, z_2)^\top$, continuous partial derivatives that commute and $k \geq 2$,

$$\left[\boldsymbol{z}^\top \nabla_x\right]^k = z_1^k \frac{\partial^k}{\partial x^k} + k z_1^{k-1} z_2 \frac{\partial^k}{\partial x^{k-1} \partial y} + \binom{k}{2} z_1^{k-2} z_2^2 \frac{\partial^k}{\partial x^{k-2} \partial y^2} + \cdots + z_2^k \frac{\partial^k}{\partial y^k}.$$

We are now ready to state the Wirtinger Taylor expansion of a complex-valued function. The proof of the expansion is given in Appendix A.3

Theorem 2.4.6 (Univariate Wirtinger Taylor expansion)**.** *A function $f : \mathbb{C} \to \mathbb{C}$ that is $C^{n+1}_{\mathbb{R}}$ with all derivatives bounded in modulus on an open set $U \subseteq \mathbb{C}$ has*

n^{th}-order Wirtinger Taylor expansion about the point $\omega \in U$ where, for all complex h such that $\omega + h \in U$,

$$f(\omega + h) = f(\omega) + \sum_{k=1}^{n} \frac{1}{k!} \left[\underline{\boldsymbol{h}}^\top \nabla_{\underline{z}}\right]^k f(\omega) + o(|h|^n). \tag{2.4.9}$$

Above, $\underline{\boldsymbol{h}} = (h, h^*)^\top$.

To clarify what is going on within the summation notation and for later use, we explicitly present an example of the second order Wirtinger Taylor expansion of a $C^3_{\mathbb{R}}$ function f about ω.

Example 2.4.7. *The second order Wirtinger Taylor expansion of a function $f : \mathbb{C} \to \mathbb{C}$ about the point ω is given by*

$$\begin{aligned} f(\omega + h) = f(\omega) &+ h\frac{\partial f}{\partial z}(\omega) + h^*\frac{\partial f}{\partial z^*}(\omega) \\ &+ \frac{1}{2}h^2\frac{\partial^2 f}{\partial z^2}(\omega) + |h|^2\frac{\partial^2 f}{\partial z \partial z^*}(\omega) + \frac{1}{2}h^{*2}\frac{\partial^2 f}{\partial z^{*2}}(\omega) + o(|h|^2). \end{aligned} \tag{2.4.10}$$

As would be expected, in the case of f being holomorphic, $\frac{\partial f}{\partial z^*} = 0$ and the Taylor expansion in (2.4.6) coincides with the complex Taylor expansion for holomorphic functions. A smooth ($C^\infty_{\mathbb{R}}$) function whose n^{th}-order Taylor expansion converges to the function itself for n tending to ∞ will from hereon be called analytic. Should a distinction be necessary from the formal definition of analytic for holomorphic functions, we will denote the formal version $\mathbb{C}$-analytic and the more general Wirtinger version $\mathbb{R}$-analytic. However as most applications will involve non-holomorphic functions, $\mathbb{R}$-analytic is assumed in the absence of qualification.

Multivariate

The Wirtinger Taylor expansion of a function can be extended to higher dimensional functions $f : \mathbb{C}^m \to \mathbb{C}$ as one would expect. The exact multivariate Wirtinger Taylor expansion is given in Appendix A.4, along with a second-order example. The multivariate version may be of particular interest in terms of optimisation search algorithms e.g. over multivariate complex density functions to numerically find maximum likelihood estimates. An application of complex gradient descent algorithms can be found in detail in Adali and Schreier (2014).

2.4.4 Maximisation and minimisation

To maximise or minimise a real-valued function of a complex variable, the Wirtinger derivatives can be used to find the critical points where $\frac{\partial f}{\partial z} = 0$, which by property of f being real will imply that $\frac{\partial f}{\partial z^*}$ is also 0. Subsequently, classification as a local maximum or minimum, or neither, can be attained by using bivariate methods on x, y (refer to e.g. "Second Derivatives Test" Stewart (2005, p.803)). Expressed

in terms of the Wirtinger derivatives, using compact derivative notation where $\partial_{zz} = \frac{\partial^2}{\partial z^2}$ and $\partial_{zz^*} = \frac{\partial^2}{\partial z \partial z^*}$, if the quantity

$$(\partial_{zz^*} f)^2 - |\partial_{zz} f|^2 \tag{2.4.11}$$

is positive, then the critical point is a minimum (resp. maximum) if $\partial_{zz^*} f + \Re(\partial_{zz} f)$ is positive (resp. negative). If the quantity in (2.4.11) is negative, then it is a saddle point. Where the quantity is zero, there may still be a maximum or minimum or saddle point, it is just not detected at the second order derivatives.

It is worth noting that in circular or elliptically symmetric probability distributions, the quantity $|z|^2$ arises frequently, in which case univariate optimisation may be performed on the variable $r \geq 0$.

In regards to gradient descent optimisation for multivariate functions, Adali and Schreier (2014, p.115) show that the maximum change in the function value should be taken in direction $\nabla_{z^*} f = (\partial_{z_1^*} f, \ldots, \partial_{z_n^*} f)^\top$.

2.5 Computation with Complex Numbers

As discussed in Section 1.1, where complex data arise, statistical analysis should remain in $\mathbb{C}$ in order to take full advantage of the field structure of $\mathbb{C}$ — in particular the complex multiplication. But a computer often stores complex numbers as a pair of real numbers, usually the Cartesian components. So it may seem that computation with complex numbers is simply computation with real numbers, understanding there to be an extra multiplication operation between vectors in $\mathbb{R}^2$. However, Jeannerod et al. (2017), Muller et al. (2018, Chapter 11) and Lefèvre and Muller (2019) demonstrate that naïve incorporation of the complex multiplication in floating point arithmetic may lead to large componentwise errors in the Cartesian components. Specifically, if one were to define a function which for $z = x + iy$ and $w = u + iv$ computes directly

$$z \times w = x \times u - y \times v + i(x \times v + y \times u),$$

then floating point arithmetic rounds at each calculation such that we end up with

$$zw = RN(RN(x \times u) - RN(y \times v)) + iRN(RN(x \times v) + RN(y \times u)),$$

where RN denotes the rounding operation (Lefèvre and Muller (2019)). To understand how rounding affects computation, a simple demonstration is in `R` or Matlab, if we store $a = 10^{16}$, $b = -10^{16}$ and $c = 1$ then both software will compute

$$a + b + c = 1, \qquad \text{(which is correct)}$$
$$a + c + b = 0. \qquad \text{(which is incorrect)}$$

This occurs because in the second instance, it performs $RN(RN(a + c) + b) = RN(a + b) = 0$. A more detailed explanation may be found in Muller et al. (2018) in terms of the radix and precision of systems. In essence, this suggests that in large computations, such as in neural networks, small rounding errors may propagate and accumulate through many layers. As such, it is desirable when employing complex

arithmetic to ensure that algorithms, in particular those governing complex multiplication and division, are established to minimise error bounds on computation.

Initial improvements on naïve manual implementations of the complex multiplication as given above are possible where a fused multiply-add (FMA) instruction is available which allows a multiplication and addition to be performed by only rounding once instead of twice. Namely, for $u, v, x \in \mathbb{R}$ one can compute $uv + x$ as

$$RN(u \times v + x),$$

without rounding $u \times v$. Further details on this behaviour are given in Muller et al. (2018) as the particulars are not the focus here, rather we remark that such an instruction can exist.

To demonstrate how complex multiplication can be adjusted to avoid excessive rounding errors, we present here Algorithm 3 from Lefèvre and Muller (2019) as Algorithm 1 below which multiplies of $z = x + iy$ and $w = u + iv$. We note that Algorithm 1 relies upon the availability of an FMA instruction and two common floating-point algorithms: 2Sum and Fast2Mult, which are given in Appendix A.5. In particular, the algorithm above works by breaking w up into $w = w_h + w_\ell$ (h for high and ℓ for low), and computing $zw_h + zw_\ell$ separately whereby certain rounding calculations for the low part are performed with Fast2Mult and 2Sum which approximates rounding errors to add back into computations.

Algorithm 1 Performing complex multiplication of $z = x + iy$ and $w = u + iv$ (Lefèvre and Muller (2019)).

Require: $z = x + iy$ floating point and $w = (u_h + u_\ell) + i(v_h + v_\ell)$ for $u_{h,\ell}, v_{h,\ell}$ double-precision (or double-word) numbers
Ensure: $|u_\ell| \leq 2^{-53}|u|$ and $|v_\ell| \leq 2^{-53}|v|$
  $t \leftarrow RN(v_\ell y)$
  $\pi_\ell \leftarrow RN(u_\ell x - t)$ ▷ Using FMA, we now have $\Re(w_\ell z)$
  $(P_h, P_\ell) \leftarrow \text{Fast2Mult}(v_h, y)$
  $r_\ell \leftarrow RN(\pi_\ell - P_\ell)$
  $(Q_h, Q_\ell) \leftarrow \text{Fast2Mult}(u_h, x)$
  $s \leftarrow RN(Q_\ell + r_\ell)$
  $(m_h, m_\ell) \leftarrow \text{2Sum}(Q_h, -P_h)$
  $\gamma_\ell \leftarrow RN(m_\ell + s)$
  return $\Re(zw) = RN(m_h + \gamma_\ell)$ ▷ The real part
  $t \leftarrow RN(v_\ell x)$
  $\pi_\ell \leftarrow RN(u_\ell y + t)$ ▷ Using FMA, we now have $\Im(w_\ell z)$
  $(P_h, P_\ell) \leftarrow \text{Fast2Mult}(v_h, x)$
  $r_\ell \leftarrow RN(\pi_\ell + P_\ell)$
  $(Q_h, Q_\ell) \leftarrow \text{Fast2Mult}(u_h, y)$
  $s \leftarrow RN(Q_\ell + r_\ell)$
  $(m_h, m_\ell) \leftarrow \text{2Sum}(Q_h, P_h)$
  $\gamma_\ell \leftarrow RN(m_\ell + s)$
  return $\Im(zw) = RN(m_h + \gamma_\ell)$ ▷ The imaginary part
  return $zw = \Re(zw) + i\Im(zw)$

In Algorithm 1, we have adjusted the notation slightly from Lefèvre and Muller (2019) and also reused variables between computation of the real and imaginary components of the multiplication. As per Lefèvre and Muller (2019), one can run the real and imaginary computations in parallel by not overwriting the same variables used in the real computation during the imaginary computation. We also assume here that the precision of the floating point system is 53 (as it is for `R` and Matlab).

Overall, we emphasise that Algorithm 1, which attains a more accurate computation of a complex multiplication is not just a naïve implementation of complex multiplication in $\mathbb{R}^2$. Likewise following Muller et al. (2018, Chapter 11), once an accurate and efficient complex multiplication procedure is in place, similar adjustments and implementations may be used to improve computation of complex division. We note that from experimental tests run in Lefèvre and Muller (2019), running the above algorithm can be 1.5–2 times slower relative to a naïve complex multiplication in $\mathbb{R}^2$. Thus, in the context of machine learning a trade-off between accuracy and speed of computation needs to be considered.

Nonetheless, where complex arithmetic is available (and one would hope optimised), we have seen that there are not only algebraic reasons, but also significant computational reasons for containing statistical procedures in $\mathbb{C}$ rather than $\mathbb{R}^2$.

2.6 Complex Integration

Finally, we briefly touch upon the notion of integration, which is crucial in probability and statistics in regards to moment calculation. Wirtinger antiderivatives can be analytically found by usual methods treating z and z^* as independent variables. However, powerful tools of complex analysis such as the residue theorem (Ahlfors (1979, p. 148)) are often non-applicable through the issue of forcing functions to be non-holomorphic by making them real-valued for interpretability — which was what motivated the application of Wirtinger calculus to complex statistics in the first place. Furthermore, when considering the distribution of a complex random variable, often the interest is in the variables taking value in regions of the complex plane, rather than specifically on contours, reducing the usefulness of the bulk of complex analysis theory developed for contour integration.

To define integration over regions of $\mathbb{C}$, we recall how the Wirtinger calculus is what is induced if we treat $\mathbb{C}$ as a two-dimensional real manifold, parameterised by $(z, z^*)^\top$. Furthermore, we introduce an inner product (metric) on $\mathbb{C}$, being the real scalar inner product such that

$$\left\langle \begin{pmatrix} z \\ z^* \end{pmatrix}, \begin{pmatrix} w \\ w^* \end{pmatrix} \right\rangle = \frac{1}{2}(zw^* + z^*w) = \Re(zw^*), \tag{2.6.1}$$

which reflects the usual complex inner product in each component whilst the factor of $\frac{1}{2}$ preserves the fact that

$$\left\langle \begin{pmatrix} z \\ z^* \end{pmatrix}, \begin{pmatrix} z \\ z^* \end{pmatrix} \right\rangle = |z|^2. \tag{2.6.2}$$

Equipped with an inner product, we now have that $\mathbb{C}$ is a Riemannian manifold, upon which we can thus define integration. The details are given in Appendix A.6,

however the outcome is that by treating $\mathbb{C}$ as the above defined Riemannian manifold, integration over a complex region is *defined* by integration with respect to the real and imaginary component variables. That is, the integral of some function $f : \mathbb{C} \to \mathbb{C}$ over the region $A_z \subseteq \mathbb{C}$ is given by

$$\iint_{A_{x,y}} f(x+iy)\,\mathrm{d}x\,\mathrm{d}y, \tag{2.6.3}$$

where $A_{x,y}$ is the corresponding region in $\mathbb{R}^2$. This is often why almost all texts on complex random variables choose at the outset to identify the expectation integral with the corresponding real bivariate integral, although without this underlying justification (van den Bos (1995); Eriksson et al. (2010); Schreier and Scharf (2010); Ollila et al. (2011); Adali and Schreier (2014); Ducharme et al. (2016)).

Further engaging notions of differential geometry, we may write the integral (2.6.3) in terms of $\mathrm{d}z \wedge \mathrm{d}z^*$ as follows (derivations also in Appendix A.6),

$$\iint_{A_{x,y}} f(x+iy)\,\mathrm{d}x\,\mathrm{d}y = \frac{i}{2}\iint_{A_z} f(z)\,\mathrm{d}z \wedge \mathrm{d}z^*, \tag{2.6.4}$$

where we employ the wedge product, assuming the standard orientation on $\mathbb{C}$ i.e. anticlockwise is from the positive real axis to the positive imaginary axis so that $\mathrm{d}x\,\mathrm{d}y = \mathrm{d}x \wedge \mathrm{d}y$ (elaboration provided in the derivation in Appendix A.6). Equation (2.6.4) also appears in Hörmander (1990) to work with functions of several complex variables, and in Olhede (2006) in a discussion on probability distributions of complex random variables. In Olhede (2006), the author explicitly writes functions of complex variables, such as f above, as $f(z, z^*)$ to make explicit the parameterisation of $\mathbb{C}$ by z and its conjugate, which we have seen induces Wirtinger calculus and the above association of integration. However, we remark that the integral with respect to $\mathrm{d}z \wedge \mathrm{d}z^*$ is understood by the corresponding integral in x and y.

Nonetheless, as reported in Olhede (2006), the use of complex differentials in the integral permits the application of the complex Stokes' formula from Hörmander (1990, p. 2) to turn a region integral into a contour integral.

Proposition 2.6.1 (Stokes' formula, Hörmander (1990))**.** *A function $f : \mathbb{C} \to \mathbb{C}$ integrated over the connected region $A \subseteq \mathbb{C}$, with simple, closed boundary ∂A (a finite number of C^1 Jordan curves) can be reduced to a contour integral around the boundary whereby*

$$\int_A f(z)\,\mathrm{d}z \wedge \mathrm{d}z^* = -\oint_{\partial A} F(z)\,\mathrm{d}z,$$

for F satisfying $\frac{\partial F}{\partial z^*}(z) = f(z)$ for all $z \in A \cup \partial A$, and ∂A oriented so that A lies to its left.

The proof is given in Appendix A.7. We recognise that familiar methods of complex analysis designed for holomorphic functions (e.g. the residue theorem) may not be applied to F on a domain where $f(z) \neq 0$, because $\frac{\partial F}{\partial z^*}(z) = f(z) \neq 0$,

which does not satisfy the Cauchy-Riemann equations. However, this will give rise to a novel form of cumulative distribution function discussed in Chapter 4. This is an important difference that is only recognisable by working entirely within $\mathbb{C}$, to employ Stokes' formula for complex variables. Although we have relaxed the differential structure of $\mathbb{C}$ to mimic $\mathbb{R}^2$, there are still familiar results to be obtained by working with complex notation under this new differential structure, rather than solving problems directly in $\mathbb{R}^2$.

Chapter 3 will now explore the notions of complex probabilities, and whether it makes sense or provides more mathematical agility to allow density functions to take complex values, rather than redefining the differential structure on $\mathbb{C}$. Such a construction may allow them be holomorphic and hence take advantage of results from complex analysis. Current real-valued interpretations of probability may then be considered as a real-valued function (e.g. $|\cdot|^2$ or $\Re(\cdot)$) composed with a holomorphic complex-density. Notions of measure theory that underpin probability will also appear.

Chapter 3

Complex Probability

Before discussing properties of complex random variables and their distributions, this chapter will discuss notions of complex probability. The motivation for this discussion is the introduction of Wirtinger calculus in the previous chapter to bypass the stringent requirements of holomorphicity in $\mathbb{C}$ and provide methods of calculus on real-valued functions of complex variables. By relaxing the definition of differentiability in $\mathbb{C}$, we lose the ability to invoke powerful tools of complex analysis, many of which were developed for holomorphic functions (Ahlfors (1979)). This chapter discusses the question: what if, instead of changing the meaning of differentiability, we defined distributions by holomorphic functions, hence: not real-valued?

3.1 Complex Probability

To obtain, for example, a holomorphic density function $f_Z(z)$ where Z is a complex random variable, we immediately require that for non-constant $f_Z(z)$, the density must take complex values. As the density is a representation of the probability, we must thus ask ourselves what is a complex probability?

In the literature, there exist several approaches to complex probability. Some attempt to generalise certain conditions of Kolmogorov's axioms of real probability (Jaoude (2016); Lin (2021); Wang (2021)). Others attack the idea from a measure theoretic perspective (Srinivasan and Sudarshan (1994); Srinivasan (1997)). There are also approaches that consider the underlying subjectivity or belief principles in a Bayesian probability framework and use decision and logic theory to generalise real probability in arguing for a more fundamental notion of complex probability (Youssef (1994); Bordley (2005)). By far, however the most established approach to complex probability is that in the discipline of quantum physics. In quantum physics, phenomena that appeared to defy real (classical) mechanics induced the necessity to describe quantum states in terms of probabilities. In turn, quantum physicists realised even before Kolmogorov's axioms were published in 1933, that complex numbers were essential in the description of quantum states, such as in Schrödinger's famous equation (Schrödinger (1926); Wigner (1932)). Since then, more recent research has reinforced that even not assuming complex numbers in the classical sense of $i^2 = -1$, the complex algebra is indeed the necessary algebra by which to mathematically represent quantum systems (Goyal et al. (2010); Renou et al. (2021)).

We now provide an overview and discuss the three broad approaches to complex probability present in the literature: axiomatic, measure theoretic and quantum physical. The main difference between the first and the third approaches comes from whether one wishes to normalise the sum of complex probabilities, or the L_1-

or L_2-norm, where in the case of real positive probabilities, the sum and the L_1-norm coincide. Finally, we provide a motivating example of why it may be statistically advantageous to consider complex probabilities, even in a setting of real random variables.

3.2 An Axiomatic Approach

Much of probability theory is based upon Kolmogorov's five axioms of probability set out in Kolmogorov (1933). We state the axioms as they are understood in modern terms (rather than the original five axioms given by Kolmogorov). For a sample space Ω and $\mathcal{A}$ a collection of subsets of Ω closed under countable unions and complementation (called a σ-algebra), the elements of $\mathcal{A}$ are called events and we define a probability measure to be such that:

1. To each $A \in \mathcal{A}$, we can assign a *non-negative real* number $\mathbb{P}(A)$. The quantity $\mathbb{P}(A)$ is called the probability of the event A.
2. $\mathbb{P}(\Omega) = 1$ (normalisation).
3. If $A_1, A_2, \ldots \in \mathcal{A}$ are disjoint, then

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

For every event $A \in \mathcal{A}$, axioms 1–3 imply that $\mathbb{P}(A) \in [0, 1]$. The probability is commonly understood in the frequentist sense of being the long-run relative frequency of the occurrence of event A in repeated trials, however other schools of thought such as Bayesian probability exist.

The discussion here considers chiefly the restriction in Axiom 1 that probability must be a non-negative real number. These notions have been challenged almost from the outset by quantum physicists such as in Wigner (1932), Bartlett (1945) and other references found in Székely (2005) which emphasise how non-negative real probabilities are insufficient to describe quantum systems. We will present the quantum physical approach to complex probability in a later section and here focus on recent mathematical research that has looked at generalising the axioms of probability by relaxing the restriction that $\mathbb{P}(A)$ be a non-negative real number to now allow for complex values. In particular we look at the approach in Wang (2021), with similar notions reflected in Lin (2021).

If we allow $\mathbb{P}$ to map events to complex probabilities, i.e. we relax Axiom 1 to assign a complex number, but still require the normalisation that

$$\mathbb{P}(\Omega) = \sum_{k} \mathbb{P}(A_k) = 1 \tag{3.2.1}$$

for any partition $\{A_k\}_k$ of Ω. Then we require that the real probabilities given by $\Re(\mathbb{P})$ must sum to 1 whilst the imaginary probabilities $\Im(\mathbb{P})$ must sum to 0. In Wang (2021), the author further imposes the restriction of $\mathbb{P}$ to be in the rectangle

of the complex plane where $\Re(\mathbb{P}) \in [0,1]$ and $\Im(\mathbb{P}) \in [-1,1]$. This facilitates the representation of any complex probability $\mathbb{P}_{\mathbb{C}}(A)$ as

$$\mathbb{P}_{\mathbb{C}}(A) = P_1(A) + i(P_2(A) - P_3(A)), \tag{3.2.2}$$

for P_1, P_2, P_3 real 'Kolmogorovian' probability measures. However, an interpretation of this complex probability quantity is not subsequently provided, rather a series of familiar formulae such as independence and conditional probability are developed, but properties to do with conditional probabilities only hold under the rather restrictive case that the conditioning event B satisfies $P_2(B) = P_3(B)$ for P_2, P_3 as above. In this regard, familiar probabilistic properties only hold when working with an event which has a usual real probability $\mathbb{P}_{\mathbb{C}}(B) \in [0,1]$.

On the other hand, Lin (2021) incorporates an interpretation of complex probabilities inspired by aspects of quantum physics, such that the modulus squared of a complex probability is interpreted as the Kolmogorovian probability of the event in question. However, as recognised in Lin (2021), immediately one encounters an issue in the additive normalisation of both complex probabilities and real probabilities jointly, suggesting

$$\sum_k \mathbb{P}_{\mathbb{C}}(A_k) = 1 = \sum_k |\mathbb{P}_{\mathbb{C}}(A_k)|^2, \tag{3.2.3}$$

for $\{A_k\}_k$ some partition of Ω. That the sum and sum of moduli squared both equal 1 is not true in general for arbitrary complex-valued functions. In Lin (2021), the author steps around this issue by adding axioms in the sense that to calculate the modulus squared of non-elementary event probabilities, one should in fact take the sum of the moduli squared of complex probabilities of elementary events, not the modulus squared of the sum. That is, for $C = A \cup B$ where $A = \{\omega_k\}, B = \{\omega_l\}$ for some $\omega_k, \omega_l \in \Omega$, then

$$|\mathbb{P}_{\mathbb{C}}(C)|^2 = |\mathbb{P}_{\mathbb{C}}(A)|^2 + |\mathbb{P}_{\mathbb{C}}(B)|^2, \tag{3.2.4}$$

which of course is uncomfortable regarding redefining the modulus, and may make it difficult to keep track of probabilities in more involved settings.

That is not to say that an additively normalised system of complex probability is incorrect or useless, rather on the contrary, that much still needs to be addressed in terms of its interpretation. We do not provide a definitive interpretation here, however leave this section with a motivating example for the consideration of probabilities which are additively normalised to 1.

Example 3.2.1 (Bidabad et al. (1992))**.** *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a real-valued, discrete-time Markov chain on two states $\{0,1\}$ with transition matrix*

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \tag{3.2.5}$$

where $p_{kl} = \mathbb{P}(X_{t+1} = l \mid X_t = k)$ for $k, l \in \{0,1\}$ and $\alpha, \beta \in (0,1)$. The chain is depicted below.

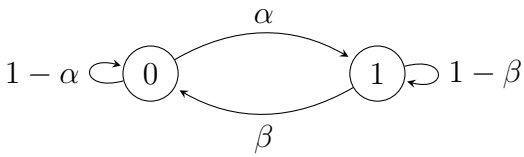

Figure 3.1: Simple two-state Markov chain

The transition probability for m-steps ahead, $\mathbb{P}\left(X_{t+m} = l \mid X_t = k\right)$*, is given by the* kl^{th} *entry in the matrix*

$$\mathbf{P}^m = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta + \alpha(1-(\alpha+\beta))^m & \alpha - \alpha(1-(\alpha+\beta))^m \\ \beta - \beta(1-(\alpha+\beta))^m & \alpha + \beta(1-(\alpha+\beta))^m \end{pmatrix}. \tag{3.2.6}$$

Letting m *tend to infinity, we observe the anticipated limiting stationary distribution of* $(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$ *as* $|1-(\alpha+\beta)| < 1$*. Now what if, in defiance of physical laws, we consider the transition probabilities at a* $\frac{1}{2}$*-step ahead (*$m = \frac{1}{2}$*)? Focusing on* $p_{00}^{1/2}$*, we obtain*

$$p_{00}^{1/2} = \frac{1}{\alpha+\beta}(\beta + \alpha\sqrt{1-(\alpha+\beta)}), \tag{3.2.7}$$

which is necessarily complex when $(\alpha+\beta) > 1$*. For example, one may consider writing, for* $(\alpha+\beta) > 1$*,*

$$p_{00}^{1/2} = \frac{1}{\alpha+\beta}(\beta + i\alpha\sqrt{\alpha+\beta-1}). \tag{3.2.8}$$

Observing all other entries in $\mathbf{P}^m$*, however, it always holds that* $p_{k0}^m + p_{k1}^m = 1$*, for all* $m \in \mathbb{R}$ *and* $k \in \{0,1\}$*. Hence we obtain complex probabilities that are normalised in sum. Further, the probabilities always exhibit moduli less than or equal to* 1*, however the sum of their moduli squared is not necessarily normalised to 1. Visually, what happens to* p_{00}^m *as* m *goes from* 0 *to infinity is depicted below in both cases* $(\alpha+\beta) < 1$ *and* $(\alpha+\beta) > 1$ *in the complex plane.*

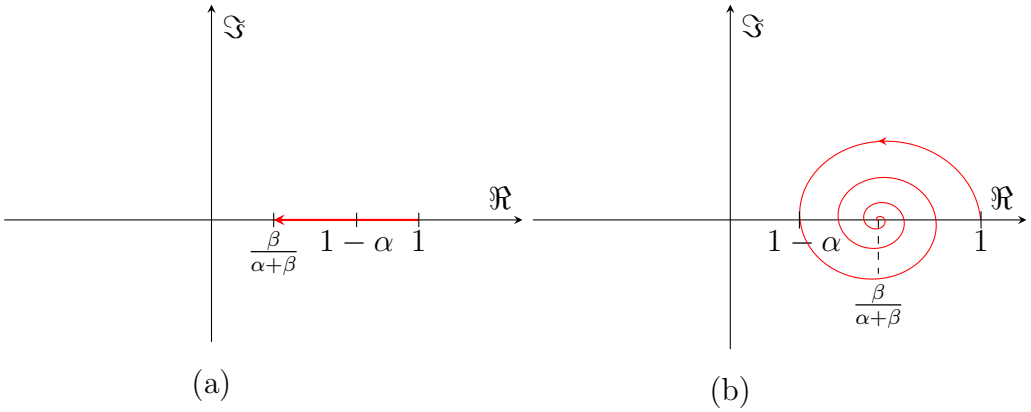

Figure 3.2: Convergence path of p_{00}^m in continuous time for (a) $\alpha+\beta < 1$ and (b) $\alpha+\beta > 1$.

The example above suggests both that complex probabilities may be considered as fractional transitional probabilities which also sum to a normalising value, and that despite being motivated here by complex random variables, complex probabilities may equally arise in contexts of real random variables. Of course, to understand these complex probabilities, we require an explanation of why it is of interest to consider fractional steps ahead 'in defiance of physical laws'. Without proclaiming a definitive interpretation, one can think of these half steps ahead as describing the randomness in the state of the chain where it cannot be observed. However, for the probability to continuously describe the state of the chain, when $\alpha + \beta > 1$, it cannot pass through $\mathbb{R}$ without passing through its stationary distribution, where it would get stuck. Hence the randomness in continuous time must leave the real numbers in order to oscillate either side of the stationary distribution point, resulting in the complex spiral. Although, to actually say what a complex probability e.g. $0.6 + 0.3i$, means however, the popular frequentist interpretation of probability appears insufficient and we require a different interpretation of this information.

3.3 A Measure Theoretic Approach

Considering probability now as an application of measure theory, we may instead reformulate Kolmogorov's axioms. Let Ω be a set called the sample space and $\mathcal{A}$ a σ-algebra on Ω. Then $(\Omega, \mathcal{A})$ forms a measurable space, upon which a probability $\mathbb{P}$ is a non-negative, real measure (by definition additive over disjoint unions) such that

$$\mathbb{P}(\Omega) = 1. \tag{3.3.1}$$

Expressed in this way allows the formulation of random variables as functions from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$ (Folland (1984, Chapter 9)).[1] That is, a random variable X is a $\mathbb{P}$-measurable function

$$X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) : \omega \mapsto X(\omega). \tag{3.3.2}$$

Then we may define the distribution function of X as the function

$$F_X : \mathbb{R} \to [0, 1], \tag{3.3.3}$$

$$F_X(x) = (\mathbb{P} \circ X^{-1})(-\infty, x], \tag{3.3.4}$$

where $\mathbb{P} \circ X^{-1}$ is itself a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The density function of an absolutely continuous real random variable, with respect to some reference measure (η), e.g., the Lebesgue measure on $[0, 1]$, is given by the Radon-Nikodym derivative, i.e., the function f_X such that $(\mathbb{P} \circ X^{-1})(\mathrm{d}x) = f(x)\,\eta(\mathrm{d}x)$ (Folland (1984, Chapter 3)). In this regard, it is relatively straightforward to define a complex random variable Z as a function

$$Z : \Omega \to \mathbb{C} : \omega \mapsto Z(\omega), \tag{3.3.5}$$

[1] One can also consider σ-algebras other than the conventional Borel σ-algebra.

which may have a probability distribution given by the measure $\mathbb{P} \circ Z^{-1}$. However, by nature of $\mathbb{P}$, this new probability measure over $\mathbb{C}$ is still real-valued. In order to obtain a complex probability, we are required to extend $\mathbb{P}$ to be a complex measure, which is a well-defined object in measure theory. For example, from Folland (1984, p. 88), a complex measure ν may be defined by two signed and bounded measures λ, μ which in turn are defined by standard, positive measures λ_+, λ_- and μ_+, μ_- respectively. That is, for signed real measures

$$\lambda = \lambda_+ - \lambda_-, \tag{3.3.6}$$

$$\mu = \mu_+ - \mu_-, \tag{3.3.7}$$

one can define the complex measure ν by

$$\nu = \lambda + i\mu. \tag{3.3.8}$$

In particular, by consequence of the Jordan decomposition for signed measures (Theorem 3.4, Folland (1984)), $\lambda_\pm, \mu_\pm$ are unique for each ν. Then, by application of the Lebesgue-Radon-Nikodym theorem (Theorem 3.12, Folland (1984)), we can obtain a complex density function $f = \frac{\mathrm{d}\nu}{\mathrm{d}\eta}$ with respect to a reference measure η.[2] It is possible for η to be a real measure, such as the familiar Lebesgue measure.

As in Wang (2021) and Lin (2021), the question of normalisation again arises. Much of the measure theoretic approach to complex probability in the literature is set out in Srinivasan and Sudarshan (1994) and Srinivasan (1997). It is worth noting that these authors motivated their investigation of complex probability from a quantum physical perspective. The normalisation condition in Srinivasan and Sudarshan (1994) requires

$$\nu(\Omega) = 1\,,\ |\nu(A)| \leq 1\ \forall\, A \in \mathcal{A}. \tag{3.3.9}$$

So again, the complex probabilities over a partition of Ω sum to 1. Srinivasan and Sudarshan (1994) then consider what they call the mod and mod^2 measures, $|\nu|$ and $|\nu|^2$ respectively. In the language of measure theory, the mod measure $|\nu|$ is the total variation measure of ν, which will lead us to call the less-standard mod^2 measure the total variation squared measure (Folland (1984)). In the context of densities, if η is a real, positive measure that takes finite values on a countable cover of Ω (i.e., σ-finite), then for $f = \frac{\mathrm{d}\nu}{\mathrm{d}\eta}$, we have $|f| = \frac{\mathrm{d}|\nu|}{\mathrm{d}\eta}$ and we define the total variation squared measure $|\nu|^2$ by its density $\frac{\mathrm{d}|\nu|^2}{\mathrm{d}\eta} = |f|^2$. These two total variation-type measures can subsequently be normalised to yield real probability measures, however we do not impose (unlike in Lin (2021)) that the normalisation of all three measures must occur simultaneously. In fact, by Exercise 3.20 in Folland (1984), if $|\nu|$ is also normalised such that $\nu(\Omega) = 1 = |\nu|(\Omega)$, then it must be that $\nu = |\nu|$, so we are already working with a real probability measure.

In Srinivasan and Sudarshan (1994), many examples are presented to explain and further motivate the use of complex measures as complex probabilities. Their

[2]Here we assume that the orthogonal measure in Theorem 3.12 of Folland (1984) is zero, so that the complex measure has a Radon-Nikodym derivative with respect to η.

examples range from common real probability distributions with complex parameters, to re-deriving Schrödinger's equation and other quantum physical laws. To illustrate computation between the three related measures, we provide their first two examples involving familiar probability distributions. Note that in Srinivasan and Sudarshan (1994), the mod (total variation) measure $|\nu|$ is defined as the measure such that

$$|\nu|(A) = \sup\left\{\left|\int_A f \,\mathrm{d}\nu\right| : |f| \leq 1\right\}, \tag{3.3.10}$$

which in Exercise 3.21 from Folland (1984) can be seen to be equivalent to the definition given earlier in terms of Radon-Nikodym derivatives.

Example 3.3.1 (Srinivasan and Sudarshan (1994))**.** *Consider an exponentially distributed real-random variable Y with complex rate parameter $\alpha+i\beta$ for $\alpha > 0, \beta \in \mathbb{R}$. Then its distribution measure ν has density with respect to the Lebesgue measure given by*

$$\nu(\mathrm{d}y) = (\alpha + i\beta)e^{-(\alpha+i\beta)y}\,\mathrm{d}y\,,\ y > 0. \tag{3.3.11}$$

We can calculate the total variation measure to be

$$|\nu|(\mathrm{d}y) = \sqrt{\alpha^2+\beta^2}\,e^{-\alpha y}\,\mathrm{d}y\,,\ y > 0, \tag{3.3.12}$$

which when normalised to become a real probability measure yields the usual exponential distribution

$$|\nu|_N(\mathrm{d}y) = \alpha e^{-\alpha y}\,\mathrm{d}y\,,\ y > 0. \tag{3.3.13}$$

Similarly, the normalised total variation squared distribution

$$|\nu|_N^2(\mathrm{d}y) = 2\alpha e^{-2\alpha y}\,\mathrm{d}y\,,\ y > 0. \tag{3.3.14}$$

Turning to a discrete example, let ξ be a real binomial distributed discrete random variable with probability mass function

$$\mathbb{P}_\nu(\xi = x) = \binom{n}{x} p_1^x p_2^{n-x}(p_1+p_2)^{-n}\,,\ x = 0, 1, \ldots, n, \tag{3.3.15}$$

for $n \in \mathbb{Z}$ and $p_1, p_2 \in \mathbb{C}$ complex probability parameters. Then the probability function with respect to the normalised *total variation measure $|\nu|_N$ is*

$$\mathbb{P}_{|\nu|_N}(\xi = x) = \binom{n}{x} |p_1|^x |p_2|^{n-x}(|p_1|+|p_2|)^{-n}\,,\ x = 0, 1, \ldots, n, \tag{3.3.16}$$

and with respect to the normalised total variation squared measure,

$$\mathbb{P}_{|\nu|_N^2}(\xi = x) = \binom{n}{x} |p_1|^{2x} |p_2|^{2n-2x}(|p_1|^2+|p_2|^2)^{-n}\,,\ x = 0, 1, \ldots, n. \tag{3.3.17}$$

In particular, the normalised total variation squared measure of (3.3.17), coupled with the following discussion of the quantum physical interpretation of complex probability will motivate the parameterisation of the complex Bernoulli distribution in Chapter 5.

3.4 A Quantum Physical Approach

Indeed, underlying many of the previously discussed approaches to complex probability is both a motivation and interpretation arising from quantum physics (synonymous here with quantum mechanics). In the eloquence of Aaronson (2013, p. 110),

> Quantum mechanics is what you would inevitably come up with if you started from probability theory, and then said, let's try to generalize it so that the numbers we used to call "probabilities" can be negative numbers. As such, the theory could have been invented by mathematicians in the nineteenth century without any input from experiment. It wasn't, but it could have been.

We do not intend this discussion to be a thorough treatment of quantum mechanics, rather we isolate the probabilistic notions and relate them to the wider context of probability and statistics. For a digestible introduction to quantum mechanics, the reader may refer to Chapter 9 of Aaronson (2013), or Bartlett (1945) for a more technical motivation.

Without delving deeply into quantum theory, the main idea is that states (often assumedly discrete in explanatory texts) of a quantum system have a probability vector attached to them, representing the complex-valued probability "amplitudes" of each "bit" or elementary event (Aaronson (2013); Knuth (2016); Jansson (2019)). To borrow an example from Aaronson (2013), consider a binary experiment with outcome 0 or 1 which is represented by the vector $(z, w)^\top \in \mathbb{C}^2$ such that

$$\left\| \begin{pmatrix} z \\ w \end{pmatrix} \right\|^2 = |z|^2 + |w|^2 = 1. \tag{3.4.1}$$

So the L_2-norm normalises to 1, with no other restrictions required. In this context, let $|z|^2$ be the real probability of a 0, and $|w|^2 = 1 - |z|^2$ the probability of a 1. This leads to calling z and w complex probability amplitudes. But why do we need complex numbers? The importance of describing the system by z and w instead of $|z|^2$ and $|w|^2$ lies in considering transformations of a state. In particular, what is called quantum interference corresponds to changing the experimental observation of the same system, which is then mathematically represented by the system which now has a complex probability amplitude vector given by a unitary linear transformation of $(z, w)^\top$. The change in a quantum state has been empirically verified to be a linear transform of the complex probability amplitudes, and only unitary linear transforms retain the unit L_2-norm (Aaronson (2013)). Such a change, in general, cannot mathematically be represented by a linear, or otherwise simple, transformation of the real probabilities $|z|^2, |w|^2$. Research by Goyal et al. (2010) goes further to show that under very weak physical assumptions, quantum

probability amplitudes behave according to the complex algebra, without assuming the existence of complex numbers in the first place.

Returning to the fact that states in a system can be represented by a complex vector, we now extend the interpretation into a continuous probability setting which links ideas from Srinivasan and Sudarshan (1994) and Lin (2021) resolved into a way of writing what quantum physics says, but in a notation familiar to probabilists and statisticians. The main idea here follows an extension of an interesting discussion in Knuth (2016) about the consideration of complex probability amplitudes as vectors, and whether vector space operations such as inner products have meaning.

3.4.1 Complex probability vectors

If a discrete probability distribution over sample space $\Omega = \{\omega_k\}_{k=1}^n$ for a partition of elementary events $E_k = \{\omega_k\}, k = 1, \ldots, n$, is given in the sense of the real probabilities

$$\mathbb{P}(E_k) = p_k. \tag{3.4.2}$$

Then the complex probability vector of the same space is given by the vector

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{C}^n, \tag{3.4.3}$$

such that $|v_k|^2 = p_k$ for all $k = 1, 2, \ldots, n$. This $\boldsymbol{v}$ is not necessarily unique thus far — we will discuss uniqueness separately soon.

Of course, we may not necessarily require a finite partition of elementary events, such as for a Poisson distribution, in which case we can obtain an infinite vector as long as the L_2-norm $\sum_{k=1}^{\infty} |v_k|^2 = \sum_{k=1}^{\infty} p_k = 1$. Knuth (2016) demonstrate how a measurable event A, which is necessarily a union of some E_k's, can be represented by the complex probability vector $\boldsymbol{v}_A$ with $(\boldsymbol{v}_A)_k = v_k$ if $E_k \cap A = E_k$ and 0 if $E_k \cap A = \emptyset$. For example, if $A = E_1 \cup E_3$, then it is represented by the vector

$$\boldsymbol{v}_A = \begin{pmatrix} v_1 \\ 0 \\ v_3 \\ 0 \\ \vdots \end{pmatrix} = \begin{pmatrix} v_1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ v_3 \\ 0 \\ \vdots \end{pmatrix} = \boldsymbol{v}_{E_1} + \boldsymbol{v}_{E_3} = \boldsymbol{v} \odot (\boldsymbol{e}_1 + \boldsymbol{e}_3), \tag{3.4.4}$$

where $\odot$ is element-wise multiplication and $\boldsymbol{e}_k$ are the usual standard basis vectors in $\mathbb{C}^n$ (with a 1 in the k^{th} entry and 0 elsewhere).

Interestingly, it is then noticed that for sets A and B, the real probability $\mathbb{P}(A \cap B)$ can be written as

$$\mathbb{P}(A \cap B) = \langle \boldsymbol{v}_A, \boldsymbol{v}_B \rangle = \boldsymbol{v}_A^\top \boldsymbol{v}_B^*, \tag{3.4.5}$$

using the standard inner product in $\mathbb{C}^n$. This is because the inner product sums the real probability (modulus squared) only of common elementary events between

A and B, which is what the additional axioms in Lin (2021) were trying to impose. The probability of a union is understood by the inclusion-exclusion principle. The conditional probability can then be given by

$$\mathbb{P}(A \mid B) = \frac{\langle \boldsymbol{v}_A, \boldsymbol{v}_B \rangle}{\langle \boldsymbol{v}_B, \boldsymbol{v}_B \rangle}, \tag{3.4.6}$$

which resembles the projection coefficient of $\text{proj}_{\boldsymbol{v}_B} \boldsymbol{v}_A$. Is this to suggest the conditional probability is in some sense a projection of the complex probability of A onto B? Furthermore, independence of A and B is now articulated by the condition

$$\langle \boldsymbol{v}_A, \boldsymbol{v}_B \rangle = \langle \boldsymbol{v}_A, \boldsymbol{v}_A \rangle \langle \boldsymbol{v}_B, \boldsymbol{v}_B \rangle, \tag{3.4.7}$$

which is equivalent to saying that, for A and B non-$\mathbb{P}$-null sets, $\boldsymbol{v}_A$ is orthogonal to $\boldsymbol{v}_A - \frac{\boldsymbol{v}_B}{\|\boldsymbol{v}_B\|^2}$ and $\boldsymbol{v}_B$ is orthogonal to $\frac{\boldsymbol{v}_A}{\|\boldsymbol{v}_A\|^2} - \boldsymbol{v}_B$. This is consistent with the statement that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and vice versa. However, to look beyond the ideas in Knuth (2016) replacing $\langle \cdot, \cdot \rangle$ with the covariance operator, we would recover what looks like a correlation coefficient in (3.4.7), although equal to 1 and with variances in the denominator instead of standard deviations. This may suggest that working with these complex probability vectors can provide a new framework for considering dependence. We mark this as an area for further research.

3.4.2 Complex probability densities

Of course, in the case of a continuous random variable, we tend to work with continuous probability density functions, where the elementary events with respect to a continuous random variable are given in terms of a differential $\eta(\mathrm{d}x)$ (often just $\mathrm{d}x$ with respect to the Lebesgue measure). Now, we have an uncountable number of elementary events. But we can extend a complex probability vector, as discussed in Aaronson (2013) and Knuth (2016), to now be a complex-valued function, ϕ. Extending also the normalisation condition under the standard function inner product, we require that

$$\langle \phi, \phi \rangle_\eta = \int \phi(x)\phi(x)^* \eta(\mathrm{d}x) = \int |\phi(x)|^2 \eta(\mathrm{d}x) = 1. \tag{3.4.8}$$

Immediately, we can recognise that $|\phi(x)|^2$ as a non-negative real function must be a real probability density with respect to the measure η. We will therefore call ϕ a complex density function with respect to the measure η. In fact, this is exactly the same intuition behind the wave function in quantum mechanics and how the solution to Schrödinger's wave equation exhibits a probability density as its modulus squared. We recall from Equation (3.4.4) that the probability vector of a given set A was just the element-wise product of the whole probability vector with the relevant sum of standard basis vectors in $\mathbb{C}^n$. In the continuous setting, this corresponds to the function product (which is the generalisation of the element-wise product) of ϕ and the indicator function

$$\phi \mathbb{1}_A. \tag{3.4.9}$$

Observing now the inner product results obtained from vectors, we see that

$$\mathbb{P}(A) = \langle \phi \mathbb{1}_A, \phi \mathbb{1}_A \rangle = \int \phi \mathbb{1}_A(x)(\phi \mathbb{1}_A(x))^* \eta(\mathrm{d}x) = \int_A |\phi(x)|^2 \eta(\mathrm{d}x), \tag{3.4.10}$$

which is exactly the probability expression we expect, understanding $|\phi(x)|^2$ as the real probability density with respect to η. The joint probability is similarly yielded by

$$\mathbb{P}(A \cap B) = \langle \phi \mathbb{1}_A, \phi \mathbb{1}_B \rangle = \int_{A \cap B} |\phi(x)|^2 \eta(\mathrm{d}x), \tag{3.4.11}$$

and the conditional probability expression is also seen to translate in terms of inner products. Thus far, we have obtained equivalent expressions to those developed in measure theory, and indeed this suggests an underlying link between the total variation squared (a.k.a. mod^2) measure from Srinivasan and Sudarshan (1994) and complex probability amplitudes in quantum physics. Furthermore, we can now express that (non-$\mathbb{P}$-null) events A and B are independent if and only if $\phi \mathbb{1}_A$ and $\phi \mathbb{1}_B$ are orthogonal to $\phi \mathbb{1}_A - \frac{\phi \mathbb{1}_B}{\|\phi \mathbb{1}_B\|^2}$ and $\frac{\phi \mathbb{1}_A}{\|\phi \mathbb{1}_A\|^2} - \phi \mathbb{1}_B$ respectively. Perhaps, again, this expression of independence as orthogonality of complex functions may provide novel methods of measuring and estimating dependence.

We return now to the idea of quantum interference where a change in how we observe a system corresponds to a unitary linear transform of the complex probability amplitudes. Arguments for the necessity of unitary transforms to represent interference are addressed in Aaronson (2013). We now provide a broader interpretation of quantum interference in measure theoretic terms. Suppose our experimental setup is such that we cannot measure all elementary events of the sample space Ω, but rather the σ-algebra is composed of a reduced collection containing composite events A_k. Then we only observe the probabilities of the form $\phi \mathbb{1}_{A_k}$ at best. In this regard, knowing ϕ provides a fundamental understanding of the whole random system, even where in any given experiment, it may be that only certain, non-elementary outcomes are observed. In the language of probability, if one knows ϕ, then one can evaluate probabilities on an arbitrary σ-algebra, or one can even move between σ-algebras which define in the sense of the experimental setup which events are measurable or observable.

3.4.3 Uniqueness

Before addressing complex probabilities from a statistical perspective, we briefly touch upon the notion of uniqueness. Given only the real probabilities of an experiment, there are infinitely many complex probability densities which yield the same real probability distribution, for example up to rotations or reflections in the complex plane. In quantum theory, the uniqueness is imposed from the behaviour of quantum interference, which is not necessarily observed in a classical mechanical or statistical setting. Whilst the function ϕ represents what appears to be a more fundamental description of a random experiment, in familiar settings of probability and statistics where one does not change the experimental setup to work over a different σ-algebra of events, then a real probabilistic description is sufficient and a complex probability function may not yield more information. Without presuming

to provide any answers here, we pose the question for future research of where may there be some probabilistic benefit to working with these complex probabilities ϕ, not just over complex random variables to obtain holomorphic densities, but even over real random variables to consider varying experimental observations over the same physical setup (changing or reducing the σ-algebra).

3.5 A Statistical Perspective

Of course, now hopping out of probability theory and across to statistics, how would one go about perhaps estimating ϕ if we are in an experimental situation with observations? Quantum physicists have developed the field of wave-function tomography, which looks at estimating the wave function (ϕ) of a system by analysing measurements of outcomes of the same system from different perspectives, i.e., under different σ-algebras (Lvovsky and Raymer (2009); Lundeen et al. (2011); Nguyen and Choi (2021)). This method is similar to wave estimation used in other fields such as geophysics and neuroscience concerning MRI data — both fields which inherently also deal with complex random variables (see, e.g., Ogawa et al. (1990); Sandoval et al. (2004) and references therein).

Statistically speaking, one could take a parametric approach by assuming that a certain form of the complex probability function ϕ governs the observed system and estimate the parameters thereof. What this amounts to in some sense is estimating the square root of probabilities. Knuth (2016) provides examples of where square roots of probabilities arise, such as in the Hellinger-Battacharyya distance. Here, however, we present a motivating example that may suggest inferential benefit from direct estimation of a complex probability function.

Example 3.5.1. *Consider estimating p, the probability parameter of a Bernoulli or binomial distribution. Statistically, one performs an estimate of p (and therefore of $q = 1 - p$) over $p \in [0, 1]$. Graphically in pq-space, it is an estimate along the red line below.*

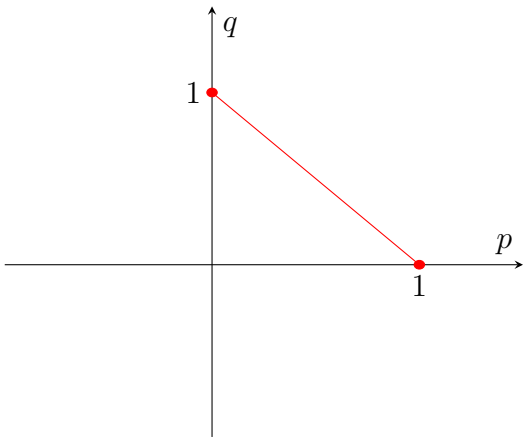

Figure 3.3: Estimation of the real binary probability p.

In particular, we acknowledge boundary issues with inference where, given a sample, if the estimate $\hat{p}$ (e.g. maximum likelihood estimator — arithmetic mean)

is close to 1 *or* 0*, then the asymptotic variance estimate of* $\hat{p}$ *used to form confidence intervals may be abruptly cut off at the boundaries in Figure 3.3, reducing the validity of inference in such cases. However, estimating a* v *such that* $|v|^2 = p$ *and hence a* w *such that* $|w|^2 = q = 1 - |v|^2$ *amounts to estimating on some sort of complex circle in the two dimensional complex* vw*-space. Visually in* $\mathbb{R}^2$ *where we consider* v *and* w *real, but allow to be negative (technically they may be complex, but four dimensions is not as easy to depict), then estimation of* v *and therefore* w *may be thought to occur on the following circle.*

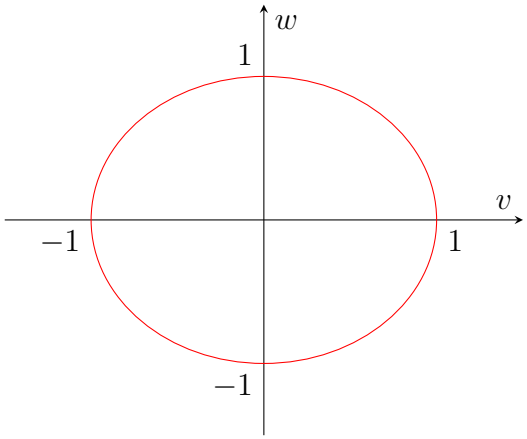

Figure 3.4: Estimation of the complex probability v.

As is graphically suggested, by smoothness of the circle at any point, directly estimating v *and* w *should not yield boundary issues, even near a modulus of* 0 *or* 1*. Inferential (e.g., confidence) intervals may now cross the axes, but still capture a parameter with a certain probability. Compared to the real case, or even a generalised* L_1*-norm case such as under the total variation measure (estimation on a diamond), statistical estimation directly of the complex probability may enhance the precision of inference by removing discontinuities and boundary problems in estimation.*

We will not pursue the notion of complex probability much further here, turning rather to complex random variables in the next chapter, which for the most part will be parameterised by real probability distributions under Wirtinger calculus. Nonetheless, we keep in mind the notion of an underlying complex probability function ϕ, which may now be holomorphic and allow the application of complex analysis to differentiate or integrate under the natural differential structure in $\mathbb{C}$. As a final remark, however, consider the real density with respect to the Lebesgue measure given by

$$f(z) = \frac{1}{\pi} e^{-|z|^2}, \ z \in \mathbb{C}. \tag{3.5.1}$$

What holomorphic function ϕ would be such that $|\phi(z)|^2 = f(z)$ (density of a complex normal distribution) for all $z \in \mathbb{C}$? Being of the form $e^{|z|^2/2} e^{i\alpha}$, we are unsure that the motivating search for holomorphic densities will be satisfied by development of complex probability theory in the quantum direction, or if perhaps other

avenues need to be considered, such as the additive probability example of Bidabad et al. (1992) (Example 3.2.1), or maybe even a different relationship between real and complex probability altogether. Nonetheless, at this stage the quantum approach appears to us to hold the most established framework and potential for further probabilistic research and statistical application.

