

# *Détection d'agrégats pour données ponctuelles*

Lionel Cucala

Mercredi 17 Janvier 2007



# *Plan de l'exposé*

*Introduction*

*1-Le cadre temporel*

*2-Le cadre spatial*

*Conclusion*

# *Plan de l'exposé*

## *Introduction*

1-Le cadre temporel

2-Le cadre spatial

Conclusion

# *Les processus ponctuels*

## *Les processus ponctuels*

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

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

## *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.

## *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.

- Intensité (de premier ordre) :

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

où  $ds$  : volume infinitésimal centré en  $s$ .

# *La détection d'agrégats*



## *La détection d'agrégats*

➡ Définition :

Agrégat = zone d'intensité anormalement élevée, non due au hasard.

## *La détection d'agrégats*

➤ Définition :

Agrégat= zone d'intensité anormalement élevée, non due au hasard.

➤ Question :

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

## *La détection d'agrégats*

➤ Définition :

Agrégat = zone d'intensité anormalement élevée, non due au hasard.

➤ Question :

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

➤ Attention !

Adaptation à la densité de population  $f(s)$ .

# *Le modèle*

## *Le modèle*

- On cherche à tester si  $(s_1, \dots, s_n)$  issu d'un processus de Poisson inhomogène d'intensité  $\lambda(s) = af(s)$ .

## *Le modèle*

- On cherche à tester si  $(s_1, \dots, s_n)$  issu d'un processus de Poisson inhomogène d'intensité  $\lambda(s) = af(s)$ .
- On se contentera de tester  $H_0 : (s_1, \dots, s_n)$  i.i.d. de densité  $f(s)$ .

## *Le modèle*

- On cherche à tester si  $(s_1, \dots, s_n)$  issu d'un processus de Poisson inhomogène d'intensité  $\lambda(s) = af(s)$ .
- On se contentera de tester  $H_0 : (s_1, \dots, s_n)$  i.i.d. de densité  $f(s)$ .
- Objectifs : test global de  $H_0$  **et** détection du (ou des) agrégats.

# *Plan de l'exposé*

*Introduction*

*1-Le cadre temporel*

*2-Le cadre spatial*

*Conclusion*



# *Plan de l'exposé*

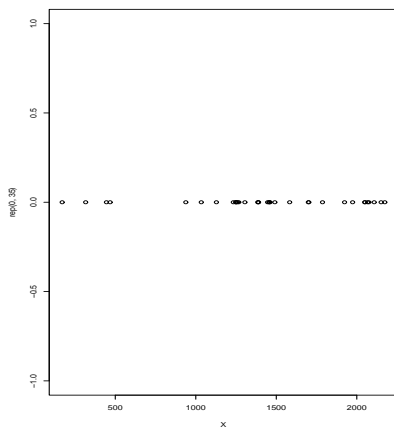
Introduction

*1-Le cadre temporel*

2-Le cadre spatial

Conclusion

# *Processus ponctuels temporels*



# *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*

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

$$a^* = \operatorname{argmax}_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a+d]).$$



## *La statistique de balayage*

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

$$a^* = \operatorname{argmax}_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a + d]).$$

- $\blacktriangleright$  Distribution de  $\Lambda(d, n)$  sous  $H_0$  exprimée et tabulée (Huntington & Naus, 1975).

## *La statistique de balayage*

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

$$a^* = \operatorname{argmax}_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a + d]).$$

$\blacktriangleright$  Distribution de  $\Lambda(d, n)$  sous  $H_0$  exprimée et tabulée (Huntington & Naus, 1975).

$\blacktriangleright$  Rejet de  $H_0$  si  $P_0(\Lambda(d, n) > \sum_{i=1}^n \mathbb{1}(X_i \in [a^*, a^* + d])) < \alpha$ .

