Théorèmes limites centraux et approximation poissonienne pour des processus ponctuels marqués - discrimination de processus et tests d'hypothèse

Contexte Ce sujet de thèse s'inscrit à l'interface des probabilités et de la statistique. Les questions étudiées sont motivées par des applications réelles issues du projet HisToGraM, dont les données proviennent de l'étude de tissus observés à différentes échelles. Les jeux de données considérés peuvent être représentés sous la forme de nuages de points dans le plan, chaque point correspondant à une entité caractérisée par diverses propriétés ou marques; par exemple, le type, la morphologie ou le transcriptome d'une cellule (figure 1).

Pour une condition donnée, on suppose que les données suivent la loi d'un processus ponctuel marqué, dont les paramètres varient selon la condition étudiée. Le problème central consiste alors à distinguer, d'un point de vue statistique, plusieurs processus ponctuels marqués et à identifier les statistiques discriminantes les plus pertinentes. Ces statistiques sont construites à partir de graphes associés aux processus ponctuels.

Théorèmes limites pour un processus ponctuel marqué Après avoir défini un modèle approprié, l'objectif de la thèse est d'établir des résultats asymptotiques d'une part en régime normal (théorèmes limites centraux) et d'autre part en régime poissonnien (étude d'extrêmes).

Concrètement, étant donné un processus ponctuel marqué $\{(x, m_x) : x \in \mathcal{P}_n\}$ d'intensité n sur $\mathbf{R}^d$ $(d \geq 1)$, et $G(\mathcal{P}_n)$ un graphe associé à $\mathcal{P}_n$, il s'agirait de montrer que, pour k fixé et à renormalisation et centrage standard près, la quantité

$$\sum_{x_{1:k} \in (\mathcal{P}_n)^k_{\neq}} \mathbf{1}_{x_{1:k} \in G(\mathcal{P}_n)} \mathbf{1}_{m_{x_{1:k}} \in A} \mathbf{1}_{z(x_{1:k}) \in [0,1]^d} \tag{1}$$

converge vers une loi gaussienne lorsque l'intensité tend vers l'infini, où A est un événement et où $z(x_{1:k})$ désigne un certain point de référence (par exemple le centre circonscrit du triangle associé à $x_{1:k}$). A titre d'exemple, $G(\mathcal{P}_n)$ peut désigner le graphe de Gilbert [11, Sec. 16.4] ou encore la triangulation de Delaunay [15, Sec. 10.2], deux objets classiques en géométrie aléatoire.

En régime poissonnien, l'idée serait de montrer que, sous de bonnes hypothèses, la statistique

$$\sum_{x_{1:k} \in (\mathcal{P}_n)^k_{\neq}} \mathbf{1}_{x_{1:k} \in G(\mathcal{P}_n)} \mathbf{1}_{m_{x_{1:k}} \in A} \mathbf{1}_{g(x_{1:k}, \mathcal{P}_n) > v_n} \mathbf{1}_{z(x_{1:k}) \in [0,1]^d} \tag{2}$$

converge vers une loi de Poisson, où v_n est un seuil convenablement choisi et g une certaine caractéristique géométrique (par exemple le rayon circonscrit à $x_{1:k}$). Une telle statistique apparaît naturellement dans l'étude d'extrêmes. En effet, obtenir une approximation poissonienne de cette dernière permet d'avoir des résultats sur les statistiques d'ordre, voir par exemple [7, Sec. 2.1].

De nombreux résultats asymptotiques ont été établis en géométrie aléatoire, sous la forme de théorèmes limites centraux ou d'approximations poissoniennes (par exemple [3, 4, 6]). Ces travaux

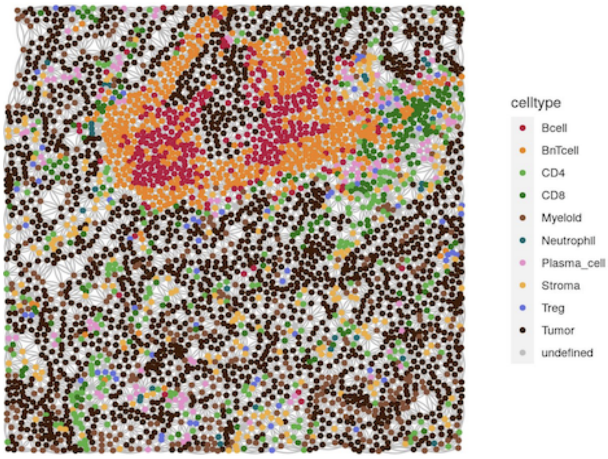

Figure 1: Génomique spatiale

concernent essentiellement des processus ponctuels simples ou des U-statistiques, pour lesquelles les théorèmes limites sont bien compris (par exemple [8, 13]). Dans un travail en cours et soumis prochainement, Błaszczyszyn, Yogeshwaran et Yukich établissent un théorème limite central associé à la statistique (1) dans le cas où les marques dépendent entièrement du processus $\mathcal{P}_n$ ou sont indépendantes de ce dernier.

