

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

AGRÉGATION D'ESTIMATEURS POUR L'ESTIMATION DE L'INTENSITÉ  
D'UN PROCESSUS PONCTUEL DE POISSON HOMOGENÈNE DANS  $R^2$

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES

PAR  
WANLING LI

OCTOBRE 2018

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Le présent travail a été réalisé grâce à l'aide de plusieurs personnes à qui je voudrais témoigner mes sincères remerciements.

Je voudrais remercier avant tout mon directeur principal de ce mémoire, Monsieur Jean-François Coeurjolly, pour ses judicieux conseils, ses avis critiques, ses corrections, sa disponibilité et sa patience.

Je voudrais également remercier mon co-directeur de ce mémoire, Monsieur François Watier, pour ses conseils académiques et professionnels, son soutien et ses encouragements pendant mes études universitaires.

Je voudrais aussi adresser mes remerciements aux professeurs du département de mathématiques, qui m'ont fourni des outils nécessaires pour réaliser mes études universitaires.

Je remercie ma famille pour son soutien. Merci à mon mari, qui m'a donné sans cesse des supports tout au long de mes études. Merci à mes parents en Chine, qui m'ont fourni toujours des encouragements.

Je remercie aussi mon superviseur à Statistique Canada, Monsieur José Gaudet, qui m'a beaucoup soutenu pendant la rédaction de ce mémoire.

Enfin, merci à Na qui m'a aidé beaucoup pour mes études et pour mon travail.

## TABLE DES MATIÈRES

LISTE DES TABLEAUX . . . . .	v
TABLE DES FIGURES . . . . .	vi
RÉSUMÉ . . . . .	vii
INTRODUCTION . . . . .	1
CHAPITRE I	
DIAGRAMME DE VORONOÏ . . . . .	4
1.1 Définition du diagramme de Voronoï ordinaire . . . . .	5
1.2 Propriétés de base du diagramme de Voronoï . . . . .	8
1.3 Génération du diagramme de Voronoï et de la tessellation de Delaunay en Matlab . . . . .	9
1.4 Diagramme de Voronoï infini . . . . .	11
1.5 Diagramme de Poisson-Voronoï . . . . .	13
CHAPITRE II	
INTENSITÉ DU PROCESSUS PONCTUEL HOMOGENÈME DE POISSON	18
2.1 Estimateur standard . . . . .	19
2.2 Estimateur Voronoï . . . . .	27
CHAPITRE III	
APPROCHE DE COMBINAISON DES ESTIMATEURS . . . . .	34
3.1 Méthode d'estimation pour un paramètre . . . . .	36
3.2 Performance de la méthode de Lavancier et Rochet . . . . .	38
CHAPITRE IV	
NOUVEL ESTIMATEUR AGRÉGÉ . . . . .	40
4.1 Etablissement du nouvel estimateur . . . . .	41
4.2 Simulation . . . . .	42
CONCLUSION . . . . .	54

ANNEXE A	
CODE . . . . .	56
BIBLIOGRAPHIE . . . . .	61

## LISTE DES TABLEAUX

Table	Page
4.1 Estimation avec eststd et estvtot . . . . .	46
4.2 Estimation avec estvtot, eststd et thetahat . . . . .	50

## TABLE DES FIGURES

Figure	Page
1.1 Diagramme de Voronoï avec 10 points pré-déterminés . . . . .	9
1.2 Triangulation de Delaunay . . . . .	10
2.1 Points réalisés avec $\lambda = 50$ , $ S  = 9$ . . . . .	30
2.2 Diagramme de Poisson-Voronoi . . . . .	31
4.1 Histogramme de la distribution des estimations avec estvtot . . .	47
4.2 Histogramme de la distribution des estimations avec eststd . . . .	48
4.3 Comparaison eststd vs estvtot . . . . .	49
4.4 Histogramme de la distribution des estimations avec thetihat . .	52
4.5 Comparaison entre estvtot, eststd et thetihat . . . . .	53

## RÉSUMÉ

Le présent travail a pour l'objectif d'établir un nouvel estimateur pour estimer l'intensité d'un processus de Poisson homogène. C'est une application de la méthode de combinaison des estimateurs initiaux et le nouvel estimateur se manifeste très bien en comparant avec les estimateurs initiaux.

Mots-Clés : intensité d'un processus de Poisson homogène, processus ponctuel spatial, estimateur Voronoï, diagramme de Poisson-Voronoï, estimateur agrégé.

## INTRODUCTION

L'objectif de ce mémoire est de réaliser une nouvelle estimation pour l'intensité d'un processus de Poisson homogène dans  $R^2$  et de comparer ses caractéristiques avec d'autres estimateurs.

De nombreuses recherches ont été déjà faites sur l'estimation de l'intensité, et nous en citons quelques-uns ici : Meijering (1953) sur l'estimateur Voronoï, Guan (2006) sur l'estimation du maximum de vraisemblance, Illian (2008) sur l'estimation de l'intensité du processus ponctuel spatial, etc. Dans le présent travail, nous voulons chercher une meilleure estimation pour l'intensité d'un processus de Poisson homogène.

Dans la première partie, la définition du diagramme de Voronoï ordinaire et celle du diagramme de Poisson-Voronoï ont été données et certaines propriétés ont été étudiées. Nous avons aussi fait une introduction pour les applications du diagramme de Voronoï. Ensuite, nous avons donné l'algorithme pour générer un diagramme de Poisson-Voronoï.

Dans la deuxième partie, une étude spéciale a été faite sur les estimateurs de l'intensité du processus de Poisson homogène. Les estimateurs existants incluent la moyenne, l'estimateur du maximum de vraisemblance, l'estimateur de la méthodes des moments, l'estimateur de la méthode de moindre carrés ordinaires, l'estimateur Voronoï, l'estimateur Centroidal Voronoï et l'estimateur Kernel, etc. Certains

estimateurs ont été étudiés dans ce chapitre et utilisés à nouveau dans le chapitre quatre dans le but d'établir un nouvel estimateur de l'intensité d'un processus de Poisson homogène.

Dans la troisième partie, une méthode de combinaison des estimateurs initiaux afin d'obtenir une meilleure estimation a été donnée. Cette méthode avait été introduite par Frédéric Lavancier et Paul Rochet (2016), qui ont beaucoup développé l'idée similaire dans les recherches existantes. La méthode de Frédéric Lavancier et Paul Rochet s'applique non seulement aux modèles paramétriques et semi-paramétriques, mais aussi aux modèles non-paramétriques. Mais ils ont mis l'accent sur les modèles paramétriques et semi-paramétriques. Un résumé sur la performance de cette méthode a été aussi donnée dans cette partie.

Dans la quatrième partie, une tentative a été faite pour établir un meilleur estimateur en base des estimateurs initiaux à l'aide de la méthode de Lavancier et Rochet (2016). Certains estimateurs présentés dans la partie deux ont été utilisés dans ce chapitre pour établir ce nouvel estimateur. Dans ce chapitre des simulations ont été faites pour illustrer et comparer les distributions des estimateurs et cela permet de donner une idée sur la performance de cette méthode appliquée pour estimer l'intensité d'un processus de Poisson homogène.

Ce mémoire est une nouvelle application de la méthode de la combinaison des estimateurs pour estimer l'intensité d'un processus de Poisson homogène, et il a sa limite. Il a mis l'accent sur le cas dans  $R^2$  et n'a pas traité les cas dans la dimension plus haute. Bien qu'on sache que cette méthode fonctionne aussi bien pour les cas dans la dimension plus haute, une étude rigoureuse devra être faite

dans la future recherche. En même temps, nous n'avons traité que l'estimation de l'intensité pour le processus de Poisson homogène. Les cas pour le processus ponctuel inhomogène de Poisson et les cas pour d'autres processus ponctuels sont à étudier dans le futur.

## CHAPITRE I

### DIAGRAMME DE VORONOÏ

Le diagramme de Voronoï existe partout dans la vie quotidienne. Des taches sur le corps des girafes aux terres fissurées à cause de la sécheresse, de la structure de l'aile de libellule à l'enveloppe de litchi, tous ces exemples mentionnés donnent l'impression que ce n'est qu'une simple image dans la vie. Cependant le diagramme de Voronoï est aussi largement trouvé et étudié dans de divers domaines de recherche, notamment quand on veut partitionner l'espace en zones d'influence.

L'histoire d'utilisation des diagrammes de Voronoï peut remonter au 17<sup>ième</sup> siècle (Descartes), et au fur et à mesure, les diagrammes de Voronoï ont été beaucoup utilisés dans les domaines tels que l'illustration astronomique, les études de forme quadratique, l'épidémiologie, la géographie, l'agriculture, etc. Par exemple, en entrant un code postal sur une page de site web d'un magasin, on peut trouver l'adresse du magasin le plus proche de ce code postal. C'est exactement un exemple d'utilisation des diagrammes de Voronoï dans notre vie quotidienne. Maintenant nous utilisons aussi les diagrammes de Voronoï dans les domaines de hautes-technologies tels que l'informatique, la télécommunication et l'apprentissage automatique. Les études sur le diagramme de Voronoï peuvent être étendues aux  $m$  dimensions. Dans ce mémoire, nous n'étudions les cas dans 2 dimensions pour

faciliter la compréhension.

### 1.1 Définition du diagramme de Voronoï ordinaire

Dans cette partie, nous donnons d'abord une définition intuitive pour le diagramme de Voronoï dans  $R^2$ . Ensuite on établit une définition plus précise avec les termes mathématiques.

Soit  $P$  un ensemble fini de  $n$  points discrets dans un plan  $R^2$  et  $n$  est plus grand ou égal à 2, on associe chaque position dans le plan avec le point le plus proche dans l'ensemble  $P$ . On obtient ainsi un ensemble de régions et le plan est donc divisé en plusieurs polygones. Cette tessellation est appelée le *diagramme de Voronoï* dans  $R^2$ . Et les régions sont appelées les *polygones de Voronoï*.

Après avoir donné la définition intuitive, nous voulons maintenant établir la définition avec les termes mathématiques.

Soit  $n$  le nombre de points dans  $R^2$  et soit  $2 \leq n < \infty$ , on a les  $n$  points  $p_1, \dots, p_n$  avec les vecteurs  $x_1, \dots, x_n$  dans le plan cartésien. Soit  $p$  un point dans  $R^2$  avec le vecteur  $x$ . Si  $p_i$  est le point le plus proche de  $p$  ou un des points les plus proches de  $p$ , on a alors la relation  $\|x - x_i\| \leq \|x - x_j\|$  pour  $j \neq i, i, j \in I_n = \{1, \dots, n\}$ . Dans ce cas,  $p$  est associé à  $p_i$ .

La définition mathématique pour le diagramme de Voronoï est donnée ci-dessous.

Soit

$$P = \{p_1, \dots, p_n\} \subset \mathbb{R}^2, \text{ où } 2 \leq n < \infty, \text{ et } x_i \neq x_j \text{ pour } j \neq i, i, j \in I_n.$$

La région donnée par

$$V(p_i) = \{x \mid \|x - x_i\| \leq \|x - x_j\| \text{ pour } j \neq i, i, j \in I_n\}$$

est appelée le *polygone de Voronoï associé avec  $p_i$*  où le *polygone de  $p_i$* , et l'ensemble donné par

$$V = \{V(p_1), \dots, V(p_n)\}$$

est appelé le *diagramme de Voronoï* généré par  $P$ , et s'écrit aussi comme  $V(P)$ .

