



HAL
open science

Quelques contributions à la modélisation et l'analyse statistique de processus spatiaux

Cécile Hardouin

► **To cite this version:**

Cécile Hardouin. Quelques contributions à la modélisation et l'analyse statistique de processus spatiaux. Statistiques [math.ST]. Université Paris Ouest Nanterre La Défense, 2011. tel-00648144

HAL Id: tel-00648144

<https://theses.hal.science/tel-00648144v1>

Submitted on 5 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Habilitation
à diriger des recherches

Cécile Hardouin

MODAL'X

Université Paris Ouest Nanterre la Défense

**Quelques contributions à la modélisation
et l'analyse statistique de processus spatiaux**

Soutenue le 11 juillet 2011 devant la commission d'examen composée de

M. Jean-Noël Bacro	Rapporteur
M. Patrice Bertail	Rapporteur
M. Joël Chadeuf	
M. Noël Cressie	Rapporteur
Mme Florence Forbes	
M. Stéphane Robin	
M. Xavier Guyon	Invité

Remerciements

Je remercie ici tous ceux qui m'ont accompagnée durant ces années. Tout d'abord Xavier Guyon, qui m'a formée à la recherche, et avec qui c'est toujours un plaisir de travailler. Après ma thèse nous avons continué à travailler ensemble régulièrement et nous avons encore des travaux à terminer ; en même temps, il a toujours été disponible pour des discussions sur des projets avec d'autres collaborateurs. Je le remercie beaucoup pour sa disponibilité, son attention, mais aussi et surtout pour son amitié. C'est lui enfin qui m'a soutenue en premier dans ce projet d'habilitation.

Je remercie Noël Cressie et Jean-Noël Bacro d'avoir accepté de rapporter ce travail, malgré leur emploi du temps chargé ; je leur en suis très reconnaissante. Je suis très heureuse que leurs conclusions m'aient été favorables, leur contribution dans cette démarche est un honneur et un plaisir.

Je souhaite aussi exprimer ma gratitude à Patrice Bertail qui a bien voulu faire l'effort de plonger dans le spatial et présenter mon dossier de recherche.

Je remercie vivement Florence Forbes, Joël Chadeuf et Stéphane Robin qui me font l'honneur de participer à ce jury.

J'exprime ma gratitude envers mes différents coauteurs, particulièrement Jian-Feng, qui m'a emmené sur les chemins des images. Chacun aura contribué à ma découverte de nouveaux sujets et applications.

Un grand merci à mes collègues, et souvent amis, du Samos et de Nanterre.

Merci aux Samosiens pour leur gentillesse, leur générosité, et leur bonne humeur indéfectible, qui m'ont permis de travailler dans une ambiance toujours très agréable et conviviale (et gourmande) ; merci à Marie qui anime cette équipe avec une énergie et une générosité hors du commun.

Merci aux Nanterrois pour leur disponibilité et leur sympathie ; j'ai partagé avec eux de nombreuses heures d'enseignement, d'éternelles discussions sur les besoins et les attentes de nos étudiants, quelques doutes et découragements devant certaines copies, quelques petits bonheurs devant d'autres, beaucoup d'attente sur les quais du RER, et quelques manifs ! Nous avons aussi nos déjeuners, et j'attends les prochaines retrouvailles.

D'une manière générale je remercie tous ceux, matheux ou pas, petits et grands, qui m'ont encouragée à rédiger ce mémoire.

Je garde mes derniers remerciements pour Roberto, qui depuis de longues années m'encourage dans mes projets et me pousse toujours plus loin, et pour mes enfants qui se remettront peut-être un jour d'avoir une mère prof de maths !

Table des matières

Présentation générale	1
1 Modèles spatiaux	5
1.1 Introduction	5
1.2 Dynamiques temporelles de champs de Markov ([Har5])	6
1.3 Généralisation des auto-modèles de Besag ([Har7, Har9])	14
1.4 Modèles spatiaux à états mixtes ([Har6, Har10, Har13])	18
1.4.1 Variables à états mixtes	18
1.4.2 Auto-modèles à états mixtes	21
1.4.3 Dynamique temporelle d’auto-modèles à états mixtes ([Har13])	23
1.5 Perspectives	27
2 Adoption de standards et diffusion de technologies	33
2.1 Introduction	33
2.2 Tests de répartition spatiale ([Har4])	35
2.2.1 Modèles sans coordination spatiale	35
2.2.2 Modèles avec coordination spatiale	36
2.2.3 Tests de non-coordination spatiale : le cas d’un lattice régulier aux ν plus proches voisins	36
2.2.4 Autres tests et étude empirique des puissances	38
2.3 Répartition spatiale avec présence initiale d’un standard ([Har8, Har11])	39
2.4 Perspectives	43
3 Calcul de constante de normalisation pour un modèle de Gibbs	47
3.1 Introduction	47
3.2 Récursions pour une distribution de Gibbs temporelle ([Har12])	48
3.2.1 Récursions sur les marginales et calcul exact de la constante de normalisation	48
3.2.2 Extension aux processus “ r -factorisables”	50
3.2.3 Récursion sur les marginales conditionnelle au futur	50
3.2.4 Cas d’un processus de Gibbs spatial	51
4 Quelques résultats en statistique paramétrique	55
4.1 Estimation par minimum de contraste en situation non ergodique	55
4.1.1 Processus de contraste fort ([Har0, Har1])	55
4.1.2 Normalité asymptotique ([Har0-1-2])	57
4.1.3 Test asymptotique de différence de contrastes ([Har0-1-2])	58

4.2	Estimation sur le log-périodogramme “régularisé” ([Har0, Har3])	60
A	Curriculum Vitae	65
A.1	Synthèse de la carrière	65
A.2	Publications et travaux	66
A.3	Communications et conférences	67
A.4	Animation recherche et responsabilités collectives	68
A.5	Enseignement et encadrement	69

Présentation générale

Les travaux présentés dans ce mémoire ont fait l'objet des publications suivantes :

[Har0] Hardouin C., (1992) Quelques résultats nouveaux en statistique des processus : contraste fort, régression à résidus à longue portée, estimation par log-périodogramme. Thèse Université Paris 7.

[Har1] Guyon X., Hardouin C. (1992) The chi-square coding test for nested Markov random field hypotheses, in *Stoch. Models, Stat. Methods and Algorithms in Image Analysis. Lecture Notes in Statistics*, 74, Springer, 165–176.

[Har2] Bayomog S., Guyon X., Hardouin C., Yao J.F. (1996) Test de différence de contrastes et somme pondérée de khi-deux, *The Canadian Journal of Statistics*, Vol 24, n° 1.

[Har3] Comte F., Hardouin C. (1997) Régression sur le log-périodogramme régularisé pour des processus gaussiens à densité spectrale bornée, *C. R. Acad. Sci. Paris, Ser. I* 325, no. 11, 1203–1206.

[Har4] Hardouin C., Guyon X. (2001) Standards adoption dynamics and test for non spatial coordination, in *Spatial statistics : methodological aspects and applications*, Lecture Notes in Statistics 159, Springer, 39–56.

[Har5] Guyon X., Hardouin C. (2002) Markov chain Markov field dynamics : models and statistics, *Statistics* 36, no. 4, 339–363.

[Har6] Bouthemy P., Hardouin C., Piriou G., Yao J.-F. (2006) Mixed state auto-models and motion texture modeling, *Journal of Mathematical Imaging and Vision*. Vol 25, n 3, 387–402.

[Har7] Hardouin C., Yao J.-F. (2007) Multi-parameter auto-models with applications to cooperative systems, *C. R. Acad. Sci. Paris, Ser. I* 345.

[Har8] Hardouin C. (2007) Dumping influence on a non iterative dynamics, in *Recent Advances in Stochastic Modelling and Data Analysis*, Editor C. H. Skiadas, World Scientific Publishing, 371–381.

[Har9] Hardouin C., Yao J.-F. (2008) Multi-parameter auto-models and their application, *Biometrika*. 95(2), 335–349.

[Har10] Hardouin C., Yao J.-F. (2008) Spatial modelling for mixed state observations, *Electronic Journal of Statistics* Vol. 2, 213–233.

