

# Test de la vraisemblance entre deux motifs de points

Application à une maladie sur une  
parcelle de vigne

Labenne Amaury, INRA, UMR 1065 SAVE, Villenave d'Ornon.

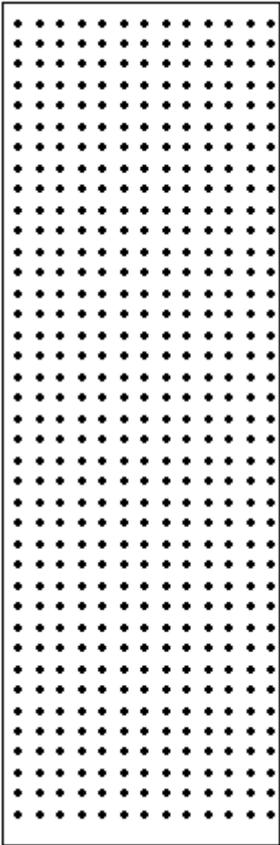
Bonneu Florent, Laboratoire de Mathématiques d'Avignon (EA 2151), Avignon.

# Introduction

- L'esca, une maladie du bois de la vigne.
- Pas de lien systématique entre les symptômes de la maladie et la mort du cep.
- Plusieurs questions :
  - Est-ce que les ceps meurent aux mêmes endroits où il y avait des ceps symptomatiques l'année passée ?
  - Est-ce que les ceps symptomatiques réapparaissent aux mêmes endroits de la parcelle d'une année à l'autre ?
- Nécessité de mettre en place un test pour comparer les densités des deux parcelles. Utilisation de  et du package *spatstat*.

# Notations

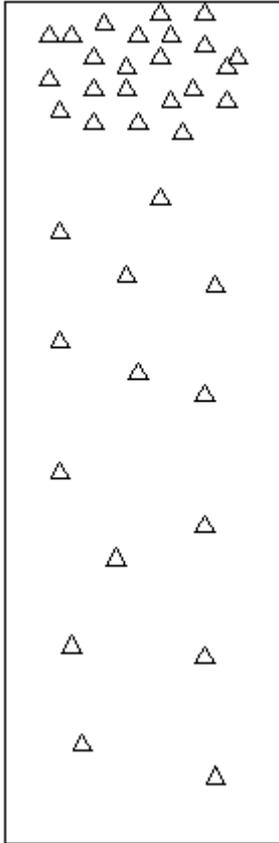
Carte de la parcelle



- Chaque cep à une position fixe  $i$ , on lui associe une variable d'état  $X_i$  où  $X = (X_1, \dots, X_n)$  est le vecteur des états des ceps.
- Si le cep  $i$  est malade, on note  $X_i = 1$  sinon  $X_i = 0$ .

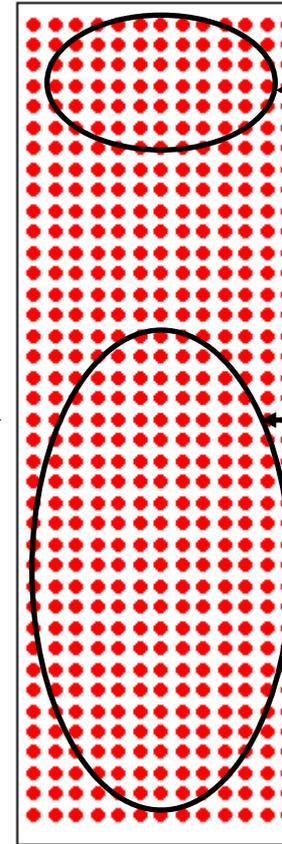
# Estimation des probabilités

Parcelle A



On estime la probabilité pour tous les ceps d'être symptomatiques par une méthode de noyaux (on choisit la taille du noyau  $H$ ).  
La probabilité est estimée grâce à la fonction *relrisk* du package *spatstat*.