Chapter 4

Complex Random Variables

This chapter presents established theory around complex-valued (hereafter just 'complex') random variables and mathematical description of their distribution, augmented with new results. The distribution of a random variable may be uniquely defined by many mathematical objects. At the fundamental level, the distribution defines the probability of any measurable event occurring. That is, the distribution is defined by $\{\mathbb{P}(Z \in A) : A \in \mathcal{B}(\mathbb{C})\}$. When handling real random variables, objects frequently used to uniquely determine the distribution include the cumulative distribution function (c.d.f.), the probability mass or density functions (p.m.f. and p.d.f. respectively), and the characteristic function (c.f.). A treatment of each of these in a complex setting follows. However, unlike much of the prominent literature on random complex numbers (Wooding (1956); Goodman (1963); Eriksson et al. (2010); Schreier and Scharf (2010); Ducharme et al. (2016), and references therein), here the development of complex-valued distributions will be motivated from within $\mathbb{C}$ itself, rather than translating notation of bivariate analogues from $\mathbb{R}^2$ to $\mathbb{C}$. That is not to say that previous results are not useful, in fact the majority of this chapter provides already established results, however the motivation is in some cases presented differently, and encourages the exploration of much ignored ideas, such as the Mellin transform.

4.1 Complex-Valued Random Variables

As with real random variables, a complex random variable (c.r.v.) is simply a measurable function from some sample measure space $(\Omega, \mathcal{A}, \mathbb{P})$ to the space $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$. Here the Borel σ-algebra is assumed, however a different σ-algebra may also be considered to define the sets upon which one wishes to be able to define a probability. That is,

$$Z : \Omega \to \mathbb{C} : \omega \mapsto Z(\omega).$$

Situations where complex values occur as representations of output of an experiment or phenomenon were discussed in Section 1.1 and Section 2.3.

The probability of events occurring in the sample space Ω of a given experiment will be expressed in terms of the random variable indicating the quantity of interest, whereby $\mathbb{P}(Z \in A)$ is short hand for $\mathbb{P}(\{\omega \in \Omega : Z(\omega) \in A\})$, because formally $\mathbb{P}$ acts upon subsets of Ω that are in $\mathcal{A}$. The bulk of this section will work with real-valued probability measures, unlike complex-valued ones discussed in Chapter 3.

As per the decomposition of a complex number, one can express a c.r.v. in terms of its Cartesian or polar forms. That is, for X, Y, R and Θ being real-valued random variables,

$$Z = X + iY = Re^{i\Theta}. \tag{4.1.1}$$

The majority of early literature, most notably Wooding (1956) and Goodman (1963), developed c.r.v.'s for interest in signal processing and modulation, furthered with later developments of complex principal component analysis (PCA) where the real component of the complex signal was the original data of interest (Horel (1984)). This has continued throughout more recent theoretical work, such as in Eriksson et al. (2010) or Schreier and Scharf (2010) by primarily tackling probabilistic and statistical problems in $\mathbb{C}^n$ as equivalent problems in $\mathbb{R}^{2n}$, making use of the conversion matrix $\mathbf{J}_n$, augmented vectors $\underline{\boldsymbol{z}}$ and the Wirtinger derivatives to translate notation.

4.1.1 Discrete, contour and continuous random variables

Just as with real probability, complex random variables may be classified as discrete, or continuous (or hybrid/semicontinuous). However in $\mathbb{C}$, we also introduce the notion of contour random variables, that is, complex random variables taking values along a contour in $\mathbb{C}$.

Definition 4.1.1. *A complex random variable Z is called:*

(a) discrete if it takes values on a countable subset of $\mathbb{C}$.
(b) continuous if it takes values (surjectively) on a connected subset of $\mathbb{C}$ containing at least one ball $\{z \in \mathbb{C} : |z - z_0| < \varepsilon\}$ for some $z_0 \in \mathbb{C}$ and real $\varepsilon > 0$.
(c) contour if it takes values on a contour (curve) in $\mathbb{C}$.
(d) hybrid or semicontinuous/semicontour if it is a mix of the above categories.

In particular, the rest of this chapter will focus upon the treatment of continuous complex random variables. Discrete random variables may be naturally thought of as having a single real probability (or complex probability per Chapter 3) assigned to each distinct outcome. Examples of discrete complex random variables are given in Chapter 5. Contour random variables are not explored here, and present an interesting topic of future research. Particularly, with contour random variables we may define a complex density function under the natural differential structure in $\mathbb{C}$, as we understand contour integration. Here we may have complex valued density functions which integrate to 1 — i.e. densities with respect to a complex Lebesgue measure. For example a contour random variable of modulus 1 distributed around the unit circle in the complex plane. Moments of such a random variable yield a form of characteristic function, closely related to directional statistics (Mardia and Jupp (2000); Halliwell (2015)).

4.2 Cumulative Distribution Functions

For a real random variable, its c.d.f. $F_X : \mathbb{R} \to [0, 1]$ is defined as the function such that $F_X(x) = \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$. In the context of real random variables, this is a convenient construction relative to the Borel σ-algebra of $\mathbb{R}$, which can be seen

to be the σ-algebra generated by all sets of the form $(-\infty, x]$ for $x \in \mathbb{R}$ (technically $\mathbb{Q}$ is sufficient). As such, any measurable event can have its probability calculated if the c.d.f. is known, by taking sums and differences of c.d.f. values.

However, it is well-known that for $d \geq 2$, the space $\mathbb{R}^d$ does not exhibit a canonical ordering that is completely transitive and asymmetric for all elements. That is to say "greater than" ($>$) and "less than" ($<$) cannot be defined to be able to compare and distinguish all unique points. This consequently also applies to the complex field, where there does not exist a natural, complete ordering of the complex numbers. Therefore, it does not make sense to naïvely write the c.d.f. of a c.r.v. as $F_Z(z) = \mathbb{P}(Z \leq z)$, without first defining the $\leq$ relation in $\mathbb{C}$. This problem of ordering in multivariate statistics has been studied in great depth by many, at least since 1966 to our knowledge, with concerns from Kendall in "Discrimination and Classification" (refer to Barnett (1976)), and the problem remains broadly open today (e.g. Hallin et al. (2021)).

Almost all treatments of c.r.v.'s to date begin with the definition of the distribution function of the complex random variable $Z = X + iY$ as $F_Z(z) \coloneqq F_{(X,Y)}(x, y)$, being the bivariate distribution of the real and imaginary component, justified by the unique Cartesian decomposition of a complex number (Eriksson et al. (2010); Schreier and Scharf (2010); Adali and Schreier (2014)). Often this lends itself to the notation $\mathbb{P}(Z \leq z) = \mathbb{P}(\Re(Z) \leq \Re(z), \Im(Z) \leq \Im(z))$. The formulation from the joint c.d.f. in $\mathbb{R}^2$ likely arises from early development of the complex normal distribution defined as the complex random variable with real and imaginary parts normally distributed — essentially a bivariate normal distribution translated into complex notation (Wooding (1956); Goodman (1963); Andersen et al. (1995)). One may consider, as in Kesemen et al. (2020), whether the distribution function could also be meaningfully defined in terms of polar coordinates. However the description of $\mathbb{P}(Z \leq z) = \mathbb{P}(|Z| \leq |z|, \operatorname{Arg}(Z) \leq \operatorname{Arg}(z))$ immediately yields problems in terms of the argument relationship exhibiting branch cuts, as angles are equivalent modulo 2π. As such, the interval to consider angles in makes little sense, especially as one would have $-\pi + \varepsilon < \pi$ as the widest angle difference, despite the angles only being separated by some small $\varepsilon > 0$ (Júlvez (2015)). That is not to say a polar description of the c.d.f. is irrelevant, rather the functional form should consider a ball-wise approach, rather than remaining fixed at the origin as we now discuss.

Defining the c.d.f. of a c.r.v. via $F_Z(z) = \mathbb{P}(\Re(Z) \leq \Re(z), \Im(Z) \leq \Im(z))$ treats probabilities in terms of sets generated by rectangles in the complex plane, which is justified being one way to generate the Borel σ-algebra of $\mathbb{C}$. However, in most complex (and multivariate real) analysis contexts, often the fundamental sets to work with are balls. As such, we can consider an equivalently unique distribution function for a given random variable Z, which can be defined by $F_Z^B(z, r) \coloneqq \mathbb{P}(|Z - z| \leq r) = \mathbb{P}\left(Z \in \overline{B}(z, r)\right)$, where $\overline{B}(z, r)$ is the closed ball in the complex plane centred at z with radius r. We will call F_Z^B the ball-wise c.d.f..

A description of $F_Z(z)$ for all $z \in \mathbb{C}$ uniquely defines the distribution of Z by describing the probability of Z falling in any generating set of the Borel σ-algebra. Likewise if we have $F_Z^B(z, r)$ for all $z \in \mathbb{C}$ for any fixed $r \in \mathbb{R}^+$ we again have a unique description of the distribution of Z. However, we will see that in some circumstances, it will be advantageous to fix z and vary r. Note that it is sufficient to be able to define F_Z and F_Z^B only on complex numbers with rational components,

and $r \in \mathbb{Q}^+$, where $\mathbb{Q}^+$ is the field of positive rational numbers (Folland (1984); Stepić and Ognjanović (2014)). Both of these distribution function descriptions are in fact explored in Stepić and Ognjanović (2014) from a perspective of handling complex probabilities, and proving that they do indeed provide complete probability logics over the complex numbers.

Whilst it may appear that F_Z^B is more complicated, with domain $\mathbb{C}\times\mathbb{R}^+$, whereas F_Z is just defined on $\mathbb{C}$, the ball-wise c.d.f. will prove to be more useful when it comes to describing circular, or even elliptically symmetric distributions as the domain of F_Z^B can be reduced, e.g. to $\mathbb{R}^+$ for circular distributions, whilst still uniquely defining the distribution of the random variable. It also can be seen to be more interpretable if one considers the stereographic projection of the complex plane onto the Riemann sphere (Ahlfors (1979, p.19)), whereby the ball-wise probability regions are also balls on the Riemann sphere.

Importantly, although these cumulative distribution functions can be theoretically defined, we see there exist multiple unique descriptions of the Borel σ-algebra, and thus multiple 'valid' analogues of the c.d.f. used in univariate real statistics once we begin to work in $\mathbb{C}$. This is primarily an issue resulting from the fact that in $\mathbb{R}$, a square and a ball take the same appearance of an interval and thus generate one obvious distribution function form, for which there are multiple analogues when extended to $\mathbb{C}$ or in higher dimensions. This issue will reappear in Chapter 7 when considering the probability integral transform which underpins many useful measures of dependence in $\mathbb{R}$. We now move away from these c.d.f. definitions interpreted by taking areas over $\mathbb{C}$ like in $\mathbb{R}^2$, to instead consider what we will call the complex cumulative distribution function.

4.2.1 Complex cumulative distribution function

We define here a complex cumulative distribution function, inspired by Stokes' formula in the complex plane (Proposition 2.6.1), letting f be a probability density (see next section). The ideas here are very similar to the discussion in Olhede (2006), however, there the author considers this function as more of an alternate density, rather than an analogue of the cumulative distribution function. Importantly, we motivate this c.d.f. description by its similarities to the real univariate c.d.f. and also remark that it is unique to complex random variables, without bivariate analogue.

Definition 4.2.1. *The complex cumulative distribution function of a continuous random variable Z is defined to be the function $F_Z^{\mathbb{C}}(z)$ which satisfies, for any region $A \subseteq \mathbb{C}$,*

$$\mathbb{P}\left(Z \in A\right) = \frac{1}{2i} \oint_{\partial A} F_Z^{\mathbb{C}}(z)\,\mathrm{d}z, \tag{4.2.1}$$

where ∂A is the boundary contour of A in $\mathbb{C}$.

In particular, it should be understood that

$$1 = \mathbb{P}(Z \in \mathbb{C}) = \lim_{R\to\infty} \frac{1}{2i} \oint_{\Gamma_R} F_Z^{\mathbb{C}}(z)\,\mathrm{d}z \tag{4.2.2}$$

$$= \lim_{R\to\infty} \frac{1}{2} \int_{-\pi}^{\pi} F_Z^{\mathbb{C}}(Re^{it})Re^{it}\,\mathrm{d}t. \tag{4.2.3}$$

That is, we consider the contour integral on a circle of radius R tending to infinity.

The complex c.d.f. is not a real valued function from which one can directly read off real probabilities, rather it describes the behaviour of the distribution on the boundary of regions. This is similar to univariate real statistics, where the usual c.d.f. is a function evaluated at x which gives the probability of a real random variable X lying in the interval $(-\infty, x]$. More precisely, if F_X is the c.d.f. of a continuous real random variable X, then $\mathbb{P}(X \in I)$ for some interval I from a to b (open, closed or a mixture) is given by $F_X(b) - F_X(a)$, where a and b form the boundary of I. Similarly, in $\mathbb{C}$, considering a region $A \subseteq \mathbb{C}$ which contains an open ball, then the probability that a complex random variable takes value in A is defined by the values of the complex c.d.f. over the boundary ∂A.

The density (as given in the next section) f_Z of a continuous complex random variable Z with respect to the Lebesgue measure can also be recovered from $F_Z^{\mathbb{C}}$ by

$$f_Z(z) = \frac{\partial F_Z^{\mathbb{C}}}{\partial z^*}. \tag{4.2.4}$$

This provides a familiar relationship as with real statistics. In particular we remark that the complex c.d.f. could prove useful to understand the distribution of arbitrary transformations of complex random variables, just as the distribution of a transformation of a real random variable can be understood by manipulating its c.d.f.. In the literature, there is currently very limited coverage of understanding transformations of random variables which may leave a given distribution family.

For example, the complex c.d.f. of a circular normal distribution with zero mean and variance σ^2 (defined in Chapter 5) is

$$\frac{1}{z\pi}\left(1 - e^{-\frac{zz^*}{\sigma^2}}\right),\ z \in \mathbb{C}. \tag{4.2.5}$$

This same function is also given in Olhede (2006). With this function, we can use contour integrals to calculate the probability of a complex normal random variable taking value in an arbitrary region on the complex plane, not just rectangles or balls as per the c.d.f.'s F_Z and F_Z^B. It is worth noting that the complex c.d.f. agrees with the rectangular and ball-wise c.d.f.'s by taking the contour integrals over rectangles or balls.

We will end the discussion of the complex c.d.f. here. Further discussion of properties and extension to discrete random variables may be found in Olhede (2006). Importantly, however, we believe that by understanding this function as an analogue of the cumulative distribution function will provide much greater insight into manipulating the distributions of complex random variables. We recommend future investigation as to the exact properties and interaction of this function with

transformations of complex random variables, and any connections it may have with notions of complex probability.

Although we have three cumulative distribution functions which may each be used to uniquely describe the distribution of a complex random variable, we will often prefer to describe the distribution locally using a p.m.f. for discrete random variables and p.d.f. in the continuous case. The following section will address description of a continuous c.r.v. with its density function, and also provide the link between its density and the ball-wise and rectangular c.d.f.'s discussed above.

4.3 Densities

Similar to how complex c.d.f.'s have been treated to date, most sources handling general c.r.v.'s establish the density function of a continuous c.r.v. as the corresponding real bivariate density (usually with respect to the Lebesgue measure) of the real and imaginary components, which arises by the treatment of $\mathbb{C}$ as a two-dimensional real manifold with integration correspondingly defined. In the case of discrete random variables, the probability mass function is just a list (e.g. vector) of probabilities corresponding to the discrete outcomes. Discrete random variables will not be explored much here, however examples are given in the next Chapter.

Being a local description of the distribution, the density, unlike the c.d.f., does not depend upon whether a ball-wise or rectangular approach is taken for the probability of generating sets. In the real case, the link between the c.d.f. and p.d.f. of a continuous random variable is that the p.d.f. (with respect to the Lebesgue measure) is the derivative of the c.d.f.. This comes from the the Radon-Nikodym theorem (see e.g. Theorems 3.8 and 3.12 in Folland (1984, pp. 83–90)), where the density function with respect to a certain measure is the limiting ratio of the probability measures over an appropriate set as the set shrinks to a point (i.e. an interval in $\mathbb{R}$). As the Lebesgue measure in $\mathbb{R}$ over an interval is just the interval length, for a certain probability measure of the distribution of random variable X on some interval $(x + \mathrm{d}x)$, we have that

$$f_X(x) = \lim_{\mathrm{d}x \to 0} \frac{F_X(x + \mathrm{d}x) - F_X(x)}{\mathrm{d}x} = F'_X(x),$$

where the difference in the c.d.f. $F_X(x+\mathrm{d}x)-F_X(x)$ is the probability $\mathbb{P}\left(X \in (x + \mathrm{d}x]\right)$. The finer notions of measure theory and conditions required for this to hold are not elaborated here and the reader can refer to many sources, such as Folland (1984).

Nonetheless, this allows us to construct a similar link between the two c.d.f. forms presented in the previous subsection and the p.d.f. of a continuous c.r.v.. In particular, considering first the rectangular approach as the usual bivariate c.d.f., F_Z, it is unsurprising that the relationship with the density function with respect to the Lebesgue measure for $z = x + iy$ is simply

$$f_Z(z) = \frac{\partial^2 F_Z}{\partial x \partial y}(z).$$

This can be verified where the Lebesgue measure in $\mathbb{C}$ is the area of the region under consideration, and identify $F_Z : \mathbb{C} \to \mathbb{R}$ with the corresponding bivariate function so that $F_Z(x+iy) = F_Z(x,y)$ for all $x+iy \in \mathbb{C}$. Thus considering the rectangle with corners given by z, $z+\mathrm{d}x$, $z+i\,\mathrm{d}y$ and $z+\mathrm{d}z$ and area given by $\mathrm{d}x\,\mathrm{d}y$, we have that with respect to the Lebesgue measure,

$$\begin{aligned} f_Z(z) &= \lim_{\mathrm{d}z\to 0} \frac{F_Z(z+\mathrm{d}z) - F_Z(z+\mathrm{d}x) - F_Z(z+i\,\mathrm{d}y) + F_Z(z)}{\mathrm{d}x\,\mathrm{d}y} \\ &= \lim_{\mathrm{d}x,\,\mathrm{d}y\to 0} \frac{F_Z(x+\mathrm{d}x, y+\mathrm{d}y) - F_Z(x+x, y+\mathrm{d}y)}{\mathrm{d}x\,\mathrm{d}y} - \frac{F_Z(x+\mathrm{d}x, y) - F_Z(x,y)}{\mathrm{d}x\,\mathrm{d}y} \\ &= \lim_{\mathrm{d}y\to 0} \frac{\frac{\partial F_Z}{\partial x}(x, y+\mathrm{d}y) - \frac{\partial F_Z}{\partial x}(x,y)}{\mathrm{d}y} \\ &= \frac{\partial^2 F_Z}{\partial x \partial y}(z), \end{aligned}$$

exactly as in the real bivariate case.

Considering now the ball-wise c.d.f., F_Z^B, we similarly have that the Lebesgue measure on the ball $\overline{B}(z,r)$ is just the area πr^2. Hence the density is given by

$$f_Z(z) = \lim_{r\to 0} \frac{F_Z^B(z,r)}{\pi r^2} = \frac{1}{\pi}\frac{\partial F_Z^B}{\partial (r^2)}(z).$$

The ball-wise construction is useful for circular, and particularly bounded, random variables.

Example 4.3.1 (Complex Uniform Distribution)**.** *Consider the complex uniform distribution (defined in Section 5.3) on the unit disc in the complex plane. Being uniform, we expect the distribution should correspond to a scaled Lebesgue measure on the unit disc. Normalisation of $F_Z^B(0,1) = 1$ implies that around any point z in the unit disc with $r \leq 1 - |z|$, we have by uniformity that $F_Z^B(z,r) = r^2$. Thus, it is clear that the density with respect to the Lebesgue measure must be $f_Z(z) = \frac{1}{\pi}$.*

On the other hand, attempting to define the rectangular c.d.f. of the complex uniform distribution on the disc requires the calculation of areas of arbitrary distorted quadrants of the disc.

4.4 Independence

Prevalent throughout real statistics, the notion of independence transfers rather directly to complex random variables given we are still working with real probability measures. That is, two complex random variables Z, W are independent if and only if their joint density is given as the product of their marginal densities,

$$f_{Z,W}(z,w) = f_Z(z) f_W(w). \tag{4.4.1}$$

Familiar results in terms of expectations (defined below) e.g. $\mathbb{E}\,(g(Z)h(W)) = \mathbb{E}\,(g(Z))\,\mathbb{E}\,(h(W))$, and conditional probability all hold directly, as despite a complex random variable Z taking complex values, independence is defined in terms

of the probability measure over the sample space Ω, and $\mathbb{P} \circ Z^{-1}$ is still a familiar real probability measure which obeys existing results of conditional probability and independence.

4.5 Moments

When handling the moments of a c.r.v. we will prefer to work within $\mathbb{C}$, taking advantage of complex integration as defined in Section 2.6. We thus define the expectation operator $\mathbb{E}$ as follows.

Definition 4.5.1. *The expected value of a function g of a continuous complex random variable Z with Lebesgue density f_Z is defined as*

$$\mathbb{E}\left(g(Z)\right) = \frac{i}{2}\iint_{\mathbb{C}} g(z) f_Z(z)\,\mathrm{d}z \wedge \mathrm{d}z^*. \tag{4.5.1}$$

One can observe by the identification of integration and the density with the bivariate counterparts that the expectation as defined above is indeed linear. In particular, for $Z = X + iY$,

$$\mathbb{E}\left(Z\right) = \frac{i}{2}\iint_{\mathbb{C}} z f_Z(z)\,\mathrm{d}z \wedge \mathrm{d}z^* = \int_{\mathbb{R}}\int_{\mathbb{R}} (x+iy) f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \mathbb{E}\left(X\right) + i\mathbb{E}\left(Y\right). \tag{4.5.2}$$

However, under the complex algebra, we obtain several variants for higher order moments which coincide in $\mathbb{R}$ or multivariate real statistics due to the lack of complex conjugation. Specifically, one may consider not only k^{th} order moments of the form

$$\mathbb{E}\left(Z^k\right)\ ,\ k \in \mathbb{N}, \tag{4.5.3}$$

but also k^{th}-order moments of the form

$$\mathbb{E}\left(|Z|^k\right)\ ,\ k \in \mathbb{N}, \tag{4.5.4}$$

or even more generally

$$\mathbb{E}\left(Z^m Z^{*n}\right)\ ,\ m,n \in \mathbb{N}, \tag{4.5.5}$$

where $m + n = k$. Of course the restriction $m, n \in \mathbb{N}$ may be relaxed to consider fractional or even complex moments, however this requires more care and is discussed separately in Section 4.6 due to the multivalued nature of such moments.

Clearly, for k even and $m = n = \frac{k}{2}$, one obtains (4.5.4), which is referred to as the k^{th}-order absolute moment,

$$\mathbb{E}\left(Z^m Z^{*m}\right) = \mathbb{E}\left(|Z|^k\right) = \mathbb{E}\left(R^k\right),$$

where R is the real random variable that describes the modulus of Z (Eriksson et al. (2010)). A useful existence theorem from Eriksson et al. (2010, Theorem 5.1, p. 5404) is the following.

Theorem 4.5.2 (Eriksson et al. (2010))**.** *For a complex random variable Z, if any k^{th} order moment exists ($k \in \mathbb{Z}$), then all moments (absolute and other) of order k or less must exist. Furthermore, for all $m, n \in \mathbb{Z}$ such that $m + n = k$,*

$$\left|\mathbb{E}\left(Z^m Z^{*n}\right)\right| \leq \mathbb{E}\left(|Z|^k\right). \tag{4.5.6}$$

The proof is provided in Appendix A.8 being similar, but more explicit, than the brief proof provided in Eriksson et al. (2010).

In higher dimensions, that is for a complex random vector $\boldsymbol{Z} \in \mathbb{C}^n$, moments may be calculated as real random vector moments are calculated. For example at the second moment, we have the second absolute moment matrix given by $\mathbb{E}\left(\boldsymbol{Z}\boldsymbol{Z}^{\mathcal{H}}\right)$ and the second non-absolute moment matrix given by $\mathbb{E}\left(\boldsymbol{Z}\boldsymbol{Z}^{\top}\right)$.

4.5.1 Second-order moments

Often of key interest (primarily in conjunction with the normal distribution) are the second order moments of a distribution. We first define the covariance operator. The covariance between two complex random vectors $\boldsymbol{Z}, \boldsymbol{W} \in \mathbb{C}^n$ is given, similar to the real covariance, however employing the complex field's inner product, by:

$$\mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}, \boldsymbol{W}\right) = \mathbb{E}\left((\boldsymbol{Z} - \mathbb{E}\left(\boldsymbol{Z}\right))(\boldsymbol{W} - \mathbb{E}\left(\boldsymbol{W}\right))^{\mathcal{H}}\right).$$

That is, it is linear in its first argument and conjugate linear in its second, and

$$\mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}, \boldsymbol{W}\right) = \mathbb{C}\mathrm{ov}\left(\boldsymbol{W}, \boldsymbol{Z}\right)^{\mathcal{H}}.$$

Due to the presence also of non-absolute moments, we also define the pseudocovariance operator to be

$$\mathrm{Ps}\mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}, \boldsymbol{W}\right) = \mathbb{E}\left((\boldsymbol{Z} - \mathbb{E}\left(\boldsymbol{Z}\right))(\boldsymbol{W} - \mathbb{E}\left(\boldsymbol{W}\right))^{\top}\right) = \mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}, \boldsymbol{W}^*\right),$$

which is just linear in both arguments.

Considering how to describe the variance of a univariate c.r.v. Z, it is usual to take the second-order *absolute* central moment as the *variance* of Z and call the *non-absolute* central moment the *pseudovariance* (Neeser and Massey (1993); Eriksson et al. (2010)). They will be denoted as,

$$\mathbb{V}\mathrm{ar}\left(Z\right) = \sigma^2 \coloneqq \mathbb{E}\left(\left|Z - \mathbb{E}\left(Z\right)\right|^2\right) = \mathbb{C}\mathrm{ov}\left(Z, Z\right) \quad \in \mathbb{R}^+, \tag{4.5.7}$$

$$\mathrm{Ps}\mathbb{V}\mathrm{ar}\left(Z\right) = \rho^2 \coloneqq \mathbb{E}\left(\left(Z - \mathbb{E}\left(Z\right)\right)^2\right) = \mathrm{Ps}\mathbb{C}\mathrm{ov}\left(Z, Z\right) \in \mathbb{C}. \tag{4.5.8}$$

It is interesting to note that the pseudovariance, which will be considered later under notions of properness in complex distributions, is effectively the covariance of the variable Z with its conjugate Z^*. Unlike much other literature on complex random variables, we will prefer to denote the pseudovariance with a squared term due to the use of its square root in the decomposition of a normal distribution later, however it should be kept in mind that ρ^2 is complex.

Corollary 4.5.3. *The pseudovariance of a random variable is bounded above in modulus by the variance (if it exists). That is, if* $\mathbb{V}\mathrm{ar}(Z) = \sigma^2$ *and* $\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = \rho^2$,

$$|\rho^2| \leq \sigma^2. \tag{4.5.9}$$

Proof. By application of theorem 4.5.2 with $k = m = 2$ on the random variable $Z - \mathbb{E}(Z)$. □

It is worth noting that conventional formulae for the variance and pseudovariance still hold in the complex case, and that they facilitate an understanding of the second-order quantities in terms of the Cartesian real random component variables. If $Z = X + iY$, then we have that

$$\mathbb{V}\mathrm{ar}(Z) = \mathbb{E}\left(|Z|^2\right) - |\mathbb{E}(Z)|^2 = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y), \tag{4.5.10}$$

$$\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = \mathbb{E}\left(Z^2\right) - \mathbb{E}(Z)^2 = \mathbb{V}\mathrm{ar}(X) - \mathbb{V}\mathrm{ar}(Y) + 2i\mathbb{C}\mathrm{ov}(X, Y). \tag{4.5.11}$$

This leads us to develop an interesting proposition off the back of Corollary 4.5.3, specifically concerning when there is equality in (4.5.9).

Proposition 4.5.4. *Let* Z *be a complex random variable with mean* μ*, variance* σ^2 *and pseudovariance* ρ^2*. Then* $|\rho|^2 = \sigma^2$ *if and only if there exists a real random variable* X *with zero mean and unit variance such that* $Z = \rho X + \mu$*. That is,* Z *is distributed along a line in the complex plane parallel to* ρ *from the point* μ*.*

Proof. First handling the forward direction, let $W = Z - \mu$ be the zero mean complex random variable which also has variance σ^2 and pseudovariance ρ^2, denoted $W \sim {}_{\mathbb{C}}(0, \sigma^2, \rho^2)$. Define ρ by its polar coordinates to be $\rho = \sigma e^{i\theta}$, noting that θ is non-random. Then, by conjugate linearity of the covariance and linearity of the pseudocovariance functions, $e^{-i\theta}W \sim {}_{\mathbb{C}}(0, |e^{-i\theta}|^2\sigma^2, e^{-2i\theta}\rho^2) = {}_{\mathbb{C}}(0, \sigma^2, \sigma^2)$. Now, using (4.5.10) and (4.5.11), by the pseudovariance being real and equal to the variance, it must be that $\mathbb{V}\mathrm{ar}(\Im(W)) = 0$, and hence $\Im(W) \overset{\text{a.s.}}{=} 0$ since $\mathbb{E}(W) = 0$. Therefore, $e^{-i\theta}W$ must be distributed along the real axis with mean 0 and variance σ^2. Already, it is clear that therefore by the rotational geometry of the complex multiplication, W must be distributed along the line parallel to $e^{i\theta}$. Specifically, if $X \sim {}_{\mathbb{R}}(0, 1)$ is such that $e^{-i\theta}W{=}\sigma X$, then it must be that $W{=}\sigma e^{i\theta}X = \rho X$. Hence $Z{=}\rho X + \mu$.

Now handling the reverse direction, it is straightforward that if $X \sim {}_{\mathbb{R}}(0, 1)$ and $Z = \rho X + \mu$, then:

$$\begin{aligned}
\mathbb{E}(Z) &= \rho\mathbb{E}(X) + \mu = \mu; \\
\mathbb{V}\mathrm{ar}(Z) &= |\rho|^2\mathbb{E}\left(|X|^2\right) = |\rho|^2; \\
\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) &= \rho^2\mathbb{E}\left(X^2\right) = \rho^2.
\end{aligned}$$

Hence $Z \sim {}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ where $\sigma^2 = |\rho|^2$, which completes the proof. □

This proposition will be useful in expressing a decomposition of the complex normal distribution in the next chapter.

4.6 Principal Expectation

As discussed in Section 2.2, arbitrary non-integer moments of complex numbers are multivalued as a result of the behaviour of the complex logarithm. Thus, a moment such as $\mathbb{E}\left(Z^m Z^{*n}\right)$ may take finitely, or even infinitely, many values whenever $m - n \notin \mathbb{Z}$. Closely linked is the identification of the distribution of Z with the distribution of its polar components $|Z|$ and $\text{Arg}(Z)$. We then overlap with theory of directional statistics in terms of the argument which deserves to be explicitly handled in order to compute non-integer moments of complex random variables.

4.6.1 Polar density identification

Just as the density function of a c.r.v. may be identified with the Cartesian bivariate density, one would expect, by elementary methods of calculus converting x, y to r, θ, to also be able to identify the distribution with that of the polar components, i.e. the modulus and argument. However, as discussed in Section 2.2, we need to be clear about which form of the argument we choose. Conventionally, we will choose to identify the distribution of the complex random variable with the joint distribution of its modulus $|Z|$ and principal argument $\text{Arg}(Z)$, which takes value only in $(-\pi, \pi]$. By the fact that the Lebesgue measure corresponds to $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$ where $x = r\cos(\theta)$ and $y = r\sin(\theta)$, then we have that the bivariate density of $|Z|, \text{Arg}(Z)$ is given by

$$f_{|Z|,\text{Arg}(Z)}(r,\theta) = r f_Z(re^{i\theta})\,,\ r \in \mathbb{R}^+,\ \theta \in (-\pi, \pi]. \tag{4.6.1}$$

Whilst the above identification holds directly for continuous and contour random variables, if we have a distribution such that $\mathbb{P}\left(Z = 0\right) > 0$, then one needs to be careful, as formally 0 does not have a defined argument. Note, however that for discrete random variables, we will tend to have a list of probabilities over these discretised outcomes, rather than a continuous function in r and θ as above.

Again, the choice of the argument interval is arbitrary and one could easily redefine the interval and obtain a new distribution. The conversion of the angular distribution between argument intervals may be done by wrapping the marginal distribution of the argument. More generally, if we have an arbitrary distribution of some real random variables R and Θ and define $Z = Re^{i\Theta}$, then

$$f_{|Z|}(r) = f_R(r), \tag{4.6.2}$$

$$f_{\text{Arg}_\alpha(Z)}(\theta) = \sum_{k\in\mathbb{Z}} f_\Theta(\theta + 2k\pi)\,,\ \theta \in (\alpha, \alpha + 2\pi], \tag{4.6.3}$$

recalling that Arg_α takes the equivalent angle in the interval $(\alpha, \alpha + 2\pi]$ and the principal argument $\text{Arg} = \text{Arg}_{-\pi}$ (Pewsey and García-Portugués (2021)).

Thus we observe that in fact the distribution of the argument enters the realms of directional statistics. In directional statistics, the distribution of a continuous real random variable describing an angle may be defined via a continuous density function f which is 2π-periodic such that

$$\int_\alpha^{\alpha+2\pi} f(\theta)\mathrm{d}\theta = 1$$

for all $\alpha \in \mathbb{R}$ (Mardia and Jupp (2000); Pewsey and García-Portugués (2021)). In this regard, the density does not integrate to 1 over the entire real line, but rather in any interval of length 2π. This indicates the requirement that for it to be truly continuous, the distribution on any interval of length 2π must have the same start and endpoints. However an important distinction we will make here is between the distribution of $\arg(Z)$, which is 2π-periodic as described above, and that of $\mathrm{Arg}(Z)$, which by definition of the Arg function is a variable that only takes value in $(-\pi, \pi]$ and thus the density of $\mathrm{Arg}(Z)$ will be taken to be zero outside this interval.

Example 4.6.1 (Kotlarski (1965))**.** *If Θ follows a triangular distribution on $[-2\pi, 2\pi]$, and is independent of R which has density f_R, then the random variable $Z = Re^{i\Theta}$ has distribution*

$$f_Z(re^{i\theta}) = \frac{1}{2\pi r} f_R(r). \tag{4.6.4}$$

This indicates that $\mathrm{Arg}(Z)$ has a uniform distribution on $(-\pi, \pi]$, i.e. wrapping the triangular distribution over $[-2\pi, 2\pi]$ onto an interval of length 2π yields a uniform distribution.

One key concern arises then in terms of complex moments of the form as given above

$$\mathbb{E}\left(Z^m Z^{*n}\right) = \mathbb{E}\left(|Z|^{m+n} e^{i(m-n)(\mathrm{Arg}(Z)+2k\pi)}\right) \tag{4.6.5}$$

$$= e^{i2k\pi(m-n)} \mathbb{E}\left(|Z|^{m+n} e^{i(m-n)\mathrm{Arg}(Z)}\right), \tag{4.6.6}$$

for all $k \in \mathbb{Z}$ by the multivalued nature of complex powers. We note that, if $|Z|$ and $\mathrm{Arg}(Z)$ were independent, then the expectation would split to yield the $(m+n)^{\text{th}}$ moment of $|Z|$ and what appears to be the characteristic function of $\mathrm{Arg}(Z)$. However, as is standard in directional statistics (Mardia and Jupp (2000); Pewsey and García-Portugués (2021)), the characteristic function of a directional random variable Θ is defined to only converge for integer inputs i.e. $m - n \in \mathbb{Z}$, which is where $(e^{i\Theta})^{m-n}$ is single valued. This can be seen by the first term in (4.6.6) becoming 1 for all $k \in \mathbb{Z}$ if $(m-n)$ is an integer. However, this is not to say that we cannot define moments of a complex random variable for arbitrary $m, n \in \mathbb{C}$, rather that as with complex analysis if we treat the multiple values of each power as an equivalence class, we can then isolate a single value in each equivalence class to uniquely characterise the moments. As with the introduction of the principal argument and principal logarithm functions, we thus introduce the principal expectation.

4.6.2 Principal expectation operator

The principal expectation operator is defined to take the expectation of a complex random variable Z by treating it in polar form as the random variable $|Z|e^{\mathrm{Arg}(Z)}$, specifically, such that the argument is a bounded real random variable taking value only on $(-\pi, \pi]$.

Definition 4.6.2. *We define the principal expectation operator* $\mathfrak{P}\mathbb{E}$ *to be*

$$\mathfrak{P}\mathbb{E}_Z\left(g(Z)\right) = \int_{-\pi}^{\pi}\int_0^{\infty} g(re^{i\theta}) f_Z(re^{i\theta})\, r\,\mathrm{d}r\,\mathrm{d}\theta. \tag{4.6.7}$$

The random variable whose principal argument is to be taken is denoted by the subscript.

The principal expectation is thus a single-valued quantity for all arbitrary moments of a complex random variable Z. This operator will make certain computations regarding the Mellin transform simpler by allowing the use of integrals instead of summations for calculation and inversion.

Basic properties of the principal expectation are that:

1. For all constants $\omega \in \mathbb{C}$,

$$\mathfrak{P}\mathbb{E}_Z\left(\omega g(Z)\right) = \omega \mathfrak{P}\mathbb{E}_Z\left(g(Z)\right). \tag{4.6.8}$$

2. For all single-valued complex functions g,

$$\mathfrak{P}\mathbb{E}_Z\left(g(Z)\right) = \mathbb{E}\left(g(Z)\right), \tag{4.6.9}$$

   where $\mathbb{E}$ is the usual expectation operator.

3. In particular, for all $m, n \in \mathbb{C}$ such that $m - n \in \mathbb{Z}$,

$$\mathfrak{P}\mathbb{E}_Z\left(Z^m Z^{*n}\right) = \mathbb{E}\left(Z^m Z^{*n}\right). \tag{4.6.10}$$

4. For all $m, n \in \mathbb{C}$,

$$\mathbb{E}\left(Z^m Z^{*n}\right) = e^{i(m-n)2\pi k} \mathfrak{P}\mathbb{E}_Z\left(Z^m Z^{*n}\right)\ ,\ \forall\, k \in \mathbb{Z}. \tag{4.6.11}$$

5. For R, Θ, independent real random variables with $R \geq 0$ and respective densities f_R, f_Θ, then if $Z = Re^{i\Theta}$,

$$\mathfrak{P}\mathbb{E}_Z\left(Z^m Z^{*n}\right) = \mathbb{E}\left(R^{m+n}\right) \sum_{k\in\mathbb{Z}} e^{i(m-n)2\pi k} \int_{(2k-1)\pi}^{(2k+1)\pi} e^{i(m-n)\theta} f_\Theta(\theta)\,\mathrm{d}\theta. \tag{4.6.12}$$

   In particular, if Θ is distributed over $(-\pi, \pi]$, or $(m - n) \in \mathbb{Z}$, then the summation term coincides with the characteristic function of Θ evaluated at $(m - n)$, which is why in directional statistics the characteristic function is given by a series evaluating the c.f. of Θ at integers (Loomis (1953, p. 141); Bochner (1955, p. 73); Mardia and Jupp (2000, pp. 26–28); Pewsey and García-Portugués (2021)). Here, however, we will not make such a restriction, using instead the principal expectation as it is defined.

6. If the random variable Z is such that $\text{Arg}(Z)$ is distributed uniformly on $(-\pi, \pi]$ independent of $|Z|$, then so is $\text{Arg}(e^{i\alpha}Z)$ and

$$\mathfrak{E}_{e^{i\alpha}Z}\left(Z^m Z^{*n}\right) = e^{-i\alpha(m-n)}\mathfrak{E}_Z\left(Z^m Z^{*n}\right), \tag{4.6.13}$$

for all $\alpha \in \mathbb{R}$.

7. For a complex random variable Z, if we define $W = |Z|^s e^{is\text{Arg}(Z)}$ for some $s \in \mathbb{C}$, it may be that

$$\mathfrak{E}_Z\left(Z^s\right) \neq \mathfrak{E}_W\left(W\right), \tag{4.6.14}$$

exactly how we may have that $\text{Arg}(W) \neq s\text{Arg}(Z)$.

One can think of the principal expectation as the expectation of Z as it is distributed on the Riemann surface of the complex logarithm (pictured as a ray spiralling in $\mathbb{R}^3$ where the height is the (non-principal) argument) between heights of $-\pi$ and π, which is homeomorphic to the whole complex plane (Kotlarski (1965)).

Alternatively, a complex power establishes an equivalence classes over $\mathbb{C}$ containing all the multiple values (and expectations thereof), and the principal expectation evaluates only a single value of this equivalence class. All other values can be subsequently found for different values of k in property 4 above. In this sense, the principal expectation uniquely determines the general expectation of a complex power. Thus, applying an analogue of the Mellin transform with the principal expectation operator will uniquely determine the distribution, but with single-valued calculations across a continuous region of $\mathbb{C}^2$, unlike the more general Mellin transform which is single-valued on a partial lattice. Both transforms are presented in Section 4.8.

4.7 Characteristic Function

The c.f. of a real random variable, X, is defined as the function $\varphi(t) = \mathbb{E}\left(e^{itX}\right)$, which itself is simply the first moment of a complex random variable that takes values on the unit circle (Lukacs (1970); Ushakov (1999)). The origins of the c.f. stem from integral transforms of the density, specifically being the Fourier transform of the density with respect to the Lebesgue measure. The c.f. for real random variables always exists (for real t) and exhibits three very useful properties:

1. (*Moment generating property*) If the c.f. is k-times differentiable at 0, then all moments of order $2\left\lfloor\frac{k}{2}\right\rfloor$ exist, and $\mathbb{E}\left(X^m\right) = (-i)^m \frac{\partial\varphi}{\partial t}\big|_{t=0}$ (Ushakov (1999)).

2. (*Unique characterisation*) The c.f. always exists and uniquely characterises the distribution of X. That is, if $\varphi_X(t) = \varphi_Y(t)$ for all $t \in \mathbb{R}$, then X and Y have the same distribution.

3. (*Distribution of sum*) The distribution of the sum of independent random variables X and Y has c.f. equal to the product of the characteristic functions, that is $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

It is argued here that these are the three most useful properties of the characteristic function which one wishes to retain in the complex analogue. A naïve guess may be to consider for c.r.v. Z, the characteristic function defined by $\mathbb{E}\left(e^{itZ}\right)$, as used by Hinds (1974). However immediately, the moment generating property breaks down as one cannot recover, through Wirtinger derivatives or otherwise, moments involving Z^*. Aiming to satisfy such a property, and keeping in mind the real scalar inner product in $\mathbb{C}$, we give the following definition.

Definition 4.7.1 (Characteristic Function)**.** *The characteristic function of a complex random variable Z is defined as the function $\varphi_Z : \mathbb{C} \to \mathbb{C}$ satisfying*

$$\varphi_Z(t) = \mathbb{E}\left(e^{\frac{i}{2}(t^*Z + tZ^*)}\right),\ t \in \mathbb{C}. \tag{4.7.1}$$

This definition coincides with the real c.f. when Z and t are purely real. Furthermore, employing the Wirtinger derivatives (where they exist) with respect to $t \in \mathbb{C}$, it can be seen that

$$\mathbb{E}\left(Z^m Z^{*n}\right) = (-2i)^{n+m} \frac{\partial^{n+m}\varphi_Z}{\partial t^n \partial t^{*m}}\bigg|_{t=0}, \tag{4.7.2}$$

which satisfies the moment generating property for a c.r.v..

Upon further investigation of the c.f. proposed in (4.7.1), one can note that $\frac{i}{2}(t^*Z + tZ^*) = i\Re(t^*Z) = i(t_1X + t_2Y)$, for $t = t_1 + it_2$. In fact, this characteristic function definition, which in $\mathbb{C}$ retains the desirable moment generating property, is directly equivalent to the bivariate real characteristic function if t and Z were represented as vectors in $\mathbb{R}^2$. By equivalence of the density with the corresponding bivariate density as demonstrated in Subsection 4.3, it follows that the complex c.f. must also satisfy the unique characterisation property. Finally, as complex addition is equivalent to addition in $\mathbb{R}^2$ (or by direct expansion of the definition in (4.7.1)) the c.f. of the sum of independent complex random variables is just the product of their c.f.'s.

In higher dimensions the c.f. can be similarly defined for the complex random vector $\boldsymbol{Z} \in \mathbb{C}^n$ by letting

$$\varphi_{\boldsymbol{Z}}(\boldsymbol{t}) = \mathbb{E}\left(e^{i\Re(\boldsymbol{t}^{\mathcal{H}}\boldsymbol{Z})}\right), \tag{4.7.3}$$

for $\boldsymbol{t} \in \mathbb{C}^n$.

Note the complex c.f. described in (4.7.1) is not new and already in use in the literature (see Wooding (1956); Andersen et al. (1995); Giri (2004); Eriksson et al. (2010); Schreier and Scharf (2010); Ducharme et al. (2016)), however here we have presented a justification for this function from a complex perspective, rather than directly translating the $\mathbb{R}^2$ form of the characteristic function into complex notation. For completeness, where the characteristic functions of $\Re(Z)$ and $\Im(Z)$ can be extended by analytic continuation to complex domains, the complex characteristic function definition can be broadened to what we call the 'generalised characteristic function' $\varphi_Z(s,t) = \mathbb{E}\left(\exp\left\{i(sZ + tZ^*)\right\}\right)$ (similar to what was originally used in Spitzer (1955)), where $\varphi_Z(t^*/2, t/2)$ recovers the $\varphi_Z(t)$ as per (4.7.1).

4.7.1 Inverse transform

As the c.f. uniquely defines the distribution of a random variable, we can recover the density from its c.f.. In particular, the c.f. is the Fourier transform of a density f_Z of the distribution of a c.r.v. Z. Thus we may use the inverse Fourier transform to yield the corresponding density function (Bracewell (1965, p. 241); Lukacs (1970)). If, for $z = x + iy$,

$$\varphi_Z(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{i}{2}(t^*z+tz^*)} f_Z(z)\,\mathrm{d}x\,\mathrm{d}y, \tag{4.7.4}$$

then for $t = u + iv$

$$f_Z(z) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{i}{2}(t^*z+tz^*)} \varphi_Z(t)\,\mathrm{d}u\,\mathrm{d}v. \tag{4.7.5}$$

Note that the density need not be with respect to the Lebesgue measure for the above result to hold.

4.7.2 Moment generating function

As a small mention, due to its prevalence in real probability and statistics, we could consider the complex analogue of the moment generating function (m.g.f.) which is essentially the Laplace transform of the distribution. Direct translation from $\mathbb{R}^2$ and similar construction to above may suggest a function of the form $m_Z(s) = \mathbb{E}\left(e^{\frac{1}{2}(s^*Z+sZ^*)}\right)$ to coincide with the real m.g.f., which one may recognise using the generalised c.f. as $\varphi_Z(-is^*/2, -it/2)$. Alternatively, considering the approach of taking $m_Z(s) = \varphi_Z(-is)$ instead yields the function $m_Z(s) = \mathbb{E}\left(e^{\frac{1}{2}(sZ^*-s^*Z)}\right)$. Whilst moments are equally recoverable from this second function for Z and s both complex numbers, if Z and s are both purely real it trivially to 1. Overall, whilst the former proposed m.g.f. derived from the generalised c.f. would be preferred, given the complex nature of Z and the variable s itself, we will avoid using the m.g.f. in favour of the c.f. for any necessary calculations.

4.8 Mellin Transform

So far, many results for the distribution of random variables are analogous to the results of bivariate statistics, translated into complex notation with the help of the Wirtinger derivatives. However, as has been mentioned numerous times, a crucial difference between $\mathbb{C}$ and $\mathbb{R}^2$ that motivates the use of complex numbers in practice is the convenient multiplication in $\mathbb{C}$. As such, given the characteristic function describes the distribution of a sum of c.r.v.'s, we now turn to describing the distribution of the product of c.r.v.'s, something which does not elicit an analogous extension from $\mathbb{R}^2$.

4.8.1 The real Mellin transform

Whereas the sum of independent real random variables is closely linked to the c.f. description of their distribution, the distribution of their product is closely related to the Mellin transform.

Definition 4.8.1 (Galambos and Simonelli (2004))**.** *The Mellin transform of a function f defined on domain $\mathbb{R}^+$ is defined as*[1]

$$\mathcal{M}(f;s) = \int_0^\infty x^s f(x)\,\mathrm{d}x, \tag{4.8.1}$$

for all complex $s \in \mathcal{S}_f$, being the vertical band of convergence $\{s \in \mathbb{C} : a < \Re(s) < b\}$ for some $a, b \in \mathbb{R}$ so that the integral in (4.8.1) converges and is holomorphic.

The couple $\mathcal{M}(f;\cdot)$ and $\mathcal{S}_f$ uniquely define the function f via the inverse Mellin transform

$$f(x) = \frac{1}{2\pi i}\int_{\Re(s)=c} \mathcal{M}(f,s)x^{-s-1}\,\mathrm{d}s, \tag{4.8.2}$$

where $c \in \mathbb{R}$ lies in the vertical strip described by $\mathcal{S}_f$ (Galambos and Simonelli (2004)).

In the context of probability and statistics, we consider the transformed function f to be a probability density f_X of a real random variable X and denote the Mellin transform $\mathcal{M}(f_X, s)$ simply by $\mathcal{M}_X(s)$. We will call $\mathcal{M}_X(s)$ the Mellin transform of X. As the density of a random variable integrates to 1, we have that the imaginary axis $\Re(s) = 0$ always lies in the band of convergence $\mathcal{S}_f$. The convergence band parameters a, b are defined by the behaviour of X in the tails of its distribution, i.e. near 0 and as it tends to ∞. Specifically, $a \leq 0$ is given by $\inf\left\{k \in \mathbb{R} : \mathbb{E}\left(X^k\right) < \infty\right\}$, and conversely $b \geq 0$ can be thought of as $\sup\left\{k \in \mathbb{R} : \mathbb{E}\left(X^k\right) < \infty\right\}$. We note that the Mellin transform presented above of a density only applies for positive random variables, however Epstein (1948) proposed the extension of the Mellin transform to an arbitrary random variable by considering the Mellin transform on the distribution of the positive and negative parts of real random variable X. That is, the distributions of $X_+ = \max\{X, 0\}$ and $X_- = -\min\{0, X\}$. This definition was extended in Galambos and Simonelli (2004, p.16) such that $\mathcal{M}_X(s) = \mathcal{M}_{X^+}(s) + \gamma\mathcal{M}_{X^-}(s)$ for a formal indeterminate γ such that $\gamma^2 = 1$ (often $(-1)^s$).

Applications of the Mellin transform in real probability and statistics have been well explored such as in: Epstein (1948); Zolotarev (1957); Bertrand et al. (2010); Cottone and Di Paola (2009); and Geenens (2021), among others. The more theoretical properties can also be found in most books covering the properties of functions of complex variables or integral transforms (see e.g. Bracewell (1965); Remmert (1991); Ablowitz and Fokas (2003); Galambos and Simonelli (2004)).

The main properties of interest are:

1. (*Characteristic function of logarithm*) for a real random variable X, its Mellin transform $\mathcal{M}_X(is) = \varphi_{\ln(X)}(s)$, where $\varphi_{\ln(X)}$ is the characteristic function of $\ln(X)$ (Galambos and Simonelli (2004)).

[1]Most authors define the integrand with x^{s-1} rather than x^s, however we prefer the x^s form that is given in Galambos and Simonelli (2004) for its coincidence with the characteristic function.

2. (*Imaginary axis in band of convergence*) The Mellin transform of a probability distribution f_X has the imaginary axis ($\Re(s) = 0$) in its band of convergence $\mathcal{S}_{f_X}$ (Zolotarev (1957)).

3. (*Moment description*) For any real random variable X, one can directly calculate the moments as $\mathbb{E}(X^s) = \mathcal{M}_X(s) = \mathcal{M}_{X^+}(s) + (-1)^s \mathcal{M}_{X^-}(s)$ for $s \in \mathcal{S}_{f_X}$ (Galambos and Simonelli (2004, p.23)).

4. (*Distribution of a product of random variables*) For independent real random variables X and Y, the distribution of their product can be uniquely determined by $\mathcal{M}_{XY}(s) = \mathcal{M}_X(s)\mathcal{M}_Y(s)$.

4.8.2 The complex Mellin transform

Extending now to complex random variables, it is less clear than the development of the c.f. as the advantage of the Mellin transform's unique characterisation of a distribution comes from the ability for the parameter s to be non-integer, and thus calculate fractional and complex moments (Cottone and Di Paola (2009)). Nonetheless, we can still generalise the Mellin transform to complex random variables, keeping in mind the moment generating property and inspired by the relationship to the generalised characteristic function we have already defined. Doing so, we arrive at the Mellin transform generalisation that has in fact been around for over half a century, presented in Kotlarski (1965) in terms of the random variable's polar decomposition and further addressed by Brock and Krutchkoff (1970), but unexplored since (to the best of our knowledge). In particular, neither Kotlarski (1965), nor Brock and Krutchkoff (1970), provided an explicit inversion formula by which one can recover a density of a continuous random variable from the Mellin transform. We present this inversion through new formulae. We also make an important distinction between the general Mellin transform and the principal Mellin transform — a distinction made clear thanks to the principal expectation operator.

