Statistiques Spatiales

Eric Marcon et Stéphane Traissac

Version : 2018-09-16

Ce document est réalisé de façon dynamique et reproductible grâce à :

- LaTeX, dans sa distribution Miktex (http://miktex.org/) et la classe memoir (http://www.ctan.org/pkg/memoir).
- R (http://www.r-project.org/) et RStudio (http://www.rstudio.com/)
- bookdown. (http://bookdown.org/)

UMR Écologie des forêts de Guyane
http://www.ecofog.gf

Les opinions émises par les auteurs sont personnelles et n'engagent ni l'UMR EcoFoG ni ses tutelles.

Photographie en couverture : Hadrien Lalagüe

Table des matières

Introduction

Pourquoi étudier les structures spatiales ?

L'analyse des structures spatiales peut être considérée tout d'abord comme un outil statistique élémentaire. Quand on observe une carte sur laquelle sont placés des objets, ils ne semblent en général pas répartis au hasard. On peut voir des variations de densité, sous la forme de gradients ou d'agrégats. Pour décrire un phénomène quantitatif à partir d'un certain nombre d'observations, disons le diamètre moyen des arbres d'un peuplement, on utilise classiquement l'outil fourni par la théorie des variables aléatoires : on mesure un certain nombre d'arbres, on considère que leur diamètre moyen est un estimateur de l'espérance de la variable aléatoire qu'est le diamètre de chaque arbre, et on calcule la variance de l'échantillon pour avoir une idée de la variabilité des diamètres. Si on s'intéresse maintenant à la répartition spatiale de ces arbres, la quantité d'information disponible est bien plus importante : la densité des arbres, leur diamètre moyen selon le lieu et les interactions éventuelles des arbres entre eux. Les outils mathématiques nécessaires sont les processus ponctuels, généralisation spatialisée (généralement à deux dimensions) des variables aléatoires.

La question fondamentale consiste à caractériser clairement ce qu'on observe. Prenons l'exemple d'un peuplement forestier dans lequel les arbres d'une espèce sont apparemment regroupés et forment des agrégats. Deux causes peuvent en être responsables : l'hétérogénéité du milieu favorise la présence de cette espèce localement, la densité des arbres est alors variable, ou bien les arbres sont regroupés par des processus d'agrégation (par exemple une faible dispersion des graines) indépendamment du milieu qu'on supposera dans ce cas homogène. Les mécanismes écologiques sont extrêmement différents dans les deux cas. Mais la distinction entre hétérogénéité et agrégation est formellement impossible sur un jeu de données unique : des hypothèses de travail sont indispensables. Cette difficulté est probablement la raison pour laquelle de nombreuses méthodes statistiques ont été mises en place pour détecter les structures spatiales. Une revue complète en sera faite (chapitre TODO), la famille d'outils la plus efficace sera étudiée en détail (chapitre TODO) ainsi que ses fondements mathématiques (les processus ponctuels, chapitre 1) et de nouveaux outils seront proposés pour repousser les limites actuelles. La relation directe entre les processus ponctuels et les phénomènes écologiques sous-jacents est difficile, à cause de la simplification excessive mais surtout parce que la réalisation d'un processus observée sous la forme d'un semis de points est instantanée et unique alors qu'un peuplement forestier est le résultat d'une histoire et de processus successifs.

La théorie des processus ponctuels est indispensable pour comprendre les méthodes présentées ici. Une présentation en est faite dans premier chapitre. Le deuxième chapitre présente l'état de l'art en termes d'analyse spatiale des processus ponctuels suivi de l'introduction de nouveaux outils permettant d'étendre son champ d'application. L'objectif est de permettre la caractérisation des structures spatiales dans un cadre théorique aussi vaste que possible et répondant aux exigences des données réelles, comme la prise en compte de l'hétérogénéité.

Chapitre 1

Processus classiques

Les processus ponctuels fournissent le cadre mathématique nécessaire à l'étude des structures spatiales. Une définition mathématique est nécessaire avant de présenter les processus les plus courants, dont le processus de Poisson. L'approche utilisée classiquement par les non-mathématiciens est locale : les propriétés d'un processus sont définies autour de chaque point. Elle a l'avantage d'être concrète et facilement compréhensible. Son inconvénient est de laisser un certain flou sur le comportement global du processus. Une définition globale est nécessaire pour aller plus loin : c'est celle que nous présenterons en première partie de ce chapitre. La définition locale est présentée en-suite pour les lecteurs rebutés par les équations.

