

Résumé de BaddEtAl16

DUGUE Clément

31 mai 2017

Table des matières

1	Introduction	3
1.1	Répartition des Points	3
1.1.1	Les Points	3
1.1.2	Les points de différents types	3
1.1.3	Répartition de point marqués	4
1.1.4	Covariables	4
1.1.5	Différentes dimensions spatiales	4
1.1.6	Reproduction de répartitions	4
1.2	Méthodes statistiques pour la répartition de point	5
1.2.1	Statistiques sommaires	5
1.2.2	Modèle statistique et inférence	5
1.2.3	Validation	5
2	Correlation (chap. 7)	6
2.1	Introduction	6
2.2	La fonction k	6
2.2.1	La fonction K empirique	6
2.2.2	La vraie fonction K d'un processus de point	7
2.2.3	Utilisation de la fonction K empirique	7
2.3	Correction des bords pour la fonction K	7
2.3.1	Estimation sans les effets de bords	7
2.3.2	Effets de bords	7
2.3.3	Correction des bords	7
2.3.4	Correction isotrope	7
2.3.5	Correction par translation	7
2.3.6	Discussion	7
2.4	Erreurs standard et interval de confiance	7
2.4.1	Variance sous CSR	7
2.4.2	Démarrage du system bloqué	7
2.4.3	Interval de confiance du syteme de démarrage de Loh	7
2.5	Tester si une répartition est complètement aléatoire	7
2.5.1	Enveloppes de point	7
2.5.2	Enveloppes globale	7
2.5.3	Tests non-graphiques	7

3	Espacement (chap. 8)	8
3.1	Introduction	8
3.2	Fonction d'espace vide F	8
3.2.1	Définitions pour un processus de point stationnaire	8
3.2.2	Valeurs pour un aléatoire complet	9
3.2.3	Interprétation de F	9
3.3	Fonction de plus proche voisins G	10
3.3.1	Définitions pour un processus de point stationnaire	10
3.3.2	Valeurs pour un aléatoire complet	10
3.3.3	Interprétation de G	10
3.4	Intervars de confiance et enveloppes de simulation	11
3.5	Hasard de l'espace vide (désavantages)	11
3.6	Fonction J	11
3.7	Théorie pour la correction des bords	11
3.7.1	Estimation de F et G sans effets de bord	11
3.7.2	Effets de bords pour F et G	11
3.7.3	Correction des bords	11
3.7.4	Correction de Kaplan-Meier	11
3.7.5	Correction de Hanisch-Chiu-Stoyan	11
3.7.6	Comparaison des méthodes	11

Chapitre 1

Introduction

1.1 Répartition des Points

1.1.1 Les Points

Une représentation spatiale de point est un ensemble de données fournissant une vue de la localisation d'événements ou de chose dans l'espace, son analyse statistique est utile pour de nombreux domaines. Il est important de pouvoir identifier des tendances dans la densité des points afin de pouvoir interpréter les représentations spatiales.

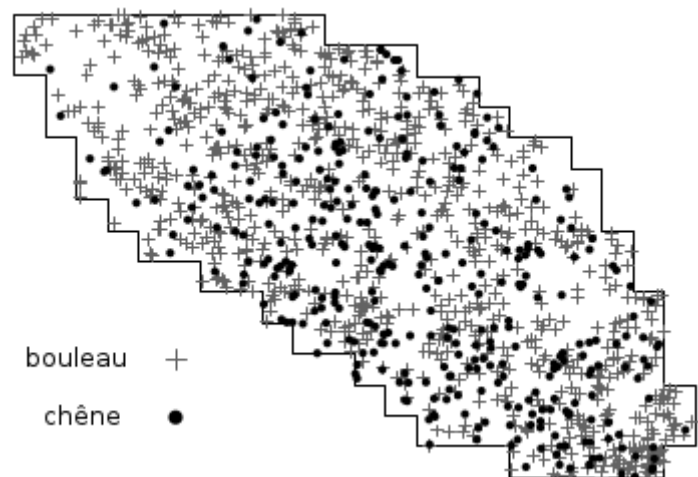
L'analyse statistique de la répartition de points permet de repérer objectivement des informations graphique que l'on n'aurait pas pu distinguer à l'œil nu. De plus une répartition spatiale de point est souvent un substitut à des variables inobservables physiquement (exposition à la pollution, événements historiques).

Il n'y a pas de solution toute faite pour l'analyse statistique des représentations spatiale de point. En effet chaque cas doit être analysé en fonction de son contexte (obtention des données, objectif de l'analyse).

