J. Math. Anal. Appl. 409 (2014) 212–227

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

Normal approximations for wavelet coefficients on spherical Poisson fields

Claudio Durastanti [a], Domenico Marinucci [a,*], Giovanni Peccati [b]

[a] Department of Mathematics, University of Rome Tor Vergata, Italy

[b] Unité de Recherche en Mathématiques, Luxembourg University, Luxembourg

ARTICLE INFO

Article history:
Received 21 September 2012
Available online 25 June 2013
Submitted by U. Stadtmueller

Keywords:
Berry–Esseen bounds
Malliavin calculus
Multidimensional normal approximation
Poisson process
Stein's method
Spherical wavelets

ABSTRACT

We compute explicit upper bounds on the distance between the law of a multivariate Gaussian distribution and the joint law of wavelet/needlet coefficients based on a homogeneous spherical Poisson field. In particular, we develop some results from Peccati and Zheng (2010) [42], based on Malliavin calculus and Stein's methods, to assess the rate of convergence to Gaussianity for a triangular array of needlet coefficients with growing dimensions. Our results are motivated by astrophysical and cosmological applications, in particular related to the search for point sources in Cosmic Rays data.

1. Introduction

The aim of this paper is to establish multidimensional normal approximation results for vectors of random variables having the form of wavelet coefficients integrated with respect to a Poisson measure on the unit sphere. The specificity of our analysis is that we require the dimension of such vectors to grow to infinity. Our techniques are based on recently obtained bounds for the normal approximation of functionals of general Poisson measures (see [40,42]), as well as on the use of the localization properties of wavelet systems on the sphere (see [36], as well as the recent monograph [30]). A large part of the paper is devoted to the explicit determination of the above quoted bounds in terms of dimension.

1.1. Motivation and overview

A classical problem in asymptotic statistics is the assessment of the speed of convergence to Gaussianity (that is, the computation of explicit Berry–Esseen bounds) for parametric and nonparametric estimation procedures—for recent references connected to the main topic of the present paper, see for instance [16,29,54]. In this area, an important novel development is given by the derivation of effective Berry–Esseen bounds by means of the combination of two probabilistic techniques, namely the *Malliavin calculus of variations* and the *Stein's method* for probabilistic approximations. The monograph [6] is the standard modern reference for Stein's method, whereas [38] provides an exhaustive discussion of the use of Malliavin calculus for proving normal approximation results on a Gaussian space. The fact that one can use Malliavin calculus to deduce normal approximation bounds (in total variation) for functionals of Gaussian fields was first exploited in [37]—where one can find several quantitative versions of the "fourth moment theorem" for chaotic random variables proved in [39]. Lower bounds can also be computed, entailing that the rates of convergence provided by these techniques are sharp in many instances—see again [38].

* Corresponding author.
E-mail address: marinucc@mat.uniroma2.it (D. Marinucci).

0022-247X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmaa.2013.06.028

In a recent series of contributions, the interaction between Stein's method and Malliavin calculus has been further exploited for dealing with the normal approximation of functionals of a general Poisson random measure. The most general abstract results appear in [40] (for one-dimensional normal approximations) and [42] (for normal approximations in arbitrary dimensions). These findings have recently found a wide range of applications in the field of stochastic geometry—see [25,26,34,28,47] for a sample of geometric applications, as well as the webpage

http://www.iecn.u-nancy.fr/~nourdin/steinmalliavin.htm

for a constantly updated resource on the subject.

The purpose of this paper is to apply and extend the main findings of [40,42] in order to study the multidimensional normal approximation of the elements of the first Wiener chaos of a given Poisson measure. Our main goal is to deduce bounds that are well-adapted to deal with applications where the dimension of a given statistic increases with the number of observations. This is a framework which arises naturally in many relevant fields of modern statistical analysis; in particular, our principal motivation originates from the implementation of *wavelet systems on the sphere*. In these circumstances, when more and more data become available, a higher number of wavelet coefficients is evaluated, as it is customarily the case when considering, for instance, thresholding nonparametric estimators. We shall hence be concerned with sequences of Poisson fields, whose intensity grows monotonically. We then exploit the wavelet localization properties to establish bounds that grow linearly with the number of functionals considered; we are then able to provide explicit recipes, for instance, for the number of joint testing procedures that can be simultaneously entertained ensuring that the Gaussian approximation may still be shown to hold, in a suitable sense.