Proba P



Zone avec une forte probabilité

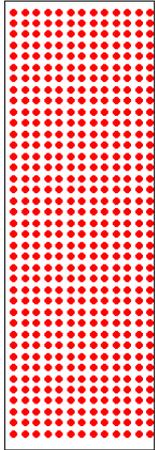
Zone avec une faible probabilité

Ici, on ne représente que les ceps symptomatiques, ie :  $X_i = 1$

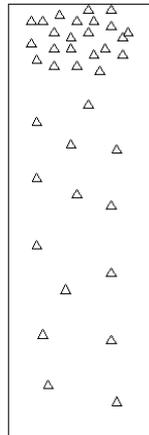
On obtient donc une probabilité  $p_i$  d'être symptomatique en chaque point de la parcelle.  $P = (p_1, \dots, p_n)$

# Principe du test et vraisemblance

Proba P



Parcelle A



$$V = \prod_{i=1}^n p_i^{X_i} (1 - p_i)^{1 - X_i}$$

Vraisemblance

Exemple :

○	△	△	○
△	○	△	△
○	△	○	○
○	○	○	○

p1	p2	p3	p4
p5	p6	p7	p8
p9	p10	p11	p12
p13	p14	p15	p16

△ : Ceps symptomatiques, i.e.,  $X_i = 1$

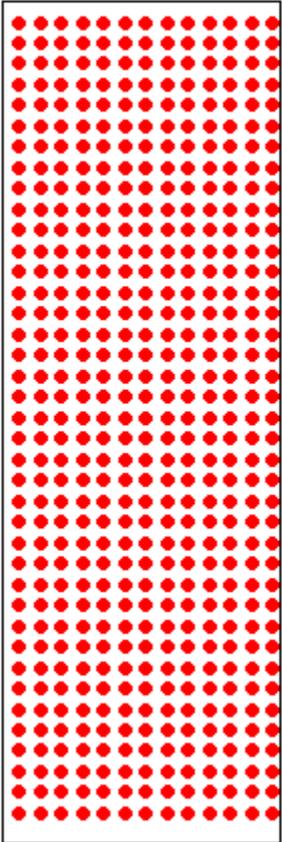
○ : Ceps sains, i.e.,  $X_i = 0$

Vraisemblance du motif :

$$V = p_2 \cdot p_3 \cdot p_5 \cdot p_7 \cdot p_8 \cdot p_{10} \cdot (1 - p_1) (1 - p_4) (1 - p_6) (1 - p_9) (1 - p_{11}) (1 - p_{12})$$

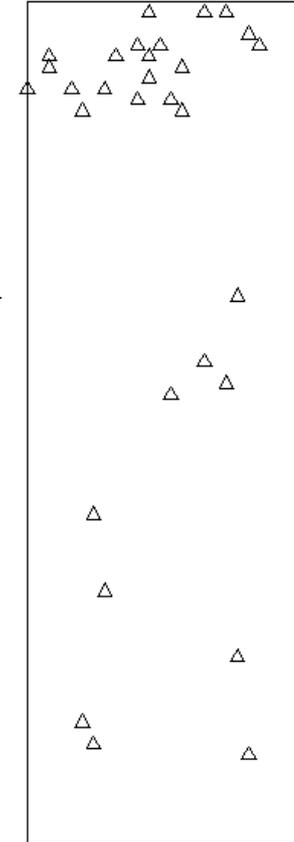
# Simulations de parcelles semblables à A

Proba P



A chaque position  $i$ , on applique une variable de Bernoulli de paramètre  $p_i$ . Ainsi, on obtient un nouveau motif A1 simulé, ressemblant au motif A.

