

# REVUE DE STATISTIQUE APPLIQUÉE

J. VAILLANT

## Échantillonnage et étude statistique de populations en milieu hétérogène

*Revue de statistique appliquée*, tome 40, n° 4 (1992), p. 15-26

[http://www.numdam.org/item?id=RSA\\_1992\\_\\_40\\_4\\_15\\_0](http://www.numdam.org/item?id=RSA_1992__40_4_15_0)

© Société française de statistique, 1992, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## ÉCHANTILLONNAGE ET ÉTUDE STATISTIQUE DE POPULATIONS EN MILIEU HÉTÉROGÈNE

J. Vaillant

*I.N.R.A., Laboratoire de Biométrie, 78000 Versailles, France*

### RÉSUMÉ

En écologie, la répartition spatiale d'individus n'ayant ni attraction, ni répulsion mutuelle mais des affinités communes avec des facteurs de l'environnement peut engendrer des lois de probabilité usuelles pour le dénombrement d'individus dans les unités d'échantillonnage. Cependant, le type d'échantillonnage pratiqué et l'existence de zones d'hétérogénéité peuvent conduire à des vraisemblances d'échantillon plus complexes. Dans cet article, nous analysons certaines de ces situations et discutons l'approche statistique permettant d'étudier l'hétérogénéité de l'environnement.

**Mots-clés** : *Echantillonnage, processus spatial de COX, vraisemblance, loi de Poisson composée*

### ABSTRACT

When individuals of an ecological population do not influence each other but have common affinity to environment factors, this can generate classical probability distributions for the counts of individuals in sampling units. However, because of the sampling procedure and existence of heterogeneity zones, the sample likelihood may be more complex. In this paper, such situations are examined and the statistical approach for analysing the environment heterogeneity is discussed.

**Key-words** : *Sampling, spatial COX process, likelihood, compound Poisson distribution*

### 1. Introduction

L'étude de la répartition spatiale d'une population écologique est souvent faite à travers une observation partielle de cette population pour des raisons évidentes de coût d'échantillonnage. Pour des individus immobiles (par exemple, des plantes, des arbres, des pontes d'insectes), il existe deux principaux types de plans d'échantillonnage. Le premier consiste à mesurer des distances : à partir de points arbitraires de la région à étudier (dits points d'échantillonnage), on calcule

la distance au plus proche individu de la population. On peut ensuite mesurer la distance au plus proche voisin de cet individu, et aller ainsi de suite jusqu'au  $k$ ème voisin (Diggle, Cox (1983)). Ces méthodes furent introduites par Cottam et Curtis (1949) et sont souvent appliquées en recherches forestières. Les plans du second type sont basés sur des dénombrements d'individus à l'intérieur d'unités d'échantillonnage disjointes (on dira par la suite des quadrats), cette façon de faire s'avérant la seule praticable dans de nombreux cas (Rogers (1974), chap. 9; Cliff, Ord (1981), chap. 4). Quand les individus sont distribués indépendamment les uns des autres et que le milieu est homogène (c'est-à-dire équiprobabilité d'occupation de n'importe quel point de la région étudiée), leur distribution spatiale peut être décrite par un processus de Poisson homogène. Les nombres d'individus dans les quadrats sont alors des réalisations indépendantes d'une même loi de Poisson. Quand le milieu est hétérogène, la distribution des individus peut être décrite par un processus spatial de Cox  $N(\cdot)$  (voir Diggle (1983), chap. 4 §7). Ce type de processus correspond à un processus de Poisson à intensité aléatoire. En notant  $X$  la région de  $\mathbb{R}^2$  étudiée et  $\lambda(x)$  la valeur de l'intensité au point  $x$  de  $X$ , alors  $(\lambda(x))_{x \in X}$  est un processus stochastique, dit processus intensité, dont la loi détermine celle du processus  $N(\cdot)$  (Daley, Vere-Jones (1988), chap. 8 §5). Le processus intensité reflète la variabilité de l'environnement. Conditionnellement à cet environnement, les individus sont répartis indépendamment les uns des autres. En termes plus probabilistes, ceci équivaut à dire que, sachant la réalisation du processus  $(\lambda(x))_{x \in X}$ ,  $N(\cdot)$  est un processus de Poisson.

Dans cet article, nous nous intéressons à l'étude statistique de l'hétérogénéité du milieu écologique dans le cas où elle se traduit par des effets aléatoires  $A_1, \dots, A_k$  associés respectivement à des sous-régions  $X_1, \dots, X_k$  de  $X$  et tels que :

$$\forall x \in X, \lambda(x) = \sum_{i=1}^k A_i 1(x \in X_i) \quad (1)$$

où  $1(\zeta)$  vaut 1 si  $\zeta$  est vrai, et vaut 0 sinon.

