Cambridge University Press
978-0-521-17561-6 - Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications
Domenico Marinucci and Giovanni Peccati
Excerpt
More information

1

Introduction

1.1 Overview

The purpose of this monograph is to discuss recent developments in the analysis of isotropic spherical random fields, with a view towards applications in cosmology. We shall be concerned in particular with the interplay among three leading themes, namely:

- the connection between isotropy, representation of compact groups and spectral analysis for random fields, including the characterization of polyspectra and their statistical estimation;
- the interplay between Gaussianity, Gaussian subordination, nonlinear statistics, and recent developments in the methods of moments and diagram formulae to establish weak convergence results;
- the various facets of high-resolution asymptotics, including the high-frequency behaviour of Gaussian subordinated random fields and asymptotic statistics in the high-frequency sense.

These basic themes will be exploited in a number of different applications, some with a probabilistic flavour and others with a more statistical focus.

On the probabilistic side, we mention, for instance, a systematic study of the connections between Gaussianity, independence of Fourier coefficients, ergodicity and high-frequency asymptotics of Gaussian subordinated fields. We will also discuss at length the role of isotropy in constraining the behaviour of angular power spectra and polyspectra, thus providing a characterization of the dependence structure of random fields, as well as a sound mathematical background for establishing meaningful estimation procedures.

Among the statistical applications, we mention the estimation of angular power spectra and polyspectra, and their use to implement tests for Gaussianity, isotropy and asymmetry. A common thread of the statistical and probabilistic results will be the derivation of asymptotic results in the high-frequency sense – a concept whose rationale we shall illustrate below. In this same framework, another contribution of these lecture notes is the analysis of the stochastic properties of a new class of spherical wavelets, called needlets. For instance, we shall discuss asymptotic uncorrelation of needlets coefficients (again in the high-frequency sense) and show how this property can be exploited to derive a number of statistical procedures, related e.g. to the estimation of angular power spectra and polyspectra, as well as to the already mentioned tests of asymmetry and Gaussianity.

The book is completed by the analysis of random fields which do not take ordinary scalar values, but have a more complex geometrical structure – i.e., the so-called spin random fields. These fields can again be modeled and interpreted in terms of group representation concepts, thus maintaining a consistent connection with the leading themes of this monograph.

All our stochastic results are strongly motivated by applications, and in fact we will refer to several papers where the previous concepts have been successfully applied to the analysis of astrophysical data. Indeed, although spherical random fields may arise in a number of different circumstances, including medical imaging, atmospheric sciences, geophysics, solar physics and many others, our motivating rationale has been very much influenced by cosmological applications, in particular in connection with the analysis of the Cosmic Microwave Background (CMB) radiation data. Thus, although we believe that the results discussed here have a general mathematical interest and may find applications to different fields, we wish to provide in this Introduction an informal presentation of the foundations of CMB data analysis, which will help the reader to better understand the relevance and motivations of our work.

1.2 Cosmological motivations

Cosmology is now developing into a mature observational science, with a vast array of different experiments yielding datasets of astonishing magnitude, and nearly as great challenges for theoretical and applied statisticians. Datasets are now available on a large variety of different phenomena, but the leading part in cosmological research has been played over the past 20 years by the analysis of the Cosmic Microwave Background (CMB) radiation, an area which has already led to Nobel Prizes for Physics in 1978 and in 2006.

The nature of CMB can be loosely explained as follows (see for instance Dodelson [51] for a textbook reference and Balbi [9] for a more popular account). According to the standard cosmological model, the Universe that we currently observe has originated approximately 13.7 billion years ago in a very hot and dense state, in what is universally known as the *Big Bang*. Neglecting fundamental physics in the first fractions of seconds, we can naively imagine a fluid state where matter was completely ionized, i.e. the kinetic energy of electrons was much stronger than the attractive potential of the protons, so that no stable atomic nuclei could form. It is a consequence of quantum principles that a free electron has a much larger *cross-section* for interaction with photons than when it is bound in an atomic nucleus. Loosely speaking, it follows that the probability of interactions between photons and electrons was so high that the mean free path of the former was very short and the Universe was consequently "opaque". As the Universe expanded, the mean energy content decreased, meaning that the fluid composed of matter and radiation cooled down. The mean kinetic energy of the electrons thus decreased until it reached a critical value where it was no longer sufficient to escape the attractive electric potential of protons. Stable (and neutral) hydrogen atoms were then formed. This change of state occurred at the so-called "age of recombination", which is currently reckoned to have taken place 3.7×10^5 years after the Big Bang, i.e. when the Universe had only the 0.003% of its current age. At the age of recombination, the probability of interactions became so small that, as a first approximation, photons could start to travel freely. Neglecting second order effects, we can assume they had no further interaction up to the present epoch.

