On Minkowski functionals of random fields

Andriy Olenko

Department of Mathematics and Statistics,

La Trobe University, Australia

COSMO21 Chania, 25 May 2016

The presentation is based on some results in the paper:

- *N. Leonenko, A. Olenko, Sojourn measures of Student and Fisher-Snedecor random fields, Bernoulli, 20(3) (2014) 1454–1483.*

Summary

- Level 1: collecting, cleaning and organizing data. Result:

$$\eta(x) = [\eta_1(x), ..., \eta_k(x)], \quad x \in \Delta \subset \mathbb{R}^d.$$

- Level 2: statistical modelling and inference. The main tool is a statistic:

$$T = \sum_{x \in \Delta} G(\eta(x)) \quad \text{or} \quad T = \int_{\Delta} G(\eta(x))dx.$$

Summary

- Level 1: collecting, cleaning and organizing data. Result:

$$\eta(x) = [\eta_1(x), ..., \eta_k(x)], \quad x \in \Delta \subset \mathbb{R}^d.$$

- Level 2: statistical modelling and inference. The main tool is a statistic:

$$T = \sum_{x \in \Delta} G(\eta(x)) \quad \text{or} \quad T = \int_{\Delta} G(\eta(x))dx.$$

Example 1.

"Good trivial" examples:

– sample mean $T = \frac{1}{|\Delta|}\sum_{x\in\Delta} \eta_1(x), \quad G(\eta(x)) = \frac{1}{|\Delta|}\eta_1(x);$

– sample second moment $T = \frac{1}{|\Delta|}\sum_{x\in\Delta} \eta_1^2(x), \quad G(\eta(x)) = \frac{1}{|\Delta|}\eta_1^2(x)$

"Bad trivial" example: $\max(\eta_1(x))$.

Problem: What is the distribution of $\sum_{x \in \Delta} G(\eta(x))$?

Assumptions:

Asymptotical results: Δ is large/high resolutions;

Dependence: $\eta(x)$ are strongly dependent in Δ.

Answers:

Normal, Rosenblatt, ..., very strange/complicated.

Introduction

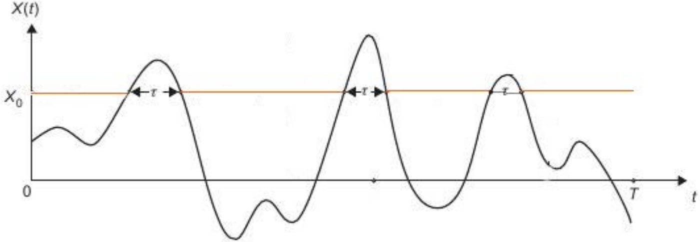
In the analysis of cosmic microwave background radiation it is common to characterise geometric properties by means of Minkowski functionals.

- Marinucci, D. (2004) Testing for non-Gaussianity on cosmic microwave background radiation: a review. *Statist. Sci.*, 2004, 19 (2), 294–307.
- Hikage, Ch., Matsubara, T. Limits on second-order non-Gaussianity from Minkowski functionals of WMAP 7-year data. *Monthly Notices of the Royal Astronomical Society*, 2012, 425 (3), 2187–2196.

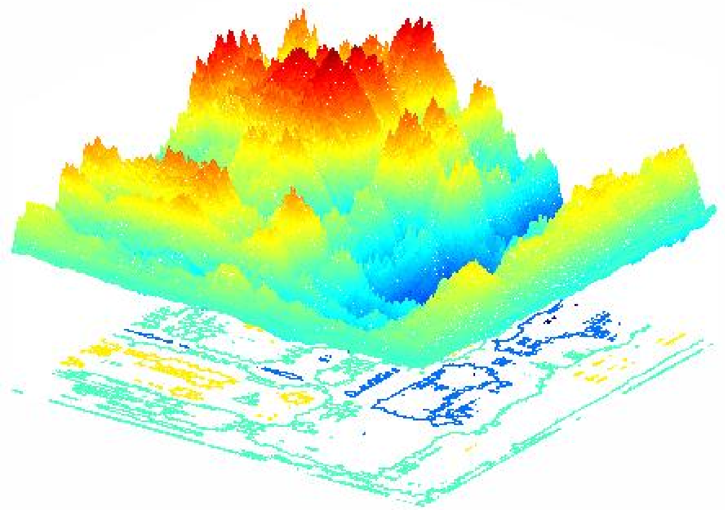
In this talk we discuss the first Minkowski functional of random fields.