Les  $X_i$  sont donc des zones d'hétérogénéité pouvant se recouper, et ne recouvrant pas forcément la région  $X$ . Une illustration est présentée en Fig. 1. Un exemple biologique pourrait être des prédateurs se répartissant en fonction des nombres  $A_i$  de proies de type  $i$ , ces dernières étant réparties sur leurs territoires respectifs  $X_i$ , avec chevauchements possibles des  $X_i$ .

Dans ce qui suit, on s'intéressera essentiellement aux problèmes liés à l'étude de l'hétérogénéité de telles distributions d'individus par le biais de tests de vraisemblance. Les propriétés distributionnelles du processus de Cox dont l'intensité est fournie par (1) ont été étudiées par Vaillant (1991) dans le cas d'effets gamma-distribués et d'ensembles  $X_i$  non nécessairement disjoints. Un rappel sur les résultats essentiels est fait au paragraphe 2. On souligne que l'aléa sur les effets  $A_i$  a pour conséquence une surdispersion des individus (plus forte variabilité du nombre d'individus d'un quadrat à l'autre par rapport à un processus de Poisson à intensité constante) et, en cas d'indépendance des  $A_i$ , une autocorrélation spatiale positive pour le nombre d'individus par unité de surface.

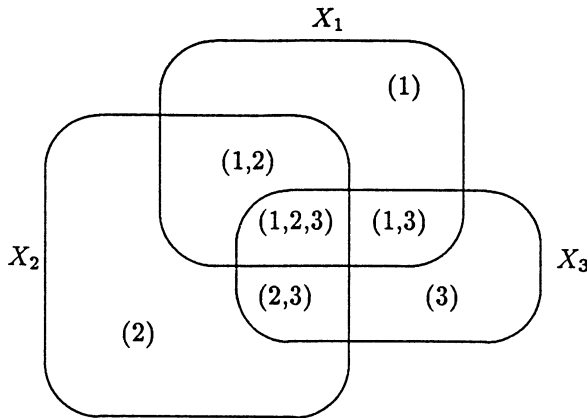


FIGURE 1

Exemple de région  $X$  ayant trois zones d'hétérogénéité  $X_1$ ,  $X_2$ ,  $X_3$  (tiré de Vaillant (1991)). Une partition de  $X$  peut être construite à partir des intersections entre les  $X_1$ ,  $X_2$ ,  $X_3$ . L'(les) effet(s) associé(s) à tout élément de cette partition est (sont) indiqué(s) entre parenthèses.

Au paragraphe 3, nous discutons deux approches pour tester l'homogénéité de répartitions spatiales d'individus sous le modèle défini par (1). La première approche consiste à se placer conditionnellement aux effets  $A_i$  et permet l'utilisation de vraisemblances d'échantillon poissonniennes. Dans la seconde approche, on cherche à tirer parti de la connaissance présumée de la loi des  $A_i$ . La difficulté d'une telle approche est discutée dans le cas d'effets gamma-distribués. On montre, en particulier que les conditions d'explicitation de la vraisemblance d'échantillon, même dans le cas d'effets  $A_i$  indépendants et identiquement distribués, sont liées aux types d'intersection entre les  $X_i$ . Ainsi, dans bien des cas, les méthodes de vraisemblance ne peuvent être appliquées qu'à travers la première approche, dite approche conditionnelle.

Dans tout ce qui suit,  $X$  est un ensemble borné de  $\mathbb{R}^2$  et l'aire de tout sous-ensemble  $Q$  de  $X$  sera noté  $|Q|$ .

## 2. Propriétés distributionnelles engendrées par les zones d'hétérogénéité

Notons  $\mathcal{B}(X)$ , l'ensemble des boréliens de  $X$ .

Un processus de Cox  $N(\cdot)$  dont le processus intensité est défini par (1) possède les propriétés suivantes :

$\forall B \in \mathcal{B}(X)$ ,  $N(B)$  suit, conditionnellement aux  $A_i$ , la loi de Poisson

$$\mathcal{P}\left(\sum_{i=1}^k A_i \mid B \cap X_i\right). \quad (2)$$

Puisque les  $A_i$  sont aléatoires,  $N(B)$  suit donc une loi de Poisson composée (Johnson, Kotz (1969), chap. 8 §2). Ces auteurs notent les lois composées de la manière symbolique suivante

$$F_1 \underset{\theta}{\wedge} F_2$$

où  $F_1$  représente la loi originale (loi conditionnelle à  $\theta$ ),  $\theta$  les paramètres aléatoires et  $F_2$  la loi de ces paramètres (loi mélangeante).