Ensuite, nous analyserons en détail quelques processus classiques, choisis pour leur utilité en écologie puis les méthodes nécessaires à leur simulation.

1.1 Définitions

Processus ponctuel

Nous nous placerons dans une partie du plan, typiquement une placette forestière, notée A, $A \subset R^2$. Les définitions données ici et la plupart des résultats sont valables dans un espace de dimension quelconque finie mais on se limitera en pratique à un espace à deux dimensions. D'autre part, travailler sur une placette de taille infinie n'a pas d'intérêt pratique, mais rien ne s'y opposerait sur le plan mathématique, sauf dans certains cas qui seront précisés dans le texte. A sera appelée aire ou zone ou fenêtre d'étude.

Nous nous intéresserons à des ensembles dénombrables de points (on dira aussi "semis de points"). Les points seront notés en minuscules, les ensembles en majuscules : $X \subset R^2$, $x \in X$. Le semis de point X sera généralement défini sur tout le plan, et son nombre de points sera infini. À l'intérieur de l'aire d'étude A, un sous ensemble de X noté X_A sera observé et cartographié : $X_A = X \cap A$.

Nous ne nous intéresserons qu'à des ensembles de points localement finis, c'est-à-dire tels que leur nombre de points dans A soit fini : $n(A) < \infty$ pour A borné. Cette restriction n'a pas de conséquences pratiques.

Il est impossible de définir directement une fonction qui attribuerait à chaque semis la probabilité de le tirer, parce que la probabilité de chaque semis est nulle. On passe donc par des ensembles de semis de points, dont la probabilité n'est pas nulle. L'ensemble des semis de points localement finis est noté N_f. N_f peut être équipé d'une σ-algèbre $\mathcal{N}_f$. Les éléments F de $\mathcal{N}_f$ sont des ensembles de semis de points localement finis. $\mathcal{N}_f$ est une tribu (ou σ-algèbre) de N_f :

- l'ensemble vide appartient à $\mathcal{N}_f$;
- pour tout $F \in \mathcal{N}_f$, le complémentaire de F dans $\mathcal{N}_f$ appartient aussi à $\mathcal{N}_f$;
- l'union de $(F_1, F_2, F_3, ...)$, une suite dénombrable d'éléments de $\mathcal{N}_f$ appartient aussi à $\mathcal{N}_f$.

$(N_f, \mathcal{N}_f)$ est un espace probabilisable. Il reste à le doter d'une mesure P, la probabilité de tirer un ensemble de semis de points particulier F. Cet ensemble est défini par le nombre de points tirés, $N(A)$, pour tout A borné. P est une mesure car :

- $P(\emptyset) = 0$
- P est σ-additive : $P(\bigcup_{i=1}^{n} F_i) = \sum_{i=1}^{n} P(F_i)$, où $\bigcup_(i = 1)^n F_i$ est une union disjointe.

P est une probabilité car on fixe $P(\mathcal{N}_f) = 1$.

Un processus ponctuel est défini comme une application d'un espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$ vers $(N_f, \mathcal{N}_f)$ tel que $N(A)$ est une variable aléatoire pour tout A borné. La définition précise de (Ω, F, P) importe peu puisqu'il est impossible à utiliser directement. Les processus sont notés en lettres grecques majuscules, par exemple Ξ. Un semis de points X est une réalisation de Xi. On note $P(X \in F)$ la probabilité que le tirage de Ξ soit un élément d'un ensemble de semis de points F particulier, par exemple défini par son nombre de points. Enfin, la surface de A sera notée $\|A\|$.

Propriété de premier ordre

Soit S une partie de A. La propriété de premier ordre $\mu(S)$, appelée également mesure d'intensité du processus Ξ, est l'espérance du nombre de points dans S :

$$\mu(S) = \mathbb{E}(N(S)) \tag{1.1}$$

Dans tous les cas que nous traiterons, la mesure d'intensité pourra être écrite comme l'intégrale d'une fonction d'intensité λ :

$$\mu(S) = \int_S \lambda(x)dx \tag{1.2}$$

Propriété de second ordre

La mesure du moment factoriel de second ordre de deux parties de A, S_1 et S_2, est l'espérance du nombre de paires de points du processus Ξ se trouvant respectivement dans S_1 et S_2 :

$$\mu_2(S_1, S_2) = \mathbb{E}\left(\sum_{x_1, x_2 \in X, x_1 \neq x_2} 1(x_1 \in S_1, x_2 \in S_2)\right) \tag{1.3}$$

