Available online at www.sciencedirect.com
SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 340 (2005) 619–622

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Estimation spatio-temporelle d'un modèle de système de particules

Xavier Guyon [a], Besnik Pumo [b]

[a] *SAMOS, université Paris 1, 90, rue de Tolbiac, 75634 Paris cedex 13, France*
[b] *SAGAH, institut national d'horticulture, 2, rue Le Nôtre, 49000 Angers, France*

Reçu le 5 juillet 2004 ; accepté après révision le 8 mars 2005

Disponible sur Internet le 22 avril 2005

Présenté par Paul Deheuvels

Résumé

Soit X un processus de contact (PC) à temps discret sur $\mathbb{Z}^2$ tel que défini par Durrett et Levin (1994). On étudie l'estimation du modèle basé sur l'évolution spatio-temporelle de X, i.e. $T+1$ observations successives de X sur un ensemble fini de sites S. Nous considérons l'estimateur de maximum de pseudo-vraisemblance marginale (PVM) et montrons, quand $T \to \infty$, que cet estimateur converge et qu'il est asymptotiquement gaussien si X survit sur S. ***Pour citer cet article : X. Guyon, B. Pumo, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Abstract

Space–time estimation of a particle system model. Let X be a discrete time contact process (CP) on $\mathbb{Z}^2$ as defined by Durrett and Levin (1994). We study the estimation of the model based on space–time evolution of X, that is, $T+1$ successive observations of X on a finite subset S of sites. We consider the maximum marginal pseudo-likelihood (MPL) estimator and show that, when $T \to \infty$, this estimator is consistent and asymptotically normal for a non vanishing supercritical CP. ***To cite this article: X. Guyon, B. Pumo, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

1. Introduction

Considérons un Processus de Contact (PC) à temps discret $X = (X_t, t = 0, 1, \ldots)$ tel qu'il a été défini par Durrett et Levin [4]. Ce processus modélise par exemple l'évolution sur $\mathbb{Z}^2$ d'une espèce végétale, $X_t(s) = 1$ (resp. $X_t(s) = 0$) signifiant la présence (resp. l'absence) d'une plante au site s à l'instant t. La dynamique de $X_t = (X_t(s), s \in \mathbb{Z}^2)$ est déterminée par des transitions $P(X_{t+1}(s) = y \mid X_t = x_t)$ invariantes par translation dans le

Adresses e-mail : Xavier.Guyon@univ-paris1.fr (X. Guyon), Besnik.Pumo@inh.fr (B. Pumo).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.03.014

temps et dans l'espace. Pour $d \geqslant 1$, soit $\mathcal{N}_d(s) = \{u \in \mathbb{Z}^2\colon \|s-u\|_1 \leqslant d\}$ le d-voisinage du site s et $\partial s = \mathcal{N}_1(s) \setminus \{s\}$ l'ensemble des 4 voisins de s. Le PC étudié ici est caractérisé par les transitions suivantes (Durrett et Levin, [4]) :

(i) Chaque plante en vie à l'instant t meurt avec une probabilité γ à l'instant $t+1$,
(ii) Une plante en vie en s donne naissance à une nouvelle plante en $u \in \partial s$ avec une probabilité λ ; les différentes reproductions pour $s \in \mathbb{Z}^2$ et $u \in \partial s$ sont indépendantes,
(iii) Au plus une plante survit en s. De plus, les événements définis en (a) et (b) sont indépendants dans le temps.

Nous considérons ici le problème de l'estimation du paramètre $\theta = (\gamma, \lambda)$ à partir d'une suite de configurations successives $\mathrm{x}(T) = (x_0, x_1, \ldots, x_T)$ observées sur $S_1 = S \cup \partial S$ où $S \subset \mathbb{Z}^2$ est un sous-ensemble fini et fixé de sites et ∂S est la 1-frontière de voisinage de S. L'estimateur $\hat{\theta}_T$ que nous proposons maximise une *pseudo-vraisemblance marginale* (PVM) de $\mathrm{x}(T)$ qui est une fonctionnelle qui s'explicite facilement là où la vraisemblance devient vite incalculable pour S grand. On montre que $\hat{\theta}_T \to \theta$ et qu'il est asymptotiquement gaussien sous la condition, notée **(I),** de non-extinction de X sur S. Ces résultats asymptotiques utilisent une propriété de sous-ergodicité de X. La condition **(I)** est satisfaite avec une probabilité positive pour un processus supercritique (Durrett, [3]).

