



UCL

COSMOLOGY IN THE PRESENCE OF NON-GAUSSIANITY

ROBERT LUDWIG SCHUHMAN

*Thesis presented for the Degree of
Doctor of Philosophy*

Department of Physics and Astronomy
Faculty of Mathematical and Physical Sciences
University College London

April 4, 2017

Robert Ludwig Schuhmann: *Cosmology in the Presence of Non-Gaussianity*

© April 4, 2017

SUPERVISORS:

Dr Benjamin Joachimi

Prof Hiranya V. Peiris

EXAMINERS:

Prof Sarah Bridle

Prof Ofer Lahav

DECLARATION

I, Robert Ludwig Schuhmann, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

London, April 4, 2017

Robert Ludwig Schuhmann

ABSTRACT

Modern observational cosmology relies on statistical inference, which models measurable quantities (including their systematic and statistical uncertainties) as random variates, examples are model parameters (‘cosmological parameters’) to be estimated via regression, as well as the observable data itself. In various contexts, these exhibit non-Gaussian distribution properties, e. g., the Bayesian joint posterior distribution of cosmological parameters from different data sets, or the random fields affected by late-time nonlinear structure formation like the convergence of weak gravitational lensing or the galaxy density contrast. Gaussianisation provides us with a powerful toolbox to model this non-Gaussian structure: a non-linear transformation from the original non-Gaussian random variate to an auxiliary random variate with (approximately) Gaussian distribution allows one to capture the full distribution structure in the first and second moments of the auxiliary.

We consider parametric families of non-linear transformations, in particular Box-Cox transformations and generalisations thereof. We develop a framework that allows us to choose the optimally-Gaussianising transformation by optimising a loss function, and propose methods to assess the quality of the optimal transform a posteriori.

First, we apply our maximum-likelihood framework to the posterior distribution of *Planck* data, and demonstrate how to reproduce the contours of credible regions without bias – our method significantly outperforms the current gold standard, kernel density estimation. Next, we use Gaussianisation to compute the model evidence for a combination of CFHTLenS and BOSS data, and compare to standard techniques. Third, we find Gaussianising transformations for simulated weak lensing convergence maps. This increases the information content accessible to two-point statistics (e. g. the power spectrum) and potentially allows for rapid production of independent mock maps with non-Gaussian correlation structure.

With these examples, we demonstrate how Gaussianisation expands our current inference toolbox, and permits us to accurately extract information from non-Gaussian contexts.

ACKNOWLEDGEMENTS

The following humans have been protecting the last glowing embers of my sanity over the last weeks, feeding and sheltering them from the storm; to my greatest regret I will never get to know any of them in person: Johann Sebastian Bach, Hanns Dieter Hüsch, Philip Glass, Georg Friedrich Händel, Claude Debussy, Carlo Gesualdo, Matthew Bellamy, and (of course) Douglas Adams.

Infinite thanks to my supervisor, Dr Benjamin Joachimi. Your mentorship has been invaluable – your careful, creative mind and your intuition balancing between theoretical and pragmatic have always been inspiring, and you have taught me a lot. Thank you for all of your guidance and support, for your smart and determined encouragements, and for putting up with all of my head-in-the-clouds antics. You have worked ceaselessly to pave my path, and for that I will be forever grateful!

Thanks also to my second supervisor, Prof Hiranya Peiris. I admire you for your endless drive, your dedication, and your brave and omnivorous mind. My gratitude for the wisdom you have shared and for the immense trust you continue to put into me – I will do my best to live up to it!

Thanks to my examiners, Prof Ofer Lahav and Prof Sarah Bridle, for spending the time and effort to make sense of my compiled scribblings.

Thanks to the supervisors of my Bachelor's and Master's theses, Prof Jochen Weller and Prof Niayesh Afshordi, for shaping my scientific path and giving tons of indispensable advice. I owe both of you a lot!

Thanks to Prof Harald Lesch for keeping the spark alive in times of misery, and for his endless passion for passing on the torch. Thanks to Prof Detlef Dürr for teaching me the importance of truthful thinking, and the *crystal clear* essence of things.

Many many thanks to all my colleagues and friends in and around London, too many to list them all! – if you, dear reader, feel that you should be mentioned here, this is the place where you are. You all have made the last three years worthwhile.

A couple of random personal shoutouts, increasingly cryptic:

Johannes Kleiner and the QH-consortium: without *Sý!* all of this would be worthless.

The PSI year of 2012: This year has been a total game changer in my life – I can honestly say that without it I would not be where I am today. Being one of 37 strangers from all continents that morph into a sworn conspiratory league like ours has been a marvellous and unique experience that I will always hold close to my heart (or my left shoulder, for that sake).

Especially boon and bunking companions Kate, Brenda, and Grisha: for an amazing time and alliance steering through so many joys, hardships, and beeping noises!

Also especially Lucas Hackl: *Things have the fortunate tendency to work themselves out.*

Also especially Katie Auchettl: I am so looking forward to standing by your side and watching the world burn together!

Also especially Mark Penney: For communism, cervisia, and C^* -algebras. And a certain pact of ours that shall not be forgot.

Cat: *WAKE UP! GRBAPNFRBLKMHBMAKEUP!*

Beaver: You actually saved my ass, do you know that?

Panther: Where to begin? This city has already become less weird without you.

Piggie: *Don't help them to bury the light!*

Bruno Moraes: It is people like you who keep the ship afloat. Not only because you have chocolate.

Michael McLeod: For things that have been, things that have not, a completely ruined liver, and *something resembling sanity.*

Dominik Thalmeier: Für eine Freundschaft, die immer ein Refugium war. Ein Raum, in dem wir miteinander ehrlich sein konnten und frei und versponnen und unangepasst und uferlos und sarkastisch und tief und durchgeknallt und fasziniert und ungefiltert und noch so viel mehr.

Regina Galambos: Für Gespräche, die nach zwölfteinhalb Stunden erst richtig anfangen. Stickstoff-Eis bei vierzig Grad. Schattenwald-Bohnen. Lästern mit Steckerlfisch am Lerchenauer See. Die Blindenschriftkröte. Brontal zerkatert zum Gegenanwählen rennen. Satanisch kochen: zum Beispiel lelótt kertitörpe. Eichhörnchen Eichhörnchen. Janáček in der Geschlossenen. Radikale Tourismusbekämpfung. Strauße mit Senf einmassieren. Dreiecksfliegen. Kokosfett-Aktionen. Panische Emails. Gemeinsam seltsame Menschen verwirren. Irgendwann doch noch das Ötz niederbrennen. Und Sanskrit lernen. Leckere Sachen aus Ländern mit lauter Verrückten drin.

Bücher, Wein und Koffein – die relevanten Dinge halt.

Urmel: Für urmelöse Urmeligkeit in urmeligster Umurmelung.

Karl: Ἄνδρα μοι ἔννεπε, Μοῦσα, πολύτροπον, ὃς μάλα πολλὰ πλάγχθη ... Du hast es nie erfahren, aber du hast mir beigebracht, mit Wahrhaftigkeit meinen eigenen Weg zu gehen, und worin Seelenruhe tatsächlich zu finden ist.

Ingeborg & Alice: Die Liebe, die wir unseren Mitmenschen geben, ist letztlich doch das einzig Wichtige.

Sabine & Georg: Denn was wäre ich ohne euch? Gar nicht da, das wäre ich ohne euch.

Das Arschkind: ÜRSEGESPRÜH!

CONTENTS

I	INTRODUCTION	12
0	ROADMAP	13
1	COSMOLOGY	15
1.1	At order zero: Friedmann-Lemaître-Robertson-Walker cosmology	15
1.2	At order one: the Cosmic Microwave Background	24
1.3	At higher order: Evolution of Structure	28
1.3.1	Weak Gravitational Lensing	30
1.3.2	Baryon Acoustic Oscillations	35
2	STATISTICAL INFERENCE FROM DATA	40
2.1	Bayesian Inference	40
2.1.1	Parameter Estimation	43
2.1.2	Model Selection	45
2.2	Sampling	47
2.2.1	Markov Chain Monte Carlo (MCMC)	48
2.2.2	Population Monte Carlo (PMC)	50
2.3	Non-Gaussian Random Variables	52
2.3.1	Classical Gaussianity Tests	53
2.3.2	Random Fields	54
2.3.3	Transforming to Gaussianity	58
II	GAUSSIANISATION	64
3	MODELLING OF POSTERIOR DENSITIES	65
3.1	Introduction	65
3.2	Gaussianisation	67
3.2.1	Finding the optimal transformation	68
3.2.2	Box–Cox transformations and their kin	69
3.2.3	Verifying the optimal transformation	70
3.2.4	Multi-pass transformations	72
3.3	A toy example	73
3.4	Performance results: Planck data	75
4	A NOVEL WAY TO COMPUTE EVIDENCE	81
4.1	Analytic formulae for evidence	81
4.2	Application to lognormal simulated data	82
4.3	Application to cosmological data	83

5	GAUSSIANISATION OF CONVERGENCE FIELDS	87
5.1	Introduction	87
5.2	Simulated convergence maps	89
5.3	The strategy: parametric transformations and loss function	93
5.3.1	Projection-based Gaussianity test	94
5.3.2	Likelihood-based Gaussianity test	97
5.4	Results	101
5.4.1	Gaussianising transformations from N -body simulations	101
5.4.2	Fast generation of weak lensing covariances	103
6	CONCLUSIONS AND OUTLOOK	108
6.1	Finite dimensions: posterior density modelling	108
6.1.1	Summary	108
6.1.2	Future work	109
6.2	Infinite dimensions: Gaussianisation of weak lensing convergence	110
6.2.1	Summary	110
6.2.2	Future work	111
III	APPENDIX	113
A	UNBOXING TRANSFORMATIONS	114
B	EVIDENCE REGRESSION	116
	BIBLIOGRAPHY	118

LIST OF FIGURES

Figure 1	Carl Friedrich Gauß	14
Figure 2	CMB temperature maps	25
Figure 3	CMB power spectra from <i>Planck</i>	26
Figure 4	linear dark matter power spectra	29
Figure 5	geometry of a thin lense	30
Figure 6	distortion components of weak lensing	31
Figure 7	galaxy images - systematics	34
Figure 8	<i>z</i> dFGRS galaxies	35
Figure 9	physical evolution of BAOs	37
Figure 10	BAO peak in BOSS data	38
Figure 11	CFHTLenS posterior density	44
Figure 12	toy model: probability densities	73
Figure 13	toy model: profile log-likelihood	74
Figure 14	toy model: contours – Gaussianisation vs KDE	75
Figure 15	<i>Planck</i> data: triangular non-Gaussian feature	76
Figure 16	<i>Planck</i> data: prominent degeneracy	76
Figure 17	<i>Planck</i> data: wall-like non-Gaussian feature (two-pass)	78
Figure 18	<i>Planck</i> data: wall-like non-Gaussian feature (one-pass)	78
Figure 19	<i>Planck</i> data: 7D marginal probability density	80
Figure 20	evidence computation on log-normal mock data set	83
Figure 21	histogram of simulated κ maps	90
Figure 22	power spectra of simulated κ maps	91
Figure 23	correlation matrices of power spectrum estimates	92
Figure 24	beta distribution	95
Figure 25	projection-based Gaussianity test	96
Figure 26	log-likelihoods on Gaussian random fields	99
Figure 27	transformed convergence histograms	100
Figure 28	correlation matrices of Box-Cox transformed spectra	102
Figure 29	power spectra of transformed mocks	104
Figure 30	Corr + Cov of transformed mocks – smoothing	105
Figure 31	Corr + Cov of transformed mocks – no smoothing	106

LIST OF TABLES

Table 1	Jeffreys' scale	46
Table 2	toy model: optimally Gaussianising parameters	74
Table 3	evidence computation on weak lensing + BAO data	85
Table 4	convergence: optimal transformations	101

ACRONYMS

2dFGRS	2-degree field galaxy redshift survey
6dFGRS	6-degree field galaxy redshift survey
ABC	Arsinh-Box-Cox
BAO	Baryon Acoustic Oscillations
BOSS	Baryon Oscillation Spectroscopic Survey
CC	Cross-Contour
CFHTLenS	Canada-France-Hawaii Telescope Lensing Survey
CMB	Cosmic Microwave Background
DES	Dark Energy Survey
FLRW	Friedmann-Lemaître-Robertson-Walker
GCH	Gaussian Copula Hypothesis
iid	independent and identically distributed
KDE	Kernel Density Estimation
ΛCDM	Λ - Cold Dark Matter
LSST	Large Synoptic Survey Telescope
KiDS	Kilo-Degree Survey
MCMC	Markov Chain Monte Carlo
MP	Model Parameter
PCA	Principal Component Analysis
PMC	Population Monte Carlo
SDDR	Savage-Dickey Density Ratio
SDSS	Sloan Digital Sky Survey
TP	Transformation Parameter
WMAP	Wilkinson Microwave Anisotropy Probe

Part I

INTRODUCTION



ROADMAP

Dirk had a tremendous propensity for getting lost when driving. This was largely because of his method of “Zen” navigation, which was simply to find any car that looked as if it knew where it was going and follow it. The results were more often surprising than successful, but he felt it was worth it for the sake of the few occasions when it was both.

*Douglas Adams
The Long Dark Tea-Time of the Soul*

During most of the 20th century, observational cosmology was dismissed as “the search for two numbers” (Sandage, 1970), referring to the instantaneous expansion speed of the Universe, and its instantaneous rate of change. This perception has radically altered within the last twenty years, with the advent of large data sets from cosmic microwave background experiments, large-scale galaxy surveys, radio interferometry and many others. We now find ourselves facing the opposite challenge – rather than experiencing a paucity of measurements, we now need to understand the statistical properties of data sets of unprecedented volume, depth, and complexity, so that we may accurately extract the information stored in them. Simultaneously, we have extended our scientific scope and focus on increasingly more refined questions: about the nature of dark matter and dark energy, the formation of the cosmic large-scale structures we observe today, and the physical nature of inflation.

The modern information-theoretical foundation of inference treats observable quantities – cosmological parameters as well as data – as random variables within probability theory. The modelling of their distributions includes all uncertainties which introduce scatter into their values, systematic biases as well as random noise. A commonplace model for random variates is the Gaussian distribution: it is purely characterised through its location and its width, and arises naturally in many contexts. Therefore it is often taken as a first assumption when modelling a random quantity; reporting measured values of parameters as mean and error bar is testament to this. However, often this assumption is not valid, and a Gaussian model does not adequately describe the observable in question – non-linear physics can be one reason of many. The results of this

mis-modelling can lead to loss of information which is stored in the non-Gaussian higher orders of the distribution, or even biases. We study the idea of Gaussianisation, to relax the assumption of Gaussianity, and to respect the non-Gaussian correlation structure of cosmological observables: mapping a non-Gaussian variable to an (approximately) Gaussian one with a suitably chosen transformation, which should be one-to-one and smooth. By recording the moments of this Gaussian distribution, we have found a model for the original (untransformed) distribution that adequately respects its non-Gaussian structure – and as long as our transformations have an analytic form, this automatically gives us explicit expressions for this distribution. In cosmology, this can be employed to two very different context in which probabilistic models with potentially non-Gaussian features can be encountered: the estimation of cosmological model parameters, as well as the observable data itself. Thus, Gaussianisation facilitates reporting a probability distribution as the outcome of an experiment, and hands us a powerful toolbox which we will present within this dissertation.

The thesis will proceed as follows: in [Chapter 1](#) we will lay the foundations of modern cosmology, and the phenomena and data sets we will use. [Chapter 2](#) introduces the apparatus of Bayesian inference, which is the predominant school of thought for cosmological data analysis. We specifically emphasize non-Gaussian distributions and the tools currently available for their treatment. In [Chapter 3](#) the main apparatus will be introduced: how to find a Gaussianising transformation, how to assess its quality in reproducing contours, and which parametric transformation families we employ. This is applied to cosmological parameters and their joint posterior distribution from *Planck* data; we demonstrate its performance on various shapes of non-Gaussianity. [Chapter 4](#) is explicitly demonstrating of the usefulness of Gaussianisation for the computation of the Bayesian evidence, a quantity crucial for model comparison; some of the algebraic fingerprint of [Chapter 3](#) and [Chapter 4](#) can be found in [Appendix A](#) and [Appendix B](#). [Chapter 5](#) details the application of Gaussianisation to weak lensing convergence –



Figure 1: Carl Friedrich Gauß together with the eponymous probability density, on an expired German banknote. From <http://bit.ly/2e106YY>.

a random field known to harbour a rich non-Gaussian correlation structure. We detail how to implement Gaussianisation for a random field, and demonstrate this with simulated convergence maps. This transformation can in principle be used to efficiently draw independent samples from the random field distribution. Finally, [Chapter 6](#) offers conclusions and lists future directions that open up for further study. [Chapter 3](#) and [Chapter 4](#) are based on Schuhmann, Joachimi, and Peiris (2016).

COSMOLOGY

In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move.

*Douglas Adams
The Restaurant at the End of the Universe*

One of the most fascinating and revolutionary scientific discoveries of the twentieth century is the quantitative physical description of the observable Universe as a whole. Ideas, myths and reflections on origin, evolution, and fate of the Cosmos and all objects therein have been around since the dawn of the human mind; for millennia they have inspired us as individuals and shaped our cultures. However, only within the last century has our collective inquisitive pursuit solidified these ruminations into a part of quantitative science and connected them to the fundamental theories of modern physics. We saw the unfolding of a reliable, coherent narrative stretching from the beginning of time over the formation of the astrophysical large-scale structures we observe until the present day – even allowing us to extrapolate into the future and fathom the eventual fate of the Universe.

Modern cosmology, the science that emerged from this quest, is far from a completed discipline: despite its towering successes, many questions remain open, as to the formation of galaxies, or the nature of dark matter and dark energy. In contrast to the more theoretically motivated approach of the first 100 years, 21st century cosmology is driven by data sets of unprecedented size and faces new statistical challenges for precise and accurate inference. Whereas cosmologists of the 20th century worked at the interface between astronomy, gravitational physics and particle physics, in the new era we find ourselves bridging astrophysics, computer science and data science.

1.1 AT ORDER ZERO: FRIEDMANN-LEMAÎTRE-ROBERTSON-WALKER COSMOLOGY

For modern cosmology, the observable Universe is not merely a collection of interacting massive bodies; it is the background spacetime itself which becomes a dynamical

physical object. The gravitational dynamics of spacetime is described by the theory of general relativity, originally formulated by Einstein (1916) – modern accounts of this bedrock of modern physics can be found in S. Weinberg (1972), Hawking and G. F. Ellis (1973), Misner, Thorne, and Wheeler (1973), Wald (1984), and Oloff (1999). To mention but the salient facts of this theory: spacetime is a four-dimensional pseudo-Riemannian orientable differential manifold whose metric tensor has Lorentzian signature. The metric tensor $g_{\mu\nu}$ obeys the Einstein field equation, a non-linear partial differential equation of second order which determines its future time evolution:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (1)$$

where $R_{\mu\nu}$ and R are Ricci tensor and Ricci scalar of the metric $g_{\mu\nu}$; and $T_{\mu\nu}$ is the energy-momentum density tensor of all matter on the manifold. We are using time and space units in which the speed of light $c = 1$. G and Λ are Newton's and the cosmological constant, respectively. Free-falling matter particles move along timelike or lightlike geodesics $X^\alpha(\tau)$, which satisfy the equation

$$\frac{d^2 X^\alpha}{d\tau^2} + \Gamma_{\rho\sigma}^\alpha [X^\mu(\tau)] \frac{dX^\rho}{d\tau} \frac{dX^\sigma}{d\tau} = 0. \quad (2)$$

Here, $\Gamma_{\rho\sigma}^\alpha$ is the unique torsion-free Levi-Civita connection which is compatible with $g_{\mu\nu}$; τ denotes the affine coordinate parametrising the trajectory – this coincides with eigen-time for timelike geodesics. Misner, Thorne, and Wheeler (1973) subsumed Equation 2 and Equation 1 into the phrase:

“Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve.”

To characterise our observable Universe over all alternative spacetimes, we take two assumptions as axioms:

- Weyl's Postulate (Weyl, 1923): The Universe is filled by a congruence of non-intersecting timelike geodesics with tangent vectors u^μ (*observers*). Further, the spacetime can be foliated into a family of spatial hypersurfaces that are orthogonal to the geodesic congruence at each point.
- Cosmological Principle: At a particular time, all points in space are physically equivalent (*spatial homogeneity*), and so are all directions (*spatial isotropy*).

The first assumption merely asserts that there is a global time coordinate – *cosmic time* – and allows us to talk about a three-dimensional space at each instant. The second one states that the curvature of each of these three-dimensional submanifolds needs to be spatially constant (though they may vary in time). These manifolds of constant curvature

have the largest possible symmetry group, and are called Einstein spaces.

The Cosmological Principle needs to be taken with a grain of salt, because for our Universe it is only valid on certain scales. The existence of ourselves and of the structures directly around us are witness to the fact that today the real Universe is very inhomogeneous indeed – only above a certain length scale, often cited to be of order ~ 300 Mpc, does the matter distribution look homogeneous. Further, the patch of the Universe we can actually observe is of finite volume, limited in the past by the Big Bang, hence there is *a priori* no way for us to test for homogeneity on scales much larger. Nevertheless, phenomena like the cosmic microwave background (CMB; see Section 1.2) and galaxy clustering (see Section 1.3) indicate that the Cosmological Principle is indeed a valid and reliable assumption on a range of scales of cosmological interest (Wu, Lahav, and Rees, 1999).

Friedmann (1922, 1924) first derived an explicit form for all metric tensors satisfying these two conditions; in the subsequent fifteen years several authors rediscovered and refined his analysis (Lemaître, 1931; Robertson, 1935, 1936a,b; Walker, 1937) – hence these metrics are known as Friedmann-Lemaître-Robertson-Walker (FLRW) cosmologies. For a more comprehensive introduction, see Dodelson (2003), Mukhanov (2005), and S. Weinberg (2008). With time coordinate t and spherical coordinates (r, ϑ, ϕ) , this metric is

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\Omega \right]. \quad (3)$$

Here, $d\Omega = d\vartheta^2 + \sin^2(\vartheta)d\phi^2$ is the standard volume element on the unit 2-sphere. K is a constant of dimension of inverse area, and is related to the curvature scalar of the corresponding three-dimensional Einstein space via ${}^{(3)}R = 6K$. Note that this is not the full Ricci scalar of the Lorentzian 4D spacetime manifold, but the Ricci scalar of the induced metric on the 3D spacelike hypersurfaces at constant cosmic time, i. e., the Einstein spaces, whose metric is described by the [...] term in Equation 3. Often, this constant is split via $K = k/r_K^2$ into a *curvature radius* $r_K > 0$ and a unitless constant $k = -1, 0, +1$, which classifies the Einstein spaces into hyperbolic, flat, or spherical geometries. A coordinate transformation from (t, r) to (η, χ) is often used, where the *conformal time* $\eta(t)$ is defined via

$$\eta = \int_0^t \frac{dt'}{a(t')} \quad (4)$$

and the *conformal radius* $\chi(r)$ as

$$\chi = r_K S_k^{-1}(r/r_K); \quad S_k(x) := \begin{cases} \sin(x) & (k = +1) \\ x & (k = 0) \\ \sinh(x) & (k = -1). \end{cases} \quad (5)$$

If we measure these coordinates in units of the curvature radius, i. e. $\eta = r_K \tilde{\eta}$, $\chi = r_K \tilde{\chi}$, the line element [Equation 3](#) can be recast as

$$ds^2 = a^2(\eta) r_K^2 [d\tilde{\eta}^2 - d\tilde{\chi}^2 - S_k^2(\tilde{\chi}) d\Omega]. \quad (6)$$

This makes the geometric classification into spherical/flat/hyperbolic completely lucid, since this line element is conformally equivalent to the time axis \times the unit 3-sphere / flat \mathbb{R}^3 / the unit 3-antisphere. The unitless scale factor $a(t)$ is the only degree of freedom left in the metric tensor: it is a time-dependent scaling of the curvature radius, and thus contains the full information about the history of expansion (or contraction). Often, it is normalised such that it reaches unity at the current time. To establish its equations of motion, we need to insert [Equation 3](#) into [Equation 1](#), and impose the Cosmological Principle once more: it constrains the energy-momentum tensor to have the shape of a perfect fluid: $T^\mu{}_\nu = (\rho + p) u^\mu u_\nu - p \delta^\mu_\nu = \text{diag}[\rho, -p, -p, -p]^\mu{}_\nu$. The functions ρ and p are the energy density and pressure of the full matter content, and depend on time only. The resulting differential equations for $a(t)$,

$$H^2(t) = \left[\frac{\dot{a}(t)}{a(t)} \right]^2 = \frac{8\pi G}{3} \rho(t) - \frac{K}{a^2} + \frac{\Lambda}{3} \quad (7)$$

$$-H^2 q(t) = \frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3} [\rho(t) + 3p(t)] + \frac{\Lambda}{3}, \quad (8)$$

are known as the Friedmann equations (Friedmann, 1922). The values of the Hubble parameter $H(t) = \dot{a}/a$ and the deceleration parameter $q(t) = -\ddot{a}/\dot{a}^2$ determine whether a [FLRW](#) model is expanding or contracting, and whether said evolution speeds up or slows down. Only a careful balance between the terms in [Equation 7](#) by artificially fine-tuning Λ to a fixed negative value would allow for a static universe. In fact, this led Einstein (1917) to introduce Λ into [Equation 1](#) in the first place – later, Hubble (1929) extended earlier measurements by Slipher (1917) of the recession velocities of galaxies, and discovered that indeed $H > 0$, our Universe does indeed expand. Although Einstein subsequently redacted Λ from the field equations, calling it “his biggest blunder”, we now have accumulated solid evidence that the cosmological constant has a positive value measurably different from zero (Riess et al., 1998; Perlmutter et al., 1999).

The Hubble parameter made its first appearance in cosmology as the proportionality factor between the distance of a remote galaxy and its apparent recession velocity from

us due to the cosmological expansion: $v_{\text{rec}} = H_0 d$ (Hubble, 1929). It can be shown that this linear relation is the only one consistent with the Cosmological Principle (see Mukhanov, 2005). The inverse of the Hubble parameter defines a length scale $r_H(z) = H^{-1}(z)$. For an observer, imagine a sphere with radius r_H (*Hubble sphere* or *Hubble patch*) – its surface consists of the points which recede with the speed of light. Points outside cannot causally influence the observer at present, and vice versa. However, they could have done so in the past, or may have the opportunity in the future, since the Hubble parameter is a function of cosmic time – hence the occasionally used moniker “Hubble constant” is a historical misnomer (Bondi and Gold, 1948; Hoyle, 1948). A universe for which $H(z)$ is in fact constant, however, is called *de Sitter* space (de Sitter, 1917, 1918). Its scale factor grows exponentially, and the spacetime manifold actually exhibits a causality horizon at distance H^{-1} . Whilst this does not hold in any other spacetime, $r_H(z)$ is often treated as a convenient placeholder for more elaborate causality scales, which are usually of the same order of magnitude, and in a slight abuse of terminology called the *Hubble horizon*.

We can now see what distinguishes the three geometrical cases we outlined earlier: momentarily setting Λ to zero in Equation 7, we can reshuffle the Friedmann equation to

$$K = \frac{8\pi G a^2}{3} (\rho - \rho_{\text{cr}}), \quad \rho_{\text{cr}} := \frac{3H^2}{8\pi G a^2}. \quad (9)$$

Hence, the sign of K is determined by the complete energy budget of the Universe ρ in comparison to the *critical density* ρ_{cr} : if it exceeds it the geometry is spherical, if it falls short it is hyperbolic, and only if it precisely balances will the Universe be flat.

To solve for the three functions (a, ρ, p) , we need to supplement Equation 7 and Equation 8 with an equation of state connecting energy density and pressure: this arises from the statistical properties of the fluid in question, and is often of the form $p = w\rho$ for a constant w . For non-relativistic matter (i. e. the average kinetic energy of one particle is small against its rest mass), this is simply $w = 0$, whereas for highly relativistic matter (particle rest mass is either vanishing or small against the average kinetic energy), we have $w = 1/3$ (see Dodelson, 2003; Mukhanov, 2005; S. Weinberg, 2008).

The cosmic inventory at the present epoch is usually defined in terms of the omega parameters¹: for each species i , define $\Omega_i = \rho_{i,0}/\rho_{\text{cr},0}$ – its fraction measured in today’s critical density. If the species is not interacting with other particles, its energy density will then evolve like $\rho_i(t) = \Omega_i a(t)^{-3(1+w_i)}$.

Often, the cosmological constant and curvature are wedged into the same formalism by defining a “cosmological constant density” component $\Omega_\Lambda = \Lambda/3H_0^2$, $w_\Lambda = -1$. In particular, this means that the energy density associated with this component, i. e. $\rho_\Lambda =$

¹ In what follows, the subscript X_0 shall denote that the quantity X is evaluated at the present moment in cosmic time.

$\Omega_{\Lambda\rho_{\text{cr},0}}$, stays constant in time. The above conclusions connecting energy budget and curvature remain true if one includes this component. Dark energy scenarios, which intend to modify either the particle content or the laws of gravity, often present an explicit component that reproduces the effect of the cosmological constant. It is possible to go further and define a “curvature density” $\Omega_K = -3K/8\pi G$, which is negative for spherical geometry and positive for hyperbolic; this hypothetical component would have an equation-of-state parameter $w_K = -1/3$. Including both of these, the Friedmann equation [Equation 7](#) at present time can now be recast into the elegant form

$$\sum_i \Omega_i = 1; \tag{10}$$

now index i runs over all species of matter and radiation, Λ , and K .

In case that one single species is present in the Universe, the equation of motion for $a(t)$ can be solved analytically, yielding the expressions

$$a(t) = \exp(t/t_0) \quad (w = -1); \tag{11}$$

$$a(t) = (t/t_0)^{2/3(1+w)} \quad (\text{otherwise}). \tag{12}$$

One noteworthy feature of the power-law solutions is that there is a time for which $a(t) = 0$: the spacetime manifold is singular at this event. Moving backwards in time towards $t = 0$, the physical energy density, pressure and temperature all increase without bounds, until the validity of our current theories of gravity and particle physics cannot be ascertained. This scenario, which appears to be an accurate model of our Universe, is known as the hot Big Bang theory; what is before it is beyond our current understanding.

The propagation of photons through the expanding homogeneous and isotropic [FLRW](#) geometry are described by the geodesic equation [Equation 2](#). This has one crucial consequence whose importance for observational cosmology can hardly be overstressed: the energy of a photon decreases along its trajectory, according to the scale factor: $E(t) \propto a^{-1}(t)$. This means that a photon that has been emitted by a galaxy at a time t_{em} , and absorbed by an observer on earth at a later time t_{obs} will have been shifted towards redder wavelengths on its way (if the Universe is expanding), or towards the blue (if it is contracting). This effect is quantified by introducing the redshift parameter z via

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} = \frac{a(t_{\text{obs}})}{a(t_{\text{em}})}. \tag{13}$$

If such a photon is observed today, and its emission wavelength is known, e. g. by virtue of coming from a known spectral line, it allows us to measure the scale factor at which it has been emitted. This way, the redshift $1 + z = a^{-1}(t)$ becomes a natural time coordinate

for looking back into the cosmos: we are situated at $z = 0$, and the Big Bang happened at $z = \infty$.

One important observable specifying the cosmic expansion history is the redshift dependence of the *angular diameter distance*, one of several cosmological distance notions (see Dodelson, 2003; S. Weinberg, 2008). In a Euclidean space, the angular size θ of a remote object (in *rad*) is just the quotient of physical size to distance: $\theta = x/d$, hence if its size is known, we can infer its distance from the angle it subtends on our field of vision. On an FLRW spacetime, the idea “distance to a remote object, which we observe at a certain redshift z ” is far from obvious – the expanding geometry influences the propagation of the light coming towards us. Nevertheless we can extend the idea to define a redshift-dependent distance measure

$$d_A(z) = \frac{x}{\theta} = \frac{1}{H_0(1+z)\sqrt{|\Omega_K|}} S_k \left[\sqrt{|\Omega_K|} H_0 \chi(z) \right], \quad (14)$$

where the comoving distance $\chi(z)$ is found via an integral over the expansion history

$$\chi(z) = \frac{1}{H_0} \int_0^z \frac{d\tilde{z}}{\sqrt{\Omega_\Lambda + \Omega_K(1+\tilde{z})^2 + \Omega_m(1+\tilde{z})^3 + \Omega_r(1+\tilde{z})^4}}. \quad (15)$$

A cornucopia of data sets from a variety of different cosmic epochs have heralded the modern age of precision cosmology, and allowed us to measure the omega parameters to sub-percent accuracy. This permits us to construct a coherent narrative about the history of the Universe since the Big Bang – this concordance model is commonly referred to by the name Λ - Cold Dark Matter (Λ CDM). Its inventory consists of the following species (values from Planck Collaboration, 2015c, the uncertainties cited correspond to one-sigma):

- baryonic matter ($\Omega_b = 0.0490 \pm 0.0019$): This contains all non-relativistic particles of the standard model - in particular, baryons and non-relativistic leptons. All visible matter is made up of these species, including galaxies and interstellar gas.
- cold dark matter ($\Omega_c = 0.264 \pm 0.012$): The baryonic content is not sufficient to fill up the entire budget for non-relativistic matter – the rest must consist of a species whose fundamental particle physics is currently unknown. We know that it is non-relativistic (*cold*) and does not participate in electromagnetic or strong interactions (*dark*). This unsettling statement matches a series of other anomalies:

(i) Dating back as far as Kapteyn (1922), Oort (1932), and Zwicky (1933), it has been noted that the velocity dispersion of galaxies in clusters is far higher than expected from their visible mass content, indicating the presence of another gravitating component. For a more recent detection of such a system, see Minchin et al. (2005).

(ii) Similarly, temperature and flux of X-rays emitted by hot gas in galaxy clusters allow for a measurement of the gravitating mass in the system. Again, the baryons can only account for a small fraction of the full mass (Vikhlinin et al., 2006).

(iii) The differential rotation velocity of many spiral galaxies, in dependence on the radius from the central bulge, does not match the curves expected if the majority of the gravitating mass were in the bulge, i. e. the location of most of the luminous matter (Rubin, Ford, and Thonnard, 1980).

(iv) Weak gravitational lensing, a powerful method that will be introduced in Section 1.3.1, is equally sensitive to all gravitating mass, luminous or not. Heymans et al. (2013a) constrain the total mass content to $\Omega_c + \Omega_b = 0.255 \pm 0.014$ by combining tomographic cosmic shear measurements with WMAP7 data (Larson et al., 2011).

(v) Several systems – the Bullet Cluster (1E 0657-558) being the most famous – exhibit a significant offset between the luminous component (as imaged with X-rays) and the gravitating component (as inferred via weak and strong lensing) – see Markevitch et al. (2004).

- radiation ($\Omega_r = (9.28 \pm 0.52) \times 10^{-5}$): This measurement includes every species that has been relativistic in the moment the CMB was created – in the standard model, these are photons and neutrinos. Cosmology is consistent with the existence of three light neutrino species, as they are observed in particle physics.
- dark energy ($\Omega_\Lambda = 0.685 \pm 0.013$): This could simply be taken as a measurement of the cosmological constant Λ ; nevertheless, there are substantial fundamental problems with the theoretical understanding of this value, located at the interface between gravitation and quantum field theory – see Carroll, Press, and Turner (1992), Rugh and Zinkernagel (2001), and Padilla (2015); but also Bianchi and Rovelli (2010) for a dissenting standpoint. The physical nature of dark energy remains one of the most important unsolved problems of cosmology and fundamental physics. In addition to the CMB, the following observations confirm the existence of this additional component (this is by no means an exhaustive list, see, e. g., D. H. Weinberg et al., 2013 for a more in-depth treatment):

(i) Direct measurements of the matter density Ω_m indicate a deficit in the total energy budget – since the Universe is measured to be flat, the density parameters Ω_i for all matter species (including all dark components) have to add up to one. First indications that this is not valid without dark energy arose from the investigation of galaxy clustering (Efstathiou, Sutherland, and Maddox, 1990; Maddox et al., 1990; Cole, D. H. Weinberg, et al., 1997).

(ii) Type-Ia supernovae can be used to trace the late-time expansion history of the Universe since they are standard candles. This probe constituted the first significant discovery that said expansion is accelerating (Riess et al., 1998; Perlmutter et al., 1999), which was awarded the 2011 Nobel Prize in Physics.

(iii) Weak gravitational lensing and baryon acoustic oscillations are probes that test the non-linear clustering of matter as well as the late-time expansion history in a complementary fashion; both confirm the existence of dark energy to high significance. Their theory and observation will be introduced in detail in [Section 1.3.1](#) and [Section 1.3.2](#).

The *Planck* data set, from which these values are derived, measures the statistical properties of the CMB radiation, whose physics will be detailed in [Section 1.2](#); they are the constraints from the temperature power spectrum on baseline flat Λ CDM (see [Section 1.2](#) for details on this observable). If the curvature parameter is allowed to vary, the connection of the radiation temperature and polarisation spectra constrain Ω_K to zero to a precision of less than 10^{-2} , hence the restriction to flat Λ CDM cosmology as our *Standard Model of Cosmology* is a good one.

Winding back cosmic time in this concordance scenario, the following eras unfold:

- $z = 0$ – today: The most dominant component in the Universe is dark energy – i. e. the cosmological constant. Having taken over only relatively recently, it will continue to do so into the far future, and the expansion of the Universe will approximate the accelerating de Sitter expansion law [Equation 11](#) ever more closely.
- $z \sim 0.7$: around this time, the sign of the deceleration parameter switches, since we are transitioning from Λ -domination into an era of matter-domination, where $a(t) = (t/t_0)^{2/3}$. The nonlinear collapse of cosmic structures (dark matter halos, galaxies) mostly takes place in this era – see [Section 1.3](#).
- $z \sim 1100$: the recombination event produces the radiation of the CMB – see [Section 1.2](#). Before, the Universe is filled with hot opaque plasma; photons cannot move freely because they scatter off free electrons.
- $z \sim 3400$: at this time, radiation takes over the role of dominating component; hence $a(t) = (t/t_0)^{1/2}$.
- $z \sim 10^{28}$: evidence is substantiating that at extremely high redshift another phase of approximately de Sitter-like expansion took place, dubbed *cosmic inflation*.
- $z \rightarrow \infty$: in classical gravitation, this would be the spacelike singularity, the *Big Bang*. Since this locus signals a breakdown of the theory, we expect that at a finite

redshift, some more fundamental theory superseding general relativity, perhaps involving quantum gravity, will take over. *Here be dragons.*

1.2 AT ORDER ONE: THE COSMIC MICROWAVE BACKGROUND

One robust prediction of the Big Bang model is the presence of a faint background of thermal electromagnetic blackbody radiation. Since the photon energy density scales with $(1+z)^4$, and the Stefan-Boltzmann law dictates $\rho_\gamma = \sigma T_\gamma^4$, the photon temperature will increase linearly with redshift; the number densities of all particle species grow as well. The radiation of this hot dense plasma should still be visible today, redshifted to longer wavelengths. Indeed, a faint isotropic microwave background radiation was discovered by Penzias and R. W. Wilson (1965), more or less by accident. We know today that this omnipresent radiation component has a density of roughly 300 photons per cubic centimetre, that it is almost perfectly isotropic, and that its energy distribution follows the Planckian distribution to high precision, with a temperature of 2.73 K (Fixsen et al., 1996) – in fact, it is the most perfect blackbody radiation known to science. This radiation has been studied by many probes, including COBE (Smoot et al., 1992), BOOMERanG (Torbet et al., 1999), WMAP (Bennett et al., 2003), and *Planck* (Planck Collaboration, 2015a).

Although these have revealed a remarkable isotropy of temperature, they have found fluctuations of the order 10^{-5} , and increased precision has revealed more and more detail on the temperature maps (see Figure 2). These structures force us to consider deviations from the perfectly homogeneous and isotropic FLRW background metric: this is done by introducing density perturbations

$$\delta_i(t, \underline{x}) = \frac{\rho_i(t, \underline{x}) - \bar{\rho}_i(t)}{\bar{\rho}_i(t)}, \quad (16)$$

and corresponding perturbations in the metric tensor (as defined in the *conformal Newtonian gauge*)

$$ds^2 = [1 + \Psi(t, \underline{x})]dt^2 - a^2(t)[1 + \Phi(t, \underline{x})] \left\{ \frac{dr^2}{1 - Kr^2} + r^2 d\Omega \right\}, \quad (17)$$

into the joint system of Einstein and Boltzmann equations, which govern the interacting species of matter and radiation, and then develop the first perturbative order around FLRW (see Peebles, 1980; Ma and Bertschinger, 1995; Dodelson, 2003, for the detailed calculation).

The aforementioned inflationary paradigm offers an intriguing explanation for these minute departures from perfect homogeneity: it postulates a period of accelerated expansion in the very early cosmos, by a factor of at least $\exp(60)$.

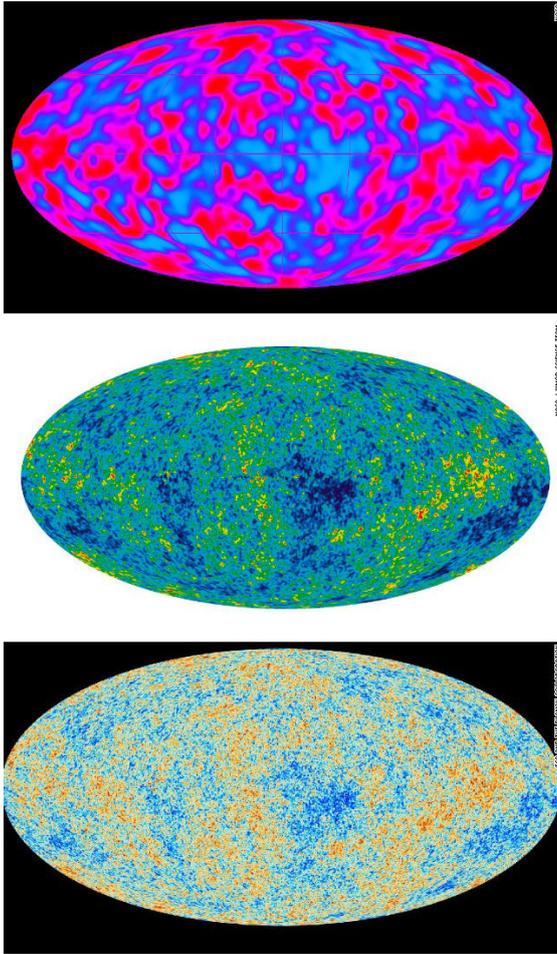


Figure 2: Comparison of CMB temperature anisotropies, as recorded by COBE, WMAP, and Planck.
From <http://bit.ly/2adJw0r>.

During this time, the quantum fluctuations in the field driving said expansion get stretched outside of the horizon; and at the end of inflation will be imprinted onto all other species. Subsequently, the modes of these density contrasts Equation 16 will collapse under their own gravity and form structures. This growth is suppressed in the early radiation-dominated epoch ($z > 3400$) and at late times dominated by dark energy ($z < 0.7$), compared to the matter-dominated period in between.

By the time that the CMB is generated, all scales still have density contrasts $\delta_i \ll 1$, and can be treated at first perturbation order – this will change in later periods of structure formation, see Section 1.3.

In the early Universe, the photons were tightly coupled to the electrons by Thomson scattering and could not propagate freely; the electron component itself was tied to the nuclei (mostly protons) via Coulomb scattering – the photons were too energetic to let electrons and nuclei form neutral atoms; protons, electrons

and photons for an opaque plasma. However, once the cosmos expanded enough and the photon temperature dropped below $\sim 3000\text{K}$, the reaction $p + e^- \leftrightarrow H + \gamma$ tipped over in favour of neutral hydrogen – this *recombination event* occurred around 380,000 years after the Big Bang ($z \simeq 1100$). As the fraction of free electrons dropped sharply, the photons were released from their tight coupling, their mean free path grew larger than the Hubble radius, and their motion became ballistic instead of diffusive. These are the photons we observe as the CMB radiation today, redshifted into the microwave regime of the electromagnetic spectrum.

Between the transition from radiation-domination to matter-domination at $z \simeq 3400$ and the recombination event at $z \simeq 1100$, the perturbations Equation 16 in the density components would have grown to about one part in 10^4 , and the perturbed gravitational potential Φ in Equation 17 alongside them. After their decoupling from baryonic matter, some photons found themselves in gravitational wells, and were redshifted once they

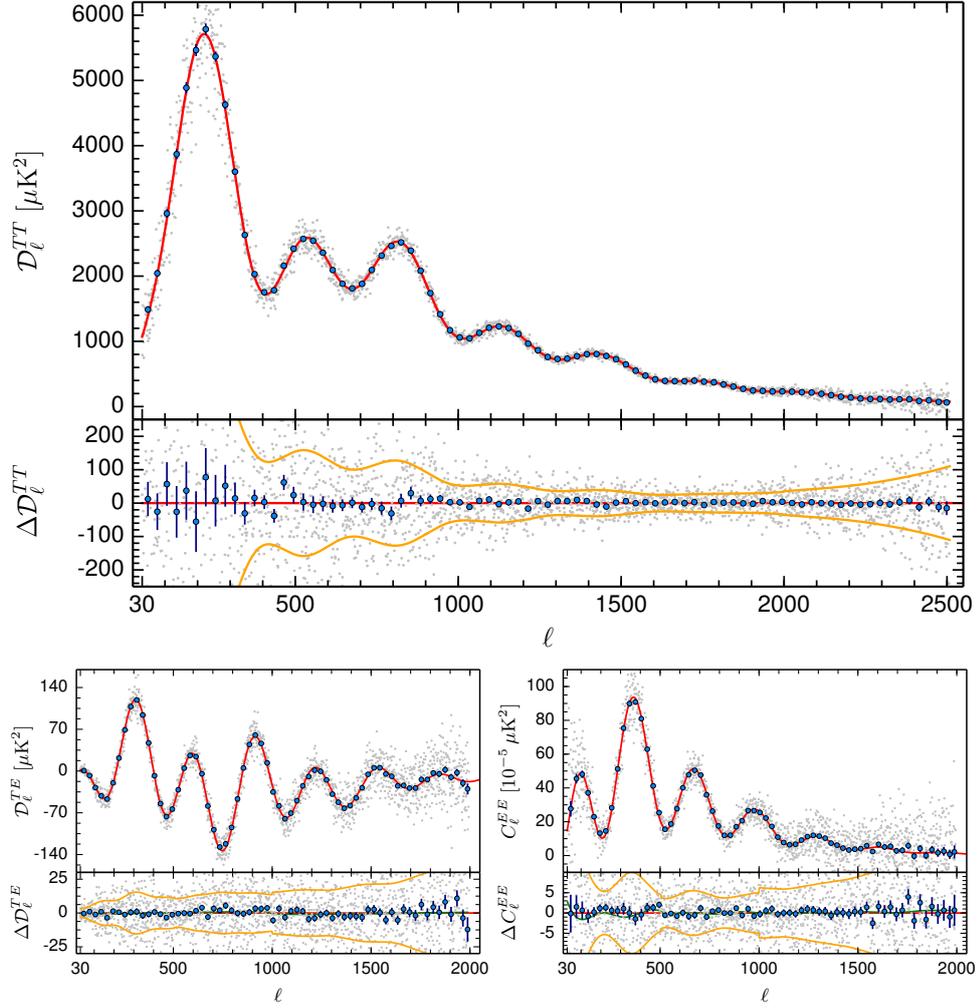


Figure 3: Temperature and polarisation power spectra, as measured by *Planck*. The red line is the best flat Λ CDM as fitted to the temperature spectra alone; the lower half for each panel shows residuals. Grey dots are the measured spectra for each multipole, blue dots are binned in ℓ , the yellow lines show the 68%-errors of the unbinned points. From (Planck Collaboration, 2015b).

climbed out; other photons experienced a corresponding blueshift. Further, the potential Ψ in Equation 17 introduces a time dilation, which also results in shifting the photon energy. Accordingly, the temperature of the photons carries the imprint of the metric perturbations – this interplay between photons and gravitational potentials is known as the Sachs-Wolfe effect (Sachs and Wolfe, 1967), and is responsible for the temperature fluctuations visible in the CMB. The anisotropies carry the imprint of a tremendous amount of physics: they are a window into the very early cosmos long before the formation of galaxies; they offer the opportunity to measure cosmological parameters with high precision. Some quantities – like Ω_K and Ω_Λ – influence the propagation of the CMB photons from recombination to us, some affect the recombination event directly –

like Ω_b , some determine the structure growth before recombination, like Ω_c . Also, we can infer properties of the primordial inhomogeneity spectrum that has been generated by inflation, like its spectral index n_s , and thus look back to the earliest time that is meaningful in a cosmological context. The amount of information increases further once we take into account the polarisation of the radiation.

On top of the *primary anisotropies* generated by the early Sachs-Wolfe effect at the recombination event, there are several processes influencing the CMB radiation on its journey through the late-time universe. To mention a few: the pattern of primary anisotropy is lensed by the gravitational potentials of the evolving dark matter (*CMB lensing*, see Lewis and Challinor, 2006), the photons experience additional shifts in frequency by these potentials (*Rees-Sciama effect*, see Rees and Sciama, 1968), and they are scattered by hot gas in galaxy clusters (*Sunyaev-Zel'dovich effect*, see Sunyaev and Zel'dovich, 1970). These secondary effects harbour a wealth of information adding to the primary structure – for a more complete treatment see Aghanim, Majumdar, and Silk (2008).

The main observable that is extracted from CMB maps is the angular power spectra, i. e. the auto- and cross-correlations of temperature and polarisation in harmonic space (see Figure 3), which quantify the statistical correlation properties of two random points separated by a specific angle. Since the evolution of the density contrasts in this regime is linear, and since non-Gaussianity in the primordial spectrum has yet to be found (Planck Collaboration, 2015d), we can expect that this two-point statistic contains the major part of the cosmological information.

The next generation of surveys will concentrate on precision measurements of CMB polarisation, with particular emphasis on a possible primordial B-mode component, its impact on inflation, physics beyond the standard model, neutrino mass constraints, and spectral distortions. Current and upcoming ground-based efforts include ACTPOL (Niemack et al., 2010; Henderson et al., 2016), POLARBEAR and the Simons Array (Arnold et al., 2010; Suzuki et al., 2016), SPTpol (Austermann et al., 2012), the Keck array and BICEP3 (Staniszewski et al., 2012; Grayson et al., 2016). Also, there are balloon-based experiments like SPIDER (Crill et al., 2008); proposals for space-based missions include CORe+ (Rubiño-Martín and CORe+ Collaboration, 2015), LiteBIRD (Matsumura et al., 2014), and PIXIE (Kogut et al., 2014).

