Extremal statistics for random geometric models

Supervision

- B. Błaszczyszyn, ENS-PSL, Département d'informatique, Bartek.Blaszczyszyn@ens.fr
- N. Chenavier, ULCO, LMPA J. Liouville, nicolas.chenavier@univ-littoral.fr

1 Introduction

1.1 The ubiquity of extremal spatial statistics

In many real-world applications, extremal statistics are crucial for understanding rare events that can have significant impacts, particularly in systems governed by randomness and spatial interactions. Random geometric models provide a natural framework to describe such systems, where objects (such as random clouds of discrete points or more complex structures derived from them) are distributed in space according to stochastic rules. By focusing on the locations and "magnitude" of extreme behaviors observed within these models, we gain valuable insights into rare but potentially catastrophic events across various fields.

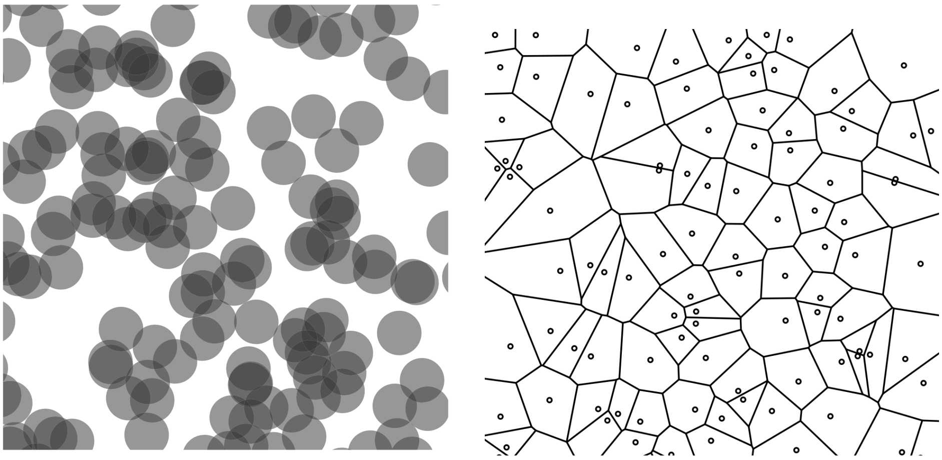
Examples include the study of environmental extremes (such as floods and storms) [19], infrastructure and urban planning (spatial congestion and potential failures) [2], natural resource exploration (rare but valuable discoveries) [15], public health (epidemics) [4], and financial risk and insurance (catastrophic losses due to natural disasters or market crashes) [17].

1.2 Stochastic Geometry

Random geometric models are often studied within the framework of Stochastic Geometry (SG) [13], a branch of probability theory. Similar in concept to Statistical Physics, SG focuses on analyzing large collections of randomly interacting particles, using ensemble averages to derive macroscopic laws from local interactions. In this context, we will be particularly interested in understanding the macroscopic laws governing extreme events.

To illustrate the problem of spatial extremal statistics in its simplest form, we can consider a homogeneous, completely random distribution of discrete points in Euclidean space — this is the well-known "spatial" Poisson point process [20]. Here, we can focus on identifying "isolated" points, which are those that have no neighbors within an "excessively" large distance. This provides a straightforward example of extreme behavior in random geometric models.

The example above helps illustrate how similar extremal statistics naturally arise in more complex models studied in Stochastic Geometry. Notably, in random geometric graphs, tessellations, and germ-grain models are extensively analyzed within this field. Figure 1 provides



(a) *(b)*

Figure 1: *(a)* A Boolean model. *(b)* The Voronoi tessellation.

examples of realizations for both Boolean models and random tessellations, highlighting the diversity of structures examined in SG.

Stochastic Geometry has been actively developed across various countries. Notable efforts are seen throughout Europe, including in the Czech Republic, Croatia, Germany, and France[1], as well as in the UK. Researchers in Australia, India, and the USA have also made significant contributions. The international community regularly gathers at the biennial conference "SGSIA", with the most recent edition held in Germany[2].

1.3 Spatial statistics and their limits

Spatial statistics are functions that capture the local characteristics of individual points or elements interacting within more complex random models. Examples include "score function" such as the distance to the closest point (mentioned in the illustrative example in the previous section), vertex degrees in random geometric graphs (corresponding to the number of balls intersecting a given ball, see Figure 1 a), and characteristics of cells in random tessellations (see Figure 1 b). These functions bridge theoretical models and the analysis of real-world phenomena.

To understand the average behaviour of these statistics, we examine their sum within an increasing window, beginning with the Law of Large Numbers (LLN). In Stochastic Geometry, this process involves assessing the behaviour of the statistic for a typical point or element of

[1] For example, the Thematic Network "MAIAGES" (`https://rt-maiages.math.cnrs.fr`) focuses on the Mathematics of Imaging, Learning, and Stochastic Geometry. This multidisciplinary research group aims to promote new mathematical methods in imaging and geometric data modeling, both stochastic and deterministic. The network encourages collaboration by supporting missions, invitations, and events, including its annual conference.