Fiocco et Zwet [5] étudient l'estimation dans un autre cadre, à savoir le PC est à temps continu et on observe une seule configuration x_t ; dans ce cas, un seul des deux paramètres du modèle est estimable.

2. Estimateur de pseudo-vraisemblance marginale (PVM)

Soit $\mathrm{x}(T) = (x_0, x_1, \ldots, x_T)$ une suite d'observations du PC sur $S_1 = S \cup \partial S$. Pour $A \subset S$, notons $P_A(x_t, x_{t+1}; \theta) = P(X_{t+1}(A) = x_{t+1}(A) \mid X_t(S_1) = x_t(S_1))$ et $m(x_t, A) = \sum_{s \in A} x_t(s)$ le nombre de sites de A occupés par x_t. La transition jointe $P_S(x_t, x_{t+1}; \theta)$ devenant incalculable pour S grand, nous utiliserons une pseudo-transition marginale $M_S(x_t, x_{t+1}; \theta) = \prod_{s \in I(x_t, S)} P_{\{s\}}(x_t, x_{t+1}; \theta)$ produit des transitions en un site sur $I(x_t, S) = \{s \in S\colon 0 < P_{\{s\}}(x_t, x_{t+1}; \theta) < 1\}$, l'ensemble des sites de S informatifs au temps t. La pseudo-vraisemblance marginale (PVM) est alors le produit des pseudo-transitions marginales aux différents instants d'observations et $\hat{\theta}_T$ est une valeur qui maximise cette PVM.

L'idée d'utiliser une pseudo-vraisemblance (PV) est classique en statistique : PV gaussienne pour un champ au second ordre (Whittle, [10]), PV conditionnelle pour un champ de Markov (Besag, [1]), PV pour estimer les paramètres d'un modèle de diffusion sur un lattice (Mollison, [9]).

La loi de $X_{t+1}(s)$ conditionnellement à x_t est une variable de Bernoulli de paramètre $P_{\{s\}}(x_t, x_{t+1}; \theta) = p(x_t, s; \theta)^{1-x_{t+1}(s)}(1 - p(x_t, s; \theta))^{x_{t+1}(s)}$ où $p(x_t, s; \theta) = \gamma^{x_t(s)} \delta^{m(x_t, \partial s)}$, $\delta = \gamma + (1-\gamma)(1-\lambda)$ étant la probabilité de non-prolifération à l'instant $t+1$ d'une plante présente à l'instant t au site s. On en déduit :

$$M_S(x_t, x_{t+1}; \theta) = \prod_{s \in I(x_t, S)} p(x_t, s; \theta)^{1-x_{t+1}(s)} \big(1 - p(x_t, s; \theta)\big)^{x_{t+1}(s)} \tag{1}$$

avec la convention $M(0, 0; \theta) = 1$ si $I(x_t, S) = \emptyset$. Notant $n(x_t)$ le nombre de sites informatifs de x_t, $n(T) = \sum_{t=0}^{T-1} n(x_t)$ le nombre total de sites informatifs, et, en utilisant (1), la log-PVM normalisée de $\mathrm{x}(T)$ est,

$$l_T(\theta) = \frac{1}{n(T)} \sum_{t=0}^{T-1} \sum_{s \in I(x_t)} \big\{\big[1 - x_{t+1}(s)\big] \log p(x_t, s; \theta) + x_{t+1}(s) \log\big[1 - p(x_t, s; \theta)\big]\big\}.$$

On suppose que $\theta \in \Theta$, un compact de $]0, 1[^2$. L'estimateur de θ est $\hat{\theta}_T = \arg\max_{\theta \in \Theta} \, l_T(\theta)$.