1.3 AT HIGHER ORDER: EVOLUTION OF STRUCTURE

Validity of perturbation theory hinges upon the density contrasts [Equation 16](#), or their Fourier transforms

$$\delta_i(t, \underline{k}) = \int d^3x \delta_i(t, \underline{x}) \exp[-i \underline{k} \cdot \underline{x}], \quad (18)$$

being $\ll 1$. We would expect that growth of cosmic structure increases the amplitude of these fluctuations, until this assumption breaks down for at least a range of k -modes. To investigate this, we define the dark matter power spectrum through the two-point correlation function in momentum space

$$\langle \bar{\delta}_c(t, \underline{k}) \delta_c(t, \underline{q}) \rangle = (2\pi)^3 \delta_D^{(3)}(\underline{k} - \underline{q}) P(k, t). \quad (19)$$

The proportionality with the Dirac delta function is enforced by the statistical homogeneity of the field in real space, and isotropy ensures that the power spectrum $P(k, t)$ depends only on the magnitude of \underline{k} . The dimensionless power spectrum

$$\Delta^2(k, t) = \frac{k^3 P(k, t)}{2\pi^2} \quad (20)$$

is a convenient measure for the mean squared fluctuations of the random field δ_c on a given scale k . As long as this is well below unity, the linear approximation holds. In the first-order perturbation calculation mentioned in [Section 1.2](#), the differential equations governing Fourier modes $\delta_c(\underline{k})$, $\delta_c(\underline{q})$ decouple for $\underline{k} \neq \underline{q}$. Hence, the amplitude of every mode grows independent of all others – the evolution from the initial amplitude at the exit from inflation $\delta_c(z = z_i, \underline{k})$ to today ($z = 0$), is encapsulated in the transfer function (except from an overall factor $D_1(z)$, see below)

$$T(k) = \frac{\delta_c(z = 0, \underline{k})}{\delta_c(z = z_i, \underline{k})} \frac{\delta_c(z = z_i, \underline{k} = \underline{q})}{\delta_c(z = 0, \underline{k} = \underline{q})}, \quad (21)$$

which encompasses every physical effect that distinguishes the growth process of one mode from another (e.g. the redshift of radiation-matter transition, or the impact of baryons, see [Section 1.3.2](#)). For a given combination of cosmological parameters, the transfer function is computed via one of the analytic fitting formulae found by numerically integrating the Boltzmann equations (Bond and Szalay, 1983; Bond and Efstathiou, 1984; Bardeen et al., 1986; Eisenstein and Hu, 1998).

The overall increase in amplitude, common to all modes, is described by the growth function $D_1(z)$. It depends on the cosmic expansion history, i.e. which component was dominating the Friedmann equations for which range of redshift. Under certain assumptions on redshift z and cosmological constant Λ , this growth function can be

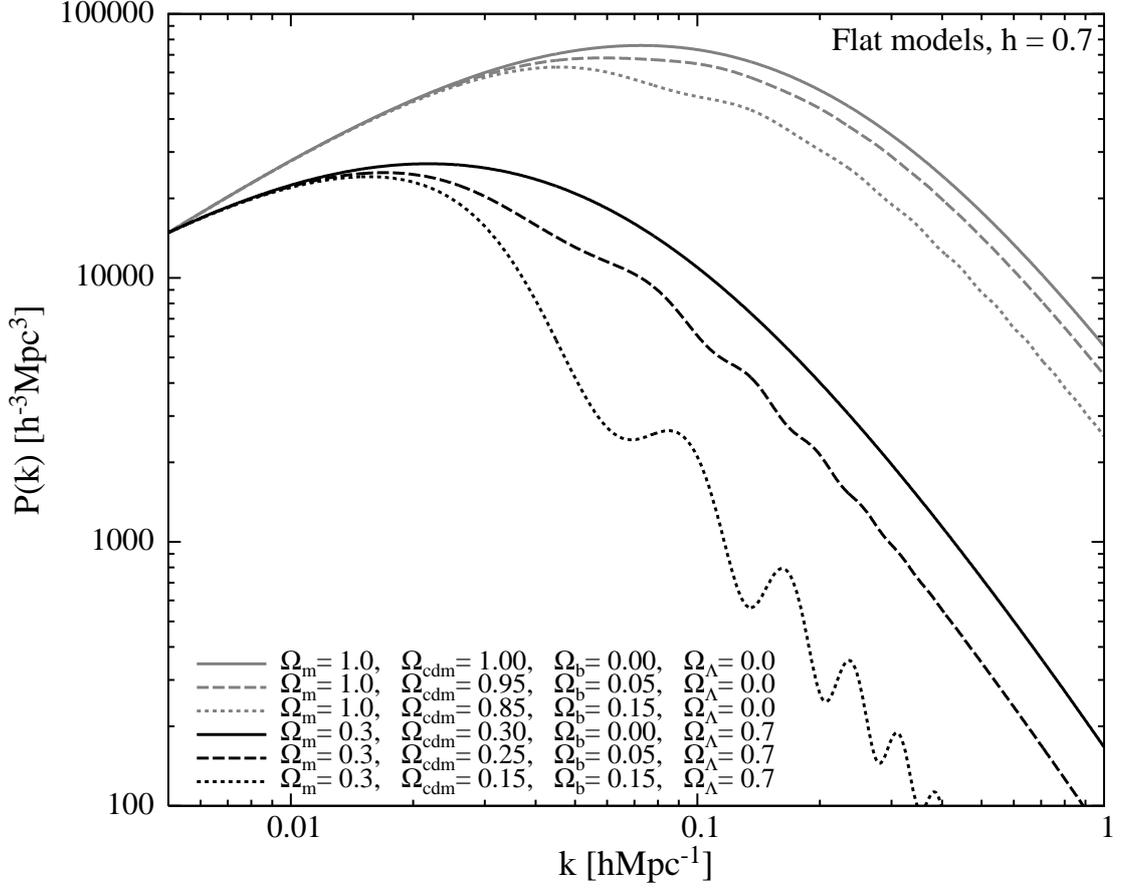


Figure 4: Linear power spectra of cold dark matter at redshift $z = 0$, computed for various combinations of cosmological parameters. From Nichol (2008).

expressed analytically as a hypergeometric function; convenient approximations for a wider range of values for Ω_m and Ω_Λ can be found in Lightman and Schechter (1990), Lahav et al. (1991), and Carroll, Press, and Turner (1992). If we assume that the initial dark matter power spectrum as seeded by inflation can be described by a simple power law $P_i(k) \propto k^{n_s}$, we can assemble an expression for the linear power spectrum of evolved cold dark matter:

$$\Delta_L^2(z, k) \propto k^{n_s+3} T^2(k) D_1^2(z), \quad (22)$$

and an analogous expression for $P_L(z, k)$. See Figure 4 for a comparison of several linear power spectra, computed in various cosmological models.

1.3.1 Weak Gravitational Lensing

Gravity – i.e. the curvature of spacetime – deflects the propagation of light. This prediction was the target of the first experimental trial to which Einstein’s theory of general relativity was subjected. Although Newton’s theory can produce a similar bending in the trajectory of corpuscular light near massive bodies, the Equivalence Principle predicts a deflection angle twice as large (Einstein, 1911) – and indeed: during the solar eclipse of May 29th, 1919, the pattern of stars surrounding the sun was observed by Dyson, Eddington, and Davidson (1920). At the time, the majority of scientists took their results as a clear confirmation for Einstein over Newton, although today it is unclear whether their methodology holds up to scrutiny (Earman and Glymour, 1980).

Further investigation of the phenomenon brought the realisation that a gravitational well can act like an optical lens by influencing the image of luminous objects (e.g. galaxies, quasars or even single stars) behind it. The result of this can be a deformation of a single galaxy into one or even multiple arcs (*strong lensing*, see Chwolson, 1924; Einstein, 1936; Walsh, Carswell, and Weymann, 1979), the statistically correlated minimal image stretching and deformation over a large field of galaxies, as induced by large scale structure (*weak lensing* – which we will be concerned with, see Gunn, 1967; Blandford et al., 1991; Miralda-Escude, 1991; Kaiser, 1992), or a temporary increase in luminosity of a single object due to focusing, without noticeable deformation of its image (*microlensing*, see Irwin et al., 1989).

Due to its historical role for the theory, most books introducing general relativity contain Einstein’s perturbative calculation of the deflection angle of a light ray grazing a single massive object, including S. Weinberg (1972), Misner, Thorne, and Wheeler (1973), Wald (1984), and Oloff (1999), see also P. Schneider, Ehlers, and Falco (1992) for a treatment geared towards lensing, and Bartelmann and P. Schneider (2001), Hoekstra and Jain (2008), Munshi et al. (2008), Bartelmann (2010), and Kilbinger (2015) for general reviews of weak gravitational lensing. It is found to be $\alpha = 2r_s/\zeta = 4GM/\zeta$ for impact parameter ζ along an object of mass M and Schwarzschild radius r_s . Given that the positions of galaxies on the sky have only small angular separation from

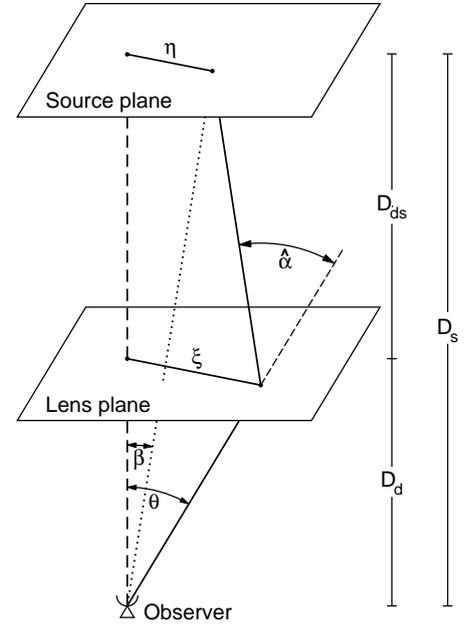


Figure 5: Shift of a source at β through a thin gravitational lens, making it appear under an angle θ . All coordinates are comoving. From Bartelmann and P. Schneider (2001).

the optical axis, we can treat the elongation angles as two-dimensional vectors and approximate the sky as flat. In Figure 5, a source is located at an angle $\underline{\beta}$, so its transverse elongation in the source plane is $\underline{\eta} = D_s \underline{\beta} = \mathcal{D}(\chi_s) \underline{\beta}$. Here $\mathcal{D}(\chi_s)$ is the comoving angular diameter distance to the source located at comoving distance χ_s and redshift z_s : $d_A(z_s) = \mathcal{D}(\chi_s)/(1+z_s)$ – see Equation 14. Assuming for an instant that the thickness of the lens plane is negligible (*thin lens approximation*), the geometry of Figure 5 dictates that $D_s \underline{\theta} = D_s \underline{\beta} + D_{ds} \underline{\alpha}$, therefore with $D_s = \mathcal{D}(\chi_s)$, $D_{ds} = \mathcal{D}(\chi_s - \chi)$

$$\delta \underline{\theta} = \underline{\theta} - \underline{\beta} = \frac{\mathcal{D}(\chi_s - \chi)}{\mathcal{D}(\chi_s)} \underline{\alpha} = \frac{\mathcal{D}(\chi_s - \chi)}{\mathcal{D}(\chi_s)} 2 \nabla_{\perp} \Phi. \quad (23)$$

	< 0	> 0
κ		
$\text{Re}[\gamma]$		
$\text{Im}[\gamma]$		

Figure 6: Effect of convergence κ and complex shear γ as linear deformations of a small circular disk (green). From <http://bit.ly/2amrZYi>.

The deflection angle can be expressed through the projection of the potential gradient onto the lens plane. Since the problem has been linearised, the deflection due to a bulk of gravitating matter can be found by integrating over a continuum of thin lenses:

$$\underline{\theta} = \underline{\beta} + 2 \int_0^{\chi_s} d\chi \frac{\mathcal{D}(\chi_s - \chi)}{\mathcal{D}(\chi_s)} \nabla_{\perp} \Phi(\chi). \quad (24)$$

See Dodelson (2003) and Bartelmann (2010) for a derivation of this result via the geodesic equation Equation 2. Since we are interested in the slight deformations in the images of galaxies, we take only the linear term in the map from apparent position $\underline{\theta}$ to actual position $\underline{\beta}$: $\underline{\beta} = (\mathbb{1} - \psi) \underline{\theta} + \mathcal{O}(\delta\theta^2)$, with the matrix

$$\psi_{ij} = -\frac{\partial}{\partial \theta_j} \delta \theta_i = -2 \int_0^{\chi_s} d\chi \frac{\mathcal{D}(\chi_s - \chi) \mathcal{D}(\chi)}{\mathcal{D}(\chi_s)} \partial_i \partial_j \Phi(\chi). \quad (25)$$

This symmetric matrix contains three real degrees of freedom describing the linear deformation of a small circular image. It is customary to introduce two quantities, *convergence* $\kappa(\theta)$ and *shear* $\gamma(\theta) = \gamma_1 + i\gamma_2$ to classify the image distortions:

$$\psi = \begin{bmatrix} -\kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & -\kappa + \gamma_1 \end{bmatrix}. \quad (26)$$

If we are interested in the shear field induced by cosmic large scale structure, rather than lensing by single objects, we refer to *cosmic shear*. The distortion effect of each of these components is illustrated in Figure 6. To make contact with power spectra of the large-

scale matter distribution which induces weak lensing, we insert the Poisson equation, which is the weak-field limit of general relativity:

$$\Delta\Phi = \frac{3}{2}\Omega_m H_0^2 (1+z)\delta_m, \quad (27)$$

into Equation 25 and Equation 26, and approximate the 2D Laplacian via the 3D Laplacian inside the χ -integration (see Limber, 1954; Peebles, 1980; Kaiser, 1992), which averages out modes parallel to the line of sight:

$$\kappa(\underline{\theta}; z_S) \simeq \int_0^{\chi_S} d\chi w(\chi, \chi_S) \delta[\chi, \mathcal{D}(\chi)\underline{\theta}], \quad (28)$$

with the definition

$$w(\chi, \chi_S) = \frac{3}{2}H_0^2\Omega_m(1+z)\frac{\mathcal{D}(\chi)\mathcal{D}(\chi_S-\chi)}{\mathcal{D}(\chi_S)}. \quad (29)$$

It should be noted that Equation 28 is practical if all source galaxies had the very same redshift z_S . Realistically, observers deal with a source distribution $p(z_S)$, which has to be modelled according to the data set at hand. One frequently used form is

$$p(z_S) = \frac{\beta z_S^2}{\Gamma(3/\beta) z_0^3} \exp\left[-\left(\frac{z_S}{z_0}\right)^\beta\right], \quad (30)$$

with parameters z_0, β suitably chosen (Brainerd, Blandford, and Smail, 1996). For such a distribution with maximal redshift z_{\max} , the convergence is

$$\kappa(\underline{\theta}) = \int_0^{z_{\max}} dz_S p(z_S) \kappa(\underline{\theta}; z_S) = \frac{3H_0^2\Omega_m}{2} \int_0^{\chi_{\max}} d\chi \frac{q[z(\chi)]\mathcal{D}(\chi)}{a(\chi)} \delta(\chi, \mathcal{D}(\chi)\underline{\theta}), \quad (31)$$

with the *lensing efficiency kernel*

$$q(z) = \int_z^{z_{\max}} dz_S p(z_S) \frac{\mathcal{D}(\chi_S - \chi)}{\mathcal{D}(\chi_S)}. \quad (32)$$

The power spectrum of the convergence field is found from the 2D versions of Equation 18 and Equation 19:

$$P_{\kappa\kappa}^{(12)}(\underline{\ell}) = \frac{9H_0^4\Omega_m^2}{4} \int_0^{\chi_{\max}} d\chi \frac{q^{(1)}(\chi)q^{(2)}(\chi)}{a^2(\chi)} P_{\delta\delta}(\underline{k})|_{\underline{k}=[\underline{\ell}/\mathcal{D}(\chi),\chi]}, \quad (33)$$

where we allowed for cross-correlating two different galaxy samples $p^{(1)}(z_S), p^{(2)}(z_S)$ with corresponding kernels $q^{(i)}(\chi)$ (Hu, 1999). It can be shown that the cosmic shear power spectra $P_{\gamma\gamma}$ are identical to spectra $P_{\kappa\kappa}$.

In an idealised setting, with all galaxies assumed spherical and spaced on a quadratic

grid, the cosmic shear can be estimated directly from the measured ellipticity. This can be derived from the quadrupole tensor of an image $I(\underline{\theta})$:

$$q_{ij} = \frac{1}{\mathcal{N}} \int d\theta w[I(\underline{\theta})](\theta_i - \langle \theta_i \rangle)(\theta_j - \langle \theta_j \rangle), \quad (34)$$

$$\text{with } \langle \theta_k \rangle = \frac{1}{\mathcal{N}} \int d\theta w[I(\underline{\theta})]\theta_k; \quad \mathcal{N} = \int d\theta w[I(\underline{\theta})]. \quad (35)$$

The weight function $w[I]$ allows down-weighting of the noisy outer regions of a galaxy, to obtain a more reliable shape measurement. Now, the (complex) ellipticity of the image is defined via

$$\epsilon = \frac{q_{11} - q_{22} + 2iq_{12}}{q_{11} + q_{22} + 2\sqrt{q_{11}q_{22} - q_{12}^2}}. \quad (36)$$

In the real world, galaxy shapes already have elliptical or more complex shapes, so upon being weakly lensed, their intrinsic ellipticity ϵ_S is related to the observed ϵ via

$$\epsilon = \frac{\epsilon_S + g}{1 + \epsilon_S g} \simeq \epsilon_S + \gamma; \quad g = \frac{\gamma}{1 - \kappa'}, \quad (37)$$

(Seitz and P. Schneider, 1997). Compared to the intrinsic ellipticity, the shear is usually an effect of a few percent, hence to collect a sufficient amount of information it is necessary to precisely image a high number of galaxies in large cosmic volumes. The following issues have to be included in the forward modelling process, to accurately extract the weak lensing signal:

- Modelling the distribution of source ellipticities is challenging, and needs to be calibrated separately. Additionally, galaxy formation can align the intrinsic ellipticity of adjacent galaxies, and this correlation can be mistaken for a shear signal (Joachimi, Cacciato, et al., 2015; Kiessling et al., 2015; Kirk et al., 2015).
- Given the number of galaxies necessary, it is too costly to take spectra for each one. The problem of inducing the source redshift z_S from the flux in several bands (*photometric redshifts*) is a vital modelling step opening up a large field of research, and can introduce further uncertainties or biases if done inadequately.
- Precise imaging is indispensable. Hence, weak lensing is particularly sensitive to proper modelling of the point-spread function, including atmospheric seeing and instrumental effects.
- The finite resolution of CCD cameras breaks the image into pixels, which further deprecates the image quality. Since we need large numbers of galaxies, we are coerced into working at the resolution limit.

Galaxies: Intrinsic galaxy shapes to measured image:

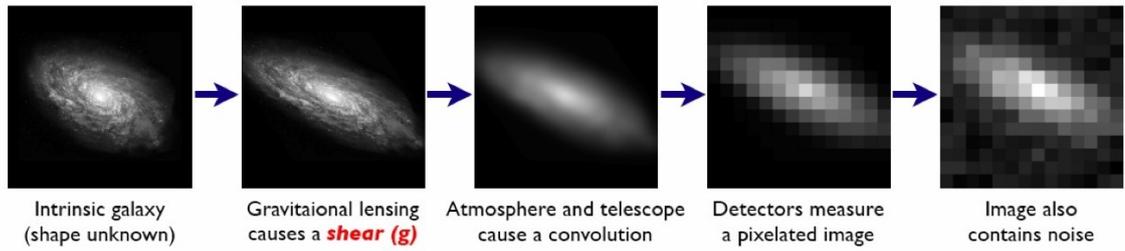


Figure 7: Schematic demonstration of the inverse problem of shear estimation, from actual galaxy to image: the weak lensing signal, confronted with some of the most important steps of the forward modelling process. From Bridle et al., 2009.

Figure 7 illustrates various distortions acting on a galaxy image, which have to be modelled appropriately to recover the weak lensing shear signal.

Despite these challenges to accurate inference, weak gravitational lensing has become one of the strongest probes of the dark sector, since it was first detected by four groups almost simultaneously (Bacon, Refregier, and R. S. Ellis, 2000; Kaiser, G. Wilson, and Luppino, 2000; Van Waerbeke et al., 2000; Wittman et al., 2000). Further methodology was developed by COSMOS (Schrabback et al., 2010) and the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) (Heymans, Van Waerbeke, et al., 2012; Kilbinger, Fu, et al., 2013; Simpson et al., 2013). One particularly promising strategy to increase the leverage on the evolution of the lensing signal with redshift is *weak lensing tomography* (see Heymans et al., 2013b): the galaxies are divided into redshift bins, such that sources in higher-redshift bins get lensed by a larger portion of mass inhomogeneities than those at lower z . The redshift bins are modelled by different source distributions $p_i(z_S)$, hence the observables are the auto- and cross-correlation power spectra Equation 33, or the corresponding correlation functions in real space.

Since the scales relevant for weak lensing surveys are non-linear, we expect the shear field to have significant non-Gaussian correlation properties. Indeed, these have been observed in Bernardeau, Mellier, and van Waerbeke (2002). Accurate inference of parameters using power spectra or an equivalent two-point statistic has to take these into account (Cooray and Hu, 2001; Semboloni et al., 2007; Hartlap, Schrabback, et al., 2009). In addition, using higher-order statistics such as skewness has the potential to break parameter degeneracies in the shear power spectra (Bernardeau, Mellier, and van Waerbeke, 2002; Kilbinger and P. Schneider, 2005); see Fu et al. (2014) and Simon et al. (2015) for recent measurements of third-order statistics on CFHTLenS data.

The precision of the most recent weak lensing surveys has reached the point where the uncertainties in cosmological parameters are limited by systematics; hence their proper modelling will become even more important with the advent of even larger and deeper data sets within the next decade, which are at least partly devoted to weak gravitational

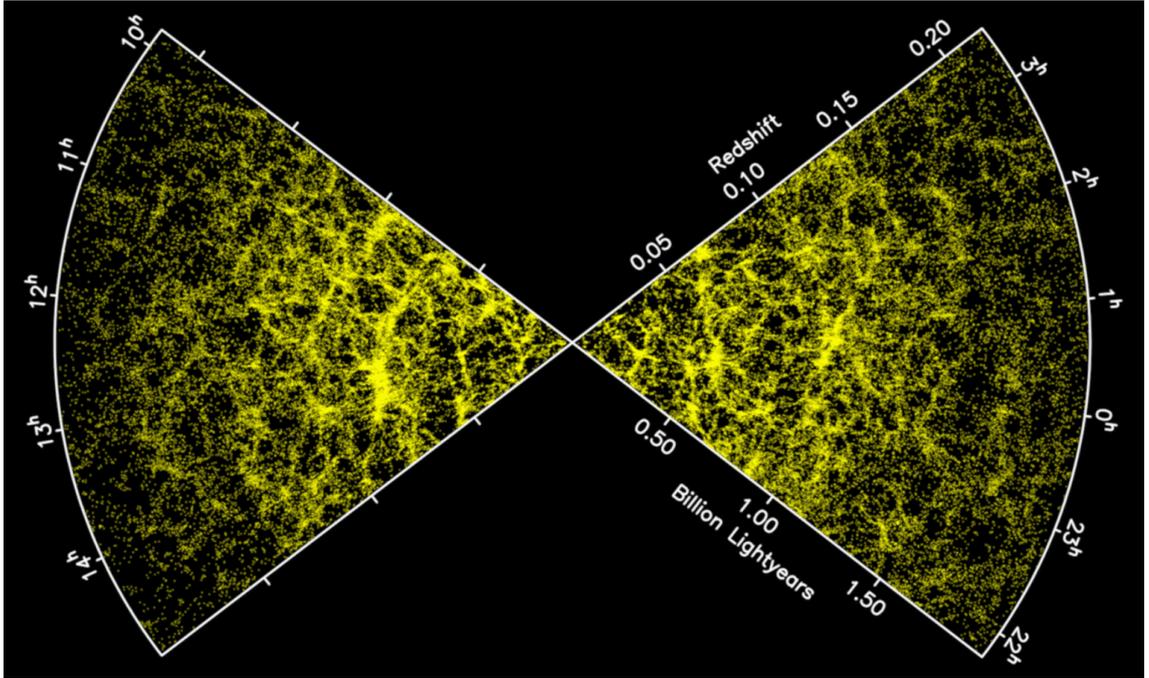


Figure 8: Galaxy clustering: a slice from the [2dFGRS](#), a spectroscopic survey mapping the 3D positions of approximately 220,000 galaxies (Cole, Percival, et al., 2005), displaying the filamentary structure in the statistical distribution of clustered objects.

From http://www.roe.ac.uk/~jap/2df/2df_slice_black_big.gif.

lensing – most notably *Euclid* (Laureijs et al., 2011); the Kilo-Degree Survey ([KiDS](#)) (de Jong et al., 2013; Hildebrandt, Viola, et al., 2016); the Dark Energy Survey ([DES](#)) (Becker et al., 2015; Dark Energy Survey Collaboration, 2016); and the Large Synoptic Survey Telescope ([LSST](#)) (LSST Dark Energy Science Collaboration, 2012).

1.3.2 Baryon Acoustic Oscillations

Before the recombination event, the baryonic part of the matter content consisted of free electrons and ionised nuclei (mostly hydrogen and helium), tied together by Coulomb scattering. Two forces act on them: gravity and the pressure of the photons, which is only felt by electromagnetically charged particles. Due to the high number densities, photons scatter off electrons efficiently, with a mean free path that is below the Hubble horizon scale – the photonic, electronic and baryonic fluids are coupled together tightly, and are constrained to move as a bulk.

If we express the perturbation in the photon monopole via the local deviations from the average temperature $\Theta(t, \underline{x}) = \delta T / \bar{T} = \delta \rho_\gamma / 4\bar{\rho}_\gamma$, and insert these into the conservation equations for photon number and photon momentum, the following linear hyperbolic

wave equation emerges (see Peebles and J. T. Yu, 1970; Hu and White, 1996; Hu and Dodelson, 2002):

$$\left[\frac{d^2}{d\eta^2} + \frac{R'}{1+R} \frac{d}{d\eta} + \frac{k^2}{3(1+R)} \right] (\Theta + \Phi) = \frac{k^2}{3} \left(\frac{\Phi}{1+R} - \Psi \right). \quad (38)$$

A similar equation can be derived for δ_b (see Eisenstein, Seo, and White, 2007). Here, a prime ' denotes a derivative with respect to conformal time η – see Equation 4. The quantity R quantifies the impact of the baryons onto the photonic component: it is defined as the quotient of baryonic to photonic spatial momentum densities:

$$R = \frac{|\pi_b|}{|\pi_\gamma|} = \frac{3\rho_b}{4\rho_\gamma}. \quad (39)$$

For every particle species in the cosmic inventory, this momentum density is the three-dimensional momentum per unit spatial volume; its components are stored inside the energy-momentum tensor $T_{\mu\nu}$ for each species: $(\pi_X)^i = (T_X)_0^i$. In the very early Universe the majority of the momentum is stored in photons ($R \ll 1$), but towards the recombination event the drag of the baryons adds a considerable amount of inertia to the photons.

Equation 38 is a damped harmonic oscillator in the Fourier domain, driven by the gravitational potentials on the right hand side. The fact that the combination $(\Theta + \Phi)$ appears is just the Sachs-Wolfe effect explained in Section 1.2. Thus, we expect solutions in the form of acoustic pressure waves propagating through the optically thick plasma, with the sound speed

$$c_s^2 = \frac{1}{3(1+R)}. \quad (40)$$

These sound waves are commonly referred to as baryon acoustic oscillations BAO. First, these waves move freely, but once baryon drag becomes important around recombination, R grows – the drop in sound speed and the increase of the friction term in Equation 38 shut off the wave propagation. Since the exit from inflation, they have travelled a comoving distance of

$$r_s = \int_{\eta_i}^{\eta_{\text{rec}}} d\eta c_s(\eta) \quad (41)$$

until the wavefronts get frozen in. This *sound horizon* is imprinted onto the baryonic matter distribution as an enhanced correlation in the statistical matter distribution across this length scale. It can also be seen in the CMB, the snapshot of the baryon-photon plasma from the last moment it was optically thick, and determines the location of the first acoustic peak in the multipoles of the temperature autocorrelation function

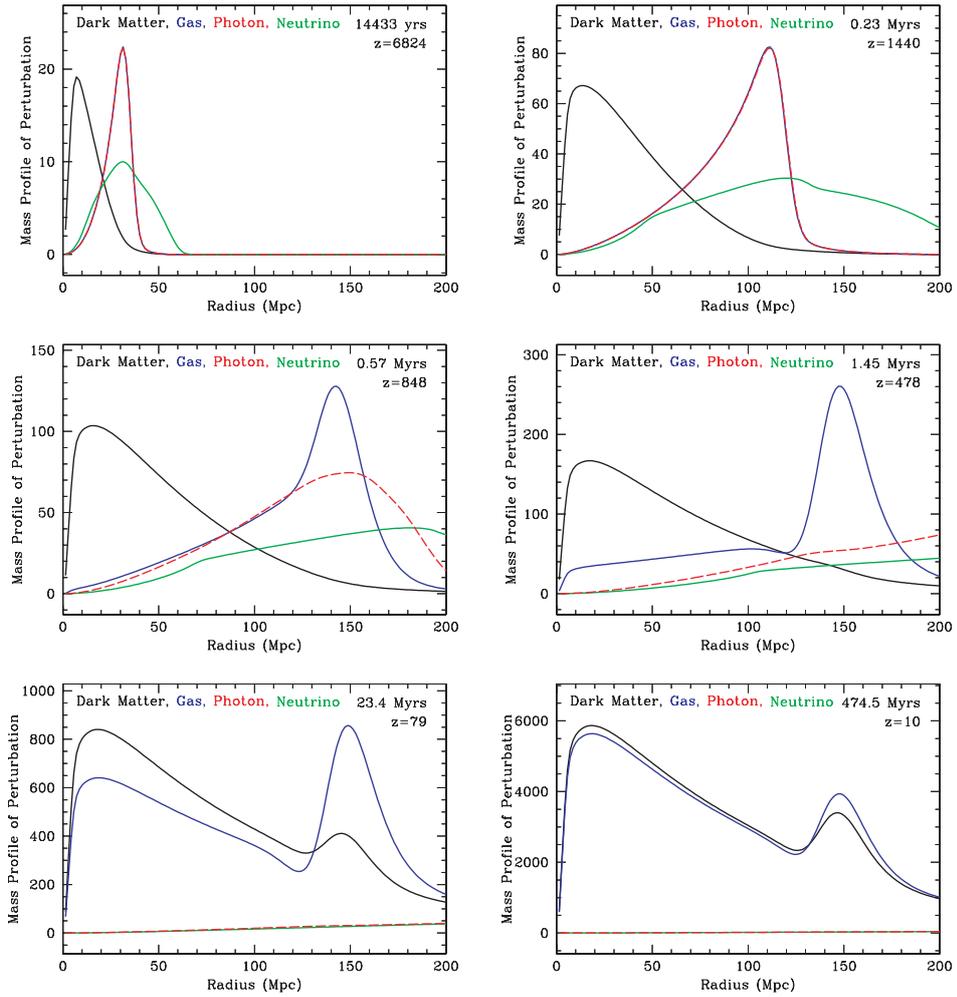


Figure 9: Snapshots of radial mass profiles vs comoving radius for various matter components. The simulation has been seeded with pointlike adiabatic overdensities at the origin at $t = 0$ yrs, the contrasts are evolving and interacting, and demonstrating the imprint of BAO onto the dark matter. From Eisenstein, Seo, and White (2007)

around $\ell \sim 220$, which corresponds to an angular scale of roughly one degree – see Figure 3. Figure 9 displays the interaction between different components before and after recombination: in the first row of panels, matter and baryonic gas are tightly coupled together, and jointly propagate outwards. Once the photons can stream freely (middle panels), an overdensity in the baryonic component remains, shaped like a spherical shell. The gravitational interaction between dark and baryonic matter imprints this structure onto the dark matter species (lower panels). In the subsequent formation of structure via gravitational collapse, the sound horizon feature persists to lower redshift. Indeed, they also show up in the transfer functions (Eisenstein and Hu, 1998), and the oscillations in the dark matter power spectra of Figure 4 are of baryon-acoustic origin – they vanish once the baryon fraction Ω_b is set to zero. Since baryons trace the dark matter distribution, the correlation excess on the acoustic scale will be transferred to the

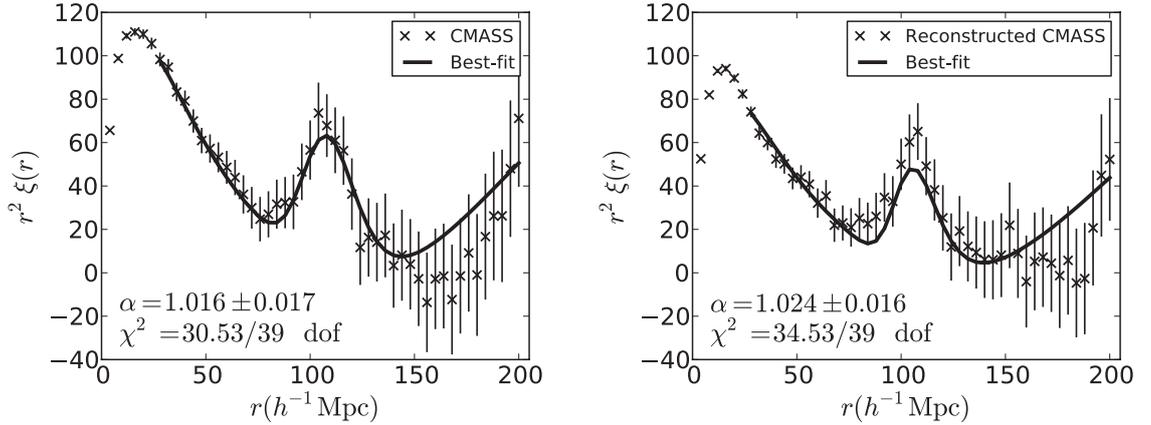


Figure 10: Galaxy autocorrelation function of the BOSS CMASS data set. The BAO peak is clearly visible. *Right panel:* it is possible to reconstruct the linear matter field from a non-linearly evolved field of galaxies by tracing their proper motions (see Eisenstein, Seo, Sirko, et al., 2007; D. H. Weinberg et al., 2013). This improves the measurement precision by reversing the non-linear evolution degrading the acoustic feature. From Anderson et al. (2012)

distribution of galaxies, and shows up as a bump in the galaxy correlation function, or as oscillations in the power spectrum. It should be noted that the galaxy density contrast δ_g need not be a faithful tracer of the underlying dark matter density contrast δ_c , or even the baryon density δ_b ; this phenomenon is known as *galaxy bias*. It can shift the location of the BAO peak by almost up to one percent, and needs to be taken into account – see Padmanabhan and White (2009) and Mehta et al. (2011) for theoretical and numerical investigations.

The first observations of this feature in galaxy clustering data were by the Sloan Digital Sky Survey (SDSS) (Eisenstein, Zehavi, et al., 2005) and almost simultaneously in the 2-degree field galaxy redshift survey (2dFGRS) data (Cole, Percival, et al., 2005). It has since been observed by its successor collaboration, the 6-degree field galaxy redshift survey (6dFGRS) (Beutler et al., 2011); by the WiggleZ Dark Energy Survey (Blake et al., 2011); by the subsequent SDSS data releases DR5 (Percival, Cole, et al., 2007) and DR7 (Percival, Reid, et al., 2010), and by the Baryon Oscillation Spectroscopic Survey (BOSS), which is a part of SDSS-III DR12 data release (Dawson, Schlegel, et al., 2013; Alam et al., 2016).

Since the sound horizon is a feature present across a wide range of redshifts, it can be used to constrain cosmological parameters (Eisenstein and Hu, 1998). Its comoving size of about 150 Mpc is fixed, and functions as a *standard ruler*: by observing the angular size of the correlation feature on the sky at different redshifts, we can infer the cosmic expansion history by mapping the redshift dependence of the angular diameter distance Equation 14, which varies with cosmology, particularly Ω_K . Since we are mapping galaxy

clustering at relatively low redshift, we expect to constrain dark energy and the Hubble parameter; also, the amplitude of the correlation excess depends on Ω_b/Ω_c .

Future BAO programmes will extend the surveyed comoving volume by covering a larger fraction of the sky and going to deeper redshifts; they will also focus on more galaxy populations of interest. Upcoming spectroscopic surveys include eBOSS (Dawson, Kneib, et al., 2016), DESI (Levi et al., 2013), PFS (Sugai et al., 2012), as well as the space-based missions *Euclid* (Laureijs et al., 2011) and *WFIRST* (Green et al., 2012).

STATISTICAL INFERENCE FROM DATA

High on a rocky promontory sat an Electric Monk on a bored horse.

From under its woven cowl the Monk gazed unblinkingly into another valley, with which it was having a problem. [...]

The Electric Monk was a labour-saving device, like a dishwasher or a video recorder. [...] Electric Monks believed things for you, thus saving you what was becoming an increasingly onerous task, that of believing all the things the world expected you to believe. [...]

The problem with the valley was this. The Monk currently believed that the valley and everything in the valley and around it, including the Monk itself and the Monk's horse, was a uniform shade of pale pink.

*Douglas Adams
Dirk Gently's Holistic Detective Agency*

2.1 BAYESIAN INFERENCE

It is integral to the scientific method to acknowledge that we are not searching for eternal certainty, and that we are not capable of acquiring complete knowledge about the Real World, even in principle. When gathering information from it, we aspire to structure it to infer some of the Real World's properties, and thus to learn about it – without ever proclaiming that what we are gathering fulfils our notion of absolute truth – without ever immunising ourselves against having to change our beliefs and theories in the future: the history of science, which so far toppled every such attempt at a Theory of Everything™, has taught the majority of scientists otherwise (Kuhn, 1970). Therefore, classical logic, which handles statements that are either True or False and nothing in between, is of limited use to the inference-based enterprise of science: when we put forward statements about the Real World, we seek to compare their quality: how *plausible* are they, in the face of the (limited) information we have gathered so far, and how does their plausibility change once new information is added? This humble approach to science, *inductive inference*, has proven tremendously successful.

The idea of using probabilities to quantify degrees of higher or lower plausibility of

a certain statement was advocated by Bayes and Price (1763), and is widely known as *Bayesianism*, although it has been disputed that Bayes was the first to publish the idea that now bears his name (Stigler, 1983). This interpretation of the term “probability” has later been rediscovered and substantially refined by de Laplace (1774, 1785, 1814). Modern treatments of Bayesian inference in science are Jeffreys (1961), Jaynes and Bretthorst (2003), and Gelman et al. (2013), see also the lecture notes by Caticha (2008), and the introduction written specifically for cosmologists by Trotta (2008).

There are different philosophical standpoints towards the precise definition and meaning of “plausibility” within the Bayesian framework – see Hacking (2001). We shall follow most data scientists in accepting the *objective Bayesian* stance: plausibility is a degree of rational belief which can only be changed by new information and not by introspection or revelation; two individuals beginning with the same assumptions, facing the same information, will end up with the identical sets of plausibilities. A seminal result is due to R. T. Cox (1946): if we make several weak assumptions on our way of consistently assigning plausibilities to statements, it will coincide with standard probability theory. Specifically, we wish these to be represented by positive real numbers $\omega(A|Z)$ – the plausibility of statement A , assuming another statement Z^1 as given. We now postulate

- that the plausibility of \bar{A} , the negation of statement A , only depends on the plausibility of A under the same assumptions: $\omega(\bar{A}|Z) = f[\omega(A|Z)]$,
- that the joint plausibility of statements A and B only depends on the plausibility of A , and the plausibility of B , given A : $\omega(A, B|Z) = g[\omega(A|Z); \omega(B|A, Z)]$.

Then, according to R. T. Cox (1946), the following can be shown: there is a bijection from $\omega(A|Z)$ to an equivalent set of real-valued plausibilities $\mathcal{P}(A|Z)$ – *Bayesian probabilities* – such that

- absolute certainty is represented by $\mathcal{P}(A|Z) = 1$, absolute falsehood by $\mathcal{P}(A|Z) = 0$. Statements of this kind obey Aristotelian logic and thus form a Boolean algebra.
- the “sum rule” of negation: $\mathcal{P}(A|Z) + \mathcal{P}(\bar{A}|Z) = 1$
- the “product rule” of conjunction:

$$\mathcal{P}(A, B|Z) = \mathcal{P}(A|Z) \mathcal{P}(B|A, Z) = \mathcal{P}(B|Z) \mathcal{P}(A|B, Z).$$

This means that the powerful machinery of classical probability theory (Kolmogorov, 1933) is now completely at the disposal of inductive inference: a belief state on a certain set of statements Ω is mathematically described by a *probability space*, i. e. a measure space $(\Omega, \mathcal{S}, \mathcal{P})$ with $\mathcal{P}(\Omega) = 1$. Furthermore, n -dimensional real random variables are measurable functions $\underline{X} : \Omega \rightarrow \mathbb{R}^n$.

¹ It is customary to always write these as conditional plausibilities, since realistically every inference always hinges upon fundamental underlying assumptions – e. g. a certain model framework.

Returning to the problem of inference about the Real World: let us assume a certain model \mathcal{M} (this could be, e. g., flat Λ CDM) which allows for different statements (or theories) Θ , all of which are inside \mathcal{M} in the sense that they make additional assumptions – e. g. certain values for the density parameters Ω_i and the Hubble parameter H_0 . We begin with a certain belief state $\mathcal{P}(\Theta|\mathcal{M})$, and wish to incorporate new information in the form of a data set \mathcal{D} . The result will be an updated belief state $\mathcal{P}(\Theta|\mathcal{D}, \mathcal{M})$, and the learning step from one to the other is referred to as *Bayesian updating*. The product rule of conjunction provides a recipe how to jump from one to the other:

Theorem 1 (Bayes' Theorem)

$$\mathcal{P}(\Theta|\mathcal{D}, \mathcal{M}) = \frac{\mathcal{P}(\mathcal{D}|\Theta, \mathcal{M}) \mathcal{P}(\Theta|\mathcal{M})}{\mathcal{P}(\mathcal{D}|\mathcal{M})}. \quad (42)$$

In standard probability theory this is just a trivial consequence of the definition of conditional probability, but adding the Bayesian interpretation it forms the cornerstone for the framework of inductive inference: it describes how we transform information into knowledge. Each ingredient has name and meaning:

- the *prior* distribution $\pi(\Theta) = \mathcal{P}(\Theta|\mathcal{M})$ on the space of theories,
- the *posterior* distribution $\Pi(\Theta) = \mathcal{P}(\Theta|\mathcal{D}, \mathcal{M})$ after updating,
- the *likelihood* $\mathcal{L}(\mathcal{D}|\Theta) = \mathcal{P}(\mathcal{D}|\Theta, \mathcal{M})$ is modelling the full process by which the data set \mathcal{D} was produced, and its dependence on the specific theory Θ ,
- the *model evidence* $E(\mathcal{M}) = \mathcal{P}(\mathcal{D}|\mathcal{M}) = \int d\Theta \mathcal{P}(\mathcal{D}|\Theta, \mathcal{M}) \mathcal{P}(\Theta|\mathcal{M})$ is, for the time being, just a factor ensuring that the posterior is properly normalised. However, in [Section 2.1.2](#) this number will prove to be crucial for comparing different models \mathcal{M}_1 and \mathcal{M}_2 .

Criticism of the Bayesian inference paradigm often attacks the incorporation of priors as subjective and dependent on the particular whims of the researcher choosing one set of prior beliefs over another. It should be noted that priors are a feature and an integral part of the whole endeavour: in situations where the data set is small, the prior will influence the posterior significantly, compared to the likelihood (this may change once more data accumulates). Then, there are good explicit reasons for choosing one distribution over another – one may want to model the outcome of another experiment as knowledge which we now attempt to build on (*informative* prior). This allows the combination of multiple data sets to merge the information inside them. Often, only certain ranges of parameters are physically sensible – particle masses need to be non-negative numbers; often, model assumptions hinge upon certain parameters not leaving a certain interval – all this can be enforced by setting the prior distribution outside of the desirable region

to zero. If Θ is confined to a finite volume, a uniform distribution can be used to model the lack of preference of any point in said volume over any other (*flat* prior). There are more sophisticated methods to construct priors modelling ignorance (*objective* priors, see Ghosh, Delampady, and Samanta, 2006), or priors which maximise information gain from the data set (*reference* priors, see Berger, Bernardo, and Sun, 2009), or priors which maximise the ignorance before taking the data (*entropic* priors, see Skilling, 1989). Since any data analysis step is only complete with data set and likelihood in conjunction with a specific prior, we can see that there is nothing perniciously subjective in the Bayesian paradigm, since jointly they uniquely determine the posterior distribution – the scientist is ultimately accountable for selection and explicit reporting of the prior they have used.

2.1.1 Parameter Estimation

For the practising Bayesian cosmologist, theory spaces tend to be bounded or unbounded domains in \mathbb{R}^n , and the distribution laws (prior and posterior) on the parameter vector $\underline{\Theta}$ are probability densities with respect to the Lebesgue measure. Bayesian updating through a data set will result in a multivariate distribution as the outcome of an observation or experiment, and it is this full distribution which is to be reported by the scientist – see Section 2.2 for the state-of-the-art methods to do this. Often, one is interested only in the *expectation value*

$$\mathbb{E}[\Theta_i] = \langle \Theta_i \rangle = \int d\Theta \Theta_i \Pi(\underline{\Theta}), \quad (43)$$

and the *variance* (square of the *error bar*)

$$\text{Var}[\Theta_i] = \sigma_{\Theta_i}^2 = \langle (\Theta_i - \langle \Theta_i \rangle)^2 \rangle = \langle \Theta_i^2 \rangle - \langle \Theta_i \rangle^2, \quad (44)$$

for each of the parameters under the posterior density $\Pi(\underline{\Theta})$, since these numbers give us an idea of the location and typical width of the distribution in parameters space. Nevertheless, it is important to take note that the full distribution $\Pi(\underline{\Theta})$ contains more information – about the cross-covariances of Θ_i and Θ_j ($i \neq j$):

$$\text{Cov}[\Theta_i \Theta_j] = \langle (\Theta_i - \langle \Theta_i \rangle)(\Theta_j - \langle \Theta_j \rangle) \rangle, \quad (45)$$

and about all the higher moments $\langle \Theta_i \Theta_j \cdots \Theta_k \rangle$. These numbers quantify, e.g., the skewness and the kurtosis of the distribution, as well as the long-distance decay behaviour – everything that is not contained in a quadratic expansion of $\ln \Pi(\underline{\Theta})$ around its maximum.

For illustrative purposes, it is useful to plot *credible regions* of posterior distributions – often called *contours*, in slight abuse of terminology. These are the surfaces of constant probability density $\Pi(\underline{\Theta})$ which happen to enclose a given probability mass $0 < p \leq 1$. Custom values are $p = 68.3\%$, 95.4% and 99.73% (*one/two/three sigma*) – see Figure 11 for credible regions of an example posterior density on the cosmological parameters $\underline{\Theta} = (\Omega_m, \sigma_8)$. These numbers stem from the paradigmatic *standard Normal distribution* which has the density

$$\mathcal{N}(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right) \quad (46)$$

for $\theta \in \mathbb{R}$. It has vanishing mean and unit variance, and the concentric intervals $[-n, n]$ contain the above probability masses of $p = 0.683, 0.954, 0.997$ for $n = 1, 2, 3$.

Its generalisation to arbitrary location, spread, and correlation is the *multivariate Gaussian distribution*, which has the property that it is fully quantified by its mean vector $\underline{\mu} = \langle \underline{\Theta} \rangle$ and its variance-covariance matrix $\Sigma = (\Sigma_{ij}) = (\text{Cov}[\Theta_i, \Theta_j])$:

$$\mathcal{G}(\underline{\Theta}; \underline{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left[-\frac{1}{2}(\underline{\Theta} - \underline{\mu})^T \Sigma^{-1}(\underline{\Theta} - \underline{\mu})\right]. \quad (47)$$

Its credible regions are always the shape of ellipsoids with coincident axes – ostensibly the 2D example in Figure 11 (pink) shows significant deviations from a Gaussian. The analysis of Heymans et al. (2013b) actually includes five parameters into its flat Λ CDM model: $\underline{\Theta} = (\Omega_m, \sigma_8, \Omega_b, h, n_s)$. The posterior on (Ω_m, σ_8) in Figure 11 has been found by *marginalisation*: if we are only interested in the distribution on a subset of parameters $\underline{\theta}$ instead of the full distribution of $\underline{\Theta} = (\underline{\theta}, \underline{\psi})$, we can integrate out the *nuisance parameters* $\underline{\psi}$:

$$\Pi_{\text{marg}}(\underline{\theta}) = \int d\psi \Pi(\underline{\theta}, \underline{\psi}). \quad (48)$$

This technique is one strength of the Bayesian approach to inference: often, it is necessary to include parameters into the analysis that are not of physical interest, but that model *systematics*, such as physical contaminations, calibration parameters, or properties of the measurement device. These can be included into the analysis in a consistent

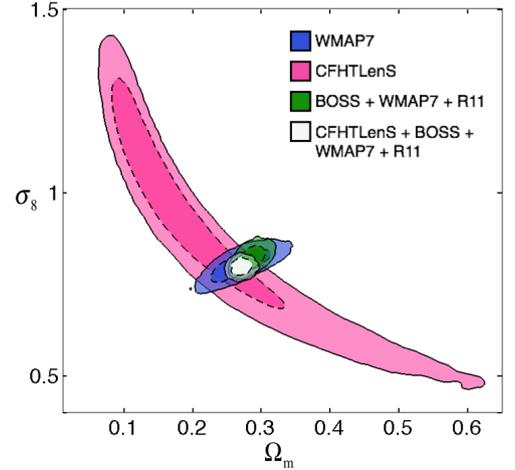


Figure 11: Example posterior density of the CFHTLenS tomography data set in combination with various other data sets. Shown are 68% and 95% credible regions. From Heymans et al. (2013b) – see there for details on the data.

manner, and afterwards removed via marginalisation. As an example, the analysis of *Planck* data included six cosmological parameters in the baseline flat Λ CDM model, and additionally 26 nuisance parameters quantifying foregrounds and calibrations (see Planck Collaboration, 2015b,c).

Another important feature of the distribution shown in Figure 11 is the presence of a near-degeneracy in the CFHTLenS-only data (pink). It is possible to move along a ridge of comparably plausible parameter values (high posterior density values), which spans a wide range for each of the parameters – hence the marginalised constraints on Ω_m and σ_8 on their own are not strong. Only a certain combination of these parameters – that which changes when moving across the ridge – is strongly constrained by the data. Note that this direction is the least well-constrained by the WMAP measurement of the CMB (blue, see Komatsu et al., 2011). Hence a combination of both data sets resolves the problem due to their mutual complementarity, and yields good marginal constraints for each of the parameters – such a complementarity is referred to as *lifting the degeneracy*.

2.1.2 Model Selection

Given the choice between two models describing the same observational phenomenon, but which are different in their basic tenets, which should be chosen to explain the underlying science? An early approach to the question stated that the model with the least amount of assumptions, or the simplest ones, is to be preferred, such that none of them is dispensable. This parsimony of hypotheses is still a vital element in scientific thinking, and is commonly attributed to the medieval monk William of Ockham (“Ockham’s Razor”), although the idea traces back to Aristotle (ca. 350 BCE). We can also use a modern Bayesian framework to address these questions: which of two models should we prefer when inferring from a data set? Is the extra complexity of one over the other justified, given the potential gain in predictivity? This shifts the focus from “least assumptions” to “most predictive assumptions”, it also carves out the notion of predictivity with more precision (Jeffreys, 1935, 1961; Kass and Raftery, 1995), see also Jaynes and Bretthorst (2003) and Gelman et al. (2013).

