

*Espacements bidimensionnels et données
entachées d'erreurs dans l'analyse des
processus ponctuels spatiaux*

Lionel Cucala

Vendredi 8 Décembre 2006



Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Introduction sur les processus ponctuels

Introduction sur les processus ponctuels

- ▶ **Jeux de données** : semis de points consistant en la localisation d'événements de même nature :
 $\{s_1, \dots, s_n\}, s_i \in D \subset \mathbb{R}^d$.

Introduction sur les processus ponctuels

- ▶ **Jeux de données** : semis de points consistant en la localisation d'événements de même nature :

$$\{s_1, \dots, s_n\}, s_i \in D \subset \mathbb{R}^d.$$

- ▶ Modélisation par un processus aléatoire : le nombre N et les positions $\{S_1, \dots, S_N\}$ des événements sont aléatoires.

Caractéristiques des processus ponctuels

Caractéristiques des processus ponctuels

➡ Intensité (de premier ordre) :

$$\lambda(s) = \lim_{\nu(ds) \rightarrow 0} \frac{\mathbb{E}N(ds)}{\nu(ds)}$$

où ds : volume infinitésimal centré en s .

Caractéristiques des processus ponctuels

- Intensité (de premier ordre) :

$$\lambda(s) = \lim_{\nu(ds) \rightarrow 0} \frac{\mathbb{E}N(ds)}{\nu(ds)}$$

où ds : volume infinitésimal centré en s .

- Caractéristiques de second-ordre traduisant les phénomènes d'attraction et de répulsion entre événements.

Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Définition de l'homogénéité spatiale

Définition de l'homogénéité spatiale

Complete Spatial Randomness \Leftrightarrow Processus de Poisson homogène :

Définition de l'homogénéité spatiale

Complete Spatial Randomness \Leftrightarrow Processus de Poisson homogène :

➡ Nombre d'événements $N \sim \mathcal{P}(\lambda\nu(D))$,

Définition de l'homogénéité spatiale

Complete Spatial Randomness \Leftrightarrow Processus de Poisson homogène :

- Nombre d'événements $N \sim \mathcal{P}(\lambda\nu(D))$,
- Sachant $N = n$, $\left((X_1, Y_1), \dots, (X_n, Y_n)\right)$ i.i.d. $\sim U(D)$.

Définition de l'homogénéité spatiale

Complete Spatial Randomness \Leftrightarrow Processus de Poisson homogène :

- Nombre d'événements $N \sim \mathcal{P}(\lambda\nu(D))$,
- Sachant $N = n$, $\left((X_1, Y_1), \dots, (X_n, Y_n)\right)$ i.i.d. $\sim U(D)$.

1^{er} réflexe : tester H_0 : CSR car

Définition de l'homogénéité spatiale

Complete Spatial Randomness \Leftrightarrow Processus de Poisson homogène :

- Nombre d'événements $N \sim \mathcal{P}(\lambda\nu(D))$,
- Sachant $N = n$, $\left((X_1, Y_1), \dots, (X_n, Y_n)\right)$ i.i.d. $\sim U(D)$.

1^{er} réflexe : tester H_0 : CSR car

- aucun paramètre à estimer,

Définition de l'homogénéité spatiale

Complete Spatial Randomness \Leftrightarrow Processus de Poisson homogène :

- Nombre d'événements $N \sim \mathcal{P}(\lambda\nu(D))$,
- Sachant $N = n$, $\left((X_1, Y_1), \dots, (X_n, Y_n)\right)$ i.i.d. $\sim U(D)$.

1^{er} réflexe : tester H_0 : CSR car

- aucun paramètre à estimer,
- aucune prédiction possible.

Tests existants

Tests existants

- ▶ tests basés sur les quadrats
(ex : tests de type Khi-deux)

Tests existants

- tests basés sur les quadrats
(ex : tests de type Khi-deux)
- tests basés sur les distances aux événements
(ex : fonction K de Ripley)

Tests existants

- tests basés sur les quadrats
(ex : tests de type Khi-deux)
- tests basés sur les distances aux événements
(ex : fonction K de Ripley)
- tests basés sur les fonctions de répartition empiriques
(ex : Kolmogorov-Smirnov et Cramer-Von Mises)

Théorie des espacements sur $[0, 1]$

Théorie des espacements sur $[0, 1]$

➡ $(U_1, \dots, U_n) \in [0, 1]^n$

Théorie des espacements sur $[0, 1]$

➤ $(U_1, \dots, U_n) \in [0, 1]^n$

➤ $\rightarrow U_{(0)} = 0 \leq U_{(1)} \leq \dots \leq \dots U_{(n)} \leq U_{(n+1)} = 1$

Théorie des espacements sur $[0, 1]$

- $(U_1, \dots, U_n) \in [0, 1]^n$
- $\rightarrow U_{(0)} = 0 \leq U_{(1)} \leq \dots \leq \dots U_{(n)} \leq U_{(n+1)} = 1$
- $\rightarrow (D_1, \dots, D_{n+1})$, où $D_i = U_{(i)} - U_{(i-1)}$

Théorie des espacements sur $[0, 1]$

- $(U_1, \dots, U_n) \in [0, 1]^n$
- $\rightarrow U_{(0)} = 0 \leq U_{(1)} \leq \dots \leq \dots U_{(n)} \leq U_{(n+1)} = 1$
- $\rightarrow (D_1, \dots, D_{n+1})$, où $D_i = U_{(i)} - U_{(i-1)}$
- H_0 : Uniformité et indépendance \Rightarrow
 $(D_i, i = 1, \dots, n+1)$ i.d. mais $\sum_{i=1}^{n+1} D_i = 1$.