De même, cette mesure pourra être écrite comme l'intégrale de λ_2, appelée densité du produit de second ordre :

$$\mu_2 S_1, S_2) = \iint_{(R^2)^2} 1(x_1 \in S_1, x_2 \in S_2)\lambda_2(x_1, x_2)dx_1dx_2 \tag{1.4}$$

On peut montrer[1] que :

$$\mathbb{E}(N(S_1)N(S_2)) = \lambda_2(S_1, S_2) + \lambda(S_1 \cap S_2)$$

1. J. Møller et R. P. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. T. 100. Chapman et Hall, p. 1-300.

1.2 Définition locale

Un processus ponctuel est l'équivalent d'une variable aléatoire dont le résultat est un ensemble de points noté X, dans un ensemble de réalisations possibles, qui sera toujours ici une surface connue et délimitée notée A.

On utilise les processus ponctuels comme outils mathématiques pour caractériser et éventuellement modéliser des événements dont on connaît la répartition spatiale, par exemple les arbres dans une forêt. Une façon intéressante de décrire un processus ponctuel dont on ne connaît pas la loi consiste à utiliser ses propriétés de premier ordre et de second ordre.

1.3 Processus de Poisson homogène

Le processus de Poisson est un processus stationnaire et isotrope, dont la réalisation donne des points à la position complètement aléatoire. Inversement, un processus ponctuel complètement aléatoire est un processus de Poisson.[2]

2. démonstration : P. J. Diggle (1983). *Statistical analysis of spatial point patterns*. London : Academic Press, p. 1-148, pp. 51-52.

Le processus de Poisson joue un rôle central en statistiques spatiales, à la fois parce que c'est le plus simple donc celui dont les propriétés ont été le mieux étudiées, et aussi parce qu'il constitue généralement le modèle nul contre lequel des semis de points peuvent être testés. C'est un processus à accroissements indépendants, il joue le rôle des marches aléatoires pour les séries temporelles à temps discret et du mouvement Brownien pour les séries à temps continu. Ces propriétés sont utilisées pour le calcul de l'intervalle de confiance de la fonction K de Ripley.

Simulation

La simulation d'un processus de Poisson (figure 1.1) est faite par la commande `rpoispp` de *spatstat*.

```
X <-rpoispp(lambda=1, win=as.owin(c(0, 10, 0, 10)))
plot(X, main="")
axis(1) ; axis(2)
```

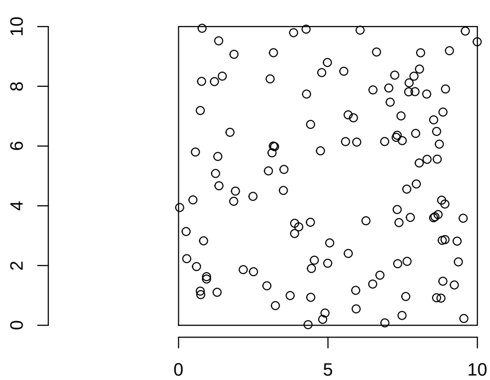

Fig. 1.1 : Simulation d'un processus de Poisson homogène d'intensité 1

Les arguments de la fonction sont l'intensité du processus λ (cf. *Propriété de premier ordre*) et la fenêtre A, ici un carré dont les coordonnées des sommets sont listées.

Propriété de premier ordre

Une réalisation d'un processus de Poisson de paramètre λA sur l'aire d'étude A est un semis de points complètement aléatoire d'intensité λ. Le nombre de points suit une loi de Poisson de paramètre λA, c'est-à-dire que :

$$P(n(A) = k) = e^{-\lambda A}\frac{(\lambda A)^k}{k!}. \tag{1.5}$$

Bibliographie

DIGGLE, P. J. (1983). *Statistical analysis of spatial point patterns.* London : Academic Press, p. 1-148 (cf. p. 4).

MØLLER, J. et R. P. WAAGEPETERSEN (2004). *Statistical Inference and Simulation for Spatial Point Processes.* T. 100. Chapman et Hall, p. 1-300 (cf. p. 3).

Résumé : Résumé en Français, en quatrième de couverture.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Mots clés : Mot clé en Français, En liste.

Abstract: English abstract, on the last page.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Keywords: Keyword in English, As a list.