1.2. Main contributions

Consider a sequence of i.i.d. random variables $\{X_i : i \geq 1\}$ with values in the unit sphere $\mathbb{S}^2$, and define $\{\psi_{jk}\}$ to be the collection of the *spherical needlets* associated with a certain constant $B > 1$, see Section 3.1 for more details and discussion. Write also $\sigma_{jk}^2 = E[\psi_{jk}(X_1)^2]$ and $b_{jk} = E[\psi_{jk}(X_1)]$, and consider an independent (possibly inhomogeneous) Poisson process $\{N_t : t \geq 0\}$ on the real line such that $E[N_t] = R(t) \to \infty$, as $t \to \infty$. Formally, our principal aim is to establish conditions on the sequences $\{j(n) : n \geq 1\}$, $\{R(n) : n \geq 1\}$ and $\{d(n) : n \geq 1\}$ ensuring that the distribution of the centered $d(n)$-dimensional vector

$$Y_n = (Y_{n,1}, \dots, Y_{n,d(n)}) = \frac{1}{\sqrt{R(n)}}\left(\sum_{i=1}^{N(n)} \frac{\psi_{j(n)k_1}(X_i)}{\sigma_{j(n)k_1}} - \frac{R(n)b_{j(n)k_1}}{\sigma_{j(n)k_1}}, \dots, \sum_{i=1}^{N(n)} \frac{\psi_{j(n)k_{d(n)}}(X_i)}{\sigma_{j(n)k_{d(n)}}} - \frac{R(n)b_{j(n)k_{d(n)}}}{\sigma_{j(n)k_{d(n)}}}\right) \tag{1.1}$$

is asymptotically close, in the sense of some smooth distance denoted d_2 (see Definition 2.6), to the law of a $d(n)$-dimensional Gaussian vector, say Z_n, with centered and independent components having unit variance. The use of a smooth distance allows one to deduce minimal conditions for this kind of asymptotic Gaussianity. The crucial point is that we allow the dimension $d(n)$ to grow to infinity, so that our results require to explicitly assess the dependence of each bound on the dimension. We shall perform our tasks through the following main steps: (i) Proposition 4.1 deals with one-dimensional normal approximations, (ii) Proposition 5.4 deals with normal approximations in a fixed dimension, and finally (iii) in Theorem 5.5 we deduce a bound that is well-adapted to the case $d(n) \to \infty$. More precisely, Theorem 5.5 contains an upper bound linear in $d(n)$, that is, an estimate of the type

$$d_2(Y_n, Z_n) \leq C(n) \times d(n). \tag{1.2}$$

It will be shown in Corollary 5.6, that the sequence $C(n)$ can be chosen to be

$$O\left(1/\sqrt{R(n)B^{-2j(n)}}\right);$$

as discussed below in Remark 4.3, $R(n) \times B^{-2j(n)}$ can be viewed as a measure of the "effective sample size" for the components of Y_n.

1.3. About de-Poissonization

Our results can be used in order to deduce the asymptotic normality of de-Poissonized linear statistics with growing dimension. To illustrate this point, assume that the random variables X_i are uniformly distributed on the sphere. Then, it is well known that $b_{jk} = 0$, whenever $j > 1$. In this framework, when $j(n) > 1$ for every n, $R(n) = n$ and $d(n)/n^{1/4} \to 0$, the conditions implying that Y_n is asymptotically close to Gaussian, automatically ensure that the law of the *de-Poissonized* vector

$$Y'_n = (Y'_{n,1}, \dots, Y'_{n,d(n)}) = \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n} \frac{\psi_{j(n)k_1}(X_i)}{\sigma_{j(n)k_1}}, \dots, \sum_{k=1}^{n} \frac{\psi_{j(n)k_{d(n)}}(X_i)}{\sigma_{j(n)k_{d(n)}}}\right) \tag{1.3}$$

is also asymptotically close to Gaussian. The reason for this phenomenon is nested in the statement of the forthcoming (elementary) Lemma 1.1 (see also [9] for similar computations).