Et  $p_i$  est appelé le *générateur* ou la *graine* du  $i^{\text{ème}}$  polygone de Voronoï.

Dans les définitions ci-dessus, nous avons défini un diagramme de Voronoï dans un plan ouvert. Cependant dans les applications, les gens ont souvent les besoins d'utiliser une région bornée  $B$ , où se trouvent les graines. Dans ce cas nous devons considérer l'ensemble

$$V_{\cap B} = \{V(p_1) \cap B, \dots, V(p_n) \cap B\},$$

ce qui est appelé le *diagramme de Voronoï borné*.

Nous pouvons, alternativement, définir le diagramme de Voronoï avec les termes de demi-plan car le diagramme de Voronoï est composé de polygones de Voronoï et qu'un polygone de Voronoï peut être défini avec les termes de demi-plan. Considérons maintenant la médiatrice du segment  $\overline{p_i p_j}$ , noté par  $b(p_i, p_j)$ . Alors

$$b(p_i, p_j) = \{x \mid \|x - x_i\| = \|x - x_j\|\}, \quad j \neq i.$$

Alors la médiatrice divise le plan en deux demi-plans et donne la région

$$H(p_i, p_j) = \{x \mid \|x - x_i\| \leq \|x - x_j\|\}, \quad j \neq i.$$

Nous appelons la région  $H(p_i, p_j)$  la *région dominante de  $p_i$  sur  $p_j$* , qui est le demi-plan délimité par la médiatrice du segment  $\overline{p_i p_j}$ . L'intersection des régions dominantes de  $p_i$  forme le polygone de Voronoï de  $p_i$ . Ainsi nous donnons une définition alternative du diagramme de Voronoï.

Soit

$$P = \{p_1, \dots, p_n\} \subset \mathbb{R}^2, \text{ où } 2 \leq n < \infty, \text{ et } x_i \neq x_j \text{ pour } j \neq i, i, j \in I_n.$$

On appelle la région

$$V(p_i) = \bigcap_{j \in I_n \setminus \{i\}} H(p_i, p_j)$$

le *polygone de Voronoï* ordinaire associé avec  $p_i$  et l'ensemble

$$V(P) = \{V(p_1), \dots, V(p_n)\}$$

le *diagramme de Voronoï* ordinaire généré par  $P$ .

Evidemment cette définition alternative équivaut à la définition mathématique car  $\|x - x_i\| \leq \|x - x_j\|$  si et seulement si  $x \in H(p_i, p_j)$  pour  $j \neq i$ .

Nous pouvons joindre toutes les paires de graines dont les polygones de Voronoï partagent la même frontière. Nous obtenons ainsi le graphe dual du diagramme

de Voronoï et on l'appelle la *triangulation de Delaunay*.

Un triangle de Delaunay est défini comme un ensemble fermé qui ne contient que les segments comme les frontières. Les points finaux d'une frontière de Delaunay sont appelés les *sommets de Delaunay*. Evidemment chaque sommet de Delaunay est une graine de  $V(P)$ , et par conséquent, l'ensemble de sommets de Delaunay est donné par  $P$ .

## 1.2 Propriétés de base du diagramme de Voronoï

Ayant donné les définitions du diagramme de Voronoï et de la triangulation de Delaunay, on veut maintenant donner quelques propriétés de base du diagramme de Voronoï et de la triangulation de Delaunay (pour la précision, vous pouvez voir Okabe (2000)).

- $V(p_i)$  est un polygone convexe non vide et le diagramme  $V(P)$  est unique dans le plan pour l'ensemble  $P$ .
- La graine la plus proche de  $p_i$  génère une frontière de Voronoï de  $V(p_i)$ . Le générateur  $p_i$  est le point de générateur le plus proche du point  $p$  si et seulement si  $V(p_i)$  contient  $p$ .
- La triangulation de Delaunay  $D(P)$  est aussi unique pour l'ensemble  $P$ . Les frontières externes de Delaunay constituent l'enveloppe convexe de  $P$ . Chaque graine du diagramme de Voronoï constitue un sommet dans la triangulation de Delaunay et les sommets du diagramme de Voronoï sont les centres des cercles circonscrits des triangles de la triangulation de Delaunay.

### 1.3 Génération du diagramme de Voronoï et de la tessellation de Delaunay en Matlab

Puisque les points discrets dans le diagramme de Voronoï sont déterministes ou non-déterministes, alors pour simuler le diagramme de Voronoï, nous pouvons suivre les étapes suivantes :

- avoir les points discrets fixés en avance ou générer des points aléatoires
- générer le diagramme de Voronoï ou la triangulation de Delaunay avec Matlab vue que les fonctions pour générer le diagramme de Voronoï et la triangulation de Delaunay sont déjà prédéfinies dans Matlab.

Nous générons d'abord un diagramme de Voronoï avec 10 points pré-déterminés.

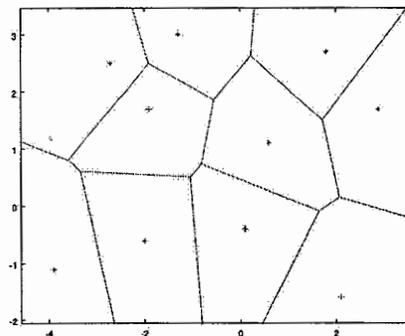
```

1 clear all
2 X = [-1.3 3.0; 1.8 2.7; -2.7 2.5; -1.9 1.7; 0.6 1.1; 2.9 1.7; ...
      -3.9 -1.1;-2.0 -0.6; 0.1 -0.4; 2.1 -1.6];
3 voronoi(X(:,1),X(:,2))

```

Nous obtenons le diagramme de Voronoï suivant :

FIGURE 1.1 Diagramme de Voronoï avec 10 points pré-déterminés



Maintenant nous générons la triangulation de Delaunay.

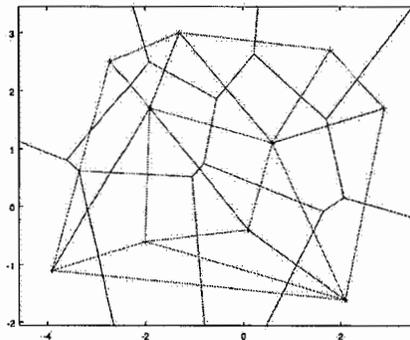
```

1 hold on
2 tri= delaunay(X(:,1),X(:,2));
3 triplot(tri, X(:,1),X(:,2), 'r');
4 hold off

```

Nous obtenons ainsi la triangulation de Delaunay (la partie rouge).

FIGURE 1.2 Triangulation de Delaunay



Nous voyons clairement que les sommets de Delaunay sont bien les générateurs de Voronoï. Et les frontières extérieurs de la triangulation forment l'enveloppe convexe de Delaunay. C'est-à-dire tous les points dans  $P$  se trouvent sur ou à l'intérieur de l'enveloppe convexe de Delaunay.

Maintenant nous voudrions générer un diagramme de Voronoï avec 10 points qui ne sont pas pré-déterminés. Nous générons d'abord 10 points aléatoires sur le plan.

```

1 clear all;
2 rand('state', 9);
3 x= rand(10,1);
4 y= rand(size(x));
5 voronoi(x,y)

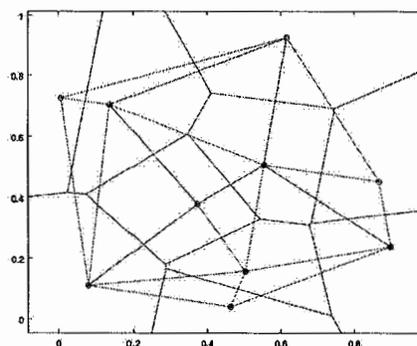
```

```

6 hold on;
7 tri=delaunay(x,y);
8 plot(x,y,'o');
9 triplot(tri, x, y, 'r');
10 hold off;

```

Nous obtenons le diagramme de Voronoï et la triangulation de Delaunay suivant :



Il est clair que l'enveloppe convexe de Delaunay forme le plus petit polygone qui englobe tous les points de germes dans l'ensemble  $P$ .

#### 1.4 Diagramme de Voronoï infini

La seule condition pour générer un diagramme de Voronoï fini est que les  $n$  points dans l'ensemble  $P$  sont finis et leur nombre est supérieur ou égale à un.

Si le nombre des points dans  $P$  sont infinis, nous obtenons alors un diagramme de Voronoï infini. Pour faciliter les études, nous supposons toujours que le nombre des points sont localement finis. Ça veut dire que le nombre des points sur la région bornée est fini. Nous supposons aussi que les points sont avec la position générale quadratique, qui est définie à la page suivante.

Pour être avec la position générale quadratique, les points doivent satisfaire aux deux hypothèses suivantes :

- hypothèse de non-colinéarité : il n'y a pas de  $k + 1$  points qui se trouvent sur l'hyperplan de  $k - 1$  dimension, où  $k = 2, \dots, m$  ;
- hypothèse de non-cosphéricité : il n'y pas de  $m + 2$  points qui se trouvent sur la même sphère dans  $R^m$ .

Dans le cas de  $R^2$ , ceci signifie qu'il n'y pas de trois points qui se trouvent sur la même droite et qu'il n'y a pas de quatre points qui se trouvent sur le même cercle.

Quel que soit  $P$  fini ou infini, il génère une tessellation de  $R^m$  si  $P$  est un ensemble dans  $R^m$ . C'est-à-dire il donne un diagramme de Voronoï de  $m$ -dimension car l'intérieur de chaque cellule de diagramme de Voronoï ne se superpose pas et les cellules remplissent collectivement l'espace.

Dans le paragraphe précédent, nous donnons une supposition que les points sont localement finis pour étudier le diagramme de Voronoï infini. Certaines propriétés du diagramme de Voronoï infini sont aussi données ici (pour la précision, veuillez voir Okabe (2000)).

- Si une cellule de Voronoï est bornée, elle a alors un nombre fini de voisins.
- Si toutes les cellules sont bornées, alors le diagramme de Voronoï est régulier.
- Chaque diagramme de Voronoï qui est généré par les points qui ont une position générale quadratique est une tessellation normale.

- Si les points sont localement finis et ont une position générale quadratique et que  $P$  est une réalisation d'un processus ponctuel simple et stationnaire,  $\Theta$ , dans  $R^m$ , alors le diagramme de Voronoï généré par ce processus ponctuel,  $V_\Theta$ , est une tessellation aléatoire de  $R^m$ .

Un processus ponctuel  $N$  sur  $S$ , qui est un sous-ensemble de  $R^m$ , est dit simple si dans la réalisation il n'y a pas de points qui coïncident. C'est-à-dire

$$N(\{\xi\}) \in \{0, 1\} \text{ presque sûrement pour tout } \xi \in S.$$

Un processus ponctuel dans  $R^m$  est dit stationnaire si sa distribution est invariante sous les translations.

Ici  $V_\Theta$  est une tessellation aléatoire régulière et normale. Et la tessellation aléatoire la plus importante est le diagramme de Poisson-Voronoï.

### 1.5 Diagramme de Poisson-Voronoï

$V_\Theta$  est le diagramme de Voronoï généré par le processus ponctuel  $\Theta$ . Une fois que  $\Theta$  génère  $V_\Theta$ , d'autres processus de points peuvent être générés par  $V_\Theta$ . Dans ce sens-là, un processus ponctuel stationnaire est très important.