Soit $\eta = \gamma + (1-\gamma)(1-\lambda)^2$ la probabilité de non-prolifération à l'instant $(t+1)$ dans $\{s, s'\}$ d'un plant présent en $u \in \partial s \cap \partial s'$ à l'instant t. On montre que si $s' \notin \mathcal{N}_2(s)$, les variables $(X_{t+1}(s) \mid x_t)$ et $(X_{t+1}(s') \mid x_t)$ sont indépendantes. Si $s' \in \mathcal{N}_2(s)$ et $s' \neq s$, $\mathrm{Cov}(X_{t+1}(s), X_{t+1}(s') \mid x_t) = p(x_t, s; \theta)\ p(x_t, s'; \theta)\ [b(x_t, s, s'; \theta) - 1]$, où

$$b(x_t, s, s'; \theta) = \begin{cases} \delta^{-m(x_t, \{s, s'\})} & \text{si } s' \in \mathcal{N}_1(s) \setminus \{s\}, \\ \delta^{-2m(x_t, \partial s \cap \partial s')} \eta^{m(x_t, \partial s \cap \partial s')} & \text{si } s' \in \mathcal{N}_2(s) \setminus \mathcal{N}_1(s). \end{cases} \tag{2}$$

3. Résultats de convergence de l'estimateur $\hat{\theta}_T$

On suppose vérifiée la condition **(I)** de non-extinction de X sur S :

$$(\mathbf{I}) : I_\infty = \{\mathbf{x} = (x_t, t \geqslant 0) \colon n(\mathrm{x}(T)) \to \infty \text{ quand } T \to \infty\}.$$

Introduisons quelques notations. Soit $\mathbf{L} = \{L_i, i \in I\}$ une partition de S et $\mathcal{L} = \{L_{i(t)}, t \geqslant 0\}$ une suite infinie d'éléments de $\mathbf{L}$. Soit $c_t(s) = (x_t(s), m(x_t, \partial s)) \in \mathcal{C}_1 = \{0, 1\} \times \{0, 1, 2, 3, 4\}$ un *résumé* de la configuration locale de x_t en s et $(c, 0)$ le résumé de la configuration de x_t sur $N_2(s)$ telle que $c_t(s) = c$ et $x_t(N_2(s) \backslash N_1(s)) \equiv 0$. Pour $c \in \mathcal{C}_1$ soit $q(c; \theta) = P(X_{t+1}(s) = 0 \mid x_t(\mathcal{N}_1(s)) = c)$ et :

$$\pi_c = \underline{\lim}_{T \to \infty} \frac{\sum_{t=0}^{T-1} n(x_t, c)}{n(T)}, \qquad \pi_{c,0}^{\mathcal{L}} = \underline{\lim}_{T \to \infty} \frac{\sum_{t=0}^{T-1} n^{L_{i(t)}}(x_t, c, 0)}{n(T)}$$

où $n(x_t, c)$ (resp. $n^L(x_t, c, 0)$) est le nombre de sites de S (resp. $S \cap L$) de configuration c (resp. $(c, 0)$) à l'instant t. Notons enfin $p^{(1)}$ le vecteur des dérivés en θ de p et $J_T(\theta_o) = A_T(\theta_o) + B_T(\theta_o)$ où :

$$A_T(\theta_o) = \frac{1}{n(T)} \sum_{t=0}^{T-1} \sum_{s \in I(x_t)} \frac{p^{(1)t}[p^{(1)}]}{p(1-p)}(x_t, s; \theta_o)$$

$$B_T(\theta_o) = \frac{1}{n(T)} \sum_{t=0}^{T-1} \sum_{s \in I(x_t)} \sum_{s' \in \mathcal{N}_2(s) \cap I(x_t), s' \neq s} \left[b(x_t, s, s'; \theta_o) - 1\right] \frac{p^{(1)}(x_t, s; \theta_o)^t [p^{(1)}(x_t, s'; \theta_o)]}{[1 - p(x_t, s; \theta_o)][1 - p(x_t, s'; \theta_o)]}$$

où $b(x_t, s, s'; \theta_o)$ est donné par (2). Le résultat principal de cet article est le suivant :

Théorème 3.1. *Supposons que $\theta_o = (\gamma_o, \lambda_o)$, la vraie valeur inconnue de θ, est un point intérieur d'un compact $\Theta \subset]0, 1[^2$. Sous la condition* (**I**) $\lim_{T \to \infty} \hat{\theta}_T \overset{\text{p.s.}}{=} \theta_o$ *et*

$$\sqrt{n(T)}\, J_T^{-1/2}(\theta_o) A_T(\theta_o)(\hat{\theta}_T - \theta_o) \xrightarrow{d} \mathbf{G}_2(0, I_2).$$