## *La statistique de balayage*

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

$$a^* = \operatorname{argmax}_{a \in [0, 1-d]} \sum_{i=1}^n \mathbb{1}(X_i \in [a, a + d]).$$

➤ Distribution de  $\Lambda(d, n)$  sous  $H_0$  exprimée et tabulée (Huntington & Naus, 1975).

➤ Rejet de  $H_0$  si  $P_0(\Lambda(d, n) > \sum_{i=1}^n \mathbb{1}(X_i \in [a^*, a^* + d])) < \alpha.$

➤ Problème : Longueur de l'agrégat  $d$  fixée à priori.

# *Introduction d'une fenêtre variable*

## *Introduction d'une fenêtre variable*

Comment comparer 2 intervalles de longueurs différentes ?

## *Introduction d'une fenêtre variable*

Comment comparer 2 intervalles de longueurs différentes ?

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

## *Introduction d'une fenêtre variable*

Comment comparer 2 intervalles de longueurs différentes ?

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

Exemple :  $I(m, d) = m/d$  (# d'événements / unité de longueur)  
⇒ Sélection du plus petit intervalle contenant 2 événements.

# *La statistique de balayage (fenêtre variable)*



## *La statistique de balayage (fenêtre variable)*

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

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

## *La statistique de balayage (fenêtre variable)*

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

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

Test de  $H_0$  contre  $H_1$  :

## *La statistique de balayage (fenêtre variable)*

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

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

Test de  $H_0$  contre  $H_1$  :

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

## *La statistique de balayage (fenêtre variable)*

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

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

Test de  $H_0$  contre  $H_1$  :

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

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

## *La statistique de balayage (fenêtre variable)*

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

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

Test de  $H_0$  contre  $H_1$  :

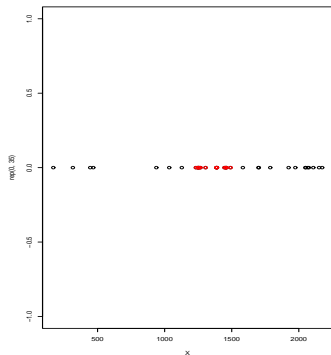
$$\begin{aligned} \frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} &= \left(\frac{m}{nd}\right)^m \left(\frac{1-m/n}{1-d}\right)^{n-m} \\ &= I_{scan}(m, d) \end{aligned}$$

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

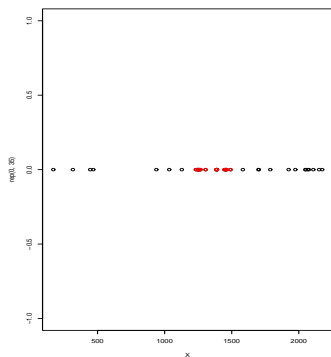
Distribution de  $\Lambda_{scan}$  sous  $H_0$  inconnue.

# *La statistique de balayage (fenêtre variable)*

# *La statistique de balayage (fenêtre variable)*



# *La statistique de balayage (fenêtre variable)*



Quantile empirique de  $\Lambda$  :

$$0.135(n_0 = 2)$$

$$0.009(n_0 = 5)$$



# *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  $\Lambda_{spac}$  sous  $H_0$  inconnue.

# *Utilisation de la statistique*

## *Utilisation de la statistique*

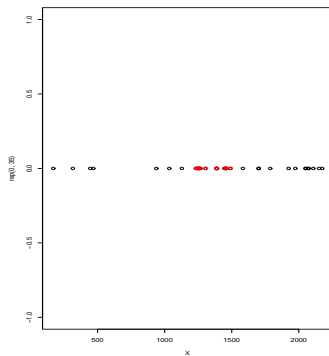
- ➡ Agrégat le plus probable :  $[X_{(j)}, X_{(k)}]$  maximisant  $I_{spac}(m, d)$  avec  $m = k - j + 1$  et  $d = X_{(k)} - X_{(j)}$ .