[Har11] Hardouin C. (2008) Influence d'une contamination initiale sur une dynamique spatiale non itérative, *Journal de la SFDS*, Tome 149, no 4, 107–129.

[Har12] Guyon X., Hardouin C (2010) Exact marginals and normalizing constant for Gibbs distributions, *C. R. Acad. Sci. Paris, Ser. I* no 348. 199–201.

[Har13] Hardouin C., Crivelli T. (2011) Mixed states spatio-temporal auto-models, *Statistics*, Vol 45 Issue 1, Special issue : Statistics on dependent data, 87–100.

Ce document présente une synthèse de mes travaux de recherche issus et postérieurs à ma thèse; le thème central est la statistique spatiale, plus précisément la modélisation et l'étude statistique de processus spatiaux ou spatio-temporels.

Lors de ma thèse ([Har0]), j'ai abordé plusieurs sujets : estimation par minimum de contraste fort, avec application aux champs de Markov ([Har1]), estimation pour une régression dont le résidu est à longue mémoire, et régression sur le log-périodogramme pour une série stationnaire.

Au sortir de ma thèse j'ai travaillé avec Fabienne Comte sur l'estimation à temps discret des processus fractionnaires, à longue mémoire ou non, par la méthode de régression sur log-périodogramme. Ce gros travail n'a pas donné lieu à publications et est synthétisé dans une note aux CRAS ([Har3]). Parallèlement, je continuais une collaboration avec des collègues du SAMOS sur les statistiques par minimum de contrastes ([Har2]).

Je me suis tournée ensuite vers le domaine de la statistique spatiale, et plus particulièrement vers l'étude des champs de Markov. C'est toujours dans ce domaine que mes recherches sont orientées; les contextes sont variés mais appartiennent fondamentalement au domaine de la statistique spatiale sur grille.

En collaboration avec Xavier Guyon, je me suis d'abord intéressée à l'étude d'un modèle spatio temporel général, consistant en une chaîne de Markov (temporelle) de champs de Markov (spatiaux) ([Har5]). Plus tard, avec Jian-Feng Yao nous avons développé une généralisation des auto-modèles de J. Besag, qui sont un type particulier de champ markovien ([Har7-9]); cette extension nous a permis de proposer une modélisation non hiérarchique pour des données spatiales de nature mixte, composées par exemple d'une masse en zéro accompagnée de valeurs réelles ([Har10]); des applications ont été réalisées en collaboration avec des chercheurs de l'équipe VISTA de l'IRISA sur la modélisation des mesures de mouvement dans des séquences vidéo ([Har6]). La modélisation spatio-temporelle de telles données était alors la suite naturelle de ce travail ([Har13]).

Parallèlement, un collègue économiste nous a proposé de travailler sur la modélisation de l'adoption des standards technologiques. Plus précisément, des agents placés sur un réseau choisissent certains standards technologiques au détriment d'autres suivant une règle contextuelle; les situations sont diverses, il peut s'agir de choix répétitifs réalisés de manière séquentielle ou synchrone, ou au contraire d'une situation de choix unique et définitif; les règles de choix elles-mêmes offrent beaucoup de possibilités. Nous avons fait plusieurs études dans ce contexte ([Har4-8-11]).

En statistique spatiale, le calcul de la constante de normalisation d'une probabilité discrète est en général un problème insoluble; par exemple, pour un modèle d'Ising sur une "petite" grille 10×10 , la constante de normalisation se calcule par sommation sur 2^{100} termes. De nombreux travaux portent sur le contournement de cette difficulté, en ayant recours à la simulation, ou, pour un but d'estimation, en remplaçant la vraisemblance par la pseudo-vraisemblance conditionnelle. Nous nous sommes intéressés avec Xavier Guyon au calcul exact de la constante pour un processus de Gibbs via un algorithme récursif sur les lois conditionnelles; nous avons rejoint des résultats obtenus par N. Friel et H. Rue qui permettent le calcul dans le cas où l'une des dimensions du lattice reste petite ([Har12]). Le problème reste encore ouvert pour des grands lattices.

Je présenterai dans un premier chapitre les résultats obtenus sur les modèles spatiaux et

spatio-temporels, et les approches menées pour la modélisation des données de nature mixte. Puis je développerai dans le chapitre 2 les travaux sur l'adoption de standards technologiques. J'aborderai ensuite le problème du calcul de la constante de normalisation. Enfin, je terminerai par un bref rappel dans le chapitre 4 des travaux plus anciens, ne s'inscrivant pas dans le domaine de la statistique spatiale, sur l'estimation par minimum de contraste et la régression sur log-périodogramme .

Notations et glossaire

$S = \{1, 2, \dots, n\}$ est un ensemble de sites fini.

Si A est un sous ensemble de S , on note $x_A = (x_i, i \in A)$ la configuration de X sur A et $x^A = (x_j, j \in S \setminus A)$ celle à l'extérieur de A .

On note $x^i = x^{\{i\}}$.

Pour un ensemble A , on note $|A|$ son cardinal.

Considérons un graphe de voisinage \mathcal{G} sur S .

$\langle i, j \rangle$ signifie que les sites i et j sont voisins pour \mathcal{G} .

Le voisinage de A est $\partial A = \{j \in S : j \notin A \text{ t.q. } \exists i \in A : \langle i, j \rangle\}$.

On note $\partial i = \partial\{i\}$.

Une *clique* est soit un singleton, soit un ensemble de sites qui sont voisins deux à deux pour le graphe de voisinage.

Un ensemble de *codage* C est un sous-ensemble de S tel que si i et j sont deux points de C avec $i \neq j$, alors i et j ne sont pas voisins. Une conséquence importante est que les variables conditionnelles $\{X_i \mid x_{\partial i}, i \in C\}$ sont indépendantes.

Chapitre 1

Modèles spatiaux

1.1 Introduction

Les champs de Gibbs-Markov fournissent un moyen commode de modéliser des données spatiales ; ils sont utilisés dans de nombreux domaines comme l'analyse d'images, l'économétrie spatiale, les sciences de l'environnement ou l'épidémiologie. Considérons un ensemble fini de sites S sur lequel est défini un graphe de voisinage symétrique \mathcal{G} . Un champ $X = (X_i, i \in S)$ vérifie la propriété de Markov relativement à ce graphe de voisinage si pour tout $A \subset S$, la loi de X_A conditionnelle à la configuration en tous les autres sites ne dépend que de la configuration $x_{\partial A}$ sur le voisinage ∂A de A : $P(X_A | x^A) = P(X_A | x_{\partial A})$. Il en découle qu'un champ de Markov va être défini via ses lois conditionnelles, c'est-à-dire ses caractéristiques locales, plutôt que par des caractéristiques globales. Le problème se pose alors de spécifier des lois conditionnelles qui se recollent effectivement en une loi jointe. Si on suppose que pour tout $A \subset S$ et toute configuration x , $P(X_A | x^A) > 0$, alors le théorème de Hammersley-Clifford ([10]) permet d'écrire le champ de Markov comme un champ de Gibbs pour lequel les potentiels sont définis sur les cliques du graphe de voisinage. Les auto-modèles de Besag ([10]) constituent une classe de champs markoviens particulièrement utilisés en statistique spatiale.

Dans tout ce qui suit, l'espace d'état E est général, il peut être qualitatif ou quantitatif, discret ou non, voire mixte. On supposera qu'il est mesurable, (E, \mathcal{E}) étant muni d'une mesure de référence m et on notera l'espace des configurations $(\Omega = E^S, \mathcal{E}^{\otimes S})$ et $\nu := m^{\otimes S}$ la mesure de référence sur Ω . Notons qu'en général $\Omega = E^S$, mais la plupart des résultats se généralisent à un produit général $\Omega = \prod_{i \in S} E_i$.

Champ de Gibbs.

Si X est un champ de Gibbs ([25]) sur un ensemble de sites S fini, sa distribution jointe s'écrit de la façon suivante :

$$\pi(x) = Z^{-1} \exp U(x).$$

U est l'énergie qui s'écrit comme une somme de *potentiels* définis sur une famille \mathcal{C} de parties W de S : $U(x) = \sum_{W \in \mathcal{C}} \Phi_W(x)$.