We proceed in the spirit of interpreting a Bayesian probability as a quantifier of the plausibility of an assertion: instead of comparing the plausibility of several theories Θ_i , given one model \mathcal{M} , we can ask for the plausibility of different models \mathcal{M}_i in the face of data \mathcal{D} – rephrasing Equation 42 on the set of models:

$$\frac{\mathcal{P}(\mathcal{M}_i|\mathcal{D})}{\mathcal{P}(\mathcal{M}_j|\mathcal{D})} = \frac{\mathcal{P}(\mathcal{D}|\mathcal{M}_i) \mathcal{P}(\mathcal{M}_i)}{\mathcal{P}(\mathcal{D}|\mathcal{M}_j) \mathcal{P}(\mathcal{M}_j)} = \frac{E(\mathcal{M}_i) \mathcal{P}(\mathcal{M}_i)}{E(\mathcal{M}_j) \mathcal{P}(\mathcal{M}_j)}. \quad (49)$$

Hence, stepping from the prior odds ratio $\mathcal{P}(\mathcal{M}_i)/\mathcal{P}(\mathcal{M}_j)$ to the posterior odds ratio $\mathcal{P}(\mathcal{M}_i|\mathcal{D})/\mathcal{P}(\mathcal{M}_j|\mathcal{D})$, the crucial quantity is the ratio of evidences $B_{ij} = E(\mathcal{M}_i)/E(\mathcal{M}_j)$,

$ \ln B_{21} $	odds ratio	interpretation
0 – 1	< 3 : 1	inconclusive
1 – 2.5	~ 3 : 1	substantial evidence
2.5 – 5	~ 12 : 1	strong evidence
> 5	> 150 : 1	decisive evidence

Table 1: Jeffreys’ scale for translating Bayes factors into verdicts.

known as the *Bayes factor*.

Each model \mathcal{M}_k has its own parameter space $\{\Theta_k\}$. The evidence is an integral over the non-normalised posterior

$$E(\mathcal{M}_k) = \int d\Theta_k \mathcal{P}(\mathcal{D}|\Theta_k, \mathcal{M}_k) \mathcal{P}(\Theta_k|\mathcal{M}_k), \quad (50)$$

which explains the frequently-(ab)used term *marginal likelihood*. It can be computationally expensive or impossible to compute, if the parameter space is high-dimensional. There are several heuristic *information criteria*, which are approximations to the Bayesian evidence and can be easier to compute in specific situations (Liddle, 2007), but only the evidence can be justified from the information theoretic formalism laid out by Cox’ axioms.

Progress can be made in the context of nested models (Dickey, 1971; Verde, Feeney, et al., 2013), by what is known as the Savage-Dickey Density Ratio (SDDR). Assume that the parameters of one model \mathcal{M}_1 can be split apart: $\Theta_1 = (\Theta_2, \Psi)$. The lower-dimensional model \mathcal{M}_2 shall be a hypersurface in this parameter space specified by setting $\Psi = \Psi_0$ constant, and its coordinates are the parameters common to both models, namely Θ_2 . It can be shown that upon weak consistency requirements the evidence ratio can be expressed as

$$B_{21} = \frac{E(\mathcal{M}_2)}{E(\mathcal{M}_1)} = \frac{\mathcal{P}_{\text{marg}}(\Psi = \Psi_0|\mathcal{D}, \mathcal{M}_1)}{\mathcal{P}_{\text{marg}}(\Psi = \Psi_0|\mathcal{M}_1)}. \quad (51)$$

Here $\mathcal{P}_{\text{marg}}(\Psi|\mathcal{D}, \mathcal{M}_1)$ and $\mathcal{P}_{\text{marg}}(\Psi|\mathcal{M}_1)$ are posterior and prior density of \mathcal{M}_1 , marginalised down to only the extra parameters Ψ . This is intuitive: if the probability mass on the hypersurface $\Psi = \Psi_0$ increases during Bayesian updating, this will tilt our preference towards the submodel.

To judge different values for $\ln B_{21}$, the terminologies laid out in Table 1, or slight variations thereof, are in frequent use to relate betting odds ratios to their intuitive meaning and to the outcome of the model comparison (see Jeffreys, 1961; Kass and Raftery, 1995; Trotta, 2008; Verde, Feeney, et al., 2013). It should be noted, however, that this classification scheme is purely a matter of semantics – the numerical value speaks for itself.

2.2 SAMPLING

One main lesson from Bayesian data analysis is that the outcome of an inferential analysis is represented by a posterior probability distribution on the space of model parameters. This begs the question – how to report it to fellow scientists? In only the easiest cases can we expect to derive an analytic form for the density, for every realistic problem we will have to rely on finite-dimensional summary statistics (like means and covariances), which necessarily discard information about the full distribution. Plots of credible regions, as illustrative as they may be, do not allow a reconstruction of the posterior density.

Currently the most practical solution is to produce large samples, i. e. sets of points in parameter space that are (ideally) independent and identically distributed (**iid**) draws from the target distribution. In this case, the point density converges almost surely to the probability density (i. e. the set of cases on which this does not happen has probability measure zero); this is a consequence of the Strong Law of Large Numbers. We will present two modern methods to produce approximately **iid** samples in [Section 2.2.1](#) and [Section 2.2.2](#).

One frequent problem is the computation of expectation values of functions on parameter space. With an **iid** sample $\{\underline{X}_i\}_{i=1}^{\mathcal{N}}$ drawn from a target distribution $\Pi(\underline{X})$, the expectation value of any function $F(\underline{X})$ weighted with the same distribution can be approximated as

$$\langle F \rangle_{\Pi} = \int d\underline{X} \Pi(\underline{X}) F(\underline{X}) \simeq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} F(\underline{X}_i), \quad (52)$$

where convergence is guaranteed almost surely for $\mathcal{N} \rightarrow \infty$.

Another advantage of **iid** samples is facilitated marginalisation: assuming that $p : \underline{X} \mapsto \underline{x}$ projects the parameter vector \underline{X} down to a subset of its entries \underline{x} , then $\{p(\underline{X}_i)\}_{i=1}^{\mathcal{N}}$ is automatically a sample of the marginal distribution $\Pi_{\text{marg}}(\underline{x})$ – in practice this is easily achieved by discarding all coordinates of the sample points that we are not interested in. How to reconstruct the probability density from a sample thereof? Many approaches can be captured in the idea of *kernel density estimation* (**KDE**) (Rosenblatt, 1956; Parzen, 1962), which is in widespread use in cosmology². To smooth out the points $\{\underline{X}_i\}_{i=1}^{\mathcal{N}}$ into a continuous density, centre a unimodal null-centered probability density $K_h(\underline{X})$ of suitably chosen width h (*kernel*) onto each sample point, and sum up the contributions from all points:

$$\hat{\Pi}_h(\underline{Y}) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} K_h(\underline{Y} - \underline{X}_i). \quad (53)$$

² See, e. g. the popular GetDist package: <http://getdist.readthedocs.io>.

Popular choices for the kernel are multivariate Gaussian, uniform (top-hat), Epanechnikov (truncated parabola), and many more (Silverman, 1986). One main advantage of this approach is that it is *non-parametric*: no assumptions have to be made about the point distribution, and the resulting density estimate is entirely data-driven.

However, the KDE method has shortcomings: often the plots of credible regions as found via KDE are still ragged and require further smoothing. The evaluation of the estimator $\hat{\Pi}_h$ can be slow for large sample size. Its value still depends on the particular choice of kernel $K_h(\underline{X})$; so do continuity, differentiability and smoothness of the estimated density function. Also – it is vital to choose the kernel bandwidth h correctly; one common approach is to minimise the *mean integrated square error* (MISE), or a suitable approximation thereof. At each point \underline{Y} , the estimator $\hat{\Pi}_h(\underline{Y})$ has a distribution due to sampling scatter, with expectation value $\mathbb{E} [\hat{\Pi}_h(\underline{Y})]$ and variance $\text{Var} [\hat{\Pi}_h(\underline{Y})]$ – all of these are bandwidth-dependent. With some algebra we find that

$$\begin{aligned} \text{MISE}_h &= \mathbb{E} \int dY \left\{ \Pi(\underline{Y}) - \hat{\Pi}_h(\underline{Y}) \right\}^2 \\ &= \int dY \text{Var} [\hat{\Pi}_h(\underline{Y})] + \int dY \left\{ \Pi(\underline{Y}) - \mathbb{E} [\hat{\Pi}_h(\underline{Y})] \right\}^2. \end{aligned} \quad (54)$$

The MISE – the expected quadratic deviation from the truth – is composed of one term quantifying the variance of the density estimator and one for its squared bias – typically, choosing h too high will cause the bias term to grow, whereas a value too low increases the variance term. This trade-off between variance and bias means that there may be a compromise in between, but the density estimate – and hence the contours – will still be somewhat biased and noisy. Under additional assumptions – multivariate Gaussianity of the target distribution, multivariate Gaussian kernel – an analytic form for the ideal bandwidth is found to be

$$h_{\text{opt}} = \left[\frac{4}{\mathcal{N}(2d+1)} \right]^{\frac{1}{d+4}}, \quad (55)$$

where d is the dimension of parameter space. Even if the assumption regarding the target density is only approximately satisfied, it is often used as a rule of thumb, known as *Silverman's rule* (Silverman, 1986).

2.2.1 Markov Chain Monte Carlo (MCMC)

Drawing iid pseudo-random samples directly from multivariate random distributions is only possible for select densities with analytic expressions, e.g. uniform or Gaussian. For generic high-dimensional distributions, like most posterior densities encountered

in cosmology, more advanced strategies are necessary – of which **MCMC** is the modern gold standard, and in wide use in many areas of science (see Gilks, Richardson, and Spiegelhalter, 1995; Brooks et al., 2011; Gelman et al., 2013) – early applications of the technique in cosmology are Christensen et al. (2001) and Lewis and Bridle (2002).

Samples $\{\underline{X}_i\}_{i=1}^{\mathcal{N}}$ are produced by a random walk through parameter space, which follows the defining property of a *discrete-time Markov chain*: the distribution of a new point \underline{X}_{k+1} does only depend on the location of the last point \underline{X}_k and is given by $\mathcal{P}(\underline{X}_{k+1}|\underline{X}_k)$. If this distribution (*transition kernel*) is designed carefully, taking into account the target distribution $\Pi(\underline{X})$, then the point density will indeed converge in distribution against $\Pi(\underline{X})$ almost surely. For an ergodic Markov chain, the sufficient criterion for this to happen in the limit $\mathcal{N} \rightarrow \infty$ is the *detailed balance* property: for all points $\underline{X}, \underline{Y}$

$$\mathcal{P}(\underline{Y}|\underline{X})\Pi(\underline{X}) = \mathcal{P}(\underline{X}|\underline{Y})\Pi(\underline{Y}). \quad (56)$$

One family of sampling strategies that satisfy this condition by design is *Metropolis-Hastings* (see Metropolis et al., 1953; Hastings, 1970): to transition from one sample point \underline{X}_k to the next, iterate the following steps:

1. Draw a proposal point \underline{Y} from a jumping distribution $J(\underline{Y}|\underline{X}_k)$.
2. Consider the ratio of target densities between the current point and the proposal point, corrected by the possible asymmetry in the jumping distribution:

$$r = \frac{\Pi(\underline{Y}) J(\underline{X}_k|\underline{Y})}{\Pi(\underline{X}_k) J(\underline{Y}|\underline{X}_k)}. \quad (57)$$

Is $r \geq 1$?

Yes: accept the proposal, $\underline{X}_{k+1} = \underline{Y}$.

No: draw a random number $\alpha \sim \mathcal{U}[0; 1]$. Is $\alpha < r$?

Yes: accept the proposal, $\underline{X}_{k+1} = \underline{Y}$.

No: reject the proposal, $\underline{X}_{k+1} = \underline{X}_k$.

Intuitively, accepting points with lower target distribution values is what allows the chain to explore the full likelihood surface representatively, including the tails (this behaviour is known as *proper mixing*), instead of just hunting for the mode. The choice of the starting point does not matter if the iteration is set up properly: the chain will quickly move towards the support of the equilibrium distribution and “forget” about its initial state. It is customary to discard several hundred steps at the beginning, until convergence has been reached (*burn-in*). Further, the points of a converged chain will still be dependent: to provide an **iid** sample, it is necessary to decorrelate the chain by retaining only every k th point, where k is an integer larger than the typical correlation

length (a common heuristic is $k = 10$).

The explicit choice of the jumping distribution requires proper care: if it is chosen too narrowly, then the chain will not be mixing; if chosen too widely, the algorithm becomes ineffective due to high rejection rates. It is custom practice to start several chains with over-dispersed starting points, to get an idea of the typical width of the target distribution, and to use that knowledge to adapt the width of the jumping kernel. There exist several tools to judge whether proper convergence and mixing have been achieved after a finite number of steps, which is a major challenge – especially in high-dimensional parameters spaces, and for likelihood surfaces with peculiar shapes or multiple modes (see Cowles and Carlin, 1996, for a comparative review).

Modern precision cosmology has profited immensely from implementing MCMC as the standard scientific tool to transfer and report probability distributions. The full data products of the *Planck* satellite are freely available online in the *Planck Legacy Archive* (<http://pla.esac.esa.int/pla/#home>) – the full grid of chains for all models and data sets considered are about 13 GB in volume (Planck Collaboration, 2015a,c). See also the Legacy Archive for Microwave Background Data Analysis (<http://lambda.gsfc.nasa.gov/>) providing chains for an eclectic collection of data sets and models.

Software implementations of MCMC that are in widespread use in cosmology include the following:

- CosmoMC: Lewis and Bridle (2002), see <http://cosmologist.info/cosmomc/>;
- AnalyzeThis!: Doran and Müller (2004), see <https://github.com/EdoardoCarlesi/cmbeasy>;
- emcee: Foreman-Mackey et al. (2013), see <http://dan.iel.fm/emcee/current/>;
- its cosmology wrapper CosmoHammer: Akeret et al. (2012), see <https://github.com/cosmo-ethz/CosmoHammer>;
- Monte Python: Audren et al. (2013), see https://github.com/audren/montepython_public/.

2.2.2 Population Monte Carlo (PMC)

An alternative sampling algorithm, PMC, is based on *importance sampling*: if we have a probability density $q(\underline{X})$ which is a reasonable approximation to the target density $\Pi(\underline{X})$, and which is easy to sample, it is possible to reweight an iid sample $\{X_i\}_{i=1}^N$ of q to produce a *weighted sample* $\{(\underline{X}_i, w_i)\}_{i=1}^N$ of Π . This works because of the observation that

the expectation value of any function $F(\underline{X})$ under q can be found directly by averaging over the sample

$$\langle F \rangle_q = \int dX q(\underline{X}) F(\underline{X}) \simeq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} F(\underline{X}_i), \quad (58)$$

hence the expectation value under Π

$$\langle F \rangle_{\Pi} = \int dX \Pi(\underline{X}) F(\underline{X}) = \int dX q(\underline{X}) \frac{\Pi(\underline{X})}{q(\underline{X})} F(\underline{X}) \simeq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} w_i F(\underline{X}_i) \quad (59)$$

can be expressed as a weighted average of F over the same points, provided we set the weights to $w_i = \Pi(\underline{X}_i) / q(\underline{X}_i)$. [PMC](#) is an adaptive importance sampling scheme, which aims to iteratively approximate the target density by a *Gaussian mixture model*, i.e. a convex combination of D multivariate Gaussian probability densities

$$q^t(\underline{X}) = \sum_{d=1}^D \lambda_d^t \mathcal{G}(\underline{X}; \underline{\mu}_d^t, \Sigma_d^t), \quad (60)$$

each with their individual mean vector $\underline{\mu}_d^t$ and covariance matrix Σ_d^t ; the component weights $\lambda_d^t \geq 0$ are required to satisfy the normalisation constraint $\sum_d \lambda_d^t = 1$. The quantities $(\lambda_d^t, \underline{\mu}_d^t, \Sigma_d^t)$ are being updated in every iteration step $t = 1, 2, \dots, t_{max}$. The objective of the iteration is to minimise the *Kullback-Leibler divergence* (also *relative entropy*) between both distributions Π and q :

$$D_{KL}(\Pi || q) = \int dX \Pi(\underline{X}) \ln \frac{\Pi(\underline{X})}{q(\underline{X})}. \quad (61)$$

This is a frequently used measure for the dissimilarity of two probability distributions, which has the properties of being positive definite – $D_{KL}(p||q) \geq 0$, (and = 0 if and only if $p \equiv q$ p -almost everywhere) – and usually asymmetric – $D_{KL}(p||q) \neq D_{KL}(q||p)$. For details on, and variants of, the optimisation procedure, see Cappé, Guillin, et al. (2004), Cappé, Douc, et al. (2007), Douc et al. (2007a,b), and Wraith et al. (2009).

As a byproduct, the [PMC](#) algorithm produces an estimate of the model evidence: if we sample the posterior distribution $\Pi(\underline{X}) = \mathcal{P}(\underline{X}|\mathcal{D}, \mathcal{M})$ of data set \mathcal{D} in model \mathcal{M} , then

$$E(\mathcal{M}) \simeq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} w_i, \quad (62)$$

where the precision of the estimate depends on the closeness of q and Π (Kilbinger, Wraith, et al., 2010). A common diagnostic for convergence of proposal density $q^t(\underline{X})$ is

the *perplexity*, a quantity derived from the Shannon entropy H^t of the distribution of the normalised weights \bar{w}_i^t :

$$p = \frac{1}{\mathcal{N}} \exp[H^t]; \quad H^t = - \sum_{i=1}^{\mathcal{N}} \bar{w}_i^t \ln \bar{w}_i^t; \quad \bar{w}_i^t = \frac{w_i^t}{\sum_{k=1}^{\mathcal{N}} w_k^t}. \quad (63)$$

In the ideal case where $q^t \equiv \Pi$, all weights $\bar{w}_i^t = 1/\mathcal{N}$, and $p = 1$; in practice, a value of $p > 0.7$ will result in reliable sampling of the posterior density.

A popular cosmology sampler using PMC is CosmoPMC (Wraith et al., 2009; Kilbinger, Benabed, et al., 2011) – see <http://www2.iap.fr/users/kilbinge/CosmoPMC/> for source code, and Kilbinger, Wraith, et al. (2010), Heymans et al. (2013b), and Kilbinger, Fu, et al. (2013) for applications of CosmoPMC to CFHTLenS and other data sets.

2.3 NON-GAUSSIAN RANDOM VARIABLES

The Gaussian probability density (Equation 47) is one of the most fundamental building blocks of stochastics and omnipresent in probabilistic inference. This stems from two fundamental reasons: for a real-valued random variable, the Gaussian probability density maximises the *differential entropy* functional

$$H[p] = -\langle \ln p \rangle_p = \int dx p(x) \ln p(x) \quad (64)$$

under the constraints of fixed mean $\mu = \langle x \rangle$ and fixed variance $\sigma^2 = \langle x^2 \rangle - \mu^2$. This functional is the generalisation of Shannon entropy for a continuous variable x (Shannon, 1948); it is commonly interpreted as measuring the amount of information about x , where high entropy corresponds to high uncertainty, or lack in information (Cover and J. A. Thomas, 2006; Caticha, 2008). In this sense, the Gaussian distribution is the most economical assumption for a probability density of which only location and width are known.

The other reason for the importance of the Normal distribution is the *Central Limit Theorem* (see de Moivre, 1738; de Laplace, 1812; Lyapunov, 1901; Lindeberg, 1922; Lévy, 1925; Turing, 1938, or any introductory text on probability theory): Let X_1, X_2, \dots be a sequence of iid real random variables with mean μ and variance $\sigma^2 < \infty$. Let

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (65)$$

be the average value of $\{X_1 \dots X_n\}$. Then, as $n \rightarrow \infty$, the random variable $\sqrt{n} \frac{S_n - \mu}{\sigma}$ will converge in distribution towards a Normal variate with mean zero and variance one.

Inference often deals with the sum or average of iid variates – e.g. the log-likelihood

$\mathcal{L}(\mathcal{D}|\theta)$ for a data set $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$, with theoretical model $Y_\theta(X)$. If every measurement Y_i of the quantity Y has an error bar σ_i , then a common ansatz to assemble the log-likelihood is

$$\mathcal{L}(\mathcal{D}|\theta) \propto \sum_{i=1}^N \frac{[Y_i - Y_\theta(X_i)]^2}{\sigma_i^2}. \quad (66)$$

If there are enough data points, and as long as the measurements are independent, the shape of the likelihood will be asymptotically Gaussian in the data even if the distribution of each measurement on its own is not. Additionally, if the model $Y_\theta(X)$ is linear in θ , the likelihood will also have a Gaussian dependence on the parameters θ . However, often enough the latter assumption is not satisfied, and random variates of cosmological interest are not well described by Gaussian variables (see, e. g., [Figure 11](#)); in these situations recording only the first two cumulants, i. e. sample mean and sample variance, will discard information about the distribution.

2.3.1 Classical Gaussianity Tests

Given a sample $\mathcal{D} = \{x_1, \dots, x_n\}$ from a probability distribution (i. e. a data set), has it been drawn from a Gaussian density? The first methods to address this question were formulated in the framework of *classical hypothesis testing*, which is a frequentist approach to statistical decision theory developed and popularised by Fisher (1925) and Neyman and Pearson (1933); for the dispute between the creators about the role of hypotheses, see e. g. Fisher (1958), Lehmann (1993), and Lenhard (2006), and also Kendall et al. (1999) for the modern consensus synthesizing their work.

Classical tests of Gaussianity proceed to compare the hypotheses

H_0 (*null hypothesis*): the distribution is Gaussian –

H_a (*alternative hypothesis*): the distribution is not Gaussian –

in the face of the data set \mathcal{D} . The decision is made via the *p-value* – i. e. the probability that a result as extreme as \mathcal{D} (or more extreme) can occur assuming H_0 is true. If this probability is less than a fixed value α that has been set before performing the test, the null hypothesis is rejected in favour of the alternative hypothesis. The number α is known as the *significance level* of the test, a commonplace heuristic is $\alpha = 0.05$.

The framework laid out by hypothesis testing and *p-values* is in widespread use across areas of science as manifold as biology, sociology, medicine and epidemiology, pharmacology, economics, criminology, political science, and psychology, and is not without controversy – partly because of the arbitrariness of the choice of α , partly because of epidemic misunderstandings about their interpretation (e. g. it is often wrongly

treated as “the probability that H_0 is true”), partly because of various other logical, methodological, statistical, and philosophical concerns – see, e. g., Bakan (1966), Goodman (1992), Schervish (1996), Hunter (1997), Goodman (1999), Ioannidis (2005), Murdoch, Tsai, and Adcock (2008), Lambdin (2012), Trafimow and Marks (2015), and Wasserstein and Lazar (2016) as well as many other contributions³.

We will focus on one Gaussianity test in detail: Shapiro and Wilk (1965) proposed the statistic W in order to test a univariate sample for Gaussianity. Given $\mathcal{D} = x_1, \dots, x_n$, define the vector containing x_i as entries, but in ascending order: $\underline{y} = [y_1, \dots, y_n]^T$, such that $y_1 \leq y_2 \leq \dots \leq y_n$, and $\{x_i\} = \{y_i\}$. Then,

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (67)$$

where $\bar{y} = \sum y_i / n$ is the sample mean, and the weight vector \underline{a} is defined via

$$\underline{a} = \frac{V^{-1} \underline{m}}{\sqrt{\underline{m}^T V^{-1} V^{-1} \underline{m}}}. \quad (68)$$

Here, \underline{m} and V are mean vector and covariance matrix of the order statistics for a sample of length n , drawn from a univariate Normal distribution; these can be approximated in closed form (Royston, 1992, 1993). These publications also present expressions for the distribution of the W statistic, provided that the null hypothesis holds; this is needed for the computation of the p -value.

The Shapiro-Wilk test has been compared to other classical tests of Gaussianity (Stephens, 1974; Pearson, D’Agostino, and Bowman, 1977; Razali and Wah, 2011). In terms of statistical power, i. e. the probability of rejecting the null-hypothesis under the assumption that it is false, it outperforms all other common Gaussianity test, including Kolmogorov-Smirnov, Anderson-Darling, D’Agostino, and Lilliefors; hence, it is generally accepted as the strongest classical test of Gaussianity.

2.3.2 Random Fields

Random fields can be thought of as the infinite-dimensional extension of random variables; or equivalently, as the multi-dimensional generalisation of stochastic processes in the time domain. They are frequently applied in statistical cosmology – see Bardeen et al. (1986), and also Adler and J. E. Taylor (2007) for a rigorous mathematical treatment on arbitrary Riemannian manifolds.

³ Accessible, non-technical discussions emphasizing the common practice and its impact on experimental research include <http://www.nature.com/news/scientific-method-statistical-errors-1.14700/>, <http://fivethirtyeight.com/features/science-isnt-broken/>, and also <http://www.youtube.com/watch?v=42QuXLuch3Q>.

Intuitively, a random field in d dimensions is a way to attach a real or complex random variable to each point in \mathbb{R}^d , such that each realisation of the field configuration is sufficiently regular (e. g. has a Fourier transform).

More precisely: let $(\Omega, \mathcal{S}, \mathbb{P})$ be a probability space. A *random field* on an open set $\Gamma \subseteq \mathbb{R}^d$ is a map $\zeta : \Gamma \times \Omega \rightarrow \mathbb{R}$ or \mathbb{C} ; $(\underline{x}, \omega) \mapsto \zeta_\omega(\underline{x})$. At each point $\underline{x} \in \Gamma$ we require the field strength $\zeta(\underline{x}) \equiv \zeta_\bullet(\underline{x})$ to be a random variable, this is satisfied if $\zeta_\bullet(\underline{x}) : \Omega \rightarrow \mathbb{R}$ or \mathbb{C} is a \mathbb{P} -measurable function. Further, we impose regularity of each realisation of the random field: $\forall \omega \in \Omega : \zeta_\omega(\bullet) \in L^1(\Gamma) \cap L^\infty(\Gamma)$ – this ensures that the Fourier transform of each random field realisation exists, and is again a realisation of another random field.

A common way to characterise random fields is via their *n-point distributions* – given finitely many points $\underline{x}_1, \dots, \underline{x}_n \in \Gamma$, they are defined as

$$\mathcal{P}_n[\zeta(\underline{x}_1) = \zeta_1, \dots, \zeta(\underline{x}_n) = \zeta_n] = \mathbb{P}(\{\omega \in \Omega : \zeta_\omega(\underline{x}_1) = \zeta_1, \dots, \zeta_\omega(\underline{x}_n) = \zeta_n\}), \quad (69)$$

Thinking as a random field as an uncountably-infinite collection of random variables indexed by a continuous variable \underline{x} , these are the marginal distributions of a finite number of points: fixing the field strengths $\zeta(\underline{x}_i)$ to the value ζ_i , what is the probability mass of all realisations satisfying this constraint? Effectively, this integrates the distribution of $\zeta(\underline{x})$ over the infinite number of points in $\Gamma \setminus \{\underline{x}_1 \dots \underline{x}_n\}$, leaving us with a probability distribution on a finite-dimensional space.

Another important quantity containing information about the correlation structure of the random field are the associated *n-point functions* – these are the generalisation of the moments of a finite-dimensional random vector:

$$\mathcal{F}_n(\underline{x}_1, \dots, \underline{x}_n) = \langle \zeta(\underline{x}_1) \cdots \zeta(\underline{x}_n) \rangle = \int_{\Omega} d\mathbb{P}(\omega) \zeta_\omega(\underline{x}_1) \cdots \zeta_\omega(\underline{x}_n).$$

A random field is *statistically homogeneous* (or *stationary*) if all *n-point functions* are shift-invariant;

$$\text{for every shift vector } \underline{r} : \mathcal{F}_n(\underline{x}_1 + \underline{r}, \dots, \underline{x}_n + \underline{r}) = \mathcal{F}_n(\underline{x}_1, \dots, \underline{x}_n), \quad (70)$$

and *statistically isotropic* if they are rotation-invariant;

$$\text{for every rotation matrix } R : \mathcal{F}_n(R\underline{x}_1, \dots, R\underline{x}_n) = \mathcal{F}_n(\underline{x}_1, \dots, \underline{x}_n). \quad (71)$$

A random field is *Gaussian* if all *n-point distributions* \mathcal{P}_n are Gaussian distributions, i. e. with densities of the form [Equation 47](#). It should be noted that this definition has

practical difficulties, since it requires control over an infinite set of n -point distributions. Transforming the random field $\zeta(\underline{x})$ into the Fourier domain yields

$$\zeta_\omega(\underline{\ell}) = \int d^d x \zeta_\omega(\underline{x}) \exp[-i \underline{\ell} \cdot \underline{x}] \Leftrightarrow \zeta_\omega(\underline{x}) = \frac{1}{(2\pi)^d} \int d^d k \zeta_\omega(\underline{\ell}) \exp[+i \underline{\ell} \cdot \underline{x}]. \quad (72)$$

We will denote a function $f(\underline{\theta})$ and its Fourier transform $f(\underline{\ell})$ by the same letter throughout this work; the argument will always clarify the context. The Fourier modes $\zeta(\underline{\ell})$ are again a set of random variables labelled by the wave vector $\underline{\ell}$, hence they form a random field on wave number space. If $\zeta(\underline{x})$ is statistically homogeneous, then the Fourier modes for different $\underline{\ell}' \neq \underline{\ell}$ are uncorrelated, though not necessarily independent. However, if additionally $\zeta(\underline{x})$ is a Gaussian random field, then so is $\zeta(\underline{\ell})$, and uncorrelatedness does indeed imply independence.

We have already introduced the power spectrum $P(\underline{\ell})$ in [Equation 19](#) – it is the statement of the famous *Wiener-Khinchin theorem* that for a homogeneous random field the power spectrum is the Fourier transform of the second-order cumulant (*correlation function*)

$$\tilde{\zeta}(\underline{r}) = \langle \zeta(\underline{x}) \zeta(\underline{x} + \underline{r}) \rangle - \langle \zeta(\underline{x}) \rangle \langle \zeta(\underline{x} + \underline{r}) \rangle = \frac{1}{(2\pi)^d} \int d^d \ell P(\underline{\ell}) \exp[i \underline{r} \cdot \underline{\ell}] \quad (73)$$

(Wiener, 1930; Khintchin, 1934). The Fourier transforms of all higher-order cumulants are called *polyspectra* (Brillinger, 2012), and are important detectors of deviations from Gaussianity: since all cumulants of order $n \geq 3$ of a Gaussian distribution must vanish, so in consequence all polyspectra vanish as well, with the sole exception of the power spectrum (Isserlis, 1918; Wick, 1950).

The averages $\langle \dots \rangle$ denote averages over the entire ensemble Ω , i. e., over all realisations ζ_ω of the random field ζ – for any measurable function $F(\zeta)$, we have

$$\langle F(\zeta) \rangle = \int_{\Omega} d\mathbb{P}(\omega) F(\zeta_\omega). \quad (74)$$

We can only observe a (finite sub-volume of) a single Universe, hence we are left with a single realisation for the random fields of interest – a priori it is hopeless to try to estimate statistical quantities like n -point functions or polyspectra from one sample alone. To salvage our enterprise, we introduce an additional property of random fields, namely *ergodicity*: if we can exchange the ensemble average with a spatial average, we can instead compute the average on a subregion $\Delta \subseteq \Gamma$

$$\langle F(\zeta) \rangle \simeq \langle F(\zeta) \rangle_{\Delta} = \frac{1}{\text{vol}(\Delta)} \int_{\Delta} dx F(\zeta_\omega(\underline{x})) \text{ for almost all } \omega. \quad (75)$$

A random field ζ satisfying this property for a sufficiently wide set of functions $F(\zeta)$ will be called *ergodic* (Boltzmann, 1884). Under certain assumptions – e. g. stationarity,

Gaussianity and continuity of the power spectrum – it can be proved (Maruyama, 1949; Grenander, 1950), but for the general case we need to postulate it. The collection of homogeneity, isotropy, and ergodicity is often called the *fair sample hypothesis* in cosmology (Peebles, 1980; Martínez and Saar, 2002). In practice, this spatial averaging amounts to partitioning the region Γ into subregions and estimating the n -point functions on each – this requires that the subregions are sufficiently independent. Further, on large scales, where only a few partitions exist, the low number of samples will lead to large uncertainties in the estimates – a phenomenon known as *cosmic variance*.

In contrast to finite-dimensional random vectors (see Section 2.3.1), the toolbox for testing Gaussianity of random fields is comparably empty. One approach is to express a random field in a given basis (pixels or harmonic functions), restrict to a finite number of coefficients, and test for Gaussianity on this finite-dimensional subspace – see Hansen et al. (2002) for an application of the Kolmogorov-Smirnov and the Cramér-von Mises tests to the harmonic coefficients a_{lm} of a toy model of the CMB, and Dineen and Coles (2005) for an entire battery of classical tests of Gaussianity applied to WMAP data – some of them focusing on the one-point distribution, some on multivariate non-Gaussianity. An approach generically addressing the (non-)Gaussianity of a stationary stochastic process (i. e. a homogeneous random field with $d = 1$) was proposed by Cuesta-Albertos et al. (2007) and Nieto-Reyes, Cuesta-Albertos, and Gamboa (2009, 2014). The foundation of the test is the intuitive observation that all projections of a Gaussian distribution will have Gaussian marginals, whereas for a non-Gaussian distributions none of them will except possibly a null set of fringe cases. The set of random fields is modelled as a separable real Hilbert space $(\mathcal{H}, \langle \cdot | \cdot \rangle)$ – this includes stochastic processes as well as finite-dimensional random variables. Cuesta-Albertos et al. (2007) proceeded to prove the following theorem:

Let μ be a *dissipative* probability measure on \mathcal{H} (see the cited paper for the technical definition), and ζ be an \mathcal{H} -valued random quantity. Then, ζ is Gaussian if and only if $\mu(E) > 0$, where

$$E(\zeta) = \{h \in \mathcal{H} : \text{the distribution of } \langle h | \zeta \rangle \text{ is Gaussian}\}. \quad (76)$$

Since any quantity ζ with a Gaussian distribution will have Gaussian-distributed projection marginals $\langle h | \zeta \rangle$ for any $h \in \mathcal{H}$, the formulation can be sharpened: with the above definition for set $E(\zeta)$, exactly one of the following statements is true:

- $\mu(E) = 1$: the distribution of ζ is Gaussian.
- $\mu(E) = 0$: the distribution of ζ is not Gaussian.

Intuitively, the property of a measure to be dissipative is the analogue of continuity with respect to the Lebesgue measure, generalised to the context of the infinite-dimensional

space of field configurations, where no Lebesgue measure exists. This ensures that the measure from which the projection directions h are drawn is sufficiently regular: the set of fringe cases that make a non-Gaussian ζ look Gaussian upon projection is indeed a set of measure zero.

Nieto-Reyes, Cuesta-Albertos, and Gamboa (2014) give practical recipes to construct dissipative measures for stochastic processes. Given a sample from a stationary process $X = (X_t)_{t \in \mathbb{Z}}$, they recommend the following procedure:

- Draw a random vector $h = (h_t)_{t \in \mathbb{N}}$ from a dissipative distribution.
- Convolve X with h : $Y_t^h = \sum_{i=0}^{\infty} h_i X_{t-i}$; $t \in \mathbb{Z}$.
- Test the one-dimensional marginals of Y^h for Gaussianity with any of the classical tests. Since non-Gaussianity of these will translate into non-Gaussianity of the full process X with probability one, a one-dimensional Gaussianity test with significance level α on the marginals will test the Gaussianity of the stochastic process on the same level of significance.

To increase the sensitivity of the test to the full range of observations (X_i) , the authors recommend choosing multiple directions h_i from dissipative distributions with varying parameters, perform the test with each of them, and combine the individual p -values via the False Discovery Rate (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

2.3.3 Transforming to Gaussianity

Any linear map $f : R \mapsto S = AR + B$ from a Gaussian random variate R (this discussion applies to finite-dimensional random vectors as well as to random fields) will again result in a Gaussian random variate S – the Fourier transform of a random field is just one example for this. However, if f is smooth but non-linear, this will lead to a non-Gaussian S , since its cumulants of third and higher order will receive additional contributions from the first- and second order cumulant of R via the Taylor expansion of f , and hence they will fail to vanish.

This observation opens up a path to model non-Gaussian random variates: given a well-motivated transformation T that remaps a non-Gaussian X to a Gaussian Y , it is simple to simulate a sample from the distribution of Y , and apply the inverse of transformation T ; this strategy is known as Gaussianisation. For consistency, T should have the properties of being bijective (i. e. a one-to-one mapping) as well as continuously differentiable and measurable, such that the probability distributions of X and Y can be transformed into each other.

Typically, the random fields of interest to observational cosmology have significant

non-Gaussian distribution properties at low redshift; specifically interesting examples include the three-dimensional density contrasts of dark matter δ_c or its galaxy counterpart δ_g , and the two-dimensional convergence κ and shear γ of weak gravitational lensing. It is still unknown whether the initial spectrum of perturbations, as seeded by inflation, is already non-Gaussian (Planck Collaboration, 2015d), and the investigation of this question is an active field of research since any deviation from Gaussianity would provide insight into the physics behind cosmic inflation (Verde, L. Wang, et al., 2000; Bartolo et al., 2004).

However, non-linear structure formation will introduce couplings between modes and thus lead to non-Gaussian distribution properties even if the initial perturbations were perfectly Gaussian. This is due to the non-linearity of the equations of motion that govern the evolution of the dark matter density contrast: gravitation is an inherently non-linear theory. The first-order perturbation theory described in Section 1.3 and Section 1.2 fails to describe the evolution once the fluctuations in the density contrast δ_c reach unity, and moments of higher than second order attain significant and measurable values. For example, in an Einstein-de Sitter universe ($\Omega_m = 1$, $\Lambda = 0$), a second-order calculation of the normalised skewness parameter

$$S_3(z) = \frac{\langle \delta_c^3(z, \underline{x}) \rangle}{\langle \delta_c^2(z, \underline{x}) \rangle^2} \quad (77)$$

reveals it to be equal to 34/7 at redshift $z = 0$, assuming Einstein gravity and Gaussian initial conditions (Peebles, 1980). See Bernardeau, Colombi, et al. (2002) for corrections to the above value accounting for smoothing in Fourier space, and also Catelan and Moscardini (1994) for a similar calculation of the fourth-order kurtosis parameter $S_4 = [\langle \delta_c^4 \rangle - 3\langle \delta_c^2 \rangle^2] / \langle \delta_c^2 \rangle^3$.

Bispectrum measurements have been reported by Bernardeau, Mellier, and van Waerbeke (2002), Fu et al. (2014), and Simon et al. (2015) for weak lensing convergence, and e. g. by Gil-Marín et al. (2015) for the clustering of BOSS galaxies. The cosmological information stored in the cumulants of higher than second order are increasingly challenging to measure directly, but are expected to hold an information content comparable to two-point statistics. This has been demonstrated by Takada and Jain (2003, 2004, 2009) and Sato and Nishimichi (2013) by comparing cosmological parameter constraints arising from weak lensing bispectra and those from power spectra.

Further, the presence of non-Gaussianity needs to be taken into account if we wish to perform inference on two-point statistics like correlation functions or power spectra, and compare them to theoretical predictions (Hu and White, 1996; Cooray and Hu, 2001; Semboloni et al., 2007). Estimated values for the power spectrum, evaluated at two different wave vectors $\underline{\ell} \neq \underline{\ell}'$ will be uncorrelated if the underlying random field

is Gaussian – this is a consequence of homogeneity, as discussed in [Section 2.3.2](#). By corollary, the correlation matrix of the power spectrum estimator

$$\text{Corr}[\hat{P}(\ell_i), \hat{P}(\ell_j)] = \frac{\text{Cov}[\hat{P}(\ell_i), \hat{P}(\ell_j)]}{\sqrt{\text{Var}[\hat{P}(\ell_i)] \text{Var}[\hat{P}(\ell_j)]}} \quad (78)$$

is equal to the unit matrix (or rather – a noisy version thereof). For non-Gaussian random fields, however, off-diagonal terms appear in the covariance matrix, in the form of trispectrum contributions (see Peebles, 1980; Meiksin and White, 1999; Scoccimarro, Zaldarriaga, and Hui, 1999; Cooray and Hu, 2001, for semi-analytical calculations and comparisons to N -body simulations). These off-diagonal contributions remain present for band-averaged power estimates even in the infinite-volume limit; see also Feldman, Kaiser, and Peacock (1994), Tegmark, Hamilton, et al. (1998), and Hu and White (2001) for technical reviews on band-averaged power spectrum estimation. These off-diagonal covariances will already affect weakly non-linear scales by inflating the error bars of band power estimates; hence the information content of these scales, when used for power spectrum estimation, is small compared to that of larger scales which are fully linear (Rimes and Hamilton, 2005, 2006; Neyrinck, Szapudi, and Rimes, 2006; Neyrinck and Szapudi, 2007; Sato, Hamana, et al., 2009). Similarly, terms containing the pentaspectrum and the trispectrum will appear in the covariance matrix of bispectrum estimators (Scoccimarro and Couchman, 2001; Kayo, Takada, and Jain, 2013; Sato and Nishimichi, 2013).

To find the covariance matrix of weak lensing two-point statistics, a vital ingredient of the likelihood, mock realisations of the random fields can be used alternatively. To this end, N -body simulations of gravitational collapse are used to produce a large number of independent convergence maps, from which the sample covariance matrix can be estimated in an unbiased fashion. Inverting this matrix, which is necessary for assembling the likelihood function, can introduce bias; this can be computed from the simulations and corrected, but the procedure comes at the price of requiring a larger number of independent realisations, thus increasing the computational cost even further (Hartlap, Simon, and P. Schneider, 2007; Dodelson and M. D. Schneider, 2013; A. N. Taylor, Joachimi, and Kitching, 2013).

Two ideas to motivate a transformation that approximately Gaussianises the evolved field of cosmological perturbations have gained popularity – one parametric, and one non-parametric. The first – known as the *lognormal* transformation – was initially motivated by the asymmetric shape of the histogram of δ_g , i.e. the one-point distribution (Hubble, 1934): it is supported on the interval $[-1, +\infty)$, and the decay behaviour for $\delta_g \rightarrow \infty$ is ostensibly slower than exponential. The approximate lognormality of the histogram of δ_g has been confirmed by Wild et al. (2005) on *2dFGRS* data, and the

approximate lognormality of the histogram of κ by Clerkin et al. (2016) on DES data; for comparisons to N -body simulations see Colombi (1994), Bernardeau and Kofman (1995), and Kayo, Taruya, and Suto (2001a).

A major advantage of the lognormal model, explaining its popularity, is that it provides a simple analytical expression for the probability, which allows one to perform calculations with pen and paper – see Hilbert, Hartlap, and P. Schneider (2011) for a compendium of useful expressions for cosmic shear. Specifically, the transformation itself takes the form

$$\delta_g \rightarrow \delta_{\ln} = \ln(1 + \delta_g) \quad (79)$$

for the galaxy density contrast and

$$\kappa \rightarrow \kappa_{\ln} = \kappa_0 \ln \left(1 + \frac{\kappa}{\kappa_0} \right) \quad (80)$$

for weak lensing convergence ($\kappa_0 > -\inf\{\kappa\}$). The hope is that the Gaussianised random fields δ_{\ln} , κ_{\ln} have approximately Gaussian distribution properties in all n -point distributions including the histogram. Coles and Jones (1991) proposed a heuristic argument for how a Gaussian distribution for the initial spectrum of peculiar velocities can lead to a lognormal distribution, but its validity has been questioned (Kayo, Taruya, and Suto, 2001b); further, the variable $\ln(1 + \delta_c)$ can also be used as an expansion parameter in cosmological perturbation theory (Szapudi and Kaiser, 2003; Szapudi, 2009; X. Wang et al., 2011).

A lognormal model for cosmological random fields has been shown to outperform simplistic Gaussian models in various settings: the one- and two-point probability distributions have been shown to be accurately described with a lognormal model (Kayo, Taruya, and Suto, 2001b); the density contrast of cold dark matter and galaxies, when evolved to low redshifts, exhibit a significantly higher information content when log-transformed, this is because some of the small, non-linear scales will be decorrelated and can be used for estimation of two-point statistics like the power spectrum (Neyrinck, Szapudi, and Szalay, 2009; Neyrinck, 2011; Neyrinck, Szapudi, and Szalay, 2011). The lognormal transformation has a similar effect on weak lensing convergence and particularly increases the sensitivity to the redshift evolution of dark energy (Hilbert, Hartlap, and P. Schneider, 2011; Seo, Sato, Dodelson, et al., 2011; Seo, Sato, Takada, et al., 2012). These information gains can be interpreted as pulling information from higher-point statistics into the two-point function.

Whilst it is significantly better than a naïve assumption of Gaussianity, however, the lognormal transformation has several downsides: the observed histogram actually has more probability mass in its long tail than a suitable lognormal distribution (Joachimi,

A. N. Taylor, and Kiessling, 2011). The correlation structure that can be attained with lognormal random variables is limited (Denuit and Dhaene, 2003; Xavier, Abdalla, and Joachimi, 2016), since any permissible Gaussian distribution must have a positive-definite covariance matrix; conversely, applying a lognormal transformation to convergence maps has in fact been observed to lead to negative eigenvalues of the sample covariance matrix (Hilbert, Hartlap, and P. Schneider, 2011; Xavier, Abdalla, and Joachimi, 2016). Further, the information gains diminish significantly once a realistic level of shape noise is added (Joachimi, A. N. Taylor, and Kiessling, 2011; Seo, Sato, Takada, et al., 2012).

The second approach to motivating a Gaussianising map is the Rosenblatt transform (Lévy, 1937; Rosenblatt, 1952). This transformation is based on the elementary observation that any real random variable X can be bijectively mapped into uniformity: if $p_X(x)$ is the probability density of X , and $C_X(x) = \int_{-\infty}^x dr p_X(r)$ is its cumulative distribution, then $Y = C_X(X)$ is a random variable with uniform distribution on the unit interval $[0, 1]$. We can now play the inverse game with the Gaussian distribution: let $\Phi(z) = \int_{-\infty}^z \frac{ds}{\sqrt{2\pi}} \exp\left[-\frac{s^2}{2}\right]$ be the cumulative distribution of the Normal distribution with mean zero and variance one. Then, $Z = \Phi^{-1}(Y) = \Phi^{-1} \circ C_X(X)$ is a random variable that is exactly Normal. Further, the transformation $X \xrightarrow{C_X} Y \xrightarrow{\Phi^{-1}} Z$ is bijective and continuously differentiable, as long as $p_X(x)$ is continuous. It should be noted, however, that unlike the lognormal transformation there is no analytic closed form for the map $\Phi^{-1} \circ C_X$, since the cumulative distribution needs to be empirically determined from the data set at hand.

This transformation, since its first application to non-Gaussian random fields of cosmological importance by D. H. Weinberg (1992), has gained popularity, and is often used synonymously with Gaussianisation (note that our use of the word is not limited to the Rosenblatt transform) – another common term is *rank-order Gaussianisation*. Like the lognormal transformation, it has been demonstrated to mitigate the polyspectra of third to sixth order for weak lensing convergence (Y. Yu et al., 2011, 2012), and to increase the information content accessible to two-point statistics for the dark matter density contrast (Neyrinck, Szapudi, and Szalay, 2009; Neyrinck, 2011; Neyrinck, Szapudi, and Szalay, 2011); it slightly outperforms the lognormal transformation in this objective. In any way, these gains are found to degrade once a realistic level of noise is added.

For a random vector $\underline{X} \in \mathbb{R}^d$, the collection of univariate marginal distributions of each component on its own $\{p_i(x)\}_{i=1}^d$ does not completely constrain the full multivariate distribution $P(\underline{X})$. Instead, the correlation structure between the components, which gets lost during marginalisation, is captured in the d -dimensional *copula* – the latter being a term for a probability distribution on the unit cube $[0, 1]^d$. In fact, *Sklar's theorem* asserts that the d one-dimensional marginal distributions combined with a copula do indeed determine the full multivariate distribution and vice versa – since the copula is exactly what remains after each component of \underline{X} has been Rosenblatt-transformed

(Sklar, 1959; Nelsen, 1999).

The random fields describing cosmological large-scale structure have an infinite sequence of n -point copulas, each and every one of which have to agree with the Gaussian copula on the n -hypercube if the random field in question can be perfectly Gaussianised. Scherrer et al. (2010) postulated exactly this under the name Gaussian Copula Hypothesis (GCH), and demonstrated its validity for $n = 2$. If the GCH is satisfied, then the Rosenblatt transformation is automatically the best Gaussianising transformation, making all polyspectra vanish identically.

Sato, Ichiki, and Takeuchi (2010, 2011) have used the Gaussian copula to construct a likelihood for band-averaged power spectrum estimates of weak lensing convergence, and demonstrate that it reproduces their two-dimensional joint distributions more faithfully than a Gaussian probability density. Lin and Kilbinger (2015a,b) and Lin, Kilbinger, and Pires (2016) used the Gaussian copula to construct a likelihood for weak lensing peak counts, which are an alternative observable known to harbour nonlinear information (Hamana, Takada, and Yoshida, 2004; Pires, Leonard, and Starck, 2012).

The status of the full GCH, however, remains unclear; as does the question of the consequences for Rosenblatt Gaussianisation, should it fail partially. We will pick up on this point in Chapter 5 to discuss it in further depth.

Part II

GAUSSIANISATION

 MODELLING OF POSTERIOR DENSITIES

The Infinite Improbability Drive is a wonderful new method of crossing vast interstellar distances in a mere nothingth of a second, without all that tedious mucking about in hyperspace. [...]

The principle of generating small amounts of finite improbability by simply hooking up the logic circuits of a Bambleweeny 57 Sub-Meson Brain to an atomic vector plotter suspended in a strong Brownian motion producer (say a nice hot cup of tea) were of course well understood – and such generators were often used to break the ice at parties by making all the molecules in the host’s undergarments leap simultaneously one foot to the left, in accordance with the Theory of Indeterminacy.

Many respectable physicists said that they weren’t going to stand for this, partly because it was a debasement of science, but mostly because they didn’t get invited to those sorts of parties.

*Douglas Adams
The Hitchhiker’s Guide to the Galaxy*

3.1 INTRODUCTION

According to the Bayesian paradigm, inference on any data set will yield a posterior probability distribution on the space of model parameters. This density function represents, in its entirety, the full knowledge gained in the attempt to infer the underlying parameters. Such distributions often depart significantly from a Gaussian form. This led to the widespread use of Monte Carlo sampling methods to report the typically non-Gaussian posterior constraints obtained from experiments, such as *Planck*¹. Reconstructing the posterior density from such a Markov Chain Monte Carlo (MCMC) sample, e.g. to visualise the multivariate parameter constraints, or to combine the constraints from multiple data sets, can be nontrivial due to the large sample size necessary to appropriately map the distribution; in addition, the contours often need further smoothing

¹ See Planck Collaboration et al. (2014) and Planck Collaboration (2015c). For the Markov chains see <http://www.cosmos.esa.int/web/planck/pla>; consult <http://lambda.gsfc.nasa.gov> for an eclectic list of data combinations in various cosmological models.

for stylistic reasons.