[2] `https://sgsia24.math.kit.edu/`

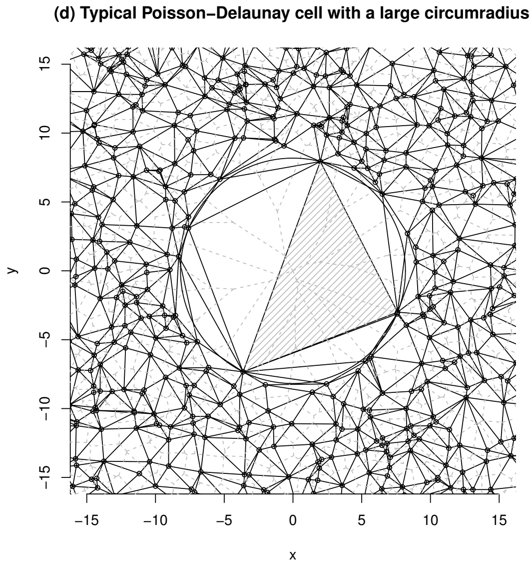


Figure 2: Large circumradius for a Poisson-Delaunay tessellation; see [12].

the model. Palm theory, as discussed in [5, Lesson 8–12], defines this concept by conditioning the model on having the typical point at the origin, linking it to the LLN through ergodicity.

Next, the Central Limit Theorem (CLT) is employed to characterize fluctuations of these statistics around the mean. A substantial body of literature addresses the CLT for geometric statistics in models driven by Poisson or binomial point processes. Various methods are used in this context, including the martingale approach and the Malliavin-Stein method, which utilizes the concept of independence [14]. For more general models, approaches based on moments and cumulants are often employed [8].

Extremal statistics is a specialized area within statistical analysis that focuses on understanding and modeling the behaviour of exceptional events or values within a dataset. In our context, it examines the exceptional values and locations of score functions (a.g. cells with large circumradius in Poisson-Delaunay, see Figure 2). Depending on the "degree of exceptionality" of these statistics, we can begin by applying the CLT to capture the fluctuations around a fixed amount above the mean. By increasing this degree of exceptionality, we can reach Poisson limits for rare occurrences of these statistics [1]. This framework leads one to the concept of large deviations,[3] where the probability of these rare events decreases exponentially fast to zero.

[3]see e.g. C. Hirsch's Large deviations project.

2 Research project

Our focus is on the extreme values of spatial statistics in large random geometric models. We propose the following areas of investigation.

Extremal characteristics of typical objects. The concept of a typical point or simple element in a large model captures the interactions of an "average" element with the rest of the model. To analyze its extremality, we explore the tail distribution of various characteristics of this typical object, which is formally defined using Palm theory. A notable example is the study by [10], which examines the structure of large cells within the typical cell of a Poisson-Voronoi tessellation. Understanding the Palm distribution is critical in such cases. *A challenging question arises when extending the analysis beyond the Poisson assumption.* In this context, capturing the spatial dependence of the statistics of interest becomes crucial — one of the key areas of investigation we address below.

Gaussian Fluctuation of Extremal Statistics The statistics of the typical object in Palm theory, through ergodic arguments, are typically related to the spatial frequency of the observed values of these characteristics within a large network. This is the essence of the spatial laws of large numbers. Consequently, the decay of the tail distributions of the characteristics of the typical object allows us to capture the rate of "sparsity" of these observations within the large model. When this rate approaches zero but does not do so too quickly (meaning the threshold value of a given statistic we are investigating in an increasing population is not rising too rapidly), we may still observe Gaussian fluctuations of these extreme value statistics around an asymptotically null rate. *A challenging question arises in establishing such Central Limit Theorems that go beyond previous results, where the law of large numbers provides a non-null rate.* In attempting to develop a new framework for such "moderate" deviations (which is similar but not identical to the works [16, 18]), we can explore a new approach proposed in [7] that offers a novel type of stabilization for the score functions, enabling us to capture their asymptotically extreme values.

Examples of this approach include carefully selected problems, such as examining large degree values in geometric models as discussed in [9], or analyzing the size of cells in tessellations within models based on point processes that exhibit attraction or repulsion.