Pour rendre ces potentiels identifiables, on se fixe un état de référence τ dans Ω , et on doit avoir pour toute partie W non vide, $\Phi_W(x) = 0$ si pour un $i \in W$, $x_i = \tau_i$.

On dit que l'énergie est admissible lorsque $Z = \sum_{x \in \Omega} \exp U(x) < \infty$ (ou $Z = \int_{\Omega} \exp U(x) \nu(dx) < \infty$) : Z est la *constante de normalisation*.

Les lois conditionnelles d'un champ de Gibbs sont les $\pi_A(x_A | x^A) = Z_A(x^A)^{-1} \exp U_A(x)$ avec $U_A(x) = \sum_{W \in \mathcal{C}: W \cap A \neq \emptyset} \Phi_W(x)$ et $Z_A(x^A) = \sum_{x_A \in \Omega_A} \exp U_A(x_A, x^A)$ (ou $\int_{\Omega_A} \exp U(x_A, x^A) \nu(dx_A)$).

Champ de Gibbs-Markov

Supposons que S est muni d'un graphe de voisinage \mathcal{G} . X est un champ de Markov sur S relativement au graphe \mathcal{G} si pour toute partie $A \subset S$, on a la propriété de Markov $\pi_A(x_A | x^A) = \pi_A(x_A | x_{\partial A})$.

On relie les deux définitions de Gibbs et Markov en prenant pour la famille \mathcal{C} les cliques associées au graphe \mathcal{G} . Le théorème de Hammersley-Clifford (Cf [10] pour une démonstration) établit l'équivalence d'un champ de Markov à un champ de Gibbs sous la *condition de positivité* qui stipule que $\pi_A(x_A | x^A) > 0$ pour toute partie A de S et configuration x de Ω .

Auto-modèles de Besag

Ces modèles particuliers dérivent de deux hypothèses ; la première porte sur les cliques qui sont au plus d'ordre 2 ; en conséquence, l'énergie s'écrit comme une somme de potentiels portant sur les singletons et de potentiels portant sur les paires. La seconde hypothèse caractérise les lois conditionnelles qui doivent appartenir à une famille exponentielle. Nous reviendrons en détail sur ces modèles ultérieurement en 1.3.

1.2 Dynamiques temporelles de champs de Markov ([Har5])

Le modèle que nous avons étudié dans [Har5] est un modèle paramétrique semi-causal appelé CMCM : Chaîne de Markov de Champ de Markov, défini sur un espace d'états général E et un ensemble fini de sites $S = \{1, 2, \dots, n\}$. Le modèle est défini comme suit : $X = (X(t), t \in \mathbb{N}^*)$ est une chaîne de Markov sur $\Omega = E^S$ et $X(t) = (X_i(t), i \in S)$ est, conditionnellement au passé $X(t-1)$, un champ de Markov sur E^S . Nous étudions la structure et les propriétés de ce modèle, ergodicité, identification, estimation et validation. Pour simplifier, nous considérons des dynamiques basées sur des chaînes de Markov d'ordre 1 homogènes dans le temps, mais aucune stationnarité dans l'espace n'est supposée. Les résultats peuvent s'étendre à des chaînes d'ordre supérieur et, hormis l'ergodicité, à des modèles inhomogènes dans le temps.

On écrit la transition de probabilité via une énergie conditionnelle c'est-à-dire

$$P(x, y) = Z^{-1}(x) \exp U(y | x) \quad \text{avec} \quad Z(x) = \int_{\Omega} \exp U(y | x) \nu^S(dy) < \infty \quad (1.1)$$

($Z(x) = \sum_{\Omega} \exp U(y | x) < \infty$ si E est discret).

Posons $X(t-1) = x$ et notons $\mathbf{0}$ un état de référence dans E^S (on prend $\mathbf{0} = 0^S$ si $E = \mathbb{N}, \mathbb{R}$ ou \mathbb{R}_+). Suivant la formule d'inversion de Moebius ([10], [36], [23]), il existe une famille minimale \mathcal{C} (choisie indépendante de x) de sous ensembles non vides de S , et des potentiels conditionnels $\Phi_W^*(\cdot | x)$ pour chaque $W \in \mathcal{C}$ tels que l'énergie s'écrive comme une somme de potentiels sur la famille \mathcal{C} , $U(y | x) = \sum_{W \in \mathcal{C}} \Phi_W^*(y | x)$, avec la condition d'identifiabilité $\Phi_W^*(y | x) = 0$ si pour

un $i \in W$, $y_i = 0$. De plus, pour chaque W de \mathcal{C} , il existe une famille \mathcal{C}_W de parties non vides de S telle que le potentiel Φ_W^* se décompose en

$$\Phi_W^*(y | x) = \Phi_W(y) + \sum_{W' \in \mathcal{C}_W} \Phi_{W',W}(x, y).$$

Ceci signifie que l'énergie U est liée à deux familles de potentiels : les potentiels d'interaction instantanée $\{\Phi_W, W \in \mathcal{C}\}$ et les potentiels d'interaction dans le temps $\{\Phi_{W',W}, W \in \mathcal{C}, W' \in \mathcal{C}_W\}$. Cette représentation semi-causale est associée à un double graphe de voisinage $\mathbf{G} = \{\mathcal{G}, \mathcal{G}_-\}$ où \mathcal{G} et \mathcal{G}_- sont définis par,

$$\langle i, j \rangle_{\mathcal{G}} \Leftrightarrow \langle (t, i), (t, j) \rangle_{\mathbf{G}} \Leftrightarrow \exists W \in \mathcal{C} \text{ t.q. } \{i, j\} \subseteq W \text{ et } \Phi_W \neq 0,$$

$$\langle j, i \rangle_{\mathcal{G}_-} \Leftrightarrow \langle (t-1, j), (t, i) \rangle_{\mathbf{G}} \Leftrightarrow \exists W \in \mathcal{C}, W' \in \mathcal{C}_W \text{ t.q. } \Phi_{W',W} \neq 0, i \in W, j \in W'.$$

On remarque ici que le graphe \mathcal{G}_- est orienté. Le voisinage d'un site i a deux composantes qui sont ses voisins instantanés (au temps t) $\partial i = \{j \in S \setminus \{i\}, \langle i, j \rangle_{\mathcal{G}}\}$, et ses voisins du passé (au temps $t-1$) $\partial i^- = \{j \in S, \langle j, i \rangle_{\mathcal{G}_-}\}$.

Finalement on écrit

$$P(x, y) = Z^{-1}(x) \exp\left\{ \sum_{W \in \mathcal{C}} \Phi_W(y) + \sum_{W_1 \in \mathcal{C}^-, W_2 \in \mathcal{C}} \Phi_{W_1, W_2}(x, y) \right\} \quad (1.2)$$

avec $\mathcal{C}^- = \bigcup_{W \in \mathcal{C}} \mathcal{C}_W$, $\Phi_{W_1, W_2} \equiv 0$ si $W_1 \notin \mathcal{C}_{W_2}$ et les conditions d'identifiabilité $\Phi_W(y) = 0$ (resp. $\Phi_{W_1, W_2}(x, y) = 0$) si pour un $i \in W$ (resp. $i \in W_2$), $y_i = 0$.

Dans un premier temps, nous étudions les propriétés de ces modèles : loi invariante, distributions marginales; puis nous établissons que la dynamique CMCM est équivalente à une représentation non causale d'un champ de Markov espace×temps.

Réversibilité et loi invariante de la chaîne CMCM

Supposons que les potentiels Φ_{W_1, W_2} s'annulent aussi si pour un $i \in W_1, x_i = 0$. On dira qu'une transition P est synchrone si elle résulte d'une relaxation synchrone et indépendante de tous les sites : $P(x, y) = \prod_{s \in S} p_s(x, y_s)$.