L'expression (2) peut donc s'écrire

$$\bigotimes_{i=1}^k \mathcal{P}(A_i | B \cap X_i) \underset{A_i}{\wedge} F_i, \quad (2')$$

$F_i$  étant la loi de  $A_i$ , et  $\bigotimes$  l'opérateur produit de convolution de lois de probabilité.

De plus, on démontre aisément que

$$\forall (B, B^*) \in \mathcal{B}(X)^2 \text{ avec } B \neq B^*,$$

$$\begin{aligned} \text{Var}(N(B)) &= E(N(B)) + \sum_{i=1}^k \sum_{j=1}^k |B \cap X_i| |B \cap X_j| \text{Cov}(A_i, A_j) \\ \text{Cov}(N(B), N(B^*)) &= \sum_{i=1}^k \sum_{j=1}^k |B \cap X_i| |B^* \cap X_j| \text{Cov}(A_i, A_j). \end{aligned} \quad (3)$$

### 2.1. Cas où les $A_i$ sont indépendants

Lorsque les  $A_i$  sont indépendants, on déduit de (3) que, pour tous éléments disjoints  $B$  et  $B^*$  de  $\mathcal{B}(X)$ ,

$$\begin{aligned} \text{Var}(N(B)) &= E(N(B)) + \sum_{i=1}^k |B \cap X_i|^2 \text{Var}(X_i) \\ \text{Cov}(N(B), N(B^*)) &= \sum_{i=1}^k \text{Var}(A_i) |B \cap X_i| |B^* \cap X_i|. \end{aligned} \quad (4)$$

Les effets aléatoires engendrent donc une surdispersion dans les quadrats (rapport variance-espérance du nombre d'individus supérieur à un) mais aussi une autocorrélation spatiale positive entre quadrats situés dans une même zone d'hétérogénéité. Cette autocorrélation n'est pas toujours prise en compte dans les analyses de populations surdispersées (Bliss, Fisher (1953); Bissel (1972a); Bissel (1972b)). L'argument, parfois utilisé, qui consiste à dire que les quadrats sont tirés par échantillonnage aléatoire simple est irrecevable dès lors que l'on n'est plus dans le cadre théorique de population fixe (Cassel, Sarndal, Wretman (1977), chap. 2).

C'est le cas ici puisqu'on a posé un modèle probabiliste sur la répartition spatiale des individus et que, par définition, on sort donc du contexte de population fixe. D'autre part, l'expression dans (4) de la covariance entre nombres d'individus de deux quadrats disjoints montre bien que ce qui entre en jeu, c'est le nombre d'effets aléatoires affectant les deux quadrats à la fois, ceci, indépendamment de la procédure d'échantillonnage utilisée.

## 2.2. Cas où les $A_i$ sont indépendants et de même loi

Le cas où les  $A_i$  sont indépendants et identiquement distribués (i.i.d) selon une loi  $F$ , correspond aux situations où les effets aléatoires sont de même nature.

Pour tous éléments disjoints  $B_1, \dots, B_r$  de  $\mathcal{B}(X)$ , la loi du  $r$ -uplet  $N(B_1), \dots, N(B_r)$  est alors

$$\bigotimes_{i=1}^k \mathcal{MP}(A_i | B_1 \cap X_i |, \dots, A_i | B_r \cap X_i |) \wedge_{A_i} F, \quad (5)$$

$\mathcal{MP}(\lambda_1, \dots, \lambda_r)$  étant la loi de Poisson multiple (Johnson, Kotz (1969), chap. 11 §4).

Ainsi, lorsque  $F$  est la loi gamma  $\mathcal{G}(\alpha, \omega)$  de fonction génératrice des moments donnée par :

$$t \longmapsto \left(1 - \frac{\omega}{\alpha} t\right)^{-\alpha} \quad (\omega \text{ étant l'espérance de } F),$$

la loi du  $r$ -uplet  $N(B_1), \dots, N(B_r)$  est un produit de convolution de lois multinomiales négatives.

Par conséquent, les  $N(B_i)$  sont parfois corrélés et ne suivent la loi binomiale négative que dans des cas bien précis (Fig. 2).

D'autre part,  $\forall B \in \mathcal{B}(X)$ , on a

$$E(N(B)) = E\left(\sum_{i=1}^k A_i | B \cap X_i |\right) = \omega \sum_{i=1}^k |B \cap X_i|.$$

On voit donc que le processus  $N(\cdot)$  n'est stationnaire à l'ordre 1 que si les  $X_i$  forment une partition de  $X$ .