Instead, we propose to redefine the underlying model parameters, so that the new posterior density approximately takes a Gaussian shape after the transformation from old to new parameters; this presupposes that we begin with a unimodal posterior density. Whereas this formalism may be extended to multimodal probability densities, we currently restrict ourselves to unimodal distributions in this work. Suggestions for how to extend it are given in [Section 6.1.2](#). Such a Gaussianising transformation would allow for enormous data compression: instead of a full [MCMC](#) sample from the posterior distribution, we only need to report the Gaussianising transformation, and the first and second moments of the resulting Gaussian distribution. From these alone, we can reconstruct an analytic expression for the full non-Gaussian posterior density, and subsequently combine it with other data sets.

Further, it becomes possible to display and compare non-ellipsoidally-shaped contours of non-Gaussian parameter constraints – whether joint or marginalised – without any smoothing. Thus, this method allows for summarising posterior densities in a versatile and efficient way, which faithfully reproduces the information contained in the full probability density.

The idea of transforming a function to a Gaussian shape is, in principle, not limited to reproducing probability densities. As the integral over a Gaussian can be performed analytically, this opens up a strategy to feasibly compute high-dimensional integrals, such as the model evidence (i.e. the marginal likelihood).

The transformed model parameters are analogous to the normal parameters of the [CMB](#): it has been highly advantageous for rapid likelihood calculation (such as [CMBfit](#), [CMBwarp](#), and [PICO](#); see [Kosowsky, Milosavljevic, and Jimenez 2002](#); [Chu, Kaplinghat, and Knox 2003](#); [Jimenez et al. 2004](#); [Sandvik et al. 2004](#); [Fendt and Wandelt 2007](#)), to redefine the cosmological model parameters such that the model is approximately linear in these newly defined *normal* parameters. Thus the likelihood approximately takes the form of a multivariate Gaussian density. For most observables, we would be at a loss to search for a linearising redefinition of the model parameter space directly motivated by the structure of the model itself. Instead, is it possible to computationally find suitable parameters, i.e. a suitable bijective transformation which approximately Gaussianises the posterior in question?

Extending the work of [Joachimi and A. N. Taylor \(2011\)](#), we present an algorithm to find and test such a non-linear Gaussianising transformation from a Markov chain sampling the posterior distribution of the original parameters. In principle, this distribution could stem from any experiment or data type. In [Section 3.2](#), we describe the details of the algorithm, verification of the reconstructed posterior distribution, and the specific transformations employed. Following an illustration of these on a toy example in [Section 3.3](#), [Section 3.4](#) demonstrates the performance of our implementation, using Markov chains

from the *Planck* satellite constraints on cosmological models (Planck Collaboration et al., 2014; Planck Collaboration, 2015c).

3.2 GAUSSIANISATION

To find the right multivariate transformation, we will at first adopt the strategy of redefining each model parameter separately, i.e. the first new model parameter will only depend on the first old model parameter, etc. In Section 3.2.4, we will drop this assumption and consider transformations which can correlate the model parameters.

The set of all multivariate Gaussianisation transformations, from which we are to pick the optimal one, will be constructed in the following way: assume a family of bijective real-valued functions $F_{\Delta} : \mathbb{R} \rightarrow \mathbb{R}$ indexed by n real transformation parameters $\Delta = (\delta^1, \dots, \delta^n)$. Given the d -dimensional vector of model parameters $\underline{X} = (X_1, \dots, X_d)$, we transform to the new (Gaussian-distributed) parameters \underline{Y} via

$$\underline{Y} = (Y_1, \dots, Y_d) = [F_{\Delta_1}(X_1), \dots, F_{\Delta_d}(X_d)], \quad (81)$$

where the full multivariate transformation is now specified by all d transformation parameter n -tuples $(\Delta_1, \dots, \Delta_d)$, i.e. one $\Delta_i = (\delta_i^1, \dots, \delta_i^n)$ for each model parameter. To avoid confusion, we shall from now on distinguish between model parameters (MP), which the posterior probability density depends on, and transformation parameters (TP), which specify one Gaussianising transformation. The algorithm can be applied to arbitrary parametrised transformation families, suitable for various forms of non-Gaussianity – in principle, we could even choose different transformations for each model parameter, instead of using the same shape $F_{\Delta_i}(X_i)$ for all of them.

Assuming such a bijective transformation $\underline{X} \mapsto \underline{Y}$, we immediately have an analytic form for the posterior density

$$\begin{aligned} \Pi(\underline{X}) &= \tilde{\Pi}(\underline{Y}) \left| \frac{d\underline{Y}}{d\underline{X}} \right| \\ &= \frac{1}{\sqrt{(2\pi)^d \det \tilde{\Sigma}}} \exp \left\{ -\frac{1}{2} [\underline{Y}(\underline{X}) - \tilde{\mu}]^T \tilde{\Sigma}^{-1} [\underline{Y}(\underline{X}) - \tilde{\mu}] \right\} \prod_{i=1}^d \left| \frac{dF_{\Delta_i}}{dX}(X_i) \right|. \end{aligned} \quad (82)$$

One still needs to find the mean vector $\tilde{\mu}$ and the covariance matrix $\tilde{\Sigma}$ of the transformed posterior density $\tilde{\Pi}$. These are estimated from the transformed sample (see Section 3.2.1).

3.2.1 Finding the optimal transformation

Given a weighted point sample $\mathcal{D} = \{(\underline{X}^a, w^a)\}_{a=1}^{\mathcal{N}}$, containing \mathcal{N} points in \mathbb{R}^d and probability weights w^a , which has been sampled from the posterior distribution in question, we wish to quantify the Gaussianisation properties of different transformations applied to this sample. To this end, we follow Box and D. R. Cox (1964) (see also Velilla, 1993; Joachimi and A. N. Taylor, 2011) in maximising the profile likelihood over TP space, i.e. depending only on the $n \times d$ real transformation parameters contained in $\Delta = (\Delta_1, \dots, \Delta_d)$. This likelihood is a function of the transformation parameters Δ , quantifying how well each transformation Gaussianises the distribution of data set \mathcal{D} ; however, it does not pertain to the posterior density in Equation 82, which is a function of the model parameters \underline{X} .

For the Gaussian parameters $\underline{\tilde{\mu}}, \tilde{\Sigma}$ in Equation 82, we insert their standard debiased weighted maximum-likelihood estimators, which depend on the transformed sample $\{(\underline{Y}^a, w^a)\}_{a=1}^{\mathcal{N}}$

$$\underline{\tilde{\mu}} = \frac{1}{W_1} \sum_{a=1}^{\mathcal{N}} w^a \underline{Y}^a; \quad (83)$$

$$\tilde{\Sigma} = \frac{W_1}{(W_1)^2 - W_2} \sum_{a=1}^{\mathcal{N}} w^a (\underline{Y}^a - \underline{\tilde{\mu}}) (\underline{Y}^a - \underline{\tilde{\mu}})^T, \quad (84)$$

with $W_1 = \sum w^a$ and $W_2 = \sum (w^a)^2$. These estimators depend on Δ indirectly, as they are computed after \mathcal{D} has been transformed with Δ . We arrive at the profile weighted log-likelihood

$$\mathcal{L}(\Delta|\mathcal{D}) = -\frac{W_1}{2} \ln \det \tilde{\Sigma}(\Delta, \mathcal{D}) + \sum_{a=1}^{\mathcal{N}} w^a \sum_{i=1}^d \ln \left| \frac{dF_{\Delta_i}}{dX}(X_i^a) \right|, \quad (85)$$

where several terms independent of Δ have been discarded. In general, both the covariance matrix of the transformed sample and the Jacobian term will depend on the transformation parameters Δ in a non-linear way, hence finding the maximum-likelihood values for the TPs will require numerical optimisation. For this purpose, we have employed the GSL implementation of the well-known Nelder–Mead simplex algorithm (Nelder and Mead 1965).

As already noted by Joachimi and A. N. Taylor (2011), log-likelihood degeneracies in TP space are common. These may jeopardise the numerical stability of the calculation of \mathcal{L} . There are generic cases where a moderately large value for one transformation parameter may already result in unmanageably large numerical values for the transformed sample, such as e.g. the power transformation $X_i \mapsto (X_i)^{\lambda_i}$ with $\lambda_i \sim 50$. Generically,

the optimisation algorithm tends to slide into these TP space regions quite easily. Hence, we include a penalty term of the form

$$\mathcal{P}(\Delta) = \epsilon \sum_{i=1}^d \sum_{s=1}^n (\delta_i^s - \delta^{s,U})^p, \quad (86)$$

where $\delta^{s,U}$ are the parameter values corresponding to the identity transformation. We minimise the function $-\mathcal{L}(\Delta|\mathcal{D}) + \mathcal{P}(\Delta)$ over the $n \times d$ real numbers in Δ . Values of $p = 4$ and $\epsilon = 10^{-4}$ have proven to be highly stabilising, and at the same time do not distort the shape of the resulting analytic posterior distribution.

In this work, we employ the Nelder–Mead algorithm just for illustrating the method – faster and more reliable algorithms to find the global minimum of the likelihood function exist (such as BOBYQA, see Powell, 2008) and can readily be applied here.

3.2.2 Box–Cox transformations and their kin

The Box–Cox transformation (Box and D. R. Cox, 1964) is a generalisation of the power map. This transformation family is widely used in statistics and econometrics, e.g. to make data approximately homoscedastic and normal. Our usage is different in that we use it to alter the distribution of model parameters, rather than the distribution of data. Including a shift parameter a , the one-dimensional version is defined as

$$x \mapsto BC_{(a,\lambda)}(x) = \begin{cases} \lambda^{-1}[(x+a)^\lambda - 1] & (\lambda \neq 0) \\ \ln(x+a) & (\lambda = 0) \end{cases} \quad (87)$$

for a single MP x , i.e. $(\delta^1, \delta^2) = (a, \lambda)$. Note that the family is continuous at $\lambda = 0$ and that the mapping requires $a < x$. Typically, an MP with a skewed distribution can be transformed to an MP with symmetric, Gaussian distribution upon the appropriate choice of the power TP λ , e.g., a log-normal distribution can be analytically transformed to a Gaussian with $a = \lambda = 0$. The identity transformation corresponds to $\delta^{1,U} = a = 1$ and $\delta^{2,U} = \lambda = 1$. Inserting this transformation family into Equation 82, we recover the formula given in Joachimi and A. N. Taylor (2011).

As an extension of the Box–Cox family, we propose the Arcsinh–Box–Cox transformation (‘ABC transformation’ hereafter):

$$x \mapsto ABC_{(a,\lambda,t)}(x) = \begin{cases} t^{-1} \sinh[t BC_{(a,\lambda)}(x)] & (t > 0) \\ BC_{(a,\lambda)}(x) & (t = 0) \\ t^{-1} \operatorname{arcsinh}[t BC_{(a,\lambda)}(x)] & (t < 0). \end{cases} \quad (88)$$

The inclusion of the TP t will prove particularly useful to remove residual kurtosis from a MP distribution. The identity transformation reads $\delta^{1,U} = a = 1$, $\delta^{2,U} = \lambda = 1$, $\delta^{3,U} = t = 0$.

The Box–Cox family does not form a group, because two subsequent transformations cannot be expressed as another Box–Cox transformation; the same holds for the ABC family. This will be of importance for Section 3.2.4.

Box–Cox transformations demonstrate that the domain of the function F_Δ – in particular its dependence on Δ – requires special attention: for given a , it is defined only for $x \in (-a, \infty)$, the same holds for ABC transformations. Thus, the optimisation procedure for the sample $\mathcal{D} = \{\underline{X}^a\}_{a=1}^{\mathcal{N}}$ requires that a_i , the shift parameter for the model parameter X_i , is bounded from below, i.e. $a_i > \min_a(-X_i^a)$. Conversely, this means that, once the optimal transformation parameters Δ^{opt} are found and inserted into the analytic expression for the original posterior density, Equation 82, it is not defined for every value possible value of the MP \underline{X} , but only for $X_i > a_i^{\text{opt}}$. This also necessitates that the normalisation needs to be adjusted, which can be done analytically. However, if the sample is large enough so that the tails of the distributions are properly represented, this truncation of the domain is not problematic.

3.2.3 Verifying the optimal transformation

Once the optimal transformation within its family is found, how do we judge the effectiveness of the resulting Gaussianisation? We adopt the following pragmatic standpoint: if the analytic posterior manages to reproduce the one-dimensional and two-dimensional marginalised contours of the sample, it is deemed acceptable. To this end, we propose the test via a cross-contour (CC) plot. The idea is to characterise a probability density by the location of its contours - the surfaces of constant density - and the probability mass stored inside, i.e. the integral of the density over the interior of a contour. If two densities $p(\underline{X})$ and $q(\underline{X})$ are identical, then they will store the same mass in any region of the parameter space; if they are different, we expect to find different probabilities for the same regions (e.g. the regions bounded by contours of p). Thus, looking at the family of contour-bounded regions of p , we can ask: does the probability for these, assigned via q , agree with the probability for them assigned via p ?

To formalise this, consider the following: given a probability density p in d dimensions, which takes function values between 0 and p_{max} , we define the contour-bounded region assigned to the density value $r \in [0, p_{\text{max}}]$ as

$$\Omega_p(r) = \{\underline{X} \in \mathbb{R}^d : p(\underline{X}) \geq r\}.$$

The probability mass enclosed in any of these is

$$\int_{\Omega_p(r)} p(\underline{X}) dX \in [0, 1].$$

Now, assuming we have two probability densities p and q in d dimensions, do the contours of q reproduce those of p ? They do in the relevant sense if for every $r \in [0, p_{\max}]$, the q -mass enclosed in the r -contour of p equals the p -mass in this contour, i.e.

$$\int_{\Omega_p(r)} q(\underline{X}) dX = \int_{\Omega_p(r)} p(\underline{X}) dX. \quad (89)$$

It should be noted that this alone is not a sufficient condition for $p \equiv q$, but the counterexamples, which can be constructed mathematically, are non-generic and can be neglected for our purposes.

To detect deviations of the contours of p and q , we could simply plot the left and the right side of [Equation 89](#) for a grid of r -values between 0 and p_{\max} , and plot the points with respect to the line $y = x$. For concrete problems, it is often more instructive to subtract the right side from the left side, and plot the excess (or deficit) probability mass of q inside the contours of p . If, in this plot, the excess for every contour is consistent with zero, we have succeeded.

In our situation, we compare a point sample \mathcal{D} with a probability density function p – the analytic posterior density as reconstructed via Gaussianisation. The right side of [Equation 89](#) is the probability mass in the region where the density is greater or equal to r ; the left side is the fraction of the point sample which lies in the same region. So, for every value r in the range of p , we find the probability mass enclosed in $\Omega_p(r)$ by gridding $p(\underline{X})$ over a region containing the sample. Similarly, we count the number of points in \mathcal{D} where the value of p is above r , to compute the fraction of points that lie within $\Omega_p(r)$. This fraction is an estimator of the actual probability mass enclosed, because \mathcal{D} is a discrete sample from the actual posterior distribution. To find the variance of this estimator, we calculate the fraction on 2,000 bootstrap realisations of \mathcal{D} , and determine the 95%-confidence intervals from these. If, for every r , the analytic posterior probability mass inside $\Omega_p(r)$ is within this confidence interval for the sample point fraction within $\Omega_p(r)$, we judge our reconstruction attempt to be successful.

It should be noted that poor [MCMC](#) sampling of the original target density will yield a poor representation of this density by our reconstructed density [Equation 82](#). Any information about the distribution lost by undersampling cannot be regained. However, as demonstrated in [Section 3.3](#), our method reproduces less biased contours than other standard methods of density estimation even in regions of low point density, i.e. where any density estimate must be an extrapolation. Hence, it can be used when the length of the input Markov chains is restricted by computational cost or file size.

3.2.4 Multi-pass transformations

If even the optimal Gaussianising transformation amongst a given family does not bring the posterior density sufficiently close to a Gaussian shape (e.g., as determined via a [CC plot](#)), we have two options. We can provide a different family of transformations and redo the optimisation; or we can repeat the process on the sample after the first transformation. As already mentioned in [Section 3.2.2](#), the transformation families employed in this work do not form groups. Hence, two subsequent transformations do not result in another transformation from that family, and transforming twice potentially provides a better Gaussianisation than transforming once. In principle, it is possible to apply multiple subsequent transformations, should the quality of the result necessitate it. In this spirit, we have implemented the following two-pass transformation protocol:

- Optimise the [TPs](#) of the first transformation.
- Linear reshaping: centring, rescaling, rotating.
- Optimise the [TPs](#) of the second transformation.

Strictly speaking, this transformation, whilst being bijective, no longer falls into the class as set up in [Equation 81](#), as different model parameters are mixed. Nonetheless, [Equation 82](#) for the analytic posterior density generalises in a straightforward way.

In the second step, the sample after the first Gaussianising transformation is subjected to the following maps (in this order): subtract the sample mean from every parameter, so that the sample is centred on the origin. Then, rescale every parameter such that the standard deviation is unity. Finally, rotate into the eigenbasis of the covariance matrix – this procedure is generally known as Principal Component Analysis ([PCA](#)). These reshaping operations not only help to avoid numerical instabilities (centring, rescaling), but also open up new directions for Gaussianisation by presenting uncorrelated parameters to the second Gaussianising transformation, since the transformations defined in [Equation 81](#) cannot mix parameters. If two parameters have substantial covariance after step 1, it can be crucial to decorrelate them.

Nevertheless, a price is to be paid for the Gaussianising power added with Step 2: it sacrifices a decisive property of the simple one-step transformation routine, namely that every transformed [MP](#) Y_i only depends on a single untransformed [MP](#) X_i . This property allows for easy marginalisation of the analytic posterior: to compute this, we can marginalise the Gaussianised sample by dropping all coordinates we wish to marginalise out and determining the mean vector and covariance matrix of the remaining ones. Transforming this marginalised Gaussian density back will then yield the marginalised posterior density on the untransformed [MPs](#). However, with linear reshaping included, this is no longer possible.

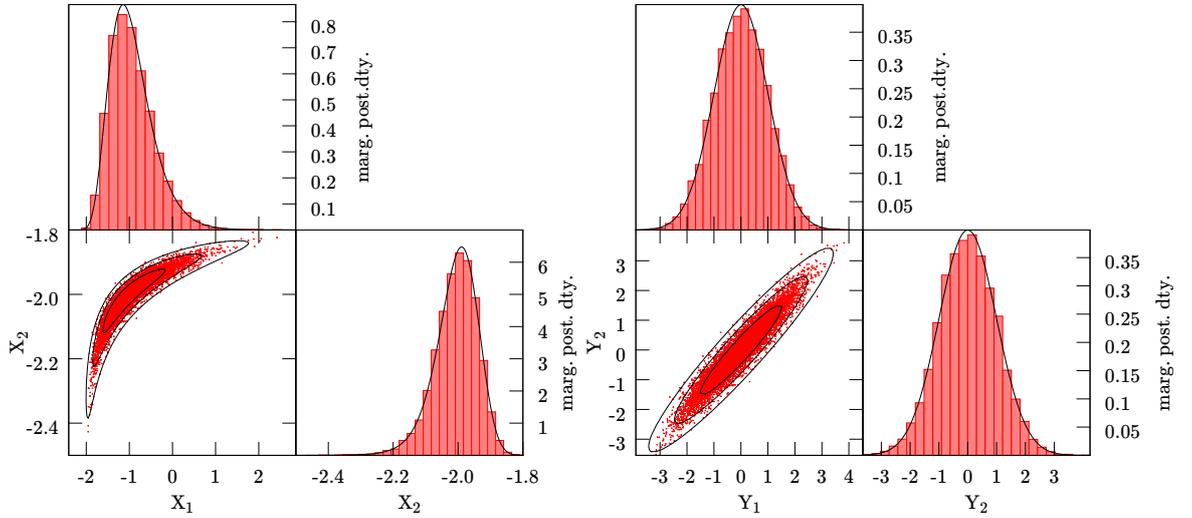


Figure 12: Bivariate sample before (left) and after Gaussianisation (right). We show the two-dimensional sample and its 1D marginals (red), and compare to the reconstructed analytic posterior density (black): the full 2D contours, and its 1D marginal distributions.

This may be problematic for some applications (such as visualisation of 1D or 2D marginal distributions, or creating a [CC](#) plot), but not for others – as long as we need only the marginal distribution of a single combination of parameters, we can marginalise by discarding all [MP](#) columns of the sample except the ones in question, prior to Gaussianising.

3.3 A TOY EXAMPLE

We illustrate these ideas on a two-dimensional example. We draw a sample of 10,000 points from a bivariate Gaussian distribution, and map it through an inverse Box–Cox transformation with known input [TP](#) values (see [Table 2](#). All weights are set to unity. This mock data sample has the advantage that there is at least one Box–Cox transformation which precisely Gaussianises the underlying probability distribution. [Figure 12](#) shows the original sample, and the one transformed with the one-pass Box–Cox transformation which was found to be optimally Gaussianising, i.e. maximising the log-likelihood in [Equation 85](#). As this is a comparably simple problem, we have set the penalty term in [Equation 86](#) to zero. The Nelder–Mead algorithm was started sixteen times independently with randomised initial conditions. The values of the recovered optimal [TPs](#) are shown in [Table 2](#); the standard deviation amongst these sixteen values is of order 10^{-7} at worst, so multiple Nelder–Mead runs are not necessary in this low-dimensional example: all of them find the same maximum of the log-likelihood. In

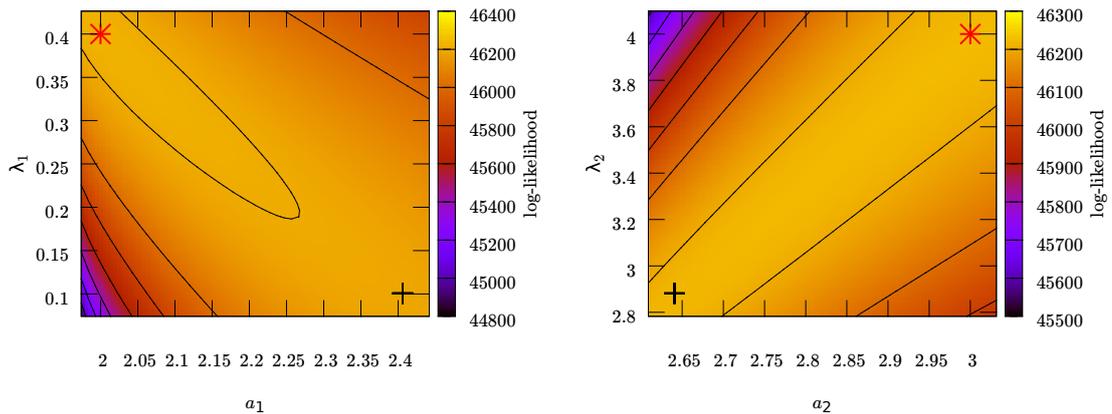


Figure 13: Profile log-likelihood for the transformation parameters relating to X_1 (top), and to X_2 (bottom). The red star shows the input values, the black cross the recovered values, as projected onto the plane. The degeneracies between different transformation parameters are apparent.

Parameter	input value	recovered value
a_1	2	2.4
λ_1	0.4	0.1
a_2	3	2.6
λ_2	4	2.9

Table 2: Optimally Gaussianising parameters for the distribution in Figure 12, as found with one-pass Box–Cox transformation.

high-dimensional cases, however, this strategy can increase the robustness of the procedure. The apparent difference between the parameters of the single inverse Box–Cox transformation and the values found for Box–Cox optimisation is due to degeneracies in parameter space. To illustrate these, we show the profile likelihood for (a_1, λ_1) where (a_2, λ_2) are held fixed at their input values, and vice versa, in Figure 13. The TPs found by the optimisation algorithm (black crosses - note that they are projected onto the plane for which the profile likelihood is shown) are degenerate with the input ones (red star). Both Box–Cox transformations map the distribution to sufficiently Gaussian form. We compare our method of reconstructing an analytic posterior density from an MCMC sample with the standard nonparametric method, Kernel Density Estimation (KDE), which also aims to find a functional form for the probability density. The $1 - 3\sigma$ -contours of the posterior density from Gaussianisation are shown jointly with those from KDE: these employ a Gaussian kernel, whose covariance matrix is estimated from the sample, and Silverman’s rule (Silverman 1986) has been used to determine the bandwidth parameter. No additional smoothing has been applied in Figure 14, top panel. The bottom panel shows the excess cross-contour probability masses between analytic posterior and

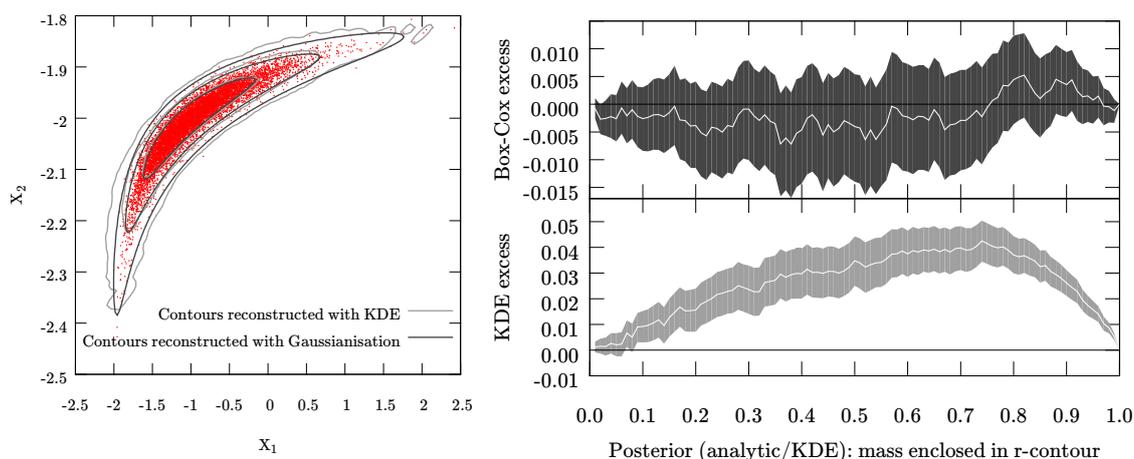


Figure 14: Comparison of the analytic posterior density, as found via Gaussianisation (dark gray), to kernel density estimation (light gray). The top panel shows the 2D contours of each density estimation method in relation to the original sample (red dots). The bottom panel (CC plots; top: Box–Cox, bottom: KDE) compares the respective contours of each function to the original sample: We determine the fraction of the point sample located inside one probability contour, and plot the excess of this fraction over the probability density mass for that same contour. The band shows the 95%-variance in the point fraction due to sampling.

sample, and KDE and sample respectively, as detailed in Section 3.2.3. Whereas the Box–Cox posterior is consistent with the sample distribution for every single contour, the KDE contours show a strong bias – the contours are wider than they should be. Given that the precision of the contour reconstruction is, for the Box–Cox method, limited only by the finite size of the sample, it has the potential to perform better than the (biased) kernel density method – see Section 2.2 for the origin of the bias of density estimates. Additionally, for applications in which frequent calls of the posterior density are a bottleneck for computation speed, our method of density reconstruction can be advantageous: the additional initial cost for finding the transformation parameters can be outweighed by the subsequent evaluation speedup.

3.4 PERFORMANCE RESULTS: PLANCK DATA

To demonstrate how the algorithm works on real data, we have employed MCMC samples from the first data release of the *Planck* mission (see Planck Collaboration et al., 2014). This satellite has measured the temperature and polarisation anisotropies in the CMB, whose power spectra are sensitive measures of the underlying cosmology. The *Planck* Collaboration has published several data products², including MCMC samples from the posterior probability densities of various cosmological models, generated with

² See <http://www.cosmos.esa.int/web/planck/pla>.

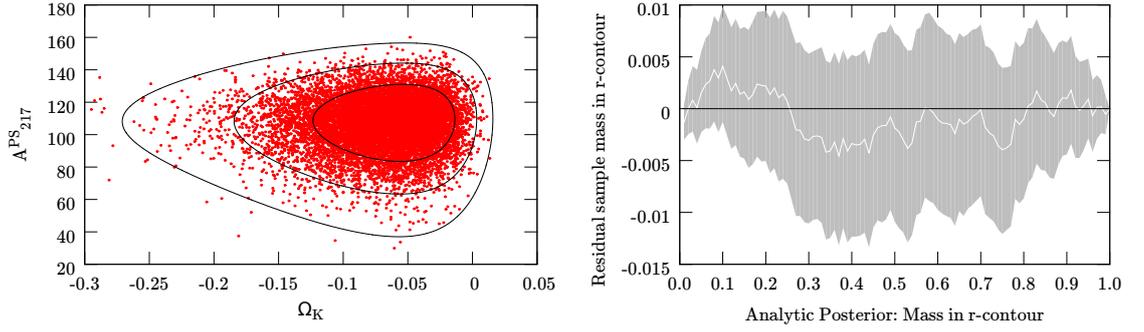


Figure 15: One-pass Gaussianisation of a triangular-shaped non-Gaussian feature in a 2D marginal *Planck* posterior via ABC transformation. *Left*: original sample (red dots) and reconstructed analytic posterior (black contours). *Right*: the CC plot shows that for every contour of the analytic posterior, the probability mass inside (white line) equals the fraction of the point sample inside, within its 95%-confidence interval (green band).

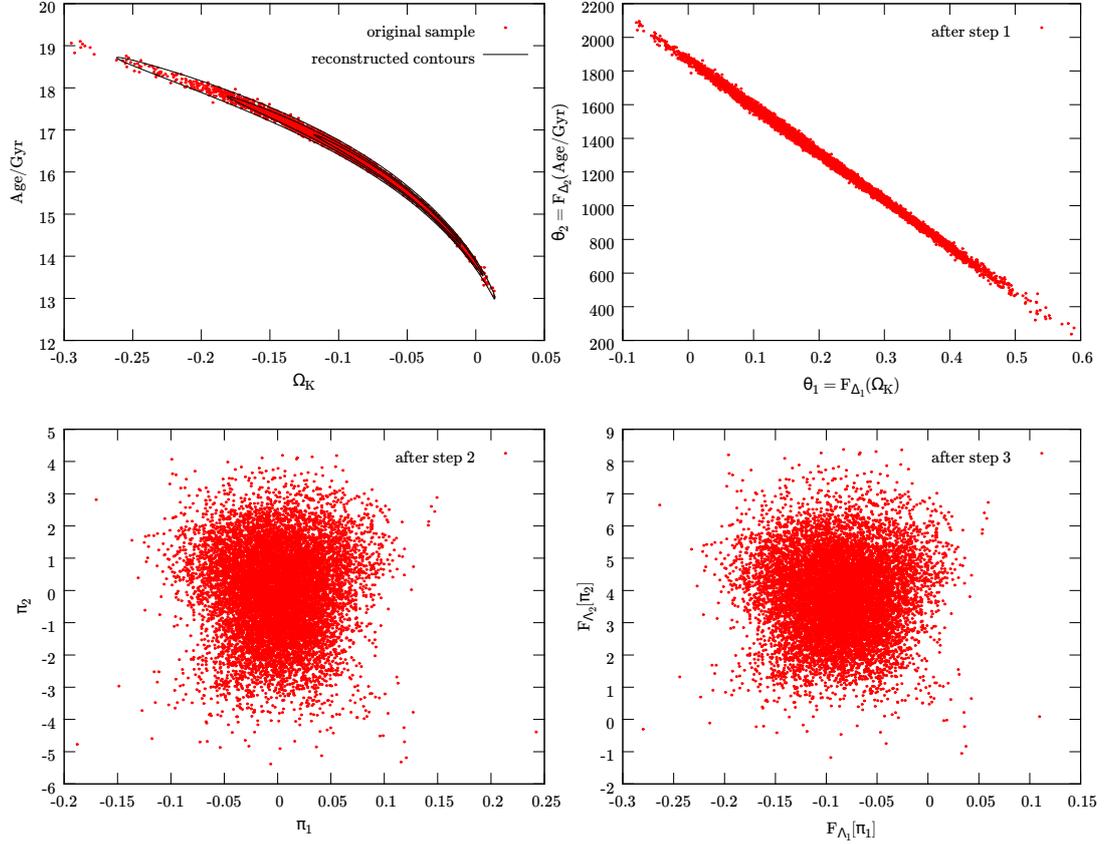


Figure 16: Two-pass Gaussianisation of a non-Gaussian model parameter degeneracy in a 2D marginal *Planck* posterior via ABC transformation, explicitly showing the protocol described in Section 3.2.4: (θ_1, θ_2) are the parameters after the first transformation; (π_1, π_2) are the coordinates after rotation into the PCA eigenbasis of the centred and rescaled (θ_1, θ_2) -sample, which are finally transformed again. (Δ_1, Δ_2) designate the transformation parameters of the first, (Λ_1, Λ_2) those of the second ABC transformation. Note how crucial the intermediate PCA step is to achieve Gaussianity.

CosmoMC (see Lewis and Bridle, 2002, also: <http://cosmologist.info/cosmomc>).

The baseline cosmology is the standard model of a flat Universe with cold dark matter and a cosmological constant, commonly known as Λ CDM. It contains six parameters: $\Omega_b h^2$ (today’s baryon density), $\Omega_c h^2$ (today’s cold dark matter density), $100 \theta_{\text{MC}}$ (scaled sound horizon), τ (reionisation optical depth), n_s (spectral index of primordial scalar perturbations), and $\ln(10^{10} A_s)$ (log power amplitude of primordial scalar perturbations). Several extensions of this baseline model are also listed, including those by adding either of the following parameters: Ω_K (curvature parameter), w (dark energy equation of state), r (primordial tensor-to-scalar amplitude ratio), and N_{eff} (effective number of relativistic degrees of freedom). The combinations $\Omega_i h^2$ have been chosen as model parameters, instead of the density parameters Ω_i , since the former are independent of the value of the Hubble parameter H_0 – the definition of the density parameters (Equation 9) involves the critical density $\rho_{\text{cr}} \propto H_0^2$. Hence, $\Omega_i h^2$ are directly proportional to the physical densities ρ_i . Note that this does not hold for Ω_K . Further, these chains list derived quantities, e.g. today’s Hubble parameter H_0 , the age of the Universe, and a variety of foreground modelling parameters, such as A_ν^{PS} and A_ν^{CIB} , modelling the amplitudes of Poisson point sources and the cosmic infrared background in the frequency bands $\nu = 100$ GHz, 143 GHz and 217 GHz. These are of particular interest to us, as the most prominent non-Gaussian features of the posterior densities can be seen in them.

The chains, as presented, are not decorrelated, so we thin them by using every 20th sample. We employ the ‘..._planck_lowl_...’ chains, which use only the temperature-temperature correlations. The plots in this section are created using the seven-parameter model including Ω_K ; the sample contains 11,546 points after thinning.

All these Markov chains assume uniform proper prior densities (i.e. being supported on compact rectangular boxes) and list the log-likelihood for every point (for further details, see Planck Collaboration et al., 2014).

We show several 2D marginalised posterior samples exhibiting different non-Gaussian features, and how well they are reproduced by the analytic posterior (Equation 82). The model parameter distributions plotted have been chosen to display various kinds of non-Gaussian morphologies, these arise either from non-linearity of the cosmological or foreground model, or from a prior-driven hard constraint on MP space. Our demonstration includes various kinds of non-Gaussianity, such as triangular shapes (see Figure 15), pronounced non-linear degeneracies (see Figure 16), and sharp boundaries (‘walls’) arising from MP space boundaries (see Figure 17).

Figure 16 demonstrates the usefulness of the intermediate PCA in between the ABC transformations: the first transformation has straightened out the curved shape of the maximum, but the distribution still appears skewed towards the upper left direction (see top right panel). This is remedied by reshaping, PCA and another ABC transformation (bottom right panel) – the second Gaussianising transformation having only little effect

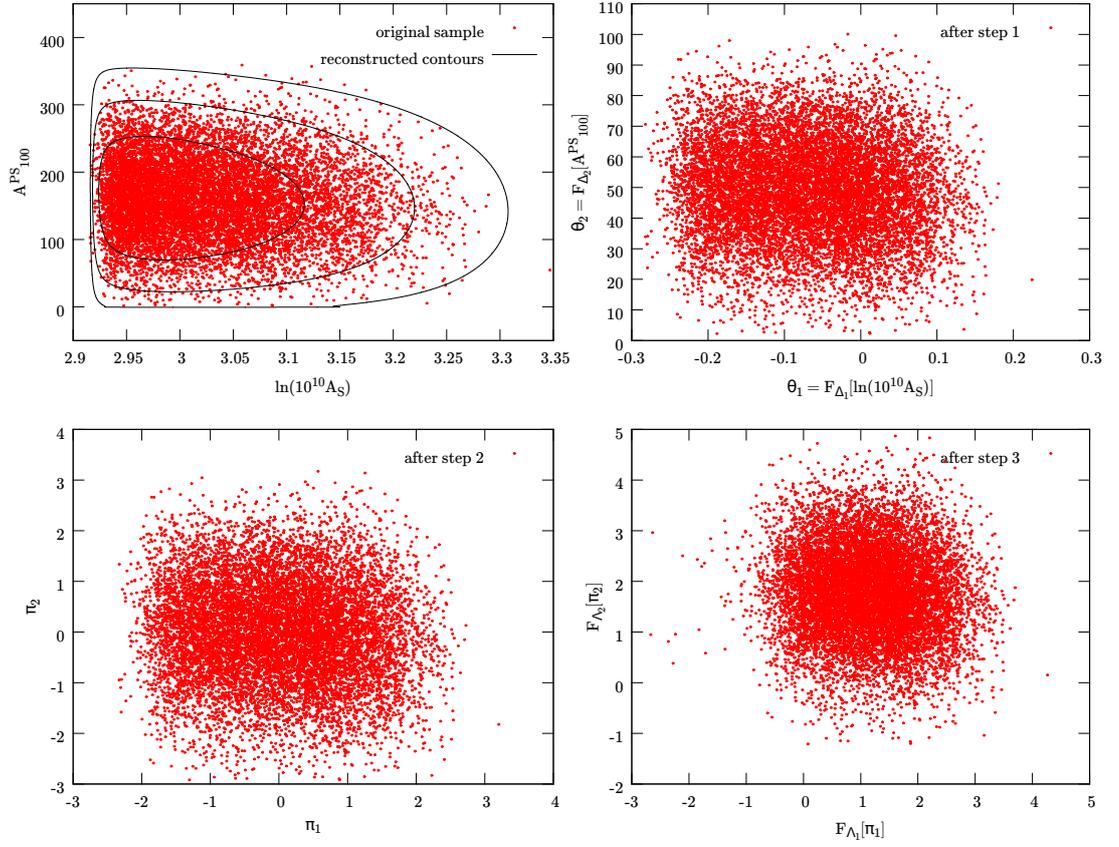


Figure 17: Two-pass Gaussianisation of a wall-like non-Gaussian feature in a 2D marginal *Planck* posterior via ABC transformation; in the same format as as Figure 16, but with the transformation parameters (Δ_1, Δ_2) and (Λ_1, Λ_2) that have been found for this sample. It is apparent that after the first transformation, the parameter θ_1 still exhibits residual kurtosis.

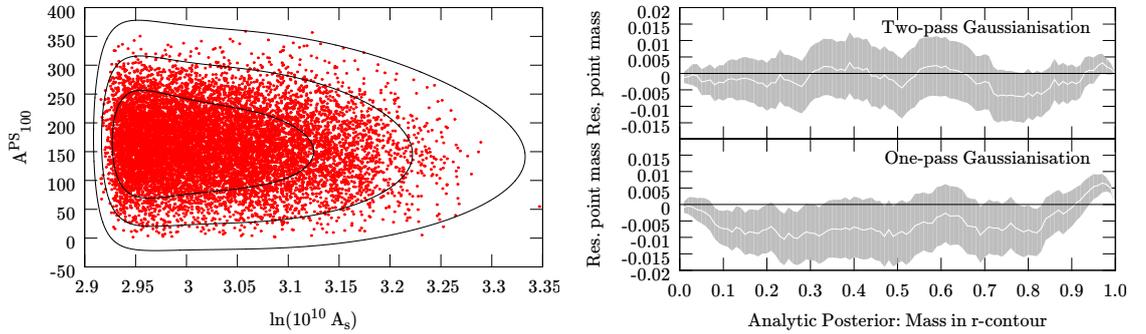


Figure 18: One-pass Gaussianisation of a 2D marginal *Planck* posterior via ABC transformation. *Left*: original sample and contours of the analytic posterior. *Top right*: CC plot for the two-pass Gaussianisation in Figure 17. *Bottom right*: CC plot for the one-pass Gaussianisation (see left panel), showing deviations of the cross-contour masses.

compared to the PCA.

The Gaussianisation of the distribution in Figure 17 (top left) shows how two concatenated transformations can be more powerful than a single one. The once-transformed sample still exhibits negative excess kurtosis, which is removed by the second transformation (bottom left to bottom right panel).

Further, we compare the CC plots of this two-pass transformation and the one-pass transformation in Figure 18, which also shows the resulting contours (left panel). The associated one-pass CC plot (bottom right) shows a significant deficit of point sample mass compared to the analytic posterior mass, for the posterior contours between ~ 0.1 and ~ 0.3 , as well as for ~ 0.8 , and between ~ 0.95 and 1 . The latter is visible between the 2σ - and 3σ -contours close to the wall-like constraint at $\ln(10^{10} A_S) \simeq 2.92$. By contrast, the CC plot for the two-pass transformation (Figure 18, top right) demonstrates good agreement between the contours of analytic posterior and point samples.

To demonstrate the algorithm working on a high-dimensional example, we Gaussianise a seven-dimensional *Planck* MCMC sample with an ABC transformation. In order to visualise the result, we show all one-dimensional and two-dimensional marginal distributions of the point sample and the full analytic posterior density (see Figure 19). We employ one-pass transformations, because, as discussed in Section 3.2.4, the marginalisation of the analytic posterior from 7D down to 2D or 1D would not be possible without explicit integration or sampling, had we chosen to use the two-pass protocol.

We have demonstrated that, equipped with multi-pass protocol and CC plot, Gaussianisation is a versatile and robust toolbox for the reconstruction of posterior probability densities that we have field-tested on cosmological data.

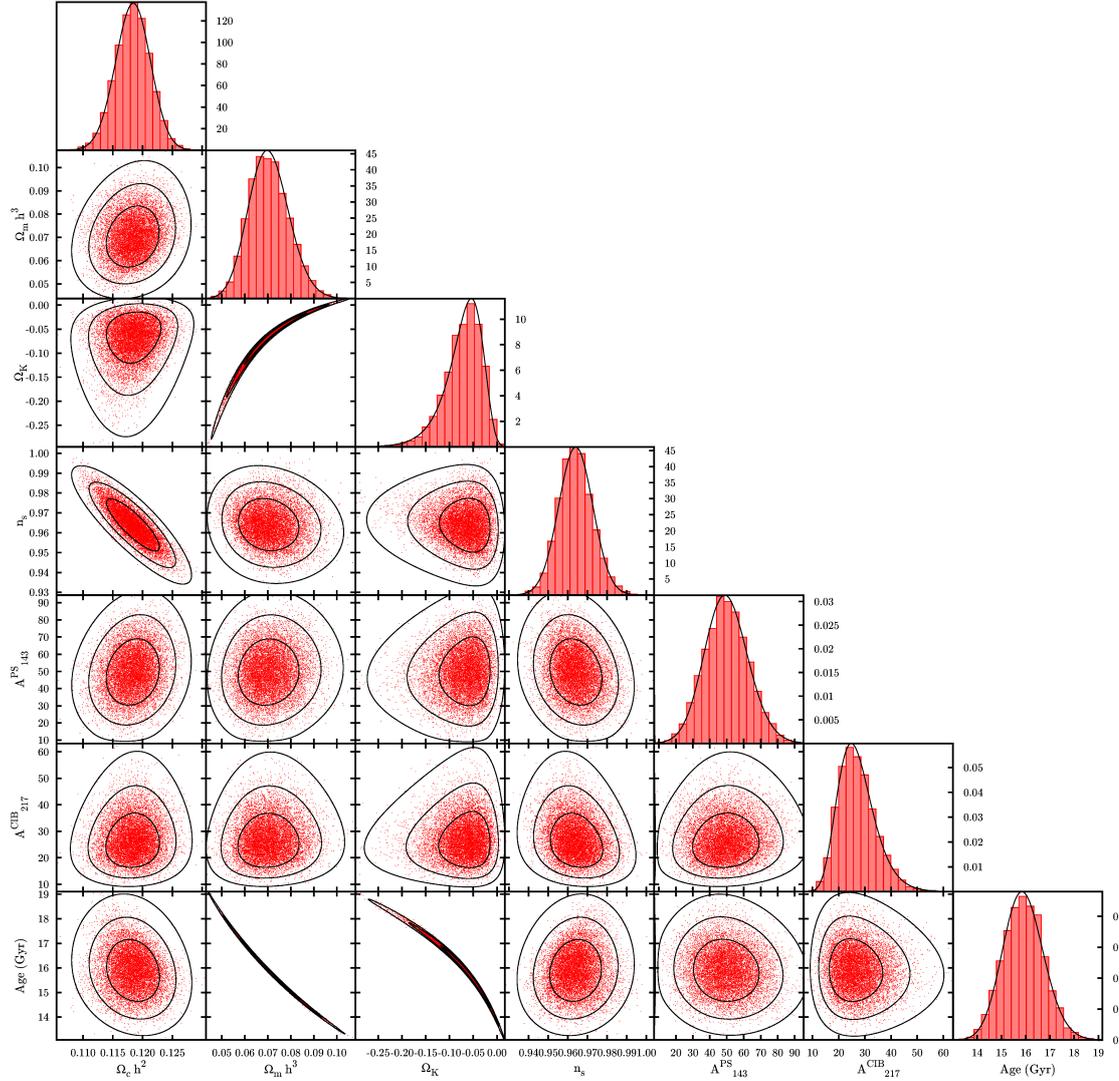


Figure 19: Reconstruction of a seven-dimensional *Planck* posterior density via a one-pass ABC transformation: 1D and 2D marginals. Black: marginal analytic posterior density (1D) or 1,2,3 σ contours. Red: marginal point sample distributions. For the 1D cases, the histograms have renormalised bar heights, to demonstrate the agreement with the value of the probability density.

4

A NOVEL WAY TO COMPUTE EVIDENCE

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another theory which states that this has already happened.

Douglas Adams

The Restaurant at the End of the Universe

The model evidence is the central tool to compare the predictivity of two different cosmological models in a Bayesian fashion – the evidence ratio allows the updating from prior to posterior odds ratios (see [Section 2.1.2](#)). In case that one model is nested in the other (just like flat Λ CDM resides inside curved Λ CDM as the hyperplane $\Omega_K = 0$), the evidence ratio is readily computed via the SDDR (Dickey, 1971; Verde, Feeney, et al., 2013). In the general case, however, no such shortcut exists. Several sampling algorithms that are popular in cosmology provide evidence estimates for the underlying likelihood and prior – these include PMC (Kilbinger, Wraith, et al., 2010) and MULTINEST (Skilling, 2006). However, the problem of evidence computation remains slow and computationally involved. We will demonstrate how to employ Gaussianisation as a new tool to access this problem, and compare to several existing methods.

4.1 ANALYTIC FORMULAE FOR EVIDENCE

The model evidence for a model \mathcal{M}_i , can be computed via

$$E_i = \mathcal{P}(\mathcal{D}|\mathcal{M}_i) = \int dX \mathcal{P}(\mathcal{D}|\underline{X}, \mathcal{M}_i) \mathcal{P}(\underline{X}|\mathcal{M}_i) \quad (90)$$

i.e. via integration of the (unnormalised) posterior density $\Pi(\underline{X}) = \mathcal{P}(\mathcal{D}|\underline{X}, \mathcal{M}_i) \mathcal{P}(\underline{X}|\mathcal{M}_i)$ over the respective parameter space of model \mathcal{M}_i – hence the term ‘marginal likelihood’ for E . If Π takes a form with non-Gaussian features, and if the model parameter space is high-dimensional, this integral itself is often difficult to calculate.

However, with a bijective transformation $T : \underline{X} \mapsto \underline{Y}$ that Gaussianises the unnormalised

posterior density $\tilde{\Pi}(\underline{Y}) = \Pi[T^{-1}(\underline{Y})] |d\underline{X}/d\underline{Y}|$, we can compute the evidence integral analytically. If $\tilde{\Pi}$ has the shape of an (unnormalised) multivariate Gaussian

$$\tilde{\Pi}(\underline{Y}) = \hat{\Pi} \exp \left[-\frac{1}{2}(\underline{Y} - \underline{\hat{\mu}})^T \tilde{\Sigma}^{-1}(\underline{Y} - \underline{\hat{\mu}}) \right] \quad (91)$$

with means $\underline{\hat{\mu}}$, covariance matrix $\tilde{\Sigma}$ and maximum $\hat{\Pi}$, the log-evidence reads

$$\ln E = \ln \hat{\Pi} + \frac{1}{2} \ln \det \tilde{\Sigma} + \frac{d}{2} \ln(2\pi). \quad (92)$$

Similar expressions for Gaussian posterior densities can be found in A. N. Taylor and Kitching (2010). To estimate $\hat{\Pi}$, we need the absolute normalisation of $\tilde{\Pi}$; hence this method can only be applied to samples which provide the values for Π (possibly also in the form of log-likelihood and log-prior). From these, we compute the values of $\ln \tilde{\Pi}(\underline{Y})$ on the optimally-Gaussianised sample by adding the logarithm of the transformation Jacobian, and then fit the parameters $\underline{\hat{\mu}}$, $\tilde{\Sigma}$, and $\hat{\Pi}$ of the Gaussian via least-squares regression. This can be performed analytically, and even be used to compute an error bar on the value of $\ln E$ – see [Appendix B](#) for details.

If the prior distribution for one [MP](#), and hence the posterior, is supported only on a finite interval, the same will hold true for the transformed [MP](#) if we restrict ourselves to one-pass transformations. If the sample size is large enough to properly represent the cutoff, the Gaussianisation transformation will alleviate this feature, but may not fully remove it. Assuming the marginal distribution to be Gaussian, when in reality we may deal with a truncated Gaussian, will lead to a systematic error in the evidence, so it is advantageous to remove these features before starting the search for optimally Gaussianising [TPs](#). [Appendix A](#) details ‘unboxing transformations’, which redefine the [MPs](#), mapping a finite open interval to the entire real line. In fact, it is also possible to use them for posterior density reconstruction, before Step 1 in [Section 3.2.4](#).

4.2 APPLICATION TO LOGNORMAL SIMULATED DATA

To demonstrate this idea, we compute the evidence integral first on a mock data set, and subsequently on real data from cosmology. For the former, we draw a random sample of length 10,000 from a ten-dimensional log-normal probability distribution, and assign to each point the value of the probability density function, multiplied with a factor of $E = \exp(5)$. All weights are set to unity. This mock sample is subjected first to the Gaussianisation procedure with one-pass [ABC](#) transformations (no unboxing), and then to the regression outlined in [Appendix B](#) to retrieve the best-fit estimate for $\ln E$ and its error bar.