La preuve utilise les résultats généraux de convergence et de normalité asymptotique des estimateurs de minimum de contrastes (Dacunha-Castelle et Duflo, [2] ; Guyon, [6]) ainsi que la convergence p.s. et en loi d'un tableau triangulaire de martingales construit à partir du gradient de la log-pseudo-vraisemblance (Dacunha-Castelle et Duflo, [2] ; Hall et Heyde, [7]. Deux pas essentiels utilisent la sous-ergodicité de X énoncée au Lemme 3.2 :

(i) Le premier pas montre que la PVM rend θ identifiable pour T grand. En effet, si $\mathrm{x}(T)$ réalise deux configurations $c_a = (u_a, v_a)$ et $c_b = (u_b, v_b)$ non-colinéaires, alors $\theta \mapsto l_T(\theta)$ est injective. Mais, sous **(I)**, la sous-ergodicité assure que $\pi_c > 0$ et donc toute configuration $c \in \mathcal{C}_1^* = \{1\} \times \{0, 1, 2, 3, 4\}$ est observée quand $T \to \infty$: la PVM rend θ identifiable.

(ii) Le deuxième pas établit que la matrice de variance-covariance de $\hat{\theta}_T$ est minorée, pour T grand, par une matrice définie positive. On montre que pour $\epsilon_o = \gamma_0(1 - \gamma_0)(1 - \lambda_0)^5[1 - (1 - \lambda_0)^4]$, $\mathcal{L}$ convenablement choisie et $J(\theta_o) = \sum_{c \neq (0,0)} \pi_{c,0}^{\mathcal{L}} \times [q(1-q)]^{-2} \times {q^{(1)}}^t [q^{(1)}](c; \theta_o)$ on a $\underline{\lim}_{T \to \infty} J_T(\theta_o) \geqslant \epsilon_o \cdot J(\theta_o)$. Pour se faire, on utilise l'idée suivante de Jensen (Jensen et Künsch, [8]) : pour $\mathcal{G}$ une tribu conditionnante générale, $\mathrm{Var}(X) \geqslant E_{\mathcal{G}}(\mathrm{Var}(X \mid \mathcal{G}))$; $l_T^{(1)}(\theta)$ s'écrivant comme une somme, on en extrait une sous-somme de termes conditionnellement indépendants et on minore chaque variance conditionnelle individuelle. Puisque $\pi^{\mathcal{L}}_{c,0} > 0$ sur $\mathcal{C}^*_1$, $J(\theta_o)$ est définie positive car, si $\alpha \neq 0$, ${}^t\alpha J(\theta_o)\alpha = \sum_{c\neq(0,0)} \pi^{\mathcal{L}}_{c,0}[q(1-q)(c;\theta_o)]^{-2}\ ({}^t\alpha q^{(1)}(c;\theta_o))^2 > 0$.

Lemme 3.2. *Sous-ergodicité de X en situation de non-extinction : Soit $\mathbf{L} = \{L_i, i \in \mathcal{I}\}$ une partition de S. Il existe $\alpha > 0$ et $\mathcal{L}$, une suite infinie d'éléments de $\mathbf{L}$, tels que, $\forall c \in \mathcal{C}^*_1$, et $\forall \mathbf{x} \in I_\infty$, on a $\pi_c \geqslant \pi^{\mathcal{L}}_{c,0} \geqslant \alpha$.*