Proposition 1 ([Har5] Proposition 1)

(i) la chaîne est réversible dans le temps si et seulement si pour tous $W_1, W_2, x, y : \Phi_{W_1, W_2}(x, y) = \Phi_{W_2, W_1}(y, x)$. Dans ce cas, P admet une unique probabilité invariante donnée par :

$$\pi(y) = \pi(0) \frac{P(0, y)}{P(y, 0)} = \pi(0) Z^{-1}(0) Z(y) \exp \sum_W \Phi_W(y)$$

(ii) cette mesure invariante π n'est pas en général un champ de Markov. Si la transition P est synchrone et réversible, alors π est markovienne.

Représentation non-causale

Si la représentation semi-causale définie précédemment fait intervenir les deux systèmes de voisinage instantané et passé, la représentation non causale fait appel aux voisins du futur (au temps $t+1$). Notons $X(t+1) = z$ et pour $i \in S$, $\partial i^+ = \{j \in S, \langle i, j \rangle_{\mathcal{G}_-}\}$.

Proposition 2 ([Har5] Proposition 2) *La représentation semi-causale (1.2) avec le système de voisinage $\{\partial i, \partial i^-, i \in S\}$ est équivalente à la représentation non causale espace \times temps suivante avec le système de voisinage $\{\partial i, \partial i^-, \partial i^+, i \in S\}$:*

$$P(y | x, z) = Z^{-1}(x, z) \exp\left\{ \sum_{W_1, W_2} \{\Phi_{W_1, W_2}(x, y) + \Phi_{W_1, W_2}(y, z)\} + \sum_W \Phi_W(y) \right\}. \quad (1.3)$$

Nous étudions ensuite quatre exemples importants en modélisation de données réelles : $E = \mathbb{R}$ et les dynamiques auto-normales, $E = \mathbb{R}_+$ et les dynamiques auto-exponentielles, $E = \mathbb{N}$ et les dynamiques auto-poissonniennes, et E qualitatif fini et les dynamiques auto-discrètes. Par exemple, on définit une dynamique auto-exponentielle par l'hypothèse de distribution suivante : pour chaque i , conditionnellement à $(X^i(t) = y^i, X(t-1) = x)$, $X_i(t)$ suit une loi exponentielle de paramètre $\lambda_i(y^i, x)$ (en fait $\lambda_i(y_{\partial i}, x_{\partial i^-})$). Dans chaque cas de distribution, les résultats d'Arnold, Castillo et Sarabia ([6], Cf aussi [4], [3] et [5]) sur le recollement des lois conditionnelles nous donnent la forme de l'écriture de l'énergie conditionnelle $U(y | x)$.

A titre d'exemple, dans le cas de la dynamique auto-exponentielle, nous avons

$$-U(y | x) = \sum_{i \in S} \alpha_i(x) y_i + \sum_{W: |W| \leq 2} \lambda_W(x) y_W, \quad (1.4)$$

avec $y_W = \prod_{i \in W} y_i$, $U(\cdot | x)$ étant admissible si $\alpha_i(x) > 0$ pour tout x et $\lambda_W(x) \geq 0$.

L'énergie d'une dynamique auto-normale s'écrit

$$-U(y | x) = \sum_{i \in S} \{\alpha_i(x) y_i + \beta_i(x) y_i^2\} + \sum_{W: |W| \leq 2} \gamma_W(x) y_W^{l_W}, \quad (1.5)$$

où $y_W^{l_W} = \prod_{i \in W} y_i^{l_i}$, $l_i = 1, 2$, et les fonctions α, β, γ étant telles que toutes les variances conditionnelles soient positives et $U(\cdot | x)$ admissible.

Enfin, la dynamique auto-logistique pour l'espace d'état $E = \{0, 1\}$ est caractérisée par

$$U(y | x) = \sum_{i \in S} \alpha_i(x) y_i + \sum_{\langle i, j \rangle} \beta_{ij} y_i y_j. \quad (1.6)$$

Puisque les potentiels conditionnels sont finis, cette énergie est admissible sans conditions. Conditionnellement à $X(t-1) = x$, $X_i(t)$ suit une loi de Bernoulli de paramètre

$$p_i(y^i, x) = \frac{\exp \delta_i(y^i, x)}{1 + \exp \delta_i(y^i, x)} \text{ avec } \delta_i(y^i, x) = \alpha_i(x) + \sum_{j \in \partial i} \beta_{ij} y_j. \quad (1.7)$$

Ergodicité et Loi des Grands Nombres

Pour chacune de ces quatre dynamiques, on donne des conditions assurant l'ergodicité de la chaîne; certains de ces résultats s'obtiennent en utilisant un critère de Lyapounov; rappelons brièvement les résultats de stabilité de Duflo [20], 6.2.2. Pour $n \geq 0$, soit $\mathcal{F}_n = \sigma(X(s), s \leq n)$ la σ -algèbre engendrée par les $X(s)$, $s \leq n$. Le critère est le suivant : on suppose que la chaîne de Markov considérée définie sur un sous ensemble fermé de \mathbb{R}^d est fortement Fellerienne, et qu'il existe une fonction de Lyapounov V telle que, pour $n > 0$, et pour des constantes $0 \leq \alpha < 1$ et $\beta < \infty$, on ait

$$\mathbb{E}[V(X_n) \mid \mathcal{F}_{n-1}] \leq \alpha V(X_{n-1}) + \beta .$$

On suppose de plus qu'il existe au plus une mesure invariante. Alors, la chaîne est positive récurrente.

De plus, la loi des grands nombres suivante s'applique : si on note μ la mesure invariante, pour toute fonction φ μ -intégrable, on a $\frac{1}{n+1} \sum_{k=0}^n \varphi(X_k) \xrightarrow{a.s.} \mu(\varphi)$. La difficulté est souvent de savoir si φ est μ -intégrable; on utilise alors le résultat suivant qui permet de contourner le problème : la même loi des grands nombres s'applique pour toute fonction φ qui est continue presque sûrement (par rapport à μ) et telle que $|\varphi| \leq aV + b$ pour des constantes a et b .

Suivant ce critère, on obtient par exemple pour la dynamique auto-exponentielle (1.4) la condition d'ergodicité suivante :

- E1** (i) $\forall i \in S, \forall W, x \rightarrow \alpha_i(x)$ et $x \rightarrow \lambda_W(x)$ sont continues.
(ii) $\exists a \in (0, \infty)$, t.q. $\forall x \in (\mathbb{R}^+)^S, \forall i \in S : \alpha_i(x) \geq a$.

Proposition 3 ([Har5] Proposition 3) *Sous l'hypothèse **E1**, la dynamique auto-exponentielle (1.4) est positive récurrente. De plus, la loi forte des grands nombres s'applique pour toute fonction f intégrable, et particulièrement pour les fonctions $x \mapsto f(x)$ vérifiant $|f(x)| \leq \alpha V_r(x) + \beta$ (pour des constantes finies α et β) avec $V_r(x) = \sum_{i \in S} x_i^r$, pour r un entier positif quelconque.*

Nous obtenons de façon analogue des conditions assurant l'ergodicité de la chaîne dans le cas de la dynamique auto-poissonienne ([Har5] Proposition 5). La chaîne auto-discrète sur un espace d'état qualitatif fini est ergodique sans conditions, ses potentiels étant finis. Dans le cas gaussien (1.5), on réécrit le modèle sous une forme autorégressive d'ordre 1, $X(t) = m + AX(t-1) + \varepsilon(t)$ avec ε gaussien, et la condition est classique, portant sur le rayon spectral $\rho(A)$ de la matrice A (Cf par exemple Duflo [20] Théorème 2.3.18).

Proposition 4 ([Har5] Proposition 4) *Si $\rho(A) < 1$, alors la matrice $(I - A)$ est régulière et la chaîne est ergodique de mesure stationnaire gaussienne μ .*

Statistique du modèle CMCM

Le travail sur ce modèle spatio-temporel se poursuit par l'estimation des paramètres; nous donnons les propriétés asymptotiques (consistance, normalité, test de sous hypothèse) des estimateurs de pseudo-vraisemblance conditionnelle; le modèle auto-exponentiel (1.4) est développé en particulier.