## *Utilisation de la statistique*

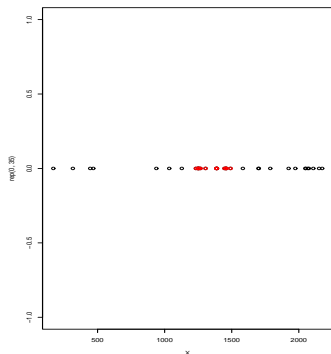
- Agrégat le plus probable :  $[X_{(j)}, X_{(k)}]$  maximisant  $I_{spac}(m, d)$  avec  $m = k - j + 1$  et  $d = X_{(k)} - X_{(j)}$ .
- Agrégat jugé significatif (rejet de  $H_0$ ) si  $\Lambda_{spac}$  supérieur au quantile d'ordre  $\alpha$  sous  $H_0$  (obtenu par simulation).

# *La statistique des espacements anormaux*

# *La statistique des espacements anormaux*



# *La statistique des espacements anormaux*



Quantile empirique de  $\Lambda$  :

$$0.004(n_0 = 2)$$

$$0.003(n_0 = 5)$$



# *Application à des données simulées*

## *Application à des données simulées*

Simulation de 100 jeux de données de 35 évènements (agrégat plat).

## *Application à des données simulées*

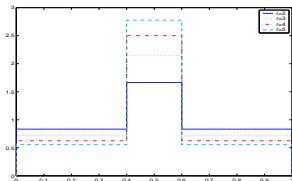
Simulation de 100 jeux de données de 35 évènements (agrégat plat).

$$f_1(x) = \begin{cases} \frac{r}{0.8+0.2r} & \text{si } 0.4 \leq x \leq 0.6, \\ \frac{1}{0.8+0.2r} & \text{si } x \in [0, 0.4] \cup [0.6, 1], \\ 0 & \text{sinon.} \end{cases}$$

## *Application à des données simulées*

Simulation de 100 jeux de données de 35 évènements (agrégat plat).

$$f_1(x) = \begin{cases} \frac{r}{0.8+0.2r} & \text{si } 0.4 \leq x \leq 0.6, \\ \frac{1}{0.8+0.2r} & \text{si } x \in [0, 0.4] \cup [0.6, 1], \\ 0 & \text{sinon.} \end{cases}$$



# *Application à des données simulées*

# *Application à des données simulées*

Erreur de première espèce : 5%.

# *Application à des données simulées*

Erreur de première espèce : 5%.

$$c(t) = \mathbb{1}_{[0.4,0.6]}(t).$$

## *Application à des données simulées*

Erreur de première espèce : 5%.

$$c(t) = \mathbb{1}_{[0.4, 0.6]}(t).$$

$$\hat{c}(t) = \begin{cases} \mathbb{1}_{[X_{(j)}, X_{(k)}]}(t) & \text{si agrégat significatif,} \\ 0 & \text{sinon.} \end{cases}$$



## *Application à des données simulées*

Erreur de première espèce : 5%.

$$c(t) = \mathbb{1}_{[0.4, 0.6]}(t).$$

$$\hat{c}(t) = \begin{cases} \mathbb{1}_{[X_{(j)}, X_{(k)}]}(t) & \text{si agrégat significatif,} \\ 0 & \text{sinon.} \end{cases}$$

Indice de concordance :  $R = \int_0^1 \mathbb{1}(c(t) = \hat{c}(t))$ .

## Application à des données simulées

**TAB.:** Empirical power of tests for CSR against a flat cluster with  $n_0 = 5$

| $r$ | Mean correlation index of the following : |                    |
|-----|---|--------------------|
|     | $\Lambda^{spac}$                          | $\Lambda$          |
| 2   | <b>81.12 % (0.19)</b>                     | 80.45 % (0.18)     |
| 3   | <b>86.50 % (0.59)</b>                     | 85.06 % (0.50)     |
| 4   | <b>88.95 % (0.83)</b>                     | 88.14 % (0.81)     |
| 5   | <b>94.09 % (0.97)</b>                     | 92.07 % (0.95)     |
| 10  | 96.95 % (1)                               | <b>97.08 % (1)</b> |

# *Application à des données simulées*

## *Application à des données simulées*

Simulation de 100 jeux de données de 35 évènements (agrégat en cloche).