To verify these, we determine the distribution of the estimator in Equation 92 by producing 1,000 bootstrap samples from the transformed sample, and computing $\ln E$ on each of them, together with its one-sigma confidence intervals. To increase the reliability of the optimally Gaussianising transformation, 24 independent Gaussianisation runs are started with randomised initial conditions; Figure 20 shows the results for these. Relative to the “true” value of 5, these agree to sub-percent

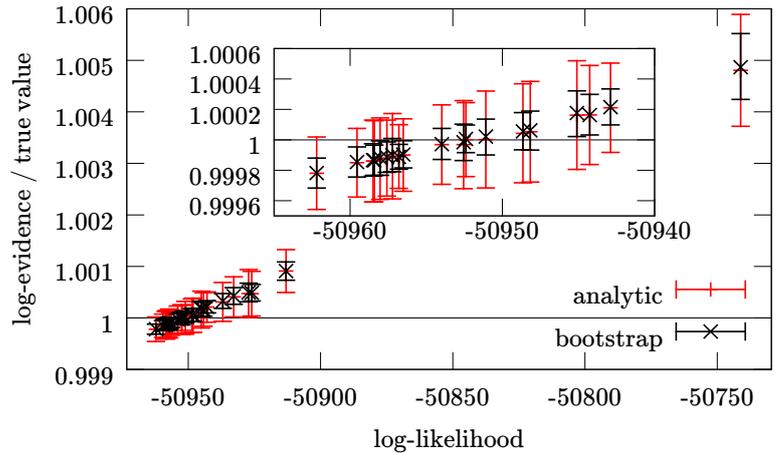


Figure 20: Ratio of log-evidence values for mock data set drawn from a 10D log-normal distribution, computed after 24 Gaussianisation runs to the true value, versus the negative log-likelihood for each run (including penalty term). Values and error bars shown are from analytic regression, as detailed in Appendix B (red), and the means and one-sigma confidence intervals from the bootstrapped distributions (black). The black line indicates the true value. *Inset*: zoomed-in version of the lower left corner.

accuracy. It is also noteworthy that a lower local maximum of the likelihood, i.e. one further to the right, will depart more from the true value. Hence, a location in TP space that is close to the exact optimum will yield a biased value for the log-evidence. This indicates that the evidence is a sensitive indicator for departures from Gaussianity. Note that a log-normal distribution can be precisely Gaussianised with Box-Cox transformations – and thus also by ABC transformations, which are a superset of these. Further, it is noteworthy that our analytic procedure yields error bars of the right magnitude, yet somewhat more conservative, compared to their bootstrapped counterparts. This discrepancy arises because the bootstrapped distributions for $\ln \det \hat{\Sigma}$ and $\ln \hat{\Pi}$ deviate slightly from Gaussianity, whereas the analytic error bars assume Gaussian error propagation (see Appendix B for details).

4.3 APPLICATION TO COSMOLOGICAL DATA

For a demonstration on real-world data, we Gaussianise the joint posterior distribution of MPs of data from weak lensing and baryon acoustic oscillations. The weak lensing data set is the 2D cosmic shear data taken by the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS; see Heymans, Van Waerbeke, et al. 2012; Kilbinger, Fu, et al. 2013). The CFHTLenS survey analysis combined weak lensing data processing with

THELI (Erben et al., 2013), shear measurement with lensfit (Miller et al., 2013), and photometric redshift measurement with PSF-matched photometry (Hildebrandt, Erben, et al., 2012). A full systematic error analysis of the shear measurements in combination with the photometric redshifts is presented in Heymans, Van Waerbeke, et al. (2012), with additional error analyses of the photometric redshift measurements presented in Benjamin et al. (2013).

The BAO data set is the Data Release 9 (DR9) CMASS sample from the Baryon Oscillation Spectroscopic Survey (BOSS), which is part of the Sloan Digital Sky Survey III (SDSS-III) – see Anderson et al. (2012). This contains 264,283 massive galaxies in a redshift range $0.43 < z < 0.7$, whose correlation function and power spectrum both exhibit the features of baryon acoustic oscillations. The quantity $d(z) = r_S(z_d)/D_V(z)$, i.e. the ratio of the comoving sound horizon r_S at the baryon drag epoch z_d and the spherically volume-averaged distance $D_V(z)$, is a probe of the underlying cosmological parameters – see Percival, Cole, et al. (2007) for details.

To draw samples from the posterior distribution, we use the CosmoPMC software package¹, which uses Population Monte Carlo (PMC), an algorithm to approximate the target distribution by a Gaussian mixture model. We compare three cosmological models: standard flat Λ CDM, curved Λ CDM, flat w CDM, and curved w CDM. The first has a four-dimensional parameter space spanned by matter density Ω_m , power spectrum normalisation σ_8 , baryon density Ω_b , and the normalised Hubble parameter h_{100} – all other parameters are set to their best fit values for flat Λ CDM, see Planck Collaboration (2015c). The latter two contain a fifth model variable each – curvature parameter Ω_K and constant dark energy equation-of-state parameter w , respectively. For all of these parameters, flat proper priors were chosen. The baseline model – flat Λ CDM is always referred to as model 1, whereas model 2 is one of the two extensions. As a byproduct of the sampling process, PMC provides the model evidence for the data set used – see Kilbinger, Wraith, et al. (2010) for further details.

In the special situation where one model is nested inside the other, the evidence ratio $B_{12} = E_1/E_2$ can be computed via the Savage-Dickey Density Ratio (SDDR) – see Section 2.1.2, and citations therein. Under mild conditions on prior and posterior densities for the full model \mathcal{M}_2 and the submodel \mathcal{M}_1 , the ratio can be derived to be

$$B_{12} = \frac{\mathcal{P}(\psi = \psi_{\text{sub}} | \mathcal{D}, \mathcal{M}_2)}{\mathcal{P}(\psi = \psi_{\text{sub}} | \mathcal{M}_2)}, \quad (93)$$

where ψ denotes the extra parameter (or parameters) contained in \mathcal{M}_2 but not in \mathcal{M}_1 , ψ_{sub} is the value of ψ that specifies the submodel \mathcal{M}_1 , and $\mathcal{P}(\psi | \mathcal{D}, \mathcal{M}_2)$ and $\mathcal{P}(\psi | \mathcal{M}_2)$ are posterior and prior densities of the full model, marginalised over all model parameters but ψ (see Section 2.1.2 for details).

¹ See <http://www2.iap.fr/users/kilbinger/CosmoPMC/>.

	flat Λ CDM	curved Λ CDM	flat w CDM
dimension	4	5	5
$\ln E$ (G)	486.96 ± 0.01	485.79 ± 0.03	486.09 ± 0.05
$\ln E$ (PMC)	487.02 ± 0.03	485.84 ± 0.01	486.00 ± 0.04
$\ln B_{12}$ (G)	n.a.	1.17 ± 0.04	0.87 ± 0.05
$\ln B_{12}$ (SDDR)	n.a.	1.23 ± 0.04	0.93 ± 0.06
$\ln B_{12}$ (PMC)	n.a.	1.19 ± 0.04	1.03 ± 0.05

Table 3: Values for evidence and Bayes factor for CFHTLenS+BOSS data set in three cosmological models, as computed with Box-Cox Gaussianisation (G) of weighted samples with 10,000 points each. For comparison: evidence value $\ln E$ and Bayes factor $\ln B_{12} = \ln E_{\text{base}} - \ln E_{\text{extension}}$ from Population Monte Carlo (PMC), and Bayes factor from Savage-Dickey Density Ratio (SDDR).

We find that the log-evidence values computed by CosmoPMC need to be offset by a factor of $n - 1$ times the log-prior density, where n is the number of data sets used. This is due to a non-standard interpretation of the prior density within CosmoPMC. Throughout this work, we apply this correction to the log-evidence values produced by CosmoPMC as well as the log-posterior values extracted from the CosmoPMC output. We follow the practice of (Kilbinger, Wraith, et al., 2010; Kilbinger, Fu, et al., 2013) of accepting a CosmoPMC run as soon as the built-in convergence diagnostic, called perplexity, exceeds a value of $p > 0.7$. Sampling to even higher values for the perplexity, up to $p \sim 0.95$, still changes the CosmoPMC value for $\ln E$ by as much as ~ 0.1 – this indicates a residual bias in the statistic. However, since the exact same offset has to appear in the CosmoPMC output values for the log-evidence $\ln E$ and for the non-normalised log-posterior $\ln \Pi(\underline{X})$, it is not of relevance to demonstrating our method, so investigating its origin is beyond the scope of this work.

Table 3 shows the log-evidences for the three models, and the Bayes factors of Λ CDM compared to either of the two extended models. The numbers in the first line were computed via one-pass Gaussianisation with ABC transformations, preceded by an unboxing transformation. To estimate the scatter of the CosmoPMC and SDDR values for $\ln E$ and $\ln B_{12}$ in the second, fourth, and fifth lines, we rerun CosmoPMC ten times for each model, and determine the mean and average for the CosmoPMC and SDDR estimators. Like for the log-normal sample, 24 independent Gaussianisation runs were started for each sample, and the one with the highest log-likelihood value chosen to transform the sample, which is then subjected to the analytic evidence computation procedure. The values in the first row of Table 3 are the weighted averages and standard deviations of all ten values, where the weights are determined from the analytic error bars as $w_i = \sigma_i^{-2}$ (see Appendix B).

The values show that the combined data favour Λ CDM over any of the extended models, although the evidence is not strong against either of the two. To sharpen the discrimination power of one model over the other, one might want to add further data sets, e. g., CMB data from the *Planck* satellite (Planck Collaboration, 2015a). This is problematic, since the posterior densities of *Planck* and CFHTLenS exhibit substantial tension in the measured values of H_0 , Ω_m , and σ_8 (MacCrann et al., 2015; Spergel, Flauger, and Hložek, 2015; Hildebrandt, Viola, et al., 2016). It is currently an unanswered question whether this discrepancy is due to physics beyond the cosmological standard model (or the standard model of particle physics, for that purpose), or due to unknown or mis-modelled systematic effects in the data analysis. However, the addition of CMB data can result in an increase of the evidence ratio, favouring Λ CDM over its extensions more strongly than it does with the combination of weak lensing and BAO data. The reason is that the normalised product of two discrepant probability densities will typically result in a density with smaller spread than the individual posteriors, even if the credible intervals for both original distributions are significantly removed from another. Since this increase in evidence does not correspond to a reliable sharpening of the parameter constraints, we will refrain from repeating this analysis including *Planck* data.

Our values agree with the numbers of SDDR and PMC within the spread between the latter two estimators, but still small deviations remain, which are larger than the error bars quoted. These may be due to residual non-Gaussianity in the transformed samples, to which the evidence is a sensitive measure.

This demonstrates the efficiency and practicality of Gaussianisation in cosmological model comparison, compared to existing algorithms.

GAUSSIANISATION OF CONVERGENCE FIELDS

“Come”, he said, sweeping through the door where Miss Janice Pearce sat glaring at a pencil, “let us go. Let us leave this festering hellhole. Let us think the unthinkable, let us do the undoable. Let us prepare to grapple with the ineffable itself, and see if we may not eff it after all.”

*Douglas Adams
Dirk Gently’s Holistic Detective Agency*

5.1 INTRODUCTION

The accurate description of random field distributions is an important and challenging problem for observational cosmology, especially if the random fields have non-Gaussian correlation properties. This is the case for the observables of weak gravitational lensing, i. e., convergence and cosmic shear. These are two-dimensional weighted projections of the dark matter density contrast δ_c , which is also non-Gaussian on small scales due to late-time nonlinear structure formation introducing mode couplings. It is possible to approximate these fields as Gaussian, and extract information only from the two-point statistics, i. e., correlation functions or power spectra (Kaiser, 1992; P. Schneider, van Waerbeke, et al., 2002; Kilbinger and P. Schneider, 2004; Joachimi, P. Schneider, and Eifler, 2008). However, the covariance matrix of a two-point estimator like the power spectrum will contain trispectrum contributions that become large on nonlinear scales, hence they need to be included into an accurate likelihood analysis of these scales. Modelling these covariances including the non-Gaussian contributions can be achieved by directly simulating the nonlinear collapse of cosmic structure and thus producing a large number of independent realisations of the convergence field, from which the sample covariance can be readily computed; these simulations, however, are computationally expensive and involved.

Additionally, the non-Gaussian correlation between modes will diminish the potential information gain from two-point estimators already on weakly nonlinear scales (Rimes and Hamilton, 2005; Neyrinck, Szapudi, and Rimes, 2006; Rimes and Hamilton, 2006;

Neyrinck and Szapudi, 2007; Carron and Szapudi, 2013; Carron, Wolk, and Szapudi, 2015), and potentially bias the power spectrum or correlation function estimates (Sato, Hamana, et al., 2009; Sato, Takada, et al., 2011).

On the other hand, the cosmological information residing in non-Gaussian higher-point observables promises to be comparable to the constraints from power spectra – works that explore bispectrum estimation as a tool to constrain parameters include Takada and Jain (2003, 2004, 2009).

It is therefore highly desirable for weak lensing inference to have a modelling toolbox for non-Gaussian random fields – to accurately access the information on non-linear scales from two-point statistics as well as higher orders. One strategy is to model the convergence as a *Gaussian-related random field*, i. e., one that can be transformed into an approximately Gaussian random field ζ via a smooth mapping $T : \kappa \mapsto \zeta$. A parametric ansatz for such a Gaussianising map which has been motivated from physics is the lognormal transformation; a non-parametric, data-driven mapping is provided by the Rosenblatt transformation (see Section 2.3.3 for definition and discussion of both).

It is noteworthy that the shapes of both transformations are inspired by the one-point distribution of convergence values, i. e., the histogram of κ -values. A priori it is not guaranteed that a mapping that approximately or even perfectly Gaussianises the one-point distribution will also Gaussianise the full random field, including all higher-dimensional marginal distributions. A well-known statistical tool to assess the higher-order correlation structure is the sequence of n -point copulas – given a random field X with one-point cumulative distribution function $C(x)$, the random field $Y = C(X)$ will have a uniform one-point distribution on $[0, 1]$. The n -point distribution of Y is a probability density on the n -dimensional hypercube, and known as copula. If X is a Gaussian field, then each n -point copula has a fixed shape known as Gaussian copula (Nelsen, 1999).

Now, if there is any such mapping that Gaussianises κ perfectly, all copulas of n th order ($n \geq 2$) have to be of Gaussian shape – this is the content of the Gaussian Copula Hypothesis (GCH) postulated by Scherrer et al. (2010). It is widely assumed to be valid, although this has never been demonstrated convincingly for higher orders, and the performance of lognormal and Rosenblatt transformation are limited (see Section 2.3.3 and Section 5.3 for details).

Therefore, it is of interest to choose a transformation based on its Gaussianising behaviour on the relevant higher-order statistics, rather than exclusively on the histogram. The benefit would be twofold: using the transformed convergence field as the basis for two-point statistic estimation increases the information content accessible to these estimators, since Gaussianisation captures the information from the higher orders. Additionally, the decorrelating effect on binned ℓ -ranges for the covariance matrix of power spectrum estimates more diagonal.

We propose a novel framework to search for parametric Gaussianising transformations

for non-Gaussian random fields, inspired by the Gaussianisation procedure of Joachimi and A. N. Taylor (2011) and Schuhmann, Joachimi, and Peiris (2016). The aim is to find an optimally Gaussianising transformation based on its performance on higher-point distributions, rather than just the histogram. This allows for efficient sampling from the random field distribution, i.e., producing mock maps that contain the full non-Gaussian correlation structure: the power spectrum of the Gaussianised maps can be used to produce Gaussian convergence maps; applying the inverse transformation to these yields independent realisations of the non-Gaussian convergence field, which can subsequently be used for, e.g., the estimation of convergence matrices. While our research was progressing, Y. Yu, Zhang, and Jing (2016) published this idea and demonstrated it using the Rosenblatt procedure. They find that this does not lead to a perfectly diagonal correlation matrix – especially on small scales, mode couplings remain, but the power spectra and covariance matrices can nevertheless be reproduced reasonably in the mildly nonlinear regime.

In contrast to their histogram-based approach, we will present a strategy that allows us to pick a transformation based on the higher-order correlations as well. We consider transformation families which are indexed by a sufficiently large parameter space; the optimally Gaussianising transformation parameters are found by optimising a loss function over these parameters. As samples from the random field distribution, we use mock convergence maps from N -body simulations. We will introduce these simulations in Section 5.2, introduce the parametric transformation family in Section 5.3, and describe two different approaches to finding a loss function for non-Gaussianity in Section 5.3.1 and Section 5.3.2. The results of our analysis will be detailed in Section 5.4, where we present the transformations themselves (Section 5.4.1) and their performance in simulating non-Gaussian mock maps (Section 5.4.2).

5.2 SIMULATED CONVERGENCE MAPS

We investigate the correlation properties of weak lensing convergence by studying two-dimensional simulated pixelised κ -maps. These have been produced with the high-performance CUBEP³M N -body simulation code¹ (Harnois-Déraps, Pen, et al., 2013). The Poisson equation of Newtonian gravity is solved on a mesh, tracking the three-dimensional clustering of dark matter particles in a cubic simulation box with a comoving width of 505 Mpc/ h . We use an extension of the simulation suite described in Harnois-Déraps and van Waerbeke (2015): flat Λ CDM has been assumed as a fiducial cosmological model, with the parameters $\Omega_m = 0.2905$, $\Omega_\Lambda = 0.7095$, $\Omega_b = 0.0473$, $n_s = 0.969$, $h = 0.6898$, and $\sigma_8 = 0.826$. The 3D dark matter distribution is collapsed into a discrete number of 2D mass sheets, which are subsequently used as thin lenses;

¹ See <https://github.com/jharno/cubep3m>.

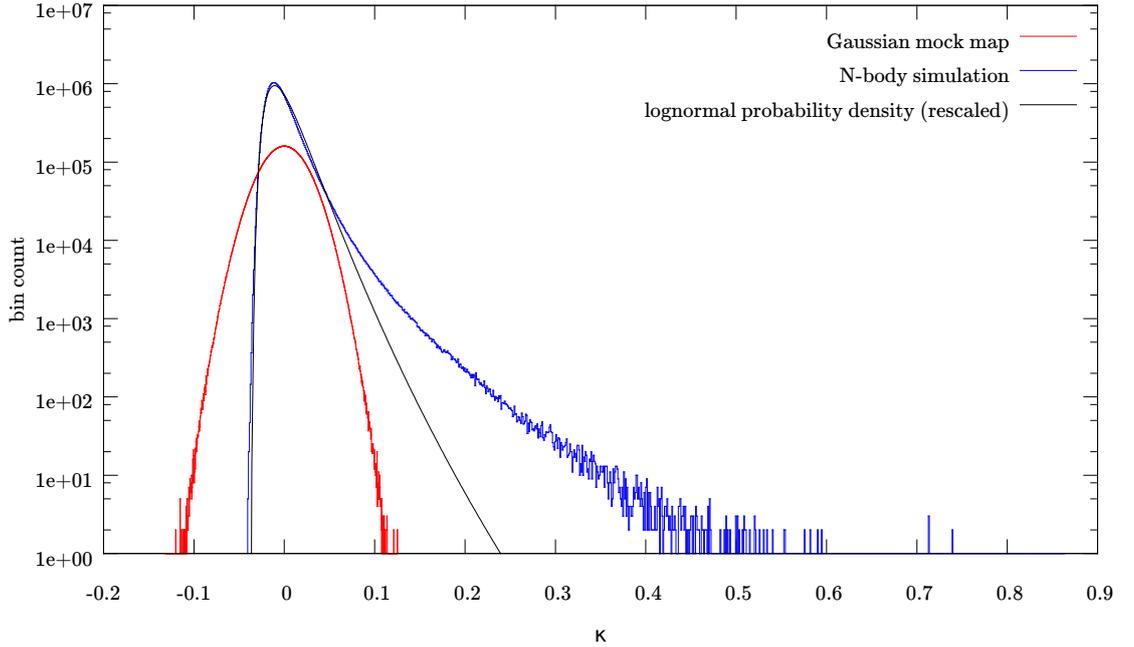


Figure 21: Histograms of one simulated convergence map from N -body simulation (blue) and one Gaussian mock map (red) – both fields have the same power spectrum and 6000^2 pixels each. The skewed deviation of the simulated map from Gaussianity in the one-point distribution is clearly visible. The black line is the lognormal probability density, rescaled to the normalisation of the histograms.

in the vacuum between these planes, light propagates along straight lines. The source galaxies are located $126.25 \text{ Mpc}/h$ behind each lens plane, the last one of which is at $z = 0.897$. This discretised ray-tracing scheme to construct the past light cone of null geodesics has been described in Martel, Premadi, and Matzner (2002), Harnois-Déraps, Vafaei, and Van Waerbeke (2012), and Harnois-Déraps and van Waerbeke (2015); it is a valid approximation to a fully three-dimensional approach (Couchman, Barber, and P. A. Thomas, 1999; Cooray and Hu, 2001; Hirata and Seljak, 2003). It should be noted that baryonic effects are absent from these mock maps since the simulated structure formation contains only dark matter. Further, shape noise is absent, i. e., the intrinsic scatter in the ellipticities of observed galaxies, which is an important modelling step for cosmic shear estimation. In spite of missing baryonic feedback, this suite is adequate for our purpose: the demonstration for our Gaussianisation method, which is not tied to a specific form or shape of non-Gaussianity. However, the influence of shape noise, which Joachimi, A. N. Taylor, and Kiessling (2011), Seo, Sato, Dodelson, et al. (2011), and Seo, Sato, Takada, et al. (2012) have shown to have a significant detrimental impact upon the Gaussianisation performance, needs to be investigated in further research.

Our simulation suite consists of 913 square maps of weak lensing convergence with an area of 60 square degrees each (this is the solid opening angle at the apex of the light

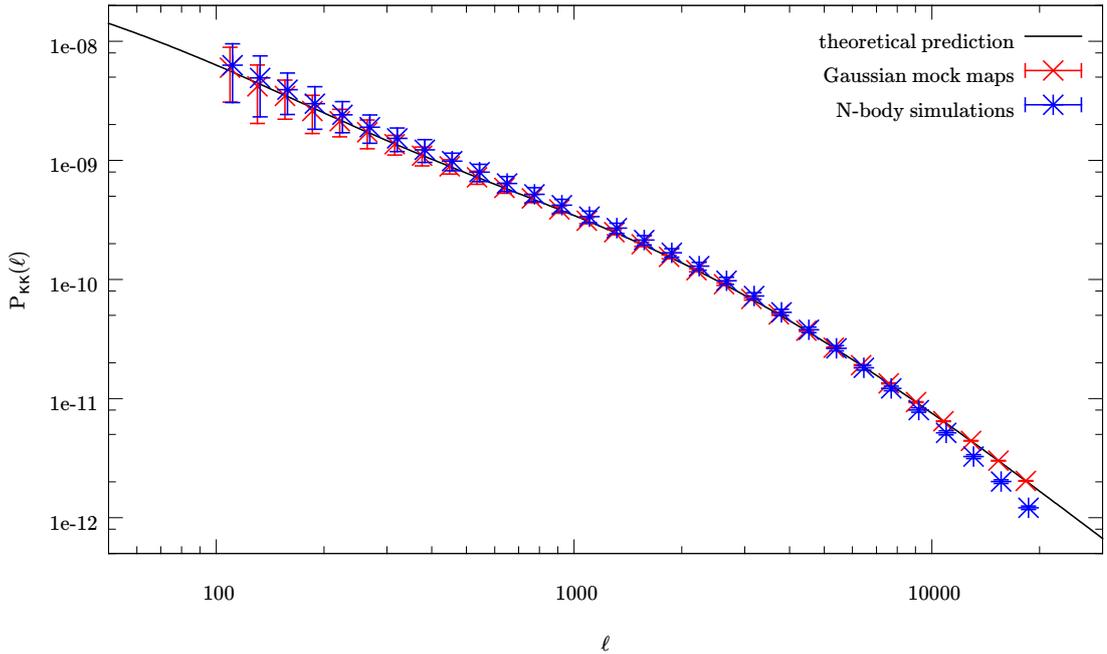


Figure 22: Convergence power spectrum estimates from N -body simulations (blue) and Gaussian realisations (red); also shown is the theoretical prediction (black). To aid visibility, the blue error bars have been offset slightly to the right.

cone), divided into 6000^2 quadratic pixels with an edge length of 4.648 arcseconds.

The histogram of κ values of a non-Gaussian convergence map is displayed in Figure 21, alongside the histogram of a Gaussian mock convergence map of identical dimensions: this has not been produced via an N -body simulation, but from the theoretically predicted convergence power spectrum for the fiducial cosmology cited above. The one-point distribution of the simulated convergence map is deviating clearly from Gaussianity, but also from a lognormal density especially in the region of large convergence values ($\kappa \gtrsim 0.1$).

Further, in Figure 22 we show power spectrum estimates averaged over all 913 non-Gaussian maps, as well as from the same number of Gaussian realisations; we also compare with the theoretical prediction for the power spectrum, including the nonlinear fitting formula described by Takahashi et al. (2012); shown is an angular frequency range of $\ell = 100 \dots 20,000$, logarithmically binned into 30 annulus-shaped radial bins.

A homogeneous random field has uncorrelated (but not necessarily independent) Fourier modes. If additionally the random field is Gaussian, then the modes need to be independent random variables. For band-averaged power spectrum estimates, this means that the covariance matrix of power bins is diagonal; any deviation from Gaussianity will correlate the power spectrum estimates in different bins. This is shown in the upper two panels of Figure 23 – whereas the Gaussian convergence field realisations indeed have a diagonal correlation matrix (upper right), the non-Gaussian maps show rich off-

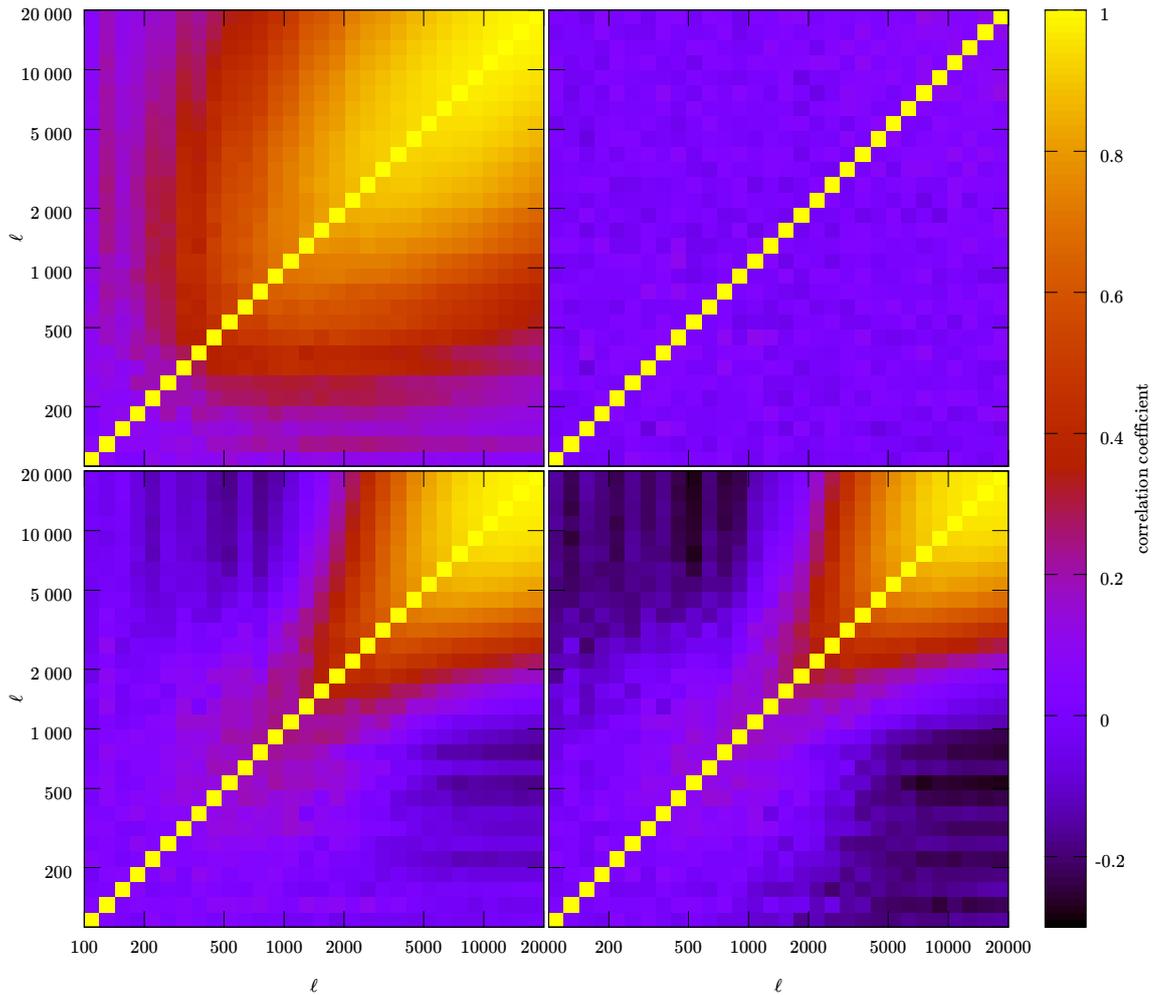


Figure 23: Sample correlation matrices of binned power spectra. *Top left*: untransformed N -body simulations; *top right*: Gaussian realisations; *bottom left*: N -body simulations after lognormal transformation; *bottom right*: N -body simulations after Rosenblatt transformation.

diagonal contributions in the power spectrum bins 10 and above (upper left) – this bin corresponds to a wave number of $\ell \sim 600$, i. e., an angular scale of roughly 36 arcminutes. In [Figure 22](#), this is slightly lower than the location of the nonlinear bump in the power spectrum (Takahashi et al., 2012); this corresponds to the findings of Meiksin and White (1999), Scoccimarro, Zaldarriaga, and Hui (1999), and Cooray and Hu (2001) that sizeable and non-negligible cross-correlations will already plague the mildly non-linear scales. We investigate the Gaussianisation performance of the transformations described in [Section 2.3.3](#), namely the lognormal map and the Rosenblatt map. We also show the correlation matrices of power spectrum estimates of the transformed N -body simulations in the lower half of [Figure 23](#). To ensure that every value of κ can be transformed, the parameter κ_0 in [Equation 80](#) is set to the absolute minimum of all maps. Both transformations succeed visibly in mitigating correlations between modes. The lognormal map

pushes the strongly correlated regime to higher wavenumbers ($\ell \gtrsim 1500$), at the price of introducing weak anticorrelations between high- ℓ and low- ℓ bins. The Rosenblatt map pushes the mode correlations higher, up to $\ell \gtrsim 2000$, but the anticorrelations in the wings are even stronger. For both Gaussianisation transformations, the off-diagonal terms in the correlation matrix between power spectrum bins become mitigated, but for none of them does the matrix become fully diagonal.

5.3 THE STRATEGY: PARAMETRIC TRANSFORMATIONS AND LOSS FUNCTION

Although widely assumed to be true in the literature, the status of the [GCH](#) is unclear. It has been demonstrated for the one-point distribution (Scherrer et al., 2010; Clerkin et al., 2016), but whether it holds even approximately for higher orders is less clear. Mathematically, given that the equivalence class of Gaussian-related random fields is only a vanishingly small subspace of the space of all random fields, it is unlikely to be valid for all orders. Also, it is easy to prove that if it holds, then the Rosenblatt transformation, which perfectly Gaussianises the one-point distribution, also must do the same with all higher orders. Thus, if the [GCH](#) held, the residual deviations from Gaussianity shown in the bottom right panel of [Figure 23](#) should be absent, indicating failure of the [GCH](#) at the critical fourth order – critical because it is the four-point distribution which needs to be captured accurately for inference on two-point statistics. Nevertheless, abandoning the hypothesis does not mean that the entire program of Gaussianisation for random fields is altogether forfeit – it opens up new directions for finding such transformations. So far, transformations have been motivated by their action on the histogram exclusively; this is especially true for the lognormal transform or the Rosenblatt transform. If the [GCH](#) holds only approximately, then it is worth searching for transformations based on their performance on higher orders: Gaussianising the histogram well is no longer an indicator for good Gaussianising performance on any higher order, or vice versa – Gaussianising the one-point distribution and the four-point distribution are different targets, and it is the latter that we aim for.

We will propose and investigate two different loss functions that have the potential to quantify non-Gaussianity of a random field in a higher-order distribution than the histogram. We then proceed to consider parametric transformation families, and minimise the non-Gaussianity of a convergence map over the transformation parameter space with standard optimisation algorithms. The first transformation family is a modification of the previously considered Box-Cox transformations

$$\kappa \mapsto BC_{(\kappa_0, \lambda)}^s(\kappa) = \begin{cases} \frac{1}{\lambda} \left[\left(1 + \frac{\kappa}{\kappa_0}\right)^\lambda - 1 \right] & (\lambda \neq 0) \\ \ln \left(1 + \frac{\kappa}{\kappa_0}\right) & (\lambda = 0). \end{cases} \quad (94)$$

This *Box-Cox-scaling transformation* is, like the usual Box-Cox transformation (Equation 87), regular across $\lambda = 0$, where it represents the mapping that perfectly Gaussianises a lognormal random field. To avoid undefined behaviour, the parameter κ_0 needs to be constrained to $\kappa_0 > \inf\{\kappa\}$.

In principle, any family of bijective and smooth transformations $T_\Delta : \mathbb{R} \rightarrow \mathbb{R}$ labelled by real parameters $\Delta = (\delta^1, \dots, \delta^n)$ could be considered, and failure to achieve Gaussianity can motivate the use of a more general family. The mapping is local, i. e., for every point $\underline{\theta}$, the non-Gaussian and (approximately) Gaussian random fields are related as $\zeta = T_\Delta[\kappa(\underline{\theta})]$. In principle, the transformation parameters Δ could vary from point to point, but we choose to restrict ourselves to global transforms – i. e., those where the same transformation parameters Δ are applied at each point. The reason is that this preserves the statistical homogeneity – if the n -point distributions of κ are shift-invariant, and the transformation is global, then so will the n -point distributions of ζ . Although this is not proven with mathematical rigour, we hypothesize that it is not only desirable but necessary to restrict ourselves to global transformations if we wish to preserve statistical homogeneity.

Parametric transformation families have been employed to Gaussianise the histogram of weak lensing convergence: Seo, Sato, Dodelson, et al. (2011) and Seo, Sato, Takada, et al. (2012) have demonstrated how the lognormal transformation can enhance the information content of convergence maps accessible to inference on two-point statistics, and tighten the constraints on dark energy. Joachimi, A. N. Taylor, and Kiessling (2011) have used the Box-Cox transformation in its original form, in order to improve upon the lognormal transform in reproducing the histogram. They found that freeing the parameter λ allows for a better fit especially of the high-convergence regions, which skew the distribution of untransformed κ values strongly to the right (see Figure 21). However, when studying the power spectra correlations of convergence maps transformed with the optimally histogram-Gaussianising Box-Cox transformations, significant residual non-Gaussianities remain in the higher orders, including the trispectrum.

5.3.1 *Projection-based Gaussianity test*

We implement the procedure of Cuesta-Albertos et al. (2007) and Nieto-Reyes, Cuesta-Albertos, and Gamboa (2009, 2014) (see Section 2.3.2) to devise a loss function quantifying deviations of a random field $\zeta(\underline{\theta})$ from Gaussianity, which takes as input a pixelised version $\{\zeta_{ij} = \zeta(\underline{\theta}_{ij})\}_{i,j=1}^N$. To this end, we will draw a pixelised random field

$h = \{h_{ij}\}_{i,j=1}^N$ from a dissipative probability distribution, and convolve the random fields to yield a new pixelised random field Y

$$Y_{ij} = (h * \zeta)_{ij} = \sum_{r,s=1}^N h_{rs} \zeta_{(i-r)(j-s)}. \quad (95)$$

The collection of all its values $\{Y_{ij}\}$ has the same distribution as the projection $\langle h | \zeta \rangle$; according to the theorem proven in Cuesta-Albertos et al. (2007), a Gaussianity test of this one-dimensional quantity at significance level α yields a Gaussianity test of the random field ζ at the same significance.

To produce the random element h , we follow Nieto-Reyes, Cuesta-Albertos, and Gamboa (2014) and draw samples from the Dirichlet distribution, which is the multivariate generalisation of the univariate Beta distribution, and dissipative in the technical sense. We implement the *stick-breaking technique* (Pitman, 2006) – here, $\beta(\alpha_1, \alpha_2)$ denotes the univariate *Beta distribution*, which is supported on the unit interval and depends on two fixed positive shape parameters. Its probability density function is

$$f_{\beta}(x; \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1}, \quad (96)$$

where $\Gamma(z) = \int_0^{\infty} dx x^{z-1} e^{-x}$ is the Gamma function. Its mean is equal to $\frac{\alpha_1}{\alpha_1 + \alpha_2}$; choosing $\alpha_1 > \alpha_2$ yields a distribution with its main mass in the lower half interval, and $\alpha_1 = 1$; $\alpha_2 \gg 1$ will result in concentrating the distribution in a small neighbourhood of zero. We implement the following algorithm:

- Draw $\tilde{h}_1 \in [0, 1]$ from $\beta(\alpha_1, \alpha_2)$.
- For $n = 2 \dots N^2$: draw $r_n \in [0, 1]$ from $\beta(\alpha_1, \alpha_2)$ and set $\tilde{h}_n = r_n \left(1 - \sum_{k=1}^{n-1} \tilde{h}_k\right)$.
- Let $\delta > 0$ fixed – for us, $\delta = 10^{-16}$. If there is an integer t ($1 < t < N^2$) such that $\sum_{k=1}^t \tilde{h}_k \geq 1 - \delta$, then find the smallest such t and set $\tilde{h}_T = 0 \forall T > t$.
- Distribute the N^2 numerical entries of $(\tilde{h}_n^{1/2})_{n=1}^{N^2}$ randomly into the field $(h_{ij})_{i,j=1}^N$.

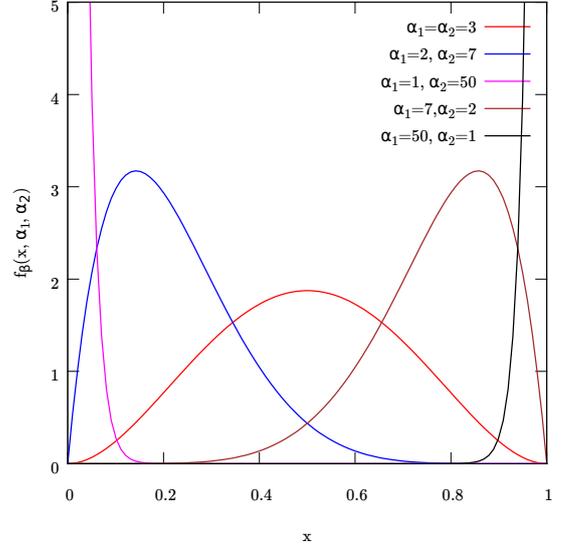


Figure 24: The probability density function of the beta distribution for different values of the shape parameters α_1 and α_2 .

The third step is necessary to ensure that h is properly normalised: $\|h\|^2 = \langle h|h \rangle = 1$. Without this procedure, round-off errors can accumulate catastrophically and lead to loss of significant digits.

The choice of the parameters α_1 and α_2 is crucial to the sensitivity of the test: if $\alpha_2 \gg \alpha_1$, then there will be many entries in h that are distinct from zero – however, performing the convolution [Equation 95](#) means building a linear combination of many components of κ , hence by the Central Limit Theorem our test can lose power to detect non-Gaussianity. On the other hand, if we choose the parameters to be too close to avoid the CLT, then there will be few entries in h , and the samples $\{Y_{ij}\}$ and $\{\zeta_{ij}\}$ will be too similar.

To avoid both pitfalls, Nieto-Reyes, Cuesta-Albertos, and Gamboa (2014) recommend producing K random fields $h^{(1,r)}$ with $(\alpha_1, \alpha_2) = (2, 7)$ and K fields $h^{(2,r)}$ with $(\alpha_1, \alpha_2) = (1, 100)$ ($r = 1 \dots K$). Then, for each of these $2K$ random fields:

- Perform the convolution ([Equation 95](#)) by multiplying $h^{a,r}$ and κ in Fourier space.
- Perform a Gaussianity test on the one-dimensional marginal sample $\{Y_{ij}^{(a,r)}\}_{i,j=1}^N$, which results in a p -value $p^{(a,r)}$. Our test of choice is Shapiro-Wilk – see [Section 2.3.1](#).

The $2K$ p -values are combined with the *False Discovery Rate* (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001) into a global p_0 for the entire test. Ranking the values in ascending order: $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(2K)}$, this is

$$p_0 = 2K \min_{j=1 \dots 2K} \left(j^{-1} p_{(j)} \right) \sum_{n=1}^{2K} \frac{1}{n}. \quad (97)$$

We will use this value p_0 as a loss function on the space of Gaussianising transformation parameters Δ : $\zeta(\underline{\theta}) = T_{\Delta}[\kappa(\underline{\theta})]$ and apply to the modified BoxCox transformations defined in [Equation 94](#).

To investigate the dependence of the false discovery rate on the transformation parameters, we choose $K = 4$, and proceed to compute p_0 for different values of λ

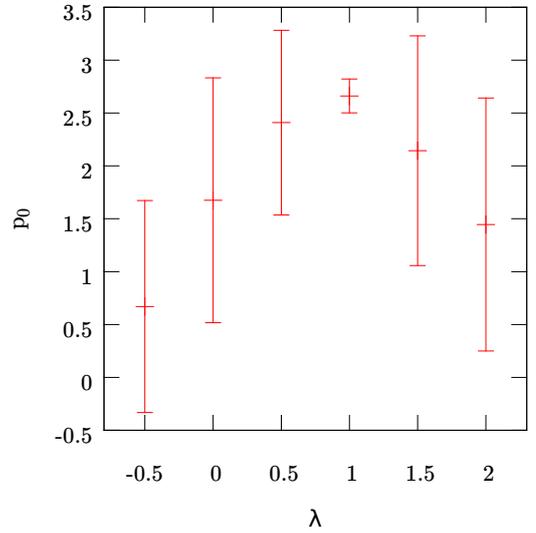


Figure 25: Testing the sensitivity of the loss function p_0 to non-Gaussianity. One Gaussian random field is transformed with a Box-Cox-Scaling transformation with varying λ ($\kappa_0 = 1$ is constant). Shown are mean and spread of eight independent evaluations of the loss function.

($\kappa_0 = 1$ is kept constant) as parameters of a Box-Cox-scaling transformation applied to a single Gaussian mock convergence map with 6000^2 pixels; for each point in parameter space we calculate the loss function eight times, each with new independent realisations of h . In [Figure 25](#), we show the mean and variance of these – whilst the average value reaches its maximum for $\lambda = 1$, the loss function cannot, within the uncertainty of a single evaluation, distinguish between a Gaussian ($\lambda = 1$) and a strongly non-Gaussian random field (e. g., $\lambda = 0$).

We therefore cannot confirm the statement of Nieto-Reyes, Cuesta-Albertos, and Gamboa (2009) that a Gaussianity test of the projection marginals, produced in the described fashion, yields a Gaussianity test of the full random field at the same significance threshold – at least not in our two-dimensional setting. We tried out other values for α_1, α_2 ; for neither of them is the sensitivity to non-Gaussianity large compared to the scatter. Increasing the number of projection directions per test might alleviate the large amount of scatter. However, the growing need for computation time and/or hard disk space does not make this method feasible, compared to the merit function we will discuss next, and will from now focus on exclusively.

5.3.2 Likelihood-based Gaussianity test

To quantify the Gaussianity of a random field, we can apply a maximum-likelihood formalism in the spirit of Box and D. R. Cox (1964), Velilla (1993), Joachimi and A. N. Taylor (2011), Joachimi, A. N. Taylor, and Kiessling (2011), and Schuhmann, Joachimi, and Peiris (2016). These authors have devised a likelihood for the parameters of Gaussianising transformations by inserting each point of a sample of a finite-dimensional random vector into a suitably transformed probability density (see [Section 3.2](#) for details), or in the case of Joachimi, A. N. Taylor, and Kiessling (2011) following the same strategy with the one-point distribution of a convergence map, i. e., all its entries pooled into one data vector. Our likelihood formalism differs from these earlier works in two important respects: the quantity whose Gaussianity is investigated is a pixelised random field in a high-dimensional space – 6000^2 in our case, whereas the number of samples – i. e., random field realisations – is significantly lower. Our log-likelihood is the logical equivalent of [Equation 85](#) for a pixelised random field $\kappa(\underline{\theta})$ instead of a low-dimensional random vector \underline{X} , hence we proceed in similar fashion: we apply a Gaussianising transformation $T_\Delta : \mathbb{R} \rightarrow \mathbb{R}$, which is a bijective map indexed by n real transformation parameters $\Delta = (\delta^1, \dots, \delta^n)$, to the full random field κ at every point: $\zeta(\underline{\theta}) = T_\Delta [\kappa(\underline{\theta})]$. Particularly, if $\Lambda = \{\underline{\theta}_{ij}\}_{i,j=1}^N$ is the set of $N \times N$ grid points, we set up the likelihood for

the transformed pixelised field $\zeta(\underline{\theta}_{ij}) = T_\Delta [\kappa(\underline{\theta}_{ij})]$. $\tilde{\Sigma}$ shall denote the full covariance matrix of the transformed convergence in all N^2 pixels, i. e.

$$\tilde{\Sigma}_{(ij)(kl)} = \text{Cov} [\zeta(\underline{\theta}_{ij}) \zeta(\underline{\theta}_{kl})]. \quad (98)$$

Then, the full log-likelihood given one convergence map $\mathcal{D} = \{\kappa(\underline{\theta}_{ij})\}$ as the data is

$$\mathcal{L}^{\text{full}}(\Delta|\mathcal{D}) = -\frac{1}{2} \ln \det \tilde{\Sigma} + \sum_{i,j=1}^N \ln \left| \frac{dT_\Delta}{d\kappa} [\kappa(\underline{\theta}_{ij})] \right|; \quad (99)$$

as in [Equation 85](#), any term without explicit or implicit Δ -dependency has been discarded. At first sight it seems difficult to estimate the Gaussianity of a distribution on $\mathbb{R}^{N \times N}$ from a single data point \mathcal{D} , or even a couple of hundred – but ergodicity of the random fields allows us to assess the full correlation structure from only a few fields, and judge its Gaussianity or deviations thereof.

In the task of estimating the determinant of the $N^2 \times N^2$ covariance matrix $\tilde{\Sigma}$, we make use of *Plancherel's theorem* (Plancherel and Mittag-Leffler, 1910), which states that the Fourier transform is a unitary map between functions – this directly relates the covariance matrix $\tilde{\Sigma}$ in position space and the covariance matrix $\tilde{\mathfrak{S}}$ in Fourier space,

$$\tilde{\mathfrak{S}}_{(ab)(cd)} = \text{Cov} [\zeta(\ell_{ab}) \zeta(\ell_{cd})], \quad (100)$$

where $\{\ell_{rs}\}$ is the grid in the Fourier domain. As a consequence, there is an $N^2 \times N^2$ unitary matrix U such that the two covariance matrices are related via

$$\tilde{\mathfrak{S}} = U \tilde{\Sigma} U^\dagger, \quad (101)$$

where U^\dagger denotes the Hermitian adjoint of U . Therefore, the determinants of both matrices $\tilde{\mathfrak{S}}$ and $\tilde{\Sigma}$ are identical. Since the random field is homogeneous, the covariance matrix of the Fourier modes is diagonal, and its log-determinant can be found by summing the logarithms of its eigenvalues:

$$\ln \det \tilde{\Sigma} = \sum_{i,j=1}^N \ln [4\pi^2 P_\zeta(\ell_{ij})]. \quad (102)$$

The evaluation of the full likelihood [Equation 99](#), given a convergence map κ and one point Δ in parameter space proceeds as follows:

- Compute the value for the transformation Jacobian $\left| \frac{dT_\Delta}{d\kappa} (\kappa_{ij}) \right|$ on each grid point, and sum their logarithms.
- Apply the transformation $T_\Delta : \kappa \rightarrow \zeta$.

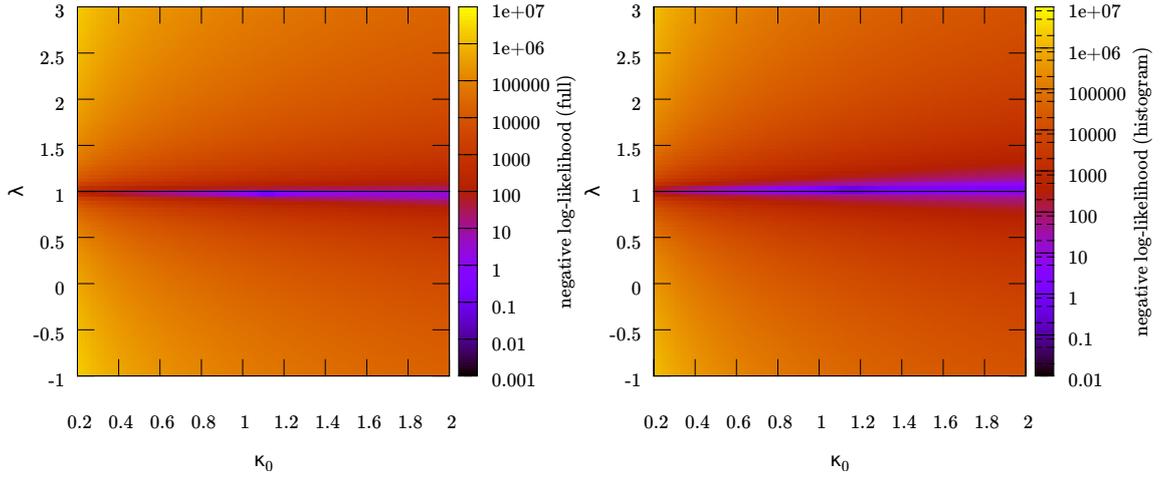


Figure 26: Negative log-likelihoods of one Gaussian random field with 6000^2 pixels, subjected to a Box-Cox transformation. *Left panel*: full log-likelihood (Equation 99); *Right panel*: histogram log-likelihood (Equation 103). The absolute values have been offset by a constant such that the minima are at $\mathcal{L} = 0$. The black lines are the linear transformations ($\lambda = 1$).

- Transform the field to $\zeta(\ell)$ via a Fast Fourier Transform (Frigo and Johnson, 2005).
- Bin the power into logarithmically spaced radial bins; find a cubic spline through these points which approximates the power spectrum $P_\zeta(\ell)$.
- Evaluate $P_\zeta(\ell)$ at each every grid point ℓ_{rs} , sum the logarithms of these numbers.

Any of these steps can be parallelised with ease; an implementation of this algorithm on 24 cores (2.7 GHz) with shared memory can evaluate the log-likelihood of Equation 99 for $N = 6000$ in under a minute.