## 3. Tests d'hétérogénéité

Dans ce paragraphe, on cherche à tester l'hypothèse  $H_0$  d'égalité des effets liés aux zones d'hétérogénéité  $X_i$  à partir d'un plan d'échantillonnage non adaptatif. Les  $X_i$  ne sont pas supposés connus forcément, mais on admet que l'intersection d'une surface échantillonnée quelconque  $S$  avec une ou plusieurs des zones  $X_i$

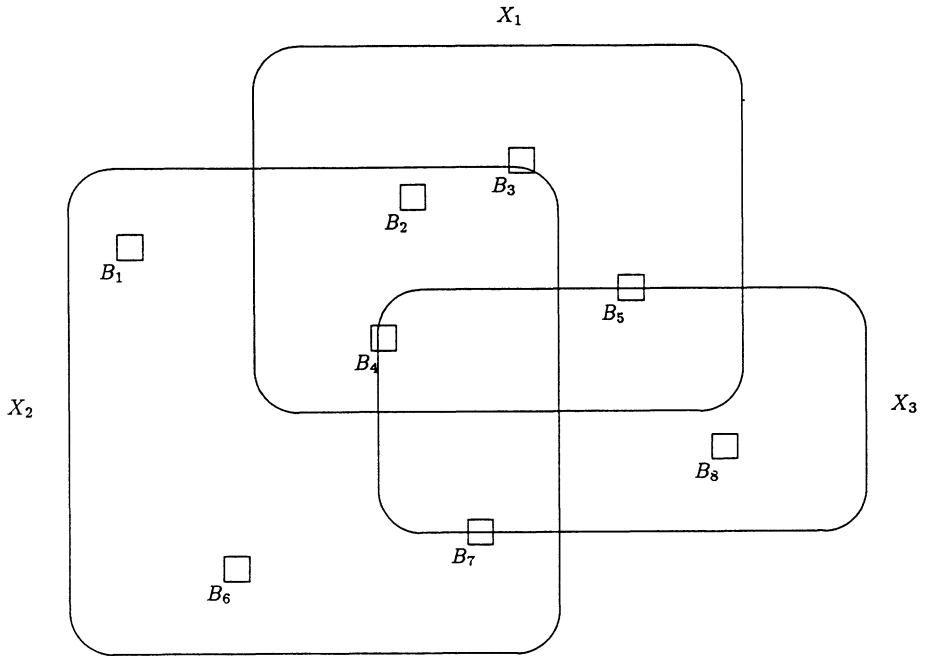


FIGURE 2

Même exemple qu'en figure 1 avec des effets indépendants et gamma-distribués (tiré de Vaillant (1991)). Les quadrats  $B_1$ ,  $B_2$ ,  $B_6$  et  $B_8$  correspondent à des binomiales négatives car ils ne chevauchent pas les frontières des zones d'hétérogénéité.  $B_5$  et  $B_7$  correspondent aussi à des binomiales négatives car ils ont des parties de même superficie dans deux zones différentes. Par contre,  $B_3$  et  $B_4$  ne correspondent pas à des binomiales négatives car ils ont des parties de superficie inégale dans deux zones différentes.

peut être confirmée après l'observation de  $S$ . Notons que  $H_0$  ne correspond à l'hypothèse de Processus de Poisson homogène que si les  $X_i$  forment une partition de  $X$ .

### 3.1. Approche conditionnelle

Quand la loi jointe des  $A_i$  n'est pas connue, il est naturel de se placer conditionnellement aux effets  $A_i$ . La vraisemblance d'échantillon s'exprime alors en fonction de paramètres  $\lambda_i$  correspondant aux réalisations des  $A_i$ . Comme on le verra maintenant, on obtient son expression en tenant compte des différentes intersections entre les zones d'hétérogénéité  $X_i$ .

Soit  $\mathcal{T}$ , l'ensemble constitué par les intersections non vides entre les  $X_i$ , de la forme

$$T_u = \left( \bigcap_{i \in u} X_i \right) \setminus \bigcup_{i \notin u} X_i$$

où  $u$  est un sous-ensemble de  $\{1, 2, \dots, k\}$ , et où  $A \setminus B = \{x \in A \text{ et } x \notin B\}$ .

$\mathcal{T}$  est une partition de  $X$  de cardinal au plus égal à  $2^k - 1$  (voir l'exemple donné dans Fig. 1).