Théorie des espacements sur $[0, 1]$

- $(U_1, \dots, U_n) \in [0, 1]^n$
- $\rightarrow U_{(0)} = 0 \leq U_{(1)} \leq \dots \leq \dots U_{(n)} \leq U_{(n+1)} = 1$
- $\rightarrow (D_1, \dots, D_{n+1})$, où $D_i = U_{(i)} - U_{(i-1)}$
- H_0 : Uniformité et indépendance \Rightarrow
 $(D_i, i = 1, \dots, n+1)$ i.d. mais $\sum_{i=1}^{n+1} D_i = 1$.
- Idée : tester H_0 en observant la dispersion des espacements.

Théorie des espacements sur $[0, 1]$

Théorie des espacements sur $[0, 1]$

Quelles mesures de dispersion ?

Théorie des espacements sur $[0, 1]$

Quelles mesures de dispersion ?

► Variance $\rightarrow V_n = \frac{1}{n+1} \sum_{i=1}^{n+1} ((n+1)D_i - 1)^2$.

Théorie des espacements sur $[0, 1]$

Quelles mesures de dispersion ?

➤ Variance $\rightarrow V_n = \frac{1}{n+1} \sum_{i=1}^{n+1} ((n+1)D_i - 1)^2$.

➤ Déviation en valeur absolue $\rightarrow R_n = \frac{1}{n+1} \sum_{i=1}^{n+1} |(n+1)D_i - 1|$.

Théorie des espacements sur $[0, 1]$

Quelles mesures de dispersion ?

➤ Variance $\rightarrow V_n = \frac{1}{n+1} \sum_{i=1}^{n+1} ((n+1)D_i - 1)^2$.

➤ Déviation en valeur absolue $\rightarrow R_n = \frac{1}{n+1} \sum_{i=1}^{n+1} |(n+1)D_i - 1|$.

Formule générale :

$$S_n = \frac{1}{n+1} \sum_{i=1}^{n+1} g((n+1)D_i)$$

Théorie des espacements sur $[0, 1]$

Quelles mesures de dispersion ?

➤ Variance $\rightarrow V_n = \frac{1}{n+1} \sum_{i=1}^{n+1} ((n+1)D_i - 1)^2$.

➤ Déviation en valeur absolue $\rightarrow R_n = \frac{1}{n+1} \sum_{i=1}^{n+1} |(n+1)D_i - 1|$.

Formule générale :

$$S_n = \frac{1}{n+1} \sum_{i=1}^{n+1} g((n+1)D_i) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \quad \text{sous } H_0 \text{ (Le Cam, 1958).}$$

Théorie des espacements sur $[0, 1]$

Quelles mesures de dispersion ?

➤ Variance $\rightarrow V_n = \frac{1}{n+1} \sum_{i=1}^{n+1} ((n+1)D_i - 1)^2$.

➤ Déviation en valeur absolue $\rightarrow R_n = \frac{1}{n+1} \sum_{i=1}^{n+1} |(n+1)D_i - 1|$.

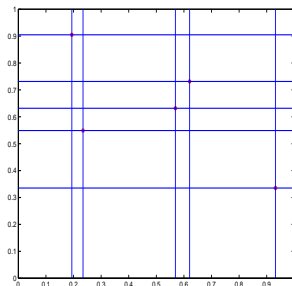
Formule générale :

$$S_n = \frac{1}{n+1} \sum_{i=1}^{n+1} g((n+1)D_i) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N} \quad \text{sous } H_0 \text{ (Le Cam, 1958).}$$

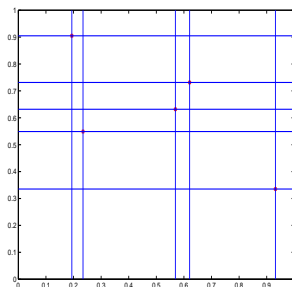
Tests basés sur S_n : puissants contre dépendance.

Espacements bidimensionnels

Espacements bidimensionnels

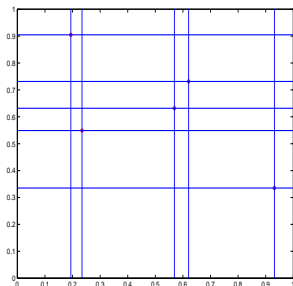


Espacements bidimensionnels



$\blacktriangleright A_{ij} = D_i^x D_j^y, \quad i = 1, \dots, n+1, j = 1, \dots, n+1.$

Espacements bidimensionnels



$$\rightarrow A_{ij} = D_i^x D_j^y, \quad i = 1, \dots, n+1, j = 1, \dots, n+1.$$

$$\rightarrow S_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} g((n+1)^2 A_{ij}).$$

Généralisation du lemme de Le Cam

Généralisation du lemme de Le Cam

- Idée 1 : $((n + 1)D_1, \dots, (n + 1)D_{n+1}) \sim \left(\frac{E_1}{\bar{E}}, \dots, \frac{E_{n+1}}{\bar{E}}\right)$,
où (E_1, \dots, E_{n+1}) *i.i.d.* $\sim \mathcal{E}(1)$.

Généralisation du lemme de Le Cam

➤ Idée 1 : $((n+1)D_1, \dots, (n+1)D_{n+1}) \sim \left(\frac{E_1}{\bar{E}}, \dots, \frac{E_{n+1}}{\bar{E}} \right)$,
 où (E_1, \dots, E_{n+1}) *i.i.d.* $\sim \mathcal{E}(1)$.