Definition 4.8.2 (Mellin Transform)**.** *The Mellin transform of a complex random variable Z is given by*

$$\mathcal{M}_Z(s,t) = \mathbb{E}\left(Z^s Z^{*t}\right) \tag{4.8.3}$$

for $s, t \in \mathbb{C}$ such that the integral converges.

The version from Kotlarski (1965) was given in terms of the polar coordinates R, Θ, using complex variables $u = s + t$ and $v = s - t$. Kotlarski (1965) also chose to consider $\Theta = \text{Arg}(Z)$, not $\arg(Z)$ such that their Mellin transform was a single valued object. Here, with the aid of the principal Mellin transform, we make a distinction between what is called the Mellin transform as defined above in Definition 4.8.2, which is multivalued whenever $s - t \notin \mathbb{Z}$, and the principal Mellin transform defined below, which coincides with that in Kotlarski (1965) and is single valued for all $s, t \in \mathbb{C}$. Both types of Mellin transform uniquely define the distribution of a continuous random variable Z, as will be shown through explicit inversion formulae. The intuition for why unique characterisation of the distribution holds

for both comes from the discussion on the principal expectation (Subsection 4.6.2), where the distribution of Z on the complex plane may be uniquely determined by its distribution on a 2π 'tall' section of the Riemann surface of the complex logarithm. This was also the argument given in Kotlarski (1965) to justify the use of $\operatorname{Arg}(Z)$ as a variable only in $(-\pi, \pi]$, which subsequently does not need to be treated as in directional statistics with a periodic density.

Definition 4.8.3 (Principal Mellin Transform)**.** *The principal Mellin transform of a complex random variable Z with $R = |Z|$ and $\Theta = \operatorname{Arg}(Z)$ is given by*

$$\mathcal{M}_Z^{\mathrm{P}}(s,t) = \mathbb{P}\mathbb{E}_Z\left(Z^s Z^{*t}\right) \tag{4.8.4}$$

$$= \mathbb{P}\mathbb{E}_Z\left(R^{s+t} e^{i(s-t)\Theta}\right) \tag{4.8.5}$$

$$= \int_{-\pi}^{\pi} \int_0^{\infty} r^{s+t+1} e^{i(s-t)\theta} f_Z(re^{i\theta}) \,\mathrm{d}r \,\mathrm{d}\theta, \tag{4.8.6}$$

for all $s, t \in \mathbb{C}$ such that the integral converges.

The principal Mellin transform is now a single-valued function for arbitrary $s, t \in \mathbb{C}$, where it exists. Furthermore, the principal Mellin transform also has the familiar correspondence to the characteristic function of the principal logarithm of Z, where for any t such that the Mellin transform converges,

$$\mathcal{M}_Z^{\mathrm{P}}\left(\frac{i}{2}t^*, \frac{i}{2}t\right) = \varphi_{\mathrm{Log}Z}(t).$$

In fact, both types of Mellin transform may be seen to satisfy properties 1, 3–4 from the properties of the real Mellin transform, where in terms of relating to the characteristic function of the logarithm, we now need to choose the principal logarithm for the principal Mellin transform. Regarding the property 2 about the imaginary axis in the region of convergence, we turn to discuss the region of convergence for the complex Mellin transform, which will now be a subset of $\mathbb{C}^2$.

4.8.3 Region of convergence

To begin the discussion of the region of convergence, we first remark that the (non-principal) Mellin transform is single-valued only for $s - t \in \mathbb{Z}$. This will be important in terms of inverting the transform later, similar to how in directional statistics, the characteristic function of an angular random variable is given by a characteristic series, rather than function — additional details on this topic regarding directional statistics are provided in Appendix A.9 (Mardia and Jupp (2000); Pewsey and García-Portugués (2021)). In this regard, we will begin with the region of convergence of the principal Mellin transform, which is a continuous connected region in $\mathbb{C}^2$, which is also the region of convergence of the more general Mellin transform by the relation of the principal expectation to the multiple values of the usual expectation.

The principal Mellin transform coincides with the bivariate characteristic function of $\ln(R), \Theta$ where $R = |Z|$ and $\Theta = \text{Arg}(Z)$ (Kotlarski (1965)), such that

$$\mathcal{M}_Z^{\text{P}}\left(\frac{ig+h}{2}, \frac{ig-h}{2}\right) = \varphi_{\ln(R),\Theta}(g,h). \tag{4.8.7}$$

Hence, we can understand the region of convergence of the Mellin transform in terms of the complex region of convergence in terms of the characteristic function, which by Theorem 7.1.1 from Lukacs (1970, p. 193), (4.8.7) converges for g and h each in horizontal strips of the complex plane. However, furthermore, by consequence of Theorem 7.2.2 from Lukacs (1970, p. 200), by the principal argument being bounded in $(-\pi, \pi]$, we have that h may in fact be anywhere in the complex plane. Thus we have that the region of convergence is only determined by the convergence of R, that is the convergence of the Mellin transform of the distribution of depends upon the behaviour of the random modulus. In particular, let

$$\alpha = \inf\left\{k \in \mathbb{R} : \mathbb{E}\left(|Z|^k\right) < \infty\right\}, \tag{4.8.8}$$

$$\beta = \sup\left\{k \in \mathbb{R} : \mathbb{E}\left(|Z|^k\right) < \infty\right\}. \tag{4.8.9}$$

That is, $\alpha \leq 0$ is determined by how the modulus distribution behaves near zero and $\beta \geq 0$ is determined by the heaviness of the tail of the modulus distribution out to infinity. The region of convergence for the complex Mellin transform is thus given by

$$\left\{(s,t) \in \mathbb{C}^2 : \alpha < \Re(s+t) < \beta\right\}. \tag{4.8.10}$$

Alternatively, one can consider that for each $s \in \mathbb{C}$, we require $t = \omega - s^*$ for a constant $\omega \in \mathbb{C}$ satisfying $\alpha < \Re(\omega) < \beta$. We depict the two interpretations of the region of convergence below imposed by the conditions. In four dimensions, this region corresponds to a dense stack of hyperplanes which when projected onto $\Re(s)\Re(t)$-space densely cover the strip in Figure 4.1(a).

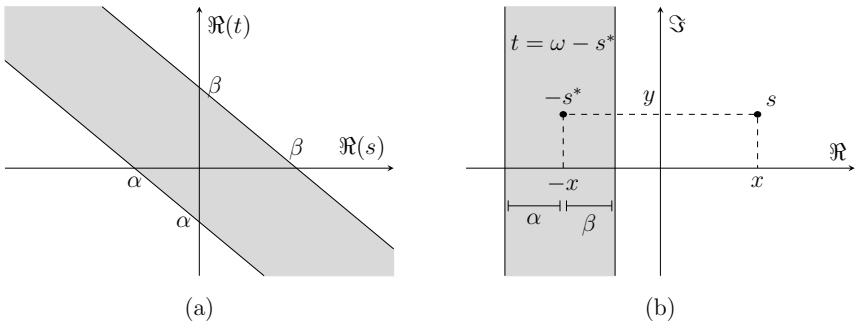

Figure 4.1: Region of convergence of the complex Mellin transform (a) in $\Re(s)\Re(t)$-space, and (b) for a fixed $s \in \mathbb{C}$.

The general Mellin transform has the same convergence region above, however it is only single-valued on a lattice of points within the strips such that $s - t \in \mathbb{Z}$, i.e. $\Im(s) = \Im(t)$ and $\Re(s) - \Re(t) \in \mathbb{Z}$.

4.8.4 Properties

We give now, several useful properties of the complex Mellin transform in terms of finding distributions of functions of complex random variables, as explored in Brock and Krutchkoff (1970). In the following we will work with complex random variables Z and W, and their respective Mellin transforms $\mathcal{M}_Z$ and $\mathcal{M}_W$. It is noted that the following also hold for the principal Mellin transform.

1. $\mathcal{M}_{Z^*}(s,t) = \mathcal{M}_Z(t,s)$.
2. For Z, W independent, $\mathcal{M}_{ZW}(s,t) = \mathcal{M}_Z(s,t)\mathcal{M}_W(s,t)$.
3. $\mathcal{M}_{1/Z}(s,t) = \mathcal{M}_Z(-s,-t)$.
4. For some $\gamma \in \mathbb{Z}$, if $W = Z^\gamma$, then $\mathcal{M}_W(s,t) = \mathcal{M}_Z(\gamma s, \gamma t)$.

4.8.5 Inversion formulae

Previous studies of the complex Mellin transform, namely Kotlarski (1965) and Brock and Krutchkoff (1970) utilised the Mellin transform in specific applications with the circular complex normal distribution to display how the Mellin transform may be used to identify the distribution of powers and products of complex random variables. However, their identification of power and product distributions relies upon first identifying known Mellin transforms. Whilst this is commonly done in real probability and statistics when using characteristic functions or the Mellin transform to identify product distributions, we also have direct inversion formulae which allow us to recover the density via an integral where required. Below, we provide inversion formulae for the Mellin transform and principal Mellin transform of the density of a continuous complex random variable. We first present the inverse for the general Mellin transform, which in many cases will be less wieldy than the principal Mellin transform inversion, specifically due to the issue of the general Mellin transform being single-valued on a lattice of points. An example from Kotlarski (1965) is then solved using these new formulae which also highlights a situation where the principal expectation and principal Mellin transform prove to be more useful than the general Mellin transform.

Theorem 4.8.4. *Let Z be a continuous random variable with a density f_Z with respect to some measure. Denote the Mellin transform and principal Mellin transforms of this density by $\mathcal{M}_Z$ and $\mathcal{M}_Z^{\mathrm{P}}$ respectively with regions of convergence for $\alpha, \beta \in \mathbb{R}$ as given above. Then for an arbitrary constant $c \in (\alpha, \beta)$, we can invert the Mellin transform by*

$$f_Z(z) = \frac{1}{4\pi^2 i} \int_{c-i\infty}^{c+i\infty} \sum_{v \in \mathbb{Z}} \mathcal{M}_Z\left(\frac{u+v}{2}, \frac{u-v}{2}\right) z^{-\frac{u+v+2}{2}} z^{*-\frac{u-v+2}{2}} \,\mathrm{d}u. \tag{4.8.11}$$

Regarding the principal Mellin transform, we can invert it by taking

$$f_Z(z) = \frac{i}{2\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{M}_Z^{\mathrm{P}}(s, \omega - s^*) z^{-s-1} z^{*-(\omega-s^*)-1} \,\mathrm{d}s \wedge \mathrm{d}s^*, \tag{4.8.12}$$

for $s = x + iy$ and ω an arbitrary complex constant satisfying $\Re(\omega) \in (\alpha, \beta)$. In the above, it should be understood that $z = re^{i\theta}$ where $\theta \in (-\pi, \pi]$.

The proof that these formulae do indeed invert the function is given in Appendix A.10. In particular, we notice that the principal Mellin transform inverse formula resembles the real inverse, where we take $\frac{i}{2\pi^2}\iint \mathcal{M}_Z^{\mathrm{P}}(s,t)z^{-s-1}z^{*-t-1}\mathrm{d}s\wedge\mathrm{d}s^*$ with $t = \omega - s^*$, as per the interpretation of the region of convergence in Figure 4.1(b).

We now present the example of finding the distribution of a reciprocal circular complex normal distribution, as given in Kotlarski (1965). However, unlike Kotlarski (1965) where particular cases needed to be established to recognise resulting Mellin transforms, we display how the inversion formula facilitates the direct recovery of the density without having established particular cases prior, and provide the general result for arbitrary variance of the distribution instead of assuming unit variance as in Kotlarski (1965). The following calculations will make use of Cauchy's residue theorem from complex analysis. Whilst we believe sufficient detail is given, the exact theorem will not be explicitly stated here.

Example 4.8.5. *Let Z be a circular complex normal with zero mean and variance σ^2. The density of Z with respect to the Lebesgue measure is given by*

$$f_Z(z) = \frac{1}{\pi\sigma^2}e^{-\frac{|z|^2}{\sigma^2}}\ ,\ z \in \mathbb{C},$$

and the corresponding principal Mellin transform is (proof in Proposition 5.4.2),

$$\mathcal{M}_Z^{\mathrm{P}}(s,t) = \frac{\sin(\pi(s-t))}{\pi(s-t)}\sigma^{s+t}\Gamma\left(\frac{s+t+2}{2}\right),$$

for all $s, t \in \mathbb{C}$ such that $s + t \notin \mathbb{Z}^- = \{-1, -2, \ldots\}$. This requirement is just to ensure that the Gamma function converges. If $s = t$, one should evaluate the above by taking the limit of $s - t \to 0$, such that the first fraction term becomes unity.

Immediately, we have an expression for all moments of the complex normal distribution, for example the absolute fourth moment (kurtosis) is given by $s = t = 2$, resulting in $2\sigma^4$. Nonetheless, suppose we wish to obtain the distribution of $\frac{1}{Z}$ (as in Section 4.1 of Kotlarski (1965)). Then we have that

$$\mathcal{M}_{1/Z}^{\mathrm{P}}(s,t) = \mathcal{M}_Z^{\mathrm{P}}(-s,-t) = \frac{\sin(\pi(t-s))}{\pi(t-s)}\sigma^{-(s+t)}\Gamma\left(\frac{2-s-t}{2}\right).$$

We may now directly invert to obtain the density $f_{1/Z}(z)$ per (4.8.12). Letting $s = x + iy$, $\omega = 0$ (which ensures convergence) and understanding $z = re^{i\theta}$, then

$$\begin{aligned} f_{1/Z}(z) &= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathcal{M}^{\mathrm{P}}_{1/Z}(s,-s^*)z^{-s-1}z^{*s^*-1}\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\sin(2\pi x)}{2\pi x}\sigma^{-2iy}\Gamma\left(1-iy\right)r^{-2iy-2}e^{-2ix\theta}\,\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{r^2\pi^2}\int_{-\infty}^{\infty}\frac{\sin(2\pi x)}{2\pi x}e^{-2ix\theta}\,\mathrm{d}x\int_{-\infty}^{\infty}(\sigma r)^{-2iy}\int_0^{\infty}m^{-iy}e^{-m}\,\mathrm{d}m\,\mathrm{d}y. \end{aligned}$$

Now, given the Gamma function and integral over y both converge, we swap the order of integration in y and m to obtain,

$$f_{1/Z}(re^{i\theta}) = \frac{1}{r^2\pi^2}\int_{-\infty}^{\infty}\frac{\sin(2\pi x)}{2\pi x}e^{-2ix\theta}\,\mathrm{d}x\int_0^{\infty}\int_{-\infty}^{\infty}(m\sigma^2r^2)^{-iy}\,\mathrm{d}y\,e^{-m}\,\mathrm{d}m. \quad (4.8.13)$$

We now make use of the Dirac delta function defined such that

$$\delta(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{iux}\,\mathrm{d}x,$$

and satisfies

$$\int_{-\infty}^{\infty}\delta(u)g(u)\,\mathrm{d}u = g(0).$$

Explicitly substituting $p = \ln(m)$ and $q = \ln(\sigma^{-2}r^{-2})$, and also letting $y \mapsto -y$ in the integral we have,

$$\begin{aligned} \int_0^{\infty}\int_{-\infty}^{\infty}(m\sigma^2r^2)^{-iy}\,\mathrm{d}y\,e^{-m}\,\mathrm{d}m &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{iy(p-q)}\,\mathrm{d}y\,e^{-e^p}e^p\,\mathrm{d}p \\ &= \int_{-\infty}^{\infty}2\pi\delta(p-q)e^{-e^p}e^p\,\mathrm{d}p \\ &= 2\pi e^{-e^q}e^q \\ &= \frac{2\pi}{r^2\sigma^2}e^{-\frac{1}{r^2\sigma^2}}. \end{aligned} \quad (4.8.14)$$

We also now use Cauchy's residue theorem to handle the integral over x, whereby

$$\begin{aligned} \int_{-\infty}^{\infty}\frac{\sin(2\pi x)}{2\pi x}e^{-2ix\theta}\,\mathrm{d}x &= \int_{-\infty}^{\infty}\frac{e^{2\pi ix}-e^{-2\pi ix}}{4\pi ix}e^{-2ix\theta}\,\mathrm{d}x \\ &= \frac{1}{4\pi i}\left[\int_{-\infty}^{\infty}\frac{e^{2ix(\pi-\theta)}}{x}\,\mathrm{d}x - \int_{-\infty}^{\infty}\frac{e^{-2ix(\pi+\theta)}}{x}\,\mathrm{d}x\right]. \end{aligned}$$

Now, we make the change of variable $u = 2x(\pi-\theta)$ in the first integral and $v = 2x(\pi+\theta)$ in the second. Importantly, as we are taking $\theta \in (-\pi, \pi]$ by use of the

principal Mellin transform, we have that $\pi - \theta \geq 0$ and $\pi + \theta \geq 0$, such that the integral bounds are still well-defined for arbitrary u, v. We thus obtain

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\sin(2\pi x)}{2\pi x} e^{-2ix\theta}\,\mathrm{d}x &= \frac{1}{4\pi i}\left[\int_{-\infty}^{\infty} \frac{e^{iu}}{u}\,\mathrm{d}u + \int_{-\infty}^{\infty} \frac{e^{iv}}{v}\,\mathrm{d}v\right] \\ &= \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{iu}}{u}\,\mathrm{d}u. \end{aligned} \tag{4.8.15}$$

To handle this integral, we consider the contour integral of $\frac{e^{iw}}{w}$ over the contour Γ depicted in Figure 4.2 below with four smooth components. By Cauchy's residue

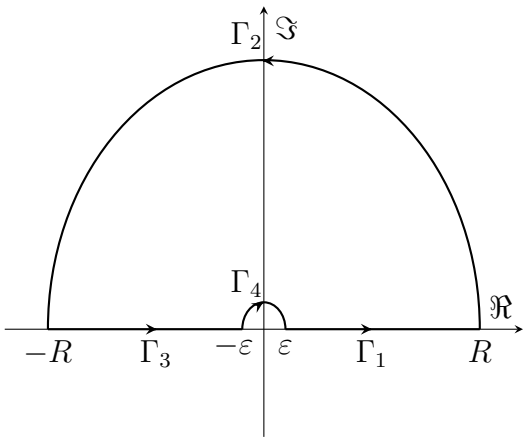

Figure 4.2: Contour of integration $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$ for $\frac{e^{iw}}{w}$.

theorem, as there are no singularities (i.e. 0) in the interior of Γ, we have that $\int_\Gamma \frac{e^{iw}}{w}\,\mathrm{d}w = 0$. Now, we can show (details in Appendix A.11) that for $R \to \infty$ and $\varepsilon \to 0$,

$$\begin{aligned}\int_{\Gamma_2} \frac{e^{iw}}{w}\,\mathrm{d}w &\to 0, \\ \int_{\Gamma_4} \frac{e^{iw}}{w}\,\mathrm{d}w &\to -i\pi, \\ \int_{\Gamma_3} \frac{e^{iw}}{w}\,\mathrm{d}w + \int_{\Gamma_1} \frac{e^{iw}}{w}\,\mathrm{d}w &\to \int_{-\infty}^{\infty} e^{iw} w\,\mathrm{d}w. \end{aligned}$$

Hence,

$$\begin{aligned} 0 &= \int_\Gamma \frac{e^{iw}}{w}\,\mathrm{d}w \\ &= \int_{\Gamma_1} \frac{e^{iw}}{w}\,\mathrm{d}w + \int_{\Gamma_2} \frac{e^{iw}}{w}\,\mathrm{d}w + \int_{\Gamma_3} \frac{e^{iw}}{w}\,\mathrm{d}w + \int_{\Gamma_4} \frac{e^{iw}}{w}\,\mathrm{d}w, \\ \Rightarrow i\pi &= \int_{-\infty}^{\infty} e^{iw} w\,\mathrm{d}w. \end{aligned}$$

Thus, returning now to (4.8.15), we have that

$$\int_{-\infty}^{\infty} \frac{\sin(2\pi x)}{2\pi x} e^{-2ix\theta} \, \mathrm{d}x = \frac{1}{2}.$$

Combining with (4.8.14) and substituting into (4.8.13), we obtain (note we explicitly separate the value at $z = 0$ which is 0 by taking the limit)

$$f_{1/Z}(z) = \begin{cases} \frac{1}{\pi|z|^4} \exp\left\{-\frac{1}{\sigma^2|z|^2}\right\}, & z \neq 0 \\ 0, & z = 0 \end{cases},$$

which is exactly the same result recognised in Kotlarski (1965) as a particular case.

The above example illustrates how to invert a complex Mellin transform, but also that tools such as the Dirac delta function and contour integration from complex analysis should be part of a statistician's toolbox if they wish to work with complex random variables. Although the example above may seem lengthy by explicitly spelling out every step, it becomes a relatively quick exercise for one well-versed in the Dirac delta function and complex analysis. Immediately, we have generalised results from Kotlarski (1965) by now taking the variance of the complex normal distribution into account. Further examples from Kotlarski (1965) and Brock and Krutchkoff (1970) may also be computed in generality using this inversion formula, demonstrating its use as an essential tool when dealing with complex random variables, and powers or products thereof. An interesting application of complex random variables and the Mellin transform to real random variables is given in Section 4.4 of Kotlarski (1965) where, for a random 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} X & Y \\ U & V \end{pmatrix}, \tag{4.8.16}$$

if one defines $Z = X + iY$ and $W = U + iV$, then $\det(\mathbf{A}) = \Im(Z^*W)$. Thus, the distribution of Z^*W may be computed via the Mellin transform and inversion where the respective transforms of Z and W are known. The distribution of $\det(\mathbf{A})$ may then be computed by marginalising the complex distribution by integrating out the real part.

As with real statistics and working with the characteristic function, it is indeed useful to be able to recognise common Mellin transforms to quickly identify the distribution of products of random variables, however in the absence of any known distribution, this inversion procedure enables the recovery of a density from a tractable Mellin transform. Furthermore, it makes explicit a connection by which the complex Mellin transform uniquely describes the distribution of continuous complex random variables.

Most importantly, the definition of the Mellin transform and its inversion for a complex random variable allows the description of product distributions — something which bivariate random variables do not naturally mimic. Indeed this complex Mellin transform is not equivalent to the extended bivariate Mellin transforms that was developed by Fox (1957) and explored further in Subrahmaniam (1970), due to

the presence of the complex multiplication and taking advantage of the polar form of a complex number.

4.9 Circularity, Properness and Elliptical Symmetry

To round out the discussion of distributions in this chapter, we now address notions of symmetry. Two concepts that are used to describe symmetry of c.r.v.'s are those of circularity and properness, with a closely related extension to elliptical symmetry.

4.9.1 Circular distributions

A c.r.v., Z, is said to be *circular* about $\mu \in \mathbb{C}$ if $Z - \mu$ and $(Z - \mu)e^{i\alpha}$ are equal in distribution for all $\alpha \in \mathbb{R}$. Letting $\mu = 0$ without loss of generality, this means that the distribution of $Re^{i\Theta}$ is the same as that of $Re^{i(\Theta+\alpha)}$ for all $\alpha \in \mathbb{R}$.

Proposition 4.9.1 (Eriksson et al. (2010))**.** *A circular complex random variable has its random modulus $|Z|$ independent of its random argument* $\mathrm{Arg}(Z)$*. Moreover, the distribution of Z only depends upon the distribution of its modulus whereby there exists a function $g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+$ such that $f_Z(z) = g(|z|)$ and* $\mathrm{Arg}(Z)$ *is uniformly distributed on $(-\pi, \pi]$.*

This is equivalent to the notion of bivariate spherical symmetry in the Cartesian components (Eriksson et al. (2010); Ollila et al. (2011)). The dependence only upon the modulus holds implication for the moments of a circular random variable.

Proposition 4.9.2 (Comon (1994))**.** *A circular random variable has all non-absolute integer-order moments (if they exist) equal to zero. That is, $\mathbb{E}\left(Z^m Z^{*n}\right) = 0$ whenever $m \neq n$ for all $m - n \in \mathbb{Z}$.*

Proof. By the definition of circularity, it must be that for any $\alpha \in (-\pi, \pi]$,

$$\mathbb{E}\left(Z^m Z^{*n}\right) = \mathbb{E}\left(\left(Ze^{i\alpha}\right)^m \left(Ze^{i\alpha}\right)^{*n}\right) = \mathbb{E}\left(Z^m Z^{*n}\right) e^{i\alpha(m-n)}.$$

For $\mathbb{E}\left(Z^m Z^{*n}\right) \neq 0$, this implies that $e^{i\alpha(m-n)} = 1$ for all $\alpha \in (-\pi, \pi]$, which is only satisfied if $m = n$. Hence if the $(m+n)^{\text{th}}$-order moments exist, the non-absolute $(m \neq n)$ moments must be 0. Note that $m - n \in \mathbb{Z}$ is required here to ensure that the equality is held on single-valued objects. □

In terms of describing the distribution of a circular random variable Z with mean μ, the ball-wise distribution function at the mean, $F_Z^B(\mu; \cdot) : \mathbb{R}^+ \to [0, 1]$ uniquely determines the circular distribution.

The dependence only upon the random variable's modulus also holds for the characteristic function, where Theorem 6.1 in Eriksson et al. (2010) gives that there must exist a function $h : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+$ such that $\varphi_Z(t) = h(|t|)$ for any circular c.r.v. Z.

For random vectors, circularity can further be broken down into marginal, weak, strong and total circularity. A brief overview of each will be given here as is required for later reference. Refer to Picinbono (1994) for a more complete description.

1. (*Marginal*) A complex random vector $\boldsymbol{Z}$ taking values in $\mathbb{C}^n$ is *marginally circular* if it's components Z_k are themselves circular complex random variables.

2. (*Weak*) A complex random vector $\boldsymbol{Z}$ is *weakly circular* if $\boldsymbol{Z}$ and $e^{i\alpha}\boldsymbol{Z}$ have the same distribution for any $\alpha \in \mathbb{R}$. That is, in terms of the density f as a function of r and θ, $f_{\boldsymbol{Z}}(\boldsymbol{r}; \theta_1, \theta_2, ..., \theta_n) = f_{\boldsymbol{Z}}(\boldsymbol{r}; \theta_1 + \alpha, \theta_2 + \alpha, ..., \theta_n + \alpha)$, for all $\alpha \in (-\pi, \pi]$. Weak circularity implies marginal circularity.

3. (*Strong*) A complex random vector $\boldsymbol{Z}$ is *strongly circular* if $\boldsymbol{Z}$ has the same distribution as the vector $(e^{i\alpha_k} Z_k)_{k=1}^n$. That is, $f_{\boldsymbol{Z}}(\boldsymbol{r}; \theta_1, \theta_2, ..., \theta_n) = f_{\boldsymbol{Z}}(\boldsymbol{r}; \theta_1 + \alpha_1, \theta_2 + \alpha_2, ..., \theta_n + \alpha_n)$, for all $\boldsymbol{\alpha} \in (-\pi, \pi]^n$. Strong circularity implies weak circularity.

4. (*Total*) Lastly, *total circularity* is where the components, Z_k, of $\boldsymbol{Z}$ are independent and circular, which implies the density can be factorised into a product of functions of the individual moduli of each component c.r.v.. Total circularity implies strong circularity. Conversely, marginal circularity coupled with independence of the Z_k's implies total circularity, in which case all four types of circularity are equivalent.

4.9.2 Properness

A c.r.v. Z is called *proper* if its pseudovariance $\rho^2 = \mathbb{E}\left((Z - \mathbb{E}(Z))^2\right) = 0$. Any circular c.r.v. must be proper by consequence of all non-absolute moments vanishing (Proposition 4.9.2). However, a c.r.v. can be proper without being circular, consider for example a complex c.r.v. which is uniformly distributed over the square $\{|\Re(z)| < 1, |\Im(z)| < 1\}$ (Eriksson et al. (2010, p.5405)). This c.r.v. can be seen to be proper, however clearly it is not circular by the nature of its support not invariant under arbitrary rotation.

The relationship between properness and circularity can be thought of like that of correlation and independence where the former are second-order properties of a distribution, whilst the latter are properties of the distribution and influence all moments, not limited to but including the second order ones. The notions also coincide in the case of the normal distribution (defined next chapter) as the second-order moments uniquely determine the distribution.

In practice, proper but non-circular data may arise in the form of quadrature amplitude modulation (QAM) data, which is distributed like QPSK data (Example 2.3.1), but on a square grid of transmission nodes centred at the origin, and hence proper, but not circular (Adali and Schreier (2014, p.118)). For a rigorous treatment of improper and non-circular data from a signal processing perspective, Schreier and Scharf (2010) and Adali et al. (2011) provide a very thorough treatment. Whilst the circular case provides many nice, simple results for complex random variables due to the univariate description of the distribution by the modulus, we will not treat improper signals as a special case, rather focus on arbitrary distributions where required.

4.9.3 Elliptical symmetry

Finally, inspired by dealing with elliptical distributions such as a normal distribution which may not be circular, a complex random variable Z is called elliptically symmetric if it can be written as an '$\mathbb{R}$-linear' function of a circular random variable. All of what follows is a summary of the formulation as provided in Ollila et al. (2011), for which the primary historical reference is Krishnaiah and Lin (1986). An $\mathbb{R}$-linear (or 'widely linear' as in Picinbono and Chevalier (1995)) transform $L : \mathbb{C} \to \mathbb{C}$ is such that

$$L(z) = \eta + \alpha z + \beta z^*,$$

for constants $\eta, \alpha, \beta \in \mathbb{C}$. The transformation is invertible as long as $|\alpha| \neq |\beta|$, with the inverse (Equation (2), Ollila et al. (2011)),

$$L^{-1}(z) = \left(\frac{\alpha^*}{|\alpha|^2 - |\beta|^2}\right)(z - \eta) - \left(\frac{\beta}{|\alpha|^2 - |\beta|^2}\right)(z - \eta)^*. \tag{4.9.1}$$

.

Hence, a c.r.v. W is elliptically symmetric if $W = L(Z)$ for some circular random variable Z with zero mean and variance σ^2. In this case, W has mean η, variance $\sigma^2(|\alpha|^2 + |\beta|^2)$ and pseudovariance $2\sigma^2\alpha\beta$. In the event that $|\alpha| = |\beta|$, the transform is non-invertible and by Proposition 4.5.4, W is distributed along a straight line in the complex plane passing through $\sqrt{\alpha\beta}$. We present an example involving the uniform distribution.

Example 4.9.3. *Let Z be uniformly distributed over the unit disc in the complex plane. It will be shown in the next chapter that Z has zero mean and variance $\frac{1}{2}$. Let $\alpha = 1 + 2i$, $\beta = -3 + 0.5i$ and $\gamma = -2 + i$. Then considering first $W = L(Z; \alpha, \beta)$, Z is now uniformly distributed over an ellipse in the complex plane as shown in Figure 4.3 with major axis parallel to $\sqrt{\alpha\beta} = \sqrt{-4 - 5.5i}$ and minor axis perpendicular. However, in the case of $W = L(Z; \alpha, \gamma)$, as $|\alpha| = |\gamma|$, then W is uniformly distributed along a line in the complex plane passing through the origin and parallel to $\sqrt{-4 - 3i} = \sqrt{5}\exp\left\{i\left(\tan^{-1}\left(\frac{3}{4}\right) - \pi\right)/2\right\}$.*

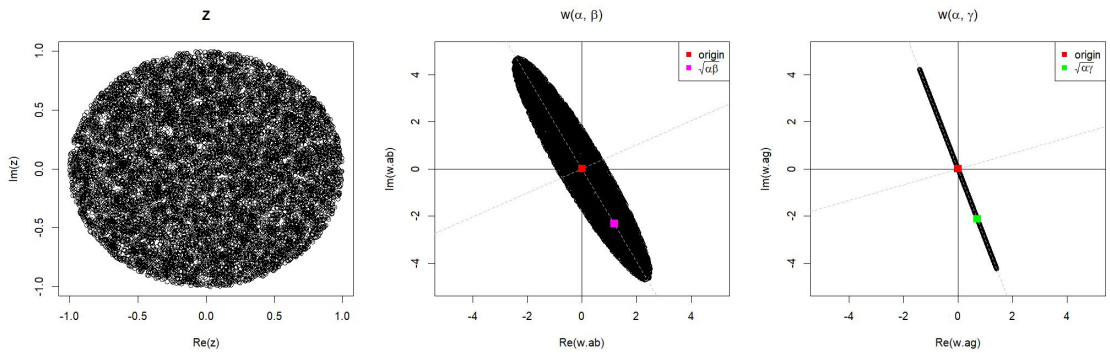

Figure 4.3: Simulations ($n = 10\,000$) of a uniformly distributed complex random variable Z and two $\mathbb{R}$-linear transformations for α, β, γ as described above.

Where $|\alpha| \neq |\beta|$, we can write the transform in terms of the desired mean η, variance σ^2 and pseudovariance ρ^2, so that by taking

$$\alpha = \frac{\sqrt{\sigma^2 + |\rho|^2} + \sqrt{\sigma^2 - |\rho|^2}}{2} e^{i\mathrm{Arg}(\rho)}, \tag{4.9.2}$$

$$\beta = \frac{\sqrt{\sigma^2 + |\rho|^2} - \sqrt{\sigma^2 - |\rho|^2}}{2} e^{i\mathrm{Arg}(\rho)}, \tag{4.9.3}$$

we have that $W = \eta + \alpha Z + \beta Z^*$ follows an elliptically symmetric distribution with the desired first and second order characteristics.

Elliptically symmetric random variables, and their generation as an $\mathbb{R}$-linear transform of standard circular random variables is how more general distributions such as the complex t-distribution may be developed, which is subsequently useful in complex linear modelling.

The next chapter will address some specific complex distributions, in particular continuous elliptically symmetric distributions as presented in the literature, as well as some novel discrete complex distributions inspired by continuation of univariate real distributions to a complex setting, rather than translation of bivariate notation.

Chapter 5

Complex Distributions

This chapter will define some common distributions of complex random variables. Continuous distributions such as the complex uniform, complex normal (Wooding (1956); Goodman (1963); van den Bos (1995)), elliptically symmetric distributions (Krishnaiah and Lin (1986); Ollila et al. (2011); Ollila et al. (2012)) and more general exponential family distributions (Ducharme et al. (2016, pp. 93–94)) are already widely explored in the literature, however predominantly from a real-valued approach by translating notation from bivariate distributions in the Cartesian decomposition of a complex random variable. Here we motivate the distributions from within the field of complex numbers $\mathbb{C}$. This chapter also motivates and defines several discrete complex distributions as natural extensions of familiar real distributions. To the best of our knowledge, the definitions of the discrete distributions are novel.

5.1 Complex Bernoulli

We begin with the complex analogue of the Bernoulli distribution, and if it may have a physically meaningful interpretation.

5.1.1 The standard complex Bernoulli

Before defining the distribution, we motivate its construction by extending the Bernoulli distribution on $\mathbb{R}$ to $\mathbb{C}$, keeping in mind how complex data are often interpreted.

The real Bernoulli

We start by recalling the real Bernoulli distribution. This is the distribution of a binary random variable, taking value 1 (on/yes) with probability p or 0 (off/no) with probability $1-p$. That is to say, $X \sim \text{Bern}_{\mathbb{R}}(p)$ gives the following distribution for X:

$$\mathbb{P}(X = 0) = 1 - p\,; \qquad \mathbb{P}(X = 1) = p,$$

or alternatively,

$$\mathbb{P}(X = x) = p^x(1-p)^{1-x}\,, \quad x \in \{0, 1\}\,.$$

This is often considered the most fundamental discrete random variable, as it can be used to simply model the occurrence of a single event — success or failure. From a measure theoretic perspective, the σ-algebra generated by a real Bernoulli random variable is the simplest non-trivial σ-algebra over the sample space of possible

outcomes i.e. for some $A \subset \Omega$ being a collection of outcomes, the only measurable sets become the collection $\{\emptyset, A, A^{\mathsf{c}}, \Omega\} = \sigma(A)$.

Basic properties are that:

$$\begin{aligned}
\mathbb{E}(X) &= p; \\
\mathbb{V}\mathrm{ar}(X) &= p(1-p); \\
\varphi_X(t) &= \mathbb{E}\left(e^{itX}\right) = 1 - p + pe^{it}; \\
\mathcal{M}_X(t) &= p\,,\ t \in \mathbb{C}.
\end{aligned}$$

Motivation and interpretation of the complex Bernoulli

How then does one consider an extension to the field of complex numbers? Considering the historical construction of the complex random normal distribution as a real bivariate normal distribution in complex notation (Wooding (1956)), one may be tempted to define the complex Bernoulli as a bivariate Bernoulli random variable. So a random complex number taking value in $\{0, 1, i, 1+i\}$?

Interpreting in a similar fashion to the real Bernoulli, 0 is still the zero element, 1 is on/success in the real direction and i is on/success in the imaginary direction. This would suggest that $1 + i$ is on/success in both directions. However, thinking back to the motivation for using complex numbers as representations of real phenomena we consider what this means in terms of representing a waveform signal. The value 1 as a multiplicative constant indicates an in-phase wave, whilst i represents a wave that is in quadrature (out-of-phase by $\frac{\pi}{2}$). The value $1 + i$ obscures the waveform interpretation, as $1 + i$ represents scaling the amplitude by $\sqrt{2}$ as well as a $\frac{\pi}{4}$ phase shift, which is a rather arbitrary operation for a fundamental case. Consequently, we propose that the complex Bernoulli distribution is in fact a discrete distribution over the three values $\{0, 1, i\}$ that distinguishes between real and imaginary results, similar to how we intuitively understand complex numbers by their real and imaginary component.

Such a Bernoulli distribution also corresponds to the next step away from non-triviality of the σ-algebra, where now the σ-algebra generated by such a random variable can be expressed as $\sigma(A, B)$ with A indicating the set related to a real value, B the set related to imaginary values, and thus$(A \cup B)^{\mathsf{c}}$ provides the remainder of the sample space, i.e. sets corresponding to a 0 outcome.

We observe several nice properties from this construction of a complex Bernoulli distribution. Firstly, we have this notion of success in the 'real-orthogonal' directions 1 and i. Secondly, we can uniquely parameterise this distribution with a single complex probability parameter (2 real degrees of freedom). Thirdly, this complex Bernoulli can be interpreted as whether a random signal is off/unreceived, in-phase or out-of-phase. Finally, thinking back to the occurrence and meaning of complex numbers in statistical application predominantly as representations of waves, the modulus and argument are the physically interpretable properties. If we let the real random variables R and Θ each be Bernoulli distributed[1] then the random variable $Re^{i\Theta}$ takes value in $\{0, 1, i\}$.

[1]Scaling Θ by $\frac{\pi}{2}$ to recover the interpretable real-orthogonal directions of 1 and i.

Definition and Properties

Based off the above, we thus define the following distribution to be the standard complex Bernoulli.

Definition 5.1.1 (Standard Complex Bernoulli)**.** *A random variable Z follows a standard complex Bernoulli distribution with parameter $\omega = p + iq \in \mathbb{C}$ if the real probability distribution is given by:*

$$\mathbb{P}(Z = 0) = 1 - |\omega|^2 = 1 - (p^2 + q^2); \tag{5.1.1}$$
$$\mathbb{P}(Z = 1) = \Re(\omega)^2 \quad = p^2; \tag{5.1.2}$$
$$\mathbb{P}(Z = i) = \Im(\omega)^2 \quad = q^2, \tag{5.1.3}$$

or alternatively

$$\mathbb{P}(Z = z) = \left(1 - |\omega|^2\right)^{1-|z|} \Re(\omega z^*)^{2|z|}, \quad z \in \{0, 1, i\}. \tag{5.1.4}$$

This is denoted by $Z \sim \mathrm{Bern}_{\mathbb{C}}(\omega)$, where $|\omega| \leq 1$.

The seemingly arbitrary decision to take squares of the components of ω reflects the discussion in Chapter 3 and behaviour in quantum physics, where if the complex probability function is such that $\mathbb{P}_{\mathbb{C}}(Z = 1) = \Re(\omega)$ and $\mathbb{P}_{\mathbb{C}}(Z = i) = i\Im(\omega)$, upon reducing the σ-algebra of measurable events such that only the event $\{Z = 1\} \cup \{Z = i\}$ is measurable[2], we recover a real $\mathrm{Bern}_{\mathbb{R}}(|\omega|^2)$ distribution with $\mathbb{P}_{\mathbb{C}}(Z = 1 \cup Z = i) = \omega$.

The properties of the complex Bernoulli distribution as above are that, for $Z \sim \mathrm{Bern}_{\mathbb{C}}(\omega)$ and $\omega = p + iq$,

$$\mathbb{E}(Z) = \Re(\omega)^2 + i\Im(\omega)^2 \quad = p^2 + iq^2; \tag{5.1.5}$$
$$\mathbb{V}\mathrm{ar}(Z) = \mathbb{E}\left(|Z|^2\right) - |\mathbb{E}(Z)|^2 \quad = p^2(1 - p^2) + q^2(1 - q^2); \tag{5.1.6}$$
$$\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = \mathbb{E}\left(Z^2\right) - \mathbb{E}(Z)^2 \quad = p^2(1 - p^2) - q^2(1 - q^2) - 2ip^2q^2; \tag{5.1.7}$$
$$\varphi_Z(t) = \mathbb{E}\left(e^{i\Re(Zt^*)}\right) \quad = 1 + p^2(e^{i\Re(t)} - 1) + q^2(e^{i\Im(t)} - 1); \tag{5.1.8}$$
$$\mathcal{M}_Z(s, t) = \Re(\omega)^2 + i^{s-t}\Im(\omega)^2 \quad = p^2 + i^{s-t}q^2, \ (s, t) \in \mathbb{C}^2. \tag{5.1.9}$$

In terms of component distributions, a complex Bernoulli random variable may be constructed as $X + iY$ where

$$X \sim \mathrm{Bern}_{\mathbb{R}}\left(\Re(\omega)^2\right), \tag{5.1.10}$$
$$Y \sim \mathrm{Bern}_{\mathbb{R}}\left(\Im(\omega)^2\right), \tag{5.1.11}$$

[2] we only care if we observe something, not what we observe

and there is the dependence structure such that only one of X or Y may be 1. In terms of polar components for $r, t \in [0, 1]$, if

$$R \sim \text{Bern}_{\mathbb{R}}(r), \tag{5.1.12}$$

$$\Theta \sim \frac{\pi}{2} \times \text{Bern}_{\mathbb{R}}(t), \tag{5.1.13}$$

and R is independent of Θ, then $Re^{i\Theta} \sim \text{Bern}_{\mathbb{C}}\left(\sqrt{r(1-t)} + i\sqrt{rt}\right)$. This allows us to simulate a complex Bernoulli random variable with the following Algorithm 2.

Algorithm 2 Simulating a standard complex Bernoulli random variable Z

Require: ω
Ensure: $|\omega| \leq 1$
Simulate: $R \sim \text{Bern}_{\mathbb{R}}(|\omega|^2)$
Simulate: $T \sim \text{Bern}_{\mathbb{R}}\left(\frac{\Im(\omega)^2}{|\omega|^2}\right)$
$\Theta \leftarrow \frac{\pi}{2}T$
$Z \leftarrow Re^{i\Theta}$

We note that the complex Bernoulli here may also be thought of as a multinomial distribution selecting one outcome over 3 possible values, identifying these 3 outcomes with $0, 1, i$ for the algebraic interpretability of 1 and i in application.

In higher dimensions, we also observe that a complex random vector $\boldsymbol{Z} \in \mathbb{C}^n$ with i.i.d. standard complex Bernoulli entries has L_2-norm $\|\boldsymbol{Z}\|_2^2 = \sum_{k=1}^n |Z_k|^2$ which follows a real binomial distribution $\text{Bin}_{\mathbb{R}}(n, |\omega|^2)$. It is worth noting that this aligns with the distribution of the L_2-norm of a random i.i.d. vector of real Bernoulli distributed random variables. This theme of real and complex L_2-norm distributions coinciding will recur when discussing the complex normal the relation of its L_2-norm to the Gamma distribution, of which the χ^2-distribution is a particular instance.

5.1.2 A thought experiment

We now consider a brief thought example that will motivate further generalisations of the complex Bernoulli distribution and potential quantities of interest. Consider a waveform signal such as a simple cosine wave travelling at a particular frequency f. This wave, $\cos(ft)$, can be represented by Euler's formula as the real component of the complex number e^{ift}. Consider now the wave passes through a random filter which may either let the wave pass through, reflect the wave, or delay it and put it into quadrature (Figure 5.1). Let there be a receiver placed after this filter to receive the subsequent wave. Then for a given single wave transmitted, the receiver will obtain one of three possible results. Either no wave at all, the transmitted wave, or the quadrature version of the transmitted wave. Mathematically, we can represent the received wave as Ze^{ift} where Z is a standard complex Bernoulli random variable.

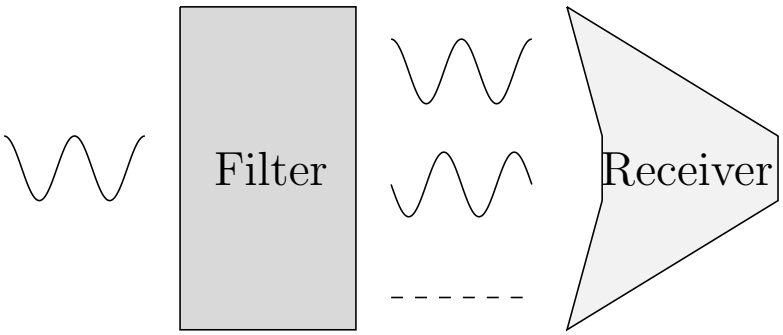

Figure 5.1: A signal passing through a random complex Bernoulli filter.

Now consider a signal passing through a series of n independent random filters in a row before reaching the receiver. Then then the received signal can thus be modelled by $\prod_{k=1}^{n} Z_k e^{ift}$, where the Z_k's are standard complex Bernoulli random variables with parameters ω_k. For $n \geq 3$, there are exactly five different signals that the may be received, corresponding to $\pm\cos(ft), \pm\sin(ft)$ and 0. Recalling the QPSK example (Example 2.3.1) from Section 2.3, the receiver in fact receives one of the four equispaced bit signals, or else it does not receive anything if the wave is reflected by any receiver. If we set each of the ω_k's such that $|\omega_k| = 1$, then we recover exactly a random QPSK signal. Hence, by construction of the standard Bernoulli random variable in terms of the polar components, we now consider the general product distribution of complex Bernoulli random variables, following on from the interpretability of the complex multiplication here.

5.1.3 Product distribution

We first consider the product distribution of i.i.d. standard complex Bernoulli random variables.

Proposition 5.1.2. *Let $Z_1, \ldots, Z_n$ be independent and identically distributed $\mathrm{Bern}_{\mathbb{C}}(\omega)$ random variables for $\omega = p + iq$. Then the distribution of $W = \prod_{k=1}^{n} Z_k$ has the following forms in the cases where $n = 2$ or $n \geq 3$.*

Case 1: $n = 2$,

$$\begin{aligned}
\mathbb{P}(W = 0) &= 1 - |\omega|^4, \\
\mathbb{P}(W = 1) &= \Re(\omega)^4, \\
\mathbb{P}(W = i) &= 2\left(\Re(\omega)\Im(\omega)\right)^2, \\
\mathbb{P}(W = -1) &= \Im(\omega)^4.
\end{aligned}$$

Case 2: $n \geq 3$, define

$$S_n = \sum_{m=0}^{\lfloor \frac{n}{4} \rfloor} \binom{n}{4m} \left(\Im(\omega)^2\right)^{4m} \left(\Re(\omega)^2\right)^{n-4m}, \tag{5.1.14}$$

then

$$\begin{aligned} \mathbb{P}(W = 0) &= 1 - |\omega|^{2n}, \\ \mathbb{P}\left(W = i^k\right) &= \Im(\omega)^{2k} S_{n-k}\,,\ k = 0, 1, 2, 3. \end{aligned}$$

The proposition can be verified by direct enumeration. Importantly, we notice the binomial distribution resemblance in the sum S_n defined for the case where $n \geq 3$. Decomposing into the polar components, which is the natural decomposition to consider under multiplication, this is equivalent in distribution to the random variable $Re^{i\Theta}$ where $R \sim \text{Bern}_{\mathbb{R}}\left(|\omega|^{2n}\right)$ and $\Theta \sim \frac{\pi}{2} \times \text{Bin}_{\mathbb{R}}\left(n, \Im(\omega)^2/|\omega|^2\right)$.

The non-identical case of the product distribution may be defined via the Mellin transform

$$\mathcal{M}_W(s,t) = \prod_{k=1}^{n} \left(\Re(\omega_k)^2 + i^{s-t}\Im(\omega_k)^2\right)\ ,\ (s,t) \in \mathbb{C}^2. \tag{5.1.15}$$

or principal Mellin transform

$$\mathcal{M}_W^{\text{P}}(s,t) = \prod_{k=1}^{n} \left(\Re(\omega_k)^2 + e^{i\frac{\pi}{2}(s-t)}\Im(\omega_k)^2\right)\ ,\ (s,t) \in \mathbb{C}^2. \tag{5.1.16}$$

and subsequent inversion, or direct enumeration of the probabilities and expansion with double binomial coefficients. However, we will not present the non-identical case here for two reasons: first, that it is unnecessarily messy in terms of notation; and secondly, as the astute reader may have noticed, in the thought experiment, why should it be that the filters are only able to shift the wave by a phase of $\pi/2$? Instead, would it not make sense to have a more general random variable that can describe random phase interference of a signal? Why not consider a random QPSK setup as a variable with a real Bernoulli distributed modulus R (or just 1 if we always receive something), and principal argument Θ distributed over the four possible values $\left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}$? This leads us to the idea of the generalised complex Bernoulli, motivated by the complex multiplication.

5.1.4 Generalised complex Bernoulli

As discussed, when interested in the standard complex Bernoulli random variable as a multiplicative discrete random variable that indicates whether a wave is in-phase or out-of-phase, the out-of-phase case does not have to be exactly in quadrature (difference of $\frac{\pi}{2}$ in the argument). This leads to the logical generalisation of the complex Bernoulli random variable to allow for an arbitrary phase shift. As such, we define the generalised complex Bernoulli random variable to take random values of either 0, or on the unit circle $\{\omega \in \mathbb{C} : |\omega| = 1\}$, where the marginal distribution of R is still a real Bernoulli random variable, but Θ is now arbitrarily distributed

(unique up to wrapping on an interval of length 2π). For example a random QPSK system can be described by letting Θ be distributed in some manner over the values $\left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}$. This leads to the following definition.

Definition 5.1.3. *A complex random variable Z follows a generalised complex Bernoulli distribution if it is equal in distribution to the random variable $Re^{i\Theta}$ where R follows a real Bernoulli distribution and Θ is a discrete real random variable.*

An example is a random QPSK system which randomly assigns an input signal to one of four equispaced bit signals.

Example 5.1.4 (Uniform Random QPSK)**.** *Consider a uniformly random QPSK signal (refer to Example 2.3.1), which can be mathematically represented by a generalised complex Bernoulli random variable $Z = Re^{i\Theta}$ where $R = 1$ and Θ is a discrete uniform distribution on the values $\{-\pi/2, 0, \pi/2, \pi\}$. Then the random variable $Z \in \{1, i, -1, -i\}$ is such that*

$$\mathbb{P}(Z = z) = \frac{1}{4}, \ \forall z \in \{1, i, -1, -i\}.$$

The equal probability masses equispaced around the unit circle in the complex plane result in the following distribution properties:

$$\begin{aligned}
\mathbb{E}(Z) &= 0; \\
\mathbb{V}\mathrm{ar}(Z) &= 1; \\
\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) &= 0; \\
\varphi_Z(t) &= \frac{\cos(\Re(t)) + \cos(\Im(t))}{2}, \ t \in \mathbb{C}; \\
\mathcal{M}_Z(s,t) &= \frac{1}{4}\left(1 + (-1)^{s+t} + i^{s-t} + i^{t-s}\right), \ s, t \in \mathbb{C}, s - t \in \mathbb{Z}.
\end{aligned}$$

The product of independent uniform random QPSK variables remains a uniform QPSK random variable, which can be verified by the fact that

$$\begin{aligned}
\left(\frac{1}{4}\left(1 + (-1)^{s+t} + i^{s-t} + i^{t-s}\right)\right)^2 &= \frac{1}{16}\left(4 + 4(-1)^{s+t} + 4i^{s-t} + 4i^{t-s}\right) \\
&= \frac{1}{4}\left(1 + (-1)^{s+t} + i^{s-t} + i^{t-s}\right),
\end{aligned}$$

and hence inductively $[\mathcal{M}_Z(s,t)]^n = \mathcal{M}_Z(s,t)$, with the same region of convergence. The sum of independent uniform random QPSK variables remains zero-mean and proper (zero pseudovariance), taking values on a diamond lattice in the complex plane with distribution similar to that of a rectangular binomial form.

Whilst the generalised complex Bernoulli distribution can be useful to model random phase shifts of a wave, more general distributions may be used in the context of random interference that can arbitrarily scale the modulus, as well as shift the phase These will be discussed as continuous complex random variables. However,

before handling continuous random variables, we first briefly discuss the bivariate Bernoulli distribution for historical reasons, and other discrete distributions related to the standard complex Bernoulli.