## *Application à des données simulées*

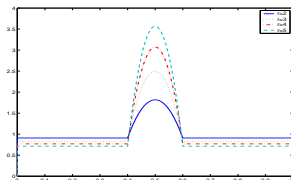
Simulation de 100 jeux de données de 35 évènements (agrégat en cloche).

$$f_2(x) = \begin{cases} \frac{10}{r+9} \{1 + (r-1) * [1 - 100(x - 0.5)^2]\} & \text{si } 0.4 \leq x \leq 0.6, \\ \frac{10}{r+9} & \text{si } x \in [0, 0.4] \cup [0.6, 1], \\ 0 & \text{sinon.} \end{cases}$$

## Application à des données simulées

Simulation de 100 jeux de données de 35 évènements (agrégat en cloche).

$$f_2(x) = \begin{cases} \frac{10}{r+9} \{1 + (r-1) * [1 - 100(x-0.5)^2]\} & \text{si } 0.4 \leq x \leq 0.6, \\ \frac{10}{r+9} & \text{si } x \in [0, 0.4] \cup [0.6, 1], \\ 0 & \text{sinon.} \end{cases}$$



## Application à des données simulées

**TAB.:** Empirical power of tests for CSR against a bell-shaped cluster with  $n_0 = 5$

| $r$ | Estimated power of the following : |                |
|-----|------------------------------------|----------------|
|     | $\Lambda^{spac}$                   | $\Lambda$      |
| 2   | <b>80.72 % (0.17)</b>              | 80.06 % (0.19) |
| 3   | <b>82.41 % (0.29)</b>              | 81.10 % (0.22) |
| 4   | <b>84.87 % (0.55)</b>              | 82.72 % (0.41) |
| 5   | <b>88.37 % (0.78)</b>              | 85.43 % (0.61) |
| 10  | <b>93.52 % (1)</b>                 | 92.96 % (0.99) |

## Application à des données simulées

**TAB.:** Empirical power of tests for CSR against a bell-shaped cluster with  $n_0 = 2$

| $r$ | Estimated power of the following : |                |
|-----|------------------------------------|----------------|
|     | $\Lambda^{spac}$                   | $\Lambda$      |
| 2   | <b>81.21 % (0.21)</b>              | 80.13 % (0.16) |
| 3   | <b>82.85 % (0.47)</b>              | 82.11 % (0.43) |
| 4   | <b>85.97 % (0.54)</b>              | 84.66 % (0.53) |
| 5   | <b>89.78 % (0.85)</b>              | 88.27 % (0.79) |
| 10  | <b>93.52 % (1)</b>                 | 93.39 % (1)    |



# *Une procédure multiple*

## *Une procédure multiple*

► A analyser :  $\{X_{(1)}, \dots, X_{(N)}\}$  sur  $[0, 1]$ .

## *Une procédure multiple*

- A analyser :  $\{X_{(1)}, \dots, X_{(N)}\}$  sur  $[0, 1]$ .
- Agrégat significatif sur  $[X_{(j)}, X_{(k)}] \Rightarrow \{X_{(1)}, \dots, X_{(j-1)}\}$  sur  $[0, X_{(j)}]$  et  $\{X_{(k+1)}, \dots, X_{(N)}\}$  sur  $[X_{(k)}, 1]$  à analyser.

## *Une procédure multiple*

- A analyser :  $\{X_{(1)}, \dots, X_{(N)}\}$  sur  $[0, 1]$ .
- Agrégat significatif sur  $[X_{(j)}, X_{(k)}] \Rightarrow \{X_{(1)}, \dots, X_{(j-1)}\}$  sur  $[0, X_{(j)}]$  et  $\{X_{(k+1)}, \dots, X_{(N)}\}$  sur  $[X_{(k)}, 1]$  à analyser.
- Itération jusqu'à épuisement.

