J. Math. Anal. Appl. 436 (2016) 1121–1148

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Gaussian approximation of nonlinear statistics on the sphere

Solesne Bourguin [a,*], Claudio Durastanti [b,1], Domenico Marinucci [b,2], Giovanni Peccati [c,3]

[a] *Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA*
[b] *Department of Mathematics, University of Rome Tor Vergata, Rome, Italy*
[c] *Unité de Recherche en Mathématiques, Luxembourg University, Luxembourg, Luxembourg*

A R T I C L E   I N F O

Article history:
Received 24 July 2014
Available online 29 December 2015
Submitted by V. Pozdnyakov

Keywords:
Malliavin calculus
Stein's method
Poisson process
U-statistics
Spherical wavelets

A B S T R A C T

We show how it is possible to assess the rate of convergence in the Gaussian approximation of triangular arrays of U-statistics, built from wavelets coefficients evaluated on a spherical Poisson field of arbitrary dimension. For this purpose, we exploit the Stein–Malliavin approach introduced in the seminal paper by Peccati, Solé, Taqqu and Utzet (2011); we focus in particular on statistical applications covering evaluation of variance in non-parametric density estimation and Sobolev tests for uniformity.

1. Introduction and overview

1.1. Motivations

The purpose of this paper is to establish quantitative central limit theorems for some U-statistics on wavelets coefficients evaluated either on spherical Poisson fields or on a vector of independent and identically distributed (i.i.d.) observations with values on a sphere. These statistics are motivated by standard problems in statistical inference, such as evaluation of the variance in density estimations and Sobolev tests of uniformity of the underlying Poisson measure. Such problems are certainly very classical in statistical inference; however, we shall investigate their solution under circumstances which are somewhat non-standard, for a number of reasons. In particular, we will focus mainly on "high-frequency" procedures, where the scale to be investigated and the number of tests to be implemented are themselves a function of the number

* Corresponding author.
E-mail addresses: solesne.bourguin@gmail.com (S. Bourguin), durastan@mat.uniroma2.it (C. Durastanti), marinucc@mat.uniroma2.it (D. Marinucci), giovanni.peccati@gmail.com (G. Peccati).

1 Supported by the ERC grant *Pascal* 277742.
2 Partially supported by the ERC grant *Pascal* 277742.
3 Partially supported by the grant F1R-MTH-PUL-12PAMP (PAMPAS).

http://dx.doi.org/10.1016/j.jmaa.2015.12.036
0022-247X/© 2015 Published by Elsevier Inc.

of observations available, according to rules to be discussed below; for these statistics, we shall establish quantitative central limit theorems by means of the so-called *Malliavin–Stein technique*. Such a technique will allow, for instance, to determine how many joint procedures can be run while maintaining a given level of accuracy in the Gaussian approximation for the sample distribution of the resulting statistics; as shown in Sections 1.4.2 and 1.4.3 below, a refined version of an argument contained in the classic paper by Dynkin and Mandelbaum [9] will allow us to extend our quantitative result (in a fully multidimensional setting) to the framework of U-statistics based on i.i.d. spherical observations.

As already mentioned, we shall assume that the domain of interest is the unit sphere $\mathbb{S}^q \subset \mathbb{R}^{q+1}$. The arguments we exploit can be extended to other compact manifolds, but we shall not pursue these generalizations here for brevity and simplicity; however, on the contrary of most of the existing literature, our procedures can also be easily adapted to cover "local" tests, i.e. the possibility that these spheres are only partially observable, as it is often the case for instance in astrophysical experiments, cf. for instance [37], see also the recent monograph [5] for several other applications of spherical data analysis.

Malliavin–Stein techniques for Poisson processes have recently drawn a lot of attention in the probabilistic literature, see for instance [6,20,21,26,28,32], as well as the textbooks [27] and [7] for background results on Gaussian approximations by means of Stein's method. As motivated above, our aim here is to apply and extend the now well-known results of [28,30] in order to deduce bounds that are well-adapted to the applications we mentioned; our principal motivation originates from the implementation of wavelet systems on the sphere in the framework of statistical analysis for Cosmic Rays data, as for instance in [16,19, 34,36]. As noted in [8], under these circumstances, when more and more data become available, higher and higher frequencies (i.e., smaller and smaller scales) can be probed. We shall hence be concerned with sequences of Poisson fields, whose intensity grows monotonically; it is then possible to exploit local Normal approximations, where the rate of convergence to the asymptotic Gaussian distribution is related to the scale parameter of the corresponding wavelet transform in a natural and intuitive way. Similar arguments were earlier exploited for linear statistics in [8]; the proofs in the nonlinear case we consider here are considerably more complicated from the technical point of view, but remarkably the main qualitative conclusions go through unaltered.

1.2. U-statistics on the Poisson space

We will now recall a few basic definitions on Poisson random measures and Stein–Malliavin bounds; we refer for instance to [29,31,33] for more discussions and details. Assuming that we are working on a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the following definition is standard:

Definition 1.1. Let $(\Theta, \mathcal{A}, \lambda)$ be a σ-finite measure space, and assume that λ has no atoms (that is, $\lambda(\{x\}) = 0$, for every $x \in \Theta$). A Poisson random measure on Θ with intensity measure (or control measure) λ is a collection of random variables $\{N(A) : A \in \mathcal{A}\}$, taking values in $\mathbb{Z}_+ \cup \{+\infty\}$, such that the following two properties hold:

1. $N(A)$ has Poisson distribution with mean $\lambda(A)$, for every $A \in \mathcal{A}$;
2. $N(A_1), \dots, N(A_n)$ are independent whenever $A_1, \dots, A_n \in \mathcal{A}$ are pairwise disjoint.

In what follows, we shall consider a special case of Definition 1.1; more precisely, we take $\Theta = \mathbb{R}_+ \times \mathbb{S}^q$, with $\mathcal{A} = \mathcal{B}(\Theta)$, the class of Borel subsets of Θ. The symbol N indicates a Poisson random measure on Θ, with homogeneous intensity given by $\lambda = \rho \times \mu$. We shall take $\rho(ds) = R \cdot \ell(ds)$, where ℓ is the Lebesgue measure and $R > 0$ is a fixed parameter, in such a way that $\rho([0, t]) := R_t = R \cdot t$. Also, we assume that μ is a probability on $\mathbb{S}^q$ of the form $\mu(dx) = f(x)dx$, where f is a density on the sphere. Given such an object, we will denote by N_t $(t > 0)$ the Poisson measure on $(\mathbb{S}^q, \mathcal{B}(\mathbb{S}^q))$ given by