5.1.5 Real bivariate Bernoulli

As historically the complex normal distribution was developed as a complex random variable with Cartesian component variables normally distributed (Wooding (1956); Goodman (1963)), being essentially a real bivariate normal distribution written in complex notation, we consider also how a real bivariate Bernoulli distribution would be written in complex notation. Identifying $\mathbb{R}^2$ with $\mathbb{C}$, we obtain a complex random variable Z taking values in $\{0, 1, i, 1+i\}$. Then we obtain the distribution where, if the real bivariate distribution is given by $\mathbb{P}(X = x, Y = y) = p_{xy}$,

$$\begin{aligned} \mathbb{P}(Z = 0) &= p_{00}\ , & \mathbb{P}(Z = i) &= p_{01}, \\ \mathbb{P}(Z = 1) &= p_{10}\ , & \mathbb{P}(Z = 1+i) &= p_{11}. \end{aligned}$$

Note that the standard complex Bernoulli is just an instance of this distribution with[3] $\omega = \pm\sqrt{p_{10}} \pm i\sqrt{p_{01}}$ and $p_{11} = 0$. The statistical properties of the bivariate real Bernoulli distribution written in complex notation are given by

$$\begin{aligned} \mathbb{E}(Z) &= (p_{10} + p_{11}) + i(p_{01} + p_{11}), \\ \mathbb{V}\mathrm{ar}(Z) &= p_{10}(1 - p_{10}) + p_{01}(1 - p_{01}) + 2p_{11}p_{00}, \\ \mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) &= p_{10}(1 - p_{10}) - p_{01}(1 - p_{01}) + 2i(p_{11}p_{00} - p_{10}p_{01}), \\ \varphi_Z(t) &= p_{00} + p_{10}e^{i\Re(t)} + p_{01}e^{i\Im(t)} + p_{11}e^{i(\Re(t)+\Im(t))}\ ,\ t \in \mathbb{C}, \\ \mathcal{M}_Z^{\mathrm{P}}(s,t) &= p_{10} + p_{01}e^{i(s-t)\frac{\pi}{2}} + p_{11}\sqrt{2}\,e^{i(s-t)\frac{\pi}{4}}\ ,\ (s,t) \in \mathbb{C}^2. \end{aligned}$$

As discussed, although it is possible to translate notation to write a bivariate Bernoulli distribution as a single complex random variable, at this stage we lack any interpretation of this random variable in the current context for which complex random variables are considered in statistical application. As such, we will not explore this distribution further as most statistical results can already be well-understood from the bivariate real case, and uniquely complex results such as the product of these random variables does not carry any interpretation in the current fields of application of complex random variables as far as we know.

5.2 Other Complex Discrete Distributions

Here we consider the binomial and Poisson distributions in the complex case, developed from the standard complex Bernoulli distribution analogously to how these distributions arise in the real case.

5.2.1 Complex binomial

The real binomial distribution arises as the sum of n independent real Bernoulli random variables (Rice (1995); Casella and Berger (2002)). The common interpretation of a binomial distributed real random variable is often of the form "the number of successes in n repeated trials of a binary experiment". We have already

[3] The signs of the real and imaginary components of ω need not be the same.

seen binomial coefficients arise in probabilities in the product distribution of the standard complex Bernoulli distribution (Proposition 5.1.3) which attributes binomial probabilities, although is restricted to five possible outcomes. Now, we consider the distribution arising from the sum of n independent standard complex Bernoulli random variables, which for example may be counting the number of in-phase and out-of-phase (in quadrature) outcomes of an experiment.

Definition 5.2.1 (Complex Binomial)**.** *A complex random variable W follows a complex binomial distribution with parameters n and ω if $W \overset{\mathrm{d}}{=} \sum_{k=1}^{n} Z_k$, where $Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} \mathrm{Bern}_{\mathbb{C}}(\omega)$. If $\omega = p + iq$, then the probability mass function is given by*

$$\mathbb{P}(W = z) = \binom{n}{\Re(z)\ \Im(z)} \Re(\omega)^{2\Re(z)} \Im(\omega)^{2\Im(z)} (1 - |\omega|^2)^{n - \Re(z(1-i))} \tag{5.2.1}$$

$$= \binom{n}{\Re(z)\ \Im(z)} (p^2)^{\Re(z)} (q^2)^{\Im(z)} (1 - |\omega|^2)^{n - \Re(z) - \Im(z)}, \tag{5.2.2}$$

where $\binom{n}{x\ y} = \frac{n!}{x!\, y!\, (n-x-y)!}$ is the multinomial coefficient and $z = x + iy \in \mathbb{C}$ such that $x, y \in \mathbb{N}$ and $x + y \leq n$. This is denoted by $W \sim \mathrm{Bin}_{\mathbb{C}}(n, \omega)$.

A complex random variable following a complex binomial distribution takes values on a triangular lattice of $\frac{(n+1)(n+2)}{2}$ points in the upper right quadrant of the complex plane. A plot of the range of possible values for a complex binomial random variable is given in Figure 5.2 in the case of $n = 4$.

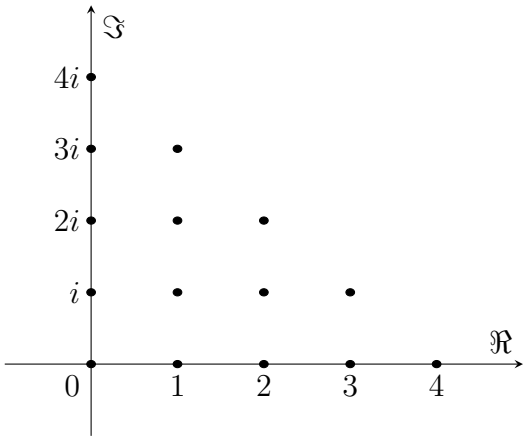

Figure 5.2: Possible values of a $\mathrm{Bin}_{\mathbb{C}}(4, \omega)$ random variable.

If $Z \sim \mathrm{Bin}_{\mathbb{C}}(n, \omega)$, then $\Re(Z) \sim \mathrm{Bin}_{\mathbb{R}}(n, \Re(\omega)^2)$ and $\Im(Z) \sim \mathrm{Bin}_{\mathbb{R}}(n, \Im(\omega)^2)$, however as in the case of the complex Bernoulli, the Cartesian components are not independent. The polar components are not investigated here as they are not a natural decomposition under summation and we currently do not have an interpretation for what the product of two complex Binomial random variables represents, just as the Cartesian component distributions of the product Bernoulli distribution were not investigated.

Basic distribution properties of $W \sim \text{Bin}_{\mathbb{C}}(n, \omega)$ for $\omega = p + iq$ in terms of $Z \sim \text{Bern}_{\mathbb{C}}(\omega)$ are:

$$\mathbb{E}(W) = n(\Re(\omega)^2 + i\Im(\omega)^2) = n\mathbb{E}(Z), \tag{5.2.3}$$

$$\mathbb{V}\text{ar}(W) = n\left(\Re(\omega)^2\left(1 - \Re(\omega)^2\right) + \Im(\omega)^2\left(1 - \Im(\omega)^2\right)\right) = n\mathbb{V}\text{ar}(Z), \tag{5.2.4}$$

$$\text{Ps}\mathbb{V}\text{ar}(W) = n\left(\Re(\omega)^2\left(1 - \Re(\omega)^2\right) - \Im(\omega)^2\left(1 - \Im(\omega)^2\right) - 2i\Re(\omega)^2\Im(\omega)^2\right) \tag{5.2.5}$$

$$= n\text{Ps}\mathbb{V}\text{ar}(Z), \tag{5.2.6}$$

$$\varphi_W(t) = \left(1 - |\omega|^2 + \Re(\omega)^2 e^{i\Re(t)} + \Im(\omega)^2 e^{i\Im(t)}\right)^n. \tag{5.2.7}$$

The characteristic function is easily derived by the fact that it is the product of n identical complex Bernoulli characteristic functions (5.1.8). Again, being a distribution inherently developed via a sum, we do not provide the Mellin transform which does not have a simple expression, nor (to our knowledge) conveys extra information.

Overall, whilst mathematically interesting due to its analogous construction relative to the real case, we do not provide a physical interpretation for the complex binomial distribution and its application is a topic of future research. It is worth reiterating, however, that the real binomial distribution (Casella and Berger (2002, p.90)) is also related to the complex Bernoulli distribution, being the L_2-norm of an i.i.d. vector of Bernoulli random variables. In the real case as $0^2 = 0$ and $1^2 = 1$, the L_2-norm coincides with the sum lending it to the usual interpretation as the number of successes (length of the n-tuple describing the repeated trials). However, more generally the L_2-norm in this case can be interpreted as the number of non-zero outcomes, which is now different to the summation distribution.

5.2.2 Complex Poisson

Continuing the analogous extension of the complex Bernoulli distribution, we now consider the complex Poisson distribution by defining the parameter $\lambda = \sqrt{n}\omega$ in the Binomial distribution, and then letting n go to infinity, assuming λ remains constant (Marshall and Olkin (1985)). Details may be found in Appendix A.12, however the result is given below.

Definition 5.2.2. *The complex random variable Z follows a complex Poisson distribution with parameter $\lambda \in \mathbb{C}$ if the probability mass function is given by*

$$\mathbb{P}(Z = z) = \frac{(\Re(\lambda)^2)^{\Re(z)}}{\Re(z)!} e^{-\Re(\lambda)^2} \cdot \frac{(\Im(\lambda)^2)^{\Im(z)}}{\Im(z)!} e^{-\Im(\lambda)^2}, \tag{5.2.8}$$

denoted $Z \sim \text{Pois}_{\mathbb{C}}(\lambda)$.

In the limit, the dependence structure that existed in the triangular lattice of the complex binomial is completely diluted such that the complex Poisson distribution is in fact a real independent bivariate Poisson distribution written in complex notation. That is, if $Z = X + iY$, then $X \sim \text{Pois}_{\mathbb{R}}(\Re(\lambda)^2)$ and $Y \sim \text{Pois}_{\mathbb{R}}(\Im(\lambda)^2)$ such that X is independent of Y.

Some properties of the complex Poisson distribution are as follows for $Z \sim \text{Pois}_{\mathbb{C}}(\lambda)$ and $\lambda = a + ib$.

$$\mathbb{E}(Z) = \Re(\lambda)^2 + i\Im(\lambda)^2 = a^2 + ib^2, \tag{5.2.9}$$
$$\mathbb{V}\text{ar}(Z) = |\lambda|^2 = a^2 + b^2, \tag{5.2.10}$$
$$\text{Ps}\mathbb{V}\text{ar}(Z) = \Re(\lambda^2) = a^2 - b^2, \tag{5.2.11}$$
$$\varphi_Z(t) = \exp\left\{\Re(\lambda)^2(e^{\Re(t)} - 1) + \Im(\lambda)^2(e^{\Im(t)} - 1)\right\} \tag{5.2.12}$$
$$= \exp\left\{a^2(e^{\Re(t)} - 1) + b^2(e^{\Im(t)} - 1)\right\}. \tag{5.2.13}$$

The complex binomial and Poisson distributions that were defined in this section define these distributions in the complex context following an analogous construction from real probability. However it is important to note that in real probability the sum and the L_2-norm coincide for the real Bernoulli distribution, however in the complex case the two operations lead to different distributions and the use of a relevant distribution should always be motivated by the context of application.

5.3 Continuous Complex Uniform

A continuous complex uniform (hereafter just complex uniform) random variable is the first of the continuous complex random variables considered. Instead of taking value on a lattice of points, or on contours in the complex plane, a continuous complex random variable is defined to take value in a region of the complex plane. A complex uniform distribution over some set $A \subset \mathbb{C}$ is a complex random variable that takes any value in A with equal likelihood. As per the definition of the Lebesgue measure in $\mathbb{C}$ as the area measure, the density of a complex uniform random variable on support A with respect to the Lebesgue measure is constant, $f_Z(z) = \frac{1}{\text{area}(A)}\mathbb{1}_A(z)$.

5.3.1 Standard complex uniform

As per discussion around cumulative distribution functions in Section 4.2, whilst in $\mathbb{R}$ the square and the ball coincide, they are different objects in $\mathbb{C}$. Again, due to the interpretability of the polar components of complex representations of real phenomena, we prefer to handle the standard case as the distribution over a ball. This also gives rise to the notion of circularity of the distribution, which has already been discussed as the rather natural symmetry in $\mathbb{C}$ under the geometric behaviour of complex multiplication. Hence we define the 'standard' complex uniform distribution as a uniform distribution over the unit disc in the complex plane.

Definition 5.3.1 (Standard Complex Uniform)**.** *A random variable Z follows a standard complex uniform distribution if Z takes values uniformly on the unit disc $D_1 = \{z \in C : |z| \leq 1\}$. The density with respect to the Lebesgue measure is given by*

$$f_Z(z) = \frac{1}{\pi}, \; z \in D_1, \tag{5.3.1}$$

and we denote the distribution $Z \sim \text{U}_{\mathbb{C}}(1)$. More generally, $U_{\mathbb{C}}^{\mu}(r)$ is a uniform distribution on the disc or radius $r > 0$ about $\mu \in \mathbb{C}$.

Some properties of the standard complex uniform distribution are that, for $Z \sim \mathrm{U}_{\mathbb{C}}(1)$, Z is circular and

$$\mathbb{E}(Z) = 0, \tag{5.3.2}$$

$$\mathbb{V}\mathrm{ar}(Z) = \frac{1}{2}, \tag{5.3.3}$$

$$\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = 0, \tag{5.3.4}$$

$$\varphi_Z(t) = \frac{4}{\pi\Im(t)} \int_0^1 \cos(\Re(t)x)\sin(\Im(t)\sqrt{1-x^2})\mathrm{d}x, \tag{5.3.5}$$

$$\mathcal{M}_Z^{\mathrm{P}}(s,t) = \begin{cases} \frac{2}{s+t+2}\,, & s = t \\ \frac{2}{s+t+2}\frac{\sin(\pi(s-t))}{\pi(s-t)}\,, & s \neq t \end{cases}, \ (s,t) \in \mathbb{C}^2. \tag{5.3.6}$$

Considering a decomposition of the complex random variable, being a circular distribution we naturally consider the polar decomposition. By Proposition 4.9.1, the random modulus must be independent of the random argument (Eriksson et al. (2010); Ollila et al. (2011)). Letting $\mathrm{U}_{\mathbb{R}}[a,b]$ denote the usual real uniform distribution on $[a,b]$ (Casella and Berger (2002, p.98)), then if

$$R^2 \sim \mathrm{U}_{\mathbb{R}}[0,1], \tag{5.3.7}$$

$$\Theta \sim \mathrm{U}_{\mathbb{R}}(-\pi,\pi], \tag{5.3.8}$$

the random variable $Re^{i\Theta} \sim \mathrm{U}_{\mathbb{C}}(1)$ for R and Θ independent. Hence a standard complex uniform random variable may be simulated as follows.

Algorithm 3 Simulating a standard complex uniform random variable Z

Simulate: $R^2 \sim \mathrm{U}_{\mathbb{R}}[0,1]$
Simulate: $\Theta \sim \mathrm{U}_{\mathbb{R}}(-\pi,\pi]$
$Z \leftarrow Re^{i\Theta}$

Other circular complex uniform distributions may be obtained by scaling the standard complex uniform. In particular, under linear transformations, if $Z \sim \mathrm{U}_{\mathbb{C}}(1)$, then for complex constants $\omega, \mu \in \mathbb{C}$,

$$\omega Z + \mu \sim U_{\mathbb{C}}^{\mu}(|\omega|). \tag{5.3.9}$$

However linear combinations of multiple complex uniform distributions are not necessarily themselves complex uniform, much as in the real case the sum of two uniform distributions is not uniform.

An exception is using the $\mathbb{R}$-linear construction from Ollila et al. (2011), where one can also construct elliptically symmetric complex uniform distributions as displayed in Example 4.3 by an $\mathbb{R}$-linear transformation such that $W = \alpha Z + \beta Z^*$ follows an elliptical uniform distribution with zero-mean and variance given by $\frac{1}{2}(|\alpha|^2 + |\beta|^2)$, but also pseudovariance (and hence it is non-circular) given by $\alpha\beta$.

5.3.2 Uniform on the square

Whilst the practical interpretation of the modulus and argument led to the definition of the standard uniform distribution on a ball one can equally consider a

uniform distribution on squares or rectangles. Again, the density function with respect to the Lebesgue measure will be constant and equal to the inverse of the area. So for example, the uniform distribution on the square with corners at the points $0, 1, i, 1+i$ will have density $f_Z(z) = 1$. This corresponds for example to a bivariate uniform distribution. Properties of the complex uniform distribution on the square can be deduced by translating notation from the real bivariate case and will not be explored further here.

5.4 The Normal Distribution

This section presents the complex normal distribution as it has been defined in the literature. As introduced by Wooding (1956) and further defined in Goodman (1963) (and generalised thereupon since), a c.r.v. Z is said to have a complex normal distribution if its real and imaginary parts follow a bivariate normal distribution. The initial definition concerned itself with the circular complex normal, that is $\Re(Z)$ and $\Im(Z)$ are independent and have the same common variance. We will first present the circular complex normal distribution, with a derivation entirely within $\mathbb{C}$, akin to that done by Gauss of the real normal distribution (Gauss (1809); Stahl (2006)), before presenting the more general non-circular and multivariate forms.

5.4.1 Univariate circular complex normal distribution

A complex derivation

Consider n independent, complex-valued measurements $\omega_1, \omega_2, \ldots, \omega_n$ of the same true quantity $p \in \mathbb{C}$. Each measurement inherently contains some error $\varepsilon_k \in \mathbb{C}$ such that $p = \omega_k + \varepsilon_k$ for $k = 1, 2, \ldots, n$. We employ three similar assumptions to those of Gauss in *Theoria Motus* (Gauss (1809, pp. 257–259)), and reported in Stahl (2006). Note that this inherently assumes an additive error. A similar derivation in the case of a multiplicative complex error will result in the complex log-normal distribution discussed later.

1. Measurements are more likely to be closer to the true value than very far away, that is errors of small modulus are more likely than those of large modulus.

2. Errors will be symmetric around the true value i.e. circular in a complex context.

3. The most likely value based off the data should be the arithmetic average, $\overline{\omega} = \frac{\sum_{k=1}^{n} \omega_k}{n}$.

Let f_ξ be the density of the complex random error ξ with a distribution from which we assume the i.i.d. sample $\{\varepsilon_1, \ldots, \varepsilon_n\}$ is drawn. For concise notation, it will be understood that $f'(z)$ is the Wirtinger derivative with respect to z. Using the assumptions in reverse order, assumption 3 suggests that if we write out the likelihood

$$\mathcal{L}(p; \varepsilon_1, \ldots, \varepsilon_n) = \prod_{k=1}^{n} f_\xi(p - \omega_k), \tag{5.4.1}$$

then this is maximised at $p = \overline{\omega}$ i.e. $\overline{\omega}$ is the maximum likelihood estimate of p (Stahl (2006)). Using the Wirtinger derivative to find the maximum, we obtain that

$$\mathcal{L}'(p; \varepsilon_1, \ldots, \varepsilon_n) = \sum_{k=1}^{n} f_\xi'(p - \omega_k) \prod_{l \neq k} f_\xi(p - \omega_l) \tag{5.4.2}$$

$$= \mathcal{L}(p; \varepsilon_1, \ldots, \varepsilon_n) \sum_{k=1}^{n} \frac{f_\xi'(p - \omega_k)}{f_\xi(p - \omega_k)}. \tag{5.4.3}$$

Setting the left-hand side equal to 0 for a potential maximum, and letting $g(z) = \frac{f_\xi'(z)}{f_\xi(z)}$, we obtain by positivity of $\mathcal{L}$,

$$\sum_{k=1}^{n} g(\overline{\omega} - \omega_k) = 0. \tag{5.4.4}$$

Now, as this should hold for any sample, consider the sample where $\omega_1 = \omega$ and $\omega_k = \omega + n\eta z$ for complex variables ω, η, z such that $|\eta| = 1$. Then the circular symmetry of assumption 2 implies that for all z, $f(\eta z) = f(z)$. By differentiation, this implies that $\eta g(\eta z) = g(z)$ for all $z \in \mathbb{C}$. Thus, using this particular sample where $\overline{\omega} = \omega + (n-1)\eta z$, we have that

$$0 = g((n-1)\eta z) + (n-1)g(-\eta z) \tag{5.4.5}$$

$$= \frac{1}{\eta} g((n-1)z) - \frac{1}{\eta}(n-1)g(z), \tag{5.4.6}$$

$$\Rightarrow g((n-1)z) = (n-1)g(z). \tag{5.4.7}$$

In Gauss' derivation as given in Stahl (2006), where z is real (having assumed continuity of f_ξ and hence g for differentiability), this would restrict g to being a linear function of z. However for g being a complex function, g may arbitrarily be an $\mathbb{R}$-linear function,

$$g(z) = \alpha z + \beta z^*. \tag{5.4.8}$$

The remaining condition 1 will help interpret the possible values of α and β once we recover a form for the density f_ξ.

Now, we take the Wirtinger antiderivative of (5.4.8) by treating z independent of its conjugate, recovering

$$\ln(f_\xi(z)) = \frac{\alpha}{2} z^2 + \beta |z|^2 + C(z^*), \tag{5.4.9}$$

for C a function of z^* only. Exponentiating yields

$$f_\xi(z) = \exp\left\{ \frac{\alpha}{2} z^2 + \beta |z|^2 + C(z^*) \right\}. \tag{5.4.10}$$

Firstly, by the restriction that f_ξ is real-valued for all z, it must be that $C(z^*) = B + \alpha z^{*2}$ for some constant $B \in \mathbb{R}$ and α, β real. Furthermore, by circular symmetry,

we require that $\alpha = 0$. Hence we recover the familiar form (for $A = e^B$)

$$f_\xi(z) = Ae^{\beta|z|^2}. \tag{5.4.11}$$

Now, by positivity of the modulus squared, condition 1 requires that the density attains a maximum at $z = 0$, which is satisfied for $\beta < 0$.[4] Using our crystal ball, let $\beta = -\frac{1}{\sigma^2}$ for some constant $\sigma \in \mathbb{R}$. Finally, we can determine the constant A by normalisation of a probability density, whereby (assuming the principal argument without consequence)

$$\begin{aligned} 1 &= \int_{-\pi}^{\pi}\int_0^\infty f_\xi(re^{i\theta})r \,\mathrm{d}r \,\mathrm{d}\theta && (5.4.12)\\ &= 2\pi A \int_0^\infty re^{-\frac{r^2}{\sigma^2}} \,\mathrm{d}r \,\mathrm{d}\theta && (5.4.13)\\ &= \pi\sigma^2 A\left[-e^{-\frac{r^2}{\sigma^2}}\right]_0^\infty && (5.4.14)\\ &= A\pi\sigma^2, && (5.4.15)\\ \Rightarrow A &= \frac{1}{\pi\sigma^2}. && (5.4.16) \end{aligned}$$

Thus we obtain the univariate complex normal density of the form

$$f(z) = \frac{1}{\pi\sigma^2}e^{-\frac{|z|^2}{\sigma^2}}\ ,\ z \in \mathbb{C}. \tag{5.4.17}$$

More generally of course, we may relocate this distribution by centering it around a point other than the origin in which case we arrive at defining the circular complex normal via its density with respect to the Lebesgue measure.

Definition

Definition 5.4.1 (Circular Univariate Complex Normal)**.** *A complex random variable Z follows a circular complex normal distribution with mean μ and variance σ^2 if its probability density function with respect to the Lebesgue measure is given by*

$$f_Z(z) = \frac{1}{\pi\sigma^2}e^{-\frac{|z-\mu|^2}{\sigma^2}}\ ,\ z \in \mathbb{C}, \tag{5.4.18}$$

and denoted $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$.

Observing (5.4.18) in terms of the Cartesian components of Z, it can be quickly recognised that the density in fact describes the density of an independent bivariate normal distribution with common variance of $\frac{\sigma^2}{2}$, which was how the original definition was motivated in Wooding (1956) (although this was handled directly in the multivariate case), rather than motivating the distribution from within $\mathbb{C}$ itself as we have done here. This similarity should not be surprising as the normal is an additive error distribution and $\mathbb{R}^2$ and $\mathbb{C}$ as additive groups are isomorphic. Thus

[4] This form of f_ξ indeed confirms that the stationary point of $\mathcal{L}$ is a maximum.

we also obtain that the circular complex normal distribution family is uniquely defined by its second-order moments and that it is closed under arbitrary complex linear combinations.

We can also choose to define the circular complex normal distribution uniquely by its characteristic function, or complex (principal) Mellin transform, both of which are given below.

Proposition 5.4.2. *A circular complex normally distributed random variable Z with mean μ and variance σ^2 exhibits the characteristic function*

$$\varphi_Z(t) = e^{i\Re(\mu t^*) - \frac{1}{4}\sigma^2|t|^2}\,,\ t \in \mathbb{C}. \tag{5.4.19}$$

Furthermore, if $\mu = 0$ (so Z is centered), then the Mellin transform of a centered circular complex normal distribution is given by

$$\mathcal{M}_Z^{\mathrm{P}}(s,t) = \frac{\sin(\pi(s-t))}{\pi(s-t)}\sigma^{s+t}\Gamma\left(\frac{s+t+2}{2}\right), \tag{5.4.20}$$

for all $s,t \in \mathbb{C}$ such that $s+t \notin 2\mathbb{Z}^- = \{-2,-4,\ldots\}$. Where $s = t$, we obtain

$$\mathcal{M}_Z^{\mathrm{P}}(s,s) = \sigma^{2s}\Gamma(s+1). \tag{5.4.21}$$

Proof. The proof of the characteristic function follows by directly translating notation from bivariate statistics and is a simplified version of the complex multivariate characteristic function purported in the literature (see e.g. Ducharme et al. (2016) and references therein). The Mellin transform follows from definition, integral manipulation and subsequently convergence of the Gamma function in the complex plane. Succinctly, for $s \neq t$,

$$\begin{aligned}
\mathcal{M}_Z^{\mathrm{P}}(s,t) &= \mathbb{E}_Z\left(Z^s Z^{*t}\right) \\
&= \int_{-\pi}^{\pi}\int_0^{\infty} r^{s+t+1}e^{i(s-t)\theta}\frac{1}{\pi\sigma^2}e^{-\frac{r^2}{\sigma^2}}\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \frac{\sin(\pi(s-t))}{\pi(s-t)}\sigma^{s+t}\int_0^{\infty}\left(\frac{r}{\sigma}\right)^2 e^{-\left(\frac{r}{\sigma}\right)^2}\frac{2r}{\sigma^2}\,\mathrm{d}r \\
&= \frac{\sin(\pi(s-t))}{\pi(s-t)}\sigma^{s+t}\int_0^{\infty} u^{\frac{s+t+2}{2}-1}e^{-u}\mathrm{d}u \qquad \left(\text{let } u = (r/\sigma)^2\right) \\
&= \frac{\sin(\pi(s-t))}{\pi(s-t)}\sigma^{s+t}\Gamma\left(\frac{s+t+2}{2}\right),
\end{aligned}$$

where the Gamma function (a.k.a. Euler's) integral in the second-last line converges for all complex values of $\frac{s+t+2}{2}$, except for the non-positive integers (Davis (1959)). Hence, as long as $\frac{s+t}{2}$ is not a negative integer, the integral and therefore Mellin transform converges. Should $s = t$, then the limit will cause the first term to approach unity, leaving only $\sigma^{s+t}\Gamma\left(\frac{s+t+2}{2}\right) = \sigma^{2s}\Gamma(s+1)$. □

It is not surprising that the Gamma function should arise in the Mellin transform of the circular normal distribution once we recognise that the modulus squared of a

centered circular complex normal distribution with variance σ^2 follows an exponential distribution with scale[5] σ^2. The Gamma function is the real Mellin transform of an exponentially distributed random variable.

Cumulative distribution function

As per Proposition 4.9.1, circular complex distributions are uniquely defined by distribution of their random modulus, and thus their ball-wise cumulative distribution function $F_Z^B(\mu, \cdot) : \mathbb{R}^+ \to [0, 1]$ such that $F_Z^B(\mu, r) = \mathbb{P}(|Z - \mu| \leq r)$, where $\mu = \mathbb{E}(Z)$. In the case of the circular complex normal distribution, we can obtain an explicit distribution function defining the probability of being in a certain region around the mean. Specifically if $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$

$$F_Z^B(\mu, r) = 1 - e^{-\frac{r^2}{\sigma^2}}, \ r > 0. \tag{5.4.22}$$

Whilst circular distributions are mathematically simple to manipulate, much of the existing literature in applied fields of complex statistics have readily recognised that circularity is a very rigid structure to impose upon observed data, which often empirically does not appear circular, such as in the case of fMRI or bivariate wind data (Eriksson et al. (2010); Schreier and Scharf (2010); Adali and Schreier (2014); Ducharme et al. (2016)). Consequently, $\mathbb{R}$-linear transformations can be used to obtain an elliptically symmetric complex normal distribution with now non-zero pseudovariance. The observant reader has probably already made the connection that the more general bivariate normal distribution is elliptically symmetric when the marginal distributions are allowed to be dependent, and indeed the complex normal distribution will always be seen to directly correspond to a real bivariate normal distribution via the Cartesian decomposition.

5.4.2 Univariate complex normal distribution

We provide two definitions of the generally non-circular complex normal distribution, which are univariate forms of the original definition given by van den Bos (1995) (notation here is that used in Ducharme et al. (2016)), firstly via the density and secondly by the characteristic function.

Definition 5.4.3. *A c.r.v. Z follows a complex normal distribution with mean $\mathbb{E}(Z) = \mu \in \mathbb{C}$, variance $\mathbb{V}\mathrm{ar}(Z) = \sigma^2 \in \mathbb{R}^+$ and pseudovariace $\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = \rho^2 \in \mathbb{C}$, if its real probability density with respect to the Lebesgue measure in $\mathbb{C}$ is*

$$f_Z(z) = \frac{1}{\pi\sqrt{\sigma^4 - |\rho|^4}} \exp\left\{-\frac{\sigma^2|z-\mu|^2 - \Re(\rho^2(z-\mu)^{*2})}{\sigma^4 - |\rho|^4}\right\}, \tag{5.4.23}$$

for all $z \in \mathbb{C}$. This is denoted $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$.

Alternatively, using the characteristic function, one can define the complex normal distribution as follows.

[5] So that $f_{R^2}(q) = \frac{1}{\sigma^2} e^{-\frac{q}{\sigma^2}}$.

Definition 5.4.4. *A c.r.v.* $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ *if and only if its characteristic function is given by*

$$\varphi_Z(t) = \exp\left\{ i\Re(t^*\mu) - \frac{1}{4}(\sigma^2|t|^2 + \Re(t^{*2}\rho^2)) \right\}, \tag{5.4.24}$$

for all $t \in \mathbb{C}$.

These definitions will come in use later for basic statistical results such as the central limit theorem (Section 6.3.4). It is also worth noting that the standard complex normal distribution is defined to be a zero-mean, unit variance or $\mathcal{N}_{\mathbb{C}}(0, 1, 0)$ distribution.

5.4.3 Properties, decomposition and simulation

We now examine the properties of the univariate complex normal distribution, which will be seen to reflect many of the properties of the real complex normal. First, we report the distributions of the Cartesian and polar components of a complex normally distributed random variable. Secondly, we will consider transformations (chiefly linear) of a complex random variable, and the degenerate case when the pseudovariance has equal modulus as the variance. Finally we will provide a novel decomposition of an arbitrary complex normal distribution which may facilitate efficient simulation of complex normal observations.

Component distributions

As previously discussed, the Cartesian components of a complex normally distributed random variable $Z = X + iY$ are themselves normally distributed. If $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$, then we have that

$$X \sim \mathcal{N}_{\mathbb{R}}\left(\Re(\mu), \frac{\sigma^2 + \Re(\rho^2)}{2} \right), \tag{5.4.25}$$

$$Y \sim \mathcal{N}_{\mathbb{R}}\left(\Im(\mu), \frac{\sigma^2 - \Re(\rho^2)}{2} \right), \tag{5.4.26}$$

$$\mathbb{C}\mathrm{ov}(X, Y) = \frac{\Im(\rho^2)}{2}. \tag{5.4.27}$$

Furthermore, we obtain the nice property that properness and circularity coincide for the complex normal distribution. That is, a complex normal random variable is circular about its mean if and only if its pseudovariance is 0. Verification of this property is immediate from setting $\rho = 0$ in the above equations, in which case we have that X and Y are independent real normal random variables with equal variance, which are the component distributions of a circular complex normal random variable about the point μ.

As may be expected, the normal distribution was inherently constructed as an additive error distribution and so the polar components are not as simple in distribution or dependence structure. We handle four cases and report distributions as given in the literature.

1. If $Z \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2, 0)$ is a centered, circular complex normally distributed random variable, then $R = |Z|$ and $\text{Arg}(Z)$ are independent with $\text{Arg}(Z) \sim \text{U}_{\mathbb{R}}(-\pi, \pi]$ and $R^2 \sim \text{Exp}(\sigma^2)$, or alternatively, R can be said to follow a Rayleigh distribution with scale parameter σ (Beckmann (1964)).

2. If instead $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$ is non-centered but still circular about its mean μ, then the modulus R follows a Rice distribution with parameters $|\mu|, \sigma$ (Sijbers et al. (1998)).

3. If instead $Z \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2, \rho^2)$ such that ρ^2 is real i.e. the Cartesian components are independent normals of differing variance, then the squared modulus R^2 follows a squared Hoyt distribution with parameters $q = \frac{\sigma^2 + \Re(\rho^2)}{\sigma^2 - \Re(\rho^2)}$ and $\gamma = \sigma^2$ (Romero-Jerez and Lopez-Martinez (2017)).

4. Finally, the marginal distribution of the modulus for an arbitrary complex normal distribution may be obtained by the methods reported in Cooper and Farid (2020), which first rotates the normal distribution such that the major axis of the elliptical probability contours lies on the coordinate axes (ρ^2 is real), and then uses the result in Weil (1954). We remark that the complex algebra somewhat simplifies the notation used in Cooper and Farid (2020) to rotate the elliptical contour axes onto the coordinate axes. Here, we report the density with respect to the Lebesgue measure of $|Z|$ where $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ for $|\rho| < \sigma$. Let,

$$a = \frac{2}{\sqrt{\sigma^4 - |\rho|^4}} \exp\left\{-\frac{\sigma^2|\mu|^2 - \Re(\rho^{*2}\mu^2)}{\sigma^4 - |\rho|^4}\right\}, \tag{5.4.28}$$

$$b = \frac{|\rho|^2}{\sigma^4 - |\rho|^4}, \tag{5.4.29}$$

$$c = \frac{2|\sigma^2\mu - \rho^2\mu^*|}{\sigma^4 - |\rho|^4}, \tag{5.4.30}$$

$$\psi = \text{Arg}(\sigma^2\mu - |\rho|^2\mu^*). \tag{5.4.31}$$

Then

$$f_{|Z|}(r) = a\, r \exp\left\{-\frac{r^2\sigma^2}{\sigma^4 - |\rho|^4}\right\}\left[I_0(br^2)I_0(cr) + 2\sum_{k=1}^{\infty} I_k(br^2)I_{2k}(cr)\cos(2k\psi)\right] \tag{5.4.32}$$

$$= a\, r \exp\left\{-\frac{r^2\sigma^2}{\sigma^4 - |\rho|^4}\right\}\sum_{k=-\infty}^{\infty} I_k(br^2)I_{2k}(cr)e^{i2k\psi}, \tag{5.4.33}$$

where I_n is the n^{th} order modified Bessel function of the second kind.[6]

[6] As much as we avoid translating notation from real bivariate statistics, the calculations involved here would have been identical to those given in Weil (1954), albeit with different constants in $\mathbb{C}$ instead of $\mathbb{R}$.

The densities of a Rayleigh, Rice and squared Hoyt distribution may be found in Appendix A.13.

In particular, point 1 above implies that if $\boldsymbol{Z} \in \mathbb{C}^n$ is a vector of i.i.d. circular complex normal random variables with zero mean and variance σ^2, then the L_2-norm $\|\boldsymbol{Z}\|^2$ follows a Gamma(n, σ^2) distribution. This aligns with real statistics, where for a vector $\boldsymbol{X}$ of i.i.d. $\mathcal{N}_{\mathbb{R}}(0, \tau^2)$ random variables, $\|\boldsymbol{X}\|^2 \sim \text{Gamma}(\frac{n}{2}, 2\tau^2)$.[7] Note that now, the χ^2-distribution which is frequently the basis of many statistical tests in real statistics is better handled by its more general family of Gamma distributions to allow for scaling and summation that arises in a complex context. Handling non-circular forms of the normal distribution in modulus requires further investigation into the properties of the general distribution given in (5.4.33). This is not explored in this thesis, however is of significance in terms of the distribution of the variance estimator for an arbitrary normal distribution, which arises in our brief exploration of linear modelling in Chapter 7.

Linearity and transformation

Here, we examine the distribution of linear combinations of complex normal random variables. As would be expected, for $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ and $W \sim \mathcal{N}_{\mathbb{C}}(\eta, \tau^2, \lambda^2)$ independent complex normal random variables and constants $\omega, \nu \in \mathbb{C}$, we have that

$$\omega Z + \nu \sim \mathcal{N}_{\mathbb{C}}\left(\omega\mu + \nu,\, |\omega|^2\sigma^2,\, \omega^2\rho^2\right), \tag{5.4.34}$$

$$Z + W \sim \mathcal{N}_{\mathbb{C}}\left(\mu + \eta,\, \sigma^2 + \tau^2,\, \rho^2 + \lambda^2\right), \tag{5.4.35}$$

preserving the linearity that we are used to in the real case. The verification of these properties is a straightforward multiplication of the characteristic function given in Definition 5.4.4 and using the linearity of the $\Re$ operator under real linear combinations.

One can in fact recover a real standard normal distribution as a particular instance of the complex normal distribution. Using the result of Proposition 4.5.4, if $|\rho|^2 = \sigma^2$, then the distribution is along the line in the complex plane through μ and parallel to ρ, as in Figure 5.3.

Hence, if $\mu, \rho \in \mathbb{R}$ such that $\rho^2 = \sigma^2$, then $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \sigma^2)$ is a real $\mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ distribution. More generally, if we allow $\mu \in \mathbb{C}$, then this will just shift the horizontal real normal distribution about $\Re(\mu)$ vertically in the complex plane by $\Im(\mu)$.

Combining the linearity and relation to the real normal distribution yields a decomposition of an arbitrary complex normal distribution in terms of a standard circular complex normal and a standard real normal random variables. This will facilitate simulation of complex normal observations where the complex mean, variance and pseudovariance parameters are known, instead of having to pass through a real elliptical bivariate normal simulation.

[7] For those expecting a χ^2-distribution, a χ^2_ν distribution is just equivalent to a Gamma$(\frac{\nu}{2}, 2)$ distribution. We are using the location-scale parameter definition of the Gamma distribution as opposed to the rate.

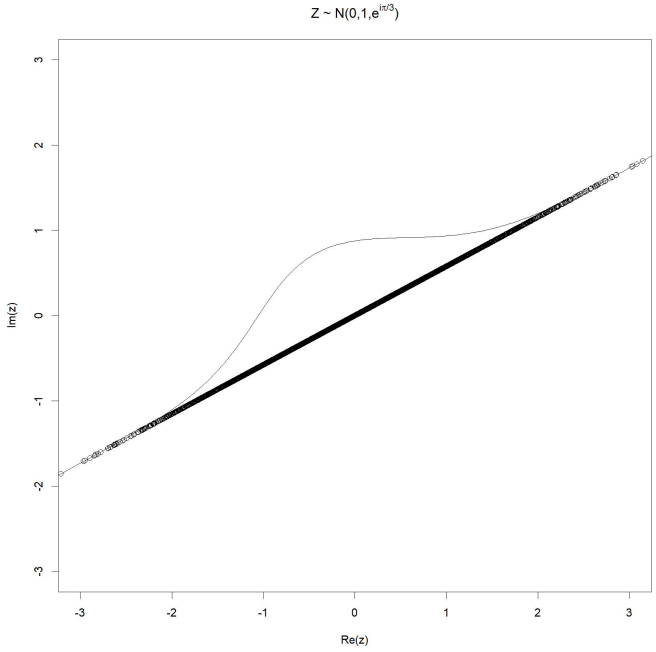

Figure 5.3: Simulations ($n = 10\,000$) from a $\mathcal{N}_{\mathbb{C}}\left(0, 1, e^{i\frac{\pi}{3}}\right)$ distribution.

A decomposition and simulation
An arbitrary complex normal random variable may be decomposed as follows.

Proposition 5.4.5. *A complex random variable $W \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ can have its distribution decomposed in terms of a standard circular complex normal random variable $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1, 0)$ independent of a standard real normal random variable $X \sim \mathcal{N}_{\mathbb{C}}(0, 1, 1)$ such that*

$$W \overset{\mathrm{d}}{=} \mu + \sqrt{\sigma^2 - |\rho|^2}\, Z + \rho X. \tag{5.4.36}$$

Ignoring a slight abuse of notation, we can also write this as

$$\mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2, \rho^2\right) = \mu + \sqrt{\sigma^2 - |\rho|^2}\,\mathcal{N}_{\mathbb{C}}(0, 1, 0) + \rho\mathcal{N}_{\mathbb{C}}(0, 1, 1). \tag{5.4.37}$$

Proof. This is a straightforward application of the linearity of the complex normal, given that we can write a real normal random variable as a complex normal random variable. Taking the right-hand side of (5.4.36), we have that

$$\begin{aligned}\mu + \sqrt{\sigma^2 - |\rho|^2}\, Z + \rho X &\sim \mathcal{N}_{\mathbb{C}}\left(\mu + 0 + 0,\ |\rho|^2 + \sigma^2 - |\rho|^2,\ (\sigma^2 - |\rho|^2) \cdot 0 + \rho^2\right) \\ &\sim \mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2, \rho^2\right).\end{aligned}$$

□

Thus, in order to simulate an arbitrary univariate complex normal random variable, it suffices to simulate three independent normal random variables.

Algorithm 4 Simulating a complex normal random variable $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$

Require: $\mu \in \mathbb{C}$, $\sigma^2 \in \mathbb{R}^+$, $\rho^2 \in \mathbb{C}$
Ensure: $|\rho|^2 \leq \sigma^2$
Simulate: $U, V, X \sim \mathcal{N}_{\mathbb{R}}(0, 1)$
$Z \leftarrow \mu + \sqrt{\frac{\sigma^2 - |\rho|^2}{2}}(U + iV) + \rho X$

Alternatively, as per construction in Ollila et al. (2011), one can also simulate an arbitrary complex normal by first simulating a standard circular complex normal and employing an $\mathbb{R}$-linear transformation.

Algorithm 5 Simulating a complex normal random variable $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ via an $\mathbb{R}$-linear transformation (Ollila et al. (2011))

Require: $\mu \in \mathbb{C}$, $\sigma^2 \in \mathbb{R}^+$, $\rho^2 \in \mathbb{C}$
Ensure: $|\rho|^2 \leq \sigma^2$
Simulate: $U \sim \mathcal{N}_{\mathbb{R}}\left(0, \frac{1}{2}\right) 0$
Simulate: $V \sim \mathcal{N}_{\mathbb{R}}\left(0, \frac{1}{2}\right) 0$
$\omega \leftarrow |\frac{\rho^2}{\sigma^2}|$
$a \leftarrow \sqrt{1 + \omega}$
$b \leftarrow \sqrt{1 - \omega}$
$W \leftarrow \frac{a+b}{2}(U + iV) + \frac{a-b}{2}(U - iV)$
$Z \leftarrow \mu + \sigma \exp\left\{\frac{i}{2}\mathrm{Arg}(\rho^2)\right\} W$

Being able to simulate independent univariate normal distributions will subsequently lend itself to simulating arbitrary multivariate distributions, however first, we must define the multivariate complex normal distribution.

5.4.4 Multivariate normal distribution

To end this section, we provide generalisations of the previous definitions of the complex normal to a complex random vector $\boldsymbol{Z} \in \mathbb{C}^n$, which was originally published by van den Bos (1995). Firstly, by considering the definition via the density presented in Ducharme et al. (2016).

Definition 5.4.6 (Ducharme et al. (2016))**.** *A complex random vector $\boldsymbol{Z}$ follows a multivariate complex normal distribution with mean $\mathbb{E}(\boldsymbol{Z}) = \boldsymbol{\mu} \in \mathbb{C}^n$, variance matrix $\mathbb{V}\mathrm{ar}(\boldsymbol{Z}) = [\mathbb{E}((Z_k - \mu_k)(Z_m - \mu_m)^*)] = \boldsymbol{\Gamma} \in M_{n\times n}(\mathbb{C})$ and pseudovariace $\mathrm{Ps}\mathbb{V}\mathrm{ar}(\boldsymbol{Z}) = [\mathbb{E}((Z_k - \mu_k)(Z_m - \mu_m))] = \mathbf{P} \in M_{n\times n}(\mathbb{C})$, if its real probability density with respect to the Lebesgue measure in $\mathbb{C}^n$ is*

$$f_{\underline{\boldsymbol{Z}}}(\underline{\boldsymbol{z}}) = \frac{1}{\pi^d |\boldsymbol{\Gamma}_{\mathbf{P}}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underline{\boldsymbol{z}} - \underline{\boldsymbol{\mu}})^{\mathcal{H}} {\boldsymbol{\Gamma}_{\mathbf{P}}}^{-1}(\underline{\boldsymbol{z}} - \underline{\boldsymbol{\mu}})\right\}, \tag{5.4.38}$$

for all $\underline{\boldsymbol{z}} = \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{z}^* \end{pmatrix} \in \mathbb{C}^{2n}$, where $\boldsymbol{\Gamma}_{\mathbf{P}}$ is the matrix

$$\boldsymbol{\Gamma}_{\mathbf{P}} = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{P} \\ \mathbf{P}^* & \boldsymbol{\Gamma}^* \end{pmatrix}.$$

This is denoted $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}^n}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{P})$.

When the pseudovariance matrix is non-zero, the density expression cannot, in our opinion, be meaningfully reduced and written in terms of $n \times n$ matrices and vectors in $\mathbb{C}^n$. This density can be derived by translating notation from the real $2n$-variate normal distribution using the conversion matrix $\mathbf{J}_n$ (Ducharme et al. (2016)). This means that to write the expanded scalar form of even a general bivariate complex normal density involves very, very excessive notation (asking Maple to expand and simplify the matrix notation, even assuming real values where necessary gave an expression a page long).

On the other hand, if the pseudocovariance matrix is indeed $\mathbf{0}$ such that we are in a circular setting, then we obtain the much simpler density form

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{\pi^d |\boldsymbol{\Gamma}|} \exp\left\{-(\boldsymbol{z}-\boldsymbol{\mu})^{\mathcal{H}} \boldsymbol{\Gamma}^{-1} (\boldsymbol{z}-\boldsymbol{\mu})\right\}, \tag{5.4.39}$$

which is indeed reminiscent of the usual multivariate real normal density (Wooding (1956); Ducharme et al. (2016)). In fact, we can rather simply write the bivariate complex normal density where we let Z_1, Z_2 have means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively, and define the covariance $\mathbb{C}\mathrm{ov}(Z_1, Z_2) = \xi\sigma\tau$.

$$f_{Z,W}(z,w) = \tag{5.4.40}$$

$$\frac{1}{\pi\sigma^2\tau^2(1-|\xi|^2)} \exp\left\{-\frac{1}{1-|\xi|^2}\left(\frac{|z-\mu|^2}{\sigma^2} - 2\Re\left(\xi^* \frac{z-\mu}{\sigma}\left(\frac{w-\eta}{\tau}\right)^*\right) + \frac{|w-\eta|^2}{\tau^2}\right)\right\}, \tag{5.4.41}$$

for all $(z,w) \in \mathbb{C}^2$. This is very similar in form to the usual real bivariate normal distribution, and we recognise that in this sense ξ would represent the correlation, something which will be discussed in further depth in Chapter 7.

An alternate definition of the complex normal distribution which only depends upon matrices and vectors of size n is given by defining the characteristic function of the distribution.

Definition 5.4.7 (Ducharme et al. (2016))**.** *A random complex vector* $\boldsymbol{Z} \in \mathbb{C}^n$ *has a* $\mathcal{N}_{\mathbb{C}^n}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{P})$ *distribution if and only if its characteristic function* $\varphi_{\boldsymbol{Z}}(\boldsymbol{t}) = \mathbb{E}\left(e^{i\Re(\boldsymbol{t}^{\mathcal{H}}\boldsymbol{Z})}\right)$ *is given by*

$$\varphi_{\boldsymbol{Z}}(\boldsymbol{t}) = \exp\left\{i\Re(\boldsymbol{t}^{\mathcal{H}}\boldsymbol{\mu}) - \frac{1}{4}\left(\boldsymbol{t}^{\mathcal{H}}\boldsymbol{\Gamma}\boldsymbol{t} + \Re(\boldsymbol{t}^{\mathcal{H}}\mathbf{P}\boldsymbol{t}^*)\right)\right\}, \tag{5.4.42}$$

for all $\boldsymbol{t} \in \mathbb{C}^n$.

Finally, of course one can also define the multivariate complex normal, similar to how a real multivariate normal distribution is defined, by saying that $\boldsymbol{Z}$ follow a complex normal distribution if every linear combination of its components $Z_1, \ldots, Z_k$ is also complex normally distributed (Ducharme et al. (2016)).

Linearity

Linearity of the multivariate complex normal distribution also follows where, for a matrix $\mathbf{A} \in M_{nn}(\mathbb{C})$ and normal random vector $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}^n}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{P})$,

$$\mathbf{A}\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}^n}\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathcal{H}}, \mathbf{A}\mathbf{P}\mathbf{A}^{\top}\right). \tag{5.4.43}$$

Simulation

Simulation from a multivariate complex normal distribution (in $\mathbb{C}^n$) is possible in practice, either by simulating from a real $2n$-variate normal distribution, or directly in $\mathbb{C}$ by simulating univariate complex normal random variables. Simulation via a real multivariate normal is achieved by simulating $\boldsymbol{X} \sim \mathcal{N}_{\mathbb{R}^{2n}}(\mathbf{J}_n\boldsymbol{\mu}, \mathbf{J}_n\boldsymbol{\Gamma}_{\mathbf{P}}\mathbf{J}_n^{\mathcal{H}})$ and identifying (Ducharme et al. (2016))

$$Z_k = X_k + iX_{n+k}\,,\ k = 1, 2, \ldots, n. \tag{5.4.44}$$

Simulating from within $\mathbb{C}$ is possible as long as simulation of independent standard circular complex normal and real normal random variables is possible, as well as an algorithm to find Cholesky and, in particular, Takagi factorisations (which may not always be feasible). The Cholesky factor of an Hermitian matrix $\boldsymbol{\Gamma}$ is the matrix $\mathbf{A}$ such that $\boldsymbol{\Gamma} = \mathbf{A}\mathbf{A}^{\mathcal{H}}$ and can be calculated using Algorithm 9 in Appendix A.14. The Takagi factorisation[8] of a symmetric matrix $\mathbf{P}$ consists of the unitary matrix $\mathbf{V}$ and the diagonal matrix $\boldsymbol{\Lambda}$ such that $\mathbf{P} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\top}$ (Xu and Qiao (2008); Ikramov (2012); Chebotarev and Teretenkov (2014)). The diagonal entries of $\boldsymbol{\Lambda}$ are the eigenvalues of an Hermitian matrix and thus non-negative. We present here an algorithm that should work in theory, however depends upon the ability to find Takagi factorisations, a discussion of which follows the algorithm. In essence, we use a multivariate version of Proposition 5.4.5, where square roots are now Cholesky or Takagi factorisations and standard circular complex and real normal random variables are now i.i.d. vectors.

Algorithm 6 Simulating a complex normal random vector $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}^n}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{P})$

Require: $\boldsymbol{\mu} \in \mathbb{C}^n$, $\boldsymbol{\Gamma} \in M_{nn}(\mathbb{C})$, $\mathbf{P} \in M_{nn}(\mathbb{C})$
Ensure: $\boldsymbol{\Gamma}$ Hermitian, $\mathbf{P}$ symmetric
Simulate: $W_1, \ldots, W_n \sim \mathcal{N}_{\mathbb{C}}(0, 1, 0)$ ▷ using Algorithm 4
Simulate: $X_1, \ldots X_n \sim \mathcal{N}_{\mathbb{C}}(0, 1, 1) = \mathcal{N}_{\mathbb{R}}(0, 1)$
  $\boldsymbol{W} \leftarrow (W_1, \ldots, W_n)^{\top}$
  $\boldsymbol{X} \leftarrow (X_1, \ldots, X_n)^{\top}$
  $\mathbf{V}, \boldsymbol{\Lambda} \leftarrow \text{Takagi}(\mathbf{P})$ ▷ assuming this is feasible
  $\mathbf{A} \leftarrow \mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}$
  $\mathbf{B} \leftarrow \text{Cholesky}(\boldsymbol{\Gamma} - \mathbf{A}\mathbf{A}^{\mathcal{H}})$ ▷ using Algorithm 9
  $\boldsymbol{Z} \leftarrow \boldsymbol{\mu} + \mathbf{B}\boldsymbol{W} + \mathbf{A}\boldsymbol{X}$

Note that there are several views in the literature regarding the Takagi factorisation feasibility. For example, Ikramov (2012) states that for $n \geq 5$, the Takagi factorisation cannot be given for a general complex symmetric matrix in terms of

[8] similar to the singular value decomposition

a finite algorithm with finitely many arithmetic operations and root extractions. Ikramov (2012) also provides a finite algorithm for the case of $n < 5$. However there also exist algorithms which deal with the general Takagi factorisation under certain matrix structures, such as Xu and Qiao (2008) where the authors provide an algorithm to compute the Takagi factorisation of a tridiagonal symmetric complex matrix. Similarly, Hahn (2007) suggests an iterative sweep algorithm based off Jacobi-type algorithms on 2×2 matrices, which does not mention a computational limit in terms of matrix dimensions. Whether the pseudocovariance function imposes a structure sufficient to develop a Takagi factorisation algorithm is a topic of future research. We expect it does, given that simulation is possible from $\mathbb{R}^{2n}$. In certain nice forms of the matrix $\mathbf{P}$, one could even use a modification of the Cholesky decomposition algorithm removing conjugation and replacing Hermitians with transposes, however this requires that one does not encounter a singular subblock of the now 'pseudoCholesky' factor at any stage of the algorithm.