Numerous real data have been modelled as Gaussian random processes or fields and studying of their sojourn measures is now a well developed subject. There is a very rich literature on the topic.

Sojourn measures of stochastic processes were studied extensively in a number of contexts and explicit formulae for their statistical characteristics were obtained for various scenarios.



- Berman, S. M. (1992) *Sojourns and Extremes of Stochastic Processes.* CA: Wadsworth & Brooks/Cole Advanced Books & Software.
- Brainina, I.S. (2013) *Applications of Random Process Excursion Analysis.* London: Elsevier.



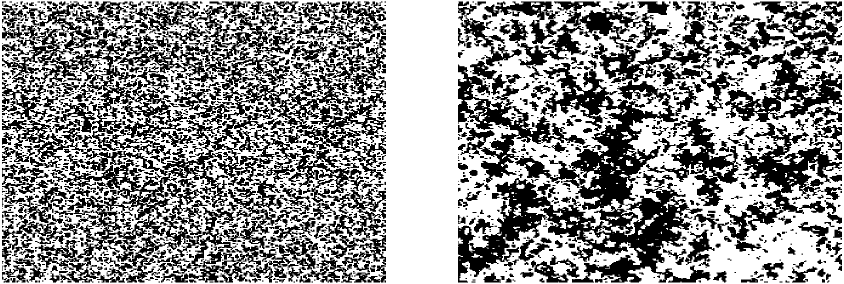
Unfortunately, one cannot expect similar simple results for the multidimensional situation. For random fields explicit formulae for the sojourn distributions are rarely known. Most published papers concern only first two moments of sojourn measures.

- Maejima, M. (1986) Some sojourn time problems for 2-dimensional Gaussian processes. *J. Multivariate Anal.*, **18** (1), 52–69.
- Adler, R. J. and Taylor, J. E. (2007) *Random Fields and Geometry.* New York: Springer.
- Borovkov, K. and McKinlay, S. (2012) The uniform law for sojourn measures of random fields. *Statist. Probab. Lett.*, **82** (9), 1745–1749.

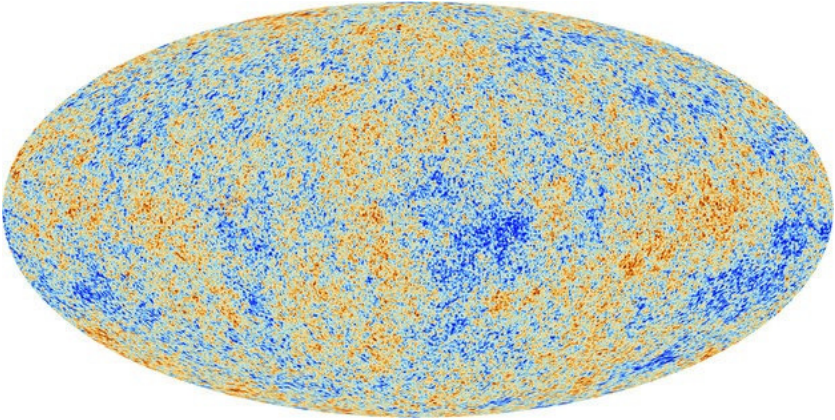
However, it turned out that there are some interesting asymptotic results in this area. Such results are usually the main tools for applications. It is natural to consider the volume of excursion sets in a bounded observation window and to study its limit behaviour as the window size grows.

Figure 1 shows two-dimensional excursion sets for realizations of two types of random fields (from left to right): short-range dependent normal scale mixture model and long-range dependent Cauchy model.