Parcelle A1

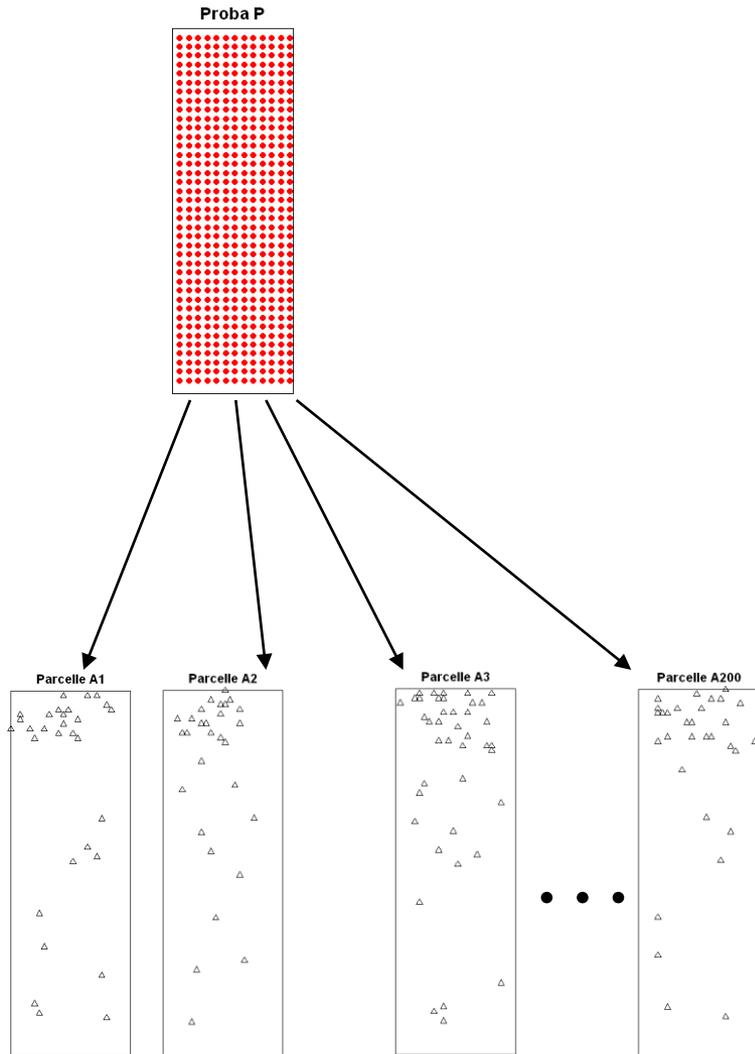


Vraisemblance

$$V = \prod_{i=1}^n p_i^{X_i} (1 - p_i)^{1 - X_i}$$

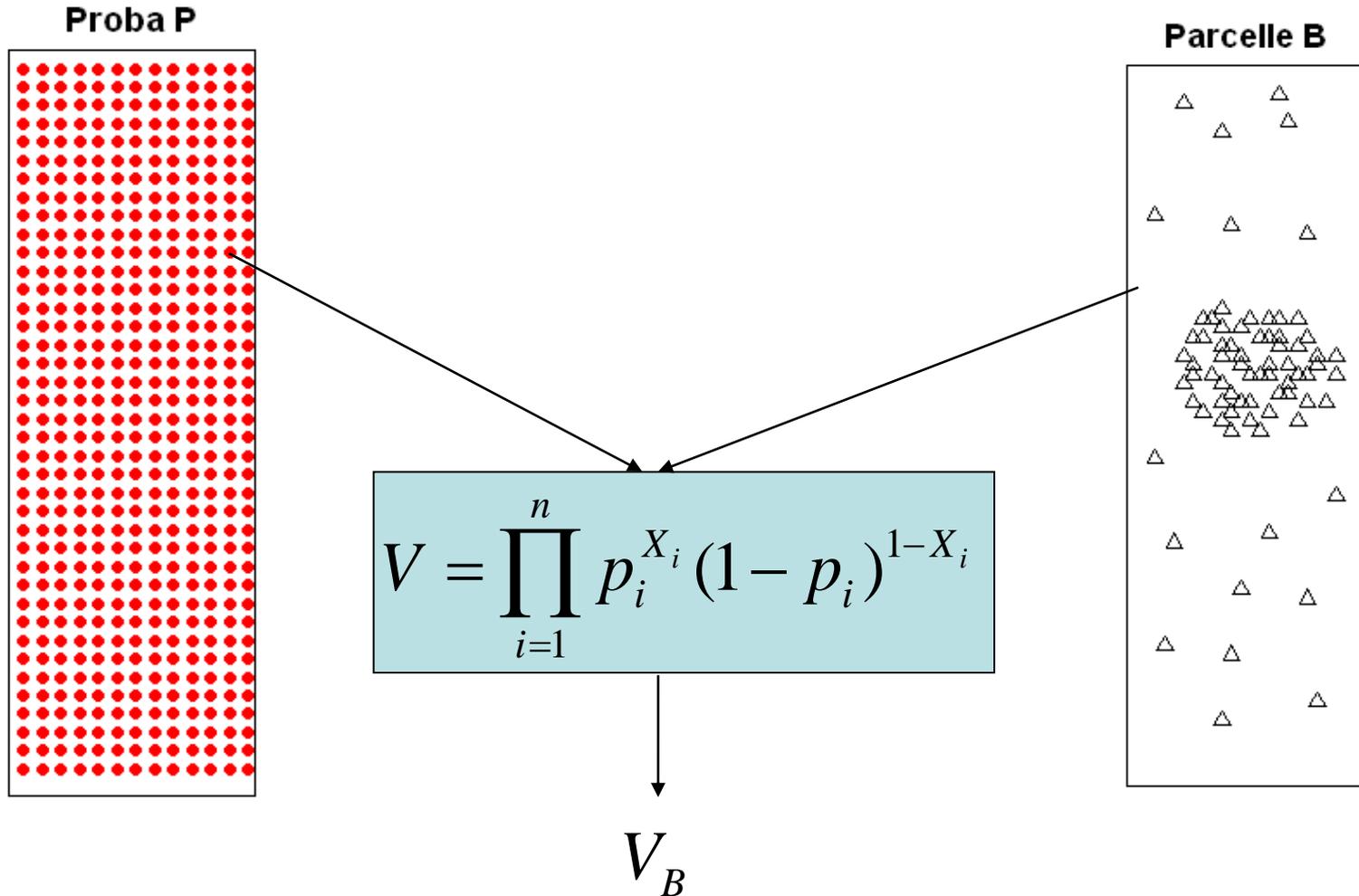
$V_{A_1}$

# Création de l'enveloppe de test



- On simule  $N_{sim}$  réalisations de la parcelle A grâce à la même carte de probabilités.
