

Statistique de Ripley et processus ponctuels

Gabriel Lang et Eric Marcon

UMR MIA,
UMR EcoFoG,
AgroParisTech

4 mai 2009

Plan

Analyse spatiale en écologie

Modèles de processus ponctuels

Statistique de Ripley et processus de Poisson homogène

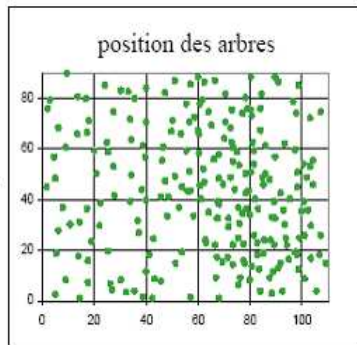
Perspectives

Méthodes d'analyses de répartitions spatiales en écologie

Motivations

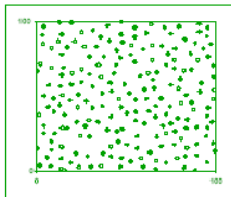
- ▶ Questions posées par les écologues :
répartitions spatiales / agrégats / interactions
- ▶ Développement des modèles de processus ponctuels sur le plan.
- ▶ Techniques de démonstration de dépendance faible

Données de peuplement forestier

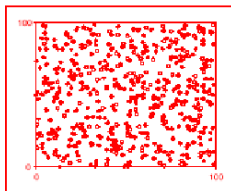


Types de répartitions spatiales

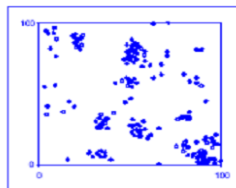
Régulière



Aléatoire

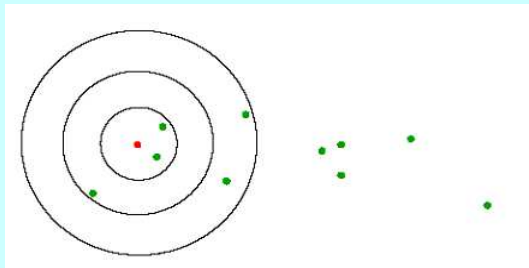


Agrégée



Mesure de l'agrégation

On calcule le nombre de voisins de chacun des points dans un cercle de rayon r . On fait varier r (statistique de Ripley). On compare avec la densité moyenne.



Statistique de Ripley

Soit A un pavé du plan et m la mesure de Lebesgue sur le plan :

Fonction K de Ripley

$$\hat{K}(r) = \frac{1}{m(A)} \sum_{x \neq y, \{x,y\} \subset S} \frac{\mathbb{1}\{\|x - y\| \leq r\}}{\rho(x)\rho(y)}$$

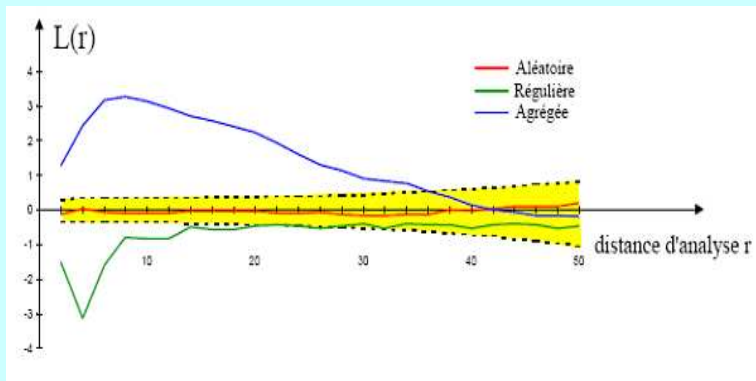
Si la répartition est "aléatoire" :

$$K(r) \simeq \pi r^2$$

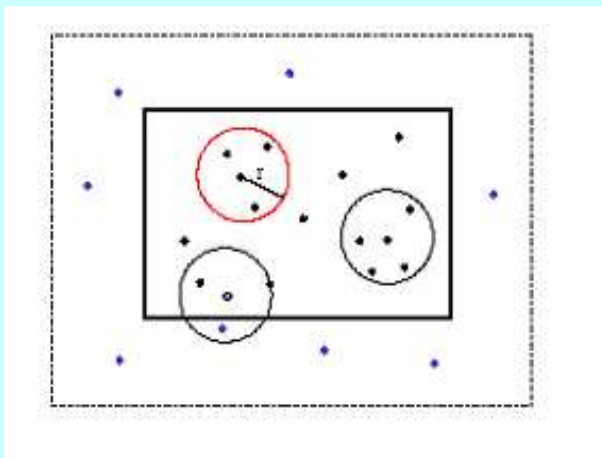
Fonction L de Ripley

$$\hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}} - r$$

Utilisation empirique de L



Effets de bord



Modèles de processus ponctuels

Processus de Poisson

Loi de Poisson $P(\lambda)$ définie sur \mathbb{N}

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Processus de Poisson homogène d'intensité ρ