1.1.2 Les points de différents types

Les points d'une représentation peuvent être classés en différents types. L'analyse de la distribution portera alors sur une comparaison des positions et densités des points de chaque type.

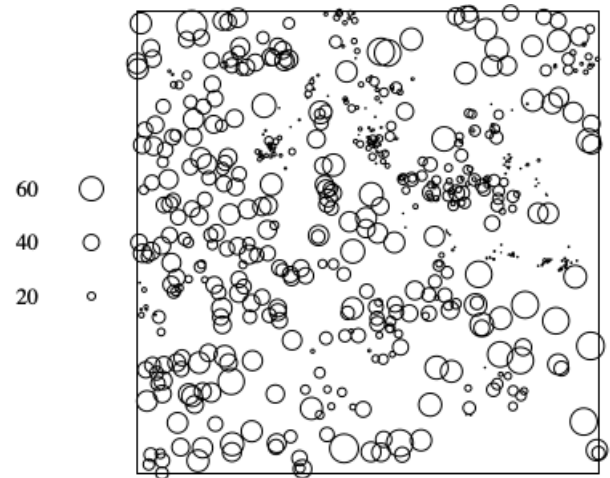
La figure suivante montre par exemple la répartition de bouleaux et de chênes dans une forêt. Une analyse pertinente serait de regarder si les 2 arbres ont la même distribution spatiale ou si la proportion relative des 2 espèces varie sur le domaine.



1.1.3 Répartition de point marqués

Marquer un point permet d'ajouter une information auxiliaire sur ce point (taille d'un arbre, masse d'une étoile)

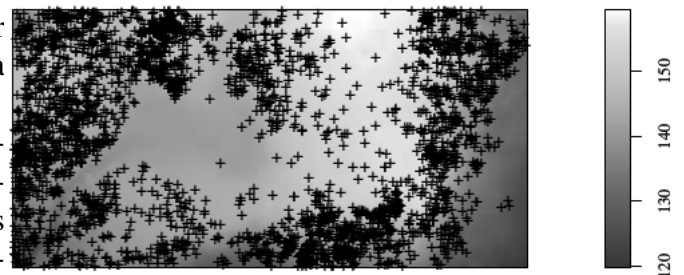
La figure suivante montre par exemple la répartition d'arbre d'une forêt, et chaque arbre est représenté par un cercle plus ou moins gros en fonction de son diamètre en cm.



1.1.4 Covariables

Les bases de données peuvent aussi inclure des covariables, c'est à dire des données utilisées pour expliquer un phénomène, plutôt que de répondre à une question.

La figure suivante montre par exemple la répartition d'arbre d'une forêt, avec une information d'altitude donnée en fond. Cette information n'est pas le but de l'étude, mais ajoute une explication possible.



1.1.5 Différentes dimensions spatiales

Les localisations des points sont généralement en 2 dimensions, mais on peut tout autant avoir une représentation spatiale en 1 dimension (accidents sur un réseau routier), en 3 dimensions (observation de cellules par un microscope 3D), ou même en espace-temps (localisation spatiale et temporel des épicentres de tremblement de terre).

1.1.6 Reproduction de répartitions

On peut comparer des données prises à des données récoltées à des moments et endroits différents pour en faire une analyse. Une reproduction d'expérience peut être traitée indépendamment ou bien apporter des informations supplémentaires à l'analyse d'une répétition d'expérience. Cela permet de séparer différentes sources de variables dans les expériences.

1.2 Méthodes statistiques pour la répartition de point

1.2.1 Statistiques sommaires

aaa

1.2.2 Modèle statistique et inférence

bbb

1.2.3 Validation

cc

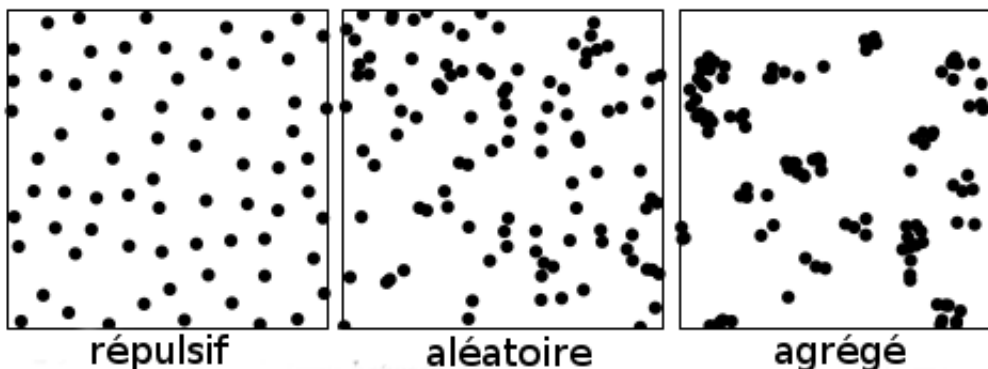
Chapitre 2

Correlation (chap. 7)