Dans cette section nous nous concentrons sur le cas où les éléments de  $P$  sont localisés selon le processus ponctuel de poisson homogène et nous voulons étudier le *diagramme de Poisson-Voronoï*  $V_P$ . Nous mettons l'accent ici sur le cas dans  $R^2$  et nous supposons que  $P$  est localement fini, c'est-à-dire fini sur la région bornée, et que le processus ponctuel est simple (la probabilité pour le cas où deux points coïncident est 0).

Le diagramme de Poisson-Voronoi est généré par une réalisation du processus ponctuel de Poisson homogène  $\Theta_P$ . Pour un diagramme de Poisson-Voronoi, le nombre des graines est une variable aléatoire discrète qui suit la loi de Poisson et les points sont distribués selon la loi uniforme.

Pour des points avec une distribution spatiale, un processus ponctuel de Poisson est un processus de comptage  $N$  tel que :

- $N(\emptyset) = 0$
- $N(A)$  est de loi de Poisson pour tout  $A$  borné où  $A$  est un sous-ensemble inclus dans  $R^m$ .
- Pour  $A_1, \dots, A_n$  disjoints deux à deux, les variables  $N(A_1), \dots, N(A_n)$  sont indépendantes.

Pour simuler le processus ponctuel homogène de Poisson dans  $A$  nous simulons donc

$$N(A) \sim P(\lambda|A|),$$

et conditionnement à

$$N(A) = n,$$

nous générons  $n$  points indépendants et de loi uniforme dans  $A$ .

Nous étudions le cas où  $m = 2$ , c'est donc un processus ponctuel de Poisson du plan  $R^2$ . Nous avons alors pour tout  $A$  borné de  $R^2$ ,

$$E[N(A)] = \lambda|A|$$

où  $\lambda$  est l'intensité du processus ponctuel homogène (nombre moyen de points par unité d'aire), qui est une constante positive, et  $|A|$  l'aire de  $A$ .

L'intensité  $\lambda$  peut donc être écrite comme

$$\lambda = \frac{E[N(A)]}{|A|}.$$

Le processus ponctuel homogène de Poisson possède des caractéristiques suivantes :

- $0 < Pr(N(A) = 0) < 1$  si  $0 < |A| < \infty$ .
- La probabilité de  $N(A)$  dépend de  $|A|$  et  $Pr(N(A) \geq 1)$  tend vers 0 lorsque  $|A|$  tend vers 0. Ceci implique que la probabilité de distribution de  $N(A)$  ne dépend pas de la localisation de  $A$ .
- Si  $A_1, \dots, A_n$  sont disjoints, alors les variables  $N(A_1), \dots, N(A_n)$  sont mutuellement indépendantes et  $N(A_1 \cup \dots \cup A_n) = N(A_1) + \dots + N(A_n)$ . Ceci signifie qu'il n'y a pas d'interaction parmi les points.
- La probabilité que deux points coïncident est de zéro.

Le diagramme de Poisson-Voronoi possède les propriétés suivantes :

*Propriété 1* : Les cellules du diagramme de Poisson-Voronoi sont des polytopes.

C'est-à-dire elles sont des polyèdres bornés et convexes.

*Propriété 2* : Le diagramme de Poisson-Voronoi est régulier et normal.

*Propriété 3* : Le diagramme de Poisson-Voronoi est stationnaire et isotropique.

Il y a deux principales utilisations du diagramme de Poisson-Voronoi. Premièrement, le diagramme de Poisson-Voronoi peut être directement utilisé comme le modèle d'une structure empirique. Deuxièmement, il peut être utilisé comme le modèle standard avec lequel les autres tessellations peuvent être comparées et ensuite être évaluées.

Plus précisément, dans les sciences naturelles, le diagramme de Poisson-Voronoi a été utilisé comme le modèle microstructural du matériau polycristallin. Par exemple, Kumar et Singh (1995) ont utilisé la forme de 3-dimension comme le modèle microstructural de matériau polycristallin dans leurs recherches sur la conductivité thermique.

Le diagramme de Poisson-Voronoi a aussi été utilisé dans le domaine géographique. Par exemple, Boots (1975) a utilisé le diagramme de Poisson-Voronoi dans  $R^2$  pour étudier les zones de services du transport commun en Angleterre.

En plus de ça, le diagramme de Poisson-Voronoi peut être utilisé dans le domaine de la théorie quantique des champs. Par exemple, Drouffe et Itzykson (1984) ont utilisé le diagramme de Poisson-Voronoi pour construire le réseau aléatoire dans leurs recherches sur la théorie quantique des champs.

Dans le domaine du réseau de télécommunication, une station est servie à la région qui est plus proche de cette station que d'autres stations. Ces stations, soit les points de concentrations, sont générés par le processus de Poisson dans  $R^2$ . Baccelli et Blaszczyzyn (2001) ont travaillé sur l'application de la tessellation de Poisson-Voronoi sur la communication sans fil.

Le diagramme de Poisson-Voronoi a aussi été appliqué dans le domaine de géométrie. Par exemple, Kumar *et al.* (1992) et Okabe *et al.* (2000) ont fait des descriptions statistiques sur les propriétés du diagramme de Poisson-Voronoi.

Les recherches sur l'intensité du processus ponctuel spatial jouent un rôle importante dans les domaines mentionnés en haut. Puisque le processus ponctuel de Poisson est tellement important et l'intensité est le seul paramètre, donc l'estimation de l'intensité est aussi très importante. Dans la partie suivante, l'accent est mis sur l'intensité du processus ponctuel homogène de Poisson qui génère ensuite le diagramme de Poisson-Voronoi.

## CHAPITRE II

### INTENSITÉ DU PROCESSUS PONCTUEL HOMOGÈNE DE POISSON

Pour le processus ponctuel homogène de Poisson, il y a déjà des méthodes communes pour estimer l'intensité, qui est une constante dans le cas homogène, par exemple, méthode du maximum de vraisemblance, méthode de moment, etc. Nous traitons ici l'estimation de l'intensité d'un processus de Poisson homogène, et une introduction pour les estimateurs existants est fournie ici pour donner un aperçu des estimateurs. Certains estimateurs seront aussi utilisés dans la partie quatre, avec la méthode présentée dans la partie trois, afin de produire un meilleur estimateur de l'intensité du processus ponctuel de Poisson homogène.

Un petit rappel pour le processus ponctuel homogène de Poisson est donné ici :

$$Pr(N(A) = x) = \frac{(\lambda|A|)^x e^{-\lambda|A|}}{x!}, x = 0, 1, 2, \dots, \forall A \subset R^m$$

Ici  $N$  est un processus ponctuel homogène de Poisson.

$N(A)$  signifie le nombre de points généré par le processus de Poisson sur la région bornée  $A$ .

$\lambda$  est l'intensité du processus et c'est le nombre de points sur une région-unité. Puisque nous étudions le cas dans le processus de Poisson homogène,  $\lambda$  est une constante ici. Le nombre des points générés sur une unité de région suit donc la distribution de Poisson avec paramètre  $\lambda$ . Et  $N(A)$ , le nombre des points produits sur la région bornée  $A$  suit donc la distribution de Poisson avec paramètre  $\lambda|A|$ , et s'écrit ci-dessous :

$$N(A) \sim \text{Poisson}(\lambda|A|)$$

où  $|A|$  est le volume ou l'aire de  $A$ .

Si  $\lambda$  est un paramètre inconnu, nous devons donc estimer  $\lambda$ .

## 2.1 Estimateur standard

Soit  $S$  un sous-ensemble de  $R^d$ . Considérons une région échantillonnée bornée  $A \subseteq S$ . Supposons que tous les points dans  $A$  sont dénombrables. Considérons maintenant un processus ponctuel de Poisson homogène  $N(A)$  défini sur  $S$ . Un estimateur naturel de  $\lambda$  est  $\frac{N(A)}{|A|}$ , où  $|A|$  signifie le volume de  $A$ . Dans le cas de  $R^2$ ,  $|A|$  = l'aire de  $A$ . Alors cet estimateur naturel s'exprime comme le suivant :

$$\hat{\lambda} = \frac{N(A)}{|A|},$$

où

$$N(A) \sim \text{Poisson}(\lambda|A|).$$

Puisque  $N(A)$  suit la loi de Poisson, nous avons donc

$$E(N(A)) = \lambda|A|$$

et

$$\text{var}(N(A)) = \lambda|A|.$$

Puisque c'est un estimateur naturel, l'intérêt est mis sur ses propriétés statistiques suivantes :

*Propriété 1* : L'estimateur naturel est un estimateur sans biais.

Le biais peut être utilisé pour évaluer un estimateur. Nous souhaitons que l'espérance de l'estimateur d'un paramètre soit égale à la vraie valeur du paramètre. Leur différence est nommé le biais :

$$\text{Biais}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Si le biais est égal à zéro, l'estimateur est dit sans biais. Dans ce cas-là, l'espérance de l'estimateur du paramètre est égale à la vraie valeur du paramètre.

Nous montrons ici que l'estimateur naturel de  $\lambda$  est un estimateur sans biais.

Preuve :

L'estimateur naturel de  $\lambda$  s'exprime comme :

$$\hat{\lambda} = \frac{N(A)}{|A|}$$

où

$$N(A) \sim \text{Poisson}(\lambda|A|).$$

Alors

$$\begin{aligned} E(\hat{\lambda}) &= E\left[\frac{N(A)}{|A|}\right] \\ &= \frac{1}{|A|} E[N(A)] \\ &= \frac{1}{|A|} \lambda|A| \\ &= \lambda. \end{aligned}$$

Donc l'estimateur naturel  $\hat{\lambda} = \frac{N(A)}{|A|}$  est un estimateur sans biais.

*Propriété 2* : La variance de l'estimateur naturel est  $\frac{\lambda}{|A|}$ .

Preuve :

L'estimateur naturel de  $\lambda$  s'exprime comme :

$$\hat{\lambda} = \frac{N(A)}{|A|}$$

où

$$N(A) \sim \text{Poisson}(\lambda|A|).$$

Alors

$$\begin{aligned} \text{Var}(\widehat{\lambda}) &= \text{Var}\left(\frac{N(A)}{|A|}\right) \\ &= \frac{1}{|A|^2} \text{Var}(N(A)) \\ &= \frac{1}{|A|^2} \lambda|A| \\ &= \frac{\lambda}{|A|}. \end{aligned}$$

*Propriété 3* : Cet estimateur naturel forme aussi l'estimateur du maximum de vraisemblance.

L'estimateur du maximum de vraisemblance est la fonction  $\widehat{\theta}(x)$  qui maximise la fonction de vraisemblance, une fois que  $X = x$  est observé, où  $X$  est une variable aléatoire réelle de loi discrète ou continue, et  $\theta$  est le paramètre à estimer.  $(x_1, \dots, x_n)$  sont des observations indépendamment et identiquement distribuées de  $X$ .

La fonction de log-vraisemblance est  $l(\theta) = \log f_n(x | \theta)$ . Parce que  $\log$  est une fonction monotone et croissante, nous pouvons utiliser la dérivée de log-vraisemblance afin de trouver le maximum de  $f_n(x | \theta)$ .