# *Une procédure à noyau (Kelsall & Diggle, 1995)*

## *Une procédure à noyau (Kelsall & Diggle, 1995)*

➡  $\hat{f}_h(t)$  estimateur à noyau issu de  $\{X_1, \dots, X_N\}$

## *Une procédure à noyau (Kelsall & Diggle, 1995)*

- ➡  $\hat{f}_h(t)$  estimateur à noyau issu de  $\{X_1, \dots, X_N\}$
- ➡  $\hat{f}_h^{(1)}(t), \dots, \hat{f}_h^{(20)}(t)$  estimateurs à noyau issus de 20 échantillons simulés sous  $H_0$ .

## *Une procédure à noyau (Kelsall & Diggle, 1995)*

- $\hat{f}_h(t)$  estimateur à noyau issu de  $\{X_1, \dots, X_N\}$
- $\hat{f}_h^{(1)}(t), \dots, \hat{f}_h^{(20)}(t)$  estimateurs à noyau issus de 20 échantillons simulés sous  $H_0$ .
- Test global : on rejette  $H_0$  si

$$\int_0^1 (\hat{f}_h(t) - 1)^2 dt > \max_{1 \leq i \leq 20} \int_0^1 (\hat{f}_h^{(i)}(t) - 1)^2 dt.$$



## *Une procédure à noyau (Kelsall & Diggle, 1995)*

- $\hat{f}_h(t)$  estimateur à noyau issu de  $\{X_1, \dots, X_N\}$
- $\hat{f}_h^{(1)}(t), \dots, \hat{f}_h^{(20)}(t)$  estimateurs à noyau issus de 20 échantillons simulés sous  $H_0$ .
- Test global : on rejette  $H_0$  si

$$\int_0^1 (\hat{f}_h(t) - 1)^2 dt > \max_{1 \leq i \leq 20} \int_0^1 (\hat{f}_h^{(i)}(t) - 1)^2 dt.$$

- Test local : agrégat en  $t$  si

$$\hat{f}_h(t) > \max_{1 \leq i \leq 20} \hat{f}_h^{(i)}(t).$$

## Application à des données simulées

**TAB.:** Empirical power of tests for CSR against a bell-shaped cluster with  $n_0 = 2$

| $r$ | Estimated power of the following : |                       |                    |
|-----|------------------------------------|-----------------------|--------------------|
|     | $\Lambda^{spac}$                   | $M\Lambda^{spac}$     | $MDig$             |
| 2   | 80.62 % (0.12)                     | <b>80.64 % (0.12)</b> | 78.97 % (0.12)     |
| 5   | 89.13 % (0.81)                     | <b>89.26 % (0.81)</b> | 88.01 % (0.53)     |
| 10  | 93.57 % (1)                        | 93.68 % (1)           | <b>96.34 % (1)</b> |

## Application à des données simulées

**TAB.:** Empirical power of tests for CSR against two bell-shaped clusters with  $n_0 = 2$

| $r$ | Estimated power of the following : |                       |                       |
|-----|------------------------------------|-----------------------|-----------------------|
|     | $\Lambda^{spac}$                   | $M\Lambda^{spac}$     | $MDig$                |
| 2   | 61.00 % (0.13)                     | <b>60.99 % (0.13)</b> | 60.59 % (0.11)        |
| 5   | 65.92 % (0.60)                     | 68.13 % (0.60)        | <b>71.05 % (0.57)</b> |
| 10  | 71.06 % (0.94)                     | 77.10 % (0.94)        | <b>83.09 % (0.96)</b> |

# *Plan de l'exposé*

*Introduction*

*1-Le cadre temporel*

*2-Le cadre spatial*

*Conclusion*

# *Plan de l'exposé*

Introduction

1-Le cadre temporel

*2-Le cadre spatial*

Conclusion

# *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  $\{s_1, \dots, s_n\}$ ,  $s_i \in D \subset \mathbb{R}^2$ .

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

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

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

Principal problème : agrégats potentiels en nombre infini.