The remarkable consequence of this mechanism is that the Universe is embedded in a uniform radiation that provides pictures of its state nearly 1.37×10^{10} years ago; this is exactly the above-mentioned CMB radiation. The existence of CMB was predicted by G. Gamow in the forties; it was later discovered fortuitously by Penzias and Wilson in 1965 – for this discovery they earned the Nobel Prize for Physics in 1978. For several years, further experiments were only able to confirm the existence of the radiation, and to test its adherence to the Planckian curve of blackbody emission, as predicted by theorists. A major breakthrough occurred with NASA satellite mission *COBE*, which was launched in 1989 and publicly released the first full-sky maps of radiation in 1992; for this experiment Smoot and Mather earned the Nobel Prize for Physics in 2006 [187]. In the Figure 1.1 below we present a CMB map (the so-called ILC, Internal Linear Combination map from *WMAP* data), see Bennett et al. [21] for more details on its construction.

Full-sky maps as ILC are constructed by weighted linear interpolation of the observations across the different channels, but they are not considered fully reliable for data analysis, especially at high frequencies. Indeed, some parts of the sky are masked by the presence of foreground emission by the Milky Way and other *foreground* sources. This is a major issue for data analysis, which we shall deal with extensively in the final chapters of this monograph. In Figure 1.2, we show the map constructed from (the Q band of) *WMAP* data, where approximately 20% of the sky has been deleted to get rid of foreground emission; the missing region around the galactic plane is immediately evident.

The nature of these maps deserves further explanation. CMB is distributed in a remarkably uniform fashion over the sky, with deviations of the order of 10^{-4} with respect to the mean value (corresponding to 2.731 Kelvin). The attempts to understand this uniformity have led to very important developments in cosmology, primarily the inflationary scenario, which now dominates the theoretical landscape (see [51]). Even more important, though, are the tiny fluctuations around this mean value, which provided the seeds for stars and galaxies to form out of gravitational instability. Measuring and understanding the nature of these fluctuations has then been the core of an enormous amount of experimental and theoretical research. In particular, their stochastic properties yield a goldmine of information about a number of extremely important issues in astrophysics and cosmology, as well as many problems at the frontier of fundamental physics.

To mention just a few of these problems, we recall the issues concerning the matter content of the Universe, its global geometry, the existence and nature of (non-baryonic) *dark matter*, the existence and nature of *dark energy*, which is related to Einstein's cosmological constant, and many others. The next experimental landmark in CMB analysis followed in 2000, when two balloon-borne experiments, *BOOMERANG* and *MAXIMA* (see de Bernardis et al. [44] and Hanany et al. [91]) yielded the first high-resolution observations on small patches of the sky (less than 10° squared). These observations led to the first constraints on the global geometry of the Universe, which was found to be (very close to) Euclidean. Another major breakthrough followed with the 2003-2010 data releases from the NASA satellite experiment *WMAP*, whose observations are publicly available on the web site http://lambda.gsfc.nasa.gov/, see for instance Bennett et al. [22]. Such data releases yielded measurement of the correlation structure of the random field up to a resolution of about 0.22 degrees, i.e., approximately 30 times better than *COBE* (7-10 degrees). Finally, a major boost in data analysis is expected from the ESA satellite mission *Planck*, which was launched on May 14, 2009; data releases for the public are due in 2011-2015. *Planck* is planned to provide datasets of nearly 5×10^{10} observations, and this will allow to settle many open questions with CMB temperature data. New challenging question are expected to arise at a faster and faster pace over the next decades; for instance, *Planck* will provide high quality *polarization* data, which will set the agenda for the experiments to come. Polarization data can be viewed as spin, or tensor-valued, rather than scalar, observations.