Considérons maintenant le cas du processus ponctuel de Poisson homogène :

$$N(A) \sim \text{Poisson}(\lambda|A|).$$

On veut montrer que l'estimateur du maximum de vraisemblance de  $\lambda$  est  $\frac{N(A)}{|A|}$ .

Preuve :

Soit  $(x_1, \dots, x_n)$  des observation réalisée du processus ponctuel homogène de Poisson  $N$  sur la région bornée  $A$ , alors c'est la vraisemblance de  $\theta$ , qui est ici le paramètre  $\lambda|A|$ . Puisque l'aire  $|A|$  est connue,  $\lambda$  est le paramètre à estimer.

Puisque  $N(A)$  suit la loi de Poisson, alors

$$P(N(A) = n) = e^{-\lambda|A|} \frac{(\lambda|A|)^n}{n!}.$$

L'estimateur du maximum de vraisemblance est une méthode acceptée pour estimer l'intensité  $\lambda$  qui maximise la vraisemblance de l'ensemble de données.

Méthode :

Supposons qu'on a des observations  $(x_1, \dots, x_n)$ , la fonction de vraisemblance

s'écrit comme

$$\begin{aligned}
 L(\lambda|A; \pi = \{x_1, \dots, x_n\}) &= f(\pi = \{x_1, \dots, x_n\}; \lambda|A) \\
 &= \prod_{i=1}^n e^{-\lambda|A|} \frac{(\lambda|A|)^{x_i}}{x_i!} \\
 &= e^{-n\lambda|A|} \frac{(\lambda|A|)^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.
 \end{aligned}$$

Nous trouvons ensuite la fonction de log-vraisemblance.

$$\begin{aligned}
 \log L &= \log e^{-n\lambda|A|} \frac{(\lambda|A|)^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\
 &= \log e^{-n\lambda|A|} + \log(\lambda|A|)^{\sum_{i=1}^n x_i} - \log \prod_{i=1}^n x_i! \\
 &= -n\lambda|A| + \sum_{i=1}^n x_i \log \lambda|A| - \sum_{i=1}^n \log x_i!
 \end{aligned}$$

Nous calculons ensuite la dérivée de log-vraisemblance et nous mettons la dérivée égale à 0 :

$$\begin{aligned}
 \frac{\partial \log L(\lambda|A)}{\partial \lambda|A|} &= -n + \frac{\sum_{i=1}^n x_i}{\lambda|A|} \\
 &= 0.
 \end{aligned}$$

Alors

$$\begin{aligned}\frac{\sum_{i=1}^n x_i}{\lambda|A|} &= n \\ \Rightarrow \lambda|A| &= \frac{\sum_{i=1}^n x_i}{n} \\ &= \bar{x}.\end{aligned}$$

Nous avons donc l'estimateur du maximum de vraisemblance de  $\lambda$

$$\begin{aligned}\hat{\lambda} &= \frac{\bar{x}}{|A|} \\ &= \frac{N(A)}{|A|}.\end{aligned}$$

Donc l'estimateur naturel  $\hat{\lambda} = \frac{N(A)}{|A|}$  est aussi l'estimateur du maximum de vraisemblance de  $\lambda$ .

En plus des propriétés mentionnées ci-dessus, l'estimateur naturel  $\hat{\lambda}$  a aussi des propriétés asymptotiques.

*Résultat 1 :*

Nous avons supposé que les points sont observés dans une région bornée  $A$ , nommée la fenêtre d'observation. Si la fenêtre s'étend à  $R^d$ ,  $\hat{\lambda}$  satisfera au théorème limite centrale.

*Théorème limite central :*

Soit  $X_1, X_2, \dots$  une suite de variables aléatoires indépendantes de même loi, de moyenne  $\mu$  et de variance  $\sigma^2$ . Alors la variable  $Z_{\bar{x}} = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma}$  tend en loi vers une loi  $N(0; 1)$ .

Cela signifie que si  $N(A)$  est assez grand, alors

$$\widehat{\lambda}|A| \sim N(\lambda|A|; \lambda|A|).$$

Et alors

$$\frac{\widehat{\lambda}|A| - \lambda|A|}{|A|^{1/2}} \sim N(0; \lambda).$$

Alors

$$|A|^{1/2}(\widehat{\lambda} - \lambda) \sim N(0; \lambda).$$

*Résultat 2*

La transformation stabilisée de la variance a été observée (Voir Heinrich et Prokesova

(2010)) :

$$2|A|^{1/2}(\sqrt{\hat{\lambda}} - \sqrt{\lambda}) \sim N(0; 1).$$

L'estimateur naturel de l'intensité  $\lambda$  sera utilisé dans la partie quatre, nommée l'estimateur standard, et sera écrit comme *eststd*.

## 2.2 Estimateur Voronoï

Considérons  $\{p_1, \dots, p_n\}$  est une réalisation finie du processus ponctuel de Poisson homogène sur la région bornée  $S$ .  $N(S)$  est le nombre des points sur  $S$  et  $N(S) = n$ .

Considérons  $T = \{C_1, \dots, C_n\}$  une tessellation de Voronoï qui correspond au processus de Poisson homogène  $N$  et qui divise  $S$  en  $n$  cellules distinctes.

Chaque cellule  $C_i$  contient toutes les positions dans  $S$  qui sont plus proches du point  $p_i$  que de tout autre point dans la réalisation du processus de Poisson homogène  $N$ .

Alors

$$C_i = \{y \in S : \|y - p_i\| \leq \|y - p_j\|, \forall p_j \in N\}$$

où

$\| \cdot \|$  est la distance euclidienne.

La définition de l'estimateur Voronoï dans le cas général est donnée ci-dessous.

Pour toute position  $y \in C_i$ ,

$$\lambda^v(y) = \frac{1}{\mu(C_i)},$$

où  $\mu(C_i)$  est l'aire de la cellule  $C_i$ .

Dans le cas homogène, l'intensité  $\lambda$  est une constante. Plusieurs chercheurs, par exemple, Meijering (1953), Gilbert (1962), Hanson (1983) et Brakke (1986), Okabe (2000) ont montré que l'aire espérée de cellule typique du diagramme de Poisson Voronoï homogène dans  $R^2$  est l'inverse de l'intensité  $\lambda$ . C'est-à-dire

$$E(C_0) = \frac{1}{\lambda}$$

où  $C_0$  est une cellule typique du diagramme de Poisson-Voronoï homogène et  $\bar{C}$  est l'aire moyenne des cellules typiques.

Une cellule typique de Voronoï, selon la définition d'Okabe (2000), est un polytope aléatoire qui a la même distribution que des cellules sélectionnées aléatoirement dans le diagramme de Voronoï, où chaque cellule a la même chance d'être sélectionnée.

Il est naturel d'estimer  $C_0$  par  $\bar{C}$  l'aire moyenne des cellules de Voronoï.

Alors

$$\hat{\lambda} = \frac{1}{\bar{C}}$$

est nommé l'estimateur Voronoï de l'intensité  $\lambda$  du processus de Poisson homogène. Cela signifie que  $\lambda$  peut être estimé en évaluant l'aire moyenne des cellules dans la région bornée et observée de la tessellation de Voronoï générée par le processus ponctuel de Poisson homogène. Plus  $\lambda$  est grand, plus l'aire moyenne des cellules typiques du diagramme de Poisson-Voronoï homogène est petite.

$\bar{C}$  est la moyenne empirique des aires de cellules de la tessellation de Voronoï. Ici nous pouvons prendre toutes les cellules observées pour calculer la moyenne des aires. Alors soit  $n$  le nombre des cellules, nous avons

$$\begin{aligned} E(\hat{\lambda}) &= E\left(\frac{1}{\bar{C}}\right) \\ &= E\left(\frac{1}{\frac{1}{n} \sum_{i=1}^n |C_i|}\right) \\ &= E\left(\frac{n}{\sum_{i=1}^n |C_i|}\right) \end{aligned}$$

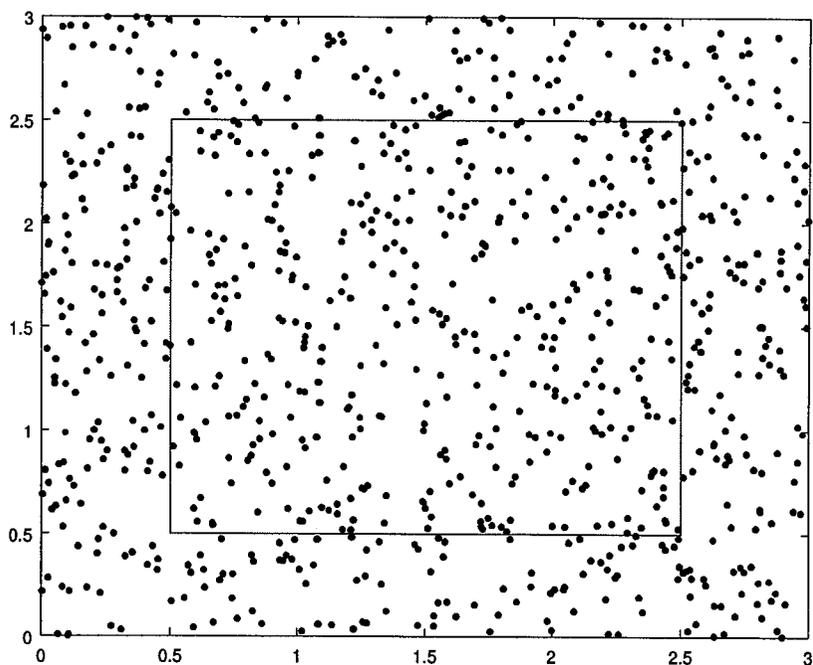
Puisqu'ici le numérateur et le dénominateur sont des estimateurs sans biais, l'estimateur Voronoï est donc un estimateur de rapport sans biais.

De façon standard, lors du calcul de l'aire moyenne, pour éviter les effets de bord, on prend toutes les cellules qui ne touchent pas le bord du domaine et on calcule  $\overline{C}$  sur ces cellules.

Nous utiliserons dans la partie quatre l'estimateur Voronoï, qui estime l'intensité en mesurant l'aire de toutes les cellules, et qui s'écrit comme *estvtot*.

Regardons maintenant les diagrammes suivants qui illustrent les deux estimateurs.