2.1 Introduction

Le but de l'analyse d'une répartition de données ponctuelles est en général de déterminer si la position d'un point est indépendante ou non de celle d'un autre. Il y a alors 3 formes typiques de représentation :

- répulsif (points ont tous une certaine distance les uns des autres) ;
- aléatoire (points n'ont pas de dépendance) ;
- agrégée (points forment des groupes de points).



Un outil standard pour mesurer cette dépendance est la covariance (2nd moment), c'est facile à calculer et à manipuler, mais cela requière une bonne estimation de la moyenne (1er moment) et cela ne permet pas de déterminer la cause des regroupement de point (car c'est une statistique sommaire).

2.2 La fonction k

2.2.1 La fonction K empirique

La fonction $\hat{K}(r)$ empirique est la moyenne cumulée des points se trouvant à une distance r de chaque point, corrigé pour les effets de bord et standardisé :

$$\hat{K}(r) = \mathbb{P}\{d(u, X) \leq r\}$$

2.2.2 La vraie fonction K d'un processus de point

2.2.3 Utilisation de la fonction K empirique

2.3 Correction des bords pour la fonction K

2.3.1 Estimation sans les effets de bords

2.3.2 Effets de bords

2.3.3 Correction des bords

2.3.4 Correction isotrope

2.3.5 Correction par translation

2.3.6 Discussion

2.4 Erreurs standard et interval de confiance

2.4.1 Variance sous CSR

2.4.2 Démarrage du system bloqué

2.4.3 Interval de confiance du syteme de démarrage de Loh

2.5 Tester si une répartition est complètement aléatoire

2.5.1 Enveloppes de point

2.5.2 Enveloppes globale

2.5.3 Tests non-graphiques

Chapitre 3

Espacement (chap. 8)

3.1 Introduction

Les fonction statistique, telle que la fonction K, mesurant la corrélation spatiale dans un processus de point sont des outils très populaire pour évaluer la dépendence. Cependant avec cette méthode, certain aspect de la dépendence ne peuvent être observés. Il faut alors utiliser les mesures sur les espaces ou les plus petites distances entre les points.

3.2 Fonction d'espace vide F

3.2.1 Définitions pour un processus de point stationnaire

Si X est un processus de point spatial, la distance :

$$d(u, X) = \min\{\|u - x_i\| : x_i \in X\}$$

d'une position $u \in \mathbb{R}$ au plus proche point d'un processus est appelé 'distance d'espace vide'.
Pour un processus de point stationnaire, la fonction de distance d'espace vide est :

$$F(r) = \mathbb{P}\{d(u, X) \leq r\}$$

définie pour toute distance $r \geq 0$, où u est une position quelconque.

Les valeurs de $F(r)$ sont les probabilités (entre 0 et 1) pour n'importe quel point u , qu'il y ait un point X présent dans à une distance r ou moins de u (qu'il y ait un point dans le cercle de rayon r autour de u).

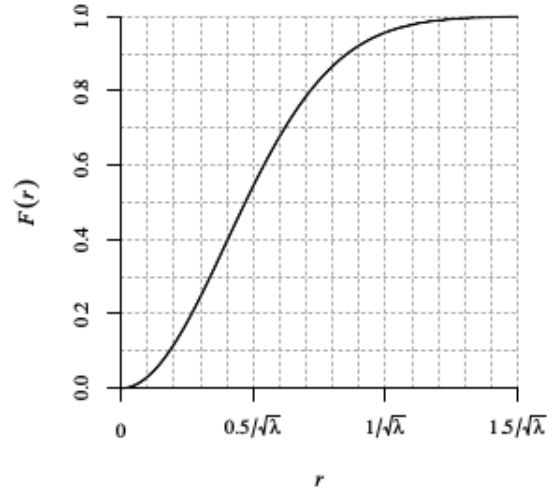
3.2.2 Valeurs pour un aléatoire complet

Pour un aléatoire complet, la fonction F suit une fonction de répartition de Poisson uniforme sur la surface d'un cercle (πr^2) soit pour une intensité λ :