Le défi central de ce sujet de thèse est d'étendre les résultats à des processus ponctuels marqués dont les marques dépendent conjointement de $\mathcal{P}_n$ et d'un processus indépendant. Une approche possible est d'adapter les méthodes de Stein [12, Sec. 3] et de Chen-Stein [1]. Ces dernières permettent notamment d'avoir des vitesses de convergence en régime normal et en régime poissonien. L'idée serait ensuite d'appliquer les résultats obtenus à diverses caractéristiques géométriques. Une autre piste naturelle serait d'établir des théorèmes limites centraux sous forme fonctionnelle, dans le même esprit que [2], où le processus est indexé par des sous-ensembles de $\mathbf{R}^d$.

Discrimination de processus ponctuels marqués Dans un deuxième temps, l'objectif serait de considérer conjointement deux processus ponctuels marqués et de tester si ces derniers proviennent de la même loi. Un premier angle d'attaque serait d'étendre les résultats obtenus dans la partie précédente pour deux processus et de localiser les points en lesquels les marques diffèrent significativement. Un second serait d'adapter des résultats de U-statistiques pour deux échantillons [10]. Ces dernières interviennent en effet de manière naturelle dans les détections de rupture.

Une autre utilisation des U-statistiques consiste à découper les observations en blocs de taille égale et à examiner l'écart entre les estimateurs sur chaque bloc, voir par exemple [9, 14] pour des estimateurs basés sur des moments empiriques. L'idée serait alors d'adapter de tels résultats pour des processus ponctuels marqués pour lesquels la statistique de test n'est pas une U-statistique classique.

L'approche pourrait également être complétée en établissant des inégalités de concentration et des résultats de grandes déviations. Une inégalité exponentielle a été obtenue pour des U-statistiques basées sur des processus de Poisson dans [5]. Une extension aux processus ponctuels marqués pourrait alors fournir une inégalité exponentielle pour la statistique définie par (2) ou encore pour la différence entre les statistiques associées aux deux processus ponctuels marqués.

Directeurs de thèse:

- Nicolas Chenavier, ULCO, LMPA J. Liouville, nicolas.chenavier@univ-littoral.fr
- Davide Giraudo, Univ. Strasbourg, IRMA, davide.giraudo@math.unistra.fr

Ce sujet de thèse s'inscrit dans le cadre du projet HIsToGraM soutenu par l'INSERM. La thèse aura lieu à l'Université du Littoral Côte d'Opale (Calais), sera financée à une hauteur de 2300 euros brut et commencera le 01/09/26.

References

[1] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4):403–434, 1990.

[2] H. Biermé and O. Durieu. Invariance principles for self-similar set-indexed random fields. *Trans. Amer. Math. Soc.*, 366(11):5963–5989, 2014.

[3] B. Bł aszczyszyn, 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.

[4] O. Bobrowski, M. Schulte, and D. Yogeshwaran. Poisson process approximation under stabilization and palm coupling. *Ann. Henri Lebesgue*, (5):1489–1534, 2022.

[5] G. Bonnet and A. Gusakova. Concentration inequalities for Poisson U-statistics, 2024.

[6] N. Chenavier and M. Otto. Compound Poisson process approximation under β-mixing and stabilization. *Ann. Appl. Probab.*, (35(4)):2544–2569, 2025.

[7] L. de Haan and A. Ferreira. *Extreme value theory. An introduction.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[8] 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.

[9] H. Dehling, D. Giraudo, and S. K. Schmidt. U-statistics of local sample moments under weak dependence. *ALEA Lat. Am. J. Probab. Math. Stat.*, 20(2):1511–1535, 2023.

[10] H. Dehling, D. Giraudo, and O. Sharipov. Convergence of the empirical two-sample U-statistics with β-mixing data. *Acta Math. Hungar.*, 164(2):377–412, 2021.

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

[12] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 2012.

[13] M. Reitzner and M. Schulte. Central limit theorems for u-statistics of Poisson point processes. *Ann. Probab.*, (41(6)):3879–3909, 2013.

[14] S. K. Schmidt, M. Wornowizki, R. Fried, and H. Dehling. An asymptotic test for constancy of the variance under short-range dependence. *Ann. Statist.*, 49(6):3460–3481, 2021.

[15] R. Schneider and W. Weil. *Stochastic and Integral Geometry.* Probability and its Applications (New York). Springer-Verlag, Berlin, 2008.