CAMBRIDGE

Cambridge University Press
978-0-521-17561-6 - Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications
Domenico Marinucci and Giovanni Peccati
Excerpt
More information

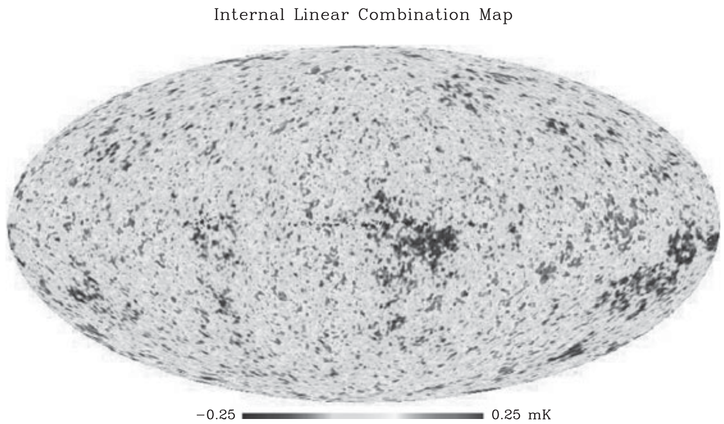

Figure 1.1 The Internal Linear Combination map from NASA-*WMAP* data

As such, they require an entirely new field of statistical research, which is still in its infancy but seems very promising for future developments - in particular, as we shall discuss in the final chapter of this book, polarization data lead to the analysis of spin random fields, which in a loose sense can be viewed as random structures that at each point take as a value a random curve (e.g. an ellipse) rather than a number. Quite interestingly, this same mathematical framework covers other astrophysical applications whose analysis is growing rapidly, such as weak gravitational lensing data [28, 115].

1.3 Mathematical framework

As introduced in the previous discussion, this monograph deals with mathematical topics at the intersection of probability, mathematical statistics and harmonic analysis on groups and homogeneous spaces. In this respect, we believe that two features of our approach deserve a special mention.

On one hand, we will provide a detailed analysis of isotropic (that is, rotationally invariant) spherical random fields by using notions from group representation theory. This approach will unveil some elegant mathematical structures, based e.g. on the properties of Clebsch-Gordan intertwining matrices and Gaunt integrals, and also places our discussion into the wider framework of

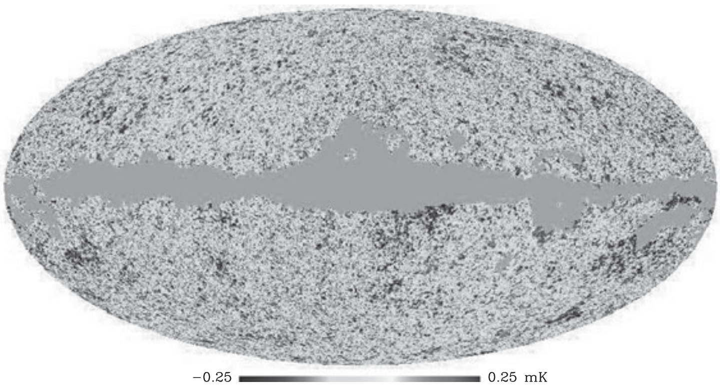

Figure 1.2 CMB radiation from the Q band of NASA-*WMAP* data

the algebraic characterization of stochastic processes defined on homogeneous spaces. See e.g. Diaconis [49], Gangolli [71], Guivarc'h, Keane and Roynette [89], among others, for examples of several fascinating interactions between probability, statistics and group theoretical concepts. Another crucial point is that the use of isotropy is directly related to the CMB analysis. In particular, following the standard literature, we can interpret the CMB radiation as a single realization of an isotropic, finite variance spherical random field. Note that, since the CMB is an image of the early universe, the underlying isotropy can be seen as a consequence of the so-called "Einstein cosmological principle", roughly stating that, on sufficiently large distance scales, the Universe looks identical everywhere in space (homogeneity) and appears the same in every direction (isotropy). One should also mention that the prevailing models for early Big Bang dynamics (the so-called *inflationary scenario*) predict the random fluctuations to be Gaussian, or polynomial (quadratic or cubic) functions of an underlying Gaussian field (see for instance [18, 19, 20]). This point justifies the fact that, in the discussion to follow, we will very often work with spherical random fields having a Gaussian, or Gaussian-subordinated, structure.