➤ $\Rightarrow S_n \sim G_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right)$,
 où $(X_1, \dots, X_{n+1}, Y_1, \dots, Y_{n+1})$ *i.i.d.* $\sim \mathcal{E}(1)$.

Généralisation du lemme de Le Cam

➤ Idée 1 : $((n+1)D_1, \dots, (n+1)D_{n+1}) \sim \left(\frac{E_1}{\bar{E}}, \dots, \frac{E_{n+1}}{\bar{E}}\right)$,
où (E_1, \dots, E_{n+1}) *i.i.d.* $\sim \mathcal{E}(1)$.

➤ $\Rightarrow S_n \sim G_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right)$,
où $(X_1, \dots, X_{n+1}, Y_1, \dots, Y_{n+1})$ *i.i.d.* $\sim \mathcal{E}(1)$.

➤ Idée 2 : Développement de Taylor

$\rightarrow g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right) \sim_{n \rightarrow \infty} g(X_i Y_j) - c(\bar{X} - 1) - c(\bar{Y} - 1)$,
où $c = \text{Cov}(g(X_1 Y_1), X_1)$.

Généralisation du lemme de Le Cam

Généralisation du lemme de Le Cam

Décomposition : $G_n = U_n + R_n$,

Généralisation du lemme de Le Cam

Décomposition : $G_n = U_n + R_n$,

$$\text{où } U_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} [g(X_i Y_j) - c(X_i - 1) - c(Y_j - 1)], \text{ et}$$

Généralisation du lemme de Le Cam

Décomposition : $G_n = U_n + R_n$,

où $U_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} [g(X_i Y_j) - c(X_i - 1) - c(Y_j - 1)]$, et

$$R_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \left[g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right) - g(X_i Y_j) + c(X_i + Y_j - 2) \right].$$

Généralisation du lemme de Le Cam

Décomposition : $G_n = U_n + R_n$,

où $U_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} [g(X_i Y_j) - c(X_i - 1) - c(Y_j - 1)]$, et

$$R_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \left[g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right) - g(X_i Y_j) + c(X_i + Y_j - 2) \right].$$

► U_n U-statistique d'ordre 2 $\Rightarrow U_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}$

Généralisation du lemme de Le Cam

Décomposition : $G_n = U_n + R_n$,

où $U_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} [g(X_i Y_j) - c(X_i - 1) - c(Y_j - 1)]$, et

$$R_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \left[g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right) - g(X_i Y_j) + c(X_i + Y_j - 2) \right].$$

➤ U_n U-statistique d'ordre 2 $\Rightarrow U_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}$

➤ $\mathbb{E}(R_n^2) \xrightarrow[n \rightarrow \infty]{} 0$.

Généralisation du lemme de Le Cam

Décomposition : $G_n = U_n + R_n$,

où $U_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} [g(X_i Y_j) - c(X_i - 1) - c(Y_j - 1)]$, et

$$R_n = \frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \left[g\left(\frac{X_i}{\bar{X}} \frac{Y_j}{\bar{Y}}\right) - g(X_i Y_j) + c(X_i + Y_j - 2) \right].$$

➤ U_n U-statistique d'ordre 2 $\Rightarrow U_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}$

➤ $\mathbb{E}(R_n^2) \xrightarrow[n \rightarrow \infty]{} 0$.

$\Rightarrow S_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}$.

Statistiques utilisées et leurs lois

Statistiques utilisées et leurs lois

➤ Variance $\rightarrow V_n =$

$$\frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \{((n+1)^2 A_{ij} - 1)^2 - 3\} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 32).$$

Statistiques utilisées et leurs lois

➤ Variance $\rightarrow V_n =$

$$\frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \{((n+1)^2 A_{ij} - 1)^2 - 3\} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 32).$$

➤ Déviation en valeur absolue $\rightarrow R_n =$

$$\frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \{|(n+1)^2 A_{ij} - 1| - \mu\} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

Statistiques utilisées et leurs lois

➤ Variance $\rightarrow V_n =$

$$\frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \{((n+1)^2 A_{ij} - 1)^2 - 3\} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 32).$$

➤ Déviation en valeur absolue $\rightarrow R_n =$

$$\frac{1}{(n+1)^{3/2}} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \{|(n+1)^2 A_{ij} - 1| - \mu\} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

Limitation : $n \leq 100 \Rightarrow$ Quantiles empiriques éloignés des quantiles de la loi limite.

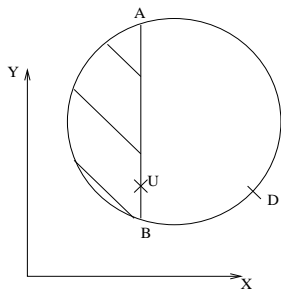
Une procédure multiple

Une procédure multiple

- Comment se ramener à un semis de points sur $[0, 1]^2$?
Utilisation de la transformation de Rosenblatt.

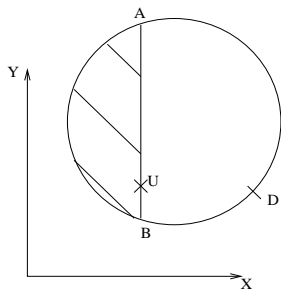
Une procédure multiple

- Comment se ramener à un semis de points sur $[0, 1]^2$?
Utilisation de la transformation de Rosenblatt.



Une procédure multiple

- Comment se ramener à un semis de points sur $[0, 1]^2$?
Utilisation de la transformation de Rosenblatt.



- Comment s'affranchir du référentiel ?
Mise en place d'une procédure multiple.