Première solution :

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

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

# *Détection d'agrégats spatiaux*

## *Détection d'agrégats spatiaux*

Exemple : la "spatial scan statistic" (Kulldorff, 1997)

## *Détection d'agrégats spatiaux*

Exemple : la "spatial scan statistic" (Kulldorff, 1997)

- ➡ Agrégats potentiels : disques dont centre=point d'une grille.

## Détection d'agrégats spatiaux

Exemple : la "spatial scan statistic" (Kulldorff, 1997)

➤ Agrégats potentiels : disques dont centre=point d'une grille.



$$I_{scan}(m, A) = \left( \frac{m}{n\nu_f(A)} \right)^m \left( \frac{1 - m/n}{1 - \nu_f(A)} \right)^{n-m}.$$



## Détection d'agrégats spatiaux

Exemple : la "spatial scan statistic" (Kulldorff, 1997)

➤ Agrégats potentiels : disques dont centre=point d'une grille.



$$I_{scan}(m, A) = \left( \frac{m}{n\nu_f(A)} \right)^m \left( \frac{1 - m/n}{1 - \nu_f(A)} \right)^{n-m}.$$

Idée : utiliser plutôt

$$I_{spac}(m, A) = 1/B_{inc}(\nu_f(A), m - 1, n + 2 - m).$$

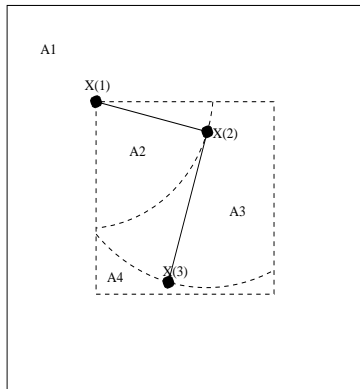
# *Détection d'agrégats spatiaux*

## *Détection d'agrégats spatiaux*

Deuxième solution : transformer les données.

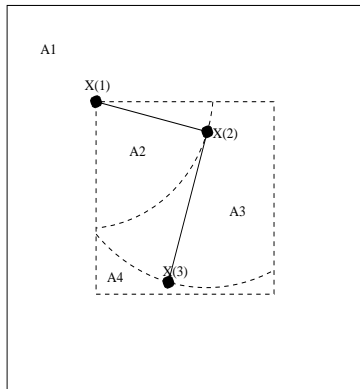
## Détection d'agrégats spatiaux

Deuxième solution : transformer les données.



## Détection d'agrégats spatiaux

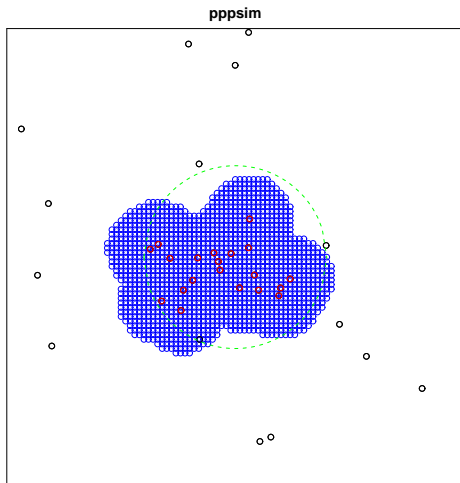
Deuxième solution : transformer les données.



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

# *Application à des données simulées*

# *Application à des données simulées*



# *Plan de l'exposé*

*Introduction*

*1-Le cadre temporel*

*2-Le cadre spatial*

*Conclusion*



# *Plan de l'exposé*

Introduction

1-Le cadre temporel

2-Le cadre spatial

*Conclusion*

# *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.