It is instructive to contrast this loss function with the log-likelihood for the one-point distribution of the convergence map κ – this has been used by Joachimi, A. N. Taylor, and Kiessling (2011) to find the Box-Cox transformation which optimally Gaussianises the histogram. It is the one-dimensional form of Equation 85, where all grid information is discarded and all κ -values are pooled into one data vector:

$$\mathcal{L}^{\text{hist}}(\Delta|\mathcal{D}) = -\frac{N^2}{2} \ln \tilde{\sigma}^2 + \sum_{i,j=1}^N \ln \left| \frac{dT_\Delta}{d\kappa} \left[\kappa(\underline{\theta}_{ij}) \right] \right|, \quad (103)$$

where the variance of the one-dimensional random quantity ζ is estimated via the standard unbiased sample variance of the sample $\{\zeta_{ij} = \zeta(\underline{\theta}_{ij})\}$:

$$\tilde{\sigma}^2 = \frac{1}{N^2 - 1} \sum_{i,j=1}^N \left(\zeta_{ij} - \frac{1}{N^2} \sum_{i,j=1}^N \zeta_{ij} \right)^2. \quad (104)$$

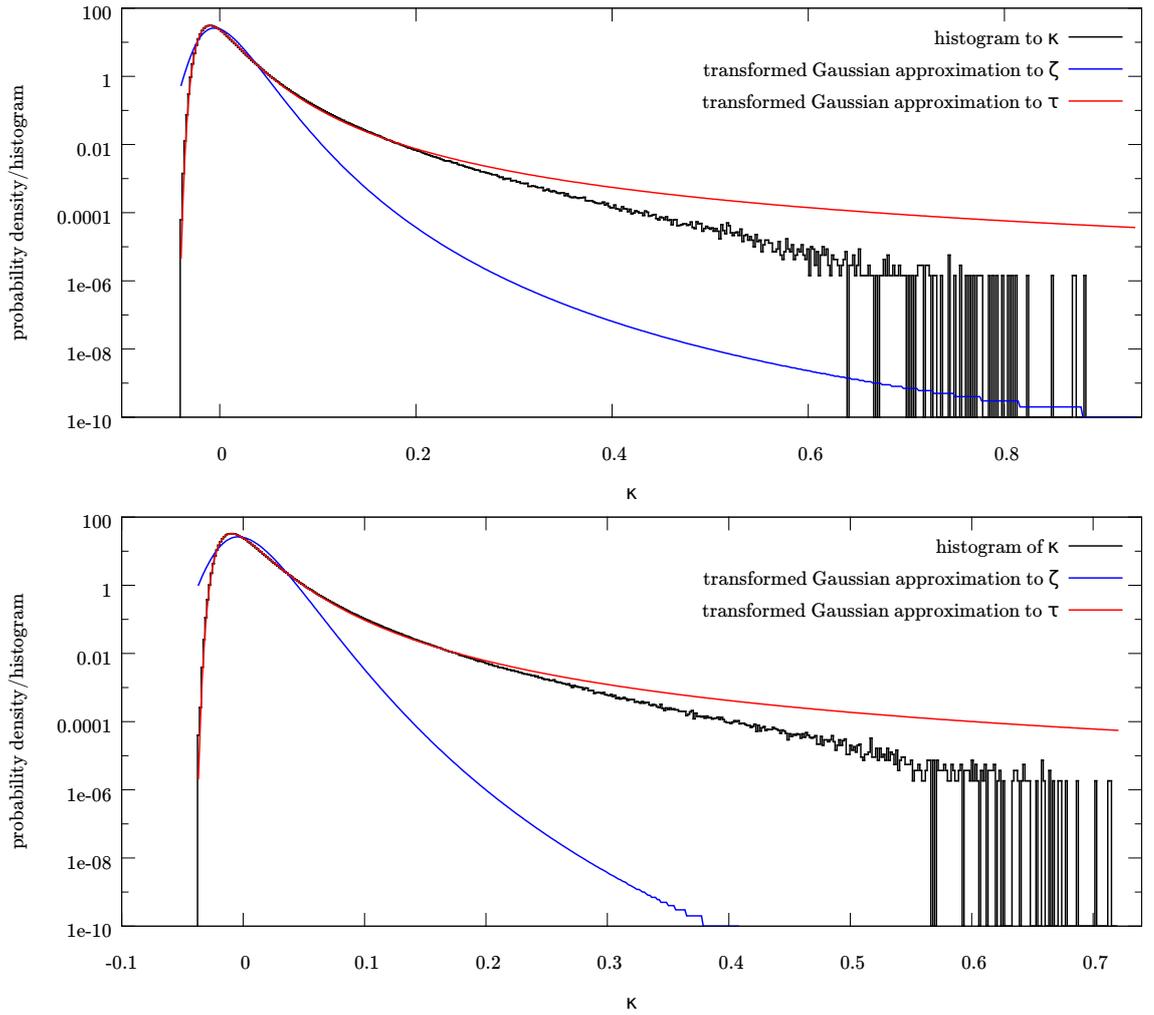


Figure 27: Untransformed histograms (black), compared to probability densities derived from assuming Gaussianity either for τ (red) or ζ (blue). The histograms have been normalised to the probability densities. *Left panel*: with smoothing at $\ell_s = 30,000$; *bottom panel*: no smoothing.

As a sanity check for our loss functions, we show the surface of the log-likelihoods [Equation 99](#) and [Equation 103](#) for a single Gaussian convergence map in [Figure 26](#). Both are gridded over the transformation parameters $\Delta = (\kappa_0, \lambda)$ of the modified Box-Cox transformation in [Equation 94](#). Since we begin with a Gaussian field, the Box-Cox transformations mapping a Gaussian into a Gaussian have to be located close to the line $\lambda = 1$, where we expect the maxima of both likelihoods – see [Figure 26](#). The full likelihood appears to be slightly more sensitive to deviations from Gaussianity, and both have the maximum ridge at the expected location.

smoothing	# fields	first:	κ_0	λ	second:	κ_0	λ
$\ell_s = 30,000$	10		0.055	-0.811		0.437	10.851
$\ell_s = 30,000$	1		0.053	-0.771		0.312	7.909
none	10		0.062	-0.925		0.322	6.347
none	1		0.061	-0.898		0.230	4.637

Table 4: Optimal transformation values, with or without smoothing, for different numbers of input convergence maps. The first transformation is found by optimising $\mathcal{L}^{\text{hist}}$, the second via \mathcal{L}^{opt} .

5.4 RESULTS

5.4.1 Gaussianising transformations from N -body simulations

We locate the transformation parameters that maximise the log-likelihood with the well-known Nelder-Mead amoeba algorithm (Nelder and Mead, 1965). Both loss functions exhibit degeneracies in the optimal Box-Cox-scaling parameters – Joachimi, A. N. Taylor, and Kiessling (2011) already explored these for $\mathcal{L}^{\text{hist}}$ in the context of the original Box-Cox transformations (Equation 87). To break these, we add a penalty term of the shape Equation 86, with the parameters $p = 6$ and $\epsilon = 1$, to the negative log-likelihood and minimize the combination of both over the transformation parameter space.

The convergence map will be Gaussianised with two subsequent transformations

$$\kappa \xrightarrow{\text{BC}_{\Delta_1}^s} \tau \xrightarrow{\text{BC}_{\Delta_2}^s} \zeta. \quad (105)$$

The first transformation parameters Δ_1 are chosen to optimise $\mathcal{L}^{\text{hist}}(\Delta|\kappa)$, and the second parameters Δ_2 such that they maximise $\mathcal{L}^{\text{full}}(\Delta|\tau)$. The first step removes the majority of non-Gaussianity in the histogram – since the one-point distribution is close to Gaussianity, this allows the second transformation to become more sensitive to the higher-point distributions, and mitigate any vestigial non-Gaussianity located there. Therefore, we expect the higher-than-first-order correlations of ζ to be more Gaussian in distribution than those of τ , which may come at the price at making the histogram of ζ less Gaussian than the histogram of τ .

To eliminate effects of the finite resolution of the N -body simulations, we apply Gaussian smoothing to the convergence maps prior to the optimisation procedure. This is performed by applying a fast Fourier transform, multiplying with a Gaussian kernel of width $\ell_s = 30,000$, and transforming back. This mode corresponds to an angular scale of about 22 arcseconds, which is the width of roughly five pixels.

Further, we test whether a single convergence map captures the random field distribution with sufficient accuracy, as postulated by ergodicity. To this end, we implement

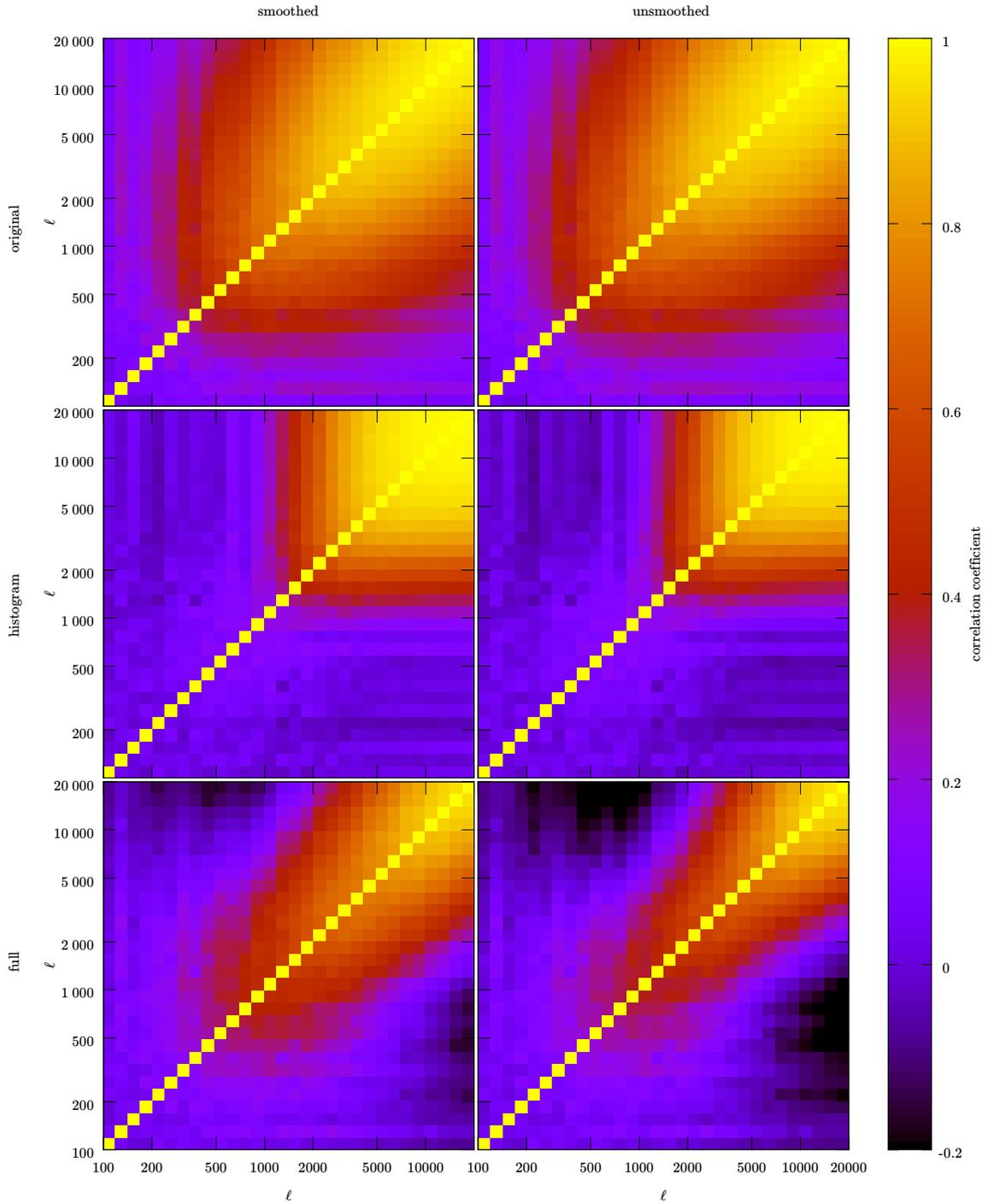


Figure 28: Sample correlation matrices of power spectra through the two-step Gaussianising transformation: *top row*: before the first transformation; *middle row*: after the histogram optimisation; *bottom row*: after the full optimisation. The spectra have been computed of 913 convergence maps, the optimisation procedure used ten. Shown are 30 power bins, logarithmically spaced between $\ell = 100$ and $\ell = 20,000$. *Left column*: Gaussian smoothing at $\ell_s = 30,000$; *right column*: no smoothing.

the log-likelihoods for multiple convergence maps and compare with the parameters for a single map: the transformation values presented in Table 4 show that the optimal parameter values do not differ significantly for the histogram Gaussianisation, whereas for the second transformation the parameters do differ between one and ten maps. Additionally, smoothing influences the optimal parameter values of both steps. From now on, we shall discuss only the transformations found by jointly Gaussianising ten maps, i. e., the first and third lines of Table 4.

In Figure 27 we show the histogram of the ten untransformed convergence maps (once with and once without smoothing), and compare with the analytic one-point density which can be derived from inverse-transforming a univariate Gaussian fitted to the histogram of either τ (red) or ζ (blue). Whereas the agreement between the first density and the histogram is excellent, the second one differs. This indicates the histogram of ζ is not as well described by a Gaussian as τ – which is just what we expected from our discussion above.

Further, Figure 28 shows the correlation matrices of the binned power spectra for 30 logarithmic bins between $\ell = 100$ and $\ell = 20,000$, with Gaussian pre-smoothing (left column), and without (right column). Top to bottom are the correlations of $P_\kappa(\ell)$, of $P_\tau(\ell)$, and of $P_\zeta(\ell)$. The first Gaussianising transformation already removes a substantial amount of non-Gaussianity – only above $\ell \sim 1600$ do off-diagonal cross-correlations between bins remain. In the subsequent step, non-Gaussianities leak back into lower power bins, but to the benefit of an overall reduction in the correlation coefficients: strong correlations above $r \sim 0.8$ remain only above $\ell \sim 7500$.

5.4.2 Fast generation of weak lensing covariances

This Gaussianisation procedure can be used to sample efficiently from the random field distribution. The distribution of a Gaussian random field is fully prescribed by its one- and two-point functions, i. e., mean and power spectrum. Given these two quantities, it is easy to produce independent realisations of the field. Using this, we can generate an arbitrary number of independent mock maps from a non-Gaussian random field κ which can be mapped to a Gaussian ζ with a bijective smooth map T , following the protocol:

- Measure the power spectrum P_ζ and the mean μ_ζ of the transformed field ζ .
- Produce \mathcal{N} independent realisations of ζ in Fourier space, by drawing the real and imaginary part of $\zeta(\ell)$ individually from a univariate Gaussian of mean zero and width $\sigma \propto \sqrt{P_\zeta}$ – with the exception of the zero mode $\zeta(\ell = 0)$;
- Apply an inverse Fourier transform to each set of modes, and shift every field by the mean μ_ζ . This will yield \mathcal{N} independent realisations of ζ in the spatial domain.

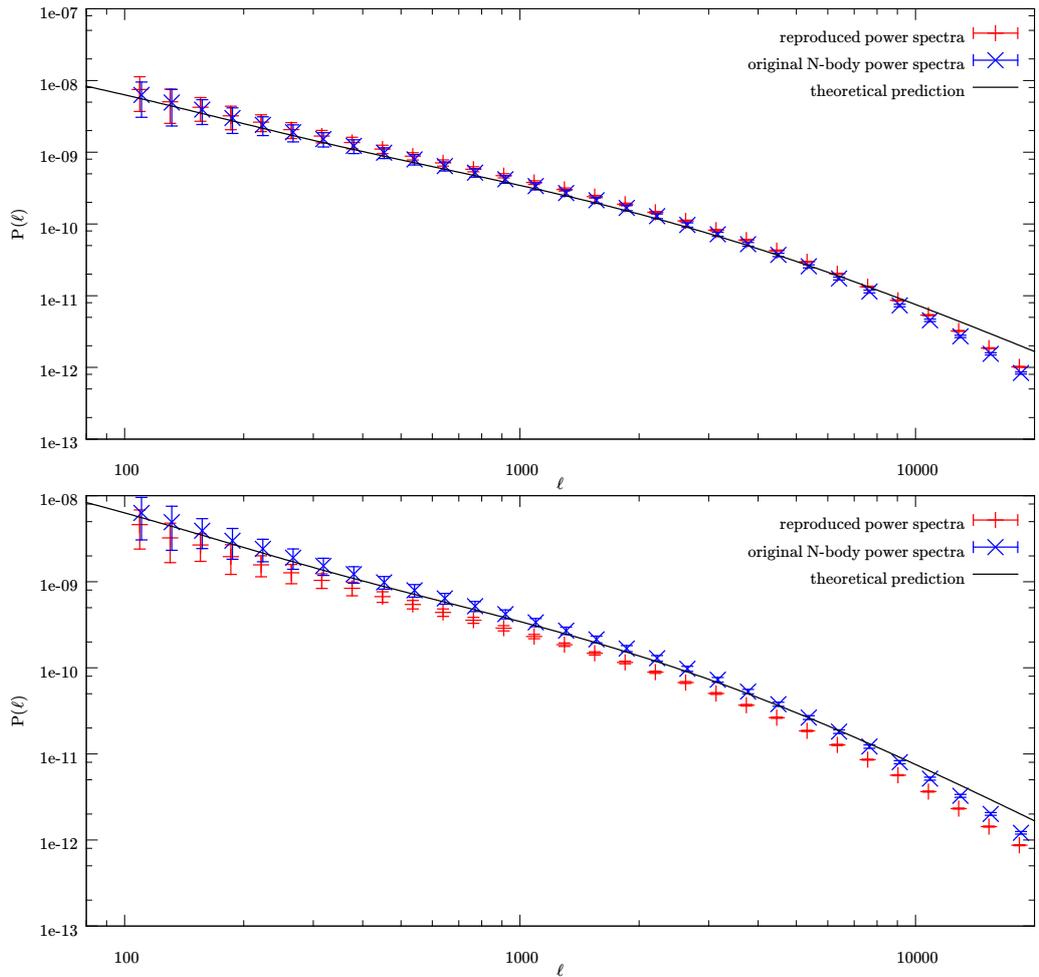


Figure 29: Estimated power spectra from back-transformed Gaussian mock convergence maps (red) and original N -body simulations (blue), in comparison to the theoretical prediction (black). *Top panel*: with Gaussian smoothing; *bottom panel*: no smoothing. The blue error bars have been offset slightly to the right to aid visibility.

- Apply T^{-1} , the inverse of the Gaussianising transformation, to each of the \mathcal{N} maps.

A very similar idea has recently been published by Y. Yu, Zhang, and Jing (2016). However, their quest for a Gaussianising transformation differs from ours in ansatz and in scope: they do not directly Gaussianise the convergence field κ , but rather the two-dimensional projected dark matter density contrast, projected onto the discrete mass sheets. Further, they restrict themselves to only the Rosenblatt map, which is histogram-based and non-parametric (see the lower right panel of Figure 23), whereas we attempt to find a mapping that Gaussianises the higher-point distributions explicitly. They find reasonable agreement between the power spectra of the N -body simulations and the back-transformed Gaussian mocks, but deviations appear on scales where non-Gaussianity becomes significant: their power spectrum estimates from mocks are biased downwards on intermediate scales $\ell \sim 500 \dots 1000$, and biased upwards for $\ell \gtrsim 1000$.

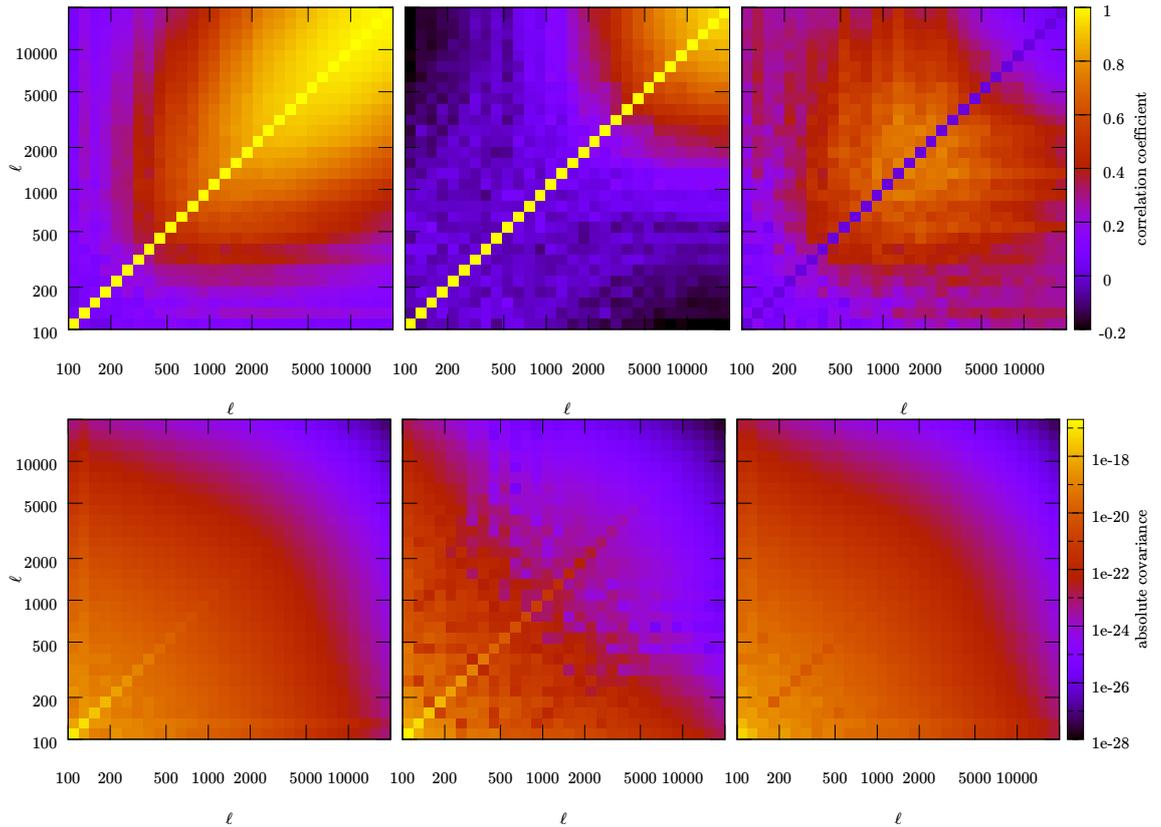


Figure 30: Comparison of the correlation matrix (top row) and covariance matrix (bottom row; absolute values) of binned power spectra. *Left panels:* from the original N -body simulations; *middle panels:* from our Gaussianisation method. *Right panels:* the absolute difference between both. These fields underwent Gaussian smoothing at $\ell_s = 30,000$.

To investigate the performance of our transformations, we measure the power spectra of the 913 convergence maps after applying the two-step transformation, find the mean power in each bin, and produce an equal number of Gaussian random fields of the same size and area. The inverse transformation is applied to these fields, and their power spectra are computed. [Figure 29](#) shows the power spectrum of the mock convergence maps which have been produced in this fashion, and compares it to the power spectrum of the original N -body simulations. With smoothing, the power spectra are well reproduced on large scales. Above $\ell \sim 5000$ the reproduced power spectra are slightly biased downwards. The deviations are more pronounced in the unsmoothed case, where the reproduced power spectra have an overall multiplicative bias. This is also present when comparing the width of the histograms – the reproduced convergence maps (unsmoothed) have a standard deviation of 0.0172, whereas the original N -body simulations have 0.0201. Note that the deviations between the measured spectra and the theoretical prediction on the very high scales are due to a discrepancy of our simulation suite and the fitting formula by Takahashi et al. (2012) – see also [Figure 22](#). Furthermore we show the covariance and correlation matrix of these fields in [Figure 30](#) and [Figure 31](#),

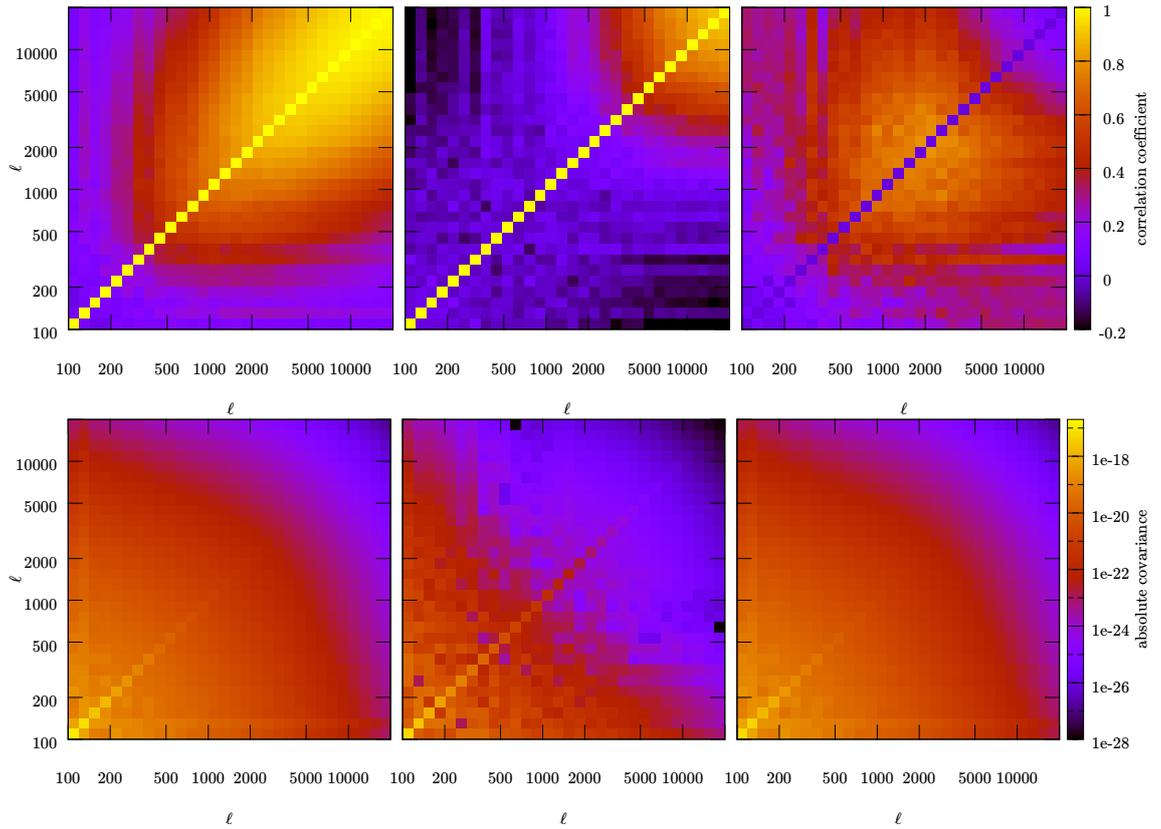


Figure 31: Like Figure 30, but without Gaussian smoothing.

next to their N -body counterparts, and the differences between the two. In both cases – smoothed and unsmoothed – the back-transformation method fails to reproduce the correlation structure. It appears that the residual post-transformation non-Gaussianity in the bottom row of Figure 28 is sufficient to cause these differences.

To find a transformation that accomplishes this, we propose the following modifications of the optimisation procedure:

- Using other, more sophisticated minimisation algorithms, such as simulated annealing (Černý 1985), or BOBYQA (Powell 2008).
- Using the log-likelihood of more than ten fields, to assure ergodicity.
- Introducing another Gaussianisation step by optimising $\mathcal{L}^{\text{hist}}$ once more, with the field ζ as input. This may remove the vestigial non-Gaussianity in the bottom row of Figure 28.

- Employing more general transformation families, e. g., a modified version of the ABC transforms defined in Equation 88:

$$\kappa \mapsto ABC_{(\kappa_0, \lambda, t)}^s(x) = \begin{cases} \frac{\kappa_0}{t} \sinh \left[\frac{t}{\kappa_0} BC_{(\kappa_0, \lambda)}^s(x) \right] & (t > 0) \\ BC_{(\kappa_0, \lambda)}^s(x) & (t = 0) \\ \frac{\kappa_0}{t} \operatorname{arcsinh} \left[\frac{t}{\kappa_0} BC_{(\kappa_0, \lambda)}^s(x) \right] & (t < 0). \end{cases} \quad (106)$$

Again, to avoid undefined behaviour, it is required to restrict the scaling parameter $\kappa_0 > \inf\{\kappa\}$.

- The balance between log-likelihood and penalty term (which is necessary to break the degeneracies in the likelihood) determines the numerical values of the optimal transformation parameters. A smaller penalty than the one we have employed may increase the sensitivity to the vestigial non-Gaussianity and thus result in better performance. At the same time, this may jeopardise the stability of the transformation found, since the inverse transformation becomes increasingly singular for growing values of λ .
- The computation of the log-likelihood Equation 99 can be restricted to the non-Gaussian scales by applying a suitable filter to the Fourier-transformed convergence maps or by a more general re-weighting scheme for different scales.
- It is possible to estimate the structure of the n -point copula empirically, i. e., from the distribution of the random fields (Charpentier, Fermanian, and Scaillet, 2007). This allows us to assess directly the quality of a transformation *a posteriori*.

CONCLUSIONS AND OUTLOOK

“An SEP”, [Ford] said, “is something that we can’t see, or don’t see, or our brain doesn’t let us see, because we think that it’s somebody else’s problem. That’s what SEP means. Somebody Else’s Problem. The brain just edits it out, it’s like a blind spot. If you look at it directly you won’t see it unless you know precisely what it is. Your only hope is to catch it by surprise out of the corner of your eye.”

*Douglas Adams
Life, the Universe, and Everything*

In this thesis, we have discussed how to transform non-Gaussian random variates approximately into Gaussianity; these variates can be finite- or infinite-dimensional. This has various applications in cosmological data analysis, which we subsequently introduced. The strength of the method stems from the fact that it gives us a model for the random variate in question, described by analytic expressions, which adequately captures the non-Gaussian structure.

6.1 FINITE DIMENSIONS: POSTERIOR DENSITY MODELLING

It is one of the fundamental tenets of the Bayesian paradigm that an experiment is a step from one probability distribution (the prior) to another (the posterior). To report this resultant probability density, it is customary to distribute large point samples, and to reconstruct the density from these – usually via KDE. Gaussianisation offers an elegant alternative that does not share many of the disadvantages that kernel density estimates exhibit.

6.1.1 Summary

Given a point sample from a posterior distribution, we describe how to find a Gaussianising transformation of the underlying probability density. From the parameters of the Gaussianised distribution and those of the transformation, we can reconstruct an ana-

lytic expression for the original probability distribution; this facilitates the combination of different data sets to obtain the joint posterior density. Further, this analytic posterior can be used to display contours of the density in question or its marginals, without the need for density estimates or smoothing procedures. Also, in reproducing the contours of the probability density reliably, it outperforms kernel density estimates (KDE), the standard tool in cosmology for reproducing a posterior density from a point sample. We suggest that, instead of distributing lengthy point samples in the form of a Markov chain, to use a Gaussianising transformation to disseminate a posterior density. Only the transformation parameters and the first and second moments of the resulting Gaussian are needed to reproduce the posterior density in its functional form; hence one can achieve substantial data compression.

We have demonstrated the posterior reproduction algorithm (Chapter 3) with our implementation in C, which employs MCMC samples from *Planck* data. The optimal Gaussianising transformation is found via a maximum-likelihood formalism. We employed Box–Cox transformations, and more flexible generalisations thereof, to Gaussianise various marginal distributions with distinctive non-Gaussian features, and showed the resulting contours. To assess the quality of a Gaussianising transformation, we have introduced the CC plot as a tool to decide whether one probability density reproduces the contours of another, or if not, to detect where they deviate.

One distinctive application of Gaussianising transformations, which we discuss and demonstrate in Chapter 4, is a novel method to compute the model evidence of a posterior distribution, given a point sample from it. We have tested this method on cosmological data from lensing and baryon acoustic oscillations, for different cosmological models, and find slight preference for Λ CDM. Compared to the numerical results from PMC and the SDDR, our new method of computing the evidence agrees well within the spread of the other two.

6.1.2 Future work

There are several possible extensions of our method as applied to posterior reconstruction, and directions to advance its scope. To optimise the Gaussianisation algorithm for speed and/or accuracy, it is possible to replace the Nelder–Mead minimum finder with more sophisticated minimisation algorithms like, e. g., BOBYQA (Powell, 2008) or simulated annealing (Černý, 1985). It is possible to engage new families of transformations, designed to cure a wider spectrum of non-Gaussian features that a multivariate probability density may possess – in our code, new families can easily be included.

Gaussianisation may be employed for fast sampling from a non-Gaussian probability density, in case that the Gaussianising parameters are either known exactly or to sufficient accuracy. Afterwards, it is possible to quickly draw a point sample from a

multivariate Gaussian distribution, and transform this sample with the inverse map. To improve the accuracy of the evidence computation, it is possible to replace the log-likelihood of [Equation 85](#)) with another loss function, which penalises deviations from Gaussianity in a sharper manner.

So far, we have been working with unimodal probability densities. We require the transformations to be bijective, hence we cannot map a multimodal distribution into a unimodal Gaussian. However, we may be able to transform such a density into a mixture of (possibly overlapping) Gaussians, where we now have to estimate the weight factor for each constituent from the transformed sample, in addition to their respective mean vectors and covariance matrices. The requisite number of components could be determined with standard clustering algorithms.

6.2 INFINITE DIMENSIONS: GAUSSIANISATION OF WEAK LENSING CONVERGENCE

Nonlinear structure formation results in random fields with non-Gaussian correlation structure, which creates several obstacles for cosmological inference: a wealth of information resides in the non-Gaussian order where it is inaccessible to two-point estimators like the power spectrum; additionally, accurate inference via two-point estimators requires knowledge of the four-point distribution – this is currently solved by simulating gravitational collapse, which is computationally costly and demanding. Gaussianising the convergence field has the potential to solve both problems. It should be noted that this application of Gaussianisation differs from the previous implementation in that we do not apply it to the posterior distribution of some model parameters, but rather directly to the probability distribution of the data itself.

6.2.1 *Summary*

In [Chapter 5](#), we present Gaussianising transformations for the distribution of simulated convergence maps, and demonstrate how they transform the correlation matrix of binned power spectra from one with rich non-Gaussian structure (represented by off-diagonal terms) into one that is close to Gaussian. Our maximum-likelihood formalism resembles the one used for posterior modelling, but is different in that we apply it to pixelised random fields, and there is only one global set of transformations. We have identified modified Box-Cox scaling transformations that bring the correlation matrix of Gaussianised convergence maps closer to Gaussianity.

We outline how these transformations can be used to efficiently produce independent mock convergence maps which capture the non-Gaussian structure. However, when the inverse transformation is applied to Gaussian realisations of the field with the same power spectrum of the Gaussianised maps, we find that the residual non-Gaussianity

does not allow for a faithful reconstruction of the non-Gaussian covariance matrix. We have listed modifications of the optimisation which can improve the performance.

6.2.2 *Future work*

Once a transformation is found which can reproduce the covariance matrix of the non-Gaussian simulated fields, various refinements, extensions and applications open up: The performance of log-normal and Rosenblatt transformations has been found to be limited once shape noise was introduced into the random field distribution. Depending on the level at which this noise is included in covariance and/or forward modelling, it is important to check the behaviour of our transformations in the presence of a realistic level of shape noise.

For a Gaussian random field with independent binned power spectrum estimates, the Fisher information (Fisher, 1935; Tegmark, A. N. Taylor, and Heavens, 1997) residing in the ℓ -ranges grows linearly with the number of modes (and quadratic if the bin spacing is logarithmic) – for the non-Gaussian convergence the information increase stalls once the non-Gaussian scales are reached. Our transformation can potentially push the location of this plateau further into the non-linear realm, or remove it altogether. Again, it is therefore important to compare its performance to lognormal and Rosenblatt models. This is tantamount to transporting information from higher-order distributions into the two-point function, where it can be accessed with power spectrum estimators. Another important check for the transformations is whether it can faithfully reproduce the bispectrum of the untransformed mocks. Being a three-point statistic, this observable is a vital part of the non-Gaussian correlation structure which is not directly accessible through power spectrum estimation, but holds a comparable amount of information. Provided that the information stems from scales that can be modelled accurately, a likelihood that is sensitive to the non-Gaussian correlation structure can be a central ingredient in a Bayesian hierarchical model.

We chose to restrict ourselves to the part of the four-point distribution that appears in the covariance matrix of binned power spectra, but there are other observables known to harbour non-Gaussian information, like weak lensing peak counts (Hamana, Takada, and Yoshida, 2004; Pires, Leonard, and Starck, 2012; Lin and Kilbinger, 2015a,b; Lin, Kilbinger, and Pires, 2016; Shirasaki, 2016). These can alternatively be used to assess the quality of our transformation.

In analogy to the analytic expression for the posterior density in the finite-dimensional case, it is possible to design a cosmological likelihood for the transformed convergence fields which is sensitive to the non-Gaussian structure and the information stored therein, and apply it to real data.

*“Every year is getting shorter, never seem to find the time.
Plans that either come to naught or half a page of scribbled lines
Hanging on in quiet desperation is the English way
The time is gone, the song is over,
Thought I'd something more to say ...”*

*Pink Floyd
Time (The Dark Side of the Moon)*

Part III

APPENDIX



UNBOXING TRANSFORMATIONS

The electronic I Ching calculator ... was much like an ordinary pocket calculator, except that the LCD screen was a little larger than usual, in order to accommodate the abridged judgements of King Wen on each of the sixty-four hexagrams, and also the commentaries of his son, the Duke of Chou, on each of the lines of each hexagram. These were unusual texts to see marching across the display of a pocket calculator, particularly as they had been translated from the Chinese via the Japanese and seemed to have enjoyed many adventures on the way.

The device also functioned as an ordinary calculator, but only to a limited degree. It could handle any calculation which returned an answer of anything up to "4". "1+1" it could manage ("2"), and "1+2" ("3") and "2+2" ("4") or "tan 74" ("3.4874145"), but anything above "4" it represented merely as "A Suffusion of Yellow".

Dirk was not certain if this was a programming error or an insight beyond his ability to fathom, but he was crazy about it anyway, enough to hand over £20 of ready cash for the thing.

*Douglas Adams
The Long Dark Tea-Time of the Soul*

A single model parameter Z , which is assumed to be constrained to an open interval (a, b) , is redefined via the unboxing transformation $U_{(a,b)} : (a, b) \rightarrow \mathbb{R}$

$$X = U_{(a,b)}(Z) = \frac{a+b}{2} + \frac{b-a}{\sqrt{2\pi}} \Phi^{-1} \left(\frac{Z-a}{b-a} \right), \quad (107)$$

where Φ^{-1} denotes the inverse of the cumulative distribution function of the Normal distribution:

$$\Phi(x) = \int_{-\infty}^x dy \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{y^2}{2} \right). \quad (108)$$

$U_{(a,b)}$, thus designed, has the following properties: it is bijective and smooth; the limits are $\lim_{Z \rightarrow a} U_{(a,b)}(Z) = -\infty$; $\lim_{Z \rightarrow b} U_{(a,b)}(Z) = +\infty$. Further, the midpoint of the inter-

val $m = \frac{1}{2}(a + b)$ is fixed: $U_{(a,b)}(m) = m$, $U'_{(a,b)}(m) = 1$. Sending the interval boundaries to infinity simultaneously will result in the identity transformation

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} U_{(a,b)}(Z) = Z. \quad (109)$$

This is a generalisation of the widely used probit transformation, which maps the unit interval onto the real numbers as $p \mapsto \Phi^{-1}(p)$. Our modified probit has one huge advantage for the subsequent search for a Gaussianising transformation: If Z , as a random variable, is uniformly distributed on (a, b) , then X is normally distributed with mean m and spread $(b - a)/\sqrt{2\pi}$.

In statistics, a frequently-used alternative to probit is the logit map $p \mapsto \ln(p/1 - p)$. For our purposes, however, the probit is preferable, since a similarly rescaled version of this logit-transformation would yield a distribution with excess kurtosis, instead of a Gaussian.

For a d -dimensional vector of model parameters $\underline{Z} = (Z_1, \dots, Z_d)$, constrained to intervals $(a_i, b_i) \ni Z_i$, we unbox each dimension separately, with the appropriate boundaries:

$$\underline{Z} \mapsto \underline{X} = \left[U_{(a_1, b_1)}(Z_1), \dots, U_{(a_d, b_d)}(Z_d) \right]. \quad (110)$$

Before starting the search for the Gaussianisation parameters, every point in the original sample is mapped through this transformation.

B

EVIDENCE REGRESSION

“Look”, said Arthur, “Would it save you a lot of time if I just gave up and went mad now?” For a while the aircar flew on in awkward silence.

*Douglas Adams
The Hitchhiker’s Guide to the Galaxy*

We outline how the computation of the log-evidence (see [Equation 92](#)) can be performed analytically, i.e. without numerical optimisation. Our data consist of a Gaussianised weighted sample of \mathcal{N} points in \mathbb{R}^d , $\{(\underline{Y}^a, w^a)\}_{a=1}^{\mathcal{N}}$ and the values of the transformed log-posterior on each of these points, $\{\ell^a\}_{a=1}^{\mathcal{N}}$. To fit a multivariate unnormalised Gaussian

$$\tilde{\Pi}(\underline{Y}) = \hat{\Pi} \exp \left[-\frac{1}{2} (\underline{Y} - \tilde{\underline{\mu}})^T \tilde{\underline{\Sigma}}^{-1} (\underline{Y} - \tilde{\underline{\mu}}) \right] \quad (111)$$

through the values of $\{\exp(\ell^a)\}_{a=1}^{\mathcal{N}}$, we use the regression model

$$\ell_{A,B,C}^{\text{model}}(\underline{Y}) = \underline{Y}^T A \underline{Y} + \underline{B}^T \underline{Y} + C, \quad (112)$$

which is linear in each of the $d(d+3)/2 + 1$ regression parameters: the upper-diagonal components of the symmetric matrix A , the components of vector \underline{B} , and the scalar C . Assuming independence and homoscedasticity, we arrive at our quantity to minimise,

$$\chi^2(A, \underline{B}, C) = \sum_{a=1}^{\mathcal{N}} w^a \left[\ell_{A,B,C}^{\text{model}}(\underline{Y}^a) - \ell^a \right]^2, \quad (113)$$

which is quadratic in every regression parameter. Thus, we can write all normal equations of the regression problem,

$$\frac{d\chi^2}{d\vartheta} \stackrel{!}{=} 0, \quad \vartheta \in \{A_{ij}, B_k, C\} \quad (114)$$

as a $[d(d+3)/2 + 1]$ -dimensional linear inhomogeneous vector equation, and solve via singular value decomposition. From the resulting values of (A, \underline{B}, C) , the parameters of the multivariate Gaussian [Equation 111](#) can readily be computed as

$$\tilde{\Sigma} = -\frac{1}{2}A^{-1}; \quad (115)$$

$$\tilde{\mu} = -\frac{1}{2}A^{-1}\underline{B}; \quad (116)$$

$$\ln \hat{\Pi} = C - \frac{1}{4}\underline{B}^T A^{-1}\underline{B}. \quad (117)$$

Furthermore, we can use the analytic regression procedure to find error bars on these estimators, and thus on $\ln E$. To this end, we can analytically find the covariance matrix Cov of all parameters $\{A_{ij}, B_i, C\}$ from the form of χ^2 and the transformed data set $\{(\underline{Y}^a, w^a, \ell^a)\}_{a=1}^N$, and then approximate the variances of $\ln \hat{\Pi}$ and of $\ln \det \tilde{\Sigma}$ by standard Gaussian error propagation. In particular:

$$\text{Var} [\ln E] = \text{Var} [\ln \hat{\Pi}] + \frac{1}{4}\text{Var} [\ln \det \tilde{\Sigma}] \simeq \underline{\Xi}^T \text{Cov} \underline{\Xi} + \frac{1}{4}\underline{\Psi}^T \text{Cov} \underline{\Psi} \quad (118)$$

with

$$\underline{\Xi} = \left(\frac{\partial \ln \hat{\Pi}}{\partial \theta_i} \right) = \begin{pmatrix} \vdots \\ \frac{1}{4}B'_m B'_n (2 - \delta_{mn}) & (m = 1\dots d, \\ \vdots & n = m\dots d) \\ \vdots \\ -\frac{1}{2}B'_k & (k = 1\dots d) \\ \vdots \\ 1 \end{pmatrix} \quad (119)$$

and

$$\underline{\Psi} = \left(\frac{\partial \ln \det \tilde{\Sigma}}{\partial \theta_i} \right) = \begin{pmatrix} \vdots \\ \frac{1}{4}(A^{-1})_{mn} (2 - \delta_{mn}) & (m = 1\dots d, \\ \vdots & n = m\dots d) \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (120)$$

where $B' = A^{-1}B$.

BIBLIOGRAPHY

- Adler, R. J. and J. E. Taylor (2007). *Random Fields and Geometry*. Springer. ISBN: 978-0-387-48112-8 (cit. on p. 54).
- Aghanim, N., S. Majumdar, and J. Silk (2008). “Secondary anisotropies of the CMB.” In: *Reports on Progress in Physics* 71.6, 066902, p. 066902. DOI: [10.1088/0034-4885/71/6/066902](https://doi.org/10.1088/0034-4885/71/6/066902). arXiv: [0711.0518](https://arxiv.org/abs/0711.0518) (cit. on p. 27).
- Akeret, J., S. Seehars, A. Amara, A. Refregier, and A. Csillaghy (2012). “CosmoHammer: Cosmological parameter estimation with the MCMC Hammer.” In: *ArXiv e-prints*. arXiv: [1212.1721](https://arxiv.org/abs/1212.1721) [[astro-ph.CO](https://arxiv.org/abs/1212.1721)] (cit. on p. 50).
- Alam, S. et al. (2016). “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample.” In: *ArXiv e-prints*. arXiv: [1607.03155](https://arxiv.org/abs/1607.03155) (cit. on p. 38).
- Anderson, L. et al. (2012). “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample.” In: *MNRAS* 427, pp. 3435–3467. DOI: [10.1111/j.1365-2966.2012.22066.x](https://doi.org/10.1111/j.1365-2966.2012.22066.x). arXiv: [1203.6594](https://arxiv.org/abs/1203.6594) (cit. on pp. 38, 84).
- Aristotle (ca. 350 BCE). *Posterior Analytics*. URL: <http://classics.mit.edu/Aristotle/posterior.html> (cit. on p. 45).
- Arnold, K. et al. (2010). “The POLARBEAR CMB polarization experiment.” In: *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V*. Vol. 7741. Proc. SPIE, pages. DOI: [10.1117/12.858314](https://doi.org/10.1117/12.858314) (cit. on p. 27).
- Audren, B., J. Lesgourgues, K. Benabed, and S. Prunet (2013). “Conservative constraints on early cosmology with MONTE PYTHON.” In: *J. Cosmology Astropart. Phys.* 2, 001, p. 001. DOI: [10.1088/1475-7516/2013/02/001](https://doi.org/10.1088/1475-7516/2013/02/001). arXiv: [1210.7183](https://arxiv.org/abs/1210.7183) (cit. on p. 50).
- Austermann, J. E. et al. (2012). “SPTpol: an instrument for CMB polarization measurements with the South Pole Telescope.” In: *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI*. Vol. 8452. Proc. SPIE, 84521E. DOI: [10.1117/12.927286](https://doi.org/10.1117/12.927286). arXiv: [1210.4970](https://arxiv.org/abs/1210.4970) [[astro-ph.IM](https://arxiv.org/abs/1210.4970)] (cit. on p. 27).
- Bacon, D. J., A. R. Refregier, and R. S. Ellis (2000). “Detection of weak gravitational lensing by large-scale structure.” In: *MNRAS* 318, pp. 625–640. DOI: [10.1046/j.1365-8711.2000.03851.x](https://doi.org/10.1046/j.1365-8711.2000.03851.x). arXiv: [astro-ph/0003008](https://arxiv.org/abs/astro-ph/0003008) (cit. on p. 34).
- Bakan, D. (1966). “The test of significance in psychological research.” In: *Psychological Bulletin* 66.6, pp. 423–437. DOI: [10.1037/h0020412](https://doi.org/10.1037/h0020412). URL: <http://dx.doi.org/10.1037/h0020412> (cit. on p. 54).

- Bardeen, J. M., J. R. Bond, N. Kaiser, and A. S. Szalay (1986). "The statistics of peaks of Gaussian random fields." In: *ApJ* 304, pp. 15–61. DOI: [10.1086/164143](https://doi.org/10.1086/164143) (cit. on pp. 28, 54).
- Bartelmann, M. (2010). "Gravitational lensing." In: *Classical and Quantum Gravity* 27.23, 233001, p. 233001. DOI: [10.1088/0264-9381/27/23/233001](https://doi.org/10.1088/0264-9381/27/23/233001). arXiv: [1010.3829](https://arxiv.org/abs/1010.3829) (cit. on pp. 30, 31).
- Bartelmann, M. and P. Schneider (2001). "Weak gravitational lensing." In: *Phys. Rep.* 340, pp. 291–472. DOI: [10.1016/S0370-1573\(00\)00082-X](https://doi.org/10.1016/S0370-1573(00)00082-X). arXiv: [astro-ph/9912508](https://arxiv.org/abs/astro-ph/9912508) (cit. on p. 30).
- Bartolo, N., E. Komatsu, S. Matarrese, and A. Riotto (2004). "Non-Gaussianity from inflation: theory and observations." In: *Phys. Rep.* 402, pp. 103–266. DOI: [10.1016/j.physrep.2004.08.022](https://doi.org/10.1016/j.physrep.2004.08.022). arXiv: [astro-ph/0406398](https://arxiv.org/abs/astro-ph/0406398) (cit. on p. 59).
- Bayes, M. and M. Price (1763). "An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S." In: *Philosophical Transactions of the Royal Society of London Series I* 53, pp. 370–418 (cit. on p. 41).
- Becker, M. R. et al. (2015). "Cosmic Shear Measurements with DES Science Verification Data." In: *ArXiv e-prints*. arXiv: [1507.05598](https://arxiv.org/abs/1507.05598) (cit. on p. 35).
- Benjamin, J. et al. (2013). "CFHTLenS tomographic weak lensing: quantifying accurate redshift distributions." In: *MNRAS* 431, pp. 1547–1564. DOI: [10.1093/mnras/stt276](https://doi.org/10.1093/mnras/stt276). arXiv: [1212.3327](https://arxiv.org/abs/1212.3327) (cit. on p. 84).
- Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, pp. 289–300. ISSN: 00359246. URL: <http://www.jstor.org/stable/2346101> (cit. on pp. 58, 96).
- Benjamini, Y. and D. Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency." In: *Ann. Statist.* 29.4, pp. 1165–1188. DOI: [10.1214/aos/1013699998](https://doi.org/10.1214/aos/1013699998). URL: <http://dx.doi.org/10.1214/aos/1013699998> (cit. on pp. 58, 96).
- Bennett, C. L. et al. (2003). "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results." In: *ApJS* 148, pp. 1–27. DOI: [10.1086/377253](https://doi.org/10.1086/377253). arXiv: [astro-ph/0302207](https://arxiv.org/abs/astro-ph/0302207) (cit. on p. 24).
- Berger, J. O., J. M. Bernardo, and D. Sun (2009). "The formal definition of reference priors." In: *ArXiv e-prints*. arXiv: [0904.0156](https://arxiv.org/abs/0904.0156) (cit. on p. 43).
- Bernardeau, F., S. Colombi, E. Gaztañaga, and R. Scoccimarro (2002). "Large-scale structure of the Universe and cosmological perturbation theory." In: *Phys. Rep.* 367, pp. 1–248. DOI: [10.1016/S0370-1573\(02\)00135-7](https://doi.org/10.1016/S0370-1573(02)00135-7). eprint: [astro-ph/0112551](https://arxiv.org/abs/astro-ph/0112551) (cit. on p. 59).