Supposons que la transition dépend d'un paramètre inconnu θ , θ appartenant à l'intérieur de Θ , un compact de \mathbb{R}^d . Lorsque la dynamique est ergodique, on obtient de façon standard les propriétés asymptotiques des estimateurs des méthodes classiques d'estimation, comme le maximum de vraisemblance. La difficulté inhérente à la procédure du maximum de vraisemblance réside dans la complexité du calcul de la constante de normalisation $Z_\theta(x)$ dans la vraisemblance. Une alternative usuelle en statistique spatiale est de considérer la pseudo-vraisemblance conditionnelle proposée par Besag en 74 ([10]). En l'absence de dépendance forte, cette procédure conduit à des estimateurs consistants, asymptotiquement normaux avec une vitesse analogue à celle obtenue en utilisant la vraisemblance, avec une perte d'efficacité assez faible en situation de dépendance spatiale modérée ([11], [23], [24]).

Nous obtenons ainsi les résultats asymptotiques de consistance et normalité asymptotique.

Supposons que la distribution conditionnelle de $X_i(t)$ sachant $X^i(t) = y^i$ et $X(t-1) = x$ est absolument continue par rapport à la mesure ν sur E (ν est la mesure de comptage si E est discret ou la mesure de Lebesgue pour $E \subseteq \mathbb{R}^p$), de densité strictement positive $f_i(y_i, y^i, x; \theta)$ (en fait $f_i(y_i, y_{\partial i}, x_{\partial i^-}; \theta)$).

On note θ_0 la vraie valeur du paramètre, et P_0 la transition associée. Le processus est observé aux temps $t = 0, \dots, T$. On considère l'estimateur $\hat{\theta}_T = \arg \min_{\theta \in \Theta} U_T(\theta)$ minimisant l'opposé du log de la pseudo-vraisemblance conditionnelle

$$U_T(\theta) = -\frac{1}{T} \sum_{t=1}^T \sum_{i \in S} \ln f_i(x_i(t), x^i(t), x(t-1); \theta)$$

Conditions pour la consistance (C) :

C1 : Pour $\theta = \theta_0$, la chaîne X est ergodique de mesure stationnaire unique μ_0 .

C2 : (i) Pour tout $i \in S$, $x, y \in E$, $\theta \mapsto f_i(y_i, y^i, x; \theta)$ est continue..

(ii) Il existe une fonction mesurable, $\mu_0 \otimes P_0$ -intégrable h sur $E \times E$ telle que pour tous $i \in S$, $\theta \in \Theta$, $x, y \in E$, $|\ln f_i(y_i, y^i, x; \theta) - \ln f_i(y_i, y^i, x; \theta_0)| \leq h(y, x)$.

C3 : Identifiabilité : si $\theta \neq \theta_0$ alors $\sum_{i \in S} \mu_0(\{x \text{ t.q. } f_i(\cdot, \cdot, x; \theta) \neq f_i(\cdot, \cdot, x; \theta_0)\}) > 0$.

Notons $f_i^{(1)}(\theta)$ et $f_i^{(2)}(\theta)$ le gradient et la matrice Hessienne de $f_i(y_i, y^i, x; \theta)$ relativement à θ . On définit les matrices de (pseudo) information conditionnelles, pour $\theta \in V_0$, un voisinage de θ_0 , $V_0 \subset \Theta$:

$$I_{ij}(y^{\{i,j\}}, x, \theta) = \mathbf{E}_{\theta_0} \left[\frac{f_i^{(1)}(\theta) f_j^{(1)}(\theta)^y}{f_i(\theta) f_j(\theta)} \mid X^{\{i,j\}}(t) = y^{\{i,j\}}, X(t-1) = x \right],$$

$$I_{ij}(x, \theta) = \mathbf{E}_{\theta_0} [I_{ij}(X^{\{i,j\}}(t), X(t-1), \theta) \mid X(t-1) = x].$$

Conditions pour la normalité asymptotique (N) :

N1 : La fonction $\theta \mapsto f_i(y_i, y^i, x; \theta)$ est deux fois continûment différentiable sur un voisinage $V_0 \subset \overset{\circ}{\Theta}$ de θ_0 , et il existe une fonction mesurable, de $\mu_0 \otimes P_0$ carré intégrable H sur $E \times E$ telle que pour tous $\theta \in V_0$, $x, y \in E$, $1 \leq u, v \leq d$:

$$\left| \frac{1}{f_i} \frac{\partial}{\partial \theta_u} f_i(y_i, y^i, x; \theta) \right| \text{ et } \left| \frac{1}{f_i} \frac{\partial^2}{\partial \theta_u \partial \theta_v} f_i(y_i, y^i, x; \theta) \right| \leq H(y, x)$$

N2 : $I_0 = \sum_{i \in S} E_{\mu_0} [I_{ii}(X(t-1), \theta_0)]$ et $J_0 = \sum_{i,j \in S} E_{\mu_0} [I_{ij}(X(t-1), \theta_0)]$ sont définies positives.

Proposition 5 ([Har5] Proposition 6)

(i) Sous les conditions **(C)**, $\hat{\theta}_T \xrightarrow[T \rightarrow \infty]{P_0} \theta_0$

(ii) Si l'on suppose de plus **(N)**, $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d(0, I_0^{-1} J_0 I_0^{-1})$

Cette propriété asymptotique est obtenue à partir des résultats généraux sur l'estimation par minimum de contraste; pour l'estimation des champs de Markov, on peut se référer en particulier à Besag ([11]), [Har1], ou [8].

Lorsque la densité conditionnelle est dans une famille exponentielle, on donne des conditions assurant l'identifiabilité **(C3)** ainsi que la régularité de la matrice I_0 ([Har5], Proposition 7). Nous reprenons également l'exemple de la dynamique auto-exponentielle et spécifions les conditions de consistance et de normalité asymptotique ([Har5], Proposition 9).

Test de sous hypothèse

A la suite des propriétés sur l'estimation, nous proposons un test de sous-modèle emboîté. Il s'agit de tester l'hypothèse $(H_q) : \theta = \varphi(\alpha)$, $\alpha \in \mathbb{R}^q$, $q < d$, où φ est une fonction de \mathbb{R}^q dans \mathbb{R}^d ; sous de bonnes conditions de régularité et de rang sur φ , on construit une statistique de test basée sur la différence des pseudo-vraisemblances conditionnelles $2T \left(U_T(\bar{\theta}_T) - U_T(\hat{\theta}_T) \right)$ où $\bar{\theta}_T$ et $\hat{\theta}_T$ sont respectivement les estimateurs de pseudo-vraisemblance conditionnelle sous les deux hypothèses (H_q) et (H_d) . Reprenant des résultats de [Har2], on sait que la loi asymptotique de cette statistique est une somme pondérée de chi-2 à un degré de liberté ([Har5] Proposition 8, Cf Chapitre 4).

Identification et validation de modèle

Enfin, nous regardons les problèmes d'identification puis de diagnostic et validation de modèles. Supposons que l'on veuille identifier un modèle CMCM semi-causal. La première étape est la détermination du double graphe de voisinage \mathcal{G} et \mathcal{G}^- ; il y a deux stratégies, le choix résultant de la complexité du problème et du nombre de sites; la première procédure est globale, en utilisant un algorithme de maximisation de la pseudo-vraisemblance conditionnelle avec un critère de pénalisation de type AIC; la seconde possibilité, moins coûteuse, est de travailler localement en déterminant site par site, les voisinages ∂i et ∂i^- pour chaque $i \in S$, et en harmonisant ensuite les voisinages instantanés obtenus pour obtenir un graphe symétrique.

Le graphe étant établi et les paramètres estimés, nous proposons une procédure de validation du modèle basée sur les résidus conditionnels dans le cas où l'espace d'état est une partie de \mathbb{R} .

Notant toujours θ_0 et $\hat{\theta}_T$ la vraie valeur du paramètre et son estimateur de pseudo-vraisemblance conditionnelle, on pose

$$\varepsilon_{it} = X_i(t) - \mu_{it} \quad \text{et} \quad \hat{\varepsilon}_{it} = X_i(t) - \hat{\mu}_{it} \quad (1.8)$$

où $\mu_{it}(\theta) = E[X_i(t) | X^i(t), X(t-1); \theta]$, $\mu_{it} = \mu_{it}(\theta_0)$, et $\hat{\mu}_{it} = \mu_{it}(\hat{\theta}_T)$.