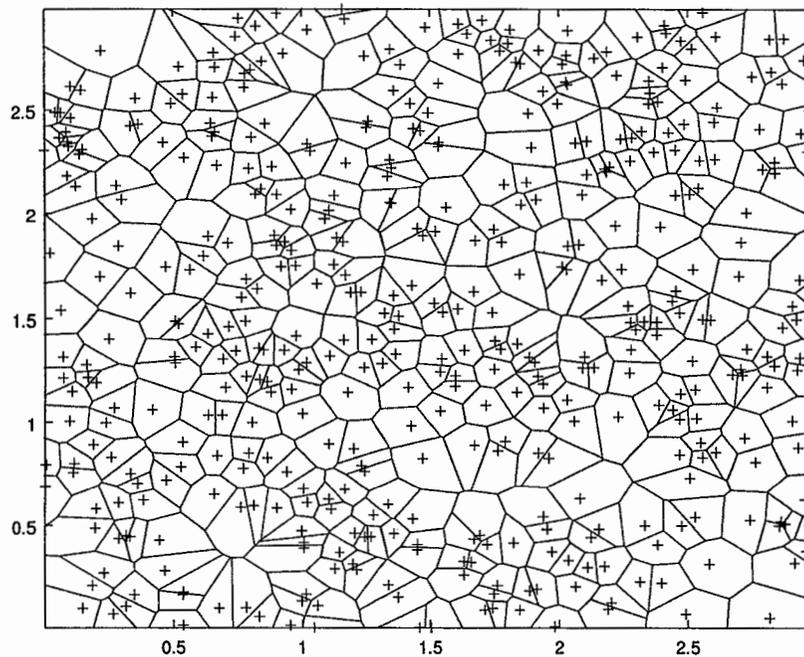
FIGURE 2.1 Points réalisés avec  $\lambda = 50$ ,  $|S| = 9$



Dans le chapitre un nous avons vu que  $N(A)$  ne dépend pas de la position de  $A$  mais de l'aire de  $A$ . Considérons la région carrée au milieu comme la fenêtre d'observation  $A$ , et nous avons  $|A| = 4$ . Nous pouvons donc compter le nombre des points dans la région bornée  $A$  et puis estimer l'intensité.

Regardons ensuite un diagramme de Poisson-Voronoi.

FIGURE 2.2 Diagramme de Poisson-Voronoi



Avec l'estimateur Voronoï, nous devons calculer l'aire totale de toutes les cellules dans le diagramme de Poisson-Voronoi borné et puis calculer l'aire moyenne des cellules. Le diagramme montre que pour la région bornée  $S : [0, 3]^2$  il y a des effets de bord pour les cellules proches du bord de  $S$ . Dans le présent travail, pour

éviter les effets de bord, nous prenons tous les points observés dans la fenêtre d'observation  $A$  (regardons le diagramme précédent), et nous calculons ensuite l'aire de toutes les cellules complètes qui contiennent les points observés (pas seulement la partie à l'intérieur de  $A$ ). Puisque nous laissons ici assez d'espace entre les bordures de  $S$  et  $A$ , les cellules considérées sont assez éloignées du bord de  $S$ , nous pouvons donc corriger les effets de bord afin d'estimer l'intensité.

Nous voulons comparer les deux estimateurs. Nous avons les mêmes points dans la région d'observation  $A$  pour les deux estimateurs. Nous avons aussi l'estimateur standard

$$\hat{\lambda} = \frac{n}{|A|}$$

et l'estimateur Voronoï

$$\hat{\lambda} = \frac{1}{\bar{C}} = \frac{1}{\frac{1}{n} \sum_{i=1}^n |C_i|}.$$

Puisque l'aire totale de toutes les cellules complètes est plus grande que  $|A|$ , nous avons donc la valeur estimée avec l'estimateur standard est plus grand que celle avec l'estimateur Voronoï.

Il y a aussi d'autres estimateurs de l'intensité  $\lambda$ . Et nous en citons seulement certains ici. Par exemple, Ord (1978) a proposé utilisé l'inverse de  $|C_i|$  pour calculer l'estimation de  $\lambda$ . Alors

$$\hat{\lambda} = \frac{1}{m} \sum_{i=1}^m |C_i|^{-1}$$

où  $m$  est la taille de l'échantillon, soit le nombre des cellules dans l'échantillon, et  $|C_i|$  l'aire de la cellule sélectionnée dans l'échantillon.

Krebs (1998) a proposé l'estimation par l'intervalle de confiance.

Coeurjolly (2017) a proposé un estimateur en base de médiane afin d'éviter l'influence des valeurs extrêmes, etc.

Dans le présent travail, nous utiliserons seulement l'estimateur standard et l'estimateur Voronoï pour illustrer la procédure d'établir un nouvel estimateur pour l'intensité avec la méthode présentée dans la partie suivante.

## CHAPITRE III

### APPROCHE DE COMBINAISON DES ESTIMATEURS

En 2016, Frédéric Lavancier et Paul Rochet ont proposé une approche de combinaison des estimateurs pour produire une meilleure estimation. Avant de commencer l'introduction de leur approche, nous donnons un aperçu des recherches auparavant pour une idée générale pour ce sujet.

Depuis longtemps des chercheurs essaient de trouver un moyen pour combiner plusieurs estimateurs d'une quantité inconnue  $\theta$  dans un modèle statistique afin de fournir une meilleure estimation. Ce type d'étude peut s'écrire comme le suivant.

Soit  $\theta$  une quantité inconnue d'un modèle statistique et supposons  $T_1, \dots, T_k$  une collection de  $k$  estimateurs du paramètre  $\theta$ . On veut combiner ces estimateurs afin d'obtenir une meilleure estimation.

Les premiers essais remontent au début du 19<sup>e</sup> siècle et ont été faits par Laplace (1812). Il a essayé de fournir une meilleure estimation pour le problème suivant :

$x_1, \dots, x_n$  sont  $n$  observations i.i.d. d'une distribution inconnue. Supposons que

la distribution est symétrique autour de paramètre  $\theta$  et on veut estimer  $\theta$ .

Il a combiné la moyenne  $T_1 = \bar{x}_n$  et la médiane  $T_2 = x_{(n/2)}$ , les deux estimateurs de la tendance centrale, et il a donné la combinaison  $\mu_1 \bar{x}_n + \mu_2 x_{(n/2)}$  avec  $\mu_1 + \mu_2 = 1$ .

Il a démontré que cette combinaison suit asymptotiquement la loi gaussienne. En plus de cela, il a aussi donné la conclusion suivante :

sous la distribution gaussienne, la meilleure combinaison est donnée avec  $\mu_1 = 1$  et  $\mu_2 = 0$ . Sinon la meilleure combinaison dépend de la distribution.

Après Laplace, les chercheurs ont proposé des moyens pour combiner une série d'estimateurs dans le but de la prévision de la série temporelle. Parmi eux Hansen (2007) a proposé un estimateur avec la moyenne des moindres carrés. C'est une méthode similaire de l'oracle, qui sera présentée dans cette partie. Et c'est une méthode de combinaison linéaire qui minimise l'erreur de moyenne quadratique.

L'idée mentionnée ci-dessus était utilisée dans le domaine de prévision. Frédéric Lavancier et Paul Rochet (2016) l'ont utilisée et étendue dans le domaine d'estimation classique. Cette procédure peut être utilisée pour les modèles non-paramétriques, les modèles paramétriques et semi-paramétriques et l'estimation de sa variance peut être réalisée par la simulation de Monte-Carlo.

Lavancier et Rochet ont aussi inclus le cas où plusieurs paramètres doivent être estimés à la fois.

Dans cette partie une introduction est donnée pour la méthode de Lavancier et Rochet et on considère la méthode d'estimation pour un paramètre car dans notre modèle il y a seulement un paramètre à estimer.

### 3.1 Méthode d'estimation pour un paramètre

Soit  $T = (T_1, \dots, T_k)^T$  est une collection d'estimateurs d'un paramètre réel. Une combinaison naturelle, soit la combinaison linéaire, est donnée ci-dessous :

$$\hat{\theta}_\mu = \mu^T T,$$

où  $\mu \in M, M \subset R^k$ .

Pour la combinaison linéaire des estimateurs, on la compare souvent avec l'*oracle* pour évaluer sa performance. L'*oracle* est défini suivant :

l'*oracle* est la combinaison linéaire  $\hat{\theta}_\mu$  qui minimise l'erreur quadratique  $E(\hat{\theta}_\mu - \theta)^2$  avec la contrainte  $\sum_{i=1}^k \mu_i = 1$ . Sous cette condition, l'*oracle* s'exprime avec l'erreur de moyenne quadratique :  $\Sigma$  de  $T$ , et peut s'écrire comme  $\hat{\theta}^*$ , qui est la meilleure combinaison linéaire de  $\hat{\theta}_\mu$ , avec l'expression suivante :

*Oracle* :

$$\hat{\theta}^* = \mu^{*T} T = \frac{\mathbf{1}^T \Sigma^{-1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} T$$

où

$$\begin{aligned}\mu^* &= \arg \min_{\mu \in \mathcal{M}} E(\mu^T T - \theta)^2 \\ &= \arg \min_{\mu \in \mathcal{M}} \mu^T \Sigma \mu \\ &= \frac{\Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}\end{aligned}$$

qui forme les poids optimaux de l'estimation.

Dans l'application,  $\mu^*$  est inconnu donc il doit être estimé par  $\hat{\mu}$ . Ceci revient à estimer  $\Sigma = E(T - \theta \mathbf{1})(T - \theta \mathbf{1})^T$ . L'estimation de  $\Sigma$  peut être réalisée par le plug-in ou la méthode de Monte-Carlo. Alors

$$\begin{aligned}\hat{\theta} &= \hat{\mu}^T T \\ &= \frac{\mathbf{1}^T \hat{\Sigma}^{-1}}{\mathbf{1}^T \hat{\Sigma}^{-1} \mathbf{1}} T,\end{aligned}$$

où

$$\hat{\mu} = \arg \min_{\mu \in \mathcal{M}} \mu^T \hat{\Sigma} \mu,$$

avec  $\mu^T \mathbf{1} = 1$ .

Deux situations sont considérées ici pour estimer  $\Sigma$ .

*Situation 1* : modèle paramétrique

Dans le modèle paramétrique,  $T$  ne dépend que de  $\theta$ . C'est-à-dire  $\Sigma = \Sigma(\theta)$ .

Si  $\Sigma(\theta)$  est connue explicitement, alors  $\widehat{\Sigma} = \Sigma(\widehat{\theta}_0)$ , où  $\widehat{\theta}_0$  est un certain estimateur de  $\theta$  ou simplement la moyenne des estimateurs, ou encore mieux la version agrégée.

Si  $\Sigma(\theta)$  n'est pas connue, alors l'estimation sera réalisée par bootstrap paramétrique.

Dans ces deux cas, l'estimation ne dépend que de  $T$  mais non des données originales pour réaliser  $T$ .

*Situation 2* : modèle non-paramétrique

Dans le modèle non-paramétrique, l'estimation de  $\Sigma$  est réalisée par le bootstrap standard non-paramétrique.

Bien que la génération de cette méthode pour estimer plusieurs paramètres soit aussi une contribution importante de la recherche de Lavancier et Rochet, nous ne mettrons pas l'accent sur ce sujet dans le présent travail.

### 3.2 Performance de la méthode de Lavancier et Rochet

L'estimation de  $\Sigma$  est un facteur principal pour la performance de l'estimation. L'estimation de  $\Sigma$  peut être réalisée avec les mêmes données que celui pour produire les estimateurs initiaux dans  $T_i$ . Et dans le modèle complet, l'ensemble

des données initiales n'est pas nécessaire pour calculer  $\hat{\theta}$ , mais seulement les estimateurs  $T_1, \dots, T_k$  qui sont nécessaires.

Lavancier et Rochet ont démontré que  $\hat{\Sigma}$  est une estimation parfaite pour  $\Sigma$  si  $|tr(\mu^T \hat{\Sigma} \mu) - tr(\mu^T \Sigma \mu)|$  est petit pour  $\mu \in M$ .

Ils ont montré que  $\hat{\theta}$  est dans certains sens, asymptotiquement équivalent à  $\hat{\theta}^*$  et approximativement fonctionne bien pour modérer la taille de données.