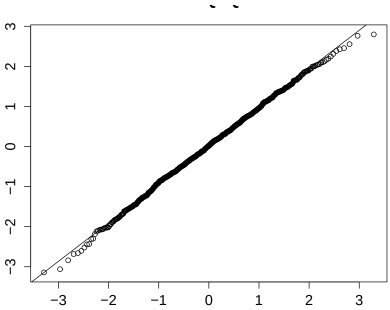
The excursion sets are shown in black colour.

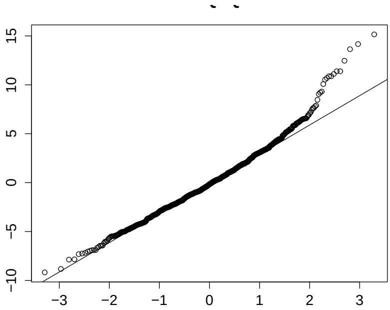


All-sky surveys from ESA's Planck space telescope:



The Q-Q plots, which correspond to the models shown above, suggest that the limit law of the short-range dependent model is normal, while for the long-range dependent model the data are not normally distributed.





Statistical Analysis shows that CMB fluctuations have a long-range correlation function.

- M. S. Movahed, F. Ghasemi, S. Rahvar, and M. R. R. Tabar. Long-range correlation in cosmic microwave background radiation. Phys. Rev. E 84, 021103, (2011).
- D.Marinucci and G. Peccati, Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications. London Mathematical Society Lecture Notes Series 389. Cambridge University Press, Cambridge, (2011).
- Current Topics in Astrofundamental Physics: The Cosmic Microwave Background. Editor N.G. Sánchez. Nato Science Series C. Springer, (2001).

First Minkowski functional

We consider a random field $S(x),\ x \in \mathbb{R}^d$.

Consider a bounded set $\Delta \subset \mathbb{R}^d$, such that $|\Delta| > 0$ and Δ contains the origin in its interior. Let $\Delta(r), r > 0$, be the homothetic image of the set Δ, with the centre of homothety in the origin and the coefficient $r > 0$, that is $|\Delta(r)| = r^d\,|\Delta|$.

Definition 2.

The first Minkowski functional is defined as

$$M_r\{S\} := |\{x \in \Delta(r) : S(x) > a(r)\}| = \int_{\Delta(r)} \chi\,(S(x) > a(r))\,dx,$$

where $\chi(\cdot)$ is an indicator function and $a(r)$ is a continuous nondecreasing function,

$$G(S(x)) = \chi\,(S(x) > a(r))\,.$$

Student and Fisher random fields

Let us consider the vector random field

$$\eta(x) = [\eta_1(x), ..., \eta_m(x), \eta_{m+1}(x), ..., \eta_{m+n}(x)]'$$

which consists of $n + m$ independent copies of a measurable mean-square continuous homogeneous isotropic zero-mean and unit variance Gaussian random field $\eta_1(x)$, $x \in \mathbb{R}^d$.

Definition 3.

The Fisher random field $F_{m,n}(x)$, $x \in \mathbb{R}^d$, is defined by

$$F_{m,n}(x) := \frac{\frac{1}{m}\left(\eta_1^2(x) + \cdots + \eta_m^2(x)\right)}{\frac{1}{n}(\eta_{m+1}^2(x) + \cdots + \eta_{m+n}^2(x))}, \quad x \in \mathbb{R}^d.$$

Definition 4.

The Student random field $T_n(x)$, $x \in \mathbb{R}^d$, is defined by

$$T_n(x) := \frac{\eta_1(x)}{\sqrt{\frac{1}{n}\left(\eta_2^2(x) + \cdots + \eta_{n+1}^2(x)\right)}}, \quad x \in \mathbb{R}^d.$$

Note that $[T_n(x)]^2 = F_{1,n}(x),\ x \in \mathbb{R}^d$.

Both Student and Fisher random fields have heavy-tailed marginal distributions.

- Worsley, K. J. (1994) Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. in Appl. Probab.*, 26 (1), 13–42.

Notations

Assumption 1.