- ▶ Pour tout rectangle B dans A , le nombre de points dans B suit une loi de Poisson de paramètre $\rho m(B)$ où $m(B)$ est l'aire de B .
- ▶ Si B et C sont deux rectangles sans intersection, les nombres de points dans B et C sont deux variables indépendantes.

Processus de Poisson inhomogène

Le processus est une modification du processus homogène dans lequel les points apparaissent préférentiellement dans certaines zones de A .

Processus de Poisson inhomogène d'intensité $\rho(x)$

Soit $\rho(x)$ une fonction intégrable définie sur A .

- ▶ Pour tout rectangle B dans A , le nombre de points dans B suit une loi de Poisson de paramètre $\rho(B)$, intégrale de la fonction $\rho(x)$ sur B .
- ▶ Si B et C sont deux rectangles sans intersection, les nombres de points dans B et C sont deux variables indépendantes.

Propriétés des processus de Poisson

- ▶ Ces processus sont à la base de la construction de la plupart des processus ponctuels.
- ▶ Ils présentent des agrégats (pr. inhomogène) ou pas (pr. homogène)
- ▶ La position des agrégats est fixe et dépend de la fonction d'intensité.
- ▶ Dans ces modèles les individus représentés par les points n'interagissent pas ; ils "ignorent" la position de leurs voisins.

Processus dépendants

On veut déformer la loi de probabilité du processus de Poisson homogène pour favoriser les agrégats (interaction attractive) ou la répartition régulière (interaction répulsive). On choisit une fonction de densité f par rapport au processus de Poisson définie sur un semis S par

$$f(S) = C \prod_{x \neq y, \{x, y\} \subset S} \Phi(\{x, y\})$$

Φ est une fonction d'interaction entre deux points, elle peut être répulsive ou attractive selon les distances entre les points

Exemples de processus définis par une densité

Processus de Strauss répulsif

$\gamma < 1$ et R est le rayon d'interaction

$$\Phi(\{x, y\}) = \gamma^{1\{\|x-y\| < R\}}$$

Les semis où deux points sont plus proches que R sont plus rares.

Processus à noyau dur

$$\Phi(\{x, y\}) = 1\{\|x - y\| > R\}$$

Les semis où deux points sont plus proches que R sont interdits.

Statistique de Ripley et processus de Poisson homogène

Intensité de deuxième ordre

X processus ponctuel sur \mathbb{R}^2 , intensité $\rho(x)$.

$X^{(2)}$ processus ponctuel sur $\mathbb{R}^2 \times \mathbb{R}^2 = \{\text{couples de points différents de } X\}$

intensité $\rho^{(2)}(x, y)$

Exemple : processus de Poisson (indépendance)

$$\rho^{(2)}(x, y) = \rho(x)\rho(y)$$

Mesure de dépendance :

$$g(x, y) = \rho^{(2)}(x, y) / \rho(x)\rho(y)$$

Fonction de Ripley

Si le processus est invariant par translation alors
 $g(x, y) = g(x - y)$.

$$\mathcal{K}(A) = \int_A g(x) dx$$

S'il est isotrope, on définit la fonction K de Ripley

$$K(r) = \mathcal{K}(B(0, r))$$

Estimateur de Ripley

A sous ensemble fini, régulier de \mathbb{R}^2 , m mesure de Lebesgue.
 S réalisation du processus

$$\hat{K}_A(r) = \frac{1}{m(A)} \sum_{X_i \neq X_j \in S} \frac{\mathbb{1}\{d(X_i, X_j) \leq r\}}{\hat{\rho}(X_i)\hat{\rho}(X_j)}$$

Hypothèse 1

X processus de Poisson homogène sur \mathbb{R}^2
intensité ρ connue.