Pour démontrer ce résultat, on définit deux suites croissantes d'instants entre 0 et $T-2$, la première, $0 \leqslant T_1 < T_2 < \cdots < T_{m_T} \leqslant T-2$ est la suite des instants informatifs i.e. telle que $x_t(\mathcal{N}_1(s)) \neq (0,0)$ pour au moins un site s de S, la seconde, $0 \leqslant T^*_1 < T^*_2 < \cdots < T^*_{m^*_T} \leqslant T-2$, est la sous-suite de la première telle que $x_t(s) = 1$ pour un site s de S. Soit $\mathcal{L} = \{L_{i(t)}, t \geqslant 0\}$ tel que $L_{i(T_k+1)}$ contienne le site informatif à l'instant T_k pour $k = 1, \ldots, m_T$. D'une part, les variables conditionnelles $(X_{T^*_{k^*}+1}(\mathcal{N}_2(s)) \mid x_{T^*_{k^*}})$ étant indépendantes pour $k^* = 1, \ldots, m^*_T$, il existe $\beta^* > 0$ tel que $\sum_{t=0}^{T-2} n^{L_{i(t+1)}}(x_{t+1}, c, 0) \geqslant \sum_{k^*=1}^{m^*_T} n^{L_{i(T^*_{k^*}+1)}}(x_{T^*_{k^*}+1}, c, 0) \geqslant m^*_T \cdot (\beta^*/2)$ quand $T \to \infty$. D'autre part, il existe $\beta > 0$ tel que $m^*_T \geqslant m_T \cdot (\beta/2)$ et $\sum_{t=0}^{T-2} n^{L_{i(t+1)}}(x_{t+1}) \leqslant \sum_{k=1}^{m_T} n(x_{T_k+1}) \leqslant |S| \cdot m_T \leqslant |S| \cdot m^*_T/(\beta/2)$. On obtient le résultat avec $\alpha = \beta \cdot \beta^*/(2 \cdot |S|)^2$ puisque $\sum_{t=0}^{T-2} n^{L_{i(t+1)}}(x_{t+1}, c, 0) \div \sum_{t=0}^{T-2} n^{L_{i(t+1)}}(x_{t+1}) \geqslant \beta \cdot \beta^*/(2 \cdot |S|)^2$ et le choix de $\mathcal{L}$ entraîne $\sum_{t=0}^{T-2} n^{L_{i(t+1)}}(x_{t+1}) \div \sum_{t=0}^{T-2} n(x_{t+1}) \geqslant 1/|S|$.

A partir de l'ergodicité de X et conditionnellement à la survie du PC, on obtiendrait directement la positivité de π_c (Durrett et Levin, [4], p. 335). Cependant l'invariance par translation espace-temps n'est pas fondamentale dans la démonstration du lemme. Cela permet d'étendre les résultats du Théorème 3.1 à des modèles non invariants dans l'espace et/ou dans le temps à la condition que ces probabilités de transition soient convenablement minorées.

Une étude expérimentale pour un PC supercritique sur le carré 64×64 et $T = 100$ confirme les résultats asymptotiques 3.1. Pour $T = 4$, on obtient de bonnes estimations des paramètres pour différents PC avec de faibles différences entre les estimations des écarts types théoriques et les écarts types empiriques. Si le processus n'est pas supercritique, on observe des biais d'estimation.

Références

[1] J. Besag, Spatial interaction and the statistical analysis of lattice systems, J. Roy. Statist. Soc. Ser. B 36 (1974) 192–225.

[2] D. Dacunha-Castelle, M. Duflo, Probability & Statistics, vol. 2, Springer, Berlin, 1986.

[3] R. Durrett, Ten lectures on particle systems, in: Cours de Saint Flour (1993), in: Lecture Notes in Math., vol. 1608, Springer, Berlin, 1995, pp. 97–201.

[4] R. Durrett, S.A. Levin, Stochastic spatial models: a user's guide to ecological applications, Philos. Trans. Roy. Soc. London Ser. B 343 (1994) 329–350.

[5] M. Fiocco, W.R. Zwet, Parameter estimation for the supercritical contact process, Bernoulli 9 (2003) 1071–1092.

[6] X. Guyon, Random Fields on a Network: Modelling, Statistics and Applications, Springer, Berlin, 1995.

[7] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, 1980.

[8] J.L. Jensen, H.R. Künsch, On asymptotic normality of pseudo-likelihood estimate for pairwise interaction processes, Ann. Inst. Statist. Math. 46 (1994) 475–486.

[9] D. Mollison, Spatial contact models for ecological and epidemic spread, with discussion, J. Roy. Statist. Soc. Ser. B 39 (1977) 283–326.

[10] P. Whittle, On stationary process in the plane, Biometrika 41 (1954) 434–449.