$$F_{pois}(r) = 1 - \exp(-\lambda \pi r^2)$$

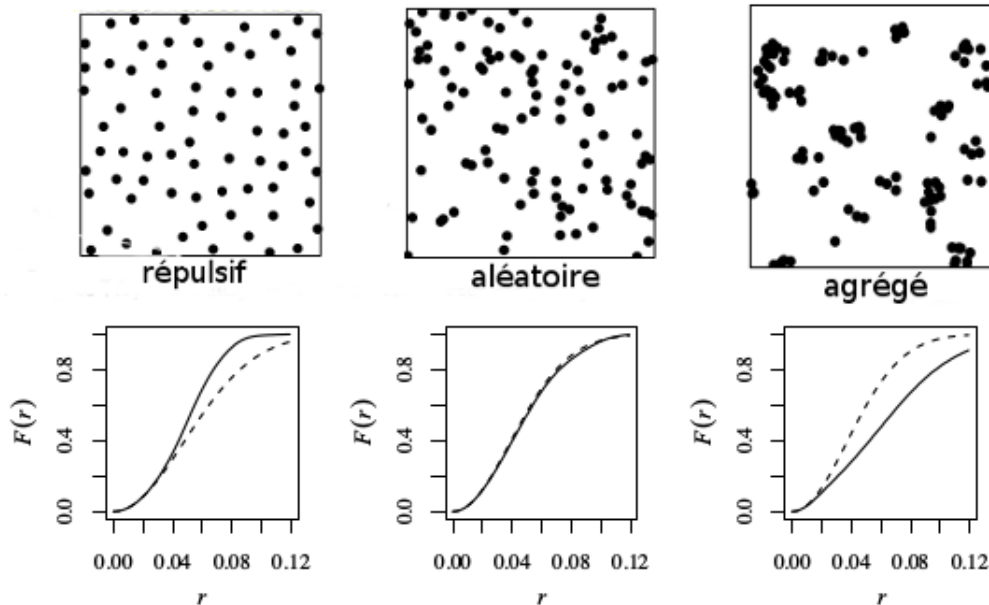
C'est la fonction d'espace vide F théorique pour un aléatoire total.

La figure suivante montre la représentation en fonction de r de cette fonction théorique.



3.2.3 Interprétation de F

La figure ci dessous montre la fonction estimée de F selon si la répartition est agrégée aléatoire ou bien répulsif. Les traits discontinues sont les représentation graphique de la fonction F théorique pour un aléatoire total.



Sur la gauche, la courbe de la fonction calculée est au dessus de celle pour une répartition totalement aléatoire ($\hat{F}(r) > F_{pois}(r)$). Ainsi, pour une distance r , la probabilité que $d(u, X) \leq r$ est plus grande qu'elle ne l'est pour une répartition aléatoire, les espaces vides sont donc plus petits que prévu. Donc la répartition est répulsive.

Sur la droite, la courbe de la fonction calculée est en dessous de celle pour une répartition totalement aléatoire ($\hat{F}(r) < F_{pois}(r)$). Ainsi, pour une distance r , la probabilité que $d(u, X) \leq r$ est plus petite qu'elle ne l'est pour une répartition aléatoire, les espaces vides sont donc plus grands que prévu. Donc la répartition est agrégée.

3.3 Fonction de plus proche voisins G

3.3.1 Définitions pour un processus de point stationnaire

Si x_i est un des points d'une représentation de points \mathbf{x} , la distance du plus proche voisin de x_i est écrite telle que :

$$d_i = d(x_i, \mathbf{x} \setminus x_i)$$

la plus courte distance de x_i à un autre point de \mathbf{x} excepté x_i . Pour un processus de point stationnaire X , la fonction de distance du plus proch voisin est :

$$G(r) = \mathbb{P}\{d(u, X \setminus u) \leq r | u \in X\}$$

définie pour toute distance $r \geq 0$, où u est une position quelconque.