5.5 Other Common Distributions

This section presents generalisations of complex distributions from the literature as well as the straightforward notion of what will be called circular continuation.

5.5.1 Complex t-distribution

Similar to the construction of the complex normal, the complex t-distribution is first defined via its circular case, which is then extended to an elliptical setting via an $\mathbb{R}$-linear transform (Krishnaiah and Lin (1986); Ollila et al. (2011)). Here we will present the definition as given in the literature (Krishnaiah and Lin (1986); Ollila et al. (2011); Ollila et al. (2012)). However, our definition has one minor difference in that we report the complex degrees of freedom that parameterise the complex t-distribution, which is equivalent to half the number of real degrees of freedom, due to better statistical interpretation in terms of the number of observations and parameters of a complex linear model (see Chapter 7).

Definition 5.5.1. *A complex random variable Z follows a circular complex t-distribution with $\nu \in \mathbb{R}^+$* complex *degrees of freedom if its density with respect to the Lebesgue measure is given by*

$$f_Z(z) = \frac{1}{\pi}\left(1 + \frac{|z|^2}{\nu}\right)^{-(\nu+1)}, \tag{5.5.1}$$

for all $z \in \mathbb{C}$, denoted $Z \sim \mathbb{C}t_\nu$. In particular, if $\nu = \frac{1}{2}$, we obtain the circular complex Cauchy distribution.

In higher dimensions, a complex random vector $\boldsymbol{Z} \in \mathbb{C}^n$ follows a circular multivariate complex t-distribution with $\nu \in \mathbb{R}^+$ complex *degrees of freedom and variance matrix (should it exist) $\frac{\nu}{\nu-1}\boldsymbol{\Gamma}$ if the density with respect to the Lebesgue measure is given by*

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{\Gamma(m+\nu)}{(\pi\nu)^n\Gamma(\nu)|\boldsymbol{\Gamma}|}\left(1 + \frac{\boldsymbol{z}^{\mathcal{H}}\boldsymbol{\Gamma}^{-1}\boldsymbol{z}}{\nu}\right)^{-(\nu+n)}, \tag{5.5.2}$$

for all $\boldsymbol{z} \in \mathbb{C}^n$, denoted $\boldsymbol{Z} \sim \mathbb{C}^n t_\nu^{\boldsymbol{\Gamma}}$.

Subsequently, elliptically symmetric t-distributions can be generated by $\mathbb{R}$-linear transformations, however conventionally, in order to remain a t-distribution, the coefficients α, β in the $\mathbb{R}$-linear transform should be such that $|\alpha|^2 + |\beta|^2 = 1$ to maintain the t-distribution variance. Of course, the t-distribution can be easily generalised to a scale-location family by scaling and relocating the standard t-distribution. The scale-location family generalisation will not be further explored here, however the interested reader may refer to Ollila et al. (2011) and Ollila et al. (2012) for further detail.

Properties

The properties of the complex t-distribution are presented here. First we start with the second-order moment properties, and then give the Cartesian and polar component distributions, the latter of which will be useful in presenting an alternative definition of the t-distribution that is useful in statistical inference for linear modelling.

Firstly, the second-order quantities for $Z \sim \mathbb{C}t_\nu$ are

$$\mathbb{E}(Z) = 0, \tag{5.5.3}$$

$$\mathbb{V}\mathrm{ar}(Z) = \frac{\nu}{\nu - 1}, \ \nu > 1 \tag{5.5.4}$$

$$\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = 0. \tag{5.5.5}$$

If $\nu \in (\frac{1}{2}, 1]$ the variance is infinite, and undefined if $\nu \leq \frac{1}{2}$. Then for $W = \alpha Z + \beta Z^*$,

$$\mathbb{E}(W) = 0, \tag{5.5.6}$$

$$\mathbb{V}\mathrm{ar}(W) = (|\alpha|^2 + |\beta|^2)\frac{\nu}{\nu - 1}, \ \nu > 1 \tag{5.5.7}$$

$$\mathrm{Ps}\mathbb{V}\mathrm{ar}(W) = 2\alpha\beta\frac{\nu}{\nu - 1}, \ \nu > 1. \tag{5.5.8}$$

The component distributions of $Z \sim \mathbb{C}t_\nu$ are given by, for $Z = X + iY = Re^{i\Theta}$,

$$\sqrt{2}X \sim t_{2\nu}, \tag{5.5.9}$$

$$\sqrt{2}Y \sim t_{2\nu}, \tag{5.5.10}$$

$$f_{R^2}(r) = \left(1 + \frac{r}{\nu}\right)^{-(\nu+1)}, \tag{5.5.11}$$

$$\mathrm{Arg}(Z) \sim \mathrm{U}_{\mathbb{R}}(-\pi, \pi], \tag{5.5.12}$$

such that R is independent of Θ (Ollila et al. (2011)). However, unlike the normal case, it is not true that X and Y are independent, which can be verified by direct multiplication of the densities. Nonetheless, we notice the appearance of 2ν as the real degrees of freedom in the Cartesian component which are reported in the literature. If one writes the densities in terms of the real degrees of freedom $\upsilon = 2\nu$, then the variance appears as the familiar $\frac{\upsilon}{\upsilon - 2}$, however we prefer to give the complex degrees of freedom for simplicity of notation in the density function and also intuitive connection with the number of parameters in a complex linear model, discussed in Chapter 7.

We now present a proposition followed by a secondary definition of the complex t-distribution, which is motivated more from within $\mathbb{C}$ itself based off the complex normal definition, rather than translating notation of a real bivariate spherically symmetric t-distribution. This is similar to how Ollila et al. (2011) define the circular complex t-distribution, and how it is defined in the real case such as in Rice (1995, p. 178).

Proposition 5.5.2. *Let $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1, 0)$ be independent of $G \sim \text{Gamma}(\nu, 1)$, then the random variable*

$$W = \frac{Z}{\sqrt{\frac{G}{\nu}}} \tag{5.5.13}$$

follows a complex circular $\mathbb{C}t_\nu$ distribution.

Proof (Sketch). The full proof is given in Appendix A.15. The sketch given here to highlight important parts is to first note that as Z is circular and G is real, then W is also circular, however the Cartesian components of W are only independent conditional upon G. Secondly, G has the same distribution as the sum of the modulus squared of ν i.i.d. copies of Z. Now as W is circular, its distribution is uniquely characterised by the distribution of its modulus squared. Letting $R^2 = |Z|^2$ (exponentially distributed), then we have that $|W|^2 = \nu R^2 G^{-1}$. Using simple methods of real probability, the distribution of G^{-1} can be determined, and thus the product distribution of $|W|^2$ can be explicitly calculated using the real Mellin transform. This distribution is seen to be exactly (5.5.11), the distribution of the modulus squared of a complex t-distribution as per Definition 5.5.1 which completes the proof. □

It is worth noting that if we choose to write the real degrees of freedom $\upsilon = 2\nu$, then we have that $G/\nu \overset{\text{d}}{=} V/\upsilon$ where $V \sim \chi^2_\upsilon$ highlighting the similarity to how a real t-distribution may be defined in terms of a standard normal and a χ^2 random variables. Nonetheless, the representation with the Gamma distribution is preferred given how it naturally arises as the L_2-norm of a circular complex normal, which is closely related to the distribution of the variance estimator.

Consequently, as with the complex normal, we can present a more general definition of an elliptically symmetric complex t-distribution similar to that given in Ollila et al. (2011), noting that an $\mathbb{R}$-linear combination of W in (5.5.13) simply yields an elliptically symmetric normal distribution divided by the square-root of the scaled Gamma distribution.

Definition 5.5.3. *A random variable W follows a complex t-distribution with $\nu > 0$ complex degrees of freedom and pseudovariance parameter ρ^2 if $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1, \rho^2)$ is independent of $G \sim \text{Gamma}(\nu, 1)$ and*

$$W = \frac{Z}{\sqrt{\frac{G}{\nu}}}. \tag{5.5.14}$$

This is denoted $W \sim \mathbb{C}t_\nu^{\rho^2}$. *If* $\nu = \frac{1}{2}$, *this will be a complex Cauchy distribution.*

The above definition amounts to W being an $\mathbb{R}$-linear transformation of the standard t-distribution defined in Definition 5.5.1 where the coefficients of the transform α, β satisfy $|\alpha|^2+|\beta|^2 = 1$ and $\rho^2 = 2\alpha\beta$. The variance of the above t-distribution is still $\frac{\nu}{\nu-1}$ for $\nu > 1$ and the pseudovariance is given by $\frac{\nu}{\nu-1}\rho^2$ where it is finite. Simulation can be obtained by dividing simulated complex normal observations from Algorithm 4 by simulations from the appropriately scaled Gamma distribution.

5.5.2 Complex log-normal distribution

Just as the complex normal distribution was derived by considering an additive error distribution with three intuitive assumptions, the log-normal distribution is developed as the multiplicative error distribution under similar assumptions, where we expect the geometric mean to be the most likely outcome. This arises as the logarithm of a product is the sum of the logarithms. So a complex random variable W follows a complex log-normal distribution if $\log(W)$ follows a complex normal distribution. Note here, we particularly use the function log which does not invoke the principal argument, as we require $\log(W)$ to be allowed to vary across the whole of $\mathbb{C}$. In that regard, we give the following definition of the log-normal distribution which is constructed by its polar components similar to the construction given in Halliwell (2015).

Definition 5.5.4. *If the complex random variable* $Z \sim \mathcal{N}_\mathbb{C}(\mu, \sigma^2, \rho^2)$, *then the random variable*

$$W = e^Z \tag{5.5.15}$$

is said to follow a complex log-normal distribution, denoted $W \sim \mathcal{LN}_\mathbb{C}(\eta, \sigma^2, \rho^2)$, *where* $\Re(\eta) = \Re(\mu)$ *and* $\Im(\eta) = (\Im(\mu) + \pi)\mathrm{mod}(2\pi) - \pi \in (-\pi, \pi]$. *The standard log-normal distribution is given where* $\mu = 0 = \rho^2$ *and* $\sigma^2 = 1$.

In terms of its component distributions, the natural decomposition to consider in the context of the complex logarithm is the polar decomposition. In particular, if R is a real log-normal$(0, 1/2)$ random variable independent of $\Theta \sim \mathcal{N}_\mathbb{R}(0, 1/2)$ (or a $\mathcal{N}_\mathbb{R}(0, 1/2)$ distribution wrapped onto $(-\pi, \pi]$), then $Re^{i\Theta}$ has a complex $\mathcal{LN}_\mathbb{C}(0, 1, 0)$ distribution. Note however that $\Theta \neq \mathrm{Arg}(Z)$.

Being a multiplicative noise distribution, the complex log-normal distribution family is closed under multiplication and powers. That is, if $W \sim \mathcal{LN}_\mathbb{C}(0, 1, 0)$ and $\Im(\mu) \in (-\pi, \pi]$, then

$$e^\mu W^\alpha W^{*\beta} \sim \mathcal{LN}_\mathbb{C}\left(\mu, |\alpha|^2 + |\beta|^2, 2\alpha\beta\right). \tag{5.5.16}$$

Furthermore, if $W \sim \mathcal{LN}_\mathbb{C}(\mu, \sigma^2, \rho^2)$ is independent of $Z \sim \mathcal{LN}_\mathbb{C}(\nu, \tau^2, \lambda^2)$, then for $\eta = \Re(\mu + \nu) + i[(\Im(\mu + \nu) + \pi)\mathrm{mod}(2\pi) - \pi]$,

$$WZ \sim \mathcal{LN}_\mathbb{C}\left(\eta, \sigma^2 + \tau^2, \rho^2 + \lambda^2\right). \tag{5.5.17}$$

Further research into the complex log-normal distribution is of interest in order to develop multiplicative complex statistical models, which provide direct interpretation in terms of noise acting upon the amplitude and phase of a signal. However, Halliwell (2015) notes that, as in the complex plane the exponential is a many-to-one function such that $e^Z = e^{Z+i2\pi k}$ for all $k \in \mathbb{Z}$, the complex log-normal may not exhibit an explicit density formula. This is why the uniqueness condition was imposed upon η in the definition. Nonetheless, further research is required to develop an explicit density which one might expect to resemble

$$\frac{1}{|z|\pi\sigma^2} \exp\left\{-\frac{|\mathrm{Log}(z)-\eta|^2}{\sigma^2}\right\}, \tag{5.5.18}$$

but a much more rigorous treatment is required to remedy the multi-valued nature of complex powers and exponentials. To calculate even the first few moments of the log-normal distribution requires the generalised characteristic function $\varphi_Z(s,t)$ of the complex normal distribution for possibly complex parameters $s,t \in \mathbb{C}$.

5.5.3 Circular continuation

Finally, we present the notion of what we will call circular continuation of well-known real distributions. This is the rather straightforward idea of taking a non-negative real distribution, such as the exponential distribution, and circularly rotating it in the complex plane. This induces a circular distribution where the modulus follows the underlying real distribution independent of the uniform argument. This idea is similar to the notion of density generators in Ollila et al. (2011). Under circular continuation, the density of Z obtained by circularly continuing the real random variable X about the point $\mu \in \mathbb{C}$ is given by

$$f_Z(z) = \frac{1}{2\pi|z-\mu|} f_X(|z-\mu|). \tag{5.5.19}$$

Example 5.5.5 (Complex Laplace distribution)**.** *Consider the distribution constructed by the circular continuation of the exponential distribution. This is similar to how in real probability, the Laplace distribution is a reflected exponential distribution. We thus define the Z to follow a complex circular Laplace distribution with location parameter $\mu \in \mathbb{C}$ and scale parameter $\lambda > 0$ if*

$$f_Z(z) = \frac{1}{2\pi\lambda|z-\mu|} e^{-\frac{|z-\mu|}{\lambda}}, \ z \in \mathbb{C}. \tag{5.5.20}$$

The mean of the complex Laplace distribution is μ and the variance is $2\lambda^2$. In particular, the complex Laplace distribution with $\mu = 0$ and $\lambda = \sigma^2$ is the distribution of W^2 where $W \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2, 0)$.

Being circular, the mean of a circularly continued distribution must be given by the location parameter μ and any integer non-absolute moments must vanish by consequence of Proposition 4.9.2. Conversely, all circular random variables may be written as the circular continuation of a particular real distribution. For example, the circular complex normal distribution is the circular continuation of the real Rayleigh distribution.

The next chapter will now briefly address basic statistical properties of complex random variables and data such as data inspection, estimation and limit theorems.

Chapter 6

Complex Random Samples and Statistical Properties

This chapter investigates properties of complex random samples and how complex data may be statistically analysed. This chapter does not claim to provide a thorough treatment of the entire theory of statistics for complex random variables, rather highlight existing developments, as well as areas that warrant further research. In particular, by nature of working with real probability measures and distributions, many familiar results for real random variables which are obtained by manipulating their distribution functions will still hold, however here results will be proven within $\mathbb{C}$ without relying upon bivariate statistics.

Regarding the language in this chapter, a consistent estimator is one that converges in probability, and a strongly consistent estimator converges almost surely, to the true parameter being estimated.

6.1 Sample Averaging

This section briefly discusses sample averaging for complex random samples.

6.1.1 Arithmetic mean

Given a random sample $\mathcal{Z} = \{Z_1, Z_2, \ldots, Z_n\}$, the arithmetic mean (hereafter 'sample mean') $\overline{Z}_n$ is given by

$$\overline{Z}_n = \frac{1}{n}\sum_{k=1}^{n} Z_k = \overline{X}_n + i\overline{Y}_n, \tag{6.1.1}$$

for $Z_k = X_k + iY_k$, $k = 1, 2, \ldots, n$.

As would be expected, the following first and second order moment conditions hold, and can easily be generalised to higher mixed moments (the case of an i.i.d. sample is also given in the second equalities):

$$\begin{aligned}
\mathbb{E}\left(\overline{Z}_n\right) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(Z_k\right) \overset{\text{i.i.d.}}{=} \mathbb{E}\left(Z\right); \\
\mathbb{V}\text{ar}\left(\overline{Z}_n\right) &= \frac{1}{n^2}\left(\sum_{k=1}^{n}\mathbb{V}\text{ar}\left(Z_k\right) + \sum\sum_{k\neq m}\mathbb{C}\text{ov}\left(Z_k, Z_m\right)\right) \overset{\text{i.i.d.}}{=} \frac{1}{n}\mathbb{V}\text{ar}\left(Z\right); \\
\text{Ps}\mathbb{V}\text{ar}\left(\overline{Z}_n\right) &= \frac{1}{n^2}\left(\sum_{k=1}^{n}\text{Ps}\mathbb{V}\text{ar}\left(Z_k\right) + \sum\sum_{k\neq m}\text{Ps}\mathbb{C}\text{ov}\left(Z_k, Z_m\right)\right) \overset{\text{i.i.d.}}{=} \frac{1}{n}\text{Ps}\mathbb{V}\text{ar}\left(Z\right).
\end{aligned}$$

Where the sample is independent (but not necessarily identically distributed), the characteristic function may be used to compute the distribution of the sum and

therefore the sample mean. Asymptotic properties of the sample mean will be given in Section 6.3 by the central limit theorem.

6.1.2 Geometric mean

When working with complex random variables, we may also consider the geometric mean of a complex random sample. Unlike the sample mean, there is no direct analogue of the complex geometric mean in real bivariate statistics as here we rely upon the complex multiplication. Given a random sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$, the geometric mean is defined by

$$\overline{Z}_n^{(g)} = \left(\prod_{k=1}^{n} Z_k \right)^{\frac{1}{n}}. \tag{6.1.2}$$

Of course, we revisit the issue of fractional powers of complex numbers which are not single-valued objects. In fact, there are exactly n distinct complex values for the geometric mean given in (6.1.2). The principal geometric mean is defined by applying the principal expectation operator $\mathfrak{E}$ in (6.1.2) instead of $\mathbb{E}$. This takes each $Z_k = R_k e^{i\Theta_k}$ for Θ_k the principle arguments. Then we have that

$$\overline{Z}_n^{(g)} = \overline{R}_n^{(g)} e^{i\overline{\Theta}_n}. \tag{6.1.3}$$

That is, the principal geometric mean of a random sample is the complex number with modulus the geometric mean of all other moduli, and argument the arithmetic mean of all principal arguments. In the absence of any qualification, we will assume the principal geometric mean. Note, however, that the geometric mean can be misleading as to the location of data. For example, the geometric mean will average $re^{i(\pi-\varepsilon)}$ and $re^{i(\varepsilon-\pi)}$ to just r on the positive real axis, despite both points lying close together around the negative real axis. Just as the geometric mean of two negative observations in $\mathbb{R}$ will be positive. In this aspect, to represent the location of the data, a change of sign may at times be required.

The following first and second order moments hold — note that further simplification in terms of the polar components hold in the case where the modulus and argument are also independent (such as for circular complex random variables). The below moments are multi-valued, however replacing expectations with principal expectations $\mathfrak{E}_{\mathcal{Z}}$ will yield the principal values.

$$\begin{aligned}
\mathbb{E}\left(\overline{Z}_n^{(g)}\right) &= \mathbb{E}\left(\sqrt[n]{Z_1 \cdot Z_2 \cdot \ldots \cdot Z_n}\right) \overset{\text{i.i.d.}}{=} \left[\mathbb{E}\left(\sqrt[n]{Z}\right)\right]^n; \\
\mathbb{V}\text{ar}\left(\overline{Z}_n^{(g)}\right) &= \mathbb{V}\text{ar}\left(\sqrt[n]{Z_1 \cdot Z_2 \cdot \ldots \cdot Z_n}\right) \overset{\text{i.i.d.}}{=} \left[\mathbb{E}\left(\sqrt[n]{|Z|^2}\right)\right]^n - \left|\mathbb{E}\left(\sqrt[n]{Z}\right)\right|^{2n}; \\
\text{Ps}\mathbb{V}\text{ar}\left(\overline{Z}_n^{(g)}\right) &= \text{Ps}\mathbb{V}\text{ar}\left(\sqrt[n]{Z_1 \cdot Z_2 \cdot \ldots \cdot Z_n}\right) \overset{\text{i.i.d.}}{=} \left[\mathbb{E}\left(\sqrt[n]{Z^2}\right)\right]^n - \left[\mathbb{E}\left(\sqrt[n]{Z}\right)\right]^{2n}.
\end{aligned}$$

Where the sample is independent, the Mellin transform is the obvious candidate to study the distribution of the geometric mean. If we relate the (principal) geometric mean to the exponential of the sample mean of the (principal) logarithm, then we can apply the central limit theorem (Subsection 6.3.4). Hence, we anticipate that the geometric mean of an i.i.d. sample will converge to a log-normal distribution

with mean $e^{\mathbb{E}(\log(Z_1))}$ (the 'population geometric mean'). Further research into the complex log-normal distribution and complex Mellin transform is required handle the multivalued logarithm and to state precisely the conditions (e.g. finite mean and variance of $\log(Z_k)$) for which this convergence holds. We also expect that a relaxed version will hold for independent samples, similar to Theorem 2.4 in Galambos and Simonelli (2004, p. 61) for real random variables.

6.2 Estimating the Distribution Function

This section discusses some basic statistical distribution properties. By identification of the complex distribution with the bivariate distribution, some results are familiar to the real bivariate case whilst others make use of the complex multiplication and the its interpretation as a rotation.

6.2.1 Empirical distribution function

As discussed in Section 4.2, there does not exist a unique cumulative distribution function for complex random variables due to the lack of a complete ordering of complex numbers. However, empirical versions of the rectangular (real bivariate) and ball-wise distribution functions exist, and may be calculated as would be expected. Here we will use the indicator function over inequalities such that $\mathbb{1}\{\cdot\}$ is equal to 1 if $\cdot$ holds and 0 otherwise.

Rectangular empirical distribution function

Given an i.i.d. sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ from a distribution with rectangular c.d.f. F_Z, the empirical rectangular distribution function is given by

$$\widehat{F}_Z(z) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}\left\{\Re(Z_k) \leq \Re(z), \Im(Z_k) \leq \Im(z)\right\}. \tag{6.2.1}$$

This counts the number of observations that ω lies above and to the right of in the complex plane, divided by the number of observations. That is, the empirical rectangular c.d.f. represents the rectangular c.d.f. of a discrete complex uniform distribution with probability mass $\frac{1}{n}$ at each of the observed sample points.

Ball-wise empirical distribution function

Given an i.i.d. sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ from a distribution with ball-wise c.d.f. F_Z^B (Section 4.2), the empirical ball-wise distribution function is given by

$$\widehat{F}_Z^B(z, r) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}\left\{|Z_k - z| \leq r\right\}. \tag{6.2.2}$$

This counts how many observations are within a closed ball of radius r about the point ω, divided by the number of obserations. This is the ball-wise c.d.f. of a complex discrete uniform distribution with mass $\frac{1}{n}$ at each observation. In particular, a circular random variable may have its distribution uniquely defined by the ball-wise distribution function at its mean. In this regard, the empirical ball-wise distribution function of a circular complex random variable around zero reduces to the empirical distribution function of the modulus $|Z|$.

6.2.2 Consistency and bootstrap

Both the rectangular and ball-wise empirical distribution functions will be strongly consistent estimators under the Glivenko-Cantelli theorem (Theorem 8.4, Wasserman (2010)). For the rectangular case, this can be seen by the fact that it holds in the bivariate case. For the ball-wise case, for each fixed z, the empirical ball-wise distribution function will be a strongly consistent estimator of the true ball-wise c.d.f. because it is just a univariate c.d.f. of the radial distribution from z.

One key application of consistency of the empirical distribution functions is that the bootstrap (Efron (1979); Wasserman (2010, Chapter 9)) still holds for complex random samples. Namely, if we can write a quantity of interest, ϑ, as a smooth (Hadamard differentiable) functional of either the rectangular or ball-wise c.d.f. — $\vartheta(F_Z)$ or $\vartheta(F_Z^B)$ respectively. Then we may estimate ϑ by $\widehat{\vartheta} = \vartheta(\widehat{F_Z})$ or $\vartheta(\widehat{F_Z^B})$ respectively and are ensured that $\widehat{\vartheta}$ is consistent (not necessarily strong).

As we have seen that the empirical distribution functions are the corresponding distribution functions of a discrete uniform distribution on the complex plane, simulation methods of drawing randomly from the sample with replacement may still be used to estimate the distribution of an estimator and thus perform inference. This approach may be generalised to other distribution functions in the complex plane, depending upon other partial orderings of the complex numbers, as long as strong consistency of the empirical counterpart is assured. In the literature, for example, other approaches to describing the distribution of complex random variables may include the consideration over half-spaces (Beran et al. (2007)) or using optimal transport theory to develop a "centre-outward" distribution function for which a Glivenko-Cantelli result also holds (Hallin et al. (2021)).

Although, whilst empirical distribution functions estimate the distribution of a random sample, they do not provide a visually intuitive representation of how the probability mass or density is spread over the complex plane. To obtain a more suitable perspective of the distribution of data, descriptive tools such as a histogram or kernel density estimators may be useful.

6.2.3 Histogram density estimates

Often when investigating data, a tool commonly used to obtain a visual impression of the distribution of observations is the histogram. In the univariate setting, a histogram provides the relative frequency of how observations fall into bins of predetermined size and location. The histogram may be interpreted as an estimator of the density f of a real random variable X from an assumedly i.i.d. sample $\mathcal{X} = \{X_1, \ldots, X_n\}$, given by

$$\widehat{f_h}(x) = \frac{1}{nh} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \mathbb{1}\left\{X_k \in B_l\right\} \mathbb{1}\left\{x \in B_l\right\}, \tag{6.2.3}$$

where B_l are bins of width h from some start point x_0,

$$B_l = [x_0 + (l-1)h, x_0 + lh)\,,\ l \in \mathbb{Z}. \tag{6.2.4}$$

The sum in (6.2.3) counts the number of sample observations that are in the same bin as whichever x is passed to the estimated function. The frequency in each bin

is scaled by the number of observations n and the bin width h so that the sum of the histogram bar areas normalises to 1, like a probability density function.

Similarly for complex random variables, we can consider a histogram estimator of the real probability density on the complex plane. By nature of identifying a complex random variable's density with the equivalent bivariate density, existing bivariate histogram estimators may be used, which generalise the univariate histogram by taking rectangular bins — usually of fixed area. In this regard, if we define our complex bins to have area h^2 and located with respect to $z_0 \in \mathbb{C}$ then

$$B_l = [z_0 + h(1+i)(l-1), z_0 + h(1+i)l]_{\text{Rect}}\,,\ l \in \mathbb{Z}, \tag{6.2.5}$$

where $[\cdot]_{\text{Rect}}$ denotes the rectangle formed with the two input points as diagonal corners. Thus the complex rectangular histogram estimator becomes

$$\widehat{f_h}(z) = \frac{1}{nh^2} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \mathbb{1}\left\{Z_k \in B_l\right\} \mathbb{1}\left\{z \in B_l\right\}. \tag{6.2.6}$$

However, as complex random variables also exhibit a polar representation, one could consider also developing a circular histogram estimator, which counts the number of observations in annulus sectors in the complex plane. That is, the bins are given by rays cutting through concentric circles in the complex plane. In terms of the parameterisation of these bins, we propose two methods. First, we can choose to define an argument and radius increment, i.e. fixed angle between rays and distance between concentric circles respectively. Secondly, we can use a single bandwidth parameter $h \in \mathbb{C}$ such that $|h| > 1$ and define the bins as annulus sectors with corners $\left\{z, z\frac{h}{h^*}, zh^2, zhh^*\right\}$. An annulus sector is given graphically in Figure 6.1 parameterised under both the first (red) and second (blue) methods. In the second method, the argument increment is always $2\text{Arg}(h)$ however the radius increment is scaled. As such, we may require an arbitrary distance from the origin to be established (like z_0 for real histograms), else an infinite number of bins of ever decreasing area would be required to approach 0 multiplicatively.

The area of an annulus sector in the first case of a fixed argument increment θ, fixed radius increment r and inner radius of the sector given by R, is $\frac{\theta r}{2}(R + r)$. In the second case of a single bandwidth parameter $h = re^{i\theta}$, the area of a sector with base $z \in \mathbb{C}$ such that $|z| = R$, is given by $R^2\theta(r^2 - 1) = R^2\text{Arg}(h)(|h|^2 - 1)$. Thus we obtain two potential circular histogram estimators from the sample $\{Z_1, \ldots, Z_n\}$ given by

$$\widehat{f}_{r,\theta}(z) = \frac{2}{n\theta r} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \mathbb{1}\left\{Z_k \in B_l\right\} \frac{1}{R_l + r} \mathbb{1}\left\{z \in B_l\right\}, \tag{6.2.7}$$

$$\widehat{f_h}(z) = \frac{1}{n\text{Arg}(h)(|h|^2 - 1)} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \frac{1}{R_l^2} \mathbb{1}\left\{Z_k \in B_l\right\} \mathbb{1}\left\{z \in B_l\right\}, \tag{6.2.8}$$

for the appropriate bins B_l with inner radius R_l under each method.

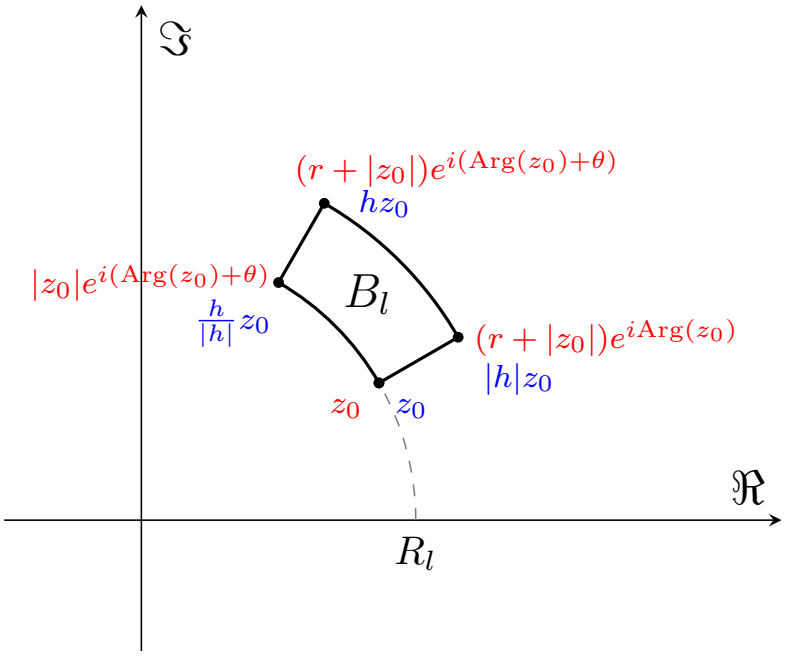

Figure 6.1: Annulus sector under the first (red) and second (blue) parameterisations. For the first parameterisation, r and θ are the radius and argument increments. Under the second we use the single complex parameter h.

A comparative plot of a circular histogram estimator of the first type (fixed argument and radius increment) is shown for circular normally distributed ($\mathcal{N}_{\mathbb{C}}(0,3,0)$) data (Figure 6.2), as well as for rectangular uniform data on square (Figure 6.3), and compared to the bivariate histogram estimator `hist2` from the `squash` package in `R` (Eklund (2020)). The raw data is plotted in Appendix B.1

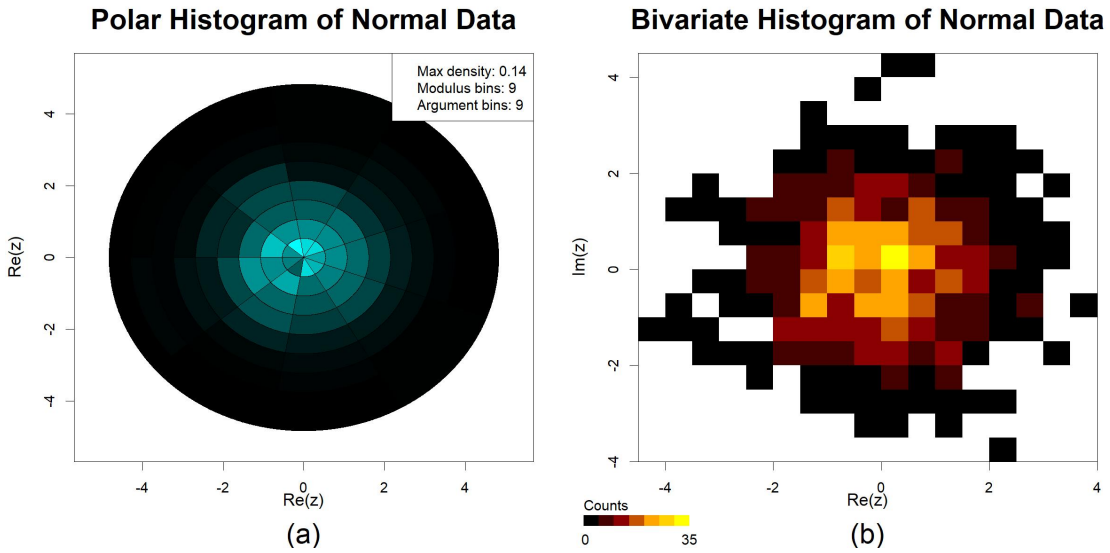

Figure 6.2: Normal data ($n = 1000$) from a $\mathcal{N}_{\mathbb{C}}(0,3,0)$ distribution, (a) a polar histogram and (b) a bivariate histogram.

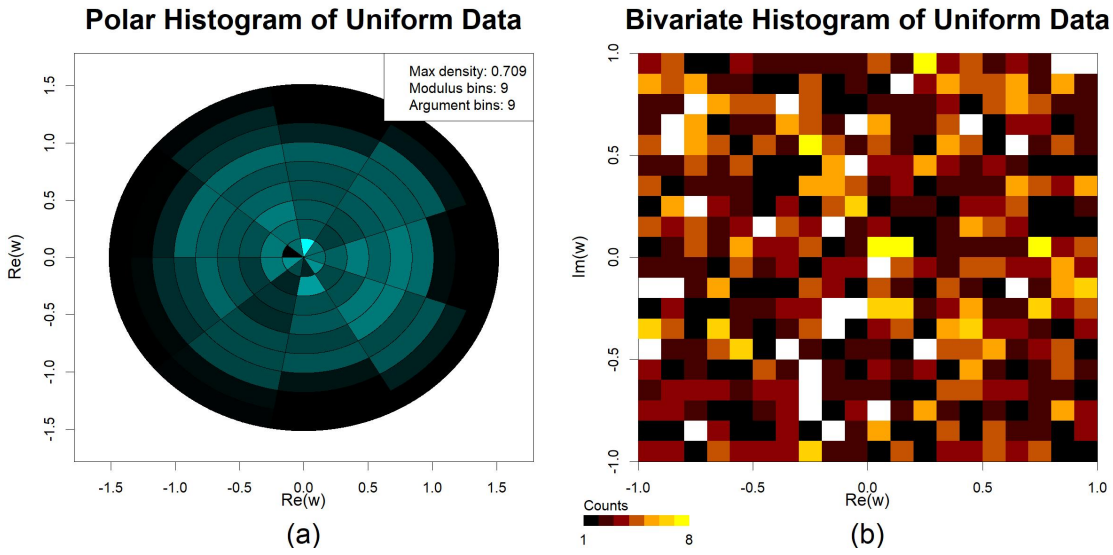

Figure 6.3: Uniform data ($n = 1000$) on a square of side length 2 centred at the origin, (a) a polar histogram and (b) a bivariate histogram.

Indeed, in the circular case the bin construction of the polar histogram indicates a clear outward radiating pattern of decreasing density (brighter colour corresponds to a larger density value). This pattern is somewhat concurred with by the bivariate histogram, however circularity of the data is not as evident using rectangular bins. On the other hand, dealing with rectangular uniform data, we observe distinct anomalies in certain sectors of the polar histogram, suggesting that the data is not from a circular distribution. Indeed, the rectangular construction of the bivariate histogram captures perfectly here the rectangular appearance of the artificial data. When handling complex data, we recommended that both existing rectangular histogram estimators and less-available polar histogram estimators should be at hand in order to obtain a more complete picture of the sampled distribution. The code used to define the circular histogram function in `R` as well as simulate the data used in the above images is provided in Appendix B.2. It would also be possible to construct $3D$ perspective plots to display the relative frequency as the height of each bin over the complex plane.

Kernel density estimation

As an extension of the histogram estimator, kernel density estimators are proposed as a topic of further research for complex random variables. Kernel density estimators provide a continuous, nonparametric estimate of the density by smoothing the probability mass of $\frac{1}{n}$ of each sample observation over an area defined by a bandwidth parameter h and according to a kernel function. An alternative view of this smoothing process is that noise distributed according to the kernel is added to a discrete uniform random variable over the sample values. For example adding zero-mean, normally distributed noise amounts to using a Gaussian kernel. From a rectangular point of view, complex kernel density estimators are expected to behave exactly like existing 2-dimensional kernel density estimators e.g. functions in `R` such as `kde2d` (Ripley et al. (2021)) or `bkde2D` (Wand et al. (2021)). However, in the

context of circular or elliptical symmetry for complex random variables, it is likely beneficial in terms of estimator efficiency to develop a kernel density estimator with a kernel function which smooths the probability mass of $\frac{1}{n}$ over an annulus sector (i.e. in the circular or elliptical shape of the distribution) based off a complex bandwidth parameter h, as given in the second description of the circular histogram (6.2.8). For example, instead of adding in noise, we could consider multiplying the data by random noise, such as log-normally distributed noise, which by the geometry of the complex multiplication will smooth the probability mass around each data point in a rotational fashion. The investigation of such kernels and estimator properties is recommended future research, in particular whether the geometry of the complex multiplication may improve estimator efficiency where rotational symmetry is present in the complex distribution.

6.3 Limit Theorems

This section develops three key limit theorems, commonly used in real statistics, for complex random variables. Firstly and secondly, the weak and strong laws of large numbers which justify the sample mean as a strongly consistent estimator of the expected value. These will rely on Chebychev's inequality which is also generalised to the complex case in a rather straightforward sense. Thirdly, the central limit theorem will be given with an explicit proof in $\mathbb{C}$ for the univariate case, as opposed to stating it holds because the real bivariate central limit theorem holds (Ollila and Koivunen (2010); Yan et al. (2019)). It is also conjectured here that the Lyapunov and Lindeberg conditions for the central limit theorem hold by correspondence with the real bivariate case, however no proof within $\mathbb{C}$ is given as yet.

6.3.1 Chebychev's inequality

Chebychev's inequality can be seen to hold for complex random variables. The statement and proof are a direct generalisation of the statement and proof given in Casella and Berger (2002, p. 122).

Theorem 6.3.1 (Chebychev's inequality)**.** *Let Z be a complex random variable and g a real, non-negative function. Then for any real $\varepsilon > 0$,*

$$\mathbb{P}\left(g(Z) > \varepsilon\right) \leq \frac{\mathbb{E}\left(g(z)\right)}{\varepsilon^2}. \tag{6.3.1}$$

Proof. Let $R = \{z : g(z) > \varepsilon\} \subseteq \mathbb{C}$ and $R_{x,y}$ the corresponding region in $\mathbb{R}^2$. Then for $z = x + iy$,

$$\mathbb{P}\left(g(Z) > \varepsilon\right) = \iint_{R_{x,y}} f_Z(z)\,\mathrm{d}x\,\mathrm{d}y.$$

Now, on the set R, we have that $\frac{g(z)}{\varepsilon} > 1$, so by positivity of f_Z,

$$\begin{aligned}\mathbb{P}\left(g(Z) > \varepsilon\right) &\leq \iint_R \frac{g(z)}{\varepsilon} f_Z(z)\,\mathrm{d}x\,\mathrm{d}y. \\ &\leq \frac{1}{\varepsilon}\int_{\mathbb{R}}\int_{\mathbb{R}} g(z) f_Z(z)\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{\mathbb{E}\left(g(Z)\right)}{\varepsilon}.\end{aligned}$$

□

One key application which will be used in proving the weak law of large numbers is by taking $g(Z) = |Z - \mathbb{E}(Z)|$ and (using the fact that $\frac{g(z)}{\varepsilon} > 1$ implies $\frac{g(z)^2}{\varepsilon^2} > 1$) therefore

$$\mathbb{P}\left(|Z - \mathbb{E}(Z)| > \varepsilon\right) \leq \frac{\mathbb{V}\mathrm{ar}(Z)}{\varepsilon^2}. \tag{6.3.2}$$

6.3.2 Weak Law of Large Numbers

We now state the weak law of large numbers for complex random variables, which justifies the consistency of the sample mean as an estimator of the mean (first moment) of a distribution. The theorem and proof here are given analogous to the real case as reported in Rice (1995, p. 163)

Theorem 6.3.2 (Weak Law of Large Numbers)**.** *Let $Z_1, Z_2, \ldots, Z_k, \ldots$ be a sequence of independent random variables with common mean $\mathbb{E}(Z_k) = \mu$ and variance $\mathbb{V}\mathrm{ar}(Z_k) = \sigma^2 < \infty$. If $\overline{Z}_n = \frac{1}{n}\sum_{k=1}^n Z_k$, then for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(|\overline{Z}_n - \mu| > \varepsilon\right) = 0. \tag{6.3.3}$$

That is, $\overline{Z}_n \xrightarrow{\mathbb{P}} \mu$.

Proof. By linearity and independence, we have that

$$\begin{aligned}\mathbb{E}\left(\overline{Z}_n\right) &= \mu, \\ \mathbb{V}\mathrm{ar}\left(\overline{Z}_n\right) &= \frac{\sigma^2}{n}.\end{aligned}$$

Then, immediately from the application of Chebychev's inequality in (6.3.2), we have that

$$\mathbb{P}\left(|\overline{Z}_n - \mu| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

which tends to 0 as n tends to infinity. □

6.3.3 Strong Law of Large Numbers

The convergence in probability of the sample mean to the true mean may be strengthened to almost sure convergence. We thus state the strong law of large

numbers for complex random variables similar to its statement for the real case in Casella and Berger (2002). A proof within $\mathbb{C}$ based off the assumption of finite variance and kurtosis (fourth absolute central moment) is given in Appendix A.16 inspired by the analogous proof for the real case in Billingsley (1995, p. 85), however we anticipate that these assumptions may be relaxed to require only $\mathbb{E}(|Z_k|) < \infty$ as stated in the theorem.

Theorem 6.3.3 (Strong Law of Large Numbers)**.** *Let $Z_1, Z_2, \ldots, Z_k, \ldots$ be an independent sequence of complex random variables with common mean $\mathbb{E}(Z_k) = \mu$ and such that $\mathbb{E}(|Z_k|) < \infty$. Then for $\overline{Z}_n = \frac{1}{n}\sum_{k=1}^{n} Z_k$ and any $\varepsilon > 0$,*

$$\mathbb{P}\left(\lim_{n\to\infty} |\overline{Z}_n - \mu| > \varepsilon\right) = 0. \tag{6.3.4}$$

That is, $\overline{Z}_n \overset{\text{a.s.}}{\to} \mu$.

6.3.4 Central Limit Theorem

A third important limit theorem that underlies much of statistical inference in the real case is the central limit theorem. This states that the distribution of the sample mean of i.i.d. random variables converges to a normal distribution as n tends to infinity. A similar result is true for complex random variables, however the pseudovariance (degree of improperness) now also needs to be taken into account. Statements of the complex central limit theorem may be found in the literature (Ollila and Koivunen (2010); Yan et al. (2019)), however here we present a proof of the univariate theorem from within $\mathbb{C}$, making use of the Wirtinger derivatives, Wirtinger Taylor expansion and characteristic function. It is unsurprising that the central limit theorem holds analogously in $\mathbb{C}$ as in $\mathbb{R}^2$ because the sample mean and normal distributions coincide and $\mathbb{C}$ and $\mathbb{R}^2$ are isomorphic additive groups. The multivariate version is also given afterwards by Equation (6.3.9).

Theorem 6.3.4 (Complex Central Limit Theorem)**.** *Let $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ be an i.i.d. complex random sample from a distribution with finite variance. Then, for $\overline{Z}_n$ the standardised sample mean,*

$$\frac{\sqrt{n}(\overline{Z}_n - \mathbb{E}(Z))}{\sqrt{\mathbb{V}\mathrm{ar}(Z)}} \quad \xrightarrow{\text{d}} \quad \mathcal{N}_{\mathbb{C}}\left(0, 1, \frac{\rho_Z^2}{\sigma_Z^2}\right),$$

where $\sigma_Z^2 = \mathbb{V}\mathrm{ar}(Z)$ and $\rho_Z^2 = \mathrm{Ps}\mathbb{V}\mathrm{ar}(Z)$.

Proof. (Inspired by Rice (1995) p. 169). Define the random variable Z to have the following first and second moment properties:

$$\begin{aligned} \mathbb{E}(Z) &= \mu \quad \in \mathbb{C}; \\ \mathbb{V}\mathrm{ar}(Z) = \mathbb{V}\mathrm{ar}(Z - \mathbb{E}(Z)) &= \sigma_Z^2 \in \mathbb{R}^+; \\ \mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) = \mathrm{Ps}\mathbb{V}\mathrm{ar}(Z - \mathbb{E}(Z)) &= \rho_Z^2 \in \mathbb{C}, \end{aligned}$$

requiring σ_z^2 (and therefore $|\mu|^2$ and $|\rho_Z|^2$ by Theorem 4.5.2) to be finite.

Define also φ_Z to be the characteristic function of the distribution of Z. Given the i.i.d. random variables $\{Z_1, \ldots, Z_n\}$ drawn from the distribution of Z, the characteristic function has the following properties for all $\omega \in \mathbb{C} \setminus \{0\}$:

$$\varphi_{\sum_{k=1}^n Z_k}(t) = [\varphi_Z(t)]^n; \tag{6.3.5}$$

$$\varphi_{\frac{Z}{\omega}}(t) = \varphi_Z\left(\frac{t}{\omega^*}\right). \tag{6.3.6}$$

Thus, considering the standardised random variable

$$W_n = \frac{\sqrt{n}(\overline{Z_n} - \mathbb{E}(Z))}{\sqrt{\mathbb{V}\mathrm{ar}(Z)}} = \frac{\sum_{k=1}^n Z_k - n\mathbb{E}(Z)}{\sqrt{n\mathbb{V}\mathrm{ar}(Z)}},$$

we have by properties (6.3.5) and (6.3.6) that

$$\varphi_{W_n}(t) = \left[\varphi_{Z-\mathbb{E}(Z)}\left(\frac{t}{\sqrt{n\mathbb{V}\mathrm{ar}(Z)}}\right)\right]^n. \tag{6.3.7}$$

From hereon, without loss of generality we will assume Z has zero mean for concise notation. Consider now the Wirtinger Taylor expansion of $\varphi_Z(s)$ around 0 as per (2.4.6), taking advantage of the moment-generating property (4.7.2) that motivated the definition of the characteristic function for complex random variables.

$$\begin{aligned}
\varphi_Z(s) &= \varphi_Z(0) + s\frac{\partial\varphi_Z}{\partial t}(0) + s^*\frac{\partial\varphi_Z}{\partial t^*}(0) + \frac{s^2}{2}\frac{\partial^2\varphi_Z}{\partial t^2}(0) + \frac{s^{*2}}{2}\frac{\partial^2\varphi_Z}{\partial t^{*2}}(0) + ss^*\frac{\partial^2\varphi_Z}{\partial t\partial t^*}(0) + o(|s|^2) \\
&= 1 + s\mathbb{E}\left(\frac{i}{2}Z^*\right) + s^*\mathbb{E}\left(\frac{i}{2}Z\right) - \frac{s^2}{8}\mathbb{E}\left(Z^{*2}\right) - \frac{s^{*2}}{8}\mathbb{E}\left(Z^2\right) - \frac{|s|^2}{4}\mathbb{E}\left(|Z|^2\right) + o(|s|^2) \\
&= 1 + 0 + 0 - \frac{s^2}{8}\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z)^* - \frac{s^{*2}}{8}\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z) - \frac{|s|^2}{4}\mathbb{V}\mathrm{ar}(Z) + o(|s|^2) \\
&= 1 - \frac{1}{4}\Re\left(s^2\rho_Z^{*2}\right) - \frac{|s|^2}{4}\sigma_Z^2 + o(|s|^2).
\end{aligned}$$

Substituting into (6.3.7) yields,

$$\begin{aligned}
\varphi_{W_n}(t) &= \left[1 - \frac{1}{4}\Re\left(\frac{t^2}{n\sigma_Z^2}\rho_Z^{*2}\right) - \frac{|t|^2}{4n\sigma_Z^2} + o\left(\frac{|t|^2}{n\sigma_Z^2}\right)\right]^n \\
&= \left[1 + \frac{-\frac{1}{4}\left(|t|^2 + \Re\left(t^2\left(\rho_Z^2/\sigma_Z^2\right)^*\right)\right)}{n} + o\left(\frac{|t|^2}{n\sigma_Z^2}\right)\right]^n.
\end{aligned}$$

Taking the limit for n going to infinity results in

$$\lim_{n\to\infty}\varphi_{W_n}(t) = \exp\left\{-\frac{1}{4}\left(|t|^2 + \Re\left(t^2\left(\rho_Z^2/\sigma_Z^2\right)^*\right)\right)\right\},$$

which can be recognised from (5.4.4) as the characteristic function of a $\mathcal{N}_{\mathbb{C}}\left(0, 1, \frac{\rho_Z^2}{\sigma_Z^2}\right)$ distributed random variable. Application of Lévy's Convergence Theorem (see e.g. Theorem 18.1 of Williams (1991, p. 185)) completes the proof. □

An alternate definition of the complex central limit theorem is given by not rescaling the distribution by its variance. That is,

$$\sqrt{n}(\overline{Z}_n - \mathbb{E}(Z)) \quad \xrightarrow{\text{d}} \quad \mathcal{N}_{\mathbb{C}}\left(0, \sigma_Z^2, \rho_Z\right). \tag{6.3.8}$$

This is how the multivariate central limit theorem is given in Ollila and Koivunen (2010, p. 106) without proof, where for i.i.d. complex random vectors $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \in \mathbb{C}^m$ with mean $\boldsymbol{\mu}$, variance $\boldsymbol{\Gamma}$ and pseudovariance $\mathbf{P}$,

$$\sqrt{n}(\overline{\boldsymbol{Z}}_n - \boldsymbol{\mu}) \quad \xrightarrow{\text{d}} \quad \mathcal{N}_{\mathbb{C}^m}(0, \boldsymbol{\Gamma}, \mathbf{P}). \tag{6.3.9}$$

As well as similarities to the real bivariate central limit theorem, we also observe the coincidence with the univariate real case. If Z is a real random variable so $\rho_Z^2 = \sigma_Z^2$, the complex central limit theorem suggests that the sample mean tends to a $\mathcal{N}_{\mathbb{C}}(0, 1, 1)$ distribution, which by Proposition 4.5.4 is equivalent to a $\mathcal{N}_{\mathbb{R}}(0, 1)$ distribution on the real axis as expected.

Additionally, proper random variables necessarily have zero pseudovariance, and hence the standardised sample mean converges to a circular standard complex normal distribution $\mathcal{N}_{\mathbb{C}}(0, 1, 0)$, despite the underlying distribution not necessarily being circular itself.

Lindeberg and Lyapunov Conditions

It is also proposed here (although without proof) that the Lindeberg-Feller and Lyapunov conditions hold for the central limit theorem, allowing it to be applied to independent but non-identical random variables, such as in weighted averages (Billingsley (1995, Section 27)). In the literature, the use of these more general central limit theorem forms is qualitatively justified by the real bivariate application and correspondence of the complex m-variate and real $2m$-variate normal distributions (Yan et al. (2019)). However, the conditions given below are subtly different to imposing the same conditions separately on the Cartesian components. A rigorous proof in $\mathbb{C}$, such as extending the real case in Billingsley (1995, pp. 359–361) is recommended for future research.

Before presenting the conditions, we first define a triangular array. Triangular arrays generalise the notion of an i.i.d. sequence of random variables, given by $Z_1, Z_2, \ldots, Z_n$. A sequence of random variables $\{Z_{nk}\}_{k=1,n\in\mathbb{Z}}^{r_n}$ form a triangular array if for each n, there are r_n independent random variables (usually with r_n increasing in n) (Billingsley (1995)). Visually, one can understand the sequence as

$$Z_{11}, \ldots Z_{1r_1} \tag{6.3.10}$$
$$Z_{21}, Z_{22}, \ldots, Z_{2r_2} \tag{6.3.11}$$
$$Z_{31}, Z_{32}, Z_{33}, \ldots, Z_{3r_3} \tag{6.3.12}$$
$$\ldots \tag{6.3.13}$$

The i.i.d. case is simply where $r_n = n$ and $Z_{nk} = Z_k$ for all $n \in \mathbb{Z}$ (Billingsley (1995, p. 359)).