- Sur chacune des parcelles  $A_j$ , on calcule la vraisemblance  $V_{A_j}$ ,  $j \in [1, N_{sim}]$
- Après avoir rangé ces valeurs dans l'ordre croissant et éliminé  $\alpha = 10\%$  des valeurs extrêmes, on obtient le vecteur  $ENV = [S_{5\%}, \dots, S_{95\%}]$ , c'est notre enveloppe de test

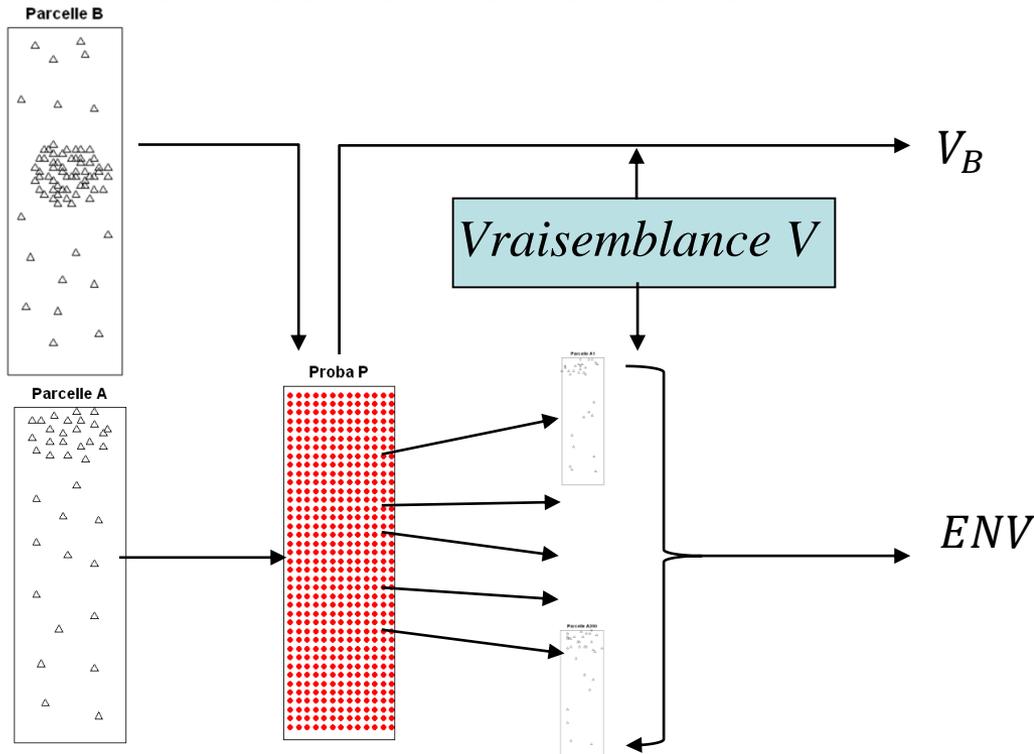
# Calcul de la vraisemblance sur la parcelle B