Conclusions et perspectives

Conclusions et perspectives

- ➡ Tests basés sur les espacements utiles pour détecter de nouvelles alternatives à la CSR.

Conclusions et perspectives

- Tests basés sur les espacements utiles pour détecter de nouvelles alternatives à la CSR.
- Les procédures à coordonnées invariantes améliorent la puissance.

Conclusions et perspectives

- Tests basés sur les espacements utiles pour détecter de nouvelles alternatives à la CSR.
- Les procédures à coordonnées invariantes améliorent la puissance.
- Différentes alternatives à la CSR \Rightarrow différents tests pour les détecter.

Conclusions et perspectives

- Tests basés sur les espacements utiles pour détecter de nouvelles alternatives à la CSR.
- Les procédures à coordonnées invariantes améliorent la puissance.
- Différentes alternatives à la CSR \Rightarrow différents tests pour les détecter.
- Généralisation à la 3D.

Conclusions et perspectives

- Tests basés sur les espacements utiles pour détecter de nouvelles alternatives à la CSR.
- Les procédures à coordonnées invariantes améliorent la puissance.
- Différentes alternatives à la CSR \Rightarrow différents tests pour les détecter.
- Généralisation à la 3D.
- Généralisation aux espacements d'ordre supérieur.

Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Plan de l'exposé

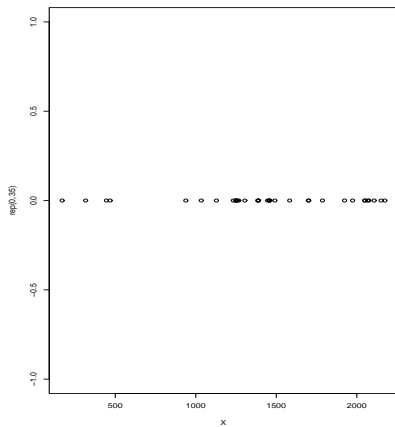
Introduction

1-Tests d'homogénéité spatiale

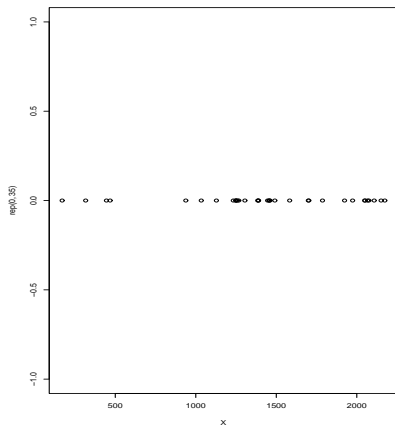
2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Processus ponctuels temporels



Processus ponctuels temporels



Y a-t-il un ou plusieurs agrégats ? Si oui, où ?

Détection d'agrégats temporels

Détection d'agrégats temporels

- ➡ Première étape : adaptation à la densité de population $f(t)$.

Détection d'agrégats temporels

- Première étape : adaptation à la densité de population $f(t)$.
- Evénements :

$$\left\{ T_1, \dots, T_n \right\} \text{ sur } [0, T] \rightarrow \left\{ X_1, \dots, X_n \right\} \text{ sur } [0, 1]$$

$$\text{où } X_i = \frac{\int_0^{T_i} f(t)}{\int_0^T f(t)}.$$

Détection d'agrégats temporels

- Première étape : adaptation à la densité de population $f(t)$.
- Evénements :

$$\left\{ T_1, \dots, T_n \right\} \text{ sur } [0, T] \rightarrow \left\{ X_1, \dots, X_n \right\} \text{ sur } [0, 1]$$

$$\text{où } X_i = \frac{\int_0^{T_i} f(t)}{\int_0^T f(t)}.$$

- Test de $H_0 : \left\{ X_1, \dots, X_n \right\}$ i.i.d. $\sim U([0, 1])$.

La statistique de balayage

La statistique de balayage

➡ $\Lambda(d, n) = \max_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a+d]).$

La statistique de balayage

- ➡ $\Lambda(d, n) = \max_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a+d])$.
- ➡ Distribution de $\Lambda(d, n)$ sous H_0 exprimée et tabulée (Huntington & Naus, 1975).

La statistique de balayage

- $\Lambda(d, n) = \max_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a + d])$.
- Distribution de $\Lambda(d, n)$ sous H_0 exprimée et tabulée (Huntington & Naus, 1975).
- Problème : Longueur de l'agrégat d fixée à priori.

La statistique de balayage (fenêtre variable)

La statistique de balayage (fenêtre variable)

Comment comparer 2 intervalles de longueurs différentes ?

La statistique de balayage (fenêtre variable)

Comment comparer 2 intervalles de longueurs différentes ?

Indice de concentration $I(m, d)$ pour un intervalle de longueur d contenant m événements.

La statistique de balayage (fenêtre variable)

Comment comparer 2 intervalles de longueurs différentes ?

Indice de concentration $I(m, d)$ pour un intervalle de longueur d contenant m événements.