Extreme statistics via Poisson germ-grain approximations. When extreme observations in the large model appear genuinely rare in terms of their low spatial frequency, their analysis might be feasible through a limiting approach using Poisson limit theory. In spatial statistics, when the regions with large values of the statistics in question are independently scattered throughout space, the limit is a compound Poisson process, see e.g. [11]. This conceptualization not only captures the spatial frequency of these events, identified by the intensity of the compound Poisson process, but also the distribution of the size of a cluster of exceedances. In this PhD thesis, we propose to also study the geometry of the clusters. The idea would then be to approximate the exceedance process using a Poisson germ-grain model. This type of result may not be directly accessible through the Palm study of the statistics of the typical point. Aldous'book [1] introduces an approach which leverages various techniques to prove Poisson convergence, including the Chen-Stein method (see e.g. [14]). Similar to the case of the CLT, the Poisson limit is expected to hold for a large class of non-Poisson models exhibiting reasonably weak dependence and/or true sparsity of extreme observations. *This approach could be applied to specific examples of extreme statistics, with a particular emphasis on addressing the geometry of regions exhibiting extreme statistical behavior.*

Capturing spatial dependence in SG models is a crucial aspect of limiting analyses, be it the CLT or Poisson limit. This connection is inherently linked to ergodicity, and several sufficient conditions are explored in this domain. These range from classical mixing considerations to various conditional properties and Brillinger mixing. Also, in recent developments in the theory of point processes, the concept of "de-correlation" was introduced in [8]. This concept, based on the evaluation of moment measures, aims to capture the joint "asymptotic independence" of the structure. It has proven useful in establishing the CLT and is expected to be beneficial in proving the Poisson limits mentioned earlier.

When not only asymptotic but also local capturing of dependence is required, a notion involving the comparison of moments and void probabilities with those of a Poisson point process was suggested in [6]. This comprehensive approach aims to capture the impact of positive and negative dependence among points and their effects on characteristics like coverage and percolation in random geometric models. Both the local and asymptotic approaches are naturally applicable to determinantal and permanental processes [3, Chapter 5]. These processes, inspired by fermions and bosons in statistical physics, respectively, exhibit inherent repulsion or attraction in point processes.

Our goal is to explore these two approaches to establish limiting or Palm results for some extreme statistics. The latter technique, directly related to moment measures and void probabilities, may be particularly relevant for studying extreme degree values in random geometric graphs and characteristics of large cells in tessellations.

References

[1] D. Aldous. *Probability approximations via the Poisson clumping heuristic*, volume 77. Springer Science & Business Media, 2013.

[2] F. Baccelli and B. Blaszczyszyn. *Stochastic Geometry and Wireless Networks, Volume II: Applications.* Now Publishers Inc., 2009.

[3] F. Baccelli, B. Błaszczyszyn, and M. Karray. *Random Measures, Point Processes, and Stochastic Geometry.* 2020. In preparation.

[4] A. Baddeley and R. Turner. *Spatial point patterns: Methodology and applications with R.* Chapman and Hall/CRC, 2005.

[5] B. Błaszczyszyn. *Lecture Notes on Random Geometric Models — Random Graphs, Point Processes and Stochastic Geometry.* 2017.

[6] B. Błaszczyszyn and D. Yogeshwaran. Clustering comparison of point processes with applications to random geometric models. In V. Schmidt, editor, *Stochastic Geometry, Spatial Statistics and Random Fields: Models and Algorithms*, volume 2120 of *Lecture Notes in Mathematics*, chapter 2, pages 31–71. Springer, 2014.

[7] B. Błaszczyszyn, D. Yogeshwaran, and J. E. Yukich. Limit theory for statistics of lipschitz-localized stochastic processes in spatial random models. private communication.

[8] B. Błaszczyszyn, D. Yogeshwaran, and J. E. Yukich. Limit theory for geometric statistics of point processes having fast decay of correlations. *Ann. Probab.*, 47(2):835–895, 2019.

[9] G. Bonnet and N. Chenavier. The maximal degree in a Poisson-Delaunay graph. *Bernoulli*, 26(2):948–979, 2020.

[10] P. Calka, Y. Demichel, and N. Enriquez. Large planar Poisson-Voronoi cells containing a given convex body. *Ann. H. Lebesgue*, 4:711–757, 2021.

[11] N. Chenavier and M. Otto. Compound poisson approximation under β-mixing and stabilization. *Available at https://arxiv.org/pdf/2310.15009.pdf*, 2023.

[12] N. Chenavier and C. Y. Robert. Cluster size distributions of extreme values for the poisson–voronoi tessellation. *The Annals of Applied Probability*, 28(6):3291–3323, 2018.

[13] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, third edition, 2013.

[14] L. Decreusefond, M. Schulte, and C. Thäle. Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Probab.*, 44(3):2147–2197, 2016.

[15] P. J. Diggle and P. J. G. Ribeiro. *Model-based Geostatistics*. Springer, 2007.

[16] P. Eichelsbacher and T. Schreiber. Process level moderate deviations for stabilizing functionals. *ESAIM: Probability and Statistics*, 14:1–15, 2010.

[17] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer Science & Business Media, 2013.

[18] M. Fenzl. Asymptotic results for stabilizing functionals of point processes having fast decay of correlations. *arXiv preprint arXiv:1909.13274*, 2019.

[19] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042–2065, 2009.

[20] G. Last and M. Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Stat. Cambridge University Press, 2017.