Let $\eta(x) = [\eta_1(x), \dots, \eta_p(x)]'$, $x \in \mathbb{R}^d$, *be a vector homogeneous isotropic Gaussian random field with covariance matrix*

$$\tilde{\mathbf{B}}(0) = \mathcal{I}, \quad \tilde{\mathbf{B}}(x) = \mathbf{E}\eta(0)\,\eta(x)' = \mathcal{I} \cdot \|x\|^{-\alpha} L(\|x\|), \quad \alpha > 0,$$

where $\mathcal{I}$ *is the unit matrix of size* p, $L(\|\cdot\|)$ *is a function slowly varying at infinity.*

Let

$$\psi(x) := \max_{1 \le i \le p} \sum_{j=1}^{p} |\mathrm{B}_{ij}(\|x\|)|, \quad \sup_{x \in R^d} \psi(x) \le 1.$$

Weakly dependent scenario for Fisher and Student random fields

We consider sojourn measures above the constant level $a(r) \equiv a$.

Result 1.

If the covariance matrix of the Student random field $T_n(x)$, $x \in \mathbb{R}^d$, *satisfies the condition* $\boxed{\psi(\cdot) \in \mathbf{L}_1(\mathbb{R}^d),}$ *then*

$$r^{-d/2} M_r \{T_n\} - |\Delta| \, r^{d/2} \left(\frac{1}{2} - \frac{1}{2} \left(1 - I_{\frac{n}{n+a^2}} \left(\frac{n}{2}, \frac{1}{2} \right) \right) \cdot \operatorname{sgn}(a) \right) \xrightarrow{\mathcal{D}} Y_\Delta,$$

where $r \to \infty$, $\boxed{|\Delta|^{-1/2} \, Y_\Delta \sim N(0, \sigma_T^2).}$

For the Fisher random field $F_{m,n}(x)$, $x \in \mathbb{R}^d$, *we get*

$$r^{-d/2} M_r \{F_{m,n}\} - |\Delta| \, r^{d/2} \left(1 - I_{\frac{ma}{n+ma}} \left(\frac{m}{2}, \frac{n}{2} \right) \right) \xrightarrow{\mathcal{D}} Y_\Delta, \ \ r \to \infty.$$

Strongly dependent scenario for Fisher and Student random fields

Result 2.

Let $\eta(x) = [\eta_1(x), \ldots, \eta_{n+1}(x)]'$, $x \in \mathbb{R}^d$, *satisfy Assumption* 1 *for* $\alpha \in (0, d).$ *Then the random variable*

$$\sqrt{2\pi}\left(1 + a^2/n\right)^{n/2} \frac{M_r\{T_n\} - |\Delta|\, r^d \left(\frac{1}{2} - \frac{1}{2}\left(1 - I_{\frac{n}{n+a^2}}\left(\frac{n}{2}, \frac{1}{2}\right)\right) \cdot \mathrm{sgn}(a)\right)}{r^{d-\alpha/2} L^{1/2}(r) \sqrt{c_2(d,\alpha)\, c_3(1,d,\alpha)}}$$

is asymptotically $\mathcal{N}(0,1),$ *as* $r \to \infty$.

$$I_{\mu}(p,q) := \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} \int_0^{\mu} t^{p-1}(1-t)^{q-1}dt, \quad \mu \in (0,1], \quad p>0,\ q>0,$$

is the incomplete beta function.

The theorem demonstrates that for Student random fields, even in the case of strong dependence, we have a normal limit law. However, for the strongly dependent case the normalization is different from $r^{-d/2}$.

Contrary to the Student case, for strongly dependent Fisher random fields we obtain a non-normal limit law.

Result 3.