$H_1 : \{X_1, \dots, X_n\}$ i.i.d. $\sim f(\cdot)$

où $f(t) = \begin{cases} \frac{m}{nd} & \text{si } t \in [a, a + d], \\ \frac{1-m/n}{1-d} & \text{si } t \in [0, a] \cup [a + d, 1]. \end{cases}$

La statistique de balayage (fenêtre variable)

La statistique de balayage (fenêtre variable)

Test de H_0 contre H_1

La statistique de balayage (fenêtre variable)

Test de H_0 contre H_1

Neyman-Pearson :

$$\begin{aligned}\frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} &= \left(\frac{m}{nd}\right)^m \left(\frac{1 - m/n}{1 - d}\right)^{n-m} \\ &= l_{scan}(m, d)\end{aligned}$$

La statistique de balayage (fenêtre variable)

Test de H_0 contre H_1

Neyman-Pearson :

$$\begin{aligned} \frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} &= \left(\frac{m}{nd}\right)^m \left(\frac{1 - m/n}{1 - d}\right)^{n-m} \\ &= l_{scan}(m, d) \end{aligned}$$

$$\Lambda_{scan} = \sup_{m \geq m_0} l_{scan}(m, d).$$

La statistique de balayage (fenêtre variable)

Test de H_0 contre H_1

Neyman-Pearson :

$$\begin{aligned} \frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} &= \left(\frac{m}{nd}\right)^m \left(\frac{1 - m/n}{1 - d}\right)^{n-m} \\ &= I_{scan}(m, d) \end{aligned}$$

$$\Lambda_{scan} = \sup_{m \geq m_0} I_{scan}(m, d).$$

Distribution de Λ_{scan} sous H_0 inconnue.

La statistique des espacements anormaux

La statistique des espacements anormaux

➡ $D_i = X_{(i)} - X_{(i-1)}, i = 1, \dots, n + 1.$

La statistique des espacements anormaux

- $D_i = X_{(i)} - X_{(i-1)}, i = 1, \dots, n + 1.$
- Idée : Observer les espacements successifs.

La statistique des espacements anormaux

- $D_i = X_{(i)} - X_{(i-1)}, i = 1, \dots, n + 1.$
- Idée : Observer les espacements successifs.
- $S_{j,k} = \sum_{i=j+1}^k D_i = X_{(k)} - X_{(j)}, 1 \leq j < k \leq n.$

La statistique des espacements anormaux

- $D_i = X_{(i)} - X_{(i-1)}, i = 1, \dots, n + 1.$
- Idée : Observer les espacements successifs.
- $S_{j,k} = \sum_{i=j+1}^k D_i = X_{(k)} - X_{(j)}, 1 \leq j < k \leq n.$
- $\Lambda_{j,k} = F_{H_0}(S_{j,k}) = B_{inc}(S_{j,k}, k - j, n + 1 - k + j).$

La statistique des espacements anormaux

La statistique des espacements anormaux

► $\forall j = 1, \dots, n-1, \forall k = j+1, \dots, n :$

$$\Lambda_{j,k} \sim U([0, 1]) \text{ sous } H_0.$$

La statistique des espacements anormaux

➡ $\forall j = 1, \dots, n-1, \forall k = j+1, \dots, n :$

$$\Lambda_{j,k} \sim U([0, 1]) \text{ sous } H_0.$$

➡ $I_{\text{spac}}(m, d) = 1/\Lambda_{j,k}$

avec $m = k - j + 1$ et $d = X_{(k)} - X_{(j)}$.

La statistique des espacements anormaux

➡ $\forall j = 1, \dots, n-1, \forall k = j+1, \dots, n :$

$$\Lambda_{j,k} \sim U([0, 1]) \text{ sous } H_0.$$

➡ $I_{spac}(m, d) = 1/\Lambda_{j,k}$

avec $m = k - j + 1$ et $d = X_{(k)} - X_{(j)}$.

➡ Statistique retenue :

$$\Lambda_{spac} = \sup_{m \geq m_0} I_{spac}(m, d).$$

La statistique des espacements anormaux

- $\forall j = 1, \dots, n-1, \forall k = j+1, \dots, n :$

$$\Lambda_{j,k} \sim U([0, 1]) \text{ sous } H_0.$$

- $I_{spac}(m, d) = 1/\Lambda_{j,k}$

avec $m = k - j + 1$ et $d = X_{(k)} - X_{(j)}$.

- Statistique retenue :

$$\Lambda_{spac} = \sup_{m \geq m_0} I_{spac}(m, d).$$

- Distribution de Λ_{spac} sous H_0 inconnue.

Utilisation de la statistique

Utilisation de la statistique

- ➡ Agrégat le plus probable : $[X_{(j)}, X_{(k)}]$ maximisant $I_{spac}(m, d)$.

Utilisation de la statistique

- Agrégat le plus probable : $[X_{(j)}, X_{(k)}]$ maximisant $I_{spac}(m, d)$.
- Agrégat jugé significatif (rejet de H_0) si Λ_{spac} supérieur au quantile d'ordre α sous H_0 (obtenu par simulation).

Détection d'agrégats spatiaux

Détection d'agrégats spatiaux

n événements observés en $\{s_1, \dots, s_n\}$, $s_i \in D \subset \mathbb{R}^2$.

Détection d'agrégats spatiaux

n événements observés en $\{s_1, \dots, s_n\}$, $s_i \in D \subset \mathbb{R}^2$.

$H_0 : \{s_1, \dots, s_n\} i.i.d. \sim f(.)$.

Détection d'agrégats spatiaux

n événements observés en $\{s_1, \dots, s_n\}$, $s_i \in D \subset \mathbb{R}^2$.

$H_0 : \{s_1, \dots, s_n\} i.i.d. \sim f(\cdot)$.

Principal problème : agrégats potentiels en nombre infini.

Détection d'agrégats spatiaux

n événements observés en $\{s_1, \dots, s_n\}$, $s_i \in D \subset \mathbb{R}^2$.

$H_0 : \{s_1, \dots, s_n\} i.i.d. \sim f(.)$.

Principal problème : agrégats potentiels en nombre infini.