Nous construisons deux tests, l'un basé sur les résidus estimés $\hat{\varepsilon}_{it}$ et l'autre sur les carrés des résidus estimés $\hat{\varepsilon}_{it}^2$.

Soit \mathcal{C} un sous-ensemble de codage de S pour le graphe instantané (Cf Notations, deux sites i, j ($i \neq j$) de cet ensemble de codage ne sont pas voisins pour le graphe \mathcal{G}): alors, conditionnellement au passé $X(t-1)$ et à la réalisation $X_{\mathcal{C}}(t)$ de X en dehors de \mathcal{C} et au temps t , $X_i(t)$ et $X_j(t)$ sont indépendants. Posant $e_{it} = \frac{\varepsilon_{it}}{\sigma(\varepsilon_{it})}$, on peut montrer que $\tilde{e}_t = \sum_{i \in \mathcal{C}} e_{it}$ est

une différence de martingale de carré intégrable par rapport à la filtration $(\mathcal{F}_t = \sigma(X(s), s \leq t), t \geq 0)$. On applique alors un théorème de limite centrale pour les martingales ([20] Corollaire 2.1.10, voir aussi [27]) ; rappelons ce résultat.

Pour une martingale de carré intégrable (M_t) , notons $(\langle M \rangle_t)$ son processus croissant défini par

$$\langle M \rangle_t = \langle M \rangle_{t-1} + E[|M_t - M_{t-1}|^2 | \mathcal{F}_{t-1}] \text{ for } t \geq 1, \text{ and } \langle M \rangle_0 = 0.$$

La première condition pour établir le TCL est de vérifier $\frac{\langle M \rangle_T}{T} \xrightarrow[T \rightarrow \infty]{P} \Gamma$, pour Γ symétrique semi définie positive. La seconde condition est une condition de Lindeberg, mais on utilise plutôt

$$\exists \alpha > 2 \text{ tel que } E[|\tilde{e}_t|^\alpha | \mathcal{F}_{t-1}] \text{ est borné.} \quad (1.9)$$

(1.9) implique la condition de Lindeberg, plus difficile à vérifier.

Dans notre cadre, posant $M_s = \sum_{t=1}^s \tilde{e}_t$, on a $\langle M \rangle_t - \langle M \rangle_{t-1} = |\mathcal{C}|$ et la première condition est vérifiée pour toute dynamique. En revanche, la seconde condition (1.9) devra être vérifiée au cas par cas. Si elle est vérifiée, on a alors le résultat suivant ([Har5] Proposition 10) :

$$\frac{1}{\sqrt{|\mathcal{C}|T}} \sum_{t=1}^T \tilde{e}_t \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1)$$

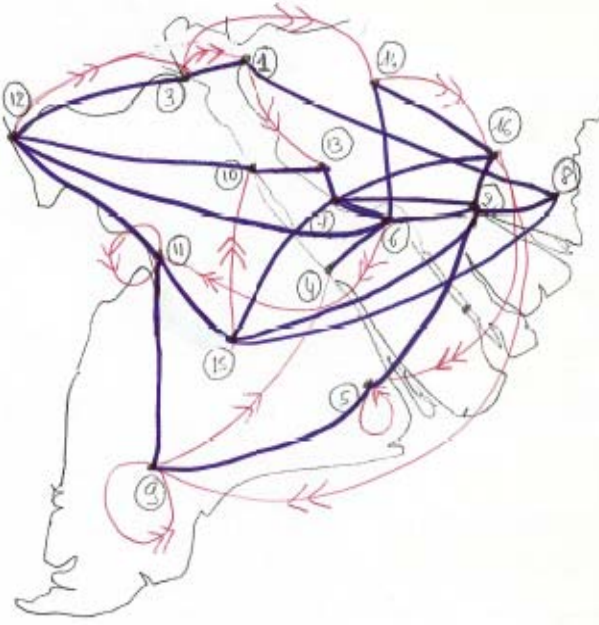
On travaille de façon analogue sur les carrés des résidus $w_{it} = \frac{e_{it}^2 - 1}{\sigma(e_{it}^2)}$, qui ont les mêmes propriétés que les e_{it} ([Har5] Proposition 11).

La vraie valeur θ_0 du paramètre étant inconnue, on la remplace par l'estimateur $\hat{\theta}_T$; en effet, lorsque celui-ci est consistant, les résultats précédents transposés aux résidus estimés restent valides.

Nous reprenons alors les exemples de dynamiques considérées (auto-exponentielles, auto-poissonniennes, auto-discrètes, auto-normales) et donnons les conditions exactes pour obtenir (1.9) pour ces modèles et les TCL qui en découlent.

Application sur une dynamique météorologique

Ces différents résultats sont confrontés à l'étude d'une application sur des données réelles. Les données consistent en des mesures journalières de pluie, effectuées sur 123 jours consécutifs de juillet à octobre 1983 en 16 sites du delta du Mekong (Vietnam), présentés Figure 1. Nous proposons un modèle CMCM auto-logistique sur les données binarisées (0 s'il ne pleut pas, 1 sinon) ; nous avons suivi la procédure décrite auparavant : identification des graphes de voisinage par procédure locale, estimation des paramètres du modèle par pseudo-vraisemblance conditionnelle, et validation par un test sur les résidus estimés calculés sur un ensemble de codage.



Site number	Name	Latitude	Longitude
1	Tan Chau	10.80	105.25
2	My Tho	10.35	106.37
3	Chau Doc	10.70	105.01
4	Can Tho	10.03	105.78
5	Soc Tran	09.60	105.97
6	Vinh Long	10.25	105.97
7	Sa Dec	10.30	105.75
8	Go Cong	10.35	106.67
9	Ca Mau	09.17	105.17
10	Long Xuyen	10.40	105.42
11	Rach Gia	10.00	105.08
12	Ha Tien	10.38	104.48
13	Cao Lanh	10.47	105.63
14	Moc Hoa	10.75	105.92
15	Vi Thanh	09.77	105.45
16	Tan An	10.53	106.40

FIG 1.1 - Stations météorologiques et graphe de voisinage

Le graphe de voisinage représenté Figure 1.1 a été estimé par régressions locales; cette procédure nous a conduit à considérer un double graphe composé de 25 arêtes d'interaction instantanée accompagnées de 17 interactions dans le temps. Nous avons alors estimé un modèle auto-logistique (1.7) : en chaque site i , conditionnellement à $X_{\partial i}(t)$ et $X_{\partial i^-}(t)$, $X_i(t)$ suit une loi de Bernoulli de paramètre $p_i(t)$ avec $p_i(t) = (1 + \exp -\lambda_i(t))^{-1}$ et $\lambda_i(t) = \delta_i + \sum_{j \in \partial i} \beta_{ij} X_j(t) + \sum_{l \in \partial i^-} \alpha_{li} X_l(t-1)$. La dimension paramétrique de ce modèle est $16+25+17=58$. Cette dimension étant importante, la procédure d'estimation-validation a été faite en deux étapes; une première procédure de minimisation des pseudo-vraisemblances conditionnelles site par site a fourni un premier jeu de paramètres et des statistiques locales de validation $V_i = \frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{e}_{it} \cdot \text{sign}(\hat{e}_{it})$. Les valeurs des paramètres ont servi d'initialisation pour l'estimation par pseudo-vraisemblance conditionnelle globale et la statistique de validation $V = \frac{1}{\sqrt{CT}} \sum_{i \in \mathcal{C}} \sum_{t=1}^T \hat{e}_{it} \text{sign}(\hat{e}_{it})$ a été calculée sur l'ensemble de codage $\mathcal{C} = \{3, 4, 5, 7, 8, 10, 11, 14\}$. Chaque modèle local a été accepté, comme le modèle global final.