- Bernardeau, F. and L. Kofman (1995). "Properties of the cosmological density distribution function." In: *ApJ* 443, pp. 479–498. DOI: [10.1086/175542](https://doi.org/10.1086/175542). arXiv: [astro-ph/9403028](https://arxiv.org/abs/astro-ph/9403028) (cit. on p. 61).
- Bernardeau, F., Y. Mellier, and L. van Waerbeke (2002). "Detection of non-Gaussian signatures in the VIRMOS-DESCART lensing survey." In: *A&A* 389, pp. L28–L32. DOI: [10.1051/0004-6361:20020700](https://doi.org/10.1051/0004-6361:20020700). arXiv: [astro-ph/0201032](https://arxiv.org/abs/astro-ph/0201032) (cit. on pp. 34, 59).
- Beutler, F., C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson (2011). "The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant." In: *MNRAS* 416, pp. 3017–3032. DOI: [10.1111/j.1365-2966.2011.19250.x](https://doi.org/10.1111/j.1365-2966.2011.19250.x). arXiv: [1106.3366](https://arxiv.org/abs/1106.3366) (cit. on p. 38).
- Bianchi, E. and C. Rovelli (2010). "Why all these prejudices against a constant?" In: *ArXiv e-prints*. arXiv: [1002.3966](https://arxiv.org/abs/1002.3966) [[astro-ph.CO](https://arxiv.org/abs/astro-ph)] (cit. on p. 22).
- Blake, C. et al. (2011). "The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations." In: *MNRAS* 418, pp. 1707–1724. DOI: [10.1111/j.1365-2966.2011.19592.x](https://doi.org/10.1111/j.1365-2966.2011.19592.x). arXiv: [1108.2635](https://arxiv.org/abs/1108.2635) (cit. on p. 38).
- Blandford, R. D., A. B. Saust, T. G. Brainerd, and J. V. Villumsen (1991). "The distortion of distant galaxy images by large-scale structure." In: *MNRAS* 251, pp. 600–627. DOI: [10.1093/mnras/251.4.600](https://doi.org/10.1093/mnras/251.4.600) (cit. on p. 30).
- Boltzmann, L. (1884). "Über die Eigenschaften monozyklischer und anderer damit verwandter Systeme." In: *Wissenschaftliche Abhandlungen*: ed. by Friedrich Hasenöhr. Vol. 3. Cambridge: Cambridge University Press, pp. 122–152. DOI: [10.1017/CB09781139381444.009](https://doi.org/10.1017/CB09781139381444.009). URL: <https://www.cambridge.org/core/books/wissenschaftliche-abhandlungen-uber-die-eigenschaften-monozyklischer-und-anderer-damit-verwandter-systeme/F66D56EDF1B20E44C1AE533DE3A6F9E3> (cit. on p. 56).
- Bond, J. R. and G. Efstathiou (1984). "Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter." In: *ApJ* 285, pp. L45–L48. DOI: [10.1086/184362](https://doi.org/10.1086/184362) (cit. on p. 28).
- Bond, J. R. and A. S. Szalay (1983). "The collisionless damping of density fluctuations in an expanding universe." In: *ApJ* 274, pp. 443–468. DOI: [10.1086/161460](https://doi.org/10.1086/161460) (cit. on p. 28).
- Bondi, H. and T. Gold (1948). "The Steady-State Theory of the Expanding Universe." In: *MNRAS* 108, p. 252. DOI: [10.1093/mnras/108.3.252](https://doi.org/10.1093/mnras/108.3.252) (cit. on p. 19).
- Box, G. E. P. and D. R. Cox (1964). "An Analysis of Transformations." In: *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2, pp. 211–252. ISSN: 00359246. URL: <http://www.jstor.org/stable/2984418> (cit. on pp. 68, 69, 97).
- Brainerd, T. G., R. D. Blandford, and I. Smail (1996). "Weak Gravitational Lensing by Galaxies." In: *ApJ* 466, p. 623. DOI: [10.1086/177537](https://doi.org/10.1086/177537). arXiv: [astro-ph/9503073](https://arxiv.org/abs/astro-ph/9503073) (cit. on p. 32).

- Bridle, S. et al. (2009). “Handbook for the GREAT08 Challenge: An image analysis competition for cosmological lensing.” In: *Annals of Applied Statistics* 3, pp. 6–37. DOI: [10.1214/08-AOAS222](https://doi.org/10.1214/08-AOAS222). arXiv: [0802.1214](https://arxiv.org/abs/0802.1214) (cit. on p. 34).
- Brillinger, D. R. (2012). “An Introduction to Polyspectra.” In: *Selected Works of David Brillinger*. Ed. by Peter Guttorp and David Brillinger. New York, NY: Springer New York, pp. 149–172. ISBN: 978-1-4614-1344-8. DOI: [10.1007/978-1-4614-1344-8_10](https://doi.org/10.1007/978-1-4614-1344-8_10). URL: http://dx.doi.org/10.1007/978-1-4614-1344-8_10 (cit. on p. 56).
- Brooks, S., A. Gelman, G. Jones, and X. L. Meng (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press. ISBN: 9781420079425 (cit. on p. 49).
- Cappé, O., R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert (2007). “Adaptive Importance Sampling in General Mixture Classes.” In: *ArXiv e-prints*. arXiv: [0710.4242](https://arxiv.org/abs/0710.4242) [[stat.CO](https://arxiv.org/abs/0710.4242)] (cit. on p. 51).
- Cappé, O., A. Guillin, J.-M. Marin, and C. P. Robert (2004). “Population Monte Carlo.” In: *Journal of Computational and Graphical Statistics* 13, pp. 907–929 (cit. on p. 51).
- Carroll, S. M., W. H. Press, and E. L. Turner (1992). “The cosmological constant.” In: *ARA&A* 30, pp. 499–542. DOI: [10.1146/annurev.aa.30.090192.002435](https://doi.org/10.1146/annurev.aa.30.090192.002435) (cit. on pp. 22, 29).
- Carron, J. and I. Szapudi (2013). “Optimal non-linear transformations for large-scale structure statistics.” In: *MNRAS* 434, pp. 2961–2970. DOI: [10.1093/mnras/stt1215](https://doi.org/10.1093/mnras/stt1215). arXiv: [1306.1230](https://arxiv.org/abs/1306.1230) (cit. on p. 88).
- Carron, J., M. Wolk, and I. Szapudi (2015). “On the information content of the matter power spectrum.” In: *MNRAS* 453, pp. 450–455. DOI: [10.1093/mnras/stv1595](https://doi.org/10.1093/mnras/stv1595). arXiv: [1412.5511](https://arxiv.org/abs/1412.5511) (cit. on p. 88).
- Catelan, P. and L. Moscardini (1994). “Kurtosis and large-scale structure.” In: *ApJ* 426, pp. 14–18. DOI: [10.1086/174034](https://doi.org/10.1086/174034). eprint: [astro-ph/9308002](https://arxiv.org/abs/astro-ph/9308002) (cit. on p. 59).
- Caticha, A. (2008). “Lectures on Probability, Entropy, and Statistical Physics.” In: *ArXiv e-prints*. arXiv: [0808.0012](https://arxiv.org/abs/0808.0012) [[physics.data-an](https://arxiv.org/abs/0808.0012)] (cit. on pp. 41, 52).
- Černý, V. (1985). “Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm.” In: *Journal of Optimization Theory and Applications* 45.1, pp. 41–51. ISSN: 1573-2878. DOI: [10.1007/BF00940812](https://doi.org/10.1007/BF00940812). URL: <http://dx.doi.org/10.1007/BF00940812> (cit. on pp. 106, 109).
- Charpentier, A., J.-D. Fermanian, and O. Scaillet (2007). “The estimation of copulas: Theory and practice.” In: *Copulas: from theory to applications in finance*, pp. 35–62 (cit. on p. 107).
- Christensen, N., R. Meyer, L. Knox, and B. Luey (2001). “Bayesian methods for cosmological parameter estimation from cosmic microwave background measurements.” In: *Classical and Quantum Gravity* 18, pp. 2677–2688. DOI: [10.1088/0264-9381/18/14/306](https://doi.org/10.1088/0264-9381/18/14/306). arXiv: [astro-ph/0103134](https://arxiv.org/abs/astro-ph/0103134) (cit. on p. 49).

- Chu, M., M. Kaplinghat, and L. Knox (2003). "Normal Parameters for an Analytic Description of the Cosmic Microwave Background Cosmological Parameter Likelihood." In: *ApJ* 596, pp. 725–733. DOI: [10.1086/378039](https://doi.org/10.1086/378039). arXiv: [astro-ph/0212466](https://arxiv.org/abs/astro-ph/0212466) (cit. on p. 66).
- Chwolson, O. (1924). "Über eine mögliche Form fiktiver Doppelsterne." In: *Astronomische Nachrichten* 221, p. 329 (cit. on p. 30).
- Clerkin, L. et al. (2016). "Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps." In: *MNRAS*. DOI: [10.1093/mnras/stw2106](https://doi.org/10.1093/mnras/stw2106). arXiv: [1605.02036](https://arxiv.org/abs/1605.02036) (cit. on pp. 61, 93).
- Cole, S., W. J. Percival, et al. (2005). "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications." In: *MNRAS* 362, pp. 505–534. DOI: [10.1111/j.1365-2966.2005.09318.x](https://doi.org/10.1111/j.1365-2966.2005.09318.x). arXiv: [astro-ph/0501174](https://arxiv.org/abs/astro-ph/0501174) (cit. on pp. 35, 38).
- Cole, S., D. H. Weinberg, C. S. Frenk, and B. Ratra (1997). "Large-scale structure in COBE-normalized cold dark matter cosmogonies." In: *MNRAS* 289, pp. 37–51. DOI: [10.1093/mnras/289.1.37](https://doi.org/10.1093/mnras/289.1.37). eprint: [astro-ph/9702082](https://arxiv.org/abs/astro-ph/9702082) (cit. on p. 22).
- Coles, P. and B. Jones (1991). "A lognormal model for the cosmological mass distribution." In: *MNRAS* 248, pp. 1–13. DOI: [10.1093/mnras/248.1.1](https://doi.org/10.1093/mnras/248.1.1) (cit. on p. 61).
- Colombi, S. (1994). "A 'skewed' lognormal approximation to the probability distribution function of the large-scale density field." In: *ApJ* 435, pp. 536–539. DOI: [10.1086/174834](https://doi.org/10.1086/174834). arXiv: [astro-ph/9402071](https://arxiv.org/abs/astro-ph/9402071) (cit. on p. 61).
- Cooray, A. and W. Hu (2001). "Power Spectrum Covariance of Weak Gravitational Lensing." In: *ApJ* 554, pp. 56–66. DOI: [10.1086/321376](https://doi.org/10.1086/321376). arXiv: [astro-ph/0012087](https://arxiv.org/abs/astro-ph/0012087) (cit. on pp. 34, 59, 60, 90, 92).
- Couchman, H. M. P., A. J. Barber, and P. A. Thomas (1999). "Measuring the three-dimensional shear from simulation data, with applications to weak gravitational lensing." In: *MNRAS* 308, pp. 180–200. DOI: [10.1046/j.1365-8711.1999.02714.x](https://doi.org/10.1046/j.1365-8711.1999.02714.x). arXiv: [astro-ph/9810063](https://arxiv.org/abs/astro-ph/9810063) (cit. on p. 90).
- Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience. ISBN: 0471241954 (cit. on p. 52).
- Cowles, M. K. and B. P. Carlin (1996). "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review." In: *Journal of the American Statistical Association* 91.434, pp. 883–904. ISSN: 01621459. URL: <http://www.jstor.org/stable/2291683> (cit. on p. 50).
- Cox, R. T. (1946). "Probability, Frequency and Reasonable Expectation." In: *American Journal of Physics* 14, pp. 1–13. DOI: [10.1119/1.1990764](https://doi.org/10.1119/1.1990764) (cit. on p. 41).

- Crill, B. P. et al. (2008). “SPIDER: a balloon-borne large-scale CMB polarimeter.” In: *Space Telescopes and Instrumentation 2008: Optical, Infrared, and Millimeter*. Vol. 7010. Proc. SPIE, 70102P. DOI: [10.1117/12.787446](https://doi.org/10.1117/12.787446). arXiv: [0807.1548](https://arxiv.org/abs/0807.1548) (cit. on p. 27).
- Cuesta-Albertos, J. A., E. del Barrio, R. Fraiman, and C. Matrán (2007). “The random projection method in goodness of fit for functional data.” In: *Computational Statistics & Data Analysis* 51.10, pp. 4814–4831. ISSN: 0167-9473. DOI: [10.1016/j.csda.2006.09.007](https://doi.org/10.1016/j.csda.2006.09.007). URL: <http://www.sciencedirect.com/science/article/pii/S0167947306003306> (cit. on pp. 57, 94, 95).
- Dark Energy Survey Collaboration (2016). “The Dark Energy Survey: more than dark energy - an overview.” In: *MNRAS* 460, pp. 1270–1299. DOI: [10.1093/mnras/stw641](https://doi.org/10.1093/mnras/stw641). arXiv: [1601.00329](https://arxiv.org/abs/1601.00329) (cit. on p. 35).
- Dawson, K. S., J.-P. Kneib, et al. (2016). “The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data.” In: *AJ* 151, 44, p. 44. DOI: [10.3847/0004-6256/151/2/44](https://doi.org/10.3847/0004-6256/151/2/44). arXiv: [1508.04473](https://arxiv.org/abs/1508.04473) (cit. on p. 39).
- Dawson, K. S., D. J. Schlegel, et al. (2013). “The Baryon Oscillation Spectroscopic Survey of SDSS-III.” In: *AJ* 145, 10, p. 10. DOI: [10.1088/0004-6256/145/1/10](https://doi.org/10.1088/0004-6256/145/1/10). arXiv: [1208.0022](https://arxiv.org/abs/1208.0022) (cit. on p. 38).
- de Jong, J. T. A., G. A. Verdoes Kleijn, K. H. Kuijken, and E. A. Valentijn (2013). “The Kilo-Degree Survey.” In: *Experimental Astronomy* 35, pp. 25–44. DOI: [10.1007/s10686-012-9306-1](https://doi.org/10.1007/s10686-012-9306-1). arXiv: [1206.1254](https://arxiv.org/abs/1206.1254) (cit. on p. 35).
- de Laplace, P.-S. (1774). “Mémoire sur la probabilité des causes par les événements.” In: *Mémoires de l'Académie royale des Sciences* 6, pp. 621–656 (cit. on p. 41).
- (1785). “Mémoire sur les approximations des formules qui sont fonctions de très grands nombres.” In: *Mémoires de l'Académie royale des Sciences*, pp. 423–467 (cit. on p. 41).
- (1812). *Théorie analytique des probabilités*. Mme Ve Courcier. URL: <https://books.google.es/books?id=nQwAAAAAMAAJ> (cit. on p. 52).
- (1814). *Essai philosophique sur les probabilités*. Bachelier Imprimeur-Libraire (cit. on p. 41).
- de Moivre, A. (1738). *The doctrine of Chances*. Woodfall. URL: https://books.google.co.uk/books?id=PII%5C_AAAAcaAJ (cit. on p. 52).
- de Sitter, W. (1917). “On the relativity of inertia. Remarks concerning Einstein’s latest hypothesis.” In: *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences* 19, pp. 1217–1225 (cit. on p. 19).
- (1918). “On the curvature of space.” In: *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences* 20, pp. 229–243 (cit. on p. 19).
- Denuit, M. and J. Dhaene (2003). “Simple characterizations of comonotonicity and countermonotonicity by extremal correlations.” In: *Belgian Actuarial Bulletin* 3 (cit. on p. 62).

- Dickey, J. M. (1971). "The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters." In: *Ann. Math. Statist.* 42.1, pp. 204–223. DOI: [10.1214/aoms/1177693507](https://doi.org/10.1214/aoms/1177693507). URL: <http://dx.doi.org/10.1214/aoms/1177693507> (cit. on pp. 46, 81).
- Dineen, P. and P. Coles (2005). "Multivariate Non-Normality in the WMAP 1st Year Data." In: *ArXiv Astrophysics e-prints*. arXiv: [astro-ph/0511802](https://arxiv.org/abs/astro-ph/0511802) (cit. on p. 57).
- Dodelson, S. (2003). *Modern cosmology*. Academic Press. ISBN: 0-12-219141-2 (cit. on pp. 17, 19, 21, 24, 31).
- Dodelson, S. and M. D. Schneider (2013). "The effect of covariance estimator error on cosmological parameter constraints." In: *Phys. Rev. D* 88.6, 063537, p. 063537. DOI: [10.1103/PhysRevD.88.063537](https://doi.org/10.1103/PhysRevD.88.063537). arXiv: [1304.2593](https://arxiv.org/abs/1304.2593) [[astro-ph](https://arxiv.org/abs/astro-ph).C0] (cit. on p. 60).
- Doran, M. and C. M. Müller (2004). "Analyse this! A cosmological constraint package for CMBEASY." In: *J. Cosmology Astropart. Phys.* 9, 003, p. 003. DOI: [10.1088/1475-7516/2004/09/003](https://doi.org/10.1088/1475-7516/2004/09/003). eprint: [astro-ph/0311311](https://arxiv.org/abs/astro-ph/0311311) (cit. on p. 50).
- Douc, R., A. Guillin, J.-M. Marin, and C. P. Robert (2007a). "Convergence of adaptive mixtures of importance sampling schemes." In: *Ann. Statist.* 35.1, pp. 420–448. DOI: [10.1214/009053606000001154](https://doi.org/10.1214/009053606000001154). URL: <http://dx.doi.org/10.1214/009053606000001154> (cit. on p. 51).
- (2007b). "Minimum variance importance sampling via Population Monte Carlo." In: *ESAIM: Probability and Statistics* 11, pp. 427–447. ISSN: 1262-3318. DOI: [10.1051/ps:2007028](https://doi.org/10.1051/ps:2007028). URL: http://www.esaim-ps.org/article_S1292810007000286 (cit. on p. 51).
- Dyson, F. W., A. S. Eddington, and C. Davidson (1920). "A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919." In: *Philosophical Transactions of the Royal Society of London Series A* 220, pp. 291–333. DOI: [10.1098/rsta.1920.0009](https://doi.org/10.1098/rsta.1920.0009) (cit. on p. 30).
- Earman, J. and C. Glymour (1980). "The Gravitational Red Shift as a Test of General Relativity: History and Analysis." In: *Studies in History and Philosophy of Science Part A* 11.3, pp. 175–214 (cit. on p. 30).
- Efstathiou, G., W. J. Sutherland, and S. J. Maddox (1990). "The cosmological constant and cold dark matter." In: *Nature* 348, pp. 705–707. DOI: [10.1038/348705a0](https://doi.org/10.1038/348705a0) (cit. on p. 22).
- Einstein, A. (1911). "Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes." In: *Annalen der Physik* 340, pp. 898–908. DOI: [10.1002/andp.19113401005](https://doi.org/10.1002/andp.19113401005) (cit. on p. 30).
- (1916). "Die Grundlage der allgemeinen Relativitätstheorie." In: *Annalen der Physik* 354, pp. 769–822. DOI: [10.1002/andp.19163540702](https://doi.org/10.1002/andp.19163540702) (cit. on p. 16).
- (1917). "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie." In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften (Berlin)*, Seite 142-152. (Cit. on p. 18).

- Einstein, A. (1936). “Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field.” In: *Science* 84, pp. 506–507. DOI: [10.1126/science.84.2188.506](https://doi.org/10.1126/science.84.2188.506) (cit. on p. 30).
- Eisenstein, D. J. and W. Hu (1998). “Baryonic Features in the Matter Transfer Function.” In: *ApJ* 496, pp. 605–614. DOI: [10.1086/305424](https://doi.org/10.1086/305424). arXiv: [astro-ph/9709112](https://arxiv.org/abs/astro-ph/9709112) (cit. on pp. 28, 37, 38).
- Eisenstein, D. J., H.-J. Seo, E. Sirko, and D. N. Spergel (2007). “Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak.” In: *ApJ* 664, pp. 675–679. DOI: [10.1086/518712](https://doi.org/10.1086/518712). arXiv: [astro-ph/0604362](https://arxiv.org/abs/astro-ph/0604362) (cit. on p. 38).
- Eisenstein, D. J., H.-J. Seo, and M. White (2007). “On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter.” In: *ApJ* 664, pp. 660–674. DOI: [10.1086/518755](https://doi.org/10.1086/518755). arXiv: [astro-ph/0604361](https://arxiv.org/abs/astro-ph/0604361) (cit. on pp. 36, 37).
- Eisenstein, D. J., I. Zehavi, et al. (2005). “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies.” In: *ApJ* 633, pp. 560–574. DOI: [10.1086/466512](https://doi.org/10.1086/466512). arXiv: [astro-ph/0501171](https://arxiv.org/abs/astro-ph/0501171) (cit. on p. 38).
- Erben, T. et al. (2013). “CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey - imaging data and catalogue products.” In: *MNRAS* 433, pp. 2545–2563. DOI: [10.1093/mnras/stt928](https://doi.org/10.1093/mnras/stt928). arXiv: [1210.8156](https://arxiv.org/abs/1210.8156) (cit. on p. 84).
- Feldman, H. A., N. Kaiser, and J. A. Peacock (1994). “Power-spectrum analysis of three-dimensional redshift surveys.” In: *ApJ* 426, pp. 23–37. DOI: [10.1086/174036](https://doi.org/10.1086/174036). arXiv: [astro-ph/9304022](https://arxiv.org/abs/astro-ph/9304022) (cit. on p. 60).
- Fendt, W. A. and B. D. Wandelt (2007). “Pico: Parameters for the Impatient Cosmologist.” In: *ApJ* 654, pp. 2–11. DOI: [10.1086/508342](https://doi.org/10.1086/508342). arXiv: [astro-ph/0606709](https://arxiv.org/abs/astro-ph/0606709) (cit. on p. 66).
- Fisher, R. A. (1925). *Statistical Methods For Research Workers*. Cosmo study guides. Cosmo Publications. ISBN: 9788130701332. URL: <https://books.google.co.uk/books?id=4bTttAJR5kEC> (cit. on p. 53).
- (1935). “The Logic of Inductive Inference.” In: *Journal of the Royal Statistical Society* 98.1, pp. 39–82. ISSN: 09528385. URL: <http://www.jstor.org/stable/2342435> (cit. on p. 111).
- (1958). “The Nature of Probability.” In: *Centennial Review* 2, pp. 151–166 (cit. on p. 53).
- Fixsen, D. J., E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright (1996). “The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set.” In: *ApJ* 473, p. 576. DOI: [10.1086/178173](https://doi.org/10.1086/178173). arXiv: [astro-ph/9605054](https://arxiv.org/abs/astro-ph/9605054) (cit. on p. 24).
- Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). “emcee: The MCMC Hammer.” In: *PASP* 125, pp. 306–312. DOI: [10.1086/670067](https://doi.org/10.1086/670067). arXiv: [1202.3665](https://arxiv.org/abs/1202.3665) [[astro-ph](https://arxiv.org/abs/astro-ph).IM] (cit. on p. 50).

- Friedmann, A. (1922). “Über die Krümmung des Raumes.” In: *Zeitschrift für Physik* 10, pp. 377–386. DOI: [10.1007/BF01332580](https://doi.org/10.1007/BF01332580) (cit. on pp. 17, 18).
- (1924). “Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes.” In: *Zeitschrift für Physik* 21, pp. 326–332. DOI: [10.1007/BF01328280](https://doi.org/10.1007/BF01328280) (cit. on p. 17).
- Frigo, M. and S. G. Johnson (2005). “The Design and Implementation of FFTW3.” In: *Proceedings of the IEEE* 93.2, pp. 216–231. ISSN: 0018-9219. DOI: [10.1109/JPROC.2004.840301](https://doi.org/10.1109/JPROC.2004.840301) (cit. on p. 99).
- Fu, L. et al. (2014). “CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations.” In: *MNRAS* 441, pp. 2725–2743. DOI: [10.1093/mnras/stu754](https://doi.org/10.1093/mnras/stu754). arXiv: [1404.5469](https://arxiv.org/abs/1404.5469) (cit. on pp. 34, 59).
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis*. Third. CRC Press (cit. on pp. 41, 45, 49).
- Ghosh, J. K., M. Delampady, and T. Samantha (2006). *An Introduction to Bayesian Analysis*. ISBN: 9780387354330 (cit. on p. 43).
- Gilks, W. R., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis. ISBN: 9780412055515 (cit. on p. 49).
- Gil-Marín, H., J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider (2015). “The power spectrum and bispectrum of SDSS DR11 BOSS galaxies - I. Bias and gravity.” In: *MNRAS* 451, pp. 539–580. DOI: [10.1093/mnras/stv961](https://doi.org/10.1093/mnras/stv961). arXiv: [1407.5668](https://arxiv.org/abs/1407.5668) (cit. on p. 59).
- Goodman, S. N. (1992). “A comment on replication, P-values and evidence.” In: *Statistics in Medicine* 11.7, pp. 875–879. ISSN: 1097-0258. DOI: [10.1002/sim.4780110705](https://doi.org/10.1002/sim.4780110705). URL: <http://dx.doi.org/10.1002/sim.4780110705> (cit. on p. 54).
- (1999). “Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy.” In: *Annals of Internal Medicine* 130.12, pp. 995–1004. DOI: [10.7326/0003-4819-130-12-199906150-00008](https://doi.org/10.7326/0003-4819-130-12-199906150-00008). URL: <http://dx.doi.org/10.7326/0003-4819-130-12-199906150-00008> (cit. on p. 54).
- Grayson, J. A. et al. (2016). “BICEP3 performance overview and planned Keck Array upgrade.” In: *ArXiv e-prints*. arXiv: [1607.04668](https://arxiv.org/abs/1607.04668) [[astro-ph.IM](https://arxiv.org/archive/astro-ph)] (cit. on p. 27).
- Green, J. et al. (2012). “Wide-Field InfraRed Survey Telescope (WFIRST) Final Report.” In: *ArXiv e-prints*. arXiv: [1208.4012](https://arxiv.org/abs/1208.4012) [[astro-ph.IM](https://arxiv.org/archive/astro-ph)] (cit. on p. 39).
- Grenander, U. (1950). “Stochastic processes and statistical inference.” In: *Arkiv för Matematik* 1.3, pp. 195–277. ISSN: 1871-2487. DOI: [10.1007/BF02590638](https://doi.org/10.1007/BF02590638). URL: <http://dx.doi.org/10.1007/BF02590638> (cit. on p. 57).
- Gunn, J. E. (1967). “On the Propagation of Light in Inhomogeneous Cosmologies. I. Mean Effects.” In: *ApJ* 150, p. 737. DOI: [10.1086/149378](https://doi.org/10.1086/149378) (cit. on p. 30).

- Hacking, I. (2001). *An Introduction to Probability and Inductive Logic*. Cambridge University Press. ISBN: 9781139643610 (cit. on p. 41).
- Hamana, T., M. Takada, and N. Yoshida (2004). “Searching for Massive Clusters in Weak Lensing Surveys.” In: *Frontier in Astroparticle Physics and Cosmology*. Ed. by K. Sato and S. Nagataki. Vol. 6, p. 375 (cit. on pp. 63, 111).
- Hansen, F. K., D. Marinucci, P. Natoli, and N. Vittorio (2002). “Testing for non-Gaussianity of the cosmic microwave background in harmonic space: An empirical process approach.” In: *Phys. Rev. D* 66.6, 063006, p. 063006. DOI: [10.1103/PhysRevD.66.063006](https://doi.org/10.1103/PhysRevD.66.063006). arXiv: [astro-ph/0206501](https://arxiv.org/abs/astro-ph/0206501) (cit. on p. 57).
- Harnois-Déraps, J., U.-L. Pen, I. T. Iliev, H. Merz, J. D. Emberson, and V. Desjacques (2013). “High-performance P³M N-body code: CUBEP³M.” In: *MNRAS* 436, pp. 540–559. DOI: [10.1093/mnras/stt1591](https://doi.org/10.1093/mnras/stt1591). arXiv: [1208.5098](https://arxiv.org/abs/1208.5098) (cit. on p. 89).
- Harnois-Déraps, J., S. Vafaei, and L. Van Waerbeke (2012). “Gravitational lensing simulations - I. Covariance matrices and halo catalogues.” In: *MNRAS* 426, pp. 1262–1279. DOI: [10.1111/j.1365-2966.2012.21624.x](https://doi.org/10.1111/j.1365-2966.2012.21624.x). arXiv: [1202.2332](https://arxiv.org/abs/1202.2332) (cit. on p. 90).
- Harnois-Déraps, J. and L. van Waerbeke (2015). “Simulations of weak gravitational lensing - II. Including finite support effects in cosmic shear covariance matrices.” In: *MNRAS* 450, pp. 2857–2873. DOI: [10.1093/mnras/stv794](https://doi.org/10.1093/mnras/stv794). arXiv: [1406.0543](https://arxiv.org/abs/1406.0543) (cit. on pp. 89, 90).
- Hartlap, J., T. Schrabback, P. Simon, and P. Schneider (2009). “The non-Gaussianity of the cosmic shear likelihood or how odd is the Chandra Deep Field South?” In: *A&A* 504, pp. 689–703. DOI: [10.1051/0004-6361/200911697](https://doi.org/10.1051/0004-6361/200911697). arXiv: [0901.3269](https://arxiv.org/abs/0901.3269) (cit. on p. 34).
- Hartlap, J., P. Simon, and P. Schneider (2007). “Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix.” In: *A&A* 464, pp. 399–404. DOI: [10.1051/0004-6361:20066170](https://doi.org/10.1051/0004-6361:20066170). arXiv: [astro-ph/0608064](https://arxiv.org/abs/astro-ph/0608064) (cit. on p. 60).
- Hastings, W. K. (1970). “Monte Carlo sampling methods using Markov chains and their applications.” In: *Biometrika* 57.1, pp. 97–109. DOI: [10.1093/biomet/57.1.97](https://doi.org/10.1093/biomet/57.1.97). eprint: <http://biomet.oxfordjournals.org/content/57/1/97.full.pdf+html>. URL: <http://biomet.oxfordjournals.org/content/57/1/97.abstract> (cit. on p. 49).
- Hawking, S. W. and G. F. R. Ellis (1973). *The large-scale structure of space-time*. Cambridge University Press. ISBN: 0-521-09906-4 (cit. on p. 16).
- Henderson, S. W. et al. (2016). “Advanced ACTPol Cryogenic Detector Arrays and Readout.” In: *Journal of Low Temperature Physics* 184, pp. 772–779. DOI: [10.1007/s10909-016-1575-z](https://doi.org/10.1007/s10909-016-1575-z). arXiv: [1510.02809](https://arxiv.org/abs/1510.02809) [[astro-ph.IM](https://arxiv.org/abs/1510.02809)] (cit. on p. 27).

- Heymans, C., L. Van Waerbeke, et al. (2012). “CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey.” In: MNRAS 427, pp. 146–166. DOI: [10.1111/j.1365-2966.2012.21952.x](https://doi.org/10.1111/j.1365-2966.2012.21952.x). arXiv: [1210.0032](https://arxiv.org/abs/1210.0032) (cit. on pp. [34](#), [83](#), [84](#)).
- Heymans, C. et al. (2013a). “CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments.” In: MNRAS 432, pp. 2433–2453. DOI: [10.1093/mnras/stt601](https://doi.org/10.1093/mnras/stt601). arXiv: [1303.1808](https://arxiv.org/abs/1303.1808) (cit. on p. [22](#)).
- (2013b). “CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments.” In: MNRAS 432, pp. 2433–2453. DOI: [10.1093/mnras/stt601](https://doi.org/10.1093/mnras/stt601). arXiv: [1303.1808](https://arxiv.org/abs/1303.1808) (cit. on pp. [34](#), [44](#), [52](#)).
- Hilbert, S., J. Hartlap, and P. Schneider (2011). “Cosmic shear covariance: the log-normal approximation.” In: A&A 536, A85, A85. DOI: [10.1051/0004-6361/201117294](https://doi.org/10.1051/0004-6361/201117294). arXiv: [1105.3980](https://arxiv.org/abs/1105.3980) (cit. on pp. [61](#), [62](#)).
- Hildebrandt, H., T. Erben, et al. (2012). “CFHTLenS: improving the quality of photometric redshifts with precision photometry.” In: MNRAS 421, pp. 2355–2367. DOI: [10.1111/j.1365-2966.2012.20468.x](https://doi.org/10.1111/j.1365-2966.2012.20468.x). arXiv: [1111.4434](https://arxiv.org/abs/1111.4434) (cit. on p. [84](#)).
- Hildebrandt, H., M. Viola, et al. (2016). “KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing.” In: *ArXiv e-prints*. arXiv: [1606.05338](https://arxiv.org/abs/1606.05338) (cit. on pp. [35](#), [86](#)).
- Hirata, C. M. and U. Seljak (2003). “Reconstruction of lensing from the cosmic microwave background polarization.” In: Phys. Rev. D 68.8, 083002, p. 083002. DOI: [10.1103/PhysRevD.68.083002](https://doi.org/10.1103/PhysRevD.68.083002). arXiv: [astro-ph/0306354](https://arxiv.org/abs/astro-ph/0306354) (cit. on p. [90](#)).
- Hoekstra, H. and B. Jain (2008). “Weak Gravitational Lensing and Its Cosmological Applications.” In: *Annual Review of Nuclear and Particle Science* 58, pp. 99–123. DOI: [10.1146/annurev.nucl.58.110707.171151](https://doi.org/10.1146/annurev.nucl.58.110707.171151). arXiv: [0805.0139](https://arxiv.org/abs/0805.0139) (cit. on p. [30](#)).
- Hoyle, F. (1948). “A New Model for the Expanding Universe.” In: MNRAS 108, p. 372. DOI: [10.1093/mnras/108.5.372](https://doi.org/10.1093/mnras/108.5.372) (cit. on p. [19](#)).
- Hu, W. (1999). “Power Spectrum Tomography with Weak Lensing.” In: ApJ 522, pp. L21–L24. DOI: [10.1086/312210](https://doi.org/10.1086/312210). arXiv: [astro-ph/9904153](https://arxiv.org/abs/astro-ph/9904153) (cit. on p. [32](#)).
- Hu, W. and S. Dodelson (2002). “Cosmic Microwave Background Anisotropies.” In: ARA&A 40, pp. 171–216. DOI: [10.1146/annurev.astro.40.060401.093926](https://doi.org/10.1146/annurev.astro.40.060401.093926). arXiv: [astro-ph/0110414](https://arxiv.org/abs/astro-ph/0110414) (cit. on p. [36](#)).
- Hu, W. and M. White (1996). “Acoustic Signatures in the Cosmic Microwave Background.” In: ApJ 471, p. 30. DOI: [10.1086/177951](https://doi.org/10.1086/177951). arXiv: [astro-ph/9602019](https://arxiv.org/abs/astro-ph/9602019) (cit. on pp. [36](#), [59](#)).
- (2001). “Power Spectra Estimation for Weak Lensing.” In: ApJ 554, pp. 67–73. DOI: [10.1086/321380](https://doi.org/10.1086/321380). arXiv: [astro-ph/0010352](https://arxiv.org/abs/astro-ph/0010352) (cit. on p. [60](#)).

- Hubble, E. (1929). "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae." In: *Proceedings of the National Academy of Science* 15, pp. 168–173. DOI: [10.1073/pnas.15.3.168](https://doi.org/10.1073/pnas.15.3.168) (cit. on pp. 18, 19).
- (1934). "The Distribution of Extra-Galactic Nebulae." In: *ApJ* 79, p. 8. DOI: [10.1086/143517](https://doi.org/10.1086/143517) (cit. on p. 60).
- Hunter, J. E. (1997). "Needed: A Ban on the Significance Test." In: *Psychological Science* 8.1, pp. 3–7. DOI: [10.1111/j.1467-9280.1997.tb00534.x](https://doi.org/10.1111/j.1467-9280.1997.tb00534.x). eprint: <http://pss.sagepub.com/content/8/1/3.full.pdf+html>. URL: <http://pss.sagepub.com/content/8/1/3.abstract> (cit. on p. 54).
- Ioannidis, J. P. A. (2005). "Why Most Published Research Findings Are False." In: *PLoS Med* 2.8. DOI: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124). URL: <http://dx.doi.org/10.1371/journal.pmed.0020124> (cit. on p. 54).
- Irwin, M. J., R. L. Webster, P. C. Hewett, R. T. Corrigan, and R. I. Jedrzejewski (1989). "Photometric variations in the Q2237 + 0305 system - First detection of a microlensing event." In: *AJ* 98, pp. 1989–1994. DOI: [10.1086/115272](https://doi.org/10.1086/115272) (cit. on p. 30).
- Isserlis, L. (1918). "On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables." In: *Biometrika* 12.1/2, pp. 134–139. ISSN: 00063444. URL: <http://www.jstor.org/stable/2331932> (cit. on p. 56).
- Jaynes, E. T. and G. L. Bretthorst (2003). *Probability Theory*. Cambridge University Press. ISBN: 0521592712 (cit. on pp. 41, 45).
- Jeffreys, H. (1935). "Some Tests of Significance, Treated by the Theory of Probability." In: *Proceedings of the Cambridge Philosophical Society* 31, p. 203. DOI: [10.1017/S030500410001330X](https://doi.org/10.1017/S030500410001330X) (cit. on p. 45).
- (1961). *The Theory of Probability*. Oxford Classic Texts in the Physical Sciences. OUP Oxford. ISBN: 9780191589676 (cit. on pp. 41, 45, 46).
- Jimenez, R., L. Verde, H. Peiris, and A. Kosowsky (2004). "Fast cosmological parameter estimation from microwave background temperature and polarization power spectra." In: *Phys. Rev. D* 70.2, 023005, p. 023005. DOI: [10.1103/PhysRevD.70.023005](https://doi.org/10.1103/PhysRevD.70.023005). arXiv: [astro-ph/0404237](https://arxiv.org/abs/astro-ph/0404237) (cit. on p. 66).
- Joachimi, B., M. Cacciato, et al. (2015). "Galaxy Alignments: An Overview." In: *Space Sci. Rev.* 193, pp. 1–65. DOI: [10.1007/s11214-015-0177-4](https://doi.org/10.1007/s11214-015-0177-4). arXiv: [1504.05456](https://arxiv.org/abs/1504.05456) (cit. on p. 33).
- Joachimi, B., P. Schneider, and T. Eifler (2008). "Analysis of two-point statistics of cosmic shear. III. Covariances of shear measures made easy." In: *A&A* 477, pp. 43–54. DOI: [10.1051/0004-6361:20078400](https://doi.org/10.1051/0004-6361:20078400). arXiv: [0708.0387](https://arxiv.org/abs/0708.0387) (cit. on p. 87).
- Joachimi, B. and A. N. Taylor (2011). "Forecasts of non-Gaussian parameter spaces using Box-Cox transformations." In: *MNRAS* 416, pp. 1010–1022. DOI: [10.1111/j.1365-2966.2011.19107.x](https://doi.org/10.1111/j.1365-2966.2011.19107.x). arXiv: [1103.3370](https://arxiv.org/abs/1103.3370) (cit. on pp. 66, 68, 69, 89, 97).

- Joachimi, B., A. N. Taylor, and A. Kiessling (2011). “Cosmological information in Gaussianized weak lensing signals.” In: *MNRAS* 418, pp. 145–169. DOI: [10.1111/j.1365-2966.2011.19472.x](https://doi.org/10.1111/j.1365-2966.2011.19472.x). arXiv: [1104.1399](https://arxiv.org/abs/1104.1399) (cit. on pp. [61](#), [62](#), [90](#), [94](#), [97](#), [99](#), [101](#)).
- Kaiser, N. (1992). “Weak gravitational lensing of distant galaxies.” In: *ApJ* 388, pp. 272–286. DOI: [10.1086/171151](https://doi.org/10.1086/171151) (cit. on pp. [30](#), [32](#), [87](#)).
- Kaiser, N., G. Wilson, and G. A. Luppino (2000). “Large-Scale Cosmic Shear Measurements.” In: *ArXiv Astrophysics e-prints*. arXiv: [astro-ph/0003338](https://arxiv.org/abs/astro-ph/0003338) (cit. on p. [34](#)).
- Kapteyn, J. C. (1922). “First Attempt at a Theory of the Arrangement and Motion of the Sidereal System.” In: *ApJ* 55, p. 302. DOI: [10.1086/142670](https://doi.org/10.1086/142670) (cit. on p. [21](#)).
- Kass, R. E. and A. E. Raftery (1995). “Bayes Factors.” In: *Journal of the American Statistical Association* 90.430, pp. 773–795. DOI: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572). eprint: <http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1995.10476572>. URL: <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572> (cit. on pp. [45](#), [46](#)).
- Kayo, I., M. Takada, and B. Jain (2013). “Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix.” In: *MNRAS* 429, pp. 344–371. DOI: [10.1093/mnras/sts340](https://doi.org/10.1093/mnras/sts340). arXiv: [1207.6322](https://arxiv.org/abs/1207.6322) [[astro-ph.CO](#)] (cit. on p. [60](#)).
- Kayo, I., A. Taruya, and Y. Suto (2001a). “Probability Distribution Function of Cosmological Density Fluctuations from a Gaussian Initial Condition: Comparison of One-Point and Two-Point Lognormal Model Predictions with N-Body Simulations.” In: *ApJ* 561, pp. 22–34. DOI: [10.1086/323227](https://doi.org/10.1086/323227). arXiv: [astro-ph/0105218](https://arxiv.org/abs/astro-ph/0105218) (cit. on p. [61](#)).
- (2001b). “Probability Distribution Function of Cosmological Density Fluctuations from a Gaussian Initial Condition: Comparison of One-Point and Two-Point Lognormal Model Predictions with N-Body Simulations.” In: *ApJ* 561, pp. 22–34. DOI: [10.1086/323227](https://doi.org/10.1086/323227). arXiv: [astro-ph/0105218](https://arxiv.org/abs/astro-ph/0105218) (cit. on p. [61](#)).
- Kendall, M. G., A. Stuart, J. K. Ord, and S. Arnold (1999). *Kendall's advanced theory of statistics. Vol.2A: Classical inference and the linear model*. ISBN: 0340662301 (cit. on p. [53](#)).
- Khintchin, A. (1934). “Korrelationstheorie der stationären stochastischen Prozesse.” In: *Mathematische Annalen* 109.1, pp. 604–615. ISSN: 1432-1807. DOI: [10.1007/BF01449156](https://doi.org/10.1007/BF01449156). URL: <http://dx.doi.org/10.1007/BF01449156> (cit. on p. [56](#)).
- Kiessling, A. et al. (2015). “Galaxy Alignments: Theory, Modelling & Simulations.” In: *Space Sci. Rev.* 193, pp. 67–136. DOI: [10.1007/s11214-015-0203-6](https://doi.org/10.1007/s11214-015-0203-6). arXiv: [1504.05546](https://arxiv.org/abs/1504.05546) (cit. on p. [33](#)).
- Kilbinger, M. (2015). “Cosmology with cosmic shear observations: a review.” In: *Reports on Progress in Physics* 78.8, 086901, p. 086901. DOI: [10.1088/0034-4885/78/8/086901](https://doi.org/10.1088/0034-4885/78/8/086901). arXiv: [1411.0115](https://arxiv.org/abs/1411.0115) (cit. on p. [30](#)).

- Kilbinger, M., K. Benabed, O. Cappé, J.-F. Cardoso, J. Coupon, G. Fort, H. J. McCracken, S. Prunet, C. P. Robert, and D. Wraith (2011). “CosmoPMC: Cosmology Population Monte Carlo.” In: *ArXiv e-prints*. arXiv: [1101.0950](#) (cit. on p. 52).
- Kilbinger, M., L. Fu, et al. (2013). “CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing.” In: *MNRAS* 430, pp. 2200–2220. DOI: [10.1093/mnras/stt041](#). arXiv: [1212.3338](#) (cit. on pp. 34, 52, 83, 85).
- Kilbinger, M. and P. Schneider (2004). “Analysis of two-point statistics of cosmic shear. II. Optimizing the survey geometry.” In: *A&A* 413, pp. 465–476. DOI: [10.1051/0004-6361:20034172](#). arXiv: [astro-ph/0308119](#) (cit. on p. 87).
- (2005). “Cosmological parameters from combined second- and third-order aperture mass statistics of cosmic shear.” In: *A&A* 442, pp. 69–83. DOI: [10.1051/0004-6361:20053531](#). arXiv: [astro-ph/0505581](#) (cit. on p. 34).
- Kilbinger, M., D. Wraith, C. P. Robert, K. Benabed, O. Cappé, J.-F. Cardoso, G. Fort, S. Prunet, and F. R. Bouchet (2010). “Bayesian model comparison in cosmology with Population Monte Carlo.” In: *MNRAS* 405, pp. 2381–2390. DOI: [10.1111/j.1365-2966.2010.16605.x](#). arXiv: [0912.1614](#) (cit. on pp. 51, 52, 81, 84, 85).
- Kirk, D. et al. (2015). “Galaxy Alignments: Observations and Impact on Cosmology.” In: *Space Sci. Rev.* 193, pp. 139–211. DOI: [10.1007/s11214-015-0213-4](#). arXiv: [1504.05465](#) (cit. on p. 33).
- Kogut, A. J. et al. (2014). “The Primordial Inflation Explorer (PIXIE).” In: *American Astronomical Society Meeting Abstracts #223*. Vol. 223. American Astronomical Society Meeting Abstracts, p. 439.01 (cit. on p. 27).
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer (cit. on p. 41).
- Komatsu, E. et al. (2011). “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation.” In: *ApJS* 192, 18, p. 18. DOI: [10.1088/0067-0049/192/2/18](#). arXiv: [1001.4538 \[astro-ph.CO\]](#) (cit. on p. 45).
- Kosowsky, A., M. Milosavljevic, and R. Jimenez (2002). “Efficient cosmological parameter estimation from microwave background anisotropies.” In: *Phys. Rev. D* 66.6, 063007, p. 063007. DOI: [10.1103/PhysRevD.66.063007](#). arXiv: [astro-ph/0206014](#) (cit. on p. 66).
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago Press. ISBN: 9780226458113 (cit. on p. 40).
- Lahav, O., P. B. Lilje, J. R. Primack, and M. J. Rees (1991). “Dynamical effects of the cosmological constant.” In: *MNRAS* 251, pp. 128–136. DOI: [10.1093/mnras/251.1.128](#) (cit. on p. 29).
- Lambdin, C. (2012). “Significance tests as sorcery: Science is empirical – significance tests are not.” In: *Theory & Psychology* 22.1, pp. 67–90. DOI: [10.1177/0959354311429854](#).

- eprint: <http://tap.sagepub.com/content/22/1/67.full.pdf+html>. URL: <http://tap.sagepub.com/content/22/1/67.abstract> (cit. on p. 54).
- Larson, D. et al. (2011). “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters.” In: *ApJS* 192, 16, p. 16. DOI: [10.1088/0067-0049/192/2/16](https://doi.org/10.1088/0067-0049/192/2/16). arXiv: [1001.4635](https://arxiv.org/abs/1001.4635) [[astro-ph.CO](https://arxiv.org/archive/astro-ph)] (cit. on p. 22).
- Laureijs, R. et al. (2011). “Euclid Definition Study Report.” In: *ArXiv e-prints*. arXiv: [1110.3193](https://arxiv.org/abs/1110.3193) [[astro-ph.CO](https://arxiv.org/archive/astro-ph)] (cit. on pp. 35, 39).
- Lehmann, E. L. (1993). “The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?” In: *Journal of the American Statistical Association* 88.424, pp. 1242–1249. DOI: [10.1080/01621459.1993.10476404](https://doi.org/10.1080/01621459.1993.10476404). eprint: <http://www.tandfonline.com/doi/pdf/10.1080/01621459.1993.10476404>. URL: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476404> (cit. on p. 53).
- Lemaître, G. (1931). “Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae.” In: *MNRAS* 91, pp. 483–490. DOI: [10.1093/mnras/91.5.483](https://doi.org/10.1093/mnras/91.5.483) (cit. on p. 17).
- Lenhard, J. (2006). “Models and Statistical Inference: The Controversy between Fisher and Neyman-Pearson.” In: *The British Journal for the Philosophy of Science* 57.1, pp. 69–91. DOI: [10.1093/bjps/axi152](https://doi.org/10.1093/bjps/axi152). eprint: <http://bjps.oxfordjournals.org/content/57/1/69.full.pdf+html>. URL: <http://bjps.oxfordjournals.org/content/57/1/69.abstract> (cit. on p. 53).
- Levi, M. et al. (2013). “The DESI Experiment, a whitepaper for Snowmass 2013.” In: *ArXiv e-prints*. arXiv: [1308.0847](https://arxiv.org/abs/1308.0847) [[astro-ph.CO](https://arxiv.org/archive/astro-ph)] (cit. on p. 39).
- Lévy, P. D. (1925). *Calcul des probabilités*. PCMI collection. Gauthier-Villars. URL: https://books.google.co.uk/books?id=8%5C_FLAAAAMAAJ (cit. on p. 52).
- (1937). *Théorie de l'Addition des Variables Aléatoires*. Monographies Des Probabilités. Paris: Gauthier-Villars. ISBN: 978-2-87647-207-5 (cit. on p. 62).
- Lewis, A. and S. Bridle (2002). “Cosmological parameters from CMB and other data: A Monte Carlo approach.” In: *Phys. Rev. D* 66.10, 103511, p. 103511. DOI: [10.1103/PhysRevD.66.103511](https://doi.org/10.1103/PhysRevD.66.103511). arXiv: [astro-ph/0205436](https://arxiv.org/abs/astro-ph/0205436) (cit. on pp. 49, 50, 77).
- Lewis, A. and A. Challinor (2006). “Weak gravitational lensing of the CMB.” In: *Phys. Rep.* 429, pp. 1–65. DOI: [10.1016/j.physrep.2006.03.002](https://doi.org/10.1016/j.physrep.2006.03.002). eprint: [astro-ph/0601594](https://arxiv.org/abs/astro-ph/0601594) (cit. on p. 27).
- Liddle, A. R. (2007). “Information criteria for astrophysical model selection.” In: *MNRAS* 377, pp. L74–L78. DOI: [10.1111/j.1745-3933.2007.00306.x](https://doi.org/10.1111/j.1745-3933.2007.00306.x). arXiv: [astro-ph/0701113](https://arxiv.org/abs/astro-ph/0701113) (cit. on p. 46).