Il est aisé de montrer (Vaillant, 1991) que la restriction à  $T_u$  du processus  $N(\cdot)$  est, conditionnellement aux  $A_i$ , un processus de Poisson homogène d'intensité

$$\lambda_u = \sum_{i \in u} \lambda_i. \quad (6)$$

Pour toute procédure d'échantillonnage non adaptative, la log-vraisemblance d'échantillon s'écrit alors

$$\mathcal{L} = \sum_{T_u \in \mathcal{T}} \left\{ n_u \log(\lambda_u S_u) - \lambda_u S_u - \log(n_u!) \right\}, \quad (7)$$

où  $n_u$  et  $S_u$  sont respectivement le nombre d'individus observés et la surface échantillonnée dans  $T_u$ .

Par conséquent, pour tout élément  $i$  de  $\{1, 2, \dots, k\}$ , on a

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \left( \sum_{u \ni i} \frac{n_u}{\lambda_u} \right) - S_i, \quad \text{où } S_i = \sum_{u \ni i} S_u \text{ est la surface échantillonnée dans } X_i.$$

Dès lors que les  $S_i$  sont non nulles, on peut obtenir les estimateurs de maximum de vraisemblance (MV) des  $\lambda_i$ . Généralement, ils sont obtenus numériquement mais on peut les obtenir explicitement. Par exemple, quand  $X_{i_0}$  a une intersection vide avec tout  $X_i$  tel que  $i \neq i_0$ , alors l'estimateur MV de  $\lambda_{i_0}$  est tout simplement  $n_{\{i_0\}}/S_{i_0}$ .

Pour tester l'hypothèse  $H_0$  d'égalité des  $\lambda_i$ , on peut utiliser le test du rapport de vraisemblance. En notant  $\lambda$  la valeur commune des  $\lambda_i$  sous  $H_0$ , son estimateur MV  $\hat{\lambda}$  nous est simplement donné par le rapport du nombre total  $n_+$  d'individus dans l'échantillon sur le total pondéré  $S_*$  des surfaces échantillonnées :

$$\hat{\lambda} = \frac{n_+}{S_*} \quad \text{avec } S_* = \sum S_u \#u, \quad \#u \text{ étant le cardinal de } u.$$

La statistique  $Y$  du test du rapport de vraisemblance est égale à deux fois le logarithme du rapport de vraisemblance. A partir de (7), on obtient

$$Y = 2 \sum_{T_u \in \mathcal{T}} \left\{ n_u \log \left( \frac{\hat{\lambda}_u S_*}{\#u n_+} \right) + \left( \frac{n_+ \#u}{S_*} - \hat{\lambda}_u \right) S_u \right\}, \quad (8)$$

où  $\hat{\lambda}_u$  est la somme des estimateurs MV des  $\lambda_i$ , pour  $i \in u$ .

Selon la théorie générale des tests de rapport de vraisemblance (Lawley (1956); Kendall et Stuart (1973, vol. 2, chap. 24)), sous  $H_0$ , la statistique  $Y$  suit asymptotiquement la loi du chi-2 à  $r - 1$  degrés de liberté,  $r$  étant le nombre de



$S_i$  non nulles. L'approximation de la loi de  $Y$  par un chi-2 est d'autant meilleure que les  $S_i$  et les  $\lambda_i$  sont grands. Quand ce n'est pas le cas, Lawley (1956) propose de diviser  $Y$  par un facteur correctif  $C$  qui permet d'améliorer la qualité de cette approximation. La méthode d'obtention de  $C$  est décrite dans son article.

Rappelons que  $H_0$  n'est équivalente à l'hypothèse de processus de Poisson homogène sur  $X$  que lorsque les  $X_i$  forment une partition de  $X$ .

*Exemple 1* : prenons  $k = 2$ ,  $X_1 \cap X_2 \neq \emptyset$  et  $S_{\{1\}} = S_{\{2\}} \neq 0$ .

On a alors  $\mathcal{T} = \{X_1 \setminus X_2, X_2 \setminus X_1, X_1 \cap X_2\}$ .

D'autre part,

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \frac{n_{\{1\}}}{\lambda_1} + \frac{n_{\{1,2\}}}{\lambda_1 + \lambda_2} - S \quad \text{et} \quad \frac{\partial \mathcal{L}}{\partial \lambda_2} = \frac{n_{\{2\}}}{\lambda_2} + \frac{n_{\{1,2\}}}{\lambda_1 + \lambda_2} - S$$

avec  $S = S_{\{1\}} + S_{\{1,2\}} = S_{\{2\}} + S_{\{1,2\}}$ .