Observations sur le carré $A_n = [0, n]^2$.

$$\hat{K}_{1,A_n}(r) = \frac{1}{n^2 \rho^2} \sum_{X_i \neq X_j \in S} \mathbb{1}\{d(X_i, X_j) \leq r\}$$

Hypothèse 2

X processus de Poisson homogène sur \mathbb{R}^2
intensité ρ inconnue.

Observations sur le carré $A_n = [0, n]^2$.

$$\hat{\rho} = \frac{\sqrt{N(N-1)}}{n^2}.$$

$$\hat{K}_{2,A_n}(r) = \frac{n^2}{N(N-1)} \sum_{X_i \neq X_j \in S} \mathbb{1}\{d(X_i, X_j) \leq r\}$$

Biais

Biais dus aux effets de bord :

$$\mathbb{E}\widehat{K}_{1,A_n}(r) - \pi r^2 = r^2 \left(-\frac{8r}{3n} + \frac{r^2}{2n^2} \right).$$

$$\begin{aligned} \mathbb{E}\widehat{K}_{2,A_n}(r) - \pi r^2 &= r^2 \left(-\frac{8r}{3n} + \frac{r^2}{2n^2} \right) \\ &\quad - r^2 e^{-\rho n^2} \left(\pi - \frac{8r}{3n} + \frac{r^2}{2n^2} \right) \left(1 + \rho n^2 e^{-\rho n^2} \right). \end{aligned}$$

Variance, intensité connue

$$\begin{aligned}\text{Var}(\widehat{K}_{1,A_n}(r)) &= \frac{1}{n^2} \left(\frac{2\pi r^2}{\rho^2} + \frac{4\pi^2 r^4}{\rho} \right) \\ &\quad - \frac{1}{n^3} \left(\frac{16}{3} \frac{r^3}{\rho^2} + \left(\frac{32\pi}{3} + \frac{1024}{45} \right) \frac{r^5}{\rho} \right) \\ &\quad + \frac{1}{n^4} \left(\frac{r^4}{\rho^2} + \left(\frac{59\pi}{12} + \frac{32}{9} \right) \frac{r^6}{\rho} \right).\end{aligned}$$

$$\text{Cov}(\widehat{K}_{1,A_n}(r), \widehat{K}_{1,A_n}(r')) = \frac{1}{n^2} \left(\frac{2\pi r^2}{\rho^2} + \frac{4\pi^2 r^2 r'^2}{\rho} \right) + O\left(\frac{1}{n^3}\right).$$

Variance, intensité inconnue

$$\begin{aligned} \text{Var}(\widehat{K}_{2,A_n}(r)) &= 2n^4 \mathbb{E} \left(\frac{\mathbb{1}\{N > 1\}}{N(N-1)} \right) (e_{r,n} - e_{r,n}^2) \\ &\quad + 4n^4 \mathbb{E} \left(\frac{\mathbb{1}\{N > 1\}(N-2)}{N(N-1)} \right) \mathbb{E}(h_1(U_j, r))^2 \\ &\quad + n^4 e^{-\rho n^2} (1 + \rho n^2) \left(1 - e^{-\rho n^2} - \rho n^2 e^{-\rho n^2} \right) e_{r,n}^2. \end{aligned}$$

$$\text{Cov}(\widehat{K}_{2,A_n}(r), \widehat{K}_{2,A_n}(r')) = \frac{2\pi r^2}{\rho^2 n^2} + O\left(\frac{1}{n^3}\right).$$

avec $e_{r,n} = r^2 \left(\pi - \frac{8r}{3n} + \frac{r^2}{2n^2} \right)$ et $\mathbb{E}(h_1(U_j, r))^2$ d'ordre $(r/n)^5$

Théorème de la limite centrale

Théorème

Pour (r_1, \dots, r_d) vecteur de réels positifs, et

$\mathcal{K}_n = (\widehat{K}_{A_n}(r_1), \dots, \widehat{K}_{A_n}(r_d))$:

$$n\sqrt{\rho}(\mathcal{K}_n - \pi(r_1^2, \dots, r_d^2)) \longrightarrow \mathcal{N}(0, \Sigma)$$

- ▶ Hypothèse 1 : $\Sigma_{s,t} = \frac{2\pi(r_s^2 \wedge r_t^2)}{\rho} + 4\pi^2 r_s^2 r_t^2$
- ▶ Hypothèse 2 : $\Sigma_{s,t} = \frac{2\pi(r_s^2 \wedge r_t^2)}{\rho}$

Statistique de test

$T = \Sigma^{-1/2} \mathcal{K}_n$ est asymptotiquement $\mathcal{N}(0, I_d)$

Sous l'hypothèse

H_0 : X est un processus de Poisson homogène d'intensité ρ .