En plus de cela, une fois  $\hat{\theta}$  obtenu, on obtient immédiatement un intervalle de confiance asymptotique, qui est de la longueur minimale parmi tous les intervalles de confiance possibles obtenu avec la combinaison linéaire de  $T$ .

Dans la partie suivante, nous construisons un nouvel estimateur pour l'intensité selon la méthode de Lavancier et Rochet.

## CHAPITRE IV

### NOUVEL ESTIMATEUR AGRÉGÉ

Selon la méthode générale de combiner des estimateurs proposée par Lavancier et Rochet, nous savons que nous pouvons établir un nouvel estimateur sur la base des estimateurs initiaux de  $\lambda$ .

Pour estimer l'intensité d'un processus de Poisson homogène, l'intensité  $\lambda$  est la quantité inconnue à estimer, et est le  $\theta$  dans la méthode de Lavancier et de Rochet. Pour faciliter la compréhension et éviter la confusion, nous gardons dans cette partie les notations de Lavancier et Rochet dans la procédure d'établissement du nouvel estimateur. C'est-à-dire :

- $\theta$  est l'intensité du processus ponctuel homogène de Poisson ( $\lambda$  dans les parties précédentes);
- $T_1$  est l'estimateur Voronoï de l'intensité ( $\widehat{\lambda}^{tot}$  dans la section précédente);
- $T_2$  est l'estimateur standard de l'intensité ( $\widehat{\lambda}^{std}$  dans les parties précédentes);
- $\mu$  est le vecteur de poids dans la combinaison linéaire des estimateurs.

#### 4.1 Etablissement du nouvel estimateur

Nous voulons maintenant construire un estimateur qui donnera une meilleure estimation pour l'intensité.

Soit

$$T = (T_1, T_2)^T$$

une collection d'estimateurs du paramètre  $\theta$  et

$$T_1 = \frac{1}{\bar{C}},$$

$$T_2 = \frac{N(A)}{|A|},$$

où

- $A$  est la région bornée observée ;
- $N(A)$  est le nombre de points générés par le processus ponctuel homogène de Poisson dans la région bornée  $A$  ;
- $|A|$  est l'aire de la région  $A$  ;
- $\bar{C}$  est l'aire moyenne de toutes les cellules complètes qui contiennent les points observés dans la région bornée  $A$ .

Avec ces deux estimateurs initiaux, nous construisons une combinaison linéaire des estimateurs

$$\hat{\theta} = \hat{\mu}^T T$$

$$= \frac{\mathbf{1}^T \hat{\Sigma}^{-1}}{\mathbf{1}^T \hat{\Sigma}^{-1} \mathbf{1}} T$$

où  $\Sigma = E[(T - \theta \mathbf{1})(T - \theta \mathbf{1})^T]$ ,  $\mu^T \mathbf{1} = 1$  et  $\mathbf{1} = (1, 1)^T$ .

La matrice  $\Sigma$  est estimée par  $\hat{\Sigma}$ . Puisque dans le modèle paramétrique,  $T$  ne dépend que de  $\theta$ , alors

$$\Sigma = \Sigma(\theta).$$

Et puisque  $\Sigma(\theta)$  n'est pas connue explicitement, l'estimation sera réalisée par bootstrap paramétrique.

Une fois nous avons obtenu  $\hat{\Sigma}$ , nous pouvons avoir  $\hat{\theta}$ , qui est l'estimateur agrégé et nous pouvons aussi écrire le vecteur de poids  $\hat{\mu}$ .

Dans la section suivante, pour faire la simulation, le nouvel estimateur est écrit comme *thetahat*.

## 4.2 Simulation

Dans cette section, nous voudrions estimer l'intensité  $\lambda$  avec les estimateurs suivants :

- l'estimateur standard ;
- l'estimateur Voronoï ;
- le nouvel estimateur agrégé.

Les simulations sont faites pour étudier les caractéristiques des estimateurs et comparer leur performance.

Pour commencer l'estimation, nous voulons d'abord générer le diagramme de

Poisson-Voronoi.

Pour le diagramme de Poisson-Voronoi, le nombre des points de générateurs est généré par le processus ponctuel de Poisson homogène. Sous cette condition, les points sont générés selon la loi uniforme. Alors, l'algorithme de la simulation du diagramme de Poisson-Voronoi est le suivant :

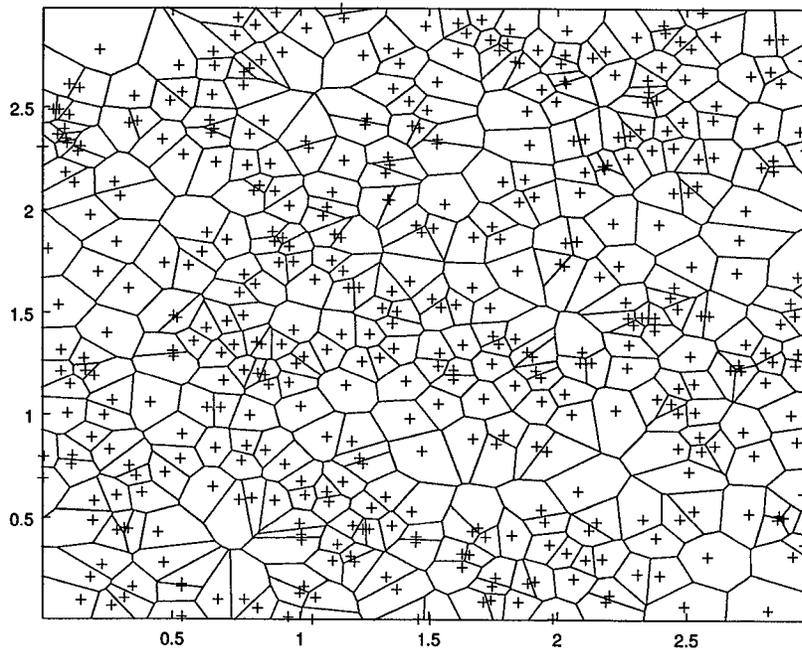
- générer le nombre de points selon le processus de Poisson homogène.
- générer les points aléatoires de loi uniforme.
- générer le diagramme de Poisson-Voronoi.
- générer la tessellation Delaunay si nécessaire.

En Matlab, nous devons d'abord générer le nombre de points avec le processus de Poisson homogène. Nous prenons l'intensité  $\lambda = 50$  par unité d'aire dans la simulation pour générer le nombre des points. En sachant que pour le processus ponctuel homogène de Poisson le nombre de points  $N(S)$  ne dépend que de  $|S|$ , nous prenons  $|S| = 9$ . Alors pour la simulation de  $\lambda = 50$ ,

- $\lambda|S| = 450$  si  $|S| = 9$

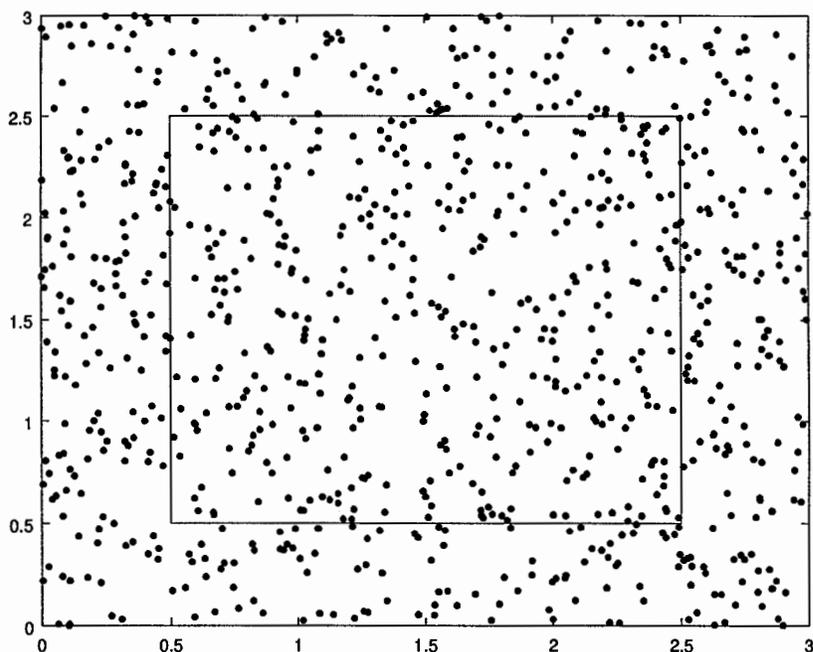
Nous générons d'abord  $n$  le nombre de points avec  $\lambda = 50$  et  $S = 9$ .

Après avoir obtenu le nombre de points  $n$ , nous devons maintenant générer le  $n$  points de la loi uniforme et nous obtenons le diagramme de Poisson-Voronoi suivant pour ces  $n$  points de loi uniforme :



Maintenant nous devons faire la simulation de l'estimation avec l'estimateur standard et l'estimateur Voronoï. Pour la simulation de cette partie, nous prenons 1000 répétitions.

Nous établissons une région bornée  $A$  avec  $|A| = 4$ . Cette région se trouve au milieu de  $S$ , et sera notre fenêtre d'observation.



Pour l'estimateur standard nous calculons le nombre des points qui se trouvent dans la région  $A$ . Et après nous pouvons estimer l'intensité.

Pour l'estimateur Voronoï, nous calculons l'aire de toutes les cellules de Voronoï dont les graines se trouvent dans la région  $A$ .

Une cellule typique de Poisson-Voronoï est un polygone sélectionné aléatoirement parmi les polygones du diagramme de Poisson-Voronoï dans la région bornée. Alors pour estimer l'intensité nous devons avoir l'aire des cellules de Poisson-Voronoï sélectionnées aléatoirement. Cependant il existe un inconvénient : il y a toujours l'effet de frontière introduit par les polygones qui sont proches de frontière de la région. Donc le nombre des polygones doit obligatoirement être suffisamment

grand pour éviter le biais.

Nous prenons ici les mêmes points que ceux pour l'estimation avec l'estimateur standard. Puisque ces points se trouvent dans la région  $A$ , nous laissons assez d'espace entre les bordures de  $S$  et  $A$ . C'est-à-dire pour les points qui se trouvent près de bord, nous prenons l'aire des cellules complètes au lieu de l'aire de la partie à l'intérieur de  $A$ . Nous pouvons donc éviter les effets de bord.

Pour calculer l'aire des cellules dans le diagramme de Poisson-Voronoi, nous commençons par trouver les coordonnées des sommets de chaque polygone qui forme le diagramme de Poisson-Voronoi. Ensuite nous calculons l'aire de chaque polygone avec logiciel. Nous prenons toutes les cellules et nous calculons la moyennes des aires de cellules  $\bar{C}$ . Et après nous estimons l'intensité par  $\hat{\lambda} = \frac{1}{\bar{C}}$ .

Ensuite nous comparons la performance de l'estimateur standard avec celle de l'estimateur Voronoi.

Nous faisons d'abord une simulation de l'estimateur standard et l'estimateur Voronoi, avec 1000 répétitions. Le tableau suivant montre le résultat.