Il en résulte que

$$\hat{\lambda}_1 = \frac{n_+}{S} \frac{n_{\{1\}}}{n_{\{1\}} + n_{\{2\}}} \quad \text{et} \quad \hat{\lambda}_2 = \frac{n_+}{S} \frac{n_{\{2\}}}{n_{\{1\}} + n_{\{2\}}}$$

(rappelons que  $n_+ = n_{\{1\}} + n_{\{2\}} + n_{\{1,2\}}$ ).

Puisque  $S_* = S_{\{1\}} + S_{\{2\}} + 2S_{\{1,2\}} = 2S$ , on obtient

$$Y = 2 \left( n_{\{1\}} \log \left( \frac{2n_{\{1\}}}{n_{\{1\}} + n_{\{2\}}} \right) + n_{\{2\}} \log \left( \frac{2n_{\{2\}}}{n_{\{1\}} + n_{\{2\}}} \right) \right).$$

*Exemple 2* : prenons  $k = 2$ ,  $X_1 \subset X_2$  et  $0 < S_{\{1,2\}}$ .

Par conséquent,  $\mathcal{T} = \{X_2 \setminus X_1, X_1 \cap X_2\}$ .

D'où

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \frac{n_{\{1,2\}}}{\lambda_1 + \lambda_2} - S_{\{1,2\}} \quad \text{et} \quad \frac{\partial \mathcal{L}}{\partial \lambda_2} = \frac{n_{\{2\}}}{\lambda_2} + \frac{n_{\{1,2\}}}{\lambda_1 + \lambda_2} - S_{\{2\}} - S_{\{1,2\}}.$$

On en déduit que

$$\hat{\lambda}_1 = \frac{n_{\{1,2\}}}{S_{\{1,2\}}} - \frac{n_{\{2\}}}{S_{\{2\}}} \quad \text{et} \quad \hat{\lambda}_2 = \frac{n_{\{2\}}}{S_{\{2\}}},$$

et  $Y$  est de nouveau obtenu aisément à partir de (8) :

$$Y = 2 \left( n_{\{1,2\}} \log \left( \frac{n_{\{1,2\}} S_*}{2S_{\{1,2\}} n_+} \right) + n_{\{2\}} \log \left( \frac{n_{\{2\}} S_*}{S_{\{2\}} n_+} \right) \right).$$

Dans ces deux exemples,  $X_1$  et  $X_2$  peuvent correspondre aux biotopes respectifs (non disjoints) de deux types de proies et,  $\lambda_1$  et  $\lambda_2$  aux degrés d'attraction d'une espèce de prédateurs sans compétition intraspécifique. Quand les deux biotopes sont identifiables, un échantillonnage par dénombrement fournit des valeurs pour  $n_{\{1\}}$ ,  $n_{\{2\}}$  et  $n_{\{1,2\}}$  et permet de tester l'hypothèse  $\lambda_1 = \lambda_2$ .

Il est à noter la possibilité de valeurs négatives pour des  $\hat{\lambda}_i$ . Cela indique alors l'inadéquation du modèle aux données observées. Ainsi, dans l'exemple 2 ci-dessus,  $\hat{\lambda}_1$  est négatif si le nombre moyen d'individus pour l'échantillon observé dans  $X_1 \cap X_2$  est inférieur à celui pour l'échantillon observé dans  $X_2 \setminus X_1$ . La supposition d'additivité  $\lambda_{1,2} = \lambda_1 + \lambda_2$  est alors à remettre en question.

Avant de tester l'égalité des  $\lambda_i$ , il est donc utile d'examiner l'adéquation du modèle fourni par (1) avec les observations effectuées par échantillonnage. Pour cela, nous proposons tout d'abord de vérifier que les  $\lambda_i$  sont tous positifs, puis de calculer le critère

$$Z = 2 \sum_{T_u \in \mathcal{T}} \left\{ n_u \log \left( \frac{n_u}{\lambda_u S_u} \right) - (n_u - \hat{\lambda}_u S_u) \right\}.$$

$Z$  est la statistique du rapport de vraisemblance entre l'hypothèse  $H_1$  pour laquelle  $\lambda_u$  est donné, pour tout  $u$ , par (6) et l'hypothèse plus vaste  $H_2$  pour laquelle les  $\lambda_u$  sont quelconques. Sous  $H_1$ , la loi asymptotique de  $Z$  est un chi-2 à  $r^* - r$  degrés de liberté (Lawley (1956); Kendall et Stuart (1973, vol. 2, chap. 24)),  $r^*$  étant le nombre de  $S_u$  non nulles. Notons que n'intervient pas dans le calcul de  $Z$  toute zone  $X_{i_0}$  ayant des intersections vides avec les autres  $X_i$ , ( $i \neq i_0$ ) car  $\lambda_{i_0} = n_{\{i_0\}}/S_{i_0}$ .