On the other hand, we shall systematically derive probabilistic limit theorems in the high-frequency (or high-resolution) sense. Roughly speaking, this means that (i) we are going to decompose a given spherical random field in terms of some deterministic basis whose elements are indexed by frequencies (i.e., *spherical harmonics*) and, (ii) we shall study the asymptotic properties (e.g. the asymptotic Gaussianity) of some relevant statistics of the frequency components by letting the frequency diverge to infinity. We will point out that this type of limit procedures yields a number of deep and difficult mathematical challenges. It is interesting to note that these difficulties arise even in the case of elementary statistics associated with random fields with a relatively simple structure (e.g., rotationally-invariant Gaussian fields). For instance, we will see in Chapter 9 that, in order to prove high-resolution CLTs for the angular bispectra of isotropic Gaussian fields, one must perform a subtle combinatorial analysis of the moments and cumulants associated with large sums of three-products of independent and complex-valued harmonic coefficients. More generally, the high-frequency asymptotic procedures undertaken in this monograph always require us to characterize the limits of linear combinations of random summands with a very complex dependence structure, which is both determined by the isotropic assumption and by the features of the underlying homogeneous space. To some extent, this situation is similar to those encountered in the analysis of large random matrices belonging to some special ensembles (for instance, the Wigner or Ginibre families). Indeed, random matrices from these ensembles can be easily described in terms of collections of i.i.d. (independent and identically distributed) random variables, but their asymptotic spectral analysis requires us to control and assess larger and larger partial sums of highly-correlated random eigenvalues (see e.g. Guionnet [88]). Finally, we stress once again that our asymptotic results are strongly motivated by physical applications. To understand this point, we observe that cosmology is (in some sense) a science based upon a single observation (i.e. our Universe) which is observed at greater and greater degrees of resolution. As we shall discuss below, high-frequency procedures provide the proper framework to describe the environment faced by applied researchers: as experiments grow more and more sophisticated, smaller and smaller scales (and hence higher and higher harmonic components) become available for data analysis. For example, *COBE* data included observations up to multipoles (equivalent to frequencies, to be defined later) in the order of a few dozens, current data from *WMAP* have raised this limit to a few hundreds, whereas the expected bound from *Planck* is in the order of a few thousands.

1.4 Plan of the book

The plan of the book is as follows. In Chapter 2, we review some basic facts from group representation theory. Part of this material is rather technical, and well beyond the classical background of most researchers working in stochastics; however we feel it is a necessary companion for the understanding of much of the ideas to follow. In fact, group representation results play a basic role in most of the material discussed in the book, from the derivation of spectral representation theorems to the analysis of high-resolution asymptotics for Gaussian subordinated random fields, from the statistical estimation of angular power spectra and bispectra (whose definition is entirely based on group theoretic ideas) to the introduction of spin random fields. The core of this chapter is the celebrated Peter-Weyl Theorem from abstract harmonic analysis, which allows generalized Fourier expansions to take place for functions defined on arbitrary compact groups. Our discussion follows well-known textbooks such as, for instance, Faraut [63] and Vilenkin and Klimyk [197].

In Chapter 3, we specialize the previous results to representation theory for the group of rotations $SO(3)$, which will allow us to introduce Fourier analysis on the sphere as well as the most important instruments for the chapters to follow, such as Wigner's D matrices, spherical harmonics, and Clebsch-Gordan coefficients. These are the fundamental tools used throughout the book in order to understand the role of rotations and isotropy, to establish asymptotic results and to analyze nonlinear transforms and higher order statistics.

Chapter 4 is also concerned with background results, but with a much stronger probabilistic flavour: in particular, we discuss at length recent developments in the analysis of central limit theorems for Gaussian subordinated sequences. We present the classical diagram formula for the analysis of nonlinear transforms of Gaussian processes, and we discuss at length much more recent results on the characterization of asymptotic Gaussianity by the use of so-called *contractions* (see Nualart and Peccati [152]). We also make use of some recent developments by Nourdin and Peccati [148] connecting Malliavin's calculus to the so-called Stein's method for probabilistic approximations. These techniques greatly simplify the derivation of Central Limit Theorems results and their use is widespread throughout the book. In this same chapter, the graphical method for dealing with convolutions of Clebsch-Gordan coefficients is also recalled (see Varshalovich, Moskalev and Khersonskii [195] for a rather exhaustive account of available results).

In Chapter 5 we introduce spectral representation results, again in connection with the abstract harmonic analysis setting of the previous chapter. These representation results can also be derived via more probabilistic techniques, and we discuss the interaction between the different approaches. Note that the group theoretic point of view is necessary in order to make some of the further developments more transparent.