Statistique de test : $T^2 = \sum T_i^2$

Zone de rejet de niveau α :

$$T^2 > \chi_\alpha^2(d).$$

$\chi_\alpha^2(d)$ quantile de la distribution $\chi^2(d)$.

Méthodes de preuve

1. Calcul intégral explicite pour le biais et la variance (pas pour la covariance)
2. Méthode de blocs pour le théorème central limite; les couples de points ne sont pas indépendants mais $2r$ -dépendants.

Travail restant à faire

1. Simulations pour valider le test et choix de l'approximation de la variance (en cours)
2. Processus de Poisson inhomogène : statistique de Ripley inadaptée
3. Processus dépendant : processus défini par une densité (e.g. Strauss), propriétés de l'estimateur et du test sous cette contre-hypothèse.

Processus de Poisson inhomogène

La statistique de Ripley est inadaptée

Elle utilise les proches voisins pour estimer l'intensité locale $\rho(x)$ par une régularisation à noyau puis une deuxième fois pour estimer les interactions.

Modification de la statistique de Ripley

s'il existe une population de référence d'intensité proportionnelle pour estimer $\rho(x)$, indépendante de la population dont on veut tester les agrégats.

Dépendance pour un processus ponctuel : mélange

Soit α le coefficient de mélange fort entre deux tribus \mathcal{F}_1 et \mathcal{F}_2 :

$$\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup\{|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$$

On définit un coefficient de mélange pour le processus X . Soit \mathcal{F}_A^X la tribu engendrée par X sur A . Soit $s > 0$ et $a > 0$

$$\alpha_a(s) = \sup\{\alpha(\mathcal{F}_A^X, \mathcal{F}_B^X)\}$$

sup sur tout A compact et convexe tel que $m(A) \leq a$, et tout B translaté de A , $B = A + x$, $x \in \mathbb{R}^2$ tel que $d(A, B) \leq s$.

Dépendance faible pour les processus ponctuels

X est un processus ponctuel ι -dépendant si pour tout u et v entiers, toute collection $(A_i)_{i=1,\dots,u}$, $(B_j)_{j=1,\dots,v}$ de sous-ensembles convexes compacts de A tels que la distance entre $\cup_{i=1,\dots,u} A_i$ et $\cup_{j=1,\dots,v} B_j$ est supérieure à r alors

$$\text{Cov} \left(\prod_{i=1,\dots,u} \mathbb{1}(N(A_i) \neq 0), \prod_{j=1,\dots,v} \mathbb{1}(N(B_j) \neq 0) \right) \leq (u + v)\iota_r.$$

Adaptation des preuves au processus de Strauss

- ▶ Prouver une relation entre fonction de potentiel markovien et faible dépendance
- ▶ Termes supplémentaires dans le calcul de la variance
- ▶ Méthode des blocs en faible dépendance pour le théorème de la limite centrale.

Relation entre dépendance et densité

$$\begin{aligned}
 & \mathbb{E}h(X_s) \\
 &= \sum_{k=0}^{\infty} \frac{\exp(-\rho n^2)}{k!} \int_{A_n} \dots \int_{A_n} h(\{x_1, \dots, x_k\}) f_k(\{x_1, \dots, x_k\}) dx_1 \dots dx_k. \\
 & \neq \sum_{i,j,k,l} \text{Cov}(h(X_i, X_j, r), h(X_k, X_l, r)) \\
 &= \sum_{k=0}^{\infty} \frac{\exp(-\rho n^2)}{k!} \int_{A_n} \dots \int_{A_n} \sum_{i,j,k,l}^{\neq} h(X_i, X_j, r) h(X_k, X_l, r) \\
 & \quad (f_4(x_i, x_j, x_k, x_l) - f_2(x_i, x_j) f_2(x_k, x_l)) dx_i dx_j dx_k dx_l.
 \end{aligned}$$

Bibliographie

- [1] Moller, J., Waagepetersen, R.(2004) *Statistical inference and simulation for spatial point processes*, Chapman and Hall.
- [2] Ripley, B. D. (1976) The second-order analysis of stationary point processes. *J. Appl. Probab.* 13, 255-266.
- [3] Ripley, B. D. (1979) Tests of 'randomness' for spatial point patterns. *J. Roy. Statist. Soc. B* 41, 368-374.