Première solution :

Détection d'agrégats spatiaux

n événements observés en $\{s_1, \dots, s_n\}$, $s_i \in D \subset \mathbb{R}^2$.

$H_0 : \{s_1, \dots, s_n\} i.i.d. \sim f(.)$.

Principal problème : agrégats potentiels en nombre infini.

Première solution :

- ➡ définir une famille finie d'agrégats potentiels.

Détection d'agrégats spatiaux

n événements observés en $\{s_1, \dots, s_n\}$, $s_i \in D \subset \mathbb{R}^2$.

$H_0 : \{s_1, \dots, s_n\} i.i.d. \sim f(.)$.

Principal problème : agrégats potentiels en nombre infini.

Première solution :

- définir une famille finie d'agrégats potentiels.
- Mêmes indicateurs de concentration avec :

$$\begin{aligned}
 I(m, d) &\rightarrow I(m, A) \\
 d &\rightarrow \frac{\int_A f(s) \nu(ds)}{\int_D f(s) \nu(ds)}
 \end{aligned}$$

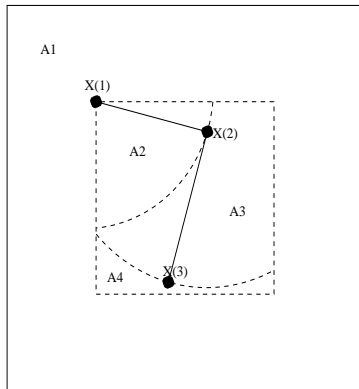
Détection d'agrégats spatiaux

Détection d'agrégats spatiaux

Deuxième solution : transformer les données.

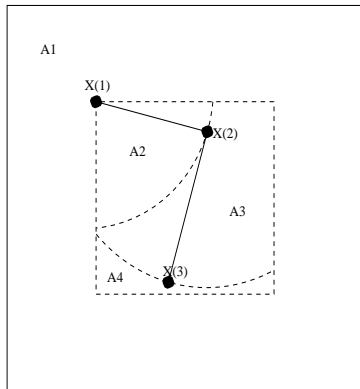
Détection d'agrégats spatiaux

Deuxième solution : transformer les données.



Détection d'agrégats spatiaux

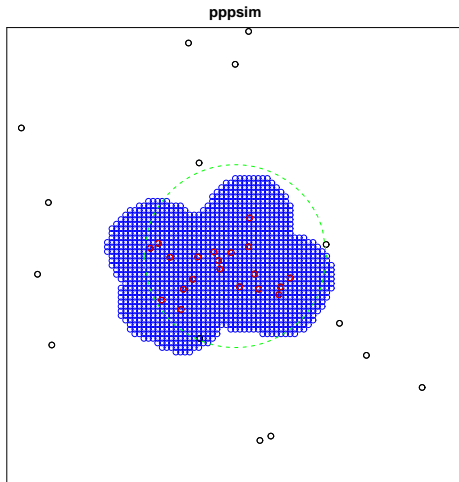
Deuxième solution : transformer les données.



$$\text{Sous } H_0 : \left\{ \frac{\int_{A_1} f(s)\nu(ds)}{\int_D f(s)\nu(ds)}, \dots, \frac{\int_{A_{n+1}} f(s)\nu(ds)}{\int_D f(s)\nu(ds)} \right\} \sim \{D_1, \dots, D_{n+1}\}.$$

Application à des données simulées

Application à des données simulées



Conclusions et perspectives

Conclusions et perspectives

- ▶ Alternative à la statistique de balayage : indépendante de H_1 , plus puissante (temporel et spatial).

Conclusions et perspectives

- Alternative à la statistique de balayage : indépendante de H_1 , plus puissante (temporel et spatial).
- Procédure de détection d'agrégats de toutes formes (spatial).

Conclusions et perspectives

- Alternative à la statistique de balayage : indépendante de H_1 , plus puissante (temporel et spatial).
- Procédure de détection d'agrégats de toutes formes (spatial).
- Nécessité de pouvoir détecter plusieurs agrégats : mise en place d'une procédure multiple.

Conclusions et perspectives

- Alternative à la statistique de balayage : indépendante de H_1 , plus puissante (temporel et spatial).
- Procédure de détection d'agrégats de toutes formes (spatial).
- Nécessité de pouvoir détecter plusieurs agrégats : mise en place d'une procédure multiple.
- Nécessité de comparer à d'autres tests.

Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Plan de l'exposé

Introduction

1-Tests d'homogénéité spatiale

2-Détection d'agrégats

3-Estimation de l'intensité (cas bruité)

Les processus ponctuels bruités

Les processus ponctuels bruités

➡ **Y** processus ponctuel défini sur $X \subseteq \mathbb{R}^2$.

Les processus ponctuels bruités

- \mathbf{Y} processus ponctuel défini sur $X \subseteq \mathbb{R}^2$.
- On observe $\{z_1, \dots, z_n\}$ sur le domaine borné $D \subseteq X$ tel que