Proposition 6.3.5 (Lindeberg-Feller Condition)**.** *Let $Z_{n1}, \ldots, Z_{nr_n}$ form a triangular array, such that $\mathbb{E}(Z_{nk}) = 0$, $\mathbb{V}\mathrm{ar}(Z_{nk}) = \sigma_{nk}^2 < \infty$ and $\mathrm{Ps}\mathbb{V}\mathrm{ar}(Z_{nk}) = \rho_{nk}^2$ for all $k \in \{1, \ldots, r_n\}$, $n \in \mathbb{Z}$. Then the Lindeberg condition is said to hold if for any $\varepsilon > 0$ and for $\tilde{\sigma}_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2$, $\tilde{\rho}_n^2 = \sum_{k=1}^{r_n} \rho_{nk}^2$,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \mathbb{E}\left(|Z_{nk}|^2 ; |Z_{nk}| > \varepsilon \tilde{\sigma}_n^2\right) = 0. \tag{6.3.14}$$

Should this hold, then

$$\frac{1}{\tilde{\sigma}_n^2} \sum_{k=1}^{r_n} Z_{nk} \xrightarrow{\mathrm{d}} \mathcal{N}_{\mathbb{C}}\left(0, 1, \frac{\tilde{\rho}_n^2}{\tilde{\sigma}_n^2}\right). \tag{6.3.15}$$

We also propose (again without proof) that the stronger Lyapunov condition holds where if, for some $\delta > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{\tilde{\sigma}_n^{2+\delta}} \mathbb{E}\left(|Z_{nk}|^{2+\delta}\right) = 0, \tag{6.3.16}$$

then convergence as per (6.3.15) occurs. Note that independence within the Z_{nk}'s for each n is still required.

Should either of these relaxed central limit theorem variants hold, then we can justify the normality of discrete Fourier transformed data which are complex data obtained by a complex weighted sum of real observations. The discrete Fourier transform is a common method by which complex data arise in application and will be discussed further in Section 7.3.1.

6.4 Parameter Estimation

Parameters of complex distributions may be estimated by similar methods as with real random variables. For example, maximum likelihood estimation is still possible because by nature of the probability density functions — and therefore likelihoods — being real-valued, one can still find a maximum of the likelihood using Wirtinger calculus. This section will not treat fundamental notions of sufficient statistics or Cramér-Rao lower bounds on estimator variance, although we do remark that these topics warrant further investigation in the complex case. A brief application of sufficient statistics and variance bounds for complex statistics is given in Schreier and Scharf (2010, Section 2.4), however there the authors state results by direct correspondence with the bivariate real case. We believe there is value in generalising this theory to a complex setting to make the field of complex statistics self-contained and so that uniquely complex results are not overlooked. We discuss first two methods of point estimation: method of moments and maximum likelihood, before discussing probability and confidence regions.

6.4.1 Method of moments

The strong law of large numbers gives strong consistency of the sample mean as an estimator of the true mean of a distribution. The sample mean is also an unbiased estimator in the sense that $\mathbb{E}\left(\overline{Z}_n\right) = \mu$. The weak law of large numbers implies consistency (convergence in probability) of the following estimators (Rice (1995, pp. 246–253); Casella and Berger (2002, pp. 312–314)) which generalise the method of moments estimators to complex random variables. For $s, t \in \mathbb{Z}$, define

$$\mu_{st} = \mathbb{E}\left(Z^s Z^{*t}\right). \tag{6.4.1}$$

Then given an i.i.d. sample $\{Z_1, \ldots, Z_n\}$, the method of moments estimators are,

$$\widehat{\mu}_{st} = \frac{1}{n}\sum_{k=1} Z_k^s Z_k^{*t}. \tag{6.4.2}$$

In the context of estimating distribution parameters which may be written in terms of moments of the distribution, plugging in the corresponding $\widehat{\mu}_{st}$ estimators gives the method of moments estimator of the parameter. If the parameter is a continuous function of the moments, then consistency of the $\widehat{\mu}_{st}$ estimators ensures consistency of the parameter estimates (Rice (1995, pp.252–253)).

Example 6.4.1. *The method of moments estimator of the mean and the second order moments are given by*

$$\widehat{\mu}_{10} = \overline{Z}_n, \tag{6.4.3}$$

$$\widehat{\mu}_{11} = \frac{1}{n}\sum_{k=1}^{n} |Z_k|^2, \tag{6.4.4}$$

$$\widehat{\mu}_{20} = \frac{1}{n}\sum_{k=1}^{n} Z_k^2, \tag{6.4.5}$$

These yield the following method of moments estimators for the variance,

$$\widehat{\sigma}^2 = \widehat{\mu}_{11} - |\widehat{\mu}_{10}|^2 \tag{6.4.6}$$

$$= \frac{1}{n}\sum_{k=1}^{n} |Z_k - \overline{Z}_n|^2, \tag{6.4.7}$$

and pseudovariance,

$$\widehat{\rho}^2 = \widehat{\mu}_{20} - \widehat{\mu}_{10}^2 \tag{6.4.8}$$

$$= \frac{1}{n}\sum_{k=1}^{n} (Z_k - \overline{Z}_n)^2. \tag{6.4.9}$$

We observe that despite being consistent, the estimates of the variance and pseudovariance given in (6.4.7) and (6.4.9) are in fact biased estimates. However,

by replacing the factor of $\frac{1}{n}$ with $\frac{1}{n-1}$ we obtain unbiased estimators, exactly as in the real case. Thus we will denote the unbiased estimators by

$$s_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} |Z_k - \overline{Z}_n|^2, \tag{6.4.10}$$

$$r_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} (Z_k - \overline{Z}_n)^2. \tag{6.4.11}$$

Proposition 6.4.2. *The variance estimator s_n^2 of an i.i.d. sample $\{Z_1, \ldots, Z_k\}$ from a $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$ distribution is such that*

$$\frac{(n-1)s_n^2}{\sigma^2} \sim \text{Gamma}(n-1, 1). \tag{6.4.12}$$

The proof is given in Appendix A.17, however we recognise the similarity to the real case as if $X \sim \text{Gamma}(n-1, 1)$, then $2X \sim \chi^2_{2n-2}$. This also motivates why in Chapter 5 we preferred to present the complex t-distribution in terms of its complex degrees of freedom, which are understood analogously to the real case in terms of the sample size, rather than constantly having to double the degrees of freedom when converting familiar results of real statistics to complex statistics.

6.4.2 Maximum likelihood

By retaining real probability and density functions, we can construct real-valued likelihoods, which may be maximised in $\mathbb{R}$ over complex parameters using Wirtinger calculus. As with real statistics (Rice (1995, Section 8.5); Casella and Berger (2002, Section 7.2.2)), given a sample of complex observations $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ we define the likelihood to be given by the joint density[1] of the random variables observed in the sample, treating the likelihood as a function of the distribution parameters (denoted here by $\boldsymbol{\vartheta} \in \mathbb{C}^p$) rather than the sample observations. The likelihood is given by

$$\mathcal{L}(\boldsymbol{\vartheta}\,;\mathcal{Z}) = f_{\mathcal{Z}}(Z_1, Z_2, \ldots, Z_n\,;\boldsymbol{\vartheta}), \tag{6.4.13}$$

and in particular if the sample is assumed to be i.i.d., then

$$\mathcal{L}(\boldsymbol{\vartheta}\,;\mathcal{Z}) = \prod_{k=1}^{n} f_Z(Z_k\,;\boldsymbol{\vartheta}). \tag{6.4.14}$$

The maximum likelihood estimator of $\boldsymbol{\vartheta}$ is thus defined to be

$$\widehat{\boldsymbol{\vartheta}} = \widehat{\boldsymbol{\vartheta}}(\mathcal{Z}) = \text{argmax}_{\boldsymbol{\vartheta}}\, \mathcal{L}(\boldsymbol{\vartheta}\,;\mathcal{Z}). \tag{6.4.15}$$

At times, it will be easier to work with the log-likelihood,

$$\ell(\boldsymbol{\vartheta}\,;\mathcal{Z}) = \ln(\mathcal{L}(\boldsymbol{\vartheta}\,;\mathcal{Z})), \tag{6.4.16}$$

[1] By using the density we assume the sampled random variables are absolutely continuous. In the discrete setting, the joint p.m.f. can be used.

recalling that ln denotes the real logarithm. As the real logarithm is monotonic increasing, maximising ℓ is equivalent to maximising $\mathcal{L}$ and can be done via the methods of Wirtinger calculus presented in Subsection 2.4.4.

Example 6.4.3. *Here we consider the maximum likelihood estimators of μ and σ^2 of a $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$ distribution, given an i.i.d. sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$, which like the real case will also be the method of moments estimators. Write the likelihood*

$$\mathcal{L}(\mu, \sigma^2 ; \mathcal{Z}) = \left(\frac{1}{\pi\sigma^2}\right)^n \exp\left\{-\frac{1}{\sigma^2}\sum_{k=1}^{n} |Z_k - \mu|^2\right\},$$

which yields the log-likelihood

$$\ell(\mu, \sigma^2 ; \mathcal{Z}) = -n\ln(\pi) - n\ln(\sigma^2) - \frac{1}{\sigma^2}\sum_{k=1}^{n}(Z_k - \mu)(Z_k^* - \mu^*).$$

Taking the Wirtinger derivative with respect to μ, we obtain

$$\partial_\mu \ell = \frac{1}{\sigma^2}\sum_{k=1}^{n}(Z_k^* - \mu^*).$$

One can verify that $\partial_{\mu\mu^*}\ell = -\frac{n}{\sigma^2} < 0$ and $\partial_{\mu\mu}\ell = 0$. So by the second-order test in (2.4.11), we have that

$$\begin{aligned}(\partial_{\mu\mu^*}\ell)^2 - |\partial_{\mu\mu}\ell|^2 &= \frac{n^2}{\sigma^2} > 0,\\ \partial_{\mu\mu^*}\ell + \Re(\partial_{\mu\mu}\ell) &= -\frac{n}{\sigma^2} < 0,\end{aligned}$$

and so setting $\partial_\mu \ell = 0$ will maximise ℓ. Thus we obtain, given $\sigma^2 > 0$, the maximum likelihood estimator

$$\widehat{\mu}^* = \overline{Z^*}_n \implies \widehat{\mu} = \overline{Z}_n.$$

Similarly maximising in σ^2,

$$\partial_{\sigma^2}\ell = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{k=1}^{n}|Z_k - \mu|^2,$$

and setting this equal to zero results in

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}|Z_k - \overline{Z}_n|^2.$$

An analogous (although algebraically heavier) result for the more general case where there is non-zero pseudovariance ρ^2 also holds in that we obtain the method of moments estimator in (6.4.9). In higher dimensions, these results generalise by

replacing $|Z_k - \mu|^2$ and $(Z_k - \mu)^2$ with $(\boldsymbol{Z}_k - \boldsymbol{\mu})^{\mathcal{H}}(\boldsymbol{Z}_k - \boldsymbol{\mu})$ and $(\boldsymbol{Z}_k - \boldsymbol{\mu})^{\top}(\boldsymbol{Z}_k - \boldsymbol{\mu})$ respectively (Ducharme et al. (2016, Proposition 2)).

6.4.3 Quantiles and confidence regions

As there is no complete ordering on $\mathbb{C}$, we do not have a canonical c.d.f., and therefore no canonical definition of the quantiles of an arbitrary distribution. Hallin et al. (2021) propose a "centre-outward" distribution and quantile function for $\mathbb{R}^d$ using notions of measure transportation which is claimed to preserve the usefulness of ranks and quantiles inasmuch as they are used for univariate real statistics. Whilst such an approach may be advantageous also in $\mathbb{C}$, we will not delve into measure transportation and optimal transport theory here.

We could use the ball-wise or rectangular c.d.f.'s to obtain ball-wise or rectangular quantiles and therefore confidence balls and rectangles for point estimates. However, in the case of non-circular or non-rectangular distributions, circular or rectangular regions of a certain probability will not necessarily be the smallest region to contain the same probability under the distribution. For example, a 95% probability ball for an elliptically symmetric distribution will necessarily be larger than a 95% probability ellipse which matches the symmetry of the distribution.

The construction of quantiles and subsequently confidence regions for arbitrary complex distributions remains an open topic of interest, for example to generalise areas of hypothesis testing and robust statistics. For circular distributions, quantiles may be defined via the ball-wise c.d.f., such that balls with radius given by the $1-\alpha$ quantile of the distribution of $|Z|$ will contain $100(1-\alpha)\%$ of the probability mass of the distribution of Z. We envisage that a quantile description should also be possible for elliptically symmetric distributions (constructed by an $\mathbb{R}$-linear transformation of a circular distribution) by cleverly adjusting the ball-wise quantiles of the underlying circular distribution. We do not presently provide a method for an arbitrary distribution, but demonstrate this idea with the complex normal distribution. This also allows the construction of asymptotic confidence regions for estimators where the central limit theorem is applicable.

Example 6.4.4. *Consider $W \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ which can be written in terms of a circular normally distributed random variable $Z_0 \sim \mathcal{N}_{\mathbb{C}}(0, 1, 0)$,*

$$W = \mu + aZ_0 + bZ_0^*,$$

for

$$a = \frac{\sqrt{\sigma^2 + |\rho|^2} + \sqrt{\sigma^2 - |\rho|^2}}{2} e^{i\mathrm{Arg}(\rho)}, \tag{6.4.17}$$

$$b = \frac{\sqrt{\sigma^2 + |\rho|^2} - \sqrt{\sigma^2 - |\rho|^2}}{2} e^{i\mathrm{Arg}(\rho)}. \tag{6.4.18}$$

Given we know a circular standard complex normal distribution has an exponentially distributed modulus squared, define $K_{1-\alpha} \coloneqq \sqrt{-\ln(\alpha)}$, which is the square

root of the $(1-\alpha)$ *quantile of an* Exp(1) *distribution. We have that for* $\overline{B}_{1-\alpha} = \{z \in \mathbb{C} : |z| \leq K_{1-\alpha}\}$,

$$\mathbb{P}\left(Z_0 \in \overline{B}_{1-\alpha}\right) = 1 - \alpha.$$

Then the filled ellipse formed by the image of the unit disc $D^1 = \{z \in \mathbb{C} : |z| \leq 1\}$ *under the map*

$$z \mapsto \mu + K_{1-\alpha} a z + K_{1-\alpha} b z^*, \tag{6.4.19}$$

contains $100(1-\alpha)\%$ *of the probability under the distribution of* W. *In Figure 6.4, we show the* 50% *and* 95% *ellipse boundaries (density contours) for a* $\mathcal{N}_{\mathbb{C}}\left(1+i, 2, 1-\frac{i}{2}\right)$ *distribution. The* `R` *code that generates these ellipses may be found in Appendix B.3.*

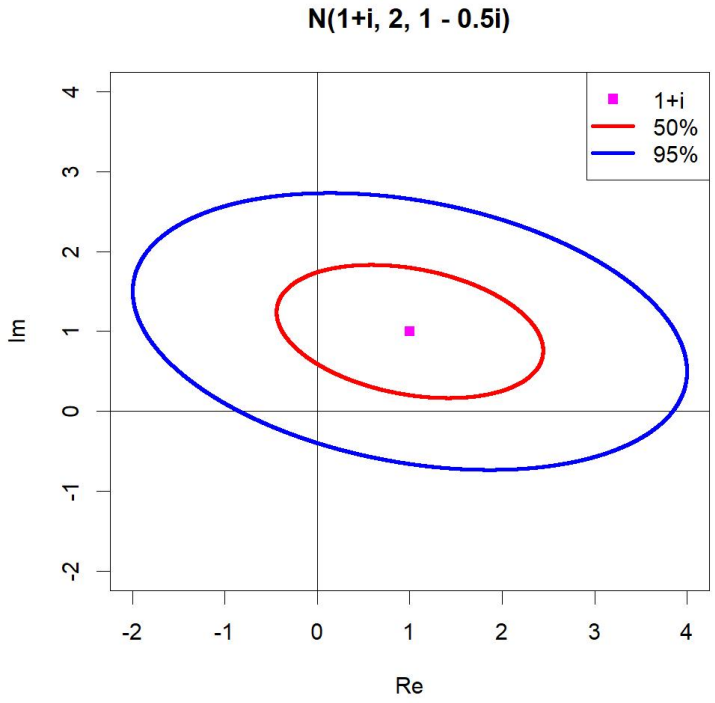

Figure 6.4: The 50% and 95% probability ellipses for a $\mathcal{N}_{\mathbb{C}}\left(1+i, 2, 1-\frac{i}{2}\right)$ distribution.

The method given above coincides (as would be expected) with the method for constructing contours of the corresponding bivariate normal density as per e.g. Johnson and Wichern (2007, pp. 153–155). If we have an i.i.d. sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ from a $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \rho^2)$ distribution, then

$$\overline{Z}_n \sim \mathcal{N}_{\mathbb{C}}\left(\mu, \frac{\sigma^2}{n}, \frac{\rho^2}{n}\right),$$

and the confidence ellipse for the mean μ is given by (6.4.19), but replacing μ with $\overline{Z}_n$ and using $\frac{a}{\sqrt{n}}, \frac{b}{\sqrt{n}}$ instead of the a, b given in (6.4.17)–(6.4.18).

6.5 Hypothesis Testing

Finally, we look at hypothesis testing for complex random variables. This section will not generalise the fundamentals of hypothesis testing from real statistics, rather it provides an overview of the testing methods that are out there and encourages future research into hypothesis testing for complex random variables. Of course hypothesis tests will depend upon the ability to define, and describe quantiles, for arbitrary distributions which itself requires further research. For example, we will see in the next chapter that a linear regression coefficient follows a circular complex t-distribution under the assumption of circular normal errors. However if the errors are assumed to follow a non-circular normal distribution, we have not yet defined the respective distribution of the coefficients, which would involve the modulus squared distribution of an arbitrary normal as given in Cooper and Farid (2020).

There are already several hypothesis tests in the literature, in particular from Eriksson et al. (2010), Schreier and Scharf (2010) and Ducharme et al. (2016). Firstly, Eriksson et al. (2010) make use of a generalised likelihood ratio test to test for properness (which in the normal case yields circularity). Secondly, Schreier and Scharf (2010) again use generalised likelihood ratio tests to test for non-zero correlation, different means and different covariances between complex valued time series. Thirdly, Ducharme et al. (2016) use the empirical characteristic function to develop a virtually omnibus goodness-of-fit test for the complex normal distribution. All of these hypothesis tests make use of the identification of the distribution of a complex random variable with the corresponding real bivariate distribution (and characteristic function) to employ familiar hypothesis testing procedures from the real setting.

Here, we present both the goodness-of-fit test for normality from Ducharme et al. (2016) and the test for properness from Eriksson et al. (2010), which are useful in practice for justifying the use of normality, or at least properness so that the central limit theorem may be invoked with a circular normal distribution.

6.5.1 A goodness-of-fit test for normality

The test from Ducharme et al. (2016) is an asymptotic test employing results of the empirical characteristic process from Csörgo (1981) and correspondence of the complex and real bivariate characteristic functions. Specifically, if the empirical characteristic function for an i.i.d. sample $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$ is given by

$$\widehat{\varphi}_n(t) = \frac{1}{n}\sum_{k=1}^{n} e^{\frac{i}{2}(tZ_k^* + t^* Z_k)}, \tag{6.5.1}$$

then the process

$$U_n(\cdot) = \sqrt{n}(\widehat{\varphi}_n(\cdot) - \varphi_0(\cdot)), \tag{6.5.2}$$

has zero mean and known covariance structure under the assumption that φ_0 is the true characteristic function of the underlying distribution. Ducharme et al. (2016) then justify by use of the central limit theorem that as $n \to \infty$, for $\boldsymbol{t} =$

$(t_1, \ldots, t_m)^\top \in \mathbb{C}^m$ the vector $\boldsymbol{U}_n(\boldsymbol{t}) = (U_n(t_1), \ldots, U_n(t_m))^\top$ tends to an m-dimensional complex normal distribution with variance and pseudovariance matrices

$$\boldsymbol{\Gamma} = [\varphi_0(t_k - t_l) - \varphi_0(t_k)\varphi_0(t_l)^*], \tag{6.5.3}$$
$$\mathbf{P} = [\varphi_0(t_k + t_l) - \varphi_0(t_k)\varphi_0(t_l)]. \tag{6.5.4}$$

Then (stepping through bivariate real statistics) if we treat the augmented vector and variance-pseudovariance matrix as

$$\underline{\boldsymbol{U}}_n = \begin{pmatrix} \boldsymbol{U}_n \\ \boldsymbol{U}_n^* \end{pmatrix}, \tag{6.5.5}$$
$$\boldsymbol{\Gamma}_{\mathbf{P}} = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{P} \\ \mathbf{P}^* & \boldsymbol{\Gamma}^* \end{pmatrix}, \tag{6.5.6}$$

we have that the test statistic

$$\xi_n = \underline{\boldsymbol{U}}_n^{\mathcal{H}} \boldsymbol{\Gamma}_{\mathbf{P}}^{+} \underline{\boldsymbol{U}}_n \xrightarrow{\text{d}} \chi^2_{\text{rank}(\boldsymbol{\Gamma}_{\mathbf{P}})}, \tag{6.5.7}$$

under the null hypothesis $H_0 : \varphi_Z = \varphi_0$ for φ_0 the characteristic function of a specific normal distribution. Note that above, $\boldsymbol{\Gamma}_{\mathbf{P}}^{+}$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{\Gamma}_{\mathbf{P}}$. In this test, we reject the null hypothesis that the data comes from the specified normal distribution at a $100\alpha\%$ significance level if the test statistic ξ_n (for n large) is greater than the $1 - \alpha$ quantile of the $\chi^2_{\text{rank}(\boldsymbol{\Gamma}_{\mathbf{P}})}$ distribution.

The above test has a simple hypothesis in the sense that we test whether the data follow a prespecified normal distribution. Ducharme et al. (2016) also present a composite hypothesis test of whether the data follow an arbitrary normal distribution. We do not provide the full detail of the composite test here, rather acknowledge that such a test exists and can be implemented where desired. For example Ducharme et al. (2016) use their goodness-of-fit test to justify assumptions of normality in terms of fMRI data in Rowe and Logan (2004), which validates subsequent inference applied regarding least-squares linear modelling.

6.5.2 A test for properness

In Eriksson et al. (2010), a test for properness, i.e. zero pseudovariance, is given in the form of a generalised likelihood ratio test based off translation of notation from a test for sphericity in real multivariate statistics. The authors present both the standard test statistics and an adjusted statistic that is robust to departures from normality, provided the distribution remains within a class of elliptically symmetric distributions with finite kurtosis (central absolute fourth moment).

Define $q = \frac{\rho^2}{\sigma^2}$ to be the "circularity quotient", namely the ratio of the pseudovariance to variance which arises in the central limit theorem as the limiting pseudovariance of a standardised complex random variable. Then if $\widehat{q} = \frac{\widehat{\rho}^2}{\widehat{\sigma}^2}$ for the method of moments estimators of ρ^2 and σ^2, we have that the test statistic

$$L_n = -(n-1)\ln(\sqrt{1 - |\widehat{q}|^2}) \xrightarrow{\text{d}} \text{Exp}(1) \tag{6.5.8}$$

under $H_0 : \rho^2 = 0$ (against $H_1 : \rho^2 \neq 0$). Again, this is an asymptotic test, relying upon large n. Note that in Eriksson et al. (2010), they report a doubled test statistic (no square root in ln) with convergence to a χ^2_2 distribution. However, for consistency with previous results in a complex setting, we prefer to work with the Exponential and Gamma distributions.

The adjusted statistic is given by

$$L_n^{\text{adj}} = \frac{2 + |\widehat{q}|^2}{\widehat{k}} L_n \xrightarrow{\text{d}} \text{Exp}(1), \tag{6.5.9}$$

where

$$\widehat{k} = \frac{\mathbb{E}\left(|\widehat{Z - \mathbb{E}(Z)}|^4\right)}{\widehat{\sigma}^4} = n \frac{\sum_{k=1}^n |Z_k - \overline{Z}_n|^4}{\left(\sum_{k=1}^n |Z_k - \overline{Z}_n|^2\right)^2}, \tag{6.5.10}$$

estimates the standardised kurtosis. In this test, we reject the null hypothesis of properness at a $100\alpha\%$ significance level if the test statistic is greater than the $1-\alpha$ quantile of an Exp(1) distribution.

Overall, many features of real statistics may be generalised to complex statistics by correspondence of the distribution and characteristic functions with their bivariate counterparts. However, we recommend future research into the theory of complex statistics should aim to develop as a self-contained theory in $\mathbb{C}$ and thus maintains an awareness of the differences to real bivariate statistics, namely the presence of complex multiplication. For example, just as the empirical characteristic process may be used to test goodness-of-fit as in Ducharme et al. (2016), we expect a similar result should be theoretically possible using the empirical counterpart of the Mellin transform, which will never be noticed if statistical procedures like hypothesis tests are only ever translated from bivariate statistics.

Chapter 7

Dependence Measures and Linear Modelling

This chapter will discuss how to describe the dependence between complex random variables, and subsequently provide a brief overview of linear modelling with complex random variables. In particular the similarity to real statistics in the context of the ordinary least squares estimator is given.

Throughout this chapter, we will refer to affine linear transforms as follows. An affine transform of z is of the form $\alpha + \beta z$ for $\alpha, \beta \in \mathbb{C}$. A conjugate affine transform of z will be of the form $\alpha + \beta z^*$ for $\alpha, \beta \in \mathbb{C}$. Finally, an $\mathbb{R}$-linear (also called "widely linear" e.g. Picinbono and Chevalier (1995)) transform of z will be of the form $\alpha + \beta z + \gamma z^*$ for $\alpha, \beta, \gamma \in \mathbb{C}$.

7.1 Correlation

We begin by discussing and interpreting the Pearson correlation (hereafter just 'correlation') between two complex random variables. Whilst heavily used in many statistical applications, we discuss the correlation as a measure of dependence with the important caveat that it is only a meaningful measure of *dependence* when we are working in a multivariate complex normal setting, i.e. the joint distribution of the complex random variables is a complex normal distribution. Whilst non-zero correlation may be useful in other settings to indicate a lack of independence, without the assumption of normality, the correlation is not very effective at measuring the strength of general dependence between two random variables.

7.1.1 Definition

Just as how complex random variables exhibit a variance and pseudovariance, we may calculate a correlation and pseudocorrelation between two complex random variables.

Definition 7.1.1. *The correlation between two complex random variables Z and W with finite variance is given by*

$$\mathbb{C}\mathrm{orr}\left(W, Z\right) \coloneqq \frac{\mathbb{C}\mathrm{ov}\left(W, Z\right)}{\sqrt{\mathbb{V}\mathrm{ar}\left(W\right)\mathbb{V}\mathrm{ar}\left(Z\right)}} \in \mathbb{C}. \tag{7.1.1}$$

The pseudocorrelation is given by

$$\mathrm{Ps}\mathbb{C}\mathrm{orr}\left(W, Z\right) \coloneqq \mathbb{C}\mathrm{orr}\left(W, Z^*\right) \frac{\mathbb{C}\mathrm{ov}\left(W, Z^*\right)}{\sqrt{\mathbb{V}\mathrm{ar}\left(W\right)\mathbb{V}\mathrm{ar}\left(Z\right)}} \in \mathbb{C}. \tag{7.1.2}$$

Furthermore, both the correlation and pseudocorrelation have modulus less than or equal to 1. These definitions are consistent with the literature, and in particular

much of what follows in regards to interpretation can be found in Chapter 4 of Schreier and Scharf (2010), motivated from a context of widely linear modelling (originally set out in Picinbono and Chevalier (1995)).

7.1.2 Interpreting the correlation — rotation

The correlation, as with the real case, measures the strength of a linear relationship between two random variables. Specifically, in the complex setting, linear combinations involve the complex multiplication, which leads to the geometric interpretation of the correlation in terms of a rotation.

The correlation can be interpreted in terms of its polar coordinates. The modulus indicates the strength of the linear relationship (as in the real case). The argument indicates the 'direction' of the linear relationship in $\mathbb{C}^2$. That is, the angle by which a complex random variable W appears as a rotation of another random variable Z, were these variables affine transforms of each other. This generalises the real correlation where a negative correlation suggests W appears as a rotation by π, i.e. negation, of Z.

For example, if we have a perfect linear relationship such that W is an affine transform of another Z. Then for some $\alpha, \beta \in \mathbb{C}$,

$$W = \alpha + \beta Z, \tag{7.1.3}$$

and

$$\mathbb{C}\text{orr}\,(W, Z) = e^{i\text{Arg}(\beta)}. \tag{7.1.4}$$

The modulus of 1 above suggests a perfect linear relationship, just like the real correlation. Also, the argument of the complex correlation indicates the angle by which W is a scaled rotation of Z. The coefficient β here is given by

$$\beta = \frac{\mathbb{C}\text{ov}\,(W, Z)}{\mathbb{V}\text{ar}\,(Z)},$$

which will be seen to appear as the simple least-squares estimator in the discussion of linear modelling later.

In fact, the modulus of the correlation is 1 if and only if the relation between W and Z is almost-surely affine linear as in (7.1.3). This is formalised and proven in Proposition 7.1.2. Importantly, where $|\mathbb{C}\text{orr}\,(W, Z)\,| = 1$, there may still be non-trivial pseudocorrelation if Z is not proper, but this does not suggest that W needs to be given by a general $\mathbb{R}$-linear combination of Z.

Importantly, as the $\mathbb{C}$ov function is an Hermitian operator, not symmetric, $\mathbb{C}\text{orr}\,(W, Z) = \mathbb{C}\text{orr}\,(Z, W)^*$, so the order of the variables now matters to interpret the argument of the correlation. Specifically the argument of the correlation indicates the angle by which the *first* variable is a linear rotation of the *second*.

The correlation may be estimated by the ratio of the corresponding method of moments estimators analogous to the real case such that, if $\xi = \mathbb{C}\text{orr}\,(W, Z)$ and

we have the samples $\mathcal{W} = \{W_1, \ldots, W_n\}$ and $\mathcal{Z} = \{Z_1, \ldots, Z_n\}$, then

$$\widehat{\xi} = \frac{\sum_{k=1}^{n}(W_k - \overline{W}_n)(Z_k - \overline{Z}_n)^*}{\sqrt{\sum_{k=1}^{n}|W_k - \overline{W}_n|^2 \sum_{k=1}^{n}|Z_k - \overline{Z}_n|^2}}. \tag{7.1.5}$$

7.1.3 Interpreting the pseudocorrelation — reflection

Unlike the real case, however, we now have a pseudocorrelation. Like where the correlation has modulus 1 when W is an affine transform of Z, the pseudocorrelation will have modulus 1 when W is a conjugate affine transform of Z. That is, if

$$W = \alpha + \beta Z^*, \tag{7.1.6}$$

we have that

$$\mathrm{Ps}\mathbb{C}\mathrm{orr}\,(W, Z) = e^{i\mathrm{Arg}(\beta)}. \tag{7.1.7}$$

Whilst the modulus of the pseudocorrelation indicates the extent to which W is a conjugate affine transform of Z, the argument does not have the same rotational interpretation due to conjugation. Rather, it is a rotation by $\mathrm{Arg}(\beta)$ composed with a reflection in the $\Re$ axis by conjugation. This composition can be summarised as a reflection of Z in the line through 0 and $e^{\frac{i}{2}\mathrm{Arg}(\beta)}$, being the (principal) square root of the pseudocorrelation (Schreier and Scharf (2010); Adali and Schreier (2014)).

Again, the pseudocorrelation has modulus 1 if and only if W is almost-surely a conjugate affine transform of Z. There may still be non-zero correlation, however this does not mean that W needs to be a general $\mathbb{R}$-linear transform of Z.

Where the pseudocorrelation has modulus less than 1, this indicates the strength of the reflectional relationship with the principal argument being twice the angle of the line of reflection. We also note that as the pseudocovariance function is symmetric, we have here that $\mathrm{Ps}\mathbb{C}\mathrm{orr}\,(W, Z) = \mathrm{Ps}\mathbb{C}\mathrm{orr}\,(Z, W)$, without need for conjugation as in the case of the correlation.

Similar to the correlation, an estimator for the pseudocorrelation $\kappa = \mathrm{Ps}\mathbb{C}\mathrm{orr}\,(W, Z)$ is given by

$$\widehat{\kappa} = \frac{\sum_{k=1}^{n}(W_k - \overline{W}_n)(Z_k - \overline{Z}_n)}{\sqrt{\sum_{k=1}^{n}|W_k - \overline{W}_n|^2 \sum_{k=1}^{n}|Z_k - \overline{Z}_n|^2}}. \tag{7.1.8}$$

7.1.4 Combining the correlations

As detailed above, when the correlation or pseudo-correlation are 1, we can understand that the variables are an affine or conjugate affine transform of each other. In fact, this leads us to the following (novel) proposition, which also highlights the real case as an instance where both correlations coincide.

Proposition 7.1.2. *For two complex random variables, W and Z with finite variance, we have the following two relationships:*

$$|\mathbb{C}\text{orr}\,(W,Z)\,| = 1 \iff W \overset{\text{a.s.}}{=} \alpha + \beta Z, \tag{7.1.9}$$

$$|\text{Ps}\mathbb{C}\text{orr}\,(W,Z)\,| = 1 \iff W \overset{\text{a.s.}}{=} \gamma + \delta Z^*, \tag{7.1.10}$$

for constants $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. Furthermore, if both the correlation and pseudocorrelation have modulus of 1, then Z is distributed along a line in the complex plane, as is W, being an affine transform (scaled rotation and translation) of Z.

Proof. The proof here is inspired by the proof of the real case in Casella and Berger (2002, pp. 172–173). Let $\xi = \mathbb{C}\text{orr}\,(W,Z)$ have modulus 1 and denote $\sigma_W^2 = \mathbb{V}\text{ar}\,(W)$, $\sigma_Z^2 = \mathbb{V}\text{ar}\,(Z)$, $\sigma_{WZ} = \mathbb{C}\text{ov}\,(W,Z)$. Then, $\sigma_{WZ} = \xi\sigma_W\sigma_Z$. Consider now the random variable $W - \xi\frac{\sigma_W}{\sigma_Z}Z$ with variance given by

$$\begin{aligned}
\mathbb{V}\text{ar}\left(W - \xi\frac{\sigma_W}{\sigma_Z}Z\right) &= \sigma_W^2 + |\xi|^2\sigma_W^2 - \xi^*\frac{\sigma_W}{\sigma_Z}\sigma_{WZ} - \xi\frac{\sigma_W}{\sigma_Z}\sigma_{ZW}^* \\
&= 2\sigma_W^2 - |\xi|^2\sigma_W^2 - |\xi|^2\sigma_W^2 \\
&= 0.
\end{aligned}$$

Hence, $W - \xi\frac{\sigma_W}{\sigma_Z}Z \overset{\text{a.s.}}{=} \alpha$ for some constant $\alpha \in \mathbb{C}$. The proof is exactly analogous if one lets $\xi = \text{Ps}\mathbb{C}\text{orr}\,(W,Z)$ and replaces Z by Z^*. The reverse direction is already given by (7.1.4) and (7.1.7) Letting now $\xi = \mathbb{C}\text{orr}\,(W,Z)$ and $\kappa = \text{Ps}\mathbb{C}\text{orr}\,(W,Z)$ be such that $|\xi| = 1 = |\kappa|$, we have by $|\xi| = 1$, that $W = \alpha + \beta Z$ for constants $\alpha, \beta \in \mathbb{C}$ (note β is as given above) and $\xi = \frac{\beta}{|\beta|}$. Then the pseudocorrelation is given by

$$\begin{aligned}
\kappa &= \text{Ps}\mathbb{C}\text{orr}\,(\alpha + \beta Z, Z) \\
&= \frac{\beta\rho_Z^2}{|\beta|\sigma_Z^2}.
\end{aligned}$$

Thus $|\kappa| = 1$ implies that $|\rho_Z^2| = \sigma^2$, which by Proposition 4.5.4 means that Z must be distributed along a line in the complex plane. Consequently, so is W being (almost-surely) an affine transform of Z. □

In particular, the proof emphasises how real random variables ($\rho^2 = \sigma^2$) under real affine linear transforms ($\alpha, \beta \in \mathbb{R}$) form a special case of complex random variables where both correlations coincide.

When neither the correlation nor pseudocorrelation have modulus 1, W is neither an affine nor conjugate affine transformation of Z. However, we may still have that W is an $\mathbb{R}$-linear transform of Z, i.e.

$$W = \alpha + \beta Z + \gamma Z^*. \tag{7.1.11}$$

In this case, Schreier and Scharf (2010, p. 90) define the "total correlation coefficient" , based off the least-squares estimator for a widely-linear model. We note that this correlation coefficient is defined to be a real quantity, bounded in $[0,1]$. In this regard, its squared value may indicate the strength of an $\mathbb{R}$-linear relationship

between Z and W, however does not provide a direct geometric interpretation as in the simpler cases. For $\xi = \mathbb{C}\text{orr}(W, Z)$ and $\kappa = \text{Ps}\mathbb{C}\text{orr}(W, Z)$, the total correlation coefficient is defined as

$$\text{Tot}\mathbb{C}\text{orr}(W, Z) = \sqrt{\frac{\sigma_Z^4(|\xi|^2 + |\kappa|^2) - 2\sigma_Z^2 \Re(\xi\kappa^* \rho_Z^2)}{\sigma_Z^4 - |\rho_Z|^4}}. \tag{7.1.12}$$

If Z is a proper random variable ($\rho_Z^2 = 0$) then this simplifies to $\sqrt{|\xi|^2 + |\kappa|^2}$. In particular Schreier and Scharf (2010) note that

$$\max\left\{|\xi|^2, |\kappa|^2\right\} \leq \text{Tot}\mathbb{C}\text{orr}(W, Z)^2 \leq \min\left\{|\xi|^2 + |\kappa|^2, 1\right\}. \tag{7.1.13}$$

This correlation is inspired by the least-squares estimator of a widely-linear regression model such that if (7.1.11) holds and $\text{Tot}\mathbb{C}\text{orr}(W, Z) = 1$, then the (absolute) mean squared error of prediction of W from Z is zero (Picinbono and Chevalier (1995); Schreier and Scharf (2010); Adali and Schreier (2014)).

However, we are unaware whether the total correlation is 1 if and only if W is an $\mathbb{R}$-linear transform of Z. It is also worth noting, as is done in Schreier and Scharf (2010), that in general $\text{Tot}\mathbb{C}\text{orr}(W, Z) \neq \text{Tot}\mathbb{C}\text{orr}(Z, W)$. This is because trying to reconstruct Z given W where (7.1.11) holds is not as straightforward an inversion as with the affine and conjugate affine transformations, and may not be possible if $|\alpha| = |\beta|$. The inversion transformation where $|\alpha| \neq |\beta|$ was given in (4.9.1).

7.1.5 Correlation and normality

As with the real case, by nature of a normal distribution being uniquely determined by its second-order properties, the correlation may be used to uniquely determine the dependence structure in the a bivariate complex normal setting. Below we give the example for the case of circular complex normal distributions without pseudocorrelation for notational compactness.

Example 7.1.3. *Let Z and W follow a jointly complex normal distribution. If the marginal distributions are such that $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$ and $W \sim \mathcal{N}_{\mathbb{C}}(\eta, \tau^2, 0)$, and the correlation $\mathbb{C}\text{orr}(Z, W) = \xi$ whilst $\text{Ps}\mathbb{C}\text{orr}(Z, W) = 0$, then we have that the joint distribution is given by*

$$f_{Z,W}(z, w) = \tag{7.1.14}$$

$$\frac{1}{\pi\sigma^2\tau^2(1 - |\xi|^2)} \exp\left\{-\frac{1}{1 - |\xi|^2}\left(\frac{|z - \mu|^2}{\sigma^2} - 2\Re\left(\xi^* \frac{z - \mu}{\sigma}\left(\frac{w - \eta}{\tau}\right)^*\right) + \frac{|w - \eta|^2}{\tau^2}\right)\right\}. \tag{7.1.15}$$

Indeed, if the correlation (and pseudocorrelation) is zero, we have that Z and W are independent as their joint density is the product of their marginal densities. Thus we again observe that uncorrelated[1] *and independent coincide in the normal case, which leads Schreier and Scharf (2010) to label their test for correlation a test for independence.*

[1] Here uncorrelated means both the correlation and pseudocorrelation are 0.

Again, we tend to have the property that under the assumption of normality, least-squares estimation and correlation coefficients tend to coincide and perform optimally. Of course, to use the correlation as a meaningful measure of *dependence*, the assumption of normality should be verified. Goodness-of-fit tests such as that of Ducharme et al. (2016) may be used to verify any assumptions of normality in complex statistical modelling and inference.

7.2 Dependence Measures

We have seen that the definition of the (Pearson) correlation may be extended to complex random variables, with corresponding estimators making use of the complex multiplication. However, again, the correlation only measures dependence in a multivariate complex normal setting. So more generally, the correlation may be zero despite the random variables being dependent. As such, to better understand the dependence structure between complex random variables, we would look to generalise other real dependence measures to a complex setting.

Many measures of dependence have been proposed for real data, from the Pearson's correlation, Spearman's ρ and Kendall's τ which are taught in undergraduate statistics courses, to distance correlation (Székely et al. (2007)), copula-based measures and the Hellinger correlation (Geenens and Lafaye de Micheaux (2020)). For a review of recent developments in dependence measures, Tjøstheim et al. (2018) provide a succinct overview and reference list. This section will not directly generalise any specific real dependence measures for complex random variables, rather discuss why the generalisation of common real dependence measures may not be so straightforward for complex random variables. We also present complex definitions of the odds ratio and local dependence functions which appear as easily generalised objects, although the statistical application of these objects is currently sparse even in literature on real random variables.

Already, it was observed that in generalising the correlation to a complex context, we encounter two possible correlations due to the occurrence of two second-order moments that coincide in the real case. Intuitively, for even higher moments of a distribution, more and more complex variants arise, all of which coincide in the real case where conjugation is equivalent to the identity map.

To look at the problem another way, many popular dependence measures in $\mathbb{R}$ rely on a sense of ordering or ranks. For example, simple robustification of Pearson's correlation into Spearman's ρ or Kendall's τ require the observations to be ranked. Similarly, other copula-based methods take advantage of the probability integral transform, i.e. that for a continuous real random variable X with cumulative distribution function F_X, $F_X(X) \sim \mathrm{U}_{\mathbb{R}}[0, 1]$, to ensure margin-freeness of related dependence measures (Grothe et al. (2014); Geenens and Lafaye de Micheaux (2020)). However for complex random variables, we do not have such a canonical transform which preserves the ranks of the complex plane, given there is no complete ordering of the complex numbers. Similar issues arise in multivariate real statistics, where some dependence measures are still generalised, albeit using the joint cumulative distribution function to provide an entry-wise partial ordering of higher dimensional Euclidean space (Grothe et al. (2014)). Similar applications may be generalised to complex random variables, by employing the probability transform with the rectangular c.d.f.. Of course the question then arises, for example in the case of a circular

distribution, why not use the ball-wise c.d.f., which would perform the probability integral transform on just the modulus, given the argument is already independent and uniformly distributed. Thus we end up with many possible variants of a dependence measure, arising from the many possible partial orderings of the complex numbers.

Stepping away from copula- and rank-based methods, other popular forms of dependence measures include the distance correlation such as that proposed by Székely et al. (2007), and generalised to general metric space by Lyons (2013), with a metric distribution function set up by Wang et al. (2021). As $\mathbb{C}$ forms a metric space, the methods of Wang et al. (2021) could in theory be applied to complex random variables, using what they define as the metric distribution function. However, whilst such a dependence measure has the nice property that it equals zero if and only if the measured variables are independent, such measures are not margin free and therefore cannot satisfy what Geenens and Lafaye de Micheaux (2020) call the "generalised Data Processing Inequality". In this sense, the marginal distributions may be diluted with independent white noise and it is possible that the dependence measure will increase its measure of dependence. So such distance- or metric-based measures of dependence may not necessarily provide accurate measurements of the level of dependence where there is a dependence structure. Beran et al. (2007) develop a nonparametric test for independence between real random vectors and acknowledge that in a multivariate context any test for independence cannot be margin-free, and as such will not satisfy the generalised Data Processing Inequality. Thinking of complex random variables as real bivariate random variables would suggest that the dependence between complex random variables is like describing the dependence between random vectors in $\mathbb{R}^2$.

Nonetheless, dependence measures for complex random variables remains a relatively unexplored area of research and whilst displaying some similarities to bivariate statistics, it may be that the algebra in $\mathbb{C}$ permits a more obvious generalisation of real dependence measures, even where two or more statistics need to be used simultaneously, such as with the correlation and pseudocorrelation. We turn now to consider instead functional descriptions of dependence. In particular, the local dependence function which, like copulas, is a function (as opposed to a single measure) that describes the dependence structure between two random variables, however does not rely on the probability integral transform which restricts the usefulness of copulas chiefly to the univariate real setting.

7.2.1 Local dependence function

Introduced in Holland and Wang (1987), the local dependence function can be used to describe the dependence structure of a random variable, similar to how a copula describes the dependence structure, independent of the marginals. The local dependence function is based off what they call the cross-product ratio function (ω) between two random variables Z, W with respect to the point (z_0, w_0) such that

$$\omega_{z_0,w_0}(z,w) = \frac{f_{Z,W}(z,w) f_{Z,W}(z_0,w_0)}{f_{Z,W}(z_0,w) f_{Z,W}(z,w_0)}, \tag{7.2.1}$$

where $f_{Z,W}$ is a density of the joint distribution of Z and W. This is also called the "odds ratio function" in Osius (2004), due to its resemblance of generalising the odds ratio that is normally given in a 2×2 contingency table and importantly it is noted that the densities need not be with respect to the Lebesgue measure to define the odds ratio function. In terms of the dependence structure, we may rewrite ω in terms of conditional distributions of $f_{Z|W}$ or $f_{W|Z}$ and recover the same function. In this sense, ω is invariant to margin replacement — but not margin free like a copula which describes how to combine two arbitrary marginal distributions. Furthermore, if Z and W are independent, we have that $\omega_{z_0,w_0}(z,w) = 1$ for all $z, w \in \mathbb{C}$. Note that here, whilst all research on the odds ratio function ω has been conducted from a perspective of real random variables, there is no restriction requiring the random variables Z, W be real. Indeed, as we still have a notion of densities for complex random variables, and understand dependence in the same way with respect to the underlying real probability space, the generalisation of the odds ratio function holds directly as given here.

To further characterise the dependence between two random variables, Holland and Wang (1987) introduce what is called the local dependence function, which removes the reliance of the odds ratio function upon the reference point (z_0, w_0). Specifically, they let the reference point be an infinitesimal increment ($z + \mathrm{d}z$ and $w + \mathrm{d}w$) and thus define the local dependence function as the derivative in each variable of the logarithm of the joint density. Of course, in the complex case, some care must be taken as we have two potential derivatives under Wirtinger calculus. We define the local dependence function for complex random variables as follows.

Definition 7.2.1. *The local dependence function between two complex random variables Z and W, with joint distribution $f_{Z,W}$ is defined as*

$$\Upsilon_{Z,W}(z,w) \coloneqq \frac{\partial^2}{\partial z \partial w^*} \ln(f_{Z,W}(z,w)). \tag{7.2.2}$$

Indeed, the local dependence function $\Upsilon_{Z,W}$ is also invariant to margin replacement where replacing the joint density with either conditional density $f_{Z|W}$ or $f_{W|Z}$ still yields the same local dependence function (Holland and Wang (1987); Wang (1993, Definition 2.2)). It is noted that parameters of the marginal distributions may still appear in the local dependence function of two random variables. The local dependence function also holds the nice property that $\Upsilon_{Z,W} \equiv 0$ if and only if Z and W are independent (provided the support of $\Upsilon_{Z,W}$ can be written as a Cartesian product set) (Holland and Wang (1987)).

The odds ratio and local dependence functions uniquely characterise the dependence structure between two random variables in the sense that if the marginal distributions f_Z and f_W are known, as well either $\omega_{z_0,w_0}(z,w)$ or $\Upsilon_{Z,W}(z,w)$, then there exists a unique bivariate density which has these marginals and ω_{z_0,w_0} or $\Upsilon_{Z,W}$ (Holland and Wang (1987); Wang (1993); Osius (2004); Osius (2009)). That is, knowing the marginals and either the odds ratio or local dependence function gives all possible information about the distribution. Furthermore, Osius (2004) prove existence for the odds ratio function, that is given two marginals and ω_{z_0,w_0}, there exists a corresponding bivariate density.

For example, a real bivariate normal distribution has a constant local dependence function. Hence, if it is known that the marginals are normally distributed and the local dependence function is a constant, then it must be that the bivariate distribution is a real bivariate normal with the correlation uniquely determined by the marginal variances and constant value of the local dependence function. Under Definition 7.2.1, the local dependence function of a bivariate complex normal distribution is also constant, and hence a similar result holds to construct a complex bivariate normal distribution. We present below the form in the case where Z and W are circular and correlated with zero pseudocorrelation, simply due to notational compactness, however it has also been verified in the case of arbitrary complex normal marginals and pseudocorrelation that the local dependence function is still constant.

Example 7.2.2. *Let Z and W follow a bivariate complex normal distribution where $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$ and $W \sim \mathcal{N}_{\mathbb{C}}(\eta, \tau^2, 0)$ such that $\mathbb{C}\mathrm{orr}(Z, W) = \xi$ and $\mathrm{Ps}\mathbb{C}\mathrm{orr}(Z, W) = 0$. Then the local dependence function is given by*

$$\Upsilon_{Z,W}(z, w) = \frac{\xi^*}{\sigma\tau(1 - |\xi|^2)}, \tag{7.2.3}$$

which is constant. Furthermore, given σ, τ and $\Upsilon_{Z,W}$, one can solve for the correlation ξ by solving a quadratic in the modulus and taking the argument to be the negative of that of $\Upsilon_{Z,W}$.

Overall, the odds ratio and local dependence functions can describe the dependence structure between complex random variables in a manner similar to how these same quantities and copulas are used for real random variables. But further research is required in terms of statistical application such as estimation from a sample, which should not rely on a canonical ordering of the complex numbers.

7.3 A Brief Look at Linear Modelling

We finish this penultimate chapter with a consideration of basic linear modelling techniques in $\mathbb{C}$, in particular from a least-squares perspective assuming circular normal errors. Complex linear modelling is used in application, for example Rowe and Logan (2004) use a least squares estimator to model the random noise present in fMRI scans of a human brain. In signal processing, least-squares estimation is performed in a widely-linear sense by assuming a received signal (represented as a complex number) is an $\mathbb{R}$-linear transform of the transmitted signal plus some normal random error (Picinbono and Chevalier (1995); Schreier and Scharf (2010); Trampitsch (2013); Adali and Schreier (2014)). Here we do not present such depth of linear modelling with complex variables and estimation as is in the aforementioned references, rather provide basic results of complex linear modelling. In particular, to our knowledge there is relatively little, if any, discussion in the literature around the distribution of estimators and ensuing statistical inference. Particularly, in the case of an ordinary least-squares coefficient estimator, we see that where errors follow a circular complex normal distribution, we obtain the familiar result that the standardised estimate follows a complex t-distribution as it was defined in Subsection 5.5.1.

7.3.1 Fourier transformed data and distribution

We first discuss the discrete Fourier transform, as this is a common method by which complex-valued data arise in practice. For example, magnetotelluric methods in geophysics involve deploying magnetometers and electrodes to measure the magnetic and electric field strengths in three orthogonal directions over time. The resulting time series are then Fourier transformed to yield a complex data set that is then analysed to understand the behaviour of the underlying electromagnetic waves travelling through the Earth, and their interaction with the composition of the Earth's crust (Cagniard (1953); Varentsov et al. (2003); personal correspondence with Dr Kate Selway and Dr Constanza Manassero).

The discrete Fourier transform of a real signal $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is given by the complex observations $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\top$ where

$$Z_k = \sum_{s=1}^{n} \exp\left\{ i \frac{2\pi(s-1)(k-1)}{n} \right\} X_s \,, \quad k = 2, 3, \ldots, n.$$

In a computational setting, this is often performed in $\mathcal{O}(n \ln(n))$ time using the fast Fourier transform (e.g. `fft` in `R`). We now examine the distribution of Fourier transformed data where the original signal observations are i.i.d. normal and more generally independent with n large both of which justify the use of a normal distribution for the data or additive errors. We note however that in practice, these signals may not be independent, which will invalidate any distributional assumptions based upon variants of the central limit theorem. In a dependent setting, methods of time series analysis such as autoregressive and moving average models should also be incorporated, however are not explored here.

Fourier transform of i.i.d. normal data

If the original real data $X_1, \ldots, X_n$ are assumed to be i.i.d. random variables from a $\mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ distribution, we have that $Z_2, \ldots, Z_n$ each follow a $\mathcal{N}_{\mathbb{C}}(0, n\sigma^2, 0)$ distribution. Note that Z_1 is excluded as it is just the sum of the real observations and hence follows a $\mathcal{N}_{\mathbb{R}}(n\mu, n\sigma^2)$ distribution. The circular normality of the Fourier transformed data may be seen from the fact that if

$$\omega_n = e^{i\frac{2\pi}{n}}, \tag{7.3.1}$$

then

$$\sum_{s=1}^{n} \omega_n^{(s-1)(k-1)} = 0, \tag{7.3.2}$$

$$\sum_{s=1}^{n} \left| \omega_n^{(s-1)(k-1)} \right|^2 = n, \tag{7.3.3}$$

$$\sum_{s=1}^{n} \left(\omega_n^{(s-1)(k-1)} \right)^2 = 0, \tag{7.3.4}$$

for all integers $k > 1$ and

$$Z_k = \sum_{s=1}^{n} \omega_n^{(k-1)(s-1)} X_s. \tag{7.3.5}$$

By equating $\mathcal{N}_{\mathbb{R}}(\mu, \sigma^2) = \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, \sigma^2)$, linearity of the complex normal and independence of the X_k's, we have that

$$Z_k \sim \mathcal{N}_{\mathbb{C}}\left(\mu \sum_{s=1}^{n} \omega_n^{(s-1)(k-1)}, \sigma^2 \sum_{s=1}^{n} |\omega_n^{(s-1)(k-1)}|^2, \sum_{s=1}^{n} (\omega_n^{(s-1)(k-1)})^2\right) \equiv \mathcal{N}_{\mathbb{C}}\left(0, n\sigma^2, 0\right). \tag{7.3.6}$$

Subsequently, one can take the scaled Fourier transform (such as in Yan et al. (2019)) by $Z_k/\sqrt{n}$ so that each Z_k follows a $\mathcal{N}_{\mathbb{C}}(0, \sigma^2, 0)$ distribution. It can even be seen that the Z_k's here are independent by the fact that

$$\mathbb{C}\text{ov}(Z_k, Z_l) = \sigma^2 \sum_{s=1}^{n} \omega_n^{(s-1)(k-l)} = \sigma^2 \delta_{kl}, \tag{7.3.7}$$

where δ_{kl} is the Kronecker delta that equals 1 if $k = l$ and 0 otherwise.