$V_B$  est notre Statistique de test

# Test et décision finale

- On teste l'hypothèse  $H_0$  : Les deux motifs de points A et B ne sont pas significativement différents.
- Pour cela, on regarde où se situe la valeur de  $V_B$  qui est notre statistique de test par rapport à l'enveloppe  $ENV$ , c'est une méthode de Monte-Carlo.

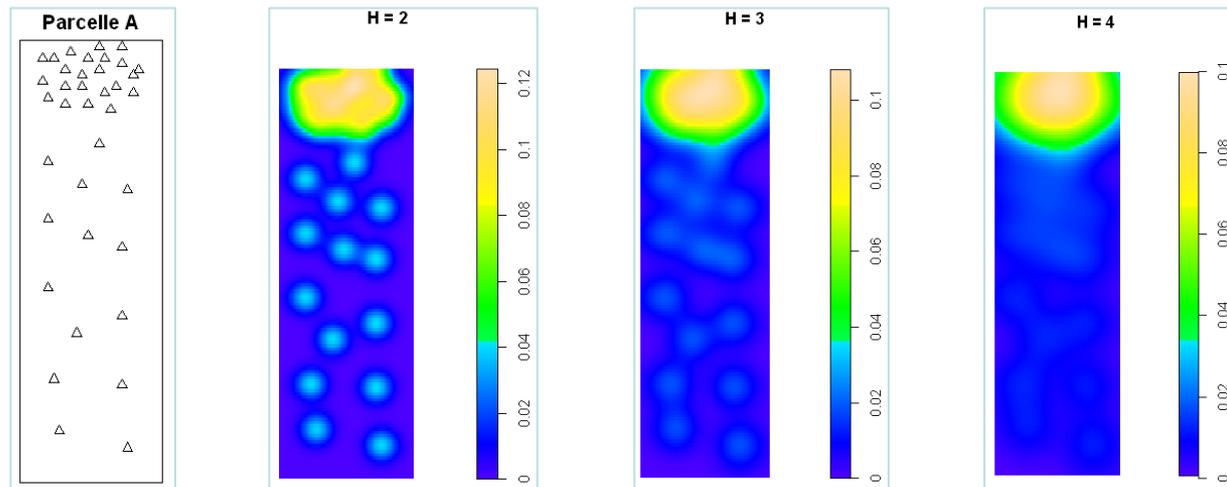


## Décision :

- Si  $V_B \in ENV$ , on ne rejette pas  $H_0$ .
- Si  $V_B \notin ENV$ , on rejette  $H_0$ , au risque de  $\alpha = 10\%$ .