- Lightman, A. P. and P. L. Schechter (1990). “The Omega dependence of peculiar velocities induced by spherical density perturbations.” In: *ApJS* 74, p. 831. DOI: [10.1086/191521](https://doi.org/10.1086/191521) (cit. on p. 29).
- Limber, D. N. (1954). “The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II.” In: *ApJ* 119, p. 655. DOI: [10.1086/145870](https://doi.org/10.1086/145870) (cit. on p. 32).
- Lin, C.-A. and M. Kilbinger (2015a). “A new model to predict weak-lensing peak counts. I. Comparison with N-body simulations.” In: *A&A* 576, A24, A24. DOI: [10.1051/0004-6361/201425188](https://doi.org/10.1051/0004-6361/201425188). arXiv: [1410.6955](https://arxiv.org/abs/1410.6955) (cit. on pp. 63, 111).
- (2015b). “A new model to predict weak-lensing peak counts. II. Parameter constraint strategies.” In: *A&A* 583, A70, A70. DOI: [10.1051/0004-6361/201526659](https://doi.org/10.1051/0004-6361/201526659). arXiv: [1506.01076](https://arxiv.org/abs/1506.01076) (cit. on pp. 63, 111).
- Lin, C.-A., M. Kilbinger, and S. Pires (2016). “A new model to predict weak-lensing peak counts. III. Filtering technique comparisons.” In: *A&A* 593, A88, A88. DOI: [10.1051/0004-6361/201628565](https://doi.org/10.1051/0004-6361/201628565). arXiv: [1603.06773](https://arxiv.org/abs/1603.06773) (cit. on pp. 63, 111).
- Lindeberg, J. W. (1922). “Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung.” In: *Mathematische Zeitschrift* 15.1, pp. 211–225. ISSN: 1432-1823. DOI: [10.1007/BF01494395](https://doi.org/10.1007/BF01494395). URL: <http://dx.doi.org/10.1007/BF01494395> (cit. on p. 52).
- LSST Dark Energy Science Collaboration (2012). “Large Synoptic Survey Telescope: Dark Energy Science Collaboration.” In: *ArXiv e-prints*. arXiv: [1211.0310](https://arxiv.org/abs/1211.0310) [[astro-ph.CO](https://arxiv.org/abs/1211.0310)] (cit. on p. 35).
- Lyapunov, A. M. (1901). “Nouvelle forme du théorème sur la limite des probabilités.” In: *Mém. Acad. Imp. Sci. St. Pétersbourg* 12, pp. 1–24 (cit. on p. 52).
- Ma, C.-P. and E. Bertschinger (1995). “Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges.” In: *ApJ* 455, p. 7. DOI: [10.1086/176550](https://doi.org/10.1086/176550). arXiv: [astro-ph/9506072](https://arxiv.org/abs/astro-ph/9506072) (cit. on p. 24).
- MacCrann, N., J. Zuntz, S. Bridle, B. Jain, and M. R. Becker (2015). “Cosmic discordance: are Planck CMB and CFHTLenS weak lensing measurements out of tune?” In: *MNRAS* 451, pp. 2877–2888. DOI: [10.1093/mnras/stv1154](https://doi.org/10.1093/mnras/stv1154). arXiv: [1408.4742](https://arxiv.org/abs/1408.4742) (cit. on p. 86).
- Maddox, S. J., G. Efstathiou, W. J. Sutherland, and J. Loveday (1990). “Galaxy correlations on large scales.” In: *MNRAS* 242, 43P–47P. DOI: [10.1093/mnras/242.1.43P](https://doi.org/10.1093/mnras/242.1.43P) (cit. on p. 22).
- Markevitch, M., A. H. Gonzalez, D. Clowe, A. Vikhlinin, W. Forman, C. Jones, S. Murray, and W. Tucker (2004). “Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56.” In: *ApJ* 606, pp. 819–824. DOI: [10.1086/383178](https://doi.org/10.1086/383178). arXiv: [astro-ph/0309303](https://arxiv.org/abs/astro-ph/0309303) (cit. on p. 22).

- Martel, H., P. Premadi, and R. Matzner (2002). "Light Propagation in Inhomogeneous Universes. III. Distributions of Image Separations." In: *ApJ* 570, pp. 17–32. DOI: [10.1086/339569](https://doi.org/10.1086/339569). arXiv: [astro-ph/0201198](https://arxiv.org/abs/astro-ph/0201198) (cit. on p. 90).
- Martínez, V. J. and E. Saar (2002). *Statistics of the Galaxy Distribution*. Chapman & Hall. ISBN: 1584880848 (cit. on p. 57).
- Maruyama, G. (1949). "The Harmonic Analysis of Stationary Stochastic Processes." In: *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics* 4.1, pp. 45–106. DOI: [10.2206/kyushumfs.4.45](https://doi.org/10.2206/kyushumfs.4.45) (cit. on p. 57).
- Matsumura, T. et al. (2014). "Mission Design of LiteBIRD." In: *Journal of Low Temperature Physics* 176, pp. 733–740. DOI: [10.1007/s10909-013-0996-1](https://doi.org/10.1007/s10909-013-0996-1). arXiv: [1311.2847](https://arxiv.org/abs/1311.2847) [[astro-ph](https://arxiv.org/abs/astro-ph).IM] (cit. on p. 27).
- Mehta, K. T., H.-J. Seo, J. Eckel, D. J. Eisenstein, M. Metchnik, P. Pinto, and X. Xu (2011). "Galaxy Bias and Its Effects on the Baryon Acoustic Oscillation Measurements." In: *ApJ* 734, 94, p. 94. DOI: [10.1088/0004-637X/734/2/94](https://doi.org/10.1088/0004-637X/734/2/94). arXiv: [1104.1178](https://arxiv.org/abs/1104.1178) [[astro-ph](https://arxiv.org/abs/astro-ph).CO] (cit. on p. 38).
- Meiksin, A. and M. White (1999). "The growth of correlations in the matter power spectrum." In: *MNRAS* 308, pp. 1179–1184. DOI: [10.1046/j.1365-8711.1999.02825.x](https://doi.org/10.1046/j.1365-8711.1999.02825.x). arXiv: [astro-ph/9812129](https://arxiv.org/abs/astro-ph/9812129) (cit. on pp. 60, 92).
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). "Equation of State Calculations by Fast Computing Machines." In: *J. Chem. Phys.* 21, pp. 1087–1092. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114) (cit. on p. 49).
- Miller, L. et al. (2013). "Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey." In: *MNRAS* 429, pp. 2858–2880. DOI: [10.1093/mnras/sts454](https://doi.org/10.1093/mnras/sts454). arXiv: [1210.8201](https://arxiv.org/abs/1210.8201) [[astro-ph](https://arxiv.org/abs/astro-ph).CO] (cit. on p. 84).
- Minchin, R. et al. (2005). "A Dark Hydrogen Cloud in the Virgo Cluster." In: *ApJ* 622, pp. L21–L24. DOI: [10.1086/429538](https://doi.org/10.1086/429538). arXiv: [astro-ph/0502312](https://arxiv.org/abs/astro-ph/0502312) (cit. on p. 21).
- Miralda-Escude, J. (1991). "The correlation function of galaxy ellipticities produced by gravitational lensing." In: *ApJ* 380, pp. 1–8. DOI: [10.1086/170555](https://doi.org/10.1086/170555) (cit. on p. 30).
- Misner, C. W., K. S. Thorne, and J. A. Wheeler (1973). *Gravitation*. W.H. Freeman and Co. ISBN: 0-7167-0334-3 (cit. on pp. 16, 30).
- Mukhanov, V. (2005). *Physical Foundations of Cosmology*. Cambridge University Press. ISBN: 9780521563987. DOI: [10.2277/0521563984](https://doi.org/10.2277/0521563984) (cit. on pp. 17, 19).
- Munshi, D., P. Valageas, L. van Waerbeke, and A. Heavens (2008). "Cosmology with weak lensing surveys." In: *Phys. Rep.* 462, pp. 67–121. DOI: [10.1016/j.physrep.2008.02.003](https://doi.org/10.1016/j.physrep.2008.02.003). arXiv: [astro-ph/0612667](https://arxiv.org/abs/astro-ph/0612667) (cit. on p. 30).
- Murdoch, D. J., Y.-L. Tsai, and J. Adcock (2008). "P-Values are Random Variables." In: *The American Statistician* 62.3, pp. 242–245. DOI: [10.1198/000313008X332421](https://doi.org/10.1198/000313008X332421). eprint:

- <http://dx.doi.org/10.1198/000313008X332421>. URL: <http://dx.doi.org/10.1198/000313008X332421> (cit. on p. 54).
- Nelder, J. A. and R. Mead (1965). "A Simplex Method for Function Minimization." In: *The Computer Journal* 7.4, pp. 308–313. DOI: [10.1093/comjnl/7.4.308](https://doi.org/10.1093/comjnl/7.4.308). eprint: <http://comjnl.oxfordjournals.org/content/7/4/308.full.pdf+html>. URL: <http://comjnl.oxfordjournals.org/content/7/4/308.abstract> (cit. on pp. 68, 101).
- Nelsen, R. B. (1999). *An Introduction to Copulas*. Lecture notes in statistics. Springer. ISBN: 978-0-387-98623-4 (cit. on pp. 63, 88).
- Neyman, J. and E. S. Pearson (1933). "On the Problem of the Most Efficient Tests of Statistical Hypotheses." In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 231.694-706, pp. 289–337. ISSN: 0264-3952. DOI: [10.1098/rsta.1933.0009](https://doi.org/10.1098/rsta.1933.0009). eprint: <http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf>. URL: <http://rsta.royalsocietypublishing.org/content/231/694-706/289> (cit. on p. 53).
- Neyrinck, M. C. (2011). "Rejuvenating the Matter Power Spectrum. III. The Cosmology Sensitivity of Gaussianized Power Spectra." In: *ApJ* 742, 91, p. 91. DOI: [10.1088/0004-637X/742/2/91](https://doi.org/10.1088/0004-637X/742/2/91). arXiv: [1105.2955](https://arxiv.org/abs/1105.2955) [[astro-ph.CO](https://arxiv.org/abs/1105.2955)] (cit. on pp. 61, 62).
- Neyrinck, M. C. and I. Szapudi (2007). "Information content in the halo-model dark-matter power spectrum - II. Multiple cosmological parameters." In: *MNRAS* 375, pp. L51–L55. DOI: [10.1111/j.1745-3933.2006.00275.x](https://doi.org/10.1111/j.1745-3933.2006.00275.x). arXiv: [astro-ph/0610211](https://arxiv.org/abs/astro-ph/0610211) (cit. on pp. 60, 87).
- Neyrinck, M. C., I. Szapudi, and C. D. Rimes (2006). "Information content in the halo-model dark-matter power spectrum." In: *MNRAS* 370, pp. L66–L70. DOI: [10.1111/j.1745-3933.2006.00190.x](https://doi.org/10.1111/j.1745-3933.2006.00190.x). arXiv: [astro-ph/0604282](https://arxiv.org/abs/astro-ph/0604282) (cit. on pp. 60, 87).
- Neyrinck, M. C., I. Szapudi, and A. S. Szalay (2009). "Rejuvenating the Matter Power Spectrum: Restoring Information with a Logarithmic Density Mapping." In: *ApJ* 698, pp. L90–L93. DOI: [10.1088/0004-637X/698/2/L90](https://doi.org/10.1088/0004-637X/698/2/L90). arXiv: [0903.4693](https://arxiv.org/abs/0903.4693) [[astro-ph.CO](https://arxiv.org/abs/0903.4693)] (cit. on pp. 61, 62).
- (2011). "Rejuvenating Power Spectra. II. The Gaussianized Galaxy Density Field." In: *ApJ* 731, 116, p. 116. DOI: [10.1088/0004-637X/731/2/116](https://doi.org/10.1088/0004-637X/731/2/116). arXiv: [1009.5680](https://arxiv.org/abs/1009.5680) [[astro-ph.CO](https://arxiv.org/abs/1009.5680)] (cit. on pp. 61, 62).
- Nichol, R. C. (2008). "Cosmology with galaxy correlations." In: *General Relativity and Gravitation* 40, pp. 249–267. DOI: [10.1007/s10714-007-0546-8](https://doi.org/10.1007/s10714-007-0546-8). arXiv: [0708.2824](https://arxiv.org/abs/0708.2824) (cit. on p. 29).
- Niemack, M. D. et al. (2010). "ACTPol: a polarization-sensitive receiver for the Atacama Cosmology Telescope." In: *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V*. Vol. 7741. Proc. SPIE, 77411S. DOI: [10.1117/12.857464](https://doi.org/10.1117/12.857464). arXiv: [1006.5049](https://arxiv.org/abs/1006.5049) [[astro-ph.IM](https://arxiv.org/abs/1006.5049)] (cit. on p. 27).

- Nieto-Reyes, F. G. A., J. A. Cuesta-Albertos, and F. Gamboa (2009). “A random-projection based procedure to test if a stationary process is Gaussian.” In: *ArXiv e-prints*. arXiv: [0911.3520 \[stat.ME\]](https://arxiv.org/abs/0911.3520) (cit. on pp. [57](#), [94](#), [97](#)).
- (2014). “A random-projection based test of Gaussianity for stationary processes.” In: *Computational Statistics & Data Analysis* 75, pp. 124–141. ISSN: 0167-9473. DOI: [10.1016/j.csda.2014.01.013](https://doi.org/10.1016/j.csda.2014.01.013). URL: <http://www.sciencedirect.com/science/article/pii/S0167947314000243> (cit. on pp. [57](#), [58](#), [94–96](#)).
- Oloff, R. (1999). *Geometrie der Raumzeit. Eine mathematische Einführung in die Relativitätstheorie*. Vieweg, Braunschweig. ISBN: 3-528-06917-1 (cit. on pp. [16](#), [30](#)).
- Oort, J. H. (1932). “The Force Exerted by the Stellar System in the Direction Perpendicular to the Galactic Plane and Some Related Problems.” In: *Bulletin of the Astronomical Institutes of the Netherlands* 6, pp. 249–287 (cit. on p. [21](#)).
- Padilla, A. (2015). “Lectures on the Cosmological Constant Problem.” In: *ArXiv e-prints*. arXiv: [1502.05296 \[hep-th\]](https://arxiv.org/abs/1502.05296) (cit. on p. [22](#)).
- Padmanabhan, N. and M. White (2009). “Calibrating the baryon oscillation ruler for matter and halos.” In: *Phys. Rev. D* 80.6, 063508, p. 063508. DOI: [10.1103/PhysRevD.80.063508](https://doi.org/10.1103/PhysRevD.80.063508). arXiv: [0906.1198 \[astro-ph.CO\]](https://arxiv.org/abs/0906.1198) (cit. on p. [38](#)).
- Parzen, E. (1962). “On Estimation of a Probability Density Function and Mode.” In: *Ann. Math. Statist.* 33.3, pp. 1065–1076. DOI: [10.1214/aoms/1177704472](https://doi.org/10.1214/aoms/1177704472). URL: <http://dx.doi.org/10.1214/aoms/1177704472> (cit. on p. [47](#)).
- Pearson, E. S., R. B. D’Agostino, and K. O. Bowman (1977). “Tests for Departure from Normality: Comparison of Powers.” In: *Biometrika* 64.2, pp. 231–246. ISSN: 00063444. URL: <http://www.jstor.org/stable/2335689> (cit. on p. [54](#)).
- Peebles, P. J. E. (1980). *The large-scale structure of the universe*. Princeton University Press. ISBN: 0-691-08239-1 (cit. on pp. [24](#), [32](#), [57](#), [59](#), [60](#)).
- Peebles, P. J. E. and J. T. Yu (1970). “Primeval Adiabatic Perturbation in an Expanding Universe.” In: *ApJ* 162, p. 815. DOI: [10.1086/150713](https://doi.org/10.1086/150713) (cit. on p. [36](#)).
- Penzias, A. A. and R. W. Wilson (1965). “A Measurement of Excess Antenna Temperature at 4080 Mc/s.” In: *ApJ* 142, pp. 419–421. DOI: [10.1086/148307](https://doi.org/10.1086/148307) (cit. on p. [24](#)).
- Percival, W. J., S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay (2007). “Measuring the Baryon Acoustic Oscillation scale using the Sloan Digital Sky Survey and 2dF Galaxy Redshift Survey.” In: *MNRAS* 381, pp. 1053–1066. DOI: [10.1111/j.1365-2966.2007.12268.x](https://doi.org/10.1111/j.1365-2966.2007.12268.x). arXiv: [0705.3323](https://arxiv.org/abs/0705.3323) (cit. on pp. [38](#), [84](#)).
- Percival, W. J., B. A. Reid, et al. (2010). “Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample.” In: *MNRAS* 401, pp. 2148–2168. DOI: [10.1111/j.1365-2966.2009.15812.x](https://doi.org/10.1111/j.1365-2966.2009.15812.x). arXiv: [0907.1660 \[astro-ph.CO\]](https://arxiv.org/abs/0907.1660) (cit. on p. [38](#)).

- Perlmutter, S. et al. (1999). “Measurements of Ω and Λ from 42 High-Redshift Supernovae.” In: *ApJ* 517, pp. 565–586. DOI: [10.1086/307221](https://doi.org/10.1086/307221). arXiv: [astro-ph/9812133](https://arxiv.org/abs/astro-ph/9812133) (cit. on pp. 18, 23).
- Pires, S., A. Leonard, and J.-L. Starck (2012). “Cosmological constraints from the capture of non-Gaussianity in weak lensing data.” In: *MNRAS* 423, pp. 983–992. DOI: [10.1111/j.1365-2966.2012.20940.x](https://doi.org/10.1111/j.1365-2966.2012.20940.x). arXiv: [1203.2877](https://arxiv.org/abs/1203.2877) (cit. on pp. 63, 111).
- Pitman, J. (2006). *Combinatorial stochastic processes*. Vol. 1875. Lecture Notes in Mathematics. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. Berlin: Springer-Verlag, pp. x+256. ISBN: 978-3-540-30990-1; 3-540-30990-X. DOI: [10.1007/b11601500](https://doi.org/10.1007/b11601500). URL: <http://bibserver.berkeley.edu/csp/april05/bookcsp.pdf> (cit. on p. 95).
- Plancherel, M. and G. Mittag-Leffler (1910). “Contribution à l’étude de la représentation d’une fonction arbitraire par des intégrales définies.” In: *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 30.1, pp. 289–335. ISSN: 0009-725X. DOI: [10.1007/BF03014877](https://doi.org/10.1007/BF03014877). URL: <http://dx.doi.org/10.1007/BF03014877> (cit. on p. 98).
- Planck Collaboration (2015a). “Planck 2015 results. I. Overview of products and scientific results.” In: *ArXiv e-prints*. arXiv: [1502.01582](https://arxiv.org/abs/1502.01582) (cit. on pp. 24, 50, 86).
- (2015b). “Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters.” In: *ArXiv e-prints*. arXiv: [1507.02704](https://arxiv.org/abs/1507.02704) (cit. on pp. 26, 45).
- (2015c). “Planck 2015 results. XIII. Cosmological parameters.” In: *ArXiv e-prints*. arXiv: [1502.01589](https://arxiv.org/abs/1502.01589) (cit. on pp. 21, 45, 50, 65, 67, 84).
- (2015d). “Planck 2015 results. XVII. Constraints on primordial non-Gaussianity.” In: *ArXiv e-prints*. arXiv: [1502.01592](https://arxiv.org/abs/1502.01592) (cit. on pp. 27, 59).
- Planck Collaboration et al. (2014). “Planck 2013 results. XVI. Cosmological parameters.” In: *A&A* 571, A16, A16. DOI: [10.1051/0004-6361/201321591](https://doi.org/10.1051/0004-6361/201321591). arXiv: [1303.5076](https://arxiv.org/abs/1303.5076) (cit. on pp. 65, 67, 75, 77).
- Powell, M. J. D. (2008). “Developments of NEWUOA for minimization without derivatives.” In: *IMA Journal of Numerical Analysis* 28.4, pp. 649–664. DOI: [10.1093/imanum/drm047](https://doi.org/10.1093/imanum/drm047). eprint: <http://imajna.oxfordjournals.org/content/28/4/649.full.pdf+html>. URL: <http://imajna.oxfordjournals.org/content/28/4/649.abstract> (cit. on pp. 69, 106, 109).
- Razali, N. M. and Y. B. Wah (2011). “Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests.” In: *Journal of Statistical Modelling and Analytics* 2.1, pp. 21–33 (cit. on p. 54).
- Rees, M. J. and D. W. Sciama (1968). “Large-scale Density Inhomogeneities in the Universe.” In: *Nature* 217, pp. 511–516. DOI: [10.1038/217511a0](https://doi.org/10.1038/217511a0) (cit. on p. 27).
- Riess, A. G. et al. (1998). “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant.” In: *AJ* 116, pp. 1009–1038. DOI: [10.1086/300499](https://doi.org/10.1086/300499). arXiv: [astro-ph/9805201](https://arxiv.org/abs/astro-ph/9805201) (cit. on pp. 18, 23).

- Rimes, C. D. and A. J. S. Hamilton (2005). “Information content of the non-linear matter power spectrum.” In: MNRAS 360, pp. L82–L86. DOI: [10.1111/j.1745-3933.2005.00051.x](https://doi.org/10.1111/j.1745-3933.2005.00051.x). arXiv: [astro-ph/0502081](https://arxiv.org/abs/astro-ph/0502081) (cit. on pp. 60, 87).
- (2006). “Information content of the non-linear power spectrum: the effect of beat-coupling to large scales.” In: MNRAS 371, pp. 1205–1215. DOI: [10.1111/j.1365-2966.2006.10710.x](https://doi.org/10.1111/j.1365-2966.2006.10710.x). arXiv: [astro-ph/0511418](https://arxiv.org/abs/astro-ph/0511418) (cit. on pp. 60, 87).
- Robertson, H. P. (1935). “Kinematics and World-Structure.” In: ApJ 82, p. 284. DOI: [10.1086/143681](https://doi.org/10.1086/143681) (cit. on p. 17).
- (1936a). “Kinematics and World-Structure II.” In: ApJ 83, p. 187. DOI: [10.1086/143716](https://doi.org/10.1086/143716) (cit. on p. 17).
- (1936b). “Kinematics and World-Structure III.” In: ApJ 83, p. 257. DOI: [10.1086/143726](https://doi.org/10.1086/143726) (cit. on p. 17).
- Rosenblatt, M. (1952). “Remarks on a Multivariate Transformation.” In: *Ann. Math. Statist.* 23.3, pp. 470–472. DOI: [10.1214/aoms/1177729394](https://doi.org/10.1214/aoms/1177729394). URL: <http://dx.doi.org/10.1214/aoms/1177729394> (cit. on p. 62).
- (1956). “Remarks on Some Nonparametric Estimates of a Density Function.” In: *Ann. Math. Statist.* 27.3, pp. 832–837. DOI: [10.1214/aoms/1177728190](https://doi.org/10.1214/aoms/1177728190). URL: <http://dx.doi.org/10.1214/aoms/1177728190> (cit. on p. 47).
- Royston, P. (1992). “Approximating the Shapiro-Wilk W-test for non-normality.” In: *Statistics and Computing* 2.3, pp. 117–119. ISSN: 1573-1375. DOI: [10.1007/BF01891203](https://doi.org/10.1007/BF01891203). URL: <http://dx.doi.org/10.1007/BF01891203> (cit. on p. 54).
- (1993). “A Toolkit for Testing for Non-Normality in Complete and Censored Samples.” In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 42.1, pp. 37–43. ISSN: 00390526, 14679884. URL: <http://www.jstor.org/stable/2348109> (cit. on p. 54).
- Rubin, V. C., W. K. J. Ford, and N. Thonnard (1980). “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4kpc) to UGC 2885 (R = 122 kpc).” In: ApJ 238, pp. 471–487. DOI: [10.1086/158003](https://doi.org/10.1086/158003) (cit. on p. 22).
- Rubiño-Martín, J. A. and CORe+ Collaboration (2015). “The CORe+ (Cosmic Origins Explorer) mission.” In: *Highlights of Spanish Astrophysics VIII*. Ed. by A. J. Cenarro, F. Figueras, C. Hernández-Monteagudo, J. Trujillo Bueno, and L. Valdivielso, pp. 329–334 (cit. on p. 27).
- Rugh, S. E. and H. Zinkernagel (2001). *The Quantum Vacuum and the Cosmological Constant Problem*. URL: <http://philsci-archive.pitt.edu/398/> (cit. on p. 22).
- Sachs, R. K. and A. M. Wolfe (1967). “Perturbations of a Cosmological Model and Angular Variations of the Microwave Background.” In: ApJ 147, p. 73. DOI: [10.1086/148982](https://doi.org/10.1086/148982) (cit. on p. 26).

- Sandage, A. R. (1970). “Cosmology: a search for two numbers.” In: *Physics Today* 23, pp. 34–41 (cit. on p. 13).
- Sandvik, H. B., M. Tegmark, X. Wang, and M. Zaldarriaga (2004). “CMBFIT: Rapid WMAP likelihood calculations with normal parameters.” In: *Phys. Rev. D* 69.6, 063005, p. 063005. DOI: [10.1103/PhysRevD.69.063005](https://doi.org/10.1103/PhysRevD.69.063005). arXiv: [astro-ph/0311544](https://arxiv.org/abs/astro-ph/0311544) (cit. on p. 66).
- Sato, M., T. Hamana, R. Takahashi, M. Takada, N. Yoshida, T. Matsubara, and N. Sugiyama (2009). “Simulations of Wide-Field Weak Lensing Surveys. I. Basic Statistics and Non-Gaussian Effects.” In: *ApJ* 701, pp. 945–954. DOI: [10.1088/0004-637X/701/2/945](https://doi.org/10.1088/0004-637X/701/2/945). arXiv: [0906.2237](https://arxiv.org/abs/0906.2237) [[astro-ph](https://arxiv.org/abs/astro-ph).C0] (cit. on pp. 60, 88).
- Sato, M., K. Ichiki, and T. T. Takeuchi (2010). “Precise Estimation of Cosmological Parameters Using a More Accurate Likelihood Function.” In: *Physical Review Letters* 105.25, 251301, p. 251301. DOI: [10.1103/PhysRevLett.105.251301](https://doi.org/10.1103/PhysRevLett.105.251301). arXiv: [1011.4996](https://arxiv.org/abs/1011.4996) [[astro-ph](https://arxiv.org/abs/astro-ph).C0] (cit. on p. 63).
- (2011). “Copula cosmology: Constructing a likelihood function.” In: *Phys. Rev. D* 83.2, 023501, p. 023501. DOI: [10.1103/PhysRevD.83.023501](https://doi.org/10.1103/PhysRevD.83.023501). arXiv: [1011.4997](https://arxiv.org/abs/1011.4997) [[astro-ph](https://arxiv.org/abs/astro-ph).C0] (cit. on p. 63).
- Sato, M. and T. Nishimichi (2013). “Impact of the non-Gaussian covariance of the weak lensing power spectrum and bispectrum on cosmological parameter estimation.” In: *Phys. Rev. D* 87.12, 123538, p. 123538. DOI: [10.1103/PhysRevD.87.123538](https://doi.org/10.1103/PhysRevD.87.123538). arXiv: [1301.3588](https://arxiv.org/abs/1301.3588) (cit. on pp. 59, 60).
- Sato, M., M. Takada, T. Hamana, and T. Matsubara (2011). “Simulations of Wide-field Weak-lensing Surveys. II. Covariance Matrix of Real-space Correlation Functions.” In: *ApJ* 734, 76, p. 76. DOI: [10.1088/0004-637X/734/2/76](https://doi.org/10.1088/0004-637X/734/2/76). arXiv: [1009.2558](https://arxiv.org/abs/1009.2558) (cit. on p. 88).
- Scherrer, R. J., A. A. Berlind, Q. Mao, and C. K. McBride (2010). “From Finance to Cosmology: The Copula of Large-Scale Structure.” In: *ApJ* 708, pp. L9–L13. DOI: [10.1088/2041-8205/708/1/L9](https://doi.org/10.1088/2041-8205/708/1/L9). arXiv: [0909.5187](https://arxiv.org/abs/0909.5187) (cit. on pp. 63, 88, 93).
- Schervish, M. J. (1996). “P Values: What They Are and What They Are Not.” In: *The American Statistician* 50.3, pp. 203–206. ISSN: 00031305. URL: <http://www.jstor.org/stable/2684655> (cit. on p. 54).
- Schneider, P., J. Ehlers, and E. E. Falco (1992). *Gravitational Lenses*. ISBN: 3-540-97070-3. DOI: [10.1007/978-3-662-03758-4](https://doi.org/10.1007/978-3-662-03758-4) (cit. on p. 30).
- Schneider, P., L. van Waerbeke, M. Kilbinger, and Y. Mellier (2002). “Analysis of two-point statistics of cosmic shear. I. Estimators and covariances.” In: *A&A* 396, pp. 1–19. DOI: [10.1051/0004-6361:20021341](https://doi.org/10.1051/0004-6361:20021341). arXiv: [astro-ph/0206182](https://arxiv.org/abs/astro-ph/0206182) (cit. on p. 87).
- Schrabback, T. et al. (2010). “Evidence of the accelerated expansion of the Universe from weak lensing tomography with COSMOS.” In: *A&A* 516, A63, A63. DOI: [10.1051/0004-6361/200913577](https://doi.org/10.1051/0004-6361/200913577). arXiv: [0911.0053](https://arxiv.org/abs/0911.0053) (cit. on p. 34).

- Schuhmann, R. L., B. Joachimi, and H. V. Peiris (2016). “Gaussianization for fast and accurate inference from cosmological data.” In: MNRAS 459, pp. 1916–1928. DOI: [10.1093/mnras/stw738](https://doi.org/10.1093/mnras/stw738). arXiv: [1510.00019](https://arxiv.org/abs/1510.00019) (cit. on pp. 14, 89, 97).
- Scoccimarro, R. and H. M. P. Couchman (2001). “A fitting formula for the non-linear evolution of the bispectrum.” In: MNRAS 325, pp. 1312–1316. DOI: [10.1046/j.1365-8711.2001.04281.x](https://doi.org/10.1046/j.1365-8711.2001.04281.x). arXiv: [astro-ph/0009427](https://arxiv.org/abs/astro-ph/0009427) (cit. on p. 60).
- Scoccimarro, R., M. Zaldarriaga, and L. Hui (1999). “Power Spectrum Correlations Induced by Nonlinear Clustering.” In: ApJ 527, pp. 1–15. DOI: [10.1086/308059](https://doi.org/10.1086/308059). arXiv: [astro-ph/9901099](https://arxiv.org/abs/astro-ph/9901099) (cit. on pp. 60, 92).
- Seitz, C. and P. Schneider (1997). “Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources.” In: A&A 318, pp. 687–699. arXiv: [astro-ph/9601079](https://arxiv.org/abs/astro-ph/9601079) (cit. on p. 33).
- Semboloni, E., L. van Waerbeke, C. Heymans, T. Hamana, S. Colombi, M. White, and Y. Mellier (2007). “Cosmic variance of weak lensing surveys in the non-Gaussian regime.” In: MNRAS 375, pp. L6–L10. DOI: [10.1111/j.1745-3933.2006.00266.x](https://doi.org/10.1111/j.1745-3933.2006.00266.x). arXiv: [astro-ph/0606648](https://arxiv.org/abs/astro-ph/0606648) (cit. on pp. 34, 59).
- Seo, H.-J., M. Sato, S. Dodelson, B. Jain, and M. Takada (2011). “Re-capturing Cosmic Information.” In: ApJ 729, L11, p. L11. DOI: [10.1088/2041-8205/729/1/L11](https://doi.org/10.1088/2041-8205/729/1/L11). arXiv: [1008.0349](https://arxiv.org/abs/1008.0349) [[astro-ph](https://arxiv.org/abs/astro-ph).C0] (cit. on pp. 61, 90, 94).
- Seo, H.-J., M. Sato, M. Takada, and S. Dodelson (2012). “Dark Energy from the Log-transformed Convergence Field.” In: ApJ 748, 57, p. 57. DOI: [10.1088/0004-637X/748/1/57](https://doi.org/10.1088/0004-637X/748/1/57). arXiv: [1109.5639](https://arxiv.org/abs/1109.5639) (cit. on pp. 61, 62, 90, 94).
- Shannon, C. E. (1948). “A mathematical theory of communication.” In: *The Bell System Technical Journal* 27.3, pp. 379–423. ISSN: 0005-8580. DOI: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x) (cit. on p. 52).
- Shapiro, S. S. and M. B. Wilk (1965). “An analysis of variance test for normality (complete samples).” In: *Biometrika* 52.3-4, pp. 591–611. DOI: [10.1093/biomet/52.3-4.591](https://doi.org/10.1093/biomet/52.3-4.591). eprint: <http://biomet.oxfordjournals.org/content/52/3-4/591.full.pdf+html>. URL: <http://biomet.oxfordjournals.org/content/52/3-4/591.short> (cit. on p. 54).
- Shirasaki, M. (2016). “Statistical connection of peak counts to power spectrum and moments in weak lensing field.” In: *ArXiv e-prints*. arXiv: [1610.00840](https://arxiv.org/abs/1610.00840) (cit. on p. 111).
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. ISBN: 0-412-24620-1 (cit. on pp. 48, 74).
- Simon, P. et al. (2015). “CFHTLenS: a Gaussian likelihood is a sufficient approximation for a cosmological analysis of third-order cosmic shear statistics.” In: MNRAS 449, pp. 1505–1525. DOI: [10.1093/mnras/stv339](https://doi.org/10.1093/mnras/stv339). arXiv: [1502.04575](https://arxiv.org/abs/1502.04575) (cit. on pp. 34, 59).

- Simpson, F. et al. (2013). “CFHTLenS: testing the laws of gravity with tomographic weak lensing and redshift-space distortions.” In: *MNRAS* 429, pp. 2249–2263. DOI: [10.1093/mnras/sts493](https://doi.org/10.1093/mnras/sts493). arXiv: [1212.3339](https://arxiv.org/abs/1212.3339) (cit. on p. 34).
- Skilling, J. (1989). “Classic Maximum Entropy.” In: *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*. Ed. by J. Skilling. Dordrecht: Springer Netherlands, pp. 45–52. ISBN: 978-94-015-7860-8. DOI: [10.1007/978-94-015-7860-8_3](https://doi.org/10.1007/978-94-015-7860-8_3). URL: http://dx.doi.org/10.1007/978-94-015-7860-8_3 (cit. on p. 43).
- (2006). “Nested sampling for general Bayesian computation.” In: *Bayesian Anal.* 1.4, pp. 833–859. DOI: [10.1214/06-BA127](https://doi.org/10.1214/06-BA127). URL: <http://dx.doi.org/10.1214/06-BA127> (cit. on p. 81).
- Sklar, A. (1959). “Fonctions de Répartition à n Dimensions et Leurs Marges.” In: *Publications de l’Institut de Statistique de L’Université de Paris* 8, pp. 229–231 (cit. on p. 63).
- Slipher, V. M. (1917). “Nebulae.” In: *Proceedings of the American Philosophical Society* 56, pp. 403–409 (cit. on p. 18).
- Smoot, G. F. et al. (1992). “Structure in the COBE differential microwave radiometer first-year maps.” In: *ApJ* 396, pp. L1–L5. DOI: [10.1086/186504](https://doi.org/10.1086/186504) (cit. on p. 24).
- Spergel, D. N., R. Flauger, and R. Hložek (2015). “Planck data reconsidered.” In: *Phys. Rev. D* 91.2, 023518, p. 023518. DOI: [10.1103/PhysRevD.91.023518](https://doi.org/10.1103/PhysRevD.91.023518). arXiv: [1312.3313](https://arxiv.org/abs/1312.3313) (cit. on p. 86).
- Staniszewski, Z. et al. (2012). “The Keck Array: A Multi Camera CMB Polarimeter at the South Pole.” In: *Journal of Low Temperature Physics* 167, pp. 827–833. DOI: [10.1007/s10909-012-0510-1](https://doi.org/10.1007/s10909-012-0510-1) (cit. on p. 27).
- Stephens, M. A. (1974). “EDF Statistics for Goodness of Fit and Some Comparisons.” In: *Journal of the American Statistical Association* 69.347, pp. 730–737. DOI: [10.1080/01621459.1974.10480196](https://doi.org/10.1080/01621459.1974.10480196). eprint: <http://www.tandfonline.com/doi/pdf/10.1080/01621459.1974.10480196>. URL: <http://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10480196> (cit. on p. 54).
- Stigler, S. M. (1983). “Who Discovered Bayes’s Theorem?” In: *The American Statistician* 37.4, pp. 290–296. ISSN: 00031305. URL: <http://www.jstor.org/stable/2682766> (cit. on p. 41).
- Sugai, H. et al. (2012). “Prime focus spectrograph: Subaru’s future.” In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Vol. 8446. Proc. SPIE, 84460Y. DOI: [10.1117/12.926954](https://doi.org/10.1117/12.926954). arXiv: [1210.2719](https://arxiv.org/abs/1210.2719) [astro-ph.IM] (cit. on p. 39).
- Sunyaev, R. A. and Y. B. Zel’dovich (1970). “Small-Scale Fluctuations of Relic Radiation.” In: *Ap&SS* 7, pp. 3–19. DOI: [10.1007/BF00653471](https://doi.org/10.1007/BF00653471) (cit. on p. 27).
- Suzuki, A. et al. (2016). “The Polarbear-2 and the Simons Array Experiments.” In: *Journal of Low Temperature Physics* 184, pp. 805–810. DOI: [10.1007/s10909-015-1425-4](https://doi.org/10.1007/s10909-015-1425-4). arXiv: [1512.07299](https://arxiv.org/abs/1512.07299) [astro-ph.IM] (cit. on p. 27).

- Szapudi, I. (2009). "Introduction to Higher Order Spatial Statistics in Cosmology." In: *Data Analysis in Cosmology*. Ed. by V. J. Martínez, E. Saar, E. Martínez-González, and M.-J. Pons-Bordería. Vol. 665. Lecture Notes in Physics, Berlin Springer Verlag, pp. 457–492. DOI: [10.1007/978-3-540-44767-2_14](https://doi.org/10.1007/978-3-540-44767-2_14) (cit. on p. 61).
- Szapudi, I. and N. Kaiser (2003). "Cosmological Perturbation Theory Using the Schrödinger Equation." In: *ApJ* 583, pp. L1–L4. DOI: [10.1086/368013](https://doi.org/10.1086/368013). arXiv: [astro-ph/0211065](https://arxiv.org/abs/astro-ph/0211065) (cit. on p. 61).
- Takada, M. and B. Jain (2003). "Three-point correlations in weak lensing surveys: model predictions and applications." In: *MNRAS* 344, pp. 857–886. DOI: [10.1046/j.1365-8711.2003.06868.x](https://doi.org/10.1046/j.1365-8711.2003.06868.x). arXiv: [astro-ph/0304034](https://arxiv.org/abs/astro-ph/0304034) (cit. on pp. 59, 88).
- (2004). "Cosmological parameters from lensing power spectrum and bispectrum tomography." In: *MNRAS* 348, pp. 897–915. DOI: [10.1111/j.1365-2966.2004.07410.x](https://doi.org/10.1111/j.1365-2966.2004.07410.x). arXiv: [astro-ph/0310125](https://arxiv.org/abs/astro-ph/0310125) (cit. on pp. 59, 88).
- (2009). "The impact of non-Gaussian errors on weak lensing surveys." In: *MNRAS* 395, pp. 2065–2086. DOI: [10.1111/j.1365-2966.2009.14504.x](https://doi.org/10.1111/j.1365-2966.2009.14504.x). arXiv: [0810.4170](https://arxiv.org/abs/0810.4170) (cit. on pp. 59, 88).
- Takahashi, R., M. Sato, T. Nishimichi, A. Taruya, and M. Oguri (2012). "Revising the Halofit Model for the Nonlinear Matter Power Spectrum." In: *ApJ* 761, 152, p. 152. DOI: [10.1088/0004-637X/761/2/152](https://doi.org/10.1088/0004-637X/761/2/152). arXiv: [1208.2701](https://arxiv.org/abs/1208.2701) (cit. on pp. 91, 92, 105).
- Taylor, A. N., B. Joachimi, and T. D. Kitching (2013). "Putting the precision in precision cosmology: How accurate should your data covariance matrix be?" In: *MNRAS* 432, pp. 1928–1946. DOI: [10.1093/mnras/stt270](https://doi.org/10.1093/mnras/stt270). arXiv: [1212.4359](https://arxiv.org/abs/1212.4359) (cit. on p. 60).
- Taylor, A. N. and T. D. Kitching (2010). "Analytic methods for cosmological likelihoods." In: *MNRAS* 408, pp. 865–875. DOI: [10.1111/j.1365-2966.2010.17201.x](https://doi.org/10.1111/j.1365-2966.2010.17201.x). arXiv: [1003.1136](https://arxiv.org/abs/1003.1136) (cit. on p. 82).
- Tegmark, M., A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay (1998). "Measuring the Galaxy Power Spectrum with Future Redshift Surveys." In: *ApJ* 499, pp. 555–576. arXiv: [astro-ph/9708020](https://arxiv.org/abs/astro-ph/9708020) (cit. on p. 60).
- Tegmark, M., A. N. Taylor, and A. F. Heavens (1997). "Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?" In: *ApJ* 480, pp. 22–35. eprint: [astro-ph/9603021](https://arxiv.org/abs/astro-ph/9603021) (cit. on p. 111).
- Torbet, E., M. J. Devlin, W. B. Dorwart, T. Herbig, A. D. Miller, M. R. Nolta, L. Page, J. Puchalla, and H. T. Tran (1999). "A Measurement of the Angular Power Spectrum of the Microwave Background Made from the High Chilean Andes." In: *ApJ* 521, pp. L79–L82. DOI: [10.1086/312197](https://doi.org/10.1086/312197). arXiv: [astro-ph/9905100](https://arxiv.org/abs/astro-ph/9905100) (cit. on p. 24).
- Trafimow, D. and M. Marks (2015). "Editorial." In: *Basic and Applied Social Psychology* 37.1, pp. 1–2. DOI: [10.1080/01973533.2015.1012991](https://doi.org/10.1080/01973533.2015.1012991). eprint: <http://dx.doi.org/10.1080/01973533.2015.1012991>. URL: <http://dx.doi.org/10.1080/01973533.2015.1012991> (cit. on p. 54).

- Trotta, R. (2008). "Bayes in the sky: Bayesian inference and model selection in cosmology." In: *Contemporary Physics* 49, pp. 71–104. DOI: [10 . 1080 / 00107510802066753](https://doi.org/10.1080/00107510802066753). arXiv: [0803.4089](https://arxiv.org/abs/0803.4089) (cit. on pp. [41](#), [46](#)).
- Turing, A. M. (1938). "Systems of Logic based on Ordinals." (unpublished). PhD thesis. University of Princeton (cit. on p. [52](#)).
- Van Waerbeke, L. et al. (2000). "Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures." In: *A&A* 358, pp. 30–44. arXiv: [astro-ph/0002500](https://arxiv.org/abs/astro-ph/0002500) (cit. on p. [34](#)).
- Velilla, S. (1993). "A note on the multivariate Box-Cox transformation to normality." In: *Statistics & Probability Letters* 17.4, pp. 259–263. ISSN: 0167-7152. DOI: [http://dx.doi.org/10.1016/0167-7152\(93\)90200-3](http://dx.doi.org/10.1016/0167-7152(93)90200-3). URL: <http://www.sciencedirect.com/science/article/pii/0167715293902003> (cit. on pp. [68](#), [97](#)).
- Verde, L., S. M. Feeney, D. J. Mortlock, and H. V. Peiris (2013). "(Lack of) Cosmological evidence for dark radiation after Planck." In: *J. Cosmology Astropart. Phys.* 9, 013, p. 013. DOI: [10 . 1088 / 1475 - 7516 / 2013 / 09 / 013](https://doi.org/10.1088/1475-7516/2013/09/013). arXiv: [1307.2904](https://arxiv.org/abs/1307.2904) [[astro-ph.CO](#)] (cit. on pp. [46](#), [81](#)).
- Verde, L., L. Wang, A. F. Heavens, and M. Kamionkowski (2000). "Large-scale structure, the cosmic microwave background and primordial non-Gaussianity." In: *MNRAS* 313, pp. 141–147. DOI: [10 . 1046 / j . 1365 - 8711 . 2000 . 03191 . x](https://doi.org/10.1046/j.1365-8711.2000.03191.x). arXiv: [astro-ph/9906301](https://arxiv.org/abs/astro-ph/9906301) (cit. on p. [59](#)).
- Vikhlinin, A., A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. Van Speybroeck (2006). "Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation." In: *ApJ* 640, pp. 691–709. DOI: [10 . 1086 / 500288](https://doi.org/10.1086/500288). arXiv: [astro-ph/0507092](https://arxiv.org/abs/astro-ph/0507092) (cit. on p. [22](#)).
- Wald, R. M. (1984). *General relativity*. University of Chicago Press. ISBN: 0-226-87033-2 (cit. on pp. [16](#), [30](#)).
- Walker, A. G. (1937). "On Milne's Theory of World-Structure." In: *Proceedings of the London Mathematical Society* s2-42.1, pp. 90–127. DOI: [10 . 1112 / plms / s2 - 42 . 1 . 90](https://doi.org/10.1112/plms/s2-42.1.90) (cit. on p. [17](#)).
- Walsh, D., R. F. Carswell, and R. J. Weymann (1979). "0957 + 561 A, B - Twin quasistellar objects or gravitational lens." In: *Nature* 279, pp. 381–384. DOI: [10 . 1038 / 279381a0](https://doi.org/10.1038/279381a0) (cit. on p. [30](#)).
- Wang, X., M. Neyrinck, I. Szapudi, A. Szalay, X. Chen, J. Lesgourgues, A. Riotto, and M. Sloth (2011). "Perturbation Theory of the Cosmological Log-density Field." In: *ApJ* 735, 32, p. 32. DOI: [10 . 1088 / 0004 - 637X / 735 / 1 / 32](https://doi.org/10.1088/0004-637X/735/1/32). arXiv: [1103.2166](https://arxiv.org/abs/1103.2166) (cit. on p. [61](#)).
- Wasserstein, R. L. and N. A. Lazar (2016). "The ASA's Statement on p-Values: Context, Process, and Purpose." In: *The American Statistician* 70.2, pp. 129–133. DOI: [10 . 1080 / 00031305 . 2016 . 1154108](https://doi.org/10.1080/00031305.2016.1154108). eprint: <http://dx.doi.org/10.1080/00031305.2016.1154108>. URL: <http://dx.doi.org/10.1080/00031305.2016.1154108> (cit. on p. [54](#)).

- Weinberg, D. H. (1992). “Reconstructing primordial density fluctuations. I - Method.” In: MNRAS 254, pp. 315–342. DOI: [10.1093/mnras/254.2.315](https://doi.org/10.1093/mnras/254.2.315) (cit. on p. 62).
- Weinberg, D. H., M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo (2013). “Observational probes of cosmic acceleration.” In: Phys. Rep. 530, pp. 87–255. DOI: [10.1016/j.physrep.2013.05.001](https://doi.org/10.1016/j.physrep.2013.05.001). arXiv: [1201.2434](https://arxiv.org/abs/1201.2434) (cit. on pp. 22, 38).
- Weinberg, S. (1972). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley-VCH. ISBN: 0-471-92567-5 (cit. on pp. 16, 30).
- (2008). *Cosmology*. Oxford University Press. ISBN: 978-0-19-852682-7 (cit. on pp. 17, 19, 21).
- Weyl, H. (1923). “Zur allgemeinen Relativitätstheorie.” In: *Physikalische Zeitschrift* 24, pp. 230–232 (cit. on p. 16).
- Wick, G. C. (1950). “The Evaluation of the Collision Matrix.” In: *Phys. Rev.* 80 (2), pp. 268–272. DOI: [10.1103/PhysRev.80.268](https://doi.org/10.1103/PhysRev.80.268). URL: <http://link.aps.org/doi/10.1103/PhysRev.80.268> (cit. on p. 56).
- Wiener, N. (1930). “Generalized harmonic analysis.” In: *Acta Mathematica* 55.1, pp. 117–258. ISSN: 1871-2509. DOI: [10.1007/BF02546511](https://doi.org/10.1007/BF02546511). URL: <http://dx.doi.org/10.1007/BF02546511> (cit. on p. 56).
- Wild, V. et al. (2005). “The 2dF Galaxy Redshift Survey: stochastic relative biasing between galaxy populations.” In: MNRAS 356, pp. 247–269. DOI: [10.1111/j.1365-2966.2004.08447.x](https://doi.org/10.1111/j.1365-2966.2004.08447.x). arXiv: [astro-ph/0404275](https://arxiv.org/abs/astro-ph/0404275) (cit. on p. 60).
- Wittman, D. M., J. A. Tyson, D. Kirkman, I. Dell’Antonio, and G. Bernstein (2000). “Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales.” In: *Nature* 405, pp. 143–148. arXiv: [astro-ph/0003014](https://arxiv.org/abs/astro-ph/0003014) (cit. on p. 34).
- Wraith, D., M. Kilbinger, K. Benabed, O. Cappé, J.-F. Cardoso, G. Fort, S. Prunet, and C. P. Robert (2009). “Estimation of cosmological parameters using adaptive importance sampling.” In: *Phys. Rev. D* 80.2, 023507, p. 023507. DOI: [10.1103/PhysRevD.80.023507](https://doi.org/10.1103/PhysRevD.80.023507). arXiv: [0903.0837 \[astro-ph.CO\]](https://arxiv.org/abs/0903.0837) (cit. on pp. 51, 52).
- Wu, K. K. S., O. Lahav, and M. J. Rees (1999). “The large-scale smoothness of the Universe.” In: *Nature* 397, pp. 225–230. DOI: [10.1038/16637](https://doi.org/10.1038/16637). arXiv: [astro-ph/9804062](https://arxiv.org/abs/astro-ph/9804062) (cit. on p. 17).
- Xavier, H. S., F. B. Abdalla, and B. Joachimi (2016). “Improving lognormal models for cosmological fields.” In: MNRAS 459, pp. 3693–3710. DOI: [10.1093/mnras/stw874](https://doi.org/10.1093/mnras/stw874). arXiv: [1602.08503](https://arxiv.org/abs/1602.08503) (cit. on p. 62).
- Yu, Y., P. Zhang, and Y. Jing (2016). “Fast generation of weak lensing maps by the inverse-Gaussianization method.” In: *ArXiv e-prints*. arXiv: [1607.05007](https://arxiv.org/abs/1607.05007) (cit. on pp. 89, 104).
- Yu, Y., P. Zhang, W. Lin, W. Cui, and J. N. Fry (2011). “Gaussianizing the non-Gaussian lensing convergence field: The performance of the Gaussianization.” In: *Phys. Rev. D*

- 84.2, 023523, p. 023523. DOI: [10 . 1103 / PhysRevD . 84 . 023523](https://doi.org/10.1103/PhysRevD.84.023523). arXiv: [1103 . 2858](https://arxiv.org/abs/1103.2858) [[astro-ph.CO](#)] (cit. on p. [62](#)).
- Yu, Y., P. Zhang, W. Lin, W. Cui, and J. N. Fry (2012). “Gaussianizing the non-Gaussian lensing convergence field II. The applicability to noisy data.” In: *Phys. Rev. D* 86.2, 023515, p. 023515. DOI: [10 . 1103 / PhysRevD . 86 . 023515](https://doi.org/10.1103/PhysRevD.86.023515). arXiv: [1201 . 4527](https://arxiv.org/abs/1201.4527) [[astro-ph.CO](#)] (cit. on p. [62](#)).
- Zwicky, F. (1933). “Die Rotverschiebung von extragalaktischen Nebeln.” In: *Helvetica Physica Acta* 6, pp. 110–127 (cit. on p. [21](#)).