Une autre validation possible est de comparer les vraies données à celles prédites par le modèle, et nous obtenons pour ce modèle un taux de similarité moyen de 77.4%. Ce modèle se révèle donc assez satisfaisant; il peut être amélioré en considérant une distribution plus complexe que l'auto-logistique : ces données de pluie exhibant une part importante de valeurs nulles, ceci nous a conduit à envisager des modèles spécifiques pour de telles données "mixtes", constituées d'une masse en zéro accompagnée de valeurs positives. La construction de ces modèles passe par celle, plus générale, des auto-modèles pour plusieurs paramètres.

1.3 Généralisation des auto-modèles de Besag ([Har7, Har9])

Ce travail a fait l'objet d'une collaboration avec Jian-Feng Yao, Université de Rennes 1.

Auto-modèles multi-paramètres

J. Besag a proposé en 1974 des modèles caractérisés par leur famille de lois conditionnelles. Le recollement des lois conditionnelles est un problème classique en statistique spatiale. Considérons un lattice $S = \{1, \dots, n\}$ muni d'un graphe de voisinage symétrique et un espace d'état $E \subset \mathbb{R}$. Si la distribution de probabilité jointe P est positive presque partout ("condition de positivité"), le Théorème de Hammersley-Clifford permet d'établir que $\log P$ est proportionnel à une somme de potentiels basés sur les sous-ensembles de S ; reprenons les notations de Besag ([10]) :

$$P(x) = Z^{-1} \exp Q(x) ,$$

$$Q(x) = \sum_{i \in S} x_i G_i(x_i) + \sum_{i,j} x_i x_j G_{i,j}(x_i, x_j) + \sum_{i,j,k} x_i x_j x_k G_{i,j,k}(x_i, x_j, x_k) + \dots$$

$$+ x_1 x_2 \dots x_n G_{1,2,\dots,n}(x_1, x_2, \dots, x_n) \quad (1.10)$$

Dans ce cadre, J. Besag définit les auto-modèles en supposant que premièrement les interactions ne concernent que les singletons et les paires, c'est-à-dire

$$Q(x) = \sum_{i \in S} x_i G_i(x_i) + \sum_{\{i,j\}} x_i x_j G_{i,j}(x_i, x_j) .$$

Cette première hypothèse est connue sous le nom de "pairwise only dependence". La seconde hypothèse est que chaque densité conditionnelle appartient à une famille exponentielle, pour chaque i ,

$$\log p_i(x_i | x^i) = A_i(x^i) B_i(x_i) + C_i(x_i) + D_i(x^i) .$$

où en fait le conditionnement en x^i est fonction de $x_{\partial i}$, les fonctions A_i et B_i désignant respectivement le paramètre naturel et la statistique exhaustive associée, et D_i définissant la constante de normalisation appropriée.

Besag montre alors que les fonctions A_i prennent nécessairement la forme

$$A_i(\cdot) = \alpha_i + \sum_j \beta_{ij} B_j(x_j)$$

avec $\beta_{ij} = \beta_{ji}$ et $\beta_{ij} = 0$ si les sites i et j ne sont pas voisins. De même, les potentiels s'écrivent

$$G_{ij}(x_i, x_j) = \beta_{ij} H_i(x_i) H_j(x_j)$$

avec $x_i H_i(x_i) = B_i(x_i) - B_i(0)$.

Enfin, Besag fait l'hypothèse additionnelle que pour chaque i , la statistique B_i est linéaire en x_i . Il en découle une famille de modèles nommés d'après la forme des distributions conditionnelles choisies : modèles auto-logistique, auto-binomial, auto-exponentiel, auto-normal...

Soulignons deux limitations de ces modèles. La première est que les lois conditionnelles en question sont en dimension 1 dans la mesure où la statistique exhaustive et le paramètre naturel associé sont scalaires. Plus précisément, si l'on considère par exemple une loi conditionnelle

gaussienne au site i , seule la moyenne dépend du voisinage, la variance étant constante ou ne dépendant que de i .

La seconde réside dans les conditions d'utilisation de ces modèles. Il faut s'assurer que $\exp Q(x)$ est sommable ou intégrable ; cette condition d'intégrabilité va entraîner des contraintes sur les paramètres qui correspondent à des comportements de compétition entre sites voisins dans beaucoup de modèles (auto-exponentiel, auto-Poisson par exemple), ce qui est une limitation pour les applications. Une solution pour pallier ce problème est de considérer des modèles tronqués.

On trouve plusieurs extensions des auto-modèles dans la littérature ; Kaiser et Cressie ([31]) ont proposé un allègement de la condition de positivité, préalable à tout modèle ; Lee, Kaiser et Cressie ([33]) ont étendu la première condition sur la dépendance entre les sites en "multiway dependence". A notre connaissance, la première tentative pour relaxer la condition sur la dimension de la famille exponentielle est dans Cressie et Lele ([13]) où le terme *multi-parameter exponential family Markov random field models* apparaît. Puis Kaiser et Cressie ([31]) analysent en détail un modèle spatial basé sur des lois conditionnelles Beta (loi dans une famille exponentielle à deux paramètres). De notre côté, nous avons travaillé sur l'idée de considérer des familles exponentielles pour lesquelles la statistique exhaustive ainsi que le paramétrage naturel associé sont en dimension quelconque. Nous donnons un résultat analogue à celui de Besag ([10]) pour un espace d'état général E . L'utilisation de lois dans une famille exponentielle *multi-paramètre* permet de résoudre les problèmes de comportement de compétition entre sites voisins. Par exemple, les lois Beta permettent de considérer des modèles aussi bien coopératifs que compétitifs. Ces modèles s'estiment par pseudo-vraisemblance conditionnelle et nous donnons un résultat de consistance pour l'estimateur associé. Présentons cette généralisation des auto-modèles multi-paramètres.

Dans tout ce qui suit, l'espace des sites $S = \{1, \dots, n\}$ est muni d'un graphe de voisinage, et on considère un espace d'état mesuré (E, \mathcal{E}, m) et l'espace des configurations associé $(\Omega = E^S, \mathcal{E}^{\otimes S})$ avec $\nu := m^{\otimes S}$. L'espace d'état E est général (pas nécessairement une partie de \mathbb{R}) et nous présentons les résultats pour $\Omega = E^S$, mais ceux-ci valent encore pour un espace des configurations plus général $\Omega = \prod_{i \in S} E_i$, que nous considérerons ultérieurement dans le cas des données de nature mixte. Nous restons dans le cadre de la condition de positivité ; l'énergie Q est alors une somme de potentiels définis sur des cliques, et retenant l'écriture générale, on a

$$P(x) = Z^{-1} \exp Q(x) , \text{ où}$$

$$\begin{aligned} Q(x) = & \sum_{i \in S} G_i(x_i) + \sum_{i,j} G_{i,j}(x_i, x_j) + \sum_{i,j,k} G_{i,j,k}(x_i, x_j, x_k) + \dots \\ & + G_{1,2,\dots,n}(x_1, x_2, \dots, x_n) . \end{aligned} \quad (1.11)$$

La généralisation des auto-modèles au cadre multi-paramètre est donnée par les trois hypothèses de base suivantes. La première est encore la "pairwise only dependence" :

$$\mathbf{[B1]} \quad Q(x) = \sum_{i \in S} G_i(x_i) + \sum_{\{i,j\}} G_{ij}(x_i, x_j) .$$

Pour s'assurer de l'identification des potentiels, fixons une configuration de référence $\tau = (\tau_i) \in \Omega$; les potentiels ci-dessus sont déterminés de façon unique si l'on suppose pour tout i, j ,

et $x \in \Omega$ que $G_{ij}(\tau_i, x_j) = G_{ij}(x_i, \tau_j) = G_i(\tau_i) = 0$. Dans la plupart des cas $\tau = (0, \dots, 0)$, mais ce choix est en fait arbitraire et par exemple, nous prendrons $\tau = (\frac{1}{2}, \dots, \frac{1}{2})$ dans le cas des lois conditionnelles Beta.

La seconde hypothèse généralise le schéma de la famille exponentielle au cadre multi-paramètre.