Les processus ponctuels bruités

- \mathbf{Y} processus ponctuel défini sur $X \subseteq \mathbb{R}^2$.
- On observe $\{z_1, \dots, z_n\}$ sur le domaine borné $D \subseteq X$ tel que

$$z_i = y_i + \epsilon_i, \quad i = 1, \dots, n$$

Les processus ponctuels bruités

- \mathbf{Y} processus ponctuel défini sur $X \subseteq \mathbb{R}^2$.
- On observe $\{z_1, \dots, z_n\}$ sur le domaine borné $D \subseteq X$ tel que

$$z_i = y_i + \epsilon_i, \quad i = 1, \dots, n$$

$$\{\epsilon_i, \quad i = 1, \dots, n\} \text{ i.i.d. } \sim g(\cdot)$$

Les processus ponctuels bruités

- \mathbf{Y} processus ponctuel défini sur $X \subseteq \mathbb{R}^2$.
- On observe $\{z_1, \dots, z_n\}$ sur le domaine borné $D \subseteq X$ tel que

$$z_i = y_i + \epsilon_i, \quad i = 1, \dots, n$$

$$\{\epsilon_i, \quad i = 1, \dots, n\} \text{ i.i.d. } \sim g(\cdot)$$

$$\epsilon_i \perp\!\!\!\perp y_i, \quad i = 1, \dots, n$$

Les processus ponctuels bruités

- \mathbf{Y} processus ponctuel défini sur $X \subseteq \mathbb{R}^2$.
- On observe $\{z_1, \dots, z_n\}$ sur le domaine borné $D \subseteq X$ tel que

$$z_i = y_i + \epsilon_i, \quad i = 1, \dots, n$$

$$\{\epsilon_i, \quad i = 1, \dots, n\} \text{ i.i.d. } \sim g(\cdot)$$

$$\epsilon_i \perp\!\!\!\perp y_i, \quad i = 1, \dots, n$$

- But : estimer l'intensité $\lambda_Y(s)$ sur D .

L'estimateur à noyau de l'intensité

L'estimateur à noyau de l'intensité

► Diggle (1985) :

$$\begin{aligned}\forall s \in \mathbb{R}^2, \hat{\lambda}_{h,EC}(s) &= \frac{\sum_{j=1}^n \frac{1}{h^2} K\left(\frac{s-z_j}{h}\right)}{\int_D \frac{1}{h^2} K\left(\frac{s-u}{h}\right) \nu(du)} \\ &= \frac{\hat{\lambda}_h(s)}{\rho_h(s)}.\end{aligned}$$

La méthode de déconvolution

La méthode de déconvolution

► On suppose $\forall t \in \mathbb{R}^2, |\mathcal{F}(g)(t)| > 0$.

La méthode de déconvolution

➡ On suppose $\forall t \in \mathbb{R}^2, |\mathcal{F}(g)(t)| > 0$.

➡

$$\begin{aligned}z_i &= y_i + \epsilon_i, \quad i = 1, \dots, n \\ \Rightarrow \lambda_Z &= \lambda_Y * g \\ \Rightarrow \mathcal{F}(\lambda_Z)(\cdot) &= \mathcal{F}(\lambda_Y)(\cdot) \mathcal{F}(g)(\cdot) \\ \Rightarrow \mathcal{F}(\lambda_Y)(\cdot) &= \mathcal{F}(\lambda_Z)(\cdot) / \mathcal{F}(g)(\cdot) \\ \Rightarrow \lambda_Y &= \mathcal{F}^{-1}(\mathcal{F}(\lambda_Z)(\cdot) / \mathcal{F}(g)(\cdot))\end{aligned}$$

L'estimateur à noyau de déconvolution

L'estimateur à noyau de déconvolution

Déconvolution+Correction de bord a posteriori :

L'estimateur à noyau de déconvolution

Déconvolution+Correction de bord a posteriori :

$$\lambda_{h,EC}^*(s) = \frac{\lambda_h^*(s)}{p_h^*(s)} = \frac{\sum_{j=1}^n \frac{1}{h^2} K_h^*\left(\frac{s-z_j}{h}\right)}{\int_D \frac{1}{h^2} K_h^*\left(\frac{s-u}{h}\right) \nu(du)}$$

L'estimateur à noyau de déconvolution

Déconvolution+Correction de bord a posteriori :

$$\lambda_{h,EC}^*(s) = \frac{\lambda_h^*(s)}{p_h^*(s)} = \frac{\sum_{j=1}^n \frac{1}{h^2} K_h^*\left(\frac{s-z_j}{h}\right)}{\int_D \frac{1}{h^2} K_h^*\left(\frac{s-u}{h}\right) \nu(du)}$$

où $K_h^*(t) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{it'y} \mathcal{F}(K)(y) / \mathcal{F}(g)(y/h) dy.$

L'étude asymptotique

L'étude asymptotique

$$\lambda_Y = \mathbb{E}(N) \times \lambda_Y^0 \text{ avec } \mathbb{E}(N) \rightarrow \infty.$$

L'étude asymptotique

$$\lambda_Y = \mathbb{E}(N) \times \lambda_Y^0 \text{ avec } \mathbb{E}(N) \rightarrow \infty.$$

Conclusion : Estimateur asymptotiquement biaisé sauf si $D = \mathbb{R}^2$.

Le choix de la largeur de bande

Le choix de la largeur de bande

On cherche h minimisant

$$MISE(h) = \mathbb{E} \int_D \{\lambda_{h,EC}^*(s) - \lambda_Y(s)\}^2 \nu(ds).$$

Le choix de la largeur de bande

On cherche h minimisant

$$MISE(h) = \mathbb{E} \int_D \{\lambda_{h,EC}^*(s) - \lambda_Y(s)\}^2 \nu(ds).$$

- ➡ On utilise le biais et la variance asymptotique lorsque $D = \mathbb{R}^2$.

Le choix de la largeur de bande

On cherche h minimisant

$$MISE(h) = \mathbb{E} \int_D \{\lambda_{h,EC}^*(s) - \lambda_Y(s)\}^2 \nu(ds).$$

- On utilise le biais et la variance asymptotique lorsque $D = \mathbb{R}^2$.
- Comment estimer $\int_{\mathbb{R}^2} (\nabla^2 \lambda_Y(s))^2 \nu(ds)$?
Règle de référence gaussienne.