Let $\eta(x) = [\eta_1(x), \ldots, \eta_{n+m}(x)]'$, $x \in \mathbb{R}^d$, *satisfy Assumption* 1 *for* $\alpha \in (0, d/2).$ *Then, for* $r \to \infty$, *the distribution of the random variable*

$$\frac{M_r\{F_{m,n}\} - |\Delta|\, r^d \left(1 - I_{\frac{ma}{n+ma}}\left(\frac{m}{2}, \frac{n}{2}\right)\right)}{c_4(a, n, m)\, r^{d-\alpha} L(r)}$$

converges to the distribution of the random variable

$$\frac{X_{2,1} + \ldots + X_{2,m}}{m} - \frac{X_{2,m+1} + \ldots + X_{2,m+n}}{n},$$

where $X_{2,j}$, $j = 1, \ldots, m+n$, *are independent copies of* $X_2.$

Let

$$K(x) := \int_{\Delta} e^{i(x,u)} du, \quad x \in \mathbb{R}^d,$$

$$X_2 := c_2(d,\alpha) \int_{\mathbb{R}^{2d}}' K(\lambda_1 + \lambda_2) \frac{W(d\lambda_1) W(d\lambda_2)}{\|\lambda_1\|^{(d-\alpha)/2} \|\lambda_2\|^{(d-\alpha)/2}},$$

where $\int_{R^{2d}}'$ denotes the multiple Wiener-Itô integral.

Definition 5.

The probability distribution of X_2 is called the Rosenblatt-type distribution.

It is possible to generalize the results of the previous section to the increasing level $a(r) \to +\infty,$ as $r \to +\infty$.

Result 4.

Let $\eta(x) = [\eta_1(x), \ldots, \eta_{n+m}(x)]'$, $x \in \mathbb{R}^d$, *satisfy Assumption* 1 *for* $\alpha \in (0, d/2)$. *If* $a(r) = o(r^{\gamma/n}),$ $\gamma \in (0, \min(\alpha, d - \alpha))$, $r \to \infty$, *then the distribution of the random variable*

$$\frac{M_r\left\{F_{m,n}\right\} - |\Delta|\, r^d \left(1 - I_{\frac{ma(r)}{n+ma(r)}}\left(\frac{m}{2}, \frac{n}{2}\right)\right)}{c_4(a(r), n, m)\, r^{d-\alpha} L(r)}$$

converges to the distribution of the random variable

$$\frac{X_{2,1} + \ldots + X_{2,m}}{m} - \frac{X_{2,m+1} + \ldots + X_{2,m+n}}{n},$$

where $X_{2,j}$, $j = 1, \ldots, m+n$, *are defined in Theorem* 3.

Rate of convergence

Definition 6.

Let Y_1 and Y_2 be arbitrary random variables. The uniform (Kolmogorov) metric for the distributions of Y_1 and Y_2 is defined by the formula

$$\rho(Y_1, Y_2) = \sup_z \left| P(Y_1 \leq z) - P(Y_2 \leq z) \right|.$$

Assumption 2.

The random field $\eta(x)$, $x \in \mathbb{R}^d$, has the spectral density

$$f(\|\lambda\|) = c_2(d,\alpha) \, \|\lambda\|^{\alpha-d} \, L\left(\frac{1}{\|\lambda\|}\right),$$

where $c_2(d,\alpha) := \frac{\Gamma\left(\frac{d-\alpha}{2}\right)}{2^\alpha \pi^{d/2} \Gamma\left(\frac{\alpha}{2}\right)}$, and $L(\|\cdot\|)$ is a locally bounded function which is slowly varying with remainder at infinity.

Result 5.

If Assumptions 1 *and* 2 *hold, then for any* $\varkappa < \frac{1}{3}\min\left(\frac{\alpha(d-2\alpha)}{d-\alpha}, \varkappa_1\right)$,

$$\rho\left(\frac{2\,K_r}{C_2\, r^{d-\alpha} L(r)}, X_2\right) = o(r^{-\varkappa}), \quad r \to \infty.$$

The order of convergence $\varkappa$ depends on the parameters α and $\varkappa_1$:

- α is a long-range dependence parameter,
- $\varkappa_1$ gives the order for the upper bound of the slowly varying (with remainder) function $L(\cdot)$.

Future work

It would be interesting:

- to obtain similar results for other geometric functionals;
- to derive analogous results under different long-range assumptions on covariance functions of random fields;
- to derive statistical estimators and tests for CMB data using the obtained distributions;
- to compare the theoretical bound $o(r^{-\varkappa})$ with numerical results for actual and simulated data.