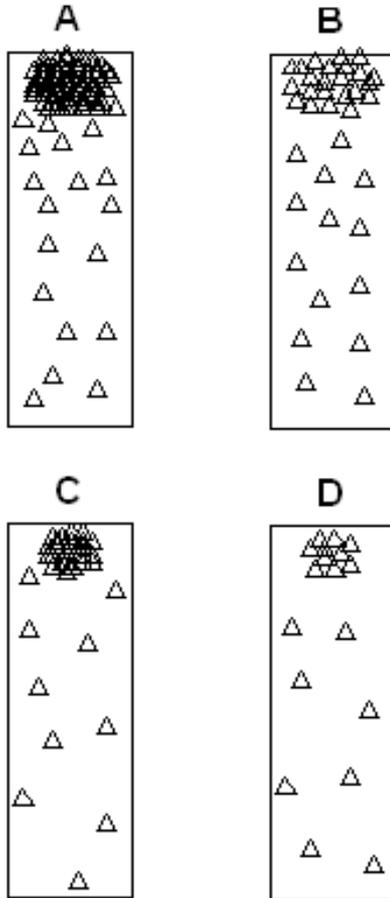
# Influence de la taille du noyau H

- Pour estimer la carte des probabilités, on utilise une méthode de noyaux. La taille du noyau H utilisé, à une influence sur la valeur des probabilités et donc sur les résultats.



- Plus H est grand, plus la carte des probabilités est « lissée ». Le test aura moins tendance à rejeter  $H_0$ .

# Résultats de quelques simulations

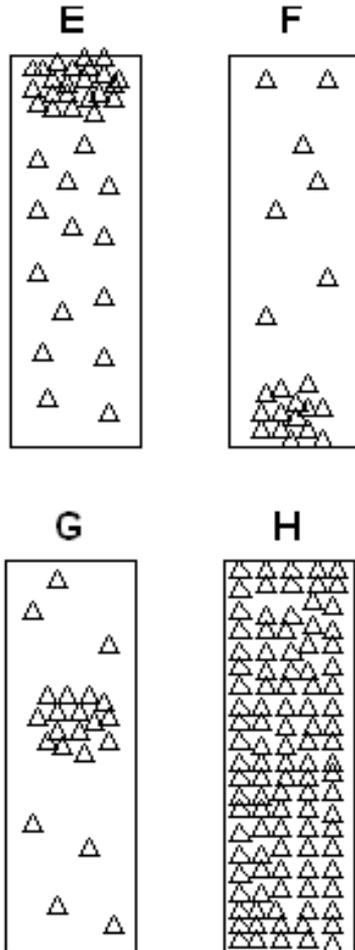


Pourcentages de rejet du test H=2				
P1(col)/P2(l)	A	B	C	D
A	4,8	58,4	11,6	31,6
B	7,2	4	7,4	9,6
C	99,5	93,6	5	51,1
D	98,5	88,5	38,2	5,3

Pourcentages de rejet du test H=3				
P1(col)/P2(l)	A	B	C	D
A	7,1	32,1	4,1	23,4
B	17,2	5,9	6,4	5,5
C	47,2	53,3	6,3	21,8
D	19,9	23,9	6	3,5

Pourcentages de rejet du test H=4				
P1(col)/P2(l)	A	B	C	D
A	4,5	23,2	3,5	17,8
B	15,3	3,7	5,2	4,9
C	9,3	29,5	4,9	12,2
D	5,6	9,3	5,3	4,5

# Résultats de quelques simulations



	Pourcentages de rejet du test H=2			
P1(col)/P2(l)	E	F	G	H
E	4	72,3	49,4	100
F	98,6	5	91,7	100
G	99,6	89,4	4,8	100
H	4,2	3,7	5,4	5,7

	Pourcentages de rejet du test H=3			
P1(col)/P2(l)	E	F	G	H
E	5,9	38,1	26	99,4
F	82,5	4,1	52,7	100
G	86	58,9	7	100
H	3,2	3,3	5,4	6,5

	Pourcentages de rejet du test H=4			
P1(col)/P2(l)	E	F	G	H
E	3,7	29,2	19,1	95,7
F	65,4	4	38	99,9
G	62,1	31,9	5,6	97,4
H	3,7	2,5	5,6	7,2

# Conclusion

- Choix de la taille du noyau  $H$  influence fortement les résultats.
- Test non symétrique, nécessité d'établir une règle de décision lorsqu'on croise les tests.
- Test non valable si le motif de départ est uniforme, en effet la probabilité est la même sur toute la parcelle.
- Test à améliorer ou à coupler avec un autre, mais qui peut être appliqué à de nombreuses problématiques impliquant des processus spatiaux sur grille.