Fourier transform of independent data

Where the data are now assumed to be independent, but non-normal, we are unable to invoke linearity to determine the exact distribution of each of the Z_k's. However, we may construct a triangular array out of

$$X_1, \tag{7.3.8}$$
$$X_1,\ \omega_2^{k-1} X_2, \tag{7.3.9}$$
$$X_1,\ \omega_3^{k-1} X_2,\ \omega_3^{2(k-1)} X_3, \tag{7.3.10}$$
$$\ldots, \tag{7.3.11}$$
$$X_1,\ \omega_n^{k-1} X_2,\ \omega_n^{2(k-1)} X_3,\ \ldots,\ \omega_n^{(n-1)(k-1)} X_n, \tag{7.3.12}$$
$$\ldots \tag{7.3.13}$$

such that (assuming it holds) the Lindeberg condition can be used to prove that for large n, $\frac{1}{\sqrt{n}} Z_k$ again asymptotically converge in distribution to a $\mathcal{N}_{\mathbb{C}}(0, \sigma^2, 0)$ distribution. This also justifies assumptions of normality when modelling complex data as the Fourier transform of a long time series, however it is again noted that the independence of the original real signal is a crucial assumption, which may not hold in practice.

7.3.2 A linear model

We now briefly touch on linear modelling, and in particular the ordinary-least squares (OLS) case with circular normal errors. The main result in this section is the explicit proof of the distribution of the OLS coefficients as following a complex t-distribution under the assumption of circular normal errors, which to the best

of our knowledge is novel. Further developments in terms of widely-linear models incorporating explanatory variables and their conjugate and related inference are considered future research, as whilst much has been done in signal processing literature to develop the least-squares estimators for such models (Picinbono and Chevalier (1995); Schreier and Scharf (2010); Trampitsch (2013); Yan et al. (2019)), there is scarcely any literature regarding statistical inference around the estimated coefficients, which is well-established in the real case.

A complex linear regression model is set up such that we have a vector $\boldsymbol{w} \in \mathbb{C}^n$ of observed responses of a complex random variable W, which we assume is a linear function of explanatory variables $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{p-1} \in \mathbb{C}^n$. Let $\mathbf{Z}$ be the $n \times p$ complex design matrix whose first column is a vector of ones, and remaining columns are the $\boldsymbol{Z}_k$'s. Then the linear regression model is such that

$$\boldsymbol{w} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{7.3.14}$$

where here we will assume that given $\mathbf{Z}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \sigma^2\mathbf{I}_n, \mathbf{0})$ i.e. that the errors are i.i.d. circular normal random variables with variance σ^2. The OLS estimator of $\boldsymbol{\beta}$, i.e. the estimator $\boldsymbol{b}$ that minimises $\left\|\boldsymbol{w} - \mathbf{Z}\tilde{\boldsymbol{b}}\right\|^2$ is given by

$$\boldsymbol{b} = (\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathcal{H}}\boldsymbol{w}. \tag{7.3.15}$$

The above formula, which is reminiscent of the real OLS estimator, but replacing transposes with Hermitians can be derived analogous to the real case, using Wirtinger calculus to minimise the sum of (modulus) squared errors (Appendix A.18). Indeed, if $p = 2$, we have that $b_2 = \widehat{\xi}\frac{\widehat{\sigma}_W}{\widehat{\sigma}_Z}$ where $\widehat{\xi}$ is the estimator of $\mathbb{C}\mathrm{orr}(W, Z_1)$ given in (7.1.5). Equation (7.3.15) is also the maximum likelihood estimate of $\boldsymbol{\beta}$ under the assumption of circular normal errors. Additionally, if $\mathbb{V}\mathrm{ar}(\boldsymbol{\varepsilon}) = \boldsymbol{\Gamma}$ is not necessarily a scaled identity matrix, but still invertible, the estimator can be adjusted by taking (Trampitsch (2013, p. 72))

$$\boldsymbol{b} = (\mathbf{Z}^{\mathcal{H}}\boldsymbol{\Gamma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathcal{H}}\boldsymbol{\Gamma}^{-1}\boldsymbol{w}. \tag{7.3.16}$$

Keeping with the i.i.d. case so that $\boldsymbol{\Gamma} = \sigma^2\mathbf{I}_n$, often in practice σ^2 is also unknown and needs to be estimated. An unbiased estimator (from rescaling the maximum likelihood estimator to remove bias) of σ^2 is given by

$$\widehat{\sigma}^2 = \frac{1}{n-p}\|\boldsymbol{\varepsilon}\|^2 \tag{7.3.17}$$

$$= \frac{1}{n-p}\|\boldsymbol{w} - \mathbf{Z}\boldsymbol{b}\|^2, \tag{7.3.18}$$

$$= \frac{1}{n-p}(\boldsymbol{w} - \mathbf{Z}\boldsymbol{b})^{\mathcal{H}}(\boldsymbol{w} - \mathbf{Z}\boldsymbol{b}). \tag{7.3.19}$$

To present the key result in this section about the distribution of each of the estimated coefficients b_k, we first set up two lemmas about the distributions of $\boldsymbol{b}$ and $\widehat{\sigma}^2$.

Lemma 7.3.1. *Given* $\mathbf{Z}$*, the OLS estimator* $\boldsymbol{b}$ *follows a* $\mathcal{N}_{\mathbb{C}^p}\left(\boldsymbol{\beta},\ \sigma^2\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1},\ \mathbf{0}\right)$ *distribution.*

Proof. The estimator is given by $\boldsymbol{b} = \left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathcal{H}}\boldsymbol{w}$, where given $\mathbf{Z}$ and assuming $\boldsymbol{\beta}$ is constant,

$$\boldsymbol{w} \sim \mathcal{N}_{\mathbb{C}^n}\left(\mathbf{Z}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n, \mathbf{0}\right). \tag{7.3.20}$$

Hence by linearity of the complex normal, we have that

$$\begin{aligned}\boldsymbol{b} &\sim \mathcal{N}_{\mathbb{C}^p}\left(\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\boldsymbol{\beta},\ \sigma^2\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1},\ \mathbf{0}\right)\\ &\sim \mathcal{N}_{\mathbb{C}^p}\left(\boldsymbol{\beta},\ \sigma^2\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1},\ \mathbf{0}\right).\end{aligned}$$

□

Lemma 7.3.2. *Given* $\mathbf{Z}$*, the distribution of the variance estimator* $\hat{\sigma}^2$ *satisfies*

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \text{Gamma}(n-p, 1), \tag{7.3.21}$$

independent of the estimator $\boldsymbol{b}$*.*

The proof is provided in Appendix A.19. We can now state and prove the familiar result for the distribution of the OLS estimator under i.i.d. circular complex normal errors.

Theorem 7.3.3. *Consider a linear model* $\boldsymbol{w} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *such that given* $\mathbf{Z}$*,* $\boldsymbol{\varepsilon} \sim \mathcal{N}_{\mathbb{C}^n}(\mathbf{0}, \sigma^2\mathbf{I}_n, \mathbf{0})$*. Then the OLS estimate* $\boldsymbol{b} = (\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathcal{H}}\boldsymbol{w}$ *of* $\boldsymbol{\beta}$ *satisfies*

$$\frac{b_k - \beta_k}{\hat{\sigma}\sqrt{\left[(\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\right]_{kk}}} \sim \mathbb{C}t_{(n-p)}, \tag{7.3.22}$$

for the variance estimator $\widehat{\sigma} = \frac{\|\boldsymbol{w}-\mathbf{Z}\boldsymbol{b}\|}{\sqrt{n-p}}$*.*

Proof. First, following from Lemma 7.3.1, we have that

$$\frac{b_j - \beta_j}{\sigma\sqrt{\left[(\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\right]_{jj}}} \sim \mathcal{N}_{\mathbb{C}}(0, 1, 0).$$

Hence, combining with Lemma 7.3.2 we obtain

$$\begin{aligned}\frac{b_k - \beta_k}{\hat{\sigma}\sqrt{\left[(\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\right]_{kk}}} &= \frac{b_k - \beta_k}{\sigma\sqrt{\left[(\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\right]_{kk}}} \Bigg/ \sqrt{\frac{(n-p)\hat{\sigma}^2}{\sigma^2}/(n-p)}\\ &= \frac{Z_0}{\sqrt{\frac{G}{(n-p)}}},\end{aligned}$$

where $Z_0 \sim \mathcal{N}_{\mathbb{C}}(0,1,0)$ is independent of $G \sim \text{Gamma}(\text{n}-\text{p},1)$. This was given as an alternate definition of the circular complex $\mathbb{C}t_{n-p}$ distribution by Proposition 5.5.3 in Chapter 5, which completes the proof. □

By consequence of Theorem 7.3.3, similar inference on linear modelling coefficients may be conducted, for example testing the significance of coefficients being non-zero. We have here that the standard complex t-distribution is circular, so we can construct confidence balls to test significance of individual coefficients. In particular, from the proof of Proposition 5.5.2, if $Z \sim \mathbb{C}t_\nu$, then the distribution of $|Z|^2$ is defined via the following density with respect to the Lebesgue measure,

$$f_{|Z|^2}(x) = \left(1+\frac{x}{\nu}\right)^{-(\nu+1)}, \; x \in \mathbb{R}^+, \tag{7.3.23}$$

and cumulative distribution function

$$F_{|Z|^2}(x) = 1 - \left(1+\frac{x}{\nu}\right)^{-\nu}, \; x \; \in \mathbb{R}^+. \tag{7.3.24}$$

Hence, if we let $K_{1-\alpha} = \sqrt{(n-p)\left(\frac{1}{\alpha^{n-p}}-1\right)}$, then the disc

$$CB_\alpha = \left\{ z \in \mathbb{C} : |z-b_k| \leq K_{1-\alpha}\widehat{\sigma}\sqrt{\left[(\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\right]_{kk}} \right\} \tag{7.3.25}$$

forms a $100(1-\alpha)\%$ confidence ball for β_k.

Furthermore, we can test the significance of coefficients by running the hypothesis test $H_0 : \beta_k = 0$ against $H_1 : \beta_k \neq 0$, such that we reject H_0 at a $100\alpha\%$ significance level if

$$\frac{b_k}{\widehat{\sigma}\sqrt{\left[(\mathbf{Z}^{\mathcal{H}}\mathbf{Z})^{-1}\right]_{kk}}} \notin \{ z \in \mathbb{C} : |z| \leq K_{1-\alpha} \}, \tag{7.3.26}$$

or equivalently if $0 \notin CB_\alpha$.

Overall, where assumptions of normality hold, statistical results for linear modelling between random variables may still be generalised to the complex case without relying upon bivariate real methods. However the ability to efficiently develop a theory of statistics for complex data that does not depend upon translated notation from $\mathbb{R}^2$ requires the foundations of probability with complex random variables to be established. Indeed, if one unpacks the linear regression model in terms of the Cartesian components of the complex variables and parameters, some form of 'linked' or 'vector' regression is observed in the real and imaginary components such that certain coefficients in the real and imaginary regressions must be equal. By using complex variables and parameters under the algebra of $\mathbb{C}$, familiar results and inference may be performed from real statistics, as well as more efficient computation compared to conducting two separate real regressions with constraints

on coefficients. Interpretability is also aided under the complex multiplication's geometrical interpretation.

Further research is required into the complex normal distribution, and in particular the modulus-squared distribution such as the general form presented in Cooper and Farid (2020) in order to be able to give the distribution of the variance estimator, and subsequently the standardised coefficients in a setting of non-circular normal errors. Considering generalised linear or nonlinear modelling techniques also offer avenues of future research that are currently unexplored in the literature to the best of our knowledge.

Chapter 8

Conclusion

In summary, complex data exist in application and require statistical tools to be analysed. Current literature approaches statistical problems with complex data by referring to equivalent real bivariate methods. However, this obfuscates the complex multiplication from statistical procedures — the very multiplication that motivates the representation of data as complex numbers in the first place. Instead, approaching complex random variables from within $\mathbb{C}$ facilitates the construction of a self-contained theory of stochastics in the complex plane, different to that of real bivariate random variables. The algebra of $\mathbb{C}$ elicits useful tools, such as the Mellin transform, without bivariate analogues. And although we relax the differential structure to mimic $\mathbb{R}^2$, working within $\mathbb{C}$ is still essential to take full advantage of integral formulae.

In this thesis, we have repositioned existing theory and extended it with new results in order to now embrace all aspects of $\mathbb{C}$. However, there is still a long way to go in terms of establishing a comprehensive theory of complex probability and statistics that rivals real probability and statistics. We do not anticipate the generalisation of all of real probability and statistics to hold analogously, for example copula- and rank-based dependence measures are not as straightforward by issue of no complete order of the complex numbers. Nonetheless, we have illuminated many areas that warrant future research to take full advantage of the complex field structure and unlock methods unique to complex random variables.

In particular, looking further into the complex cumulative distribution function, contour random variables and the incorporation in statistics of complex probabilities are largely unexplored gaps in the literature with potential to yield novel methods.

References

Aaronson, S. (2013). *Quantum Computing since Democritus*. Cambridge University Press, Cambridge.

Ablowitz, M. J. and Fokas, A. S. (2003). *Complex Variables: Introduction and Applications*. Cambridge University Press, Cambridge, 2nd edition.

Adali, T. and Schreier, P. (2014). Optimization and Estimation of Complex-Valued Signals: Theory and applications in filtering and blind source separation. *IEEE Signal Processing Magazine*, 31(5):112–128.

Adali, T., Schreier, P., and Scharf, L. (2011). Complex-valued Signal Processing: The Proper Way to Deal with Impropriety. *IEEE Transactions on Signal Processing*, 59(11):4516–4528.

Adrian, D., Maitra, R., and Rowe, D. (2018). Complex-valued Time Series Modeling for Improved Activation Detection in fMRI Studies. *The Annals of Applied Statistics*, 12(3):1451–1478.

Ahlfors, L. V. (1979). *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable*. McGraw-Hill Inc., New York, 3rd edition.

Amry, Z. (2020). Bayesian Estimate of Parameters for ARMA Model Forecasting. *Tatra Mountains Mathematical Publications*, 75(1):23–32.

Andersen, H., Højbjerre, M., Sørensen, D., and Eriksen, P. (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution*. Springer-Verlag New York, Inc., New York.

Bandettini, P., Jesmanowicz, A., Wong, E., and Hyde, J. (1993). Processing strategies for time-course data sets in functional mri of the human brain. *Magnetic Resonance in Medicine*, 30(2):161–173.

Barnett, V. (1976). The Ordering of Multivariate Data. *Journal of the Royal Statistical Society. Series A (General)*, 139(3):318–355.

Bartlett, M. S. (1945). Negative probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, 41(1):71–73.

Beckmann, P. (1964). Rayleigh Distribution and Its Generalizations. *Research of the National Bureau of Standards. Section D: Radio Science*, 68D(9):927–932.

Beran, R., Bilodeau, M., and Lafaye de Micheaux, P. (2007). Nonparametric tests of independence between random vectors. *Journal of Multivariate Analysis*, 98(9):1805–1824.

Bertrand, J., Bertrand, P., and Ovarlez, J.-P. (2010). The Mellin Transform. In Poularikas, A. D., editor, *The Transforms and Applications Handbook*, The Electrical Engineering Handbook Series, chapter 12, pages 12.1–12.38. CRC Press Inc., CRC Press,Taylor & Francis Group,6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742, 3rd edition.

Bidabad, B., Bidabad, B., and Mastorakis, N. (1992). Complex Probability and Markov Stochastic Process. In *Proceeding of the 1st Iranian Statistics Conference*, pages 198–201, Isfahan University of Technology.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, Inc., New York, NY, 3rd edition.

Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. University of California Press, Berkeley and Los Angeles.

Bordley, R. F. (2005). Econophysics and individual choice. *Physica A: Statistical Mechanics and its Applications*, 354:479–495.

Bracewell, R. (1965). *The Fourier Transform and Its Applications*. Electrical and Electronic Engineering Series. McGraw-Hill Inc., United States of America.

Brock, A. and Krutchkoff, R. (1970). Distributions of Z^γ and Z^* with Results Applied to the Complex Normal Distribution. *The Annals of Mathematical Statistics*, 41(5):1563–1567.

Brown, J. W. and Churchill, R. V. (2009). *Complex Variables and Applications*. McGraw-Hill, New York, NY, 8th edition.

Burt, W., Cummings, T., and Paulson, C. (1979). The Mesoscale Wind Field Over the Ocean. *Journal of Geophysical Research*, 79(36):5625–5632.

Cagniard, L. (1953). Basic Theory of the Magneto-telluric Method of Geophysical Prospecting. *Geophysics*, 18(3):605–635.

Caldwell, G., Bibby, H., and Brown, C. (2004). The magnetotelluric phase tensor. *Geophysical Journal International*, 158(2):457–469.

Can, V. H., Duong, M. H., and Pham, V. V. H. (2019). Persistence probability of a random polynomial arising from evolutionary game theory. *Journal of Applied Probability*, 56(3):870–890.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Wadsworth Group, Pacific Grove, CA.

Chakraborty, R. and Yu, J. W. S. X. (2019). Sur-Real: Fréchet Mean and Distance Transform for Complex-Valued Deep Learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 889–897.

Chebotarev, A. M. and Teretenkov, A. E. (2014). Singular value decomposition for the Takagi factorization of symmetric matrices. *Applied Mathematics and Computation*, 234:380–384.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with ARMA (p, q) errors. *Journal of Econometrics*, 64(1):183–206.

Comon, P. (1994). Circularité et signaux aléatoires à temps discret. *Traitement du Signal*, 11(5):417–420.

Cooper, E. and Farid, H. (2020). A Toolbox for the Radial and Angular Marginalization of Bivariate Normal Distributions. *arXiv preprint*, pages 1–18.

Copson, E. T. (1935). *An introduction to the theory of functions of a complex variable*. Oxford : Clarendon Press, 1st edition.

Cottone, G. and Di Paola, M. (2009). On the use of fractional calculus for the probabilistic characterization of random variables. *Probabilistic Engineering Mechanics*, 24(3):321–330.

Csörgo, S. (1981). Limit Behaviour of the Empirical Characteristic Function. *The Annals of Probability*, 9(1):130 – 144.

Davis, P. J. (1959). Leonhard Euler's Integral: A Historical Profile of the Gamma Function: In Memoriam: Milton Abramowitz. *The American Mathematical Monthly*, 66(10):849–869.

Descartes, R. (1637). *La Géométrie*. A. Hermann, Librairie Scientifique, Paris. Edited by A. Hermann, republished in 1887 and this version again in 2008.

Ducharme, G., Lafaye de Micheaux, P., and Marchina, B. (2016). The complex multinormal distribution, quadratic forms in complex random vectors and an omnibus goodness-of-fit test for the complex normal distribution. *Annals of the Institute of Statistical Mathematics*, 68:77–104.

Edelman, A. and Wang, Y. (2013). Random Matrix Theory and Its Innovative Applications. In Melnik, R. and I, I. K., editors, *Advances in Applied Mathematics, Modeling, and Computational Science*, volume 66 of *Fields Institute Communications*. Springer, Boston, MA.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26.

Eklund, A. C. (2020). *squash: Color-Based Plots for Multivariate Visualization*. R package version 1.0.9.

Epstein, B. (1948). Some Applications of the Mellin Transform in Statistics. *The Annals of Mathematical Statistics*, 19(3):370–379.

Erdös, P. and Turán, P. (1950). On the Distribution of Roots of Polynomials. *The Annals of Mathematics Second Series*, 51(1):105–119.

Eriksson, J., Ollila, E., and Koivunen, V. (2010). Essential Statistics and Tools for Complex Random Variables. *IEEE Transactions on Signal Processing*, 58(10):5400–5408.

Folland, G. (1984). *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, Inc., United States.

Folland, G. B. (2005). Higher-order derivatives and taylor's formula in several variables. *Preprint*, pages 1–4.

Foreman, M. and Henry, R. (1989). The harmonic analysis of tidal model time series. *Advances in Water Resources*, 12(3):109–120.

Fox, C. (1957). Some applications of Mellin transforms to the theory of bivariate statistical distributions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 53(3):620–628.

Fuhrmann, D. (1999). Complex Random Variables and Stochastic Processes. In Madisetti, V. and Williams, D., editors, *The Digital Signal Processing Handbook*, volume 196, chapter 60. CRC Press LLC, Boca Raton.

Galambos, J. and Simonelli, I. (2004). *Products of Random Variables: Applications to Problems of Physics and to Arithmetical Functions*. Pure and Applied Mathematics: A Dekker series of monographs and textbooks. CRC Press, Boca Raton.

Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium*. Little, Brown and Company. Translation by Charles Henry Davis in 1857.

Geenens, G. (2021). Mellin-Meijer kernel density estimation on $\mathbb{R}^+$. *Annals of the Institute of Statistical Mathematics*, 73:953–977.

Geenens, G. and Lafaye de Micheaux, P. (2020). The Hellinger Correlation. *Journal of the American Statistical Association*, ahead-of-print(ahead-of-print):1–15.

Giri, N. C. (2004). *Multivariate Statistical Analysis*. Marcel Dekker, Inc., USA, 2nd edition.

Goh, S. L., Chen, M., ć, D. P., Aihara, K., Obradovic, D., and Mandic, D. (2006). Complex-valued forecasting of wind profile. *Renewable Energy*, 31(11):1733 – 1750.

Gonella, J. (1972). A rotary-component method for analysing meteorological and oceanographic vector time series. *Deep-Sea Research*, 19(12):833 – 846.

Goodman, N. R. (1963). Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction). *Annals of Mathematical Statistics*, 34(1):152–177.

Goyal, P., Knuth, K. H., and Skilling, J. (2010). Origin of complex quantum amplitudes and Feynman's rules. *Physical Review A*, 81(2):022109.

Grothe, O., Schnieders, J., and Segers, J. (2014). Measuring association and dependence between random vectors. *Journal of Multivariate Analysis*, 123:96–110.

Hahn, T. (2007). Routines for the diagonalization of complex matrices. *arXiv*:0607103 [physics.comp-ph].

Hallin, M., del Barrio, E., Albertos, J. C., and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165.

Halliwell, L. J. (2015). Complex Random Variables. In *Casualty Actuarial Society E-Forum*.

Hamilton, W. (1837). Theory of Conjugate Functions or Algebraic Couples; with a Preliminary and Elementary Essay on Algebra as the Science of Pure Time. *Transactions of the Royal Irish Academy*, 17(1):293–422. Edited by David R. Wilkins 2000.

Hammersley, J. (1956). The Zeros of a Random Polynomial. *Berkeley Symposium on Mathematical Statistics and Probability*, 3.2:89–111.

Hinds, W. (1974). Moments of Complex Random Variables Related to a Certain Class of Characteristic Functions. *The Indian Journal of Statistics*, 36(2):219 – 222.

Hirose, A., editor (2013). *Complex-Valued Neural Networks: Advances and Applications*, chapter 1, pages 1–31. John Wiley & Sons, Inc., Hoboken, New Jersey, 1st edition.

Holland, P. and Wang, Y. (1987). Dependence function for continuous bivariate densities. *Communication in Statistics - Theory and Methods*, 16(3):863–876.

Horel, J. (1984). Complex Principal Component Analysis: Theory and Examples. *Journal of Applied Meteorology and Climatology*, 23(12):1660–1673.

Hörmander, L. (1990). *An Introduction to Complex Analysis In Several Variables*. Elsevier Science Publishers B.V., Amsterdam, 3rd edition.

Ikramov, K. D. (2012). Takagi's decomposition of a symmetric unitary matrix as a finite algorithm. *Computational Mathematics and Mathematical Physics*, 52(1):1–3.

Jansson, C. (2019). *Quantum Information Theory for Engineers: Free Climbing through Physics and Probability*.

Jaoude, A. A. (2016). The paradigm of complex probability and Chebyshev's inequality. *Systems Science and Control Engineering*, 4(1):99 – 137.

Jeannerod, C.-P., Kornerup, P., Louvet, N., and Muller, J.-M. (2017). Error bounds on complex floating-point multiplication with an FMA. *Mathematics of Computation*, 86:881–898.

Jiang, J. (2010). *Large Sample Techniques for Statistics*. Springer, New York, NY.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Education, Inc., New Jersey.

Júlvez, J. (2015). A straightforward method to compute average stochastic oscillations from data samples. *BMC Bioinformatics*, 16(333):1–17.

Kac, M. (1943). On the Average Number of Real Roots of a Random Algebraic Equation. *Bulletin of the American Mathematical Society*, 49(4):314–320.

Kesemen, O., Tiryaki, B. K., Tezel, O., and Pak, A. (2020). Some Statistics in Polar Coordinate System with Uniform Angles. *Gazi University Journal of Science*, 33(3):846–867.

Knuth, K. (2016). Why Square Roots of Probabilities? *arXiv*:1603.08427v1 [math.GM].

Kolmogorov, A. N. (1933). *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York. 2nd edition published in 1956 translated by Nathan Morrison.

Kotlarski, I. (1965). On the Generalized Mellin Transform of a Complex Random Variable and Its Applications. *The Annals of Mathematical Statistics*, 36(5):1459–1467.

Krishnaiah, P. R. and Lin, J. (1986). Complex elliptically symmetric distributions. *Communications in Statistics - Theory and Methods*, 15(12):3693–3718.

Lefèvre, V. and Muller, J.-M. (2019). Accurate Complex Multiplication in Floating-Point Arithmetic. In *2019 IEEE 26th Symposium on Computer Arithmetic (ARITH)*, pages 23–29.

Lin, Y. (2021). Exploring Exotic Probability Theory. *AMSI Vacation Research Scholarships 2020-21*, pages 1–20.

Loomis, L. H. (1953). *An Introduction to Abstract Harmonic Analysis*. D. Van Nostrand Company, Inc., Princeton, New Jersey.

Lukacs, E. (1970). *Characteristic Functions*. Charles Griffin and Company Limited, London.

Lundeen, J., Sutherland, B., Patel, A., Stewart, C., and Bamber, C. (2011). Direct measurement of the quantum wavefunction. *Nature*, 474:188–191.

Lvovsky, A. I. and Raymer, M. G. (2009). Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81:299–332.

Lyons, R. (2013). Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284–3305.

Mardia, K. and Jupp, P. (2000). *Directional Statistics*. John Wiley & Sons Ltd, Chichester, West Sussex.

Marshall, A. W. and Olkin, I. (1985). A Family of Bivariate Distributions Generated by the Bivariate Bernoulli Distribution. *Journal of the American Statistical Association*, 80(390):332–338.

Merzbach, U. and Boyer, C. (2011). *A History of Mathematics*. John Wiley & Sons, Inc., Hoboken, New Jersey, 3rd edition.

Mooers, C. N. K. (1973). A technique for the cross spectrum analysis of pairs of complex-valued time series, with emphasis on properties of polarized components and rotational invariants. *Deep Sea Research and Oceanographic Abstracts*, 20(12):1129 – 1141.

Muller, J.-M., Brunie, N., de Dinechin, F., Jeannerod, C.-P., Joldes, M., Lefèvre, V., Melquiond, G., Revol, N., and Torres, S. (2018). *Handbook of Floating-Point Arithmetic*, chapter 11, pages 435–452. Birkhäuser, Cham, 2nd edition.

Neeser, F. and Massey, J. (1993). Proper complex random processes with applications to information theory. *IEEE Transactions on Information Theory*, 39(4):1293–1302.

Nguyen, X.-H. T. and Choi, M.-S. (2021). Ultimate precision of direct tomography of wave functions. *Quantum Information Processing*, 20(221).

Ogawa, S., Lee, T.-M., Nayak, A. S., and Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, 14(1):68–78.

Olhede, S. (2006). Probability Density Functions for Complex Variables. *IEEE Transactions on Information Theory*, 52(3):1212–1217.

Ollila, E., Eriksson, J., and Koivunen, V. (2011). Complex Elliptically Symmetric Random Varibales — Generation, Characterization, and Circularity Tests. *IEEE Transactions On Signal Processing*, 59(1):58–69.

Ollila, E. and Koivunen, V. (2010). Robust Estimation Techniques for Complex-Valued Random Vectors. In Adali, T. and Haykin, S., editors, *Adaptive Signal Processing, Next Generation Solutions*, Wiley Series in Adaptive and Learning Systems for Signal Processing, Communication, and Control, pages 87–141. John Wiley & Sons, Inc., Hoboken, New Jersey.

Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012). Complex Elliptically Symmetric Distributions: Survey, New Results and Applications. *IEEE Transactions on Signal Processing*, 60(11):5597–5625.

Osius, G. (2004). The association between two random elements: A complete characterization and odds ratio models. *Metrika*, 60:261–277.

Osius, G. (2009). Asymptotic Inference for Semiparametric Association Models. *The Annals of Statistics*, 37(1):459–489.

Pewsey, A. and García-Portugués, E. (2021). Recent advances in directional statistics. *TEST*, 30:1–58.

Picinbono, B. (1994). On circularity. *IEEE Transactions on Signal Processing*, 42(12):3473–3482.

Picinbono, B. and Chevalier, P. (1995). Widely linear estimation with complex data. *IEEE Transactions on Signal Processing*, 43(8):2030–2033.

Remmert, R. (1991). *Theory of Complex Functions*. Springer-Verlag, New York, NY. Translated by Robert B. Burckel.

Renou, M., Trillo, D., Weilenmann, M., Thinh, L. P., Tavakoli, A., Gisin, N., Acin, A., and Navascues, M. (2021). Quantum physics needs complex numbers. *arXiv*:2101.10873 [quant-ph].

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis*. Wadsworth, Inc., Belmont, CA, 2nd edition.

Ripley, B., Venables, B., Bates, D. M., Honik, K., Gebhardt, A., and Firth, D. (2021). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-54.

Romero-Jerez, J. M. and Lopez-Martinez, F. J. (2017). A New Framework for the Performance Analysis of Wireless Communications under Hoyt (Nakagami-q) Fading. *IEEE Transactions on Information Theory*, 63(3):1693 – 1702.

Rowe, D. and Logan, B. (2004). A complex way to compute fMRI activation. *NeuroImage*, 23(3):1078 – 1092.

Sandoval, S., Kissling, E., Ansorge, J., and the SVEKALAPKO Seismic Tomography Working Group (2004). High-resolution body wave tomography beneath the SVEKALAPKO array — II. Anomalous upper mantle structure beneath the central Baltic Shield. *Geophysical Journal International*, 157(1):200–214.

Schreier, P. J. and Scharf, L. L. (2010). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, Cambridge.

Schrödinger, E. (1926). An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, 28:1049–1070.

Sijbers, J., den Dekker, A., Scheunders, P., and Van Dyck, D. (1998). Maximum-likelihood estimation of Rician distribution parameters. *IEEE Transactions on Medical Imaging*, 17(3):357–361.

Sob, U. M., Bester, H. L., Smirnov, O. M., Kenyon, J. S., and Grobler, T. L. (2020). Radio interferometric calibration using a complex Student's t-distribution and Wirtinger derivatives. *Monthy Notices of the Royal Astronomical Society*, 491(1):1026–1042.

Spitzer, F. (1955). On a Class of Random Variables. *Proceedings of the American Mathematical Society*, 6(3):494 – 505.

Srinivasan, S. and Sudarshan, E. (1994). Complex measures and amplitudes, generalized stochastic processes and their applications to quantum mechanics. *Journal of Physics A: Mathematical and General*, 27(2):517–537.

Srinivasan, S. K. (1997). Quantum mechanics via extended measures. *Journal of Physics A: Mathematical and General*, 30(23):8297–8315.

Stahl, S. (2006). The Evolution of the Normal Distribution. *Mathematics Magazine*, 79(2):96 – 113.

Stepić, A. I. and Ognjanović, Z. (2014). Complex valued probability logics. *Publications de l'Institut Mathématique*, 95(109):73 – 86.

Stewart, J. (2005). *Multivariable Calculus: Concepts and Contexts*. Thomson Brooks/Cole, Belmont, CA.

Subrahmaniam, K. (1970). On Some Applications of Mellin Transforms to Statistics: Dependent Random Variables. *SIAM Journal on Applied Mathematics*, 19(4):658–662.

Suparman (2020). Hierarchical Bayesian Choice of Laplacian ARMA Models Based on Reversible Jump MCMC Computation. *International Journal of Computational Intelligence Systems*, 13:310–317.

Svetunkov, S. (2012). *Complex-Valued Modeling in Economics and Finance*. Springer-Verlag New York, 1st edition.

Székely, G. (2005). Half of a Coin: Negative Probabilities. *Wilmott Magazine*, Technical Article 3:66 – 68.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769 – 2794.

Tao, T. (2012). *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society.

Tjøstheim, D., Otneim, H., and Støve, B. (2018). Statistical dependence: Beyond Pearson's ρ. *arXiv*:1809.10455 [math.ST].

Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., and Pal, C. J. (2018). Deep Complex Networks. In *International Conference on Learning Representations*.

Trampitsch, S. (2013). Complex-Valued Data Estimation: Second-order statistics and widely linear estimators. Master's thesis, The University of Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austria.

Ushakov, N. G. (1999). *Selected Topics in Characteristic Functions*. VSP BV, Zeist, NL.

van den Bos, A. (1995). The multivariate complex normal distribution—a generalization. *IEEE Transactions on Information Theory*, 41(2):537–539.

Varentsov, I. M., Sokolova, E., Martanus, E. R., and Nalivaiko, K. V. (2003). System of electromagnetic field transfer operators for the BEAR array of simultaneous soundings: Methods and results. *Izvestiya-physics of The Solid Earth*, 39:118–148.

Wand, M., Moler, C., and Ripley, B. (2021). *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. R package version 2.23-20.

Wang, M. (2021). Axiomatic definition of complex probability. *Zenodo*:4695269.

Wang, S., Cheng, H., Ying, L., Xiao, T., Ke, Z., Zheng, H., and Liang, D. (2020). DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magnetic Resonance Imaging*, 68(1):136–147.

Wang, X., Zhu, J., Pan, W., Zhu, J., and Zhang, H. (2021). Nonparametric Statistical Inference via Metric Distribution Function in Metric Spaces. *arXiv Pre-print Archive*, arXiv:2107.07317:1–31.

Wang, Y. J. (1993). Construction of continuous bivariate density functions. *Statistica Sinica*, 3(1):173–187.

Wasserman, L. (2010). *All of Statistics*. Springer, New York, NY.

Weil, H. (1954). The Distribution of Radial Error. *The Annals of Mathematical Statistics*, 25(1):168 – 170.

Weingarten, D. (2002). Complex probabilities on R^N as Real Probabilities on C^N and an Application to Path Integrals. *Physical Review Letters*, 89(240201).

Wigner, E. (1932). On the Quantum Correction For Thermodynamic Equilibrium. *Physical Review*, 40:749–759.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press, Cambridge.

Wirtinger, W. (1927). Zur formalen Theorie der Funktionen von mehr komplexen Veränderlichen. *Mathmatische Annalen*, 97:357–376.

Wooding, R. A. (1956). The Multivariate Distribution of Complex Normal Variables. *Biometrika*, 43(1/2):212–215.

Xu, W. and Qiao, S. (2008). A Divide-and-Conquer Method for the Takagi Factorization. *SIAM Journal on Matrix Analysis and Applications*, 30(1):142–153.

Yan, W., Yang, L., Yang, X., and Ren, W. (2019). Statistical modeling for fast Fourier transform coefficients of operational vibration measurements with non-Gaussianity using complex-valued t distribution. *Mechanical Systems and Signal Processing*, 132:291–314.

Youssef, S. (1994). Quantum mechanics as Bayesian complex probability theory. *Modern Physics Letters A*, 9(28):2571 – 2586.

Zolotarev, V. M. (1957). Mellin-Stieltjes Transforms in Probability Theory. *Theory of Probability and Its Applications*, 2(4):433–460.

Appendix A

Proofs and Derivations

In each appendix section, there will be a linking word (either the first word, or the actual theorem/proposition number) back to the corresponding section in the body.

A.1 Wirtinger Derivative Conjugate Identity

Proof of the claim that $\frac{\partial f}{\partial z} = \left(\frac{\partial f}{\partial z^*}\right)^*$ if and only if f is real-valued up to a constant

Proof. First handling the if direction, if f is real valued (up to a constant), then $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are both strictly real, as x and y are themselves real. Then it is evident by the form of the Wirtinger derivatives that

$$\left(\frac{\partial f}{\partial z^*}\right)^* = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right)^* = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) = \frac{\partial f}{\partial z}.$$

Conversely, if $\frac{\partial f}{\partial z} = \left(\frac{\partial f}{\partial z^*}\right)^*$, let $f = u + iv$, for u, v real valued functions. Then the equality of the Wirtinger derivatives implies that

$$\begin{aligned}
\frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} - i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right) &= \frac{1}{2}\left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} + i\frac{\partial u}{\partial y} - \frac{\partial v}{\partial y}\right)^*, \\
\Rightarrow \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} - i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} &= \frac{\partial u}{\partial x} - i\frac{\partial v}{\partial x} - i\frac{\partial u}{\partial y} - \frac{\partial v}{\partial y}, \\
\Rightarrow i\frac{\partial v}{\partial x} &= \frac{\partial v}{\partial y}.
\end{aligned}$$

However as v is real, so are $\frac{\partial v}{\partial x}$ and $\frac{\partial v}{\partial y}$. Hence the only solution to $i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y}$ must be for $\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} = 0$. Hence $f = u + K$ for some complex constant K and so f is real up to a constant. □

A.2 Derivation of Wirtinger Derivatives

Derivation of the Wirtinger derivatives, and in particular the novel polar form from the Cartesian versions which were given in Wirtinger (1927). The derivation of the Cartesian form follows Ahlfors (1979, p. 27). Let $f : \mathbb{C} \to \mathbb{C} : z \mapsto f(z)$ be

a complex function. By identifying $\mathbb{C}$ with $\mathbb{R}^2$ via $z = x + iy$ so that we have a function $\tilde{f} : \mathbb{R}^2 \to \mathbb{C}$, we can use the change of variable

$$\begin{aligned} x &= \frac{z + z^*}{2}, \\ y &= -\frac{i(z - z^*)}{2}, \end{aligned}$$

to obtain the Cartesian Wirtinger derivatives, treating z and z^* as independent variables. Specifically, by identifying f and $\tilde{f}$ as functions of x, y,

$$\begin{aligned} \frac{\partial f}{\partial z} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z} \\ &= \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right). \\ \frac{\partial f}{\partial z^*} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial z^*} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z^*} \\ &= \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right). \end{aligned}$$

Now, let $x = r\cos(\theta)$ and $y = r\sin(\theta)$ such that $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}\left(\frac{y}{x}\right) \pm \pi$ (depending upon quadrant in the complex plane). Then,

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial x} \\ &= \frac{\partial f}{\partial r}\frac{x}{r} - \frac{\partial f}{\partial \theta}\frac{y}{r^2} \\ &= \frac{\partial f}{\partial r}\cos(\theta) - \frac{\partial f}{\partial \theta}\frac{\sin(\theta)}{r}, \\ \frac{\partial f}{\partial y} &= \frac{\partial f}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial y} \\ &= \frac{\partial f}{\partial r}\frac{y}{r} + \frac{\partial f}{\partial \theta}\frac{x}{r^2} \\ &= \frac{\partial f}{\partial r}\sin(\theta) + \frac{\partial f}{\partial \theta}\frac{\cos(\theta)}{r}. \end{aligned}$$

Hence

$$\begin{aligned} \frac{\partial f}{\partial z} &= \frac{1}{2}\left(\frac{\partial f}{\partial r}\cos(\theta) - \frac{\partial f}{\partial \theta}\frac{\sin(\theta)}{r} - i\left(\frac{\partial f}{\partial r}\sin(\theta) + \frac{\partial f}{\partial \theta}\frac{\cos(\theta)}{r}\right)\right) \\ &= \frac{1}{2}\left(\frac{\partial f}{\partial r}(\cos(\theta) - i\sin(\theta)) - \frac{i}{r}\frac{\partial f}{\partial \theta}(\cos(\theta) - i\sin(\theta))\right) \\ &= \frac{e^{-i\theta}}{2}\left(\frac{\partial f}{\partial r} - \frac{i}{r}\frac{\partial f}{\partial \theta}\right). \end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial f}{\partial z^*} &= \frac{1}{2}\left(\frac{\partial f}{\partial r}\cos(\theta) - \frac{\partial f}{\partial \theta}\frac{\sin(\theta)}{r} + i\left(\frac{\partial f}{\partial r}\sin(\theta) + \frac{\partial f}{\partial \theta}\frac{\cos(\theta)}{r}\right)\right) \\ &= \frac{1}{2}\left(\frac{\partial f}{\partial r}(\cos(\theta) - i\sin(\theta)) + \frac{i}{r}\frac{\partial f}{\partial \theta}(\cos(\theta) - i\sin(\theta))\right) \\ &= \frac{e^{i\theta}}{2}\left(\frac{\partial f}{\partial r} + \frac{i}{r}\frac{\partial f}{\partial \theta}\right).\end{aligned}$$

A.3 Proof of the Univariate Wirtinger Taylor Expansion

Proof of the univariate Wirtinger Taylor expansion given in Theorem 2.4.6.

Proof.[1] Let $\omega = a + ib$ and $h = p + iq$. Now introduce the functions $u, v : \mathbb{C} \to \mathbb{R}$ defined such that $f = u + iv$. Also, by representing any input complex number $z = x + iy$ by $(x, y) \in \mathbb{R}^2$, we let $\tilde{f} : \mathbb{R}^2 \to \mathbb{C}$ be such that $\tilde{f}(x, y) = f(z)$ for all $z = x + iy \in \mathbb{C}$. Subsequently, we have the functions $\tilde{u}, \tilde{v} : \mathbb{R}^2 \to \mathbb{R}$ such that $\tilde{f} = \tilde{u} + i\tilde{v}$, which by linearity of the Wirtinger derivatives must also be C^{n+1} in the usual sense on the equivalent open set $U_{\mathbb{R}} \subseteq \mathbb{R}^2$. Now, for $\boldsymbol{\omega}_{\mathbb{R}} = (a, b)^\top$ and $\boldsymbol{h}_{\mathbb{R}} = (p, q)^\top$, consider the bivariate real Taylor expansion for $\tilde{u}$ (that of $\tilde{v}$ will also follow) in $U_{\mathbb{R}}$ near $\boldsymbol{\omega}_{\mathbb{R}}$ given by

$$\tilde{u}(\boldsymbol{\omega}_{\mathbb{R}} + \boldsymbol{h}_{\mathbb{R}}) = \tilde{u}(\boldsymbol{\omega}_{\mathbb{R}}) + \sum_{k=1}^{n} \frac{1}{k!}\big[\boldsymbol{h}_{\mathbb{R}}^\top \nabla_x\big]\tilde{u}(\boldsymbol{\omega}_{\mathbb{R}}) + \frac{1}{(n+1)!}\big[\boldsymbol{h}_{\mathbb{R}}^\top \nabla_x\big]^{n+1}\tilde{u}(\boldsymbol{\omega}_{\mathbb{R}} + c\boldsymbol{h}_{\mathbb{R}}),$$

for some $c \in [0, 1]$ (Folland (2005); Jiang (2010)).

We can now use the conversion matrix $\mathbf{J}_1$ to involve the augmented complex vector $\underline{\boldsymbol{h}} = (h, h^*)^\top$, by the equality from (2.1.1) that $\boldsymbol{h}_{\mathbb{R}} = \mathbf{J}_1\underline{\boldsymbol{h}}$. By also recognising from the definition of the Cartesian Wirtinger derivatives that $\nabla_{\underline{z}} = \mathbf{J}_1^\top \nabla_x$, we have that

$$\boldsymbol{h}_{\mathbb{R}}^\top \nabla_x = \underline{\boldsymbol{h}}^\top \mathbf{J}_1^\top \nabla_x = \underline{\boldsymbol{h}}^\top \nabla_{\underline{z}}. \tag{A.3.1}$$

Now that we are employing the Wirtinger derivatives, we formally enact these upon the function u as opposed to $\tilde{u}$, evaluated at ω instead of $\boldsymbol{\omega}_{\mathbb{R}}$. Thus we resolve the Taylor expansion in (A.3.1) to complex notation where

$$u(\omega + h) = u(\omega) + \sum_{k=1}^{n} \frac{1}{k!}\big[\underline{\boldsymbol{h}}^\top \nabla_{\underline{z}}\big]u(\omega) + \frac{1}{(n+1)!}\big[\underline{\boldsymbol{h}}^\top \nabla_{\underline{z}}\big]^{n+1}u(\omega + ch),$$

for some $c \in [0, 1]$. An analogous result holds for $v(\omega + h)$.

Finally, as $f = u + iv$, it remains that around $\omega \in U$,

$$f(\omega + h) = f(\omega) + \sum_{k=1}^{n} \frac{1}{k!}\big[\underline{\boldsymbol{h}}^\top \nabla_{\underline{z}}\big]f(\omega) + \frac{1}{(n+1)!}\big[\underline{\boldsymbol{h}}^\top \nabla_{\underline{z}}\big]^{n+1}\big(u(\omega + ch) + iv(\omega + dh)\big),$$

[1] Based off working by Pierre Lafaye de Micheaux and Gilles Ducharme

for some $c, d \in [0,1]$ which may not necessarily be equal. However, provided all the derivatives are bounded on an open neighbourhood of ω containing $\omega + h$, then it can be verified that the remainder term involving the $(n+1)^{\text{th}}$-order derivatives is $o(|h|^n)$. Note that this is already a property of the real Taylor expansions of $\tilde{u}$ and $\tilde{v}$ and consequently given by the fact that $||\boldsymbol{h}_{\mathbb{R}}||^n = |h|^n$ (see e.g. Corollary 1 in Folland (2005) or Jiang (2010, p.87)). This completes the proof. □

A.4 Multivariate Wirtinger Taylor Expansion

Theorem A.4.1 (Multivariate Wirtinger Taylor expansion)**.** *For a function $f : \mathbb{C}^m \to \mathbb{C}$ that is $C_{\mathbb{R}}^{k+1}$ (now across all mixed derivatives) and has bounded derivatives in an open set $U \subseteq \mathbb{C}^m$, the Wirtinger Taylor expansion about $\boldsymbol{\omega} \in \mathbb{C}^n$ is given, for $\boldsymbol{\omega} + \boldsymbol{h} \in U$, by*

$$f(\boldsymbol{\omega} + \boldsymbol{h}) = f(\boldsymbol{\omega}) + \sum_{k=1}^{n} \frac{1}{k!} \left[\underline{\boldsymbol{h}}^{\top} \nabla_{\underline{\boldsymbol{z}}}\right]^k f(\boldsymbol{\omega}) + o(\|\boldsymbol{h}\|^n),$$

noting now that $\underline{\boldsymbol{h}} = \left(\boldsymbol{h}^{\top}, \boldsymbol{h}^{\mathcal{H}}\right)^{\top} \in \mathbb{C}^{2m}$ and $\nabla_{\underline{\boldsymbol{z}}} = \left(\frac{\partial}{\partial z_1}, \ldots, \frac{\partial}{\partial z_m}, \frac{\partial}{\partial z_1^*}, \ldots, \frac{\partial}{\partial z_m^*}\right)^{\top}$.

The second-order univariate expansion given in Example 2.4.7 for the univariate case can also be extended to aid the understanding of what happens in the multivariate case, involving the gradient operators

$$\nabla_{\boldsymbol{z}} = \begin{pmatrix} \frac{\partial}{\partial z_1} \\ \vdots \\ \frac{\partial}{\partial z_m} \end{pmatrix} \quad \text{and} \quad \nabla_{\boldsymbol{z}^*} = \begin{pmatrix} \frac{\partial}{\partial z_1^*} \\ \vdots \\ \frac{\partial}{\partial z_m^*} \end{pmatrix},$$

and the Hessian matrices

$$\mathbf{H}_z = \left[\frac{\partial^2}{\partial z_k \partial z_l}\right], \quad \mathbf{H}_{|z|} = \left[\frac{\partial^2}{\partial z_k \partial z_l^*}\right], \quad \mathbf{H}_{z^*} = \left[\frac{\partial^2}{\partial z_k^* \partial z_l^*}\right].$$

Note that whilst $\mathbf{H}_z$ and $\mathbf{H}_{z^*}$ will be symmetric if the derivatives commute, $\mathbf{H}_{|z|}$ will only be Hermitian if f is real valued (up to a constant).

The second order Wirtinger Taylor expansion of $f : \mathbb{C}^m \to \mathbb{C}$ about the point $\boldsymbol{\omega} \in \mathbb{C}^m$ is given by

$$\begin{aligned} f(\boldsymbol{\omega} + \boldsymbol{h}) = f(\boldsymbol{\omega}) &+ \boldsymbol{h}^{\top} \nabla_{\boldsymbol{z}} f(\boldsymbol{\omega}) + \boldsymbol{h}^{\mathcal{H}} \nabla_{\boldsymbol{z}^*} f(\boldsymbol{\omega}) \\ &+ \boldsymbol{h}^{\top} \mathbf{H}_z(f)(\boldsymbol{\omega}) \boldsymbol{h} + \boldsymbol{h}^{\mathcal{H}} \mathbf{H}_{|z|}(f)(\boldsymbol{\omega}) \boldsymbol{h} + \boldsymbol{h}^{\mathcal{H}} \mathbf{H}_{z^*}(f)(\boldsymbol{\omega}) \boldsymbol{h}^* + o(\|\boldsymbol{h}\|^n), \end{aligned}$$

where $\|\boldsymbol{h}\|^n = (|h_1|^2 + \ldots + |h_m|^2)^{\frac{n}{2}}$.

A.5 Floating-point algorithms 2Sum and Fast2Mult

Supplementary algorithms to Algorithm 1.

A.5.1 2Sum

Computes the sum and rounding error of two floating point numbers.

Algorithm 7 2Sum(a,b) (Lefèvre and Muller (2019)).

Require: a, b two floating point (real) numbers.
$s \leftarrow RN(a + b)$
$a' \leftarrow RN(s - b)$
$b' \leftarrow RN(s - a')$
$\delta_a \leftarrow RN(a - a')$
$\delta_b \leftarrow RN(b - b')$
$t \leftarrow RN(\delta_a + \delta_b)$
return (s,t)

A.5.2 Fast2Mult

Computes the multiplication and associated rounding error.

Algorithm 8 Fast2Mult(a,b) (Lefèvre and Muller (2019)). Requires an FMA instruction.

Require: a, b two floating point (real) numbers.
$\pi \leftarrow RN(ab)$
$\rho \leftarrow RN(ab - \pi)$ ▷ using FMA
return (π, ρ)

A.6 Complex area integration

Details around the differential geometry of identifying $\mathbb{C}$ as a real two-dimensinoal Riemannian manifold. A smooth manifold is a topological space equipped with functions called charts mapping to Euclidean space. We can treat $\mathbb{C}$ as a topological space, understanding the topology which defines open sets to be generated by open balls. Then, treating $\mathbb{C}$ as a two-dimensional real manifold, we parameterise it with two variables. Here, we will choose (z, z^*), which is a natural parameterisation corresponding to a single complex number. Then, we can construct the single chart where, for $z = x + iy$,

$$\varphi : \mathbb{C} \to \mathbb{R}^2 : \begin{pmatrix} z \\ z^* \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}.$$

This indeed defines a smooth structure on $\mathbb{C}$ because the transition function $\varphi\varphi^{-1} = \mathrm{id} : \mathbb{R}^2 \to \mathbb{R}^2$ is a diffeomorphism. Under this smooth structure, any function $f : \mathbb{C} \to \mathbb{C}$ is said to be smooth if $\varphi f \varphi^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is smooth (or more loosely differentiable). Hence, we observe that being differentiable here requires differentiability in x and y — and so we recover Wirtinger calculus.

Now, to define integration, we define a metric on $\mathbb{C}$ to make it a Riemannian manifold. The metric, given the pairwise parameterisation will be the real scalar inner product such that for $z, w \in \mathbb{C}$,

$$\left\langle \begin{pmatrix} z \\ z^* \end{pmatrix}, \begin{pmatrix} w \\ w^* \end{pmatrix} \right\rangle = \frac{1}{2}(zw^* + z^*w) = \Re(zw^*).$$

This is a Riemannian metric which we will denote G and its associated matrix is represented by the identity matrix $\underline{G} = \mathbf{I}_2$. Hence, for some region $A \subseteq \mathbb{C}$ with corresponding $A_{x,y} = \varphi(A)$, we define integration on $\mathbb{C}$ by

$$\begin{aligned}\iint_A f(z)\,\mathrm{d}\mathbb{C} &= \iint_{A_{x,y}} f(x+iy)\sqrt{\det(G(x,y))}\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{A_{x,y}} f(x+iy)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

as $\det(G) = \det \underline{G} = 1$.