Ainsi, si on considère les exemples ci-dessus, on aura pour l'exemple 1

$$Z = 2 \left( (n_{\{1\}} + n_{\{2\}}) \log \left( \frac{n_{\{1\}} + n_{\{2\}}}{S - S_{\{1,2\}}} \right) + n_{\{1,2\}} \log \left( \frac{n_{\{1,2\}}}{S_{\{1,2\}}} \right) - n_+ \log \left( \frac{n_+}{S} \right) \right),$$

avec  $r^* = 3$  et  $r = 2$ .

Par contre, pour l'exemple 2, on aura  $r^* = r = 2$  donc impossibilité de tester  $H_2$  contre  $H_1$ .

Soulignons que lorsque les  $X_i$  forment une partition de  $X$ ,  $H_1$  et  $H_2$  sont équivalentes. D'autre part, d'une façon générale,  $X_{i_0} \subset X_{i_1}$  pour  $i_0 \neq i_1$  correspond à l'absence de  $T_{\{i_0\}}$  dans  $\mathcal{T}$ . Donc,  $r^*$  sera d'autant plus proche de  $r$  qu'il y aura des zones d'hétérogénéité incluses les unes dans les autres, et la puissance du test basé sur  $Z$  diminuera d'autant.

### 3.2. Cas où les $A_i$ sont gamma-distribués

Lorsque les  $A_i$  sont i.i.d selon la loi gamma  $\mathcal{G}(\alpha, \omega)$ , on a vu au §2.2 que les nombres d'individus dans des quadrats échantillonnés pouvaient être binomiaux négatifs ou quasi-binomiaux négatifs et pas nécessairement indépendants. Cependant, on démontre que pour tout élément  $T_u$  de  $\mathcal{T}$ , la restriction à  $T_u$  du processus  $N(\cdot)$

est processus binomial négatif (Diggle, Milne (1983); Vaillant (1991)) d'intensité  $\omega \#u$  et d'exposant  $\#u\alpha$ . Par conséquent,  $n_u$  suit la binomiale négative d'espérance  $\omega \#u S_u$  et d'exposant  $\#u\alpha$ .

D'autre part, soit  $\underline{n}$ , le vecteur des  $n_u$  correspondant à des  $S_u$  non nulles. D'après (4), la matrice de variance-covariance de  $\underline{n}$  est

$$V = \left( \omega \#u S_u \delta_{uv} + \frac{\omega^2}{\alpha} \#(u \cap v) S_u S_v \right),$$

$\delta_{uv}$  valant 1 si  $u = v$ , et 0 sinon.

Dans un tel cadre, l'hypothèse nulle  $H_0$  correspond à  $\alpha = \infty$ . Sous cette hypothèse, les  $n_u$  sont donc indépendants,  $n_u$  suivant la loi de Poisson  $\mathcal{P}(\omega \#u S_u)$ . L'estimateur MV de  $\omega$  est alors  $n_+/S_*$ .

Quand les  $X_i$  forment une partition de  $X$ , la matrice  $V$  est diagonale et les estimateurs MV de  $\omega$  et de  $\alpha$  sont obtenus numériquement de la façon décrite par Bissel (1972a). Le test du rapport de vraisemblance est alors appliqué sans difficulté pour tester l'hypothèse  $H_0$ .

Quand les  $X_i$  ne sont pas mutuellement disjoints, la vraisemblance n'est pas toujours explicitable. En effet, on montre (voir le cas général dans Vaillant (1991)) que la fonction génératrice des probabilités de  $\underline{n}$  est

$$(z_u) \mapsto \prod_{i=1}^k \left( 1 - \frac{\omega}{\alpha} \sum_{u \ni i} S_u (z_u - 1) \right)^{-\alpha}.$$

Plus il y a de chevauchements entre les  $X_i$ , plus l'obtention de la vraisemblance d'échantillon est compliquée, ceci pour toute procédure d'échantillonnage non-adaptative. Il est alors plus intéressant d'utiliser l'approche conditionnelle présentée au §3.1 puisqu'elle permet d'évacuer les problèmes de vraisemblance complexe engendrée par la loi des  $A_i$ .