Chapter 6 is more directly concerned with isotropic spherical random fields and their characterization. We start by discussing results from Baldi and Marinucci [10], which prove that the random harmonic coefficients of a spherical random field are independent if and only if the field is Gaussian. We then characterize the higher order moments of these coefficients in the general (not necessarily Gaussian) isotropic case by means of Clebsch-Gordan coefficients, which allow us to evaluate multiple integrals of spherical harmonics in a neat and sharp way. We can then introduce the definition and main properties of angular polyspectra, a fundamental tool in the statistical analysis of the chapters to follow (see Marinucci and Peccati [136]).

In Chapter 7 we present our first results on high-frequency asymptotics. In particular, the question we try to address is the following: given a nonlinear transform of a Gaussian field, what can be said about the limiting distribution of its Fourier components? This question was investigated by the authors in [134, 135] and is again strongly motivated by physical applications. Indeed, as discussed in the book, a major theme of cosmological research is related to the investigation of non-Gaussian features in CMB radiation. These non-Gaussian features typically take the form of nonlinear transformation of a subordinating Gaussian field, (see [18, 19, 20, 67, 68, 117, 184]); the exact form of the nonlinear terms depend on different scenarios for the Big Bang dynamics, i.e. different versions of the celebrated inflationary model first introduced by Guth in 1981 (see again [51]). As mentioned above, statistical inference is restricted to a single realization of the Universe, observed at higher and higher frequencies as the experiments get more sophisticated. The issue is then, whether the components at these frequencies allow for the consistent estimation of non-Gaussian components. We provide some answers to this question, in terms of conditions on the decay of the angular power spectra and random walks on the representations of the group of rotations $SO(3)$.

In Chapter 8 we dwell more directly into mathematical statistics issues. We start by reviewing very briefly some background issues on the construction of CMB maps and the analysis of instrumental noise. We then focus more directly on angular power spectra estimation, bias testing and bias correction, and we prove consistency and weak convergence under Gaussianity assumptions, in the idealistic circumstances of fully observed CMB maps. We then consider angular power spectra estimation under non-Gaussianity. Our main results (which follow [137]) can be connected in a surprising way to those of Chapter 7. In particular, the possibility of consistently estimating (as always, in the high-frequency sense) the angular power spectrum of the random fields turns out to be closely related to the discussion in Chapter 7 on asymptotic Gaussianity. Loosely speaking, we show in fact that under broad conditions, ergodicity of the empirical spectral measure (i.e., high-frequency consistency of the angular power spectrum) and asymptotic Gaussianity are equivalent. These results suggest a more general connection between ergodicity and Gaussianity, in a high-resolution sense.

Statistical analysis is further developed in Chapter 9, where the results on the characterization of higher moments under isotropy are exploited to introduce the (sample) bispectrum and its asymptotic behaviour. We review recent results on CLT and FCLT for these statistics (compare [131], [132]), with a heavy use of the diagram formula machinery which was introduced in previous chapters. The definition and asymptotic analysis of the bispectrum is entirely based on the properties of Clebsch-Gordan (or Wigner's) coefficients, so that the group-theory point of view proves once again to be mandatory in this field. We discuss some remarkable statistical features such as consistency, i.e. the divergence to infinity of bispectrum-based statistics under non-Gaussianity, even in the presence of a single realization of the CMB field.

In Chapter 10 we start to discuss needlets and their stochastic properties. Needlets are a new form of spherical wavelets which were introduced into the mathematical literature by Narcowich, Petrushev and Ward [143, 144]; the derivation of their stochastic properties is first due to Baldi, Kerkyacharian, Marinucci and Picard [13, 14] (see also [123, 124, 140]), while the first applications to CMB data are due to [162, 139], with many further developments provided for instance by [46, 65, 163, 164, 165, 175, 176], among others. We also discuss the wider class of Mexican needlets, introduced by Geller and Mayeli in [76, 77, 78], and applied to CMB in [179]. Needlets enjoy a number of important analytical properties that we shall discuss in some detail; from the perspective of this book, however, the key feature that makes them valuable for stochastic analysis is the asymptotic uncorrelation of needlets coefficients, in the high-resolution sense. In a loose sense, the uncorrelation property is stating that in the Gaussian case we can derive a growing array of asymptotically i.i.d. coefficients (after normalization) out of a single spherical random field. This of course, opens the way to the implementation of several statistical procedures.