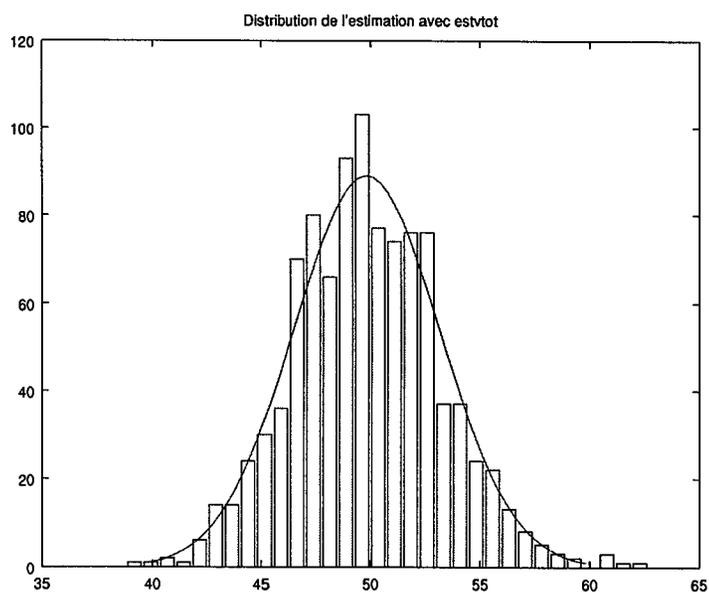
TABLE 4.1 Estimation avec eststd et estvtot

	estvtot	eststd
mean	49.812	49.791
std	3.3387	3.5074
MSE	11.136	12.29

Le tableau 1 montre que les deux estimateurs fonctionnent très bien pour l'estimation de  $\lambda$ . Après 1000 répétitions, l'estimateur Voronoi a la moyenne de 49.812 et

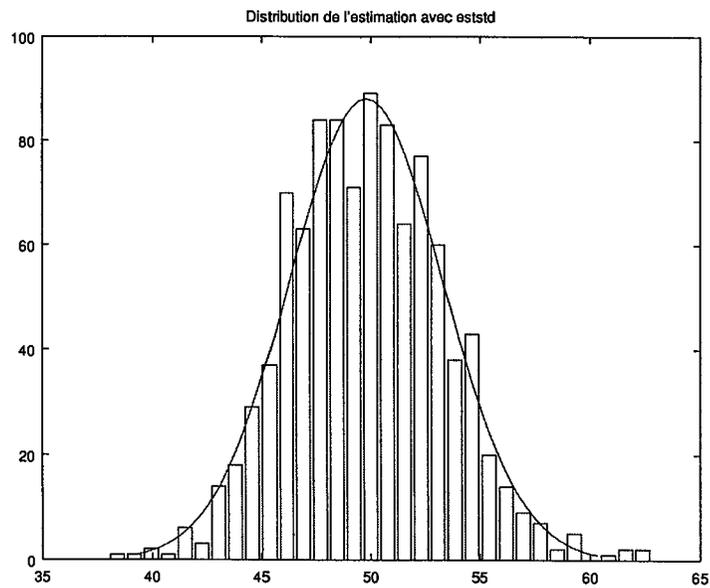
l'estimateur standard a la moyenne de 49.791. La moyenne de l'estimateur Voronoï est plus proche de la vraie valeur de  $\lambda$ . En comparant les écarts-types, nous voyons que les estimations de l'estimateur standard sont plus dispersées que celles de l'estimateur Voronoï. Quant à l'erreur moyenne quadratique, l'estimateur Voronoï a la valeur plus petite que l'estimateur standard.

FIGURE 4.1 Histogramme de la distribution des estimations avec estvtot



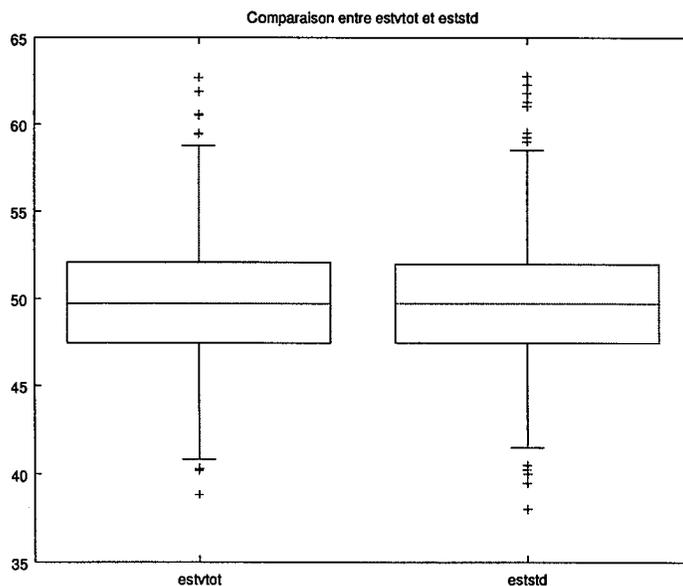
Selon l'histogramme, nous voyons que la distribution des estimations de l'estimateur Voronoï est presque symétrique autour de 50. La dispersion de l'estimateur Voronoï est plus petite que l'estimateur standard.

FIGURE 4.2 Histogramme de la distribution des estimations avec eststd



L'historgramme de l'estimateur standard montre que la distribution est presque normale et symétrique autour de 50. La dispersion est un peu plus large que celle de l'estimateur Voronoï.

FIGURE 4.3 Comparaison eststd vs estvtot



La boîte à moustaches montre que les deux estimateurs ont presque la même médiane, qui est un peu plus basse que 50. Et l'estimateur standard a plus de valeurs extrêmes que l'estimateur Voronoï. C'est une raison pour laquelle l'estimateur standard a la dispersion un peu plus large que l'estimateur Voronoï.

Donc nous pouvons dire que les deux estimateurs se manifestent tous très bien. Cependant selon les résultats de comparaison nous pensons que l'estimateur Voronoï est meilleur que l'estimateur standard.

Nous prenons l'estimateur Voronoï comme la base de la comparaison avec d'autres estimateurs. Et nous allons combiner l'estimateur standard avec l'estimateur Voronoï. Nous pouvons construire le nouvel estimateur  $\hat{\theta}$  selon la méthode de Lavancier et Rochet. Pour le faire, nous devons estimer  $\Sigma$ .

Voici la procédure pour l'estimation de  $\Sigma$  pour chaque itération  $i$ , où  $i = 1, \dots, 1000$ .

- Pour chaque  $i$ , nous devons estimer  $\Sigma$  par  $\widehat{\Sigma}_i$ .
- Nous simulons  $B = 1000$  processus de Poisson  $Y_{i,1}, \dots, Y_{i,B}$  sur  $S$  d'intensité  $\widehat{\lambda}_i^{std}$ .
- Ensuite pour chaque  $Y_{i,j}$  nous estimons  $\lambda$  avec les deux façons et nous obtenons alors  $\widehat{\lambda}_{i,j}^{std}$  et  $\widehat{\lambda}_{i,j}^{vtot}$  pour  $j = 1, \dots, B$ .
- Et puis  $\widehat{\Sigma}_i$  est la matrice de covariance empirique des  $(\widehat{\lambda}_{i,j}^{std}, \widehat{\lambda}_{i,j}^{vtot})^T$  pour  $j = 1, \dots, B$ .
- Enfin nous pouvons estimer l'intensité.

Avec 1000 répétitions, le tableau 2 montre le résultat de la simulation.

TABLE 4.2 Estimation avec estvtot, eststd et thetahat

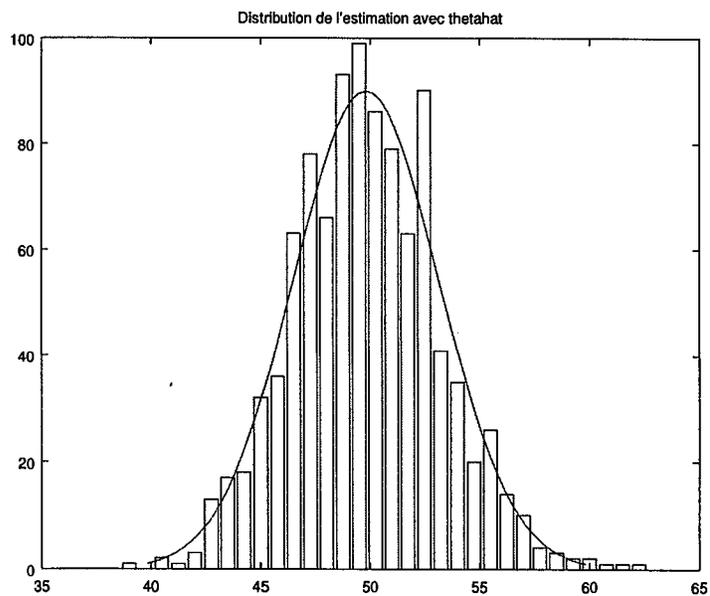
	estvtot	eststd	thetahat
mean	49.812	49.791	49.816
std	3.3387	3.5074	3.3322
MSE	11.136	12.29	11.093

Le résultat du tableau 2 montre que l'estimateur standard a la moyenne la plus éloignée de la valeur réelle de  $\lambda$  parmi les trois estimateurs. Son écart-type est aussi plus grand que les deux autres estimateurs, ceci signifie que les estimations avec cet estimateur sont plus dispersées que d'autres estimateurs. L'erreur quadratique moyenne de l'estimateur standard est plus élevée que les deux autres. Donc sa performance est moins efficace que l'estimateur Voronoï et l'estimateur agrégé.

Le nouvel estimateur agrégé donne une moyenne plus proche de la valeur réelle de  $\lambda$  que les deux estimateurs initiaux. Par conséquent, sa moyenne est meilleure que l'estimateur standard et que l'estimateur Voronoï. Son écart-type est plus petit que celui des deux estimateurs initiaux, et son erreur quadratique moyenne est un peu plus basse que celle de l'estimateur Voronoï, et beaucoup plus basse que celle de l'estimateur standard. Il minimise donc l'erreur quadratique moyenne.

Maintenant nous voudrions regarder l'histogramme pour vérifier la distribution des estimations.

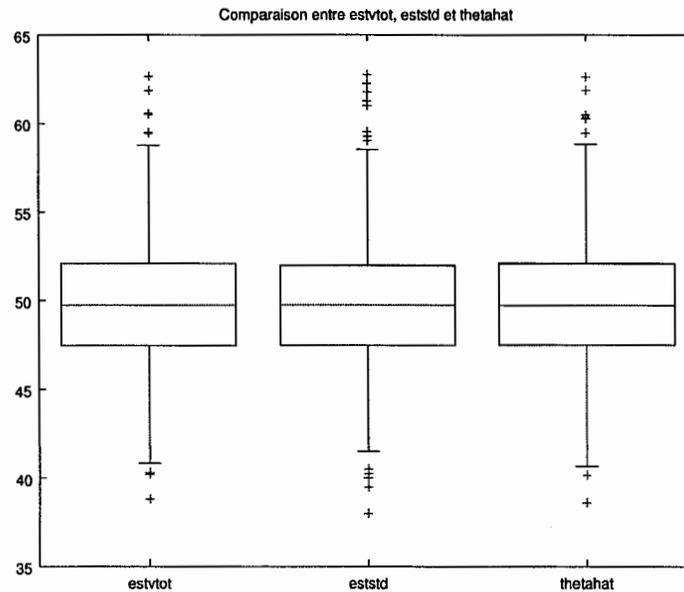
FIGURE 4.4 Histogramme de la distribution des estimations avec thetihat



Selon l'histogramme de la distribution de l'estimateur  $\hat{\theta}$ , nous voyons que la distribution est presque normale et son écart-type est plus petit que l'estimateur standard.