Il est important de rappeler que le processus spatial de Cox décrit des répartitions spatiales hétérogènes (présentation du §1.). On a, par conséquent, considéré dans cet article des individus n'ayant ni répulsion, ni attraction mutuelle, et l'apparente agrégativité qu'on pourrait constater dans leur répartition spatiale est due à leur affinité commune pour des paramètres de l'environnement. Le problème de distinction entre agrégativité apparente et agrégativité vraie est relaté par beaucoup d'auteurs (par exemple Feller (1943), Chessel (1978), Taylor (1984)), et plus particulièrement dans le cas de comptages binomiaux négatifs qui peuvent correspondre aux deux types d'agrégativité. A partir de simples données de comptage, faire cette distinction s'avère le plus souvent impossible. Soulignons donc que notre modélisation suppose a priori que les individus sont distribués indépendamment les uns des autres conditionnellement à l'environnement. Ceci explique le fait que, dans l'approche conditionnelle (§3.1), la vraisemblance puisse s'exprimer simplement en fonction des  $n_u$  et des  $S_u$ . D'autre part, en cas d'effets gamma-distribués, l'expression (5) et Fig. 2 montrent bien que c'est selon leur taille et leur emplacement par rapport aux zones d'hétérogénéité, que certains quadrats correspondront à des binomiales négatives et d'autres pas.

#### 4. Conclusion

Des individus se répartissant dans l'espace, indépendamment les uns des autres mais avec des affinités communes pour certains facteurs du milieu, engendrent des lois de Poisson composées pour ce qui concerne leur dénombrement dans des quadrats disjoints. Dans cet article, nous soulignons le fait que les observations dans ces quadrats ne sont pas forcément indépendantes, et que par conséquent, certaines méthodes statistiques ne sont plus appropriées pour leur analyse. Dans le cas d'effets aléatoires associés à des zones d'hétérogénéité, travailler conditionnellement aux réalisations des effets permet de tester des hypothèses sur l'hétérogénéité du milieu. L'approche non conditionnelle, pour laquelle on pourrait tenir compte de la loi de distribution des effets aléatoires, pose des problèmes quand les zones d'hétérogénéité se chevauchent. Notons enfin que l'on a raisonné à partir de procédures d'échantillonnage non adaptatives quelconques mais que les plans aléatoires systématiques (voir par exemple, Iachan (1982)) sont souvent préconisés pour l'étude de la répartition spatiale d'individus car ils explorent de façon régulière l'espace.

#### Bibliographie

- BISSEL, A.F. (1972a). A Negative Binomial model with varying elements sizes. *Biometrika* 59, 435-441.
- BISSEL, A.F. (1972b). Another Negative Binomial model with varying elements sizes. *Biometrika* 59, 691-693.
- BLISS, C.I., FISHER, R.A. (1953). Fitting the Negative Binomial distribution to biological data. *Biometrics* 9, 176-200.
- CASSEL, C.M., SARNDAL, C.E., WRETMAN, J.H. (1977). *Foundation of inference in statistical survey*. Wiley, New-York.
- CHESSSEL, D. (1978). Description non paramétrique de la dispersion spatiale des individus d'une espèce. *Biométrie et Ecologie*, 1, 45-135.
- CLIFF, A.D. and ORD, J.K. (1981). *Spatial processes, Models and Applications*. Pion, London.
- COTTAM, G., CURTIS, J.T. (1949). A method for making rapid surveys of woodlands, by means of pairs of randomly selected trees. *Ecology*, 30, 101-104.
- DALEY, D.J., VERE-JONES, D. (1988). *An introduction to the theory of point processes*. Springer-Verlag, New-York.
- DIGGLE, P. J. (1983). *Statistical analysis of spatial point patterns*. Ed. Academic Press, London, 148p.
- DIGGLE, P. J., COX, T.F. (1983). Some distance-based tests of independence for sparsely-sampled multivariate spatial point patterns. *Intern. Statist. Rev.* 51, 11-23.
- DIGGLE, P.J., MILNE, R.K. (1983). Negative Binomial Quadrat Counts and Point Processes. *Scand. J. Statist.* 10, 257-267.

- FELLER, W. (1943). On a general class of contagious distributions. *Ann. Math. Stat.*, 14, 389-400.
- IACHAN, R. (1982). Systematic sampling : A critical review. *Internat. Statist. Rev.* 50, 293-303.
- JOHNSON, N.L., KOTZ, S. (1969). *Discrete distributions*. Houghton Mifflin, Boston.
- KENDALL, M.G., STUART, A. (1973). *The advanced theory of statistics*, vol 2. Griffin, London.
- LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* 59, 295-303.
- ROGERS, A. (1974). *Statistical analysis of spatial pattern*. Pion Limited, London, 164 p.
- TAYLOR, L.R. (1984) Assessing and interpreting the spatial distribution of insect populations. *Ann. Rev. Entomol.*, 29, 321-357.
- VAILLANT, J. (1991). Negative binomial distributions of individuals and spatio-temporal Cox processes. *Scand. J. Statist.*, 18, 235-248.