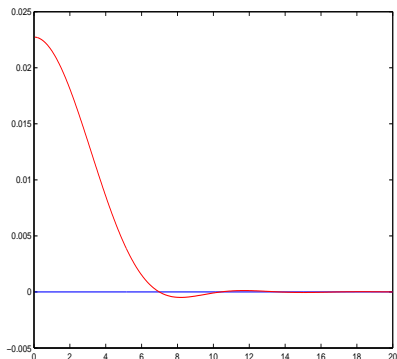
Un noyau à bande limitée

Un noyau à bande limitée

$$K_0(t) = \frac{48 t^3 \cos(t) - 6t^2 \sin(t) + 15 \sin(t) - 15t \cos(t)}{\pi t^7}$$

Un noyau à bande limitée

$$K_0(t) = \frac{48 t^3 \cos(t) - 6t^2 \sin(t) + 15 \sin(t) - 15t \cos(t)}{\pi t^7}$$



Les transformées de Fourier

Les transformées de Fourier

$$\blacktriangleright \mathcal{F}(K)(t) = (1 - t_1^2)^3(1 - t_2^2)^3 \mathbb{1}_{[-1,1]^2}(t).$$

Les transformées de Fourier

➡ $\mathcal{F}(K)(t) = (1 - t_1^2)^3(1 - t_2^2)^3 \mathbb{1}_{[-1,1]^2}(t).$

➡ $\mathcal{F}(g)$ généralement explicite.

Les transformées de Fourier

- $\mathcal{F}(K)(t) = (1 - t_1^2)^3(1 - t_2^2)^3 \mathbb{1}_{[-1,1]^2}(t)$.
- $\mathcal{F}(g)$ généralement explicite.
- Les transformées de Fourier inverses sont obtenues par procédure numérique (Simpson).

Une étude de simulation

Une étude de simulation

- ▶ $\{y_i, i = 1, \dots, n\}$ issues d'un P.P. inhomogène d'intensité $\lambda_Y(s) = C [1 + 0.7 \cos(2\pi(\|s\| - 0.5))]$.

Une étude de simulation

- ▶ $\{y_i, i = 1, \dots, n\}$ issues d'un P.P. inhomogène d'intensité $\lambda_Y(s) = C [1 + 0.7 \cos(2\pi(\|s\| - 0.5))]$.
- ▶ $\{\epsilon_i, i = 1, \dots, n\}$ i.i.d. $\sim g$

Une étude de simulation

- $\{y_i, i = 1, \dots, n\}$ issues d'un P.P. inhomogène d'intensité $\lambda_Y(s) = C[1 + 0.7 \cos(2\pi(\|s\| - 0.5))]$.
- $\{\epsilon_i, i = 1, \dots, n\}$ i.i.d. $\sim g$
- $z_i = y_i + \epsilon_i, \quad i = 1, \dots, n$

Une étude de simulation

Une étude de simulation

Notons :

Une étude de simulation

Notons :

$$ISE = \int_{[0,1]^2} (\hat{\lambda}_{h_{opt}, EC} - \lambda_Y(s))^2 \nu(ds)$$

Une étude de simulation

Notons :

$$ISE = \int_{[0,1]^2} (\hat{\lambda}_{h_{opt}, EC} - \lambda_Y(s))^2 \nu(ds)$$

$$ISE^* = \int_{[0,1]^2} (\lambda_{h^*}^*(s) - \lambda_Y(s))^2 \nu(ds)$$

Une étude de simulation

Notons :

$$ISE = \int_{[0,1]^2} (\hat{\lambda}_{h_{opt}, EC} - \lambda_Y(s))^2 \nu(ds)$$

$$ISE^* = \int_{[0,1]^2} (\lambda_{h^*}^*(s) - \lambda_Y(s))^2 \nu(ds)$$

$$ISE^{**} = \int_{[0,1]^2} (\lambda_{h^*, EC}^*(s) - \lambda_Y(s))^2 \nu(ds)$$

Une étude de simulation

Une étude de simulation

TAB.: Erreur gaussienne, $\sigma=0.05$

	<i>ISE</i>	<i>ISE*</i>	<i>ISE**</i>
1er quartile ($*10^3$)	0.8185	1.4153	0.6655
médiane ($*10^3$)	1.2474	1.7199	0.9298
3ème quartile ($*10^3$)	1.5281	1.8908	1.2138

Une étude de simulation

TAB.: Erreur gaussienne, $\sigma=0.05$

	<i>ISE</i>	<i>ISE*</i>	<i>ISE**</i>
1er quartile ($\cdot 10^3$)	0.8185	1.4153	0.6655
médiane ($\cdot 10^3$)	1.2474	1.7199	0.9298
3ème quartile ($\cdot 10^3$)	1.5281	1.8908	1.2138

TAB.: Erreur laplacienne, $\sigma=0.05$

	<i>ISE</i>	<i>ISE*</i>	<i>ISE**</i>
1er quartile ($\cdot 10^3$)	0.7869	1.1814	0.7689
médiane ($\cdot 10^3$)	1.4859	1.4223	1.1308
3ème quartile ($\cdot 10^3$)	2.0375	1.5114	1.4210

Conclusions et perspectives

Conclusions et perspectives

- ➡ Nécessité de combiner déconvolution et correction des effets de bord.

Conclusions et perspectives

- ➡ Nécessité de combiner déconvolution et correction des effets de bord.
- ➡ Extension de la méthode à l'étude des caractéristiques de second ordre.