Regardons maintenant la boîte à moustaches pour les trois estimateurs.

FIGURE 4.5 Comparaison entre estvtot, eststd et thetahat



La boîte à moustaches montre clairement que l'estimateur standard a plus de valeurs extrêmes que l'estimateur Voronoï et l'estimateur agrégé, qui augmentent la variance de l'estimation.

Le nouvel estimateur agrégé a la distribution similaire de l'estimateur Voronoï. Sa moyenne est encore plus proche de la vraie valeur de  $\lambda$  en comparant avec les deux estimateurs initiaux. Avec la méthode de Lavancier, ce nouvel estimateur diminue efficacement la variance et l'erreur quadratique moyenne. Il est donc un meilleur estimateur.

## CONCLUSION

Nous utilisons les estimateurs pour évaluer les paramètres inconnus. Les estimateurs ne donnent pas toujours le résultat satisfaisant. Plusieurs critères peuvent qualifier la qualité d'un estimateur. Parmi les critères nous citons le biais, l'efficacité, l'erreur quadratique moyenne, etc. Dans le monde idéal, nous voulons un estimateur sans biais ou asymptotiquement sans biais, avec la variance la plus faible possible et l'erreur quadratique moyenne minimale. Cependant ce n'est pas toujours facile de trouver un estimateur qui satisfait à tous les critères. Nous cherchons donc avec tous les efforts un meilleur estimateur pour s'approcher au monde idéal.

Le diagramme de Voronoï et le diagramme de Poisson-Voronoï sont partout dans la vie quotidienne et sont très important dans de divers domaines scientifiques et de recherches. Notamment le diagramme de Poisson-Voronoï est souvent utilisé comme un modèle de base avec lequel les gens font des recherches ou des comparaisons afin d'étudier les caractéristiques d'autres modèles. Ils jouent donc un rôle important dans plusieurs domaines.

Nous avons mis l'accent sur l'estimation de l'intensité d'un processus de Poisson homogène. Les estimateurs déjà existant incluent l'estimateur standard, l'estimateur Voronoï, etc. Une approche pour combinaison des estimateurs a été introduite dans ce travail et nous avons construit selon cette approche un nouvel estimateur pour l'intensité du processus de Poisson homogène. Ce nouvel estimateur est une combinaison linéaire des estimateurs initiaux et se manifeste très bien pour estimer

l'intensité. Le résultat de simulation a montré que ce nouvel estimateur est plus proche que la vraie valeur de l'intensité que les deux autres estimateurs initiaux et a la variance plus faible et l'erreur quadratique moyenne minimale. Donc nous avons trouvé un meilleur estimateur.

Le présent travail est une application de la méthode de combinaison des estimateurs et il a ses limites. Nous avons estimé l'intensité du processus ponctuel homogène de Poisson. Le cas inhomogène n'a pas été traité dans ce travail. Deuxièmement, nous avons traité le processus ponctuel de Poisson. Le cas des autres processus ponctuels n'a pas été traité dans ce travail. Ces limites sont à étudier dans le futur.

## ANNEXE A

### CODE

Code pour générer  $n$  points sur  $S$  et générer le diagramme de Voronoï

```
1 clear all;
2 figure();
3 lambda=50;
4 S=9;
5 lambdaS=lambda.*S;
6 npoints=poissrnd(lambdaS);
7 npoints
8 pproc=rand(npoints,2).*sqrt(S);
9 x=pproc(:,1);
10 y=pproc(:,2);
11 plot(x,y, '.')
12 voronoi(x, y) % avoir le diagramme de Poisson voronoi
13 xlim([min(x) max(x)])
14 ylim([min(y) max(y)])
```

## Code pour simuler les estimateurs

```

1 pkg load statistics
2 clear all;
3 N=1000;
4 lambda=50;
5 S=9;
6 A=4;
7 I=ones(2,1);
8 lambdaS=lambda.*S;
9 npoints=zeros(N,1);
10 numin=zeros(N,1);
11 eststd=zeros(N,1);
12 estvtot=zeros(N,1);
13 xv=[0.5, 2.5, 2.5, 0.5, 0.5];
14 yv=[0.5, 0.5, 2.5, 2.5, 0.5];
15 for i=1:N
16 npoints(i,1)=poissrnd(lambdaS);
17 pproc=rand(npoints(i,1),2).*sqrt(S);
18 x=pproc(:,1);
19 y=pproc(:,2);
20 [in,on] = inpolygon(x, y, xv, yv);
21 numin(i,1)=numel(x(in));
22 [C,F]=voronoin(pproc);% retourner F qui contient les indices ...
    des sommets pour construire un polygone
23 area=zeros(npoints(i,1),1);
24 for j=1:npoints(i,1)
25 area(j,1)=polyarea(C(F{j},1),C(F{j},2));
26 end
27 areain=[in,area];
28 row_index=areain(:,1)==1;
29 cellarea=areain(row_index,2);
30 estvtot(i,1)=1/mean(cellarea);

```

```
31 eststd(i,1)=numin(i,1)/A;
32 end
33
34 mean(eststd)
35 mean(estvtot)
36
37 T=[eststd, estvtot]';
38 thetahat=zeros(N,1);
39
40 B=1000;
41 for i=1:N
42 lambdaB=eststd(i,1);
43 S=9;
44 A=4;
45 lambdaBS=lambdaB.*S;
46 npointsB=zeros(B,1);
47 numinB=zeros(B,1);
48 eststdB=zeros(B,1);
49 estvtotB=zeros(B,1);
50
51 for k=1:B
52 npointsB(k,1)=poissrnd(lambdaBS);
53 pproc=rand(npointsB(k,1),2).*sqrt(S);
54 x=pproc(:,1);
55 y=pproc(:,2);
56 [in,on] = inpolygon(x, y, xv, yv);
57 numinB(k,1)=numel(x(in));
58 [C,F]=voronoin(pproc);% retourner F qui contient les indices ...
    des sommets pour construire un polygone
59 area=zeros(npointsB(k,1),1);
60 for j=1:npointsB(k,1)
61 area(j,1)=polyarea(C(F{j},1),C(F{j},2));
62 end
```

```
63 areain=[in,area];
64 row_index=areain(:,1)==1;
65 cellarea=areain(row_index,2);
66 estvtotB(k,1)=1/mean(cellarea);
67 eststdB(k,1)=numinB(k,1)/A;
68 end
69
70 est=[eststdB, estvtotB];
71 E=est-repmat(eststd(i,1),B,1);
72 sigmahati=(E'*E)./B;
73
74 lambdahattr=I'*inv(sigmahati)/(I'*inv(sigmahati)*I);
75 thetahat(i,1)=lambdahattr*T(:,i);
76 end
```

Code pour calculer la moyenne, l'écart-type, l'erreur de moyenne quadratique

```
1 mean(eststd)
2 mean(estvtot)
3 mean(thetahat)
4 std(eststd)
5 std(estvtot)
6 std(thetahat)
7 mean(abs(eststd-mean(eststd)).^2)
8 mean(abs(estvtot-mean(estvtot)).^2)
9 mean(abs(thetahat-mean(thetahat)).^2)
```

## BIBLIOGRAPHIE

- Babu, G. J. & Feigelson, E. D. (1996). *Astrostatistics*. London, UK.
- Baccelli, F. & Błaszczyszyn, B. (2001). On a coverage process ranging from the boolean model to the poisson–voronoi tessellation with applications to wireless communications. *Advances in Applied Probability*, 33(2), 293–323.
- Boots, B. N. (1975). Patterns of urban settlements revisited. *The Professional Geographer*, 27(4), 426–431.
- Brakke, K. (1986). Plane voronoi tessellation second order statistics. *American Mathematical Society*, 1, 235–236.
- Chakravarti, A. & Archibold, O. (1993). Patterns of diurnal variation of growing season precipitation on the canadian prairies : a harmonic analysis. *The Canadian Geographer/Le Géographe canadien*, 37(1), 16–28.
- Clausel, M., Coeurjolly, J.-F., Lelong, J. et al. (2016). Stein estimation of the intensity of a spatial homogeneous poisson point process. *The Annals of Applied Probability*, 26(3), 1495–1534.
- Coeurjolly, J.-F. (2017). Median-based estimation of the intensity of a spatial point process. *Annals of the Institute of Statistical Mathematics*, 69(2), 303–331.
- Diggle, P. (2003). *Statistical analysis of spatial point patterns* : edward arnold.
- Drouffe, J. & Itzykson, C. (1984). Random geometry and the statistics of two-dimensional cells. *Nuclear Physics B*, 235(1), 45–53.
- Ferenc, J.-S. & Nédá, Z. (2007). On the size distribution of poisson voronoi cells. *Physica A : Statistical Mechanics and its Applications*, 385(2), 518–526.
- Gavrikov, V. & Stoyan, D. (1995). The use of marked point processes in ecological and environmental forest studies. *Environmental and ecological statistics*, 2(4), 331–344.

- Getis, A. & Boots, B. (1975). Models of spatial processes : Point, line, and area patterns in geography.
- Gilbert, E. (1962). Random subdivisions of space into crystals. *The Annals of mathematical statistics*, 33(3), 958–972.
- Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476), 1502–1512.
- Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, 75(4), 1175–1189.
- Hanson, H. G. (1983). Voronoi cell properties from simulated and real random spheres and points. *Journal of statistical physics*, 30(3), 591–605.
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons.
- Keeler, H. P. (2016). Notes on the poisson point process. *Weierstrass Inst., Berlin, Germany, Tech. Rep.*
- Krebs, C. J. et al. (1989). *Ecological methodology*. Technical report, Harper & Row New York.
- Kumar, S. & Singh, R. N. (1995). Thermal conductivity of polycrystalline materials. *Journal of the American Ceramic Society*, 78(3), 728–736.
- Lavancier, F. & Rochet, P. (2016). A general procedure to combine estimators. *Computational Statistics & Data Analysis*, 94, 175–192.
- Ludwig, J. A. & Reynolds, J. F. (1988). *Statistical ecology : a primer in methods and computing*. John Wiley & Sons, New York.
- Meijering, J. (1953). Interface area, edge length, and number of vertices in crystal aggregates with random nucleation. *Philips Res. Rep.*, 8.
- Møller, J. & Stoyan, D. (2007). *Stochastic geometry and random tessellations*. Technical report, Department of Mathematical Sciences, Aalborg University.
- Møller, J. & Waagepetersen, R. (2004). Statistics and simulations of spatial point processes. *World Sci, Singapore*.
- Okabe, A., Boots, B., Sugihara, K. & Chiu, S. (2000). *Spatial tessellations : concepts and applications of Voronoi diagrams*. Wiley.
- Ord, J. (1978). How many trees in a forest. *Mathematical Scientist*, 3, 23–33.

Serra, J. (1982). *Image analysis and mathematical morphology*. Academic press.

Tanemura, M. (2003). Statistical distributions of poisson voronoi cells in two and three dimensions. *FORMA-TOKYO-*, 18(4), 221–247.

Van de Weygaert, R. (1994). Fragmenting the universe. 3 : The constructions and statistics of 3-d voronoi tessellations. *Astronomy and Astrophysics*, 283, 361–406.

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die reine und angewandte Mathematik*, 133, 97–178.