Now, in terms of differentials, we have by simple manipulation of the Jacobian (such as that done in Appendix A.2) that $\mathrm{d}z \wedge \mathrm{d}z^* = -2i\mathrm{d}x \wedge \mathrm{d}y$ (Hörmander (1990, p. 2)). Note that the wedge product $\mathrm{d}x \wedge \mathrm{d}y$ of differentials here may be thought of as a signed version of the area differential $\mathrm{d}x\,\mathrm{d}y$, which we frequently integrate with respect to. Basic properties are

$$\begin{aligned}\mathrm{d}x \wedge \mathrm{d}y &= -\mathrm{d}y \wedge \mathrm{d}x, \\ |\mathrm{d}x \wedge \mathrm{d}y| &= \mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

In particular the second property is why in several variable calculus at an undergraduate level, one changes variables by multiplying by the absolute value of the Jacobian determinant. Using the wedge products, instead we multiply by just the determinant, maintaining the sign. We will not delve into the deep differential geometric details here. Rather, we will assume a standard orientation on $\mathbb{C}$, namely that 'anticlockwise' is defined from the positive real axis to the positive negative axis. Under this standard orientation, we take $\mathrm{d}x \wedge \mathrm{d}y = \mathrm{d}x\,\mathrm{d}y$. Thus, an integral with respect to $\mathrm{d}x\,\mathrm{d}y$ may be written as $\frac{i}{2}$ times the integral with respect to $\mathrm{d}z \wedge \mathrm{d}z^*$. Note that we do not reduce this to $\frac{1}{2}\mathrm{d}z\,\mathrm{d}z^*$ in order to make explicit the choice of orientation on $\mathbb{C}$. We also do not know what it means to integrate with respect to the differential $\mathrm{d}z\,\mathrm{d}z^* = |\mathrm{d}z|^2$(?). That is an overview of why Wirtinger calculus and integration with respect to x and y are induced by this identification of $\mathbb{C}$ as a two-dimensional Riemannian manifold, which subsequently leads to the identification of distribution functions, such as densities, with their bivariate counterparts.

Note also that the use of $\mathrm{d}z \wedge \mathrm{d}z^*$ notation is advantageous to employ Stokes' formula and recover a complex contour integral around the boundary of A in $\mathbb{C}$, which can then itself be evaluated by usual methods of contour integration which map the contour to a one-dimensional manifold where

$$\int_{\partial A} F(z)\,\mathrm{d}z = \int_I F(z(t))z'(t)\,\mathrm{d}t,$$

for some interval $I \subseteq \mathbb{R}$ such that $z : I \to \partial A \subseteq \mathbb{C}$ traces out ∂A as t moves in I.

A.7 Proof of complex Stokes' formula

Proof of Proposition 2.6.1

Proof. The general notions are provided here at a more explicit elementary level than in Hörmander (1990). In $\mathbb{R}^2$, Stokes' or more directly Green's, theorem (Stewart (2005, p.933)) states that for partially differentiable functions g and h, and connected region $D \subseteq \mathbb{R}^2$ with simple closed boundary ∂D,

$$\iint_D \left(\frac{\partial h}{\partial x} - \frac{\partial g}{\partial y} \right) \mathrm{d}x \wedge \mathrm{d}y = \oint_{\partial D} g \, \mathrm{d}x + h \, \mathrm{d}y.$$

Now, instead let $F(z) = h(x, y) + ig(x, y)$ for all $z = x + iy$ and take $A \subseteq \mathbb{C}$ to be the region corresponding to $D \subseteq \mathbb{R}^2$. Then, keeping in mind that $\mathrm{d}z \wedge \mathrm{d}z^* = -2i \, \mathrm{d}x \wedge \mathrm{d}y$,

$$\begin{aligned}
\int_A \frac{\partial F}{\partial z^*} \mathrm{d}z \wedge \mathrm{d}z^* &= -2i \iint_D \frac{1}{2} \left[\frac{\partial h}{\partial x} - \frac{\partial g}{\partial y} + i \left(\frac{\partial g}{\partial x} + \frac{\partial h}{\partial y} \right) \right] \mathrm{d}x \wedge \mathrm{d}y \\
&= -i \oint_{\partial D} g \, \mathrm{d}x + h \, \mathrm{d}y - ih \, \mathrm{d}x + ig \, \mathrm{d}y \\
&= -i \oint_{\partial D} h(-i(\mathrm{d}x + i\mathrm{d}y)) + g(\mathrm{d}x + i\mathrm{d}y) \\
&= - \oint_{\partial D} h \, \mathrm{d}(x + iy) + ig \, \mathrm{d}(x + iy) \\
&= - \oint_{\partial A} F(z) \, \mathrm{d}z.
\end{aligned}$$

□

A.8 Existence of moments

Proof of Theorem 4.5.2, that for all m, n such that $m - n \in \mathbb{Z}$ and $m + n = k$, then $|\mathbb{E}\left(Z^m Z^{*n}\right)| \leq \mathbb{E}\left(|Z|^k\right)$.

Proof. First, we prove the inequality $\mathbb{E}\left(|Z|\right) \geq |\mathbb{E}\left(Z\right)|$ holds. Noting that $\mathbb{E}\left(\Re(Z)\right) = \Re(\mathbb{E}\left(Z\right))$, we being by expanding

$$\begin{aligned}
0 &\leq \mathbb{E}\left(|\sqrt{Z} \pm \sqrt{W}|^2 \right) \\
&= \mathbb{E}\left(|Z| + |W| \pm 2\Re\left(\sqrt{ZW^*} \right) \right), \\
\Rightarrow \mp 2\Re\left(\mathbb{E}\left(\sqrt{ZW^*} \right) \right) &\leq \mathbb{E}\left(|Z|\right) + \mathbb{E}\left(|W|\right).
\end{aligned}$$

Now, letting $W^* = Z$ and taking the absolute value of both sides, we have that

$$|\Re\left(\mathbb{E}\left(Z\right)\right)| \leq \mathbb{E}\left(|Z|\right). \tag{A.8.1}$$

Now, for a given random variable Z, $\mathbb{E}\left(Z\right) = |\mathbb{E}\left(Z\right)| e^{i\theta}$ for a non-random real constant θ. Define $\tilde{Z} = Ze^{-i\theta}$ such that $\mathbb{E}(\tilde{Z}) = |\mathbb{E}\left(Z\right)| \in \mathbb{R}$. Then continuing

(A.8.1) with $\tilde{Z}$, we obtain that

$$\begin{aligned} \left|\Re\left(\mathbb{E}(\tilde{Z})\right)\right| &\leq \mathbb{E}(|\tilde{Z}|), \\ \Rightarrow \left|\mathbb{E}(\tilde{Z})\right| &\leq \mathbb{E}(|\tilde{Z}|). \end{aligned}$$

But by construction, $|\tilde{Z}| = |Z|$ and $|\mathbb{E}(\tilde{Z})| = |\mathbb{E}(Z)|$, thus

$$|\mathbb{E}(Z)| \leq \mathbb{E}(|Z|). \tag{A.8.2}$$

Applying (A.8.2) to the random variable $Z^m Z^{*n}$, we obtain that

$$|\mathbb{E}(Z^m Z^{*n})| \leq \mathbb{E}(|Z|^m |Z^*|^n) = \mathbb{E}\left(|Z|^k\right).$$

Conversely, consider $Z^m Z^{*n} = U + iV$ for real random variables U, V. Then if $\mathbb{E}(Z^m Z^{*n})$ exists, this implies that so do $\mathbb{E}(U)$ and $\mathbb{E}(V)$. This is equivalent to saying $\mathbb{E}(|U|)$ and $\mathbb{E}(|V|)$ exist. But as $|Z^m Z^{*n}| \leq |U| + |V|$, we have that

$$\mathbb{E}\left(|Z|^k\right) = \mathbb{E}(|Z^m Z^{*n}|) \leq \mathbb{E}(|U|) + \mathbb{E}(|V|) < \infty.$$

Hence, if any k^{th}-order moment exists then all exist. This completes the proof as existence of lower order moments follows by non-negativity of the modulus and boundedness. □

A.9 Characteristic Functions in Directional Statistics

Background information about the characteristic function in directional statistics and why it is given as a series over the integers rather than a continuous function. The following is based off the explanation given in Mardia and Jupp (2000, Section 3.3). Let X be a continuous random variable representing an angle in $\mathbb{R}$. By the 2π-periodicity of angles, its density function is also 2π-periodic such that over any interval of length 2π, the density integrates to 1. By further consequence, considering a characteristic function

$$\varphi_X(t) = \mathbb{E}\left(e^{itX}\right),$$

since X and $2\pi + X$ represent the same direction, one would anticipate

$$\mathbb{E}\left(e^{itX}\right) = \mathbb{E}\left(e^{it(X+2\pi)}\right) = e^{it2\pi}\mathbb{E}\left(e^{itX}\right),$$

for all $t \in \mathbb{R}$. Thus, for the algebraic equality to hold, we require that $t \in \mathbb{Z}$. Hence, we instead obtain the characteristic series

$$\varphi_X(t) = \mathbb{E}\left(e^{itX}\right),\ t = 0, \pm 1, \pm 2, \ldots.$$

The density of X may still be recovered from this series, by a sum rather than a product whereby

$$f_X(x) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \varphi_X(t) e^{-itx}.$$

This will be important as a similar inversion will appear for the general Mellin transform.

A.10 Mellin Inversion Formulae

Proof that the Mellin inversion formulae given in Theorem 4.8.4 do indeed recover the density. Throughout these calculations, we will interchange the order of integration and summation, given all integrals and sums are assumed to converge. We will also use the Dirac delta defined by

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \, \mathrm{d}t = \frac{1}{\pi} \lim_{n\to\infty} \frac{\sin(n(x-y))}{x-y},$$

such that

$$\int_{-\infty}^{\infty} \delta(x) f(x) \, \mathrm{d}x = f(0).$$

A.10.1 General Mellin transform

To invert the Mellin transform, as we have not identified the principal argument, to identify the polar density, we must consider an arbitrary interval of length 2π beginning at φ for the argument and require $s - t \in \mathbb{Z}$ for the mixed moment to be single-valued. We have that

$$\mathcal{M}_Z(s,t) = \int_{\varphi}^{\varphi+2\pi} \int_0^{\infty} r^{s+t+1} e^{i(s-t)\theta} f_Z(re^{i\theta}) \mathrm{d}r \, \mathrm{d}\theta.$$

We now aim to recover the density $f_Z(w)$ for $w = te^{i\alpha}$. Following the inversion formula letting $u = s + t$ and $v = s - t$, we have for some constant $c \in (\alpha, \beta)$,

$$\begin{aligned}
&\frac{1}{4\pi^2 i} \int_{c-i\infty}^{c+i\infty} \sum_{v\in\mathbb{Z}} \int_{\varphi}^{\varphi+2\pi} \int_0^{\infty} r^{u+1} e^{iv\theta} \mathrm{d}r \, \mathrm{d}\theta \, t^{-u-2} e^{-i\alpha} \, \mathrm{d}u \\
&= \frac{1}{4\pi^2 i t^2} \int_0^{\infty} r \int_{c-i\infty}^{c+i\infty} \int_{\varphi}^{\varphi+2\pi} \sum_{v\in\mathbb{Z}} \left(\frac{r}{t}\right)^u e^{iv(\theta-\alpha)} \, \mathrm{d}u \, f_Z(re^{i\theta}) \, \mathrm{d}\theta \, \mathrm{d}r \\
&= \frac{1}{4\pi^2 i t^2} \int_0^{\infty} \int_{\varphi}^{\varphi+2\pi} r \int_{c-i\infty}^{c+i\infty} e^{u(\ln(r)-\ln(t))} \, \mathrm{d}u \, \frac{\theta-\alpha}{\sin\left(\frac{\theta-\alpha}{2}\right)} \pi\delta(\theta-\alpha) \, f_Z(re^{i\theta}) \, \mathrm{d}\theta \, \mathrm{d}r.
\end{aligned}$$

Note, here we have used the fact that

$$\begin{aligned}\sum_{v\in\mathbb{Z}} e^{iv(\theta-\alpha)} &= \lim_{n\to\infty} \frac{\sin\left(\frac{2n+1}{2}(\theta-\alpha)\right)}{\sin\left(\frac{\theta-\alpha}{2}\right)} \\ &= \frac{\theta-\alpha}{\sin\left(\frac{\theta-\alpha}{2}\right)} \lim_{m\to\infty} \frac{\sin(m(\theta-\alpha))}{\theta-\alpha} \\ &= \frac{\theta-\alpha}{\sin\left(\frac{\theta-\alpha}{2}\right)} \pi\delta(\theta-\alpha).\end{aligned}$$

Now making three substitutions: $\ln(r) = p$; $\mathrm{d}r = e^p\,\mathrm{d}p$ and $\ln(t) = q$ and $u = c+im$, and by fact that $\alpha \in (\varphi, \varphi+2\pi]$ we continue with

$$\begin{aligned}&\frac{1}{2\pi t^2}\int_0^\infty \int_\varphi^{\varphi+2\pi} e^{2p} \int_{-\infty}^{\infty} e^{c(p-q)} e^{im(p-q)}\,\mathrm{d}m\, \frac{\frac{\theta-\alpha}{2}}{\sin\left(\frac{\theta-\alpha}{2}\right)} \delta(\theta-\alpha)\, f_Z(e^p e^{i\theta})\,\mathrm{d}\theta\,\mathrm{d}p \\ &= \frac{1}{2\pi t^2}\int_0^\infty e^{2p} e^{c(p-q)} 2\pi\delta(p-q)\, f_Z(e^p e^{i\alpha})\,\mathrm{d}p \\ &= \frac{1}{t^2} e^{2q} e^{c(q-q)} f_Z(e^q e^{i\alpha}) \\ &= f_Z(te^{i\alpha}),\end{aligned}$$

as desired. Note that in the above calculations, we also made use of the limit of $\lim_{x\to 0}\frac{\sin(x)}{x} = 1$. Note that a more general notation, not assuming α takes value in an interval of length 2π is also valid, however we will again observe the behaviour of the Dirac delta will single out a specific value of alpha, and so assuming α as an argument ranges within one interval of length 2π for arbitrary φ is sufficient.

A.10.2 Principal Mellin transform

We have that, by identification with the polar density for $z = re^{i\theta}$,

$$\mathcal{M}_Z^{\mathrm{P}}(s,t) = \int_{-\pi}^{\pi}\int_0^\infty z^s z^{*t} f_Z(z)\, r\,\mathrm{d}r\,\mathrm{d}\theta.$$

Then, let $s = x+iy$ and $t = \omega - s^*$ for some constant ω such that it converges. Let also, $z = re^{i\theta}$ and $w = ue^{i\alpha}$. Then, we have that

$$\begin{aligned}&\frac{i}{2\pi^2}\iint_{\mathbb{C}} \mathcal{M}_Z^{\mathrm{P}}(s,t) z^{-s-1} z^{*s^*-\omega-1}\,\mathrm{d}s\wedge\mathrm{d}s^* \\ &= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{M}_Z(x+iy, \omega-(x-iy))\, z^{-(x+iy)-1} z^{*(x-iy)-\omega-1}\,\mathrm{d}x\wedge\mathrm{d}y \\ =&\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\pi}^{\pi}\int_0^\infty w^{x+iy} w^{*\omega-x+iy} f_Z(u,\alpha) u\,\mathrm{d}u\,\mathrm{d}\alpha\, z^{-(x+iy)-1} z^{*(x-iy)-\omega-1}\,\mathrm{d}x\,\mathrm{d}y \\ =&\frac{z^{*-\omega}}{zz^*\pi^2}\int_{-\pi}^{\pi}\int_0^\infty\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{w}{z}\right)^{x+iy}\left(\frac{w^*}{z^*}\right)^{-x+iy}\,\mathrm{d}x\,\mathrm{d}y\, w^{*\omega} f_Z(u,\alpha) u\,\mathrm{d}u\,\mathrm{d}\alpha.\end{aligned} \tag{A.10.1}$$

We pause here to expand and simplify the inner integrand

$$\begin{aligned}
&\left(\frac{w}{z}\right)^{x+iy}\left(\frac{w^*}{z^*}\right)^{-x+iy}\\
=&e^{x(\ln(u)-\ln(r))}e^{iy(\ln(u)-\ln(r))}e^{ix(\alpha-\theta)}e^{-y(\alpha-\theta)}e^{-x(\ln(u)-\ln(r))}e^{iy(\ln(u)-\ln(r))}e^{ix(\alpha-\theta)}e^{y(\alpha-\theta)}\\
=&e^{2iy(\ln(u)-\ln(r))}e^{2ix(\alpha-\theta)}.
\end{aligned}$$

Returning to (A.10.1), we can continue by breaking apart the integral in terms of x and y, making the substitutions $q = \ln(u)$; $\mathrm{d}u = e^q\,\mathrm{d}q$ and $p = \ln(r)$, and also remark that by substitution $\int_{-\infty}^{\infty} e^{2ix(\alpha-\theta)}\,\mathrm{d}x = \pi\delta(\alpha-\theta)$.

$$\begin{aligned}
(A.10.1) &= \frac{z^{*-\omega}}{zz^*\pi^2}\int_{-\pi}^{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{2ix(\alpha-\theta)}\,\mathrm{d}x\int_{-\infty}^{\infty} e^{2iy(q-p)}\,\mathrm{d}y\, e^{q\omega}e^{i\omega\alpha}f_Z(e^q,\alpha)e^{2q}\,\mathrm{d}q\,\mathrm{d}\alpha\\
&= \frac{z^{*-\omega}}{zz^*\pi^2}\int_{-\pi}^{\pi}\int_{-\infty}^{\infty}\pi\delta(\alpha-\theta)\pi\delta(q-p)e^{q\omega}e^{-i\omega\alpha}f_Z(e^q,\alpha)e^{2q}\,\mathrm{d}q\,\mathrm{d}\alpha\\
&= \frac{r^{-\omega}e^{i\omega\theta}}{r^2}\left(e^{p\omega}e^{-i\omega\theta}f_Z(e^p,\theta)e^{2p}\right)\\
&= \frac{r^{-\omega}r^{\omega}}{r^2}f_Z(r,\theta)r^2\\
&= f_Z(z),
\end{aligned}$$

as desired.

A.11 Contour Integral Details

Details for the integral

$$\int_{\Gamma}\frac{e^{iw}}{w}\,\mathrm{d}w.$$

On Γ_2, $w = Re^{it}$ for $t : 0 \to \pi$. So,

$$\begin{aligned}
\left|\int_{\Gamma_2}\frac{e^{iw}}{w}\,\mathrm{d}w\right| &= \left|\int_0^{\pi}\frac{e^{iR\cos(t)-R\sin(t)}}{Re^{it}}Rie^{it}\,\mathrm{d}t\right|\\
&= \left|\int_0^{\pi} i\,e^{iR\cos(t)-R\sin(t)}\,\mathrm{d}t\right|\\
&\le \int_0^{\pi}\left|i\,e^{iR\cos(t)-R\sin(t)}\right|\,\mathrm{d}t\\
&= \int_0^{\pi} e^{-R\sin(t)}\,\mathrm{d}t.
\end{aligned}$$

By Jordan's inequality (Brown and Churchill (2009, p. 273)), for $R > 0$,

$$\int_0^{\pi} e^{-R\sin(t)}\,\mathrm{d}t \le \frac{\pi}{R}.$$

Hence,

$$0 \leq \left| \int_{\Gamma_2} \frac{e^{iw}}{w} \, \mathrm{d}w \right| \leq \frac{\pi}{R} \overset{R \to \infty}{\longrightarrow} 0.$$

So as $R \to \infty$, $\int_{\Gamma_2} \frac{e^{iw}}{w} \, \mathrm{d}w \to 0$.

Now, for Γ_4, we have that $w = \varepsilon e^{it}$ where $t : \pi \to 0$. So,

$$\begin{aligned} \int_{\Gamma_4} \frac{e^{iw}}{w} \, \mathrm{d}w &= \int_{\pi}^{0} \frac{e^{\varepsilon e^{it}}}{\varepsilon e^{it}} i\varepsilon e^{it} \, \mathrm{d}t \\ &= i \int_{\pi}^{0} e^{i\varepsilon e^{it}} \, \mathrm{d}t. \end{aligned}$$

As $\varepsilon \to 0$, we have that the integrand tends to 1 (and we can swap the limit and the integral because the integral is bounded and converges). Hence,

$$\int_{\Gamma_4} \frac{e^{iw}}{w} \, \mathrm{d}w \overset{\varepsilon \to 0}{\longrightarrow} -i\pi.$$

A.12 Complex Poisson Derivation

Here we define the complex Poisson similar to that of the real Poisson distribution, by letting n in the binomial distribution tend to ∞ but such that $\lambda = \sqrt{n}\,\omega$ is finite. Then, from the definition of the complex Binomial distribution, it can be seen that

$$\begin{aligned} &\lim_{n \to \infty} \binom{n}{\Re(z)\ \Im(z)} (\Re(\omega))^{2\Re(z)} (\Im(\omega))^{2\Im(z)} (1 - |\omega|^2)^{n - \Re(z(1-i))} \\ &= \lim_{n \to \infty} \frac{n!}{\Re(z)!\ \Im(z)!\ (n - \Re(z) - \Im(z))!} \left(\frac{\Re(\lambda)^2}{n} \right)^{\Re(z)} \left(\frac{\Im(\lambda)^2}{n} \right)^{\Im(z)} \left(1 - \frac{|\lambda|^2}{n} \right)^{n - \Re(z) - \Im(z)} \\ &= \lim_{n \to \infty} \frac{(\Re(\lambda)^2)^{\Re(z)}}{\Re(z)!} \cdot \frac{(\Im(\lambda)^2)^{\Im(z)}}{\Im(z)!} \cdot \left(1 - \frac{|\lambda|^2}{n} \right)^{n} \cdot \frac{n!}{n^{\Re(z) + \Im(z)} (n - (\Re(z) + \Im(z)))!} \cdot \\ &\qquad \left(1 - \frac{|\lambda|^2}{n} \right)^{-\Re(z)} \cdot \left(1 - \frac{|\lambda|^2}{n} \right)^{-\Im(z)} \\ &= \frac{(\Re(\lambda)^2)^{\Re(z)}}{\Re(z)!} \frac{(\Im(\lambda)^2)^{\Im(z)}}{\Im(z)!} e^{-|\lambda|^2} \\ &= \frac{(\Re(\lambda)^2)^{\Re(z)}}{\Re(z)!} e^{-\Re(\lambda)^2} \cdot \frac{(\Im(\lambda)^2)^{\Im(z)}}{\Im(z)!} e^{-\Im(\lambda)^2}. \end{aligned}$$

Hence, the complex Poisson appears to exhibit the same density as if the Real and Imaginary components are independent Pois($\Re(\lambda)^2$) and Pois($\Im(\lambda)^2$) respectively, despite the real and imaginary components not being independent as Bernoulli or Binomial random variables.

A.13 Rayleigh, Rice and Hoyt Distributions

A.13.1 Rayleigh distribution

Following the definition in Beckmann (1964), a real random variable R follows a Rayleigh distribution with parameter s if the density with respect to the Lebesgue measure is given by

$$f_R(r) = \frac{2r}{s} e^{-\frac{r^2}{s}}\ ,\ r \geq 0. \tag{A.13.1}$$

A.13.2 Rice distribution

Following the definition in Sijbers et al. (1998), a real random variable R follows a Rice distribution with parameters a, s if the density with respect to the Lebesgue measure is given by

$$f_R(r) = \frac{r}{\sigma^2} I_0\left(\frac{ar}{\sigma^2}\right) \exp\left\{-\frac{r^2+a^2}{2\sigma^2}\right\}\ ,\ r \geq 0. \tag{A.13.2}$$

A.13.3 Hoyt distribution

Following the definition in Romero-Jerez and Lopez-Martinez (2017), a real random variable R^2 is said to follow a squared Hoyt distribution with parameters q and γ if the density with respect to the Lebesgue measure is given by

$$f_{R^2}(r) = \frac{1+q}{s\sqrt{q}\gamma} I_0\left(\frac{(1-q^2)r}{4q\gamma}\right) \exp\left\{-\frac{(1+q)^2 r}{4q\gamma}\right\}\ ,\ r \geq 0. \tag{A.13.3}$$

A.14 Cholesky Decomposition Algorithm

We present an algorithm that, given a positive definite Hermitian matrix $\mathbf{\Gamma}$ (such as a complex variance matrix) gives the factor $\mathbf{A}$ which is such that $\mathbf{A}\mathbf{A}^{\mathcal{H}} = \mathbf{\Gamma}$.

Algorithm 9 Cholesky factor of an Hermitian, positive-definite matrix

Require: $n \times n$ complex Hermitian matrix $\mathbf{\Gamma} = [\gamma_{k,l}]$
Ensure: $\mathbf{\Gamma}$ is positive-definite
  $\mathbf{A}_1 \leftarrow \sqrt{\gamma_{1,1}}$
  for $k = 2, \dots n$ **do**
    $\boldsymbol{\gamma}_k \leftarrow (\gamma_{1,k}, \dots, \gamma_{k-1,k})^\top$ ▷ the first $k-1$ entries of the k^{th} column of $\mathbf{\Gamma}$
    $\boldsymbol{a}_k^* \leftarrow \mathbf{A}_{k-1}^{-1}\boldsymbol{\gamma}_k$
    $a_{k,k} \leftarrow \sqrt{\gamma_{k,k} - \boldsymbol{a}_k^{\mathcal{H}}\boldsymbol{a}_k}$
    $L \leftarrow \begin{pmatrix} \mathbf{A}_{k-1} & \mathbf{0} \\ \boldsymbol{a}_k^\top & a_{k,k} \end{pmatrix}$
  end for
  return $\mathbf{A}_n$ lower triangular such that $\mathbf{A}_n\mathbf{A}_n^{\mathcal{H}} = \mathbf{\Gamma}$

Note that as the $\mathbb{C}\mathrm{ov}$ function is positive definite, so is a complex variance-covariance matrix.

A.15 Proof of the Alternate Definition of a Complex t-distribution

Proof of Proposition 5.5.2.

Proof. As Z is circular, and G is real, W is also circular. Thus, we can uniquely define the distribution of W by its modulus squared. Consider the distribution of

$$T = \left| \frac{Z}{\sqrt{\frac{G}{\nu}}} \right|^2 = \nu \frac{R^2}{G},$$

where $R^2 \sim \text{Exp}(1)$ is independent of $G \sim \text{Gamma}(\nu, 1)$. Using the real Mellin transforms, we have that

$$\mathcal{M}_{R^2}(s) = \int_0^\infty r^s e^{-r} \, \mathrm{d}r = \Gamma(s+1),$$

for all $s \notin \mathbb{Z}^- = \{-1, -2, \ldots\}$. Likewise, we have that

$$\mathcal{M}_{1/G}(s) = \mathcal{M}_G(-s) = \frac{\Gamma(\nu - s)}{\Gamma(\nu)}.$$

Hence, we have that

$$\begin{aligned} \mathcal{M}_T(s) &= \nu^s \mathcal{M}_{R^2}(s) \mathcal{M}_{1/G}(s) \\ &= \nu^s \frac{\Gamma(s+1)\Gamma(\nu - s)}{\Gamma(\nu)} \\ &= \nu^{s+1} B(s+1, \nu - s), \end{aligned}$$

where $B(\cdot, \cdot)$ is the beta function. Inverting this transform, we obtain

$$\begin{aligned} f_T(x) &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s-1} \nu^{s+1} \int_0^1 t^s (1-t)^{\nu - s - 1} \, \mathrm{d}t \, \mathrm{d}s \\ &= \frac{1}{2\pi i} \int_0^1 (1-t)^{\nu-1} \frac{\nu}{x} \int_{c-i\infty}^{c+i\infty} \left(\frac{\nu t}{x(1-t)} \right)^s \mathrm{d}s \, \mathrm{d}t. \end{aligned}$$

Now, $c = 0$ is a constant such that the integral converges. Performing the substitutions $iu = s$, $q = \ln(\nu/x)$ and $p = \ln((1-t)/t)$, and using the Dirac delta integral definition, we obtain that

$$f_T(x) = \left(1 + \frac{x}{\nu}\right)^{-(\nu+1)}, \; x \geq 0.$$

This is exactly Equation (5.5.11), which is the distribution of the modulus squared of the t-distribution as it was originally defined. Hence by circularity and equality in distribution of the modulus-squared,

$$\frac{Z}{\sqrt{\frac{G}{\nu}}} \sim \mathbb{C}t_\nu.$$

□

A.16 Proof of the Strong Law of Large Numbers

Proof of the strong law of large numbers, generalising the real proof from Billingsley (1995, p. 85).

Proof. Without loss of generality, we assume that the Z_k's have zero mean. We will also assume for this proof that the Z_k's have common finite variance $\mathbb{E}\left(|Z_k|^2\right) = \sigma^2$, pseudovariance $\mathbb{E}\left(Z_k^2\right) = \rho^2$ and kurtosis $\mathbb{E}\left(|Z_k|^4\right) = \kappa^4$. Let $S_n = \sum_{k=1}^n Z_k$. To prove convergence almost-surely, it suffices to prove that $\mathbb{P}\left(\frac{|S_n|}{n} \geq \varepsilon \text{ i.o.}\right) = 0$, for each $\varepsilon > 0$. Here, i.o. means infinitely often. Then, we have from independence and zero-mean that

$$\mathbb{E}\left(|S_n|^4\right) = \sum_{k=1}^n \mathbb{E}\left(|Z_k|^4\right) + \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}\left(|Z_k|^2|Z_l|^2\right) + \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}\left(Z_k^2 Z_l^2\right).$$

All other terms in the expansion involve a singular $\mathbb{E}\left(Z_k\right)$ which causes that term to vanish. Thus, we have

$$\mathbb{E}\left(|S_n|^4\right) = n\kappa^4 + 2n(n-1)\sigma^2 + n(n-1)|\rho|^4 \leq Kn^2,$$

for some constant $K \in \mathbb{R}$ independent of n. We then have by the Markov inequality (which we claim holds in the complex setting) that

$$\mathbb{P}\left(|S_n| \geq n\varepsilon\right) \leq \frac{Kn^2}{n^4\varepsilon^4} = \frac{K}{n^2\varepsilon^4}.$$

Then, by the (first) Borel-Cantelli lemma (which we claim also naturally extends to complex random variables), the above implies that

$$\mathbb{P}\left(\frac{|S_n|}{n} \geq \varepsilon \text{ i.o.}\right) = 0,$$

for each $\varepsilon > 0$. This completes the proof, however we note that further research may be conducted to relax some of the restrictive assumptions at the beginning of the proof. □

A.17 Proof of the Distribution of the Variance Estimator

Proof of Proposition 6.4.2.

Proof. Consider,

$$\begin{aligned}(n-1)\widehat{\sigma}^2 &= \sum_{k=1}^n |Z_k - \overline{Z}_n|^2 \\ &= \sum_{k=1}^n |Z_k - \mu|^2 - n|\overline{Z}_n - \mu|^2.\end{aligned}$$

Now as the Z_k's are i.i.d. from a $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, 0)$ distribution,

$$\sum_{k=1}^{n} |Z_k - \mu|^2 \sim \text{Gamma}(n, \sigma^2),$$
$$n|\overline{Z}_n - \mu|^2 \sim \text{Exp}(\sigma^2).$$

We now claim here without explicit proof (although it does hold, and a more rigorous argument can be seen in the context of a linear model in Appendix A.19 by considering just a linear model over an intercept), that the sample mean and variance estimator are independent. As a result, we see that this 'difference' of a Gamma and Exponential distributions can be verified (e.g. through use of characteristic functions) to result in a Gamma$(n-1, \sigma^2)$. Hence, scaling by σ^2 completes the proof. □

A.18 Proof of the Complex OLS Estimator

Verification of the complex OLS estimator. Firstly, we need some vector Wirtinger derivatives (Trampitsch (2013); Adali and Schreier (2014)). Namely, for a function $\boldsymbol{z} \in \mathbb{C}^n$ and $\mathbf{A} \in M_{nn}(\mathbb{C})$,

$$\frac{\partial \boldsymbol{z}^{\mathcal{H}}\mathbf{A}}{\partial \boldsymbol{z}^*} = \mathbf{A},$$
$$\frac{\partial \boldsymbol{z}^{\mathcal{H}}\mathbf{A}\boldsymbol{z}}{\partial \boldsymbol{z}^*} = \mathbf{A}\boldsymbol{z},$$
$$\frac{\partial \mathbf{A}\boldsymbol{z}}{\partial \boldsymbol{z}^*} = \mathbf{0}.$$

Thus, consider the sum of squared errors

$$\begin{aligned}\sum_{k=1}^{n} |\varepsilon_k|^2 &= \boldsymbol{\varepsilon}^{\mathcal{H}}\boldsymbol{\varepsilon} \\ &= (\boldsymbol{w} - \mathbf{Z}\tilde{\boldsymbol{b}})^{\mathcal{H}}(\boldsymbol{w} - \mathbf{Z}\tilde{\boldsymbol{b}}) \\ &= \boldsymbol{w}^{\mathcal{H}}\boldsymbol{w} - \tilde{\boldsymbol{b}}^{\mathcal{H}}\mathbf{Z}^{\mathcal{H}}\boldsymbol{w} - \boldsymbol{w}^{\mathcal{H}}\mathbf{Z}\tilde{\boldsymbol{b}} - \tilde{\boldsymbol{b}}^{\mathcal{H}}\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\tilde{\boldsymbol{b}}.\end{aligned}$$

Now, given this is a real-valued function, we can seek a maximum or minimum by differentiating with respect to $\tilde{\boldsymbol{b}}^*$ and setting the result equal to 0. We will not explicitly verify it here, but the stationary point in this case will be a minimum (check mixed derivative with respect to $\tilde{\boldsymbol{b}}$). Taking the derivative in $\tilde{\boldsymbol{b}}^*$ results in,

$$\frac{\partial \boldsymbol{\varepsilon}^{\mathcal{H}}\boldsymbol{\varepsilon}}{\partial \tilde{\boldsymbol{b}}^*} = \mathbf{Z}^{\mathcal{H}}\boldsymbol{w} - \mathbf{Z}\mathbf{Z}^{\mathcal{H}}\tilde{\boldsymbol{b}}.$$

Setting equal to zero yields

$$\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\boldsymbol{b} = \mathbf{Z}^{\mathcal{H}}\boldsymbol{w},$$
$$\Rightarrow \boldsymbol{b} = \left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathcal{H}}\boldsymbol{w},$$

assuming $\mathbf{Z}^{\mathcal{H}}\mathbf{Z}$ is invertible.

A.19 Proof of the Distribution of the Linear Model Variance Estimator

Proof of Lemma 7.3.2.

Proof.[1] First, consider the numerator

$$(n-p)\hat{\sigma}^2 = (\boldsymbol{w} - \mathbf{Z}\boldsymbol{b})^{\mathcal{H}}(\boldsymbol{w} - \mathbf{Z}\boldsymbol{b}).$$

Define now the $n \times n$ complex matrix $\mathbf{Q} = \mathbf{I}_n - \mathbf{Z}\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathcal{H}}$ so that

$$(n-p)\hat{\sigma}^2 = \|\mathbf{Q}\boldsymbol{w}\|^2.$$

Examining this matrix $\mathbf{Q}$, it can be seen that it is idempotent ($\mathbf{Q}^2 = \mathbf{Q}$), Hermitian and

$$\begin{aligned}
\operatorname{trace}(\mathbf{Q}) &= \operatorname{trace}(\mathbf{I}_n) - \operatorname{trace}\left(\mathbf{Z}\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathcal{H}}\right) \\
&= n - \operatorname{trace}\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\left(\mathbf{Z}^{\mathcal{H}}\mathbf{Z}\right)^{-1}\right) \\
&= n - \operatorname{trace}(\mathbf{I}_p) \\
&= n - p.
\end{aligned}$$

Being Hermitian means that $\mathbf{Q}$ must have real eigenvalues and is diagonalisable. Being idempotent means that the eigenvalues must either be 1 or 0. Finally, the trace of $n-p$, which is also the sum of the eigenvalues, indicates that 1 is an eigenvalue with algebraic multiplicity $n-p$. We can thus produce the eigendecomposition of the matrix $\mathbf{Q}$ in terms of a unitary matrix $\mathbf{U}$ and diagonal matrix $\mathbf{\Lambda}$, where

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{I}_{n-p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

and

$$\mathbf{U}^{\mathcal{H}}\mathbf{Q}\mathbf{U} = \mathbf{\Lambda}.$$

Returning to our variance estimator, by linearity we have that

$$\begin{aligned}
\mathbf{Q}\boldsymbol{w} &\sim \mathcal{N}_{\mathbb{C}^n}\left(\mathbf{0}, \sigma^2\mathbf{Q}\mathbf{Q}^{\mathcal{H}}, \mathbf{0}\right) \\
&\sim \mathcal{N}_{\mathbb{C}^n}\left(\mathbf{0}, \sigma^2\mathbf{Q}, \mathbf{0}\right).
\end{aligned}$$

[1] This proof and much of its notation is inspired by the approach taken to proving the real case by the user ocram on Cross Validated https://stats.stackexchange.com/questions/20227/why-is-rss-distributed-chi-square-times-n-p

Consider now,

$$\begin{aligned}\mathbf{U}^{\mathcal{H}}(\boldsymbol{w}-\mathbf{Z}\boldsymbol{b}) = \mathbf{U}^{\mathcal{H}}\mathbf{Q}\boldsymbol{w} &\sim \mathcal{N}_{\mathbb{C}^n}\left(\mathbf{0}, \sigma^2\mathbf{U}^{\mathcal{H}}\mathbf{Q}\mathbf{U}, \mathbf{0}\right)\\ &\sim \mathcal{N}_{\mathbb{C}^n}\left(\mathbf{0}, \sigma^2\boldsymbol{\Lambda}, \mathbf{0}\right).\end{aligned}$$

By property of the normal distribution with diagonal variance matrix, $\mathbf{U}^{\mathcal{H}}(\boldsymbol{w}-\mathbf{Z}\boldsymbol{\beta})$ is such that the first $n-p$ component vectors are independent $\mathcal{N}_{\mathbb{C}}\left(0, \sigma^2, 0\right)$ random variables, whilst the last p components are almost surely 0. Hence, the sum of the modulus squared of this random vector will yield the sum of $n-p$ independent $\text{Exp}(\sigma^2)$ random variables, so

$$\left\|\mathbf{U}^{\mathcal{H}}(\boldsymbol{w}-\mathbf{Z}\boldsymbol{b})\right\|^2 \sim \text{Gamma}(n-p, \sigma^2).$$

However, since $\mathbf{U}$ is unitary so $\mathbf{U}\mathbf{U}^{\mathcal{H}} = \mathbf{I}_n$,

$$\begin{aligned}\left\|\mathbf{U}^{\mathcal{H}}(\boldsymbol{w}-\mathbf{Z}\boldsymbol{b})\right\|^2 &= (\boldsymbol{w}-\mathbf{Z}\boldsymbol{b})^{\mathcal{H}}\mathbf{U}\mathbf{U}^{\mathcal{H}}(\boldsymbol{w}-\mathbf{Z}\boldsymbol{b})\\ &= (\boldsymbol{w}-\mathbf{Z}\boldsymbol{b})^{\mathcal{H}}(\boldsymbol{w}-\mathbf{Z}\boldsymbol{b})\\ &= (n-p)\hat{\sigma}^2.\end{aligned}$$

Thus by scaling the Gamma distribution, we obtain

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \text{Gamma}(n-p, 1).$$

Independence follows from the fact that $(\boldsymbol{w}-\mathbf{Z}\boldsymbol{b}) = \mathbf{Q}\boldsymbol{w}$ is independent of $\boldsymbol{b} = (\mathbf{I}_n-\mathbf{Q})\boldsymbol{w}$. This can be seen by the fact that, as $\boldsymbol{w}$ is normally distributed, $\mathbf{Q}\boldsymbol{w}$ and $(\mathbf{I}_n-\mathbf{Q})\boldsymbol{w}$ must also be jointly normally distributed as any linear combination of the components will be a linear combination of the components of $\boldsymbol{w}$ and therefore normally distributed. It thus suffices to see that

$$\begin{aligned}\mathbb{C}\text{ov}\left(\mathbf{Q}\boldsymbol{w}, (\mathbf{I}_n-\mathbf{Q})\boldsymbol{w}\right) &= \mathbf{Q}\sigma^2\mathbf{I}_n(\mathbf{I}_n-\mathbf{Q})^{\mathcal{H}}\\ &= \sigma^2(\mathbf{Q}-\mathbf{Q}\mathbf{Q}^{\mathcal{H}})\\ &= \sigma^2(\mathbf{Q}-\mathbf{Q}^2)\\ &= \mathbf{0},\end{aligned}$$

by the fact that $\mathbf{Q}$ is Hermitian and idempotent, and uncorrelated implies independence given we are in a multivariate complex normal setting. □

Appendix B

Plots and Code

B.1 Data for the Polar Histogram Comparison

Plots of the data for the histogram estimators displayed in the body.

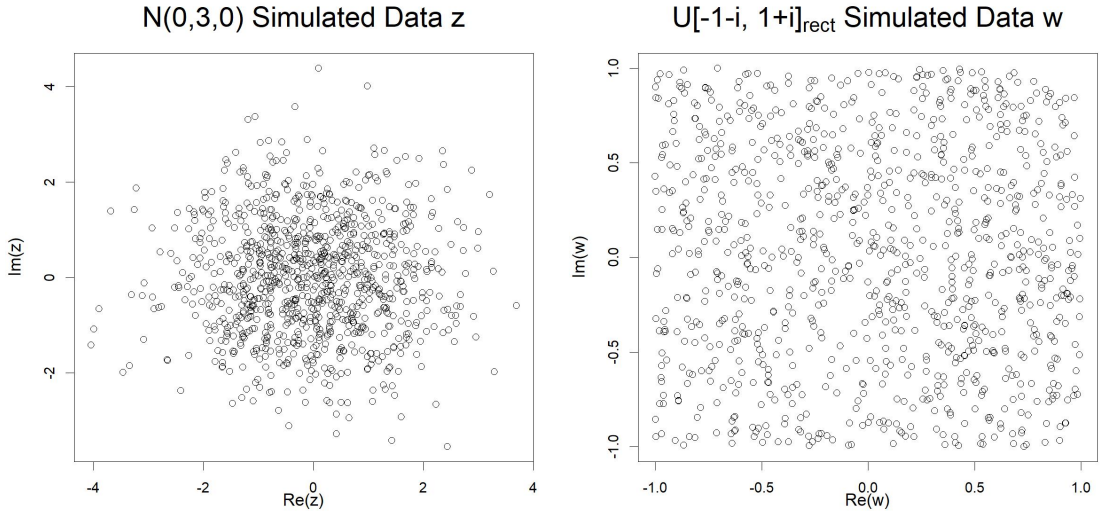

Figure B.1: Normal and uniform data ($n = 1000$) from a $\mathcal{N}_{\mathbb{C}}(0, 3, 0)$ and rectangular uniform distribution respectively.

B.2 R Code for Polar Histograms

```
i <- complex(imaginary = 1)
cnorm <- function(n, mu = 0, s2 = 1, rho2 = 0, polar = FALSE)
    {# Simulate complex normal distribution.
    i <- complex(imaginary = 1)
    rho2 <- as.complex(rho2)
    N011 <- rnorm(n, 0, 1)
    if(polar){
        R010 <- sqrt(rexp(n, 1))
        T010 <- runif(n, -pi,pi)
        N010 <- complex(modulus = R010, argument = T010)
    } else {
        X010 <- rnorm(n, 0, sqrt(1/2))
```

```
        Y010 <- rnorm(n, 0, sqrt(1/2))
        N010 <- complex(real = X010, imaginary = Y010)
    }
    W <- mu + sqrt(rho2)*N011 +
        sqrt(s2 - Mod(rho2) + 0*i)*N010 # +0*i to make complex
    if(Mod(rho2) > s2){warning("pseudocovariance of larger modulus than
    the variance")}
    return(W)
}

# Arc function to draw arcs of given radius from one angle to another ####
arc <- function(radius, argstart, argend, n = 500){
    argseq <- seq(from = argstart, to = argend, length.out = n)
    if(argseq[n] != argend){
        argseq <- c(argseq, argend)
    }
    z <- complex(length(argseq))
    for(ind in 1:length(argseq)){
        z[ind] <- complex(modulus = radius, argument = argseq[ind])
    }
    return(z)
}

# Add concentric circles to a plot ####
add_circle <- function(r, centre = 0, colr = "black"){
    colind <- 1
    colr <- c(colr[length(colr)], colr[-length(colr)])
    for(rad in r){
        lines(arc(rad, -pi, pi, n = 1000),
        col = colr[(colind%%length(colr)) + 1])
        colind <- colind + 1
    }
}

# Add rays from the origin to a plot ####
add_ray <- function(theta, centre = 0, maxi = 1e10, col = "black"){
    theta <- theta/pi
    theta <- pi*((theta+1)%%(2) - 1)
    for(tt in theta){
        lines(c(centre, centre + complex(modulus = maxi, argument = tt)))
        #lines(c(0,ifelse(abs(tt)>pi/2,-1,1)*maxi),
        centre + c(0,tan(tt)*maxi), col = col)
    }
}

# Complex polar Histogram function ####
```

```
polar_hist <- function(dat, recentre = FALSE,
                    radbreaks = length(dat)^(1/3),
                    radwidth = NULL, argbreaks = length(dat)^(1/3),
                    dens_shading = TRUE, opacity = 1, colr = "black",
                    contrast = 1, black_low = FALSE, pointcol = "black",
                    add = FALSE, tiff_out = FALSE, plot_range = NULL){
####
# dat is your complex data as a single complex vector or bivariate
matrix of real and imaginary components.
# recentre subtracts the mean from the data to centre it if TRUE
# radbreaks is how many breaks in the radius (how many radius bins).
# radwidth also allows you to fix a specific width of radius increment
if desired - leave null otherwise.
# argbreaks is how many argument breaks (bins).
# dens_shading is a logical if you want to see the density shaded
over the data.
# opacity is the opacity of the shading in [0,1] - set equal to 1
to just see the histogram.
# colr gives the colour of the shading. Choice of:
c("black", "red", "blue", "green", "yellow", "cyan", "magenta").
# contrast dictates the contrast between low and high density regions.
Set closer to zero (< 1) for more contrast (inverted if black_low = TRUE).
# black_low is a logical, if true then the empty region colour is black
rather than white. Can be good with high opacity and lower
contrast (e.g. 0.5).
# pointcol gives the colour of the data points plotted (not visible
if opacity is 1, but useful if overlaying the plot)
# add will add the histogram to the existing plot window.
# tiff_out is ignored for now, but the idea is that it will write
an image rather than plot the output.
####
    if(!is.complex(dat)){
        if(ncol(dat) == 2){
            dat <- complex(real = dat[,1], imaginary = dat[,2])
        } else{
            stop("did not receive complex or bivariate data")
        }
    }
    if(recentre){
        dat <- dat - mean(dat)
    }

    n <- length(dat)

    if(is.null(plot_range)){
        max_frame <- 1.2*max(max(Mod(dat)))
    } else {
```

```
    max_frame <- plot_range
}

if(add){
    points(dat, pch = 19, cex = 0.5, col = pointcol)
} else{
    plot(dat, ylim = c(-max_frame, max_frame),
    xlim = c(-max_frame, max_frame), pch = 19, cex = 0.5,
    col = pointcol)
}
maxrad <- 1.1*max(Mod(dat))

radbreaks <- floor(radbreaks)
if(is.null(radwidth)){
    radwidth <- maxrad/radbreaks
}
rads <- seq(from = radwidth, by = radwidth, length.out = radbreaks)
add_circle(rads)

argbreaks <- floor(argbreaks)
argwidth <- 2*pi/argbreaks
argorigin <- -pi # initial argument for first ray
                  (lowest argument value of interval)
argss <- seq(from = argorigin, by = argwidth, length.out = argbreaks)
add_ray(argss, col = "black", maxi = maxrad)

counts <- matrix(nrow = radbreaks, ncol = argbreaks)
dens <- matrix(nrow = radbreaks, ncol = argbreaks)
rads.ext <- c(0,rads)
argss.ext <- c(argss, argss[1]+2*pi)
h.vec <- numeric(nrow(counts))
for(i in 1:nrow(counts)){
    h <- argwidth/(2) * (rads.ext[i+1]^2 - rads.ext[i]^2)
    h.vec[i] <- h
    for(j in 1:ncol(counts)){
        counts[i,j] <- sum( (Mod(dat) >= rads.ext[i] & Mod(dat) <
        rads.ext[i+1]) & (Arg(dat) >= argss.ext[j] & Arg(dat) <
        argss.ext[j+1]) )
        dens[i,j] <- counts[i,j] / (n*h)
    }
}
counts
sum(counts)
sum(dens)
sum(apply(dens, 1, sum)*h.vec)
if(dens_shading){
    for(i in 1:nrow(dens)){
```

```
            for(j in 1:ncol(dens)){
                if(black_low){
                    anncolour <- rgb(red = ifelse(colr %in% c("magenta",
                    "yellow", "red", "black"),
                    (dens[i,j]/max(dens))^contrast,0),
                    blue = ifelse(colr %in% c("magenta", "cyan",
                    "blue", "black"),(dens[i,j]/max(dens))^contrast,0)
                    , green = ifelse(colr %in% c("cyan", "yellow", "green",
                    "black"), (dens[i,j]/max(dens))^contrast,0),
                    alpha = opacity)
                }else{
                    anncolour <- rgb(red = ifelse(colr %in% c("magenta",
                    "yellow", "red"),1,
                    (1 - dens[i,j]/max(dens))^contrast),
                    blue = ifelse(colr %in% c("magenta", "cyan", "blue"),
                    1, (1 - dens[i,j]/max(dens))^contrast)
                    , green = ifelse(colr %in% c("cyan", "yellow", "green"),
                    1, (1 - dens[i,j]/max(dens))^contrast), alpha = opacity)
                }
                r1 <- rads.ext[i]
                r2 <- rads.ext[i+1]
                ang1 <- argss.ext[j]
                ang2 <- argss.ext[j+1]
                polygon(c(arc(r1, ang1, ang2), arc(r2, ang2, ang1) ),
                col = anncolour )
            }
        }

        legend("topright", legend = c(paste("Max density:",
               signif(max(dens),3)), paste("Modulus bins:",
               radbreaks), paste("Argument bins:", argbreaks)))
    }
}

library(squash)

n <- 1000
z <- cnorm(n, 0,3,0)
plot(z)
polar_hist(z, black_low = TRUE, colr = "cyan", opacity = 1)
hist2(cbind(Re(z),Im(z)))

x <- runif(n,-1,1)
y <- runif(n,-1,1)
w <- complex(real = x, imaginary = y)

par(mfrow = c(1,3))
```

```
plot(z)
polar_hist(z, black_low = TRUE, colr = "cyan", opacity = 1)
hist2(cbind(Re(z),Im(z)))

par(mfrow = c(1,3))
plot(w)
polar_hist(w, black_low = TRUE, colr = "cyan", opacity = 1)
hist2(cbind(Re(w),Im(w)))
par(mfrow = c(1,1))
```

B.3 R Code for Example 6.4.4

```
n <- 10000 # make this divisible by 4 to get exact axis distances.
u <- seq(from = -pi, to = pi, by = 2*pi/n)
z <- exp(i*u) # unit circle
s2 <- 2 # sigma^2
r2 <- (2-i)/2 # rho^2. Ensure Mod(r2) < s2
alpha <- 0.05 # alpha-level of interval
mu <- 1+i
a <- (sqrt(s2 + Mod(r2)) + sqrt(s2-Mod(r2)))/2 * exp(i*Arg(r2)/2)
# coefficients as per Ollila2011 simulation.
b <- (sqrt(s2 + Mod(r2)) - sqrt(s2-Mod(r2)))/2 * exp(i*Arg(r2)/2)

K <- sqrt(-log(alpha)) # alternatively sqrt(qexp(1-alpha, 1))
K50 <- sqrt(-log(0.5)) # alternatively sqrt(qexp(1 - 0.5, 1))

ell <- mu + K*a*z + K*b*Conj(z)
ell50 <- mu + K50*a*z + K50*b*Conj(z)

plot(ell, type = "l", col = "blue", xlim = c(-2,4),
ylim = c(-2,4), lwd = 3, main = "N(1+i, 2, 1 - 0.5i)",
xlab = "Re", ylab = "Im")
abline(v=0)
abline(h =0)
lines(ell50, lwd = 3, col = "red")
lines(ell, lwd = 3, col = "blue")
points(1+i, pch = 15, col = "magenta")
legend("topright", pch = c(15,NA,NA), lty = c(0,1,1), lwd = 3,
c("magenta", "red", "blue"), legend = c("1+i", "50%", "95%"))

# Test probability

ell.flat <- (ell - mu)*exp(-i*Arg(r2)/2) # rotate the
ellipse to have axes on coordinate axes

w <- cnorm(n, mu, s2, r2) # test data
w.flat <- (w-mu)*exp(-i*Arg(r2)/2) # rotate the data
```

```
ax1 <- Mod(ell.flat[1])
ax2 <- Mod(ell.flat[n/4])

sum( (Re(w.flat)/ax1)^2 + (Im(w.flat)/ax2)^2 <= 1 )/n
# proportion of data points in the ellipse.

ind <- which((Re(w.flat)/ax1)^2 + (Im(w.flat)/ax2)^2 <= 1 )
plot(w) # plot the data
points(w[ind], col = "gold", pch = 19) # colour the points inside ellipse
points(w[-ind], col = "red", pch = 19) # colour the points outside ellipse
lines(ell, col = "blue", lwd = 3) # include the ellipse
```