3.3.2 Valeurs pour un aléatoire complet

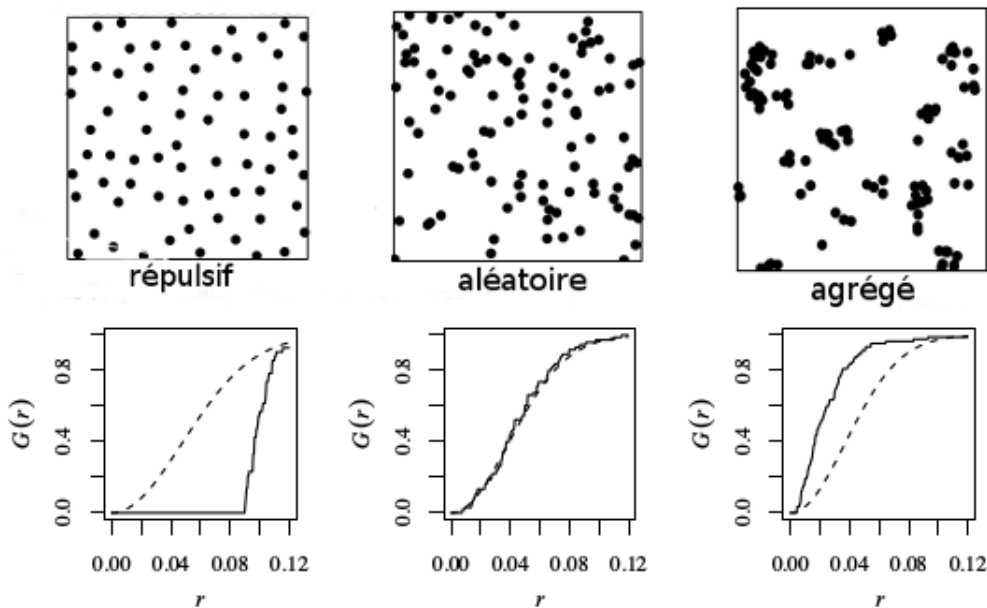
Pour un aléatoire complet, la fonction G est égale à la fonction d'espace vide F . Soit pour un processus homogène de Poisson :

$$G_{pois}(r) = 1 - \exp(-\lambda\pi r^2)$$

Cela n'est vrai que pour une répartition totalement aléatoire, en général F et G seront des fonctions différentes.

3.3.3 Interprétation de G

La figure ci dessous montre la fonction estimé de G selon si la répartition est agrégée aléatoire ou bien répulsif. Les traits discontinues sont les représentation graphique de la fonction F théorique pour un aléatoire total.



Sur la gauche, la courbe de la fonction calculée est en dessous de celle pour une répartition totalement aléatoire ($\hat{G}(r) < G_{pois}(r)$). Ainsi, les distances des plus proches voisins sont plus grandes que celles prévues pour une répartition aléatoire. Donc la répartition est répulsive.

Sur la droite, la courbe de la fonction calculée est au dessus de celle pour une répartition totalement aléatoire ($\hat{G}(r) > G_{pois}(r)$). Ainsi, les distances des plus proches voisins sont plus petites que celles prévues pour une répartition aléatoire. Donc la répartition est agrégée.

3.4 Intervals de confiance et enveloppes de simulation

3.5 Hasard de l'espace vide (désavantages)

3.6 Fonction J

Les distances du plus proche voisin et les distances d'espace-vidé ont la même distribution de probabilité si la répartition des point est totalement aléatoire. Pour les départs des expériences avec une répartition aléatoire complète, ces distances ont tendance à répondre dans des directions opposées : l'un devient plus grand et l'autre plus petit. Cela suggère qu'une comparaison de ces 2 types de distance pourrait s'avérer utile pour évaluer les départ d'une répartition aléatoire.

Une combinaison pratique de G et F, suggérée par la théorie fondamentale, est la fonction J d'un processus stationnaire ponctuel :

$$J(r) = \frac{1-G(r)}{1-F(r)}$$

définie pour tout $r \geq 0$ telle que $F(r) < 1$. Pour un processus de Poisson homogène, $F_{pois} \equiv G_{pois}$ ainsi les valeurs de $J(r) > 1$ son cohérent avec une représentation répulsive, tandis que les valeurs de $J(r) < 1$ sont cohérente avec une représentation agrégée.

La fonction J empirique est insensible aux effets de bords, car ceux-ci s'annulent dans le ratio $(1 - G)/(1 - F)$.

3.7 Théorie pour la correction des bords

3.7.1 Estimation de F et G sans effets de bord

3.7.2 Effets de bords pour F et G

3.7.3 Correction des bords

3.7.4 Correction de Kaplan-Meier

3.7.5 Correction de Hanisch-Chiu-Stoyan

3.7.6 Comparaison des méthodes