[B2] $\log p_i(x_i | x^i) = \langle A_i(x^i), B_i(x_i) \rangle + C_i(x_i) + D_i(x^i)$, $A_i(x^i) \in \mathbb{R}^l$, $B_i(x_i) \in \mathbb{R}^l$,
avec $B_i(\tau_i) = C_i(\tau_i) = 0$.

Nous ajoutons une condition de régularité vectorielle sur les statistiques exhaustives ; cette condition est peu restrictive et est satisfaite dans la plupart des exemples.

[C] Pour tout $i \in S$, $\text{Span}\{B_i(x_i) : x_i \in E\} = \mathbb{R}^l$.

Alors, le résultat suivant généralise celui de Besag ([10]) et donne la forme nécessaire des paramètres canoniques et l'écriture des potentiels.

Théorème 6 ([Har9] Théorème 1) *Supposons que les trois conditions [B1], [B2] et [C] soient vérifiées ; alors il existe une famille $\{\alpha_i : i \in S\}$ de vecteurs de dimension l et une famille de matrices $\{\beta_{ij} : i, j \in S, i \neq j\}$ de taille $l \times l$ vérifiant $\beta_{ji} = \beta_{ij}^T$ telles que*

$$A_i(x^i) = \alpha_i + \sum_{j:j \neq i} \beta_{ij} B_j(x_j), \quad i \in S \quad (1.12)$$

De plus, les potentiels s'écrivent :

$$\begin{aligned} G_i(x_i) &= \langle \alpha_i, B_i(x_i) \rangle + C_i(x_i), \\ G_{ij}(x_i, x_j) &= B_i^T(x_i) \beta_{ij} B_j(x_j). \end{aligned} \quad (1.13)$$

Nous donnons également un résultat de réciprocity : si on considère un champ défini par des potentiels de la forme (1.13), et si $\int_{\Omega} e^{Q(x)} \nu(dx) < \infty$, alors la famille de lois conditionnelles $p_i(x_i | x^i)$ est dans une famille exponentielle multi-paramètre dont le paramètre naturel satisfait (1.12) ([Har9] Proposition 1).

Notons ici que l'on n'impose ici aucune symétrie sur les β_{ij} et ce modèle peut s'appliquer à des graphes orientés ou non. Une illustration simple est la suivante : considérons un couple de variables (X_1, X_2) tel que la distribution conditionnelle de X_1 sachant $X_2 = x_2$ est une loi Gamma, et celle de X_2 sachant $X_1 = x_1$ est gaussienne. Dans cet exemple, $S = \{1, 2\}$, les deux espaces d'état sont différents et le modèle n'est pas symétrique. Ici, on prend pour configuration de référence $\tau = (1, 0)$, les statistiques exhaustives sont respectivement $B_1(x) = (-x + 1, \log x)^T$ et $B_2(x) = (x, x^2)^T$, et la condition [C] est vérifiée. D'après le résultat précédent, il existe deux vecteurs α_1, α_2 de \mathbb{R}^2 et une matrice β 2×2 telle que la densité de la loi jointe s'écrive $P(x_1, x_2) = P(\tau) \exp Q(x_1, x_2)$ avec $Q(x_1, x_2) = \langle \alpha_1, B_1(x_1) \rangle + \langle \alpha_2, B_2(x_2) \rangle + B_1^T(x_1) \beta B_2(x_2)$. Si la matrice β n'est pas symétrique (et rien n'exige qu'elle le soit), ce modèle a 8 paramètres, qui doivent vérifier que l'énergie Q est intégrable. Ce résultat est déjà connu, présenté dans [6], §4.8, mais son obtention ici est très simple via l'application du théorème.

Auto-modèles de lois conditionnelles Beta

Comme nous l'avons déjà souligné, un inconvénient de certains auto-modèles est que les conditions d'admissibilité sur les paramètres impliquent la compétition spatiale entre sites voisins, ce qui est souvent contraire à la situation observée (modèles auto-exponentiels par exemple). Une façon de contourner le problème est d'envisager la troncature ou la censure de la loi conditionnelle concernée. Une autre possibilité est de considérer directement une loi conditionnelle qui n'induit pas de contraintes ; c'est le cas des lois Beta. Nous étudions en détail ce modèle dans [Har9] ; nous donnons une condition suffisante pour son existence ([Har9] Proposition 2), puis nous examinons sous quelles hypothèses il exhibe un comportement de coopération ou de compétition spatiale. Nous examinons ensuite l'exemple d'un lattice rectangulaire $S = [1, M] \times [1, N]$ avec le système de voisinage des 4 ou des 8 plus proches voisins ; dans les deux cas, nous supposons le modèle invariant par translation, avec symétrie spatiale mais anisotropie dans les deux directions horizontales et verticales, et coopération spatiale.

Estimation

On sait que la méthode du maximum de vraisemblance est difficile à mettre en oeuvre dans le cadre des champs markoviens du fait du calcul de la constante de normalisation ; une alternative est d'utiliser la pseudo-vraisemblance conditionnelle (Besag [10]). Nous donnons dans [Har9]-Théorème 2 un résultat de consistance presque sûre pour les estimateurs des auto-modèles multi-paramètres sur un lattice rectangulaire sous hypothèse de spécification invariante par translation, l'asymptotique portant sur la taille du lattice ; les hypothèses du théorème sont classiques, et la preuve repose notamment sur la convergence des estimateurs pour des contrastes convexes (Cf [37], ou [23] Théorème 3.4.4).

Nous appliquons ce résultat général aux auto-modèles Beta définis précédemment ([Har9] Proposition 3).

Application sur données simulées

Nous avons simulé sur un lattice 64×64 un auto-modèle de lois conditionnelles Beta, invariant par translation, symétrique, anisotropique, spatialement coopératif et avec le voisinage des 4 ou des 8 plus proches voisins. Dans chaque cas nous calculons les estimateurs par pseudo-vraisemblance conditionnelle pour 1600 simulations et obtenons des estimations consistantes.

Puis nous reprenons un modèle aux 4 plus proches voisins utilisé dans [32] pour la modélisation de données réelles d'arbres malades. Nous mesurons empiriquement la vitesse de convergence des estimateurs en considérant des lattices de taille croissante ; nous nous intéressons également à la distribution des estimateurs en traçant des graphes quantiles quantiles (gaussiens) pour des tailles croissantes du lattice. Ces deux dernières expérimentations concluent à une convergence gaussienne des estimateurs à la vitesse \sqrt{n} .

1.4 Modèles spatiaux à états mixtes ([Har6, Har10, Har13])

1.4.1 Variables à états mixtes

Le travail d'application des chaînes MCMC pour les données de pluviométrie nous a montré qu'il était difficile de traiter ces données du fait de l'importance des valeurs nulles : comment construire des modèles spécifiques pour ce genre de données ? De son côté, en collaboration avec des chercheurs de l'équipe VISTA de l'IRISA/INRIA, Jian-Feng Yao s'intéressait aussi à ce type de données dans un contexte différent, celui des analyses de textures de mouvement, pour lesquelles on mesure dans une séquence vidéo le mouvement des pixels ou leur absence de mouvement, ce qui nous a conduit à travailler ensemble sur la modélisation des données de nature mixte. Ce travail a fait l'objet d'une publication collective [Har6] avec une application en imagerie, et d'une autre, plus théorique et générale, [Har10].

Les données à états mixtes se rencontrent dans divers domaines, par exemple en modélisation économétrique (absence de dépenses « culturelles » d'un ménage ou montant des dépenses et modèle TOBIT [2]), en météorologie (absence ou hauteur de précipitation), en analyse du mouvement d'une séquence vidéo (pixel immobile ou vitesse d'un pixel), en épidémiologie (absence d'une maladie et gravité si la maladie est présente)... La figure 1.2 ci-dessous montre des exemples d'histogrammes pour de telles données ; à gauche, l'histogramme pour 123 jours consécutifs de pluie dans une station météorologique du delta du Mékong ; à droite l'histogramme d'une vitesse de mouvement pour une séquence vidéo d'une foule (grossièrement la différence renormalisée de l'intensité d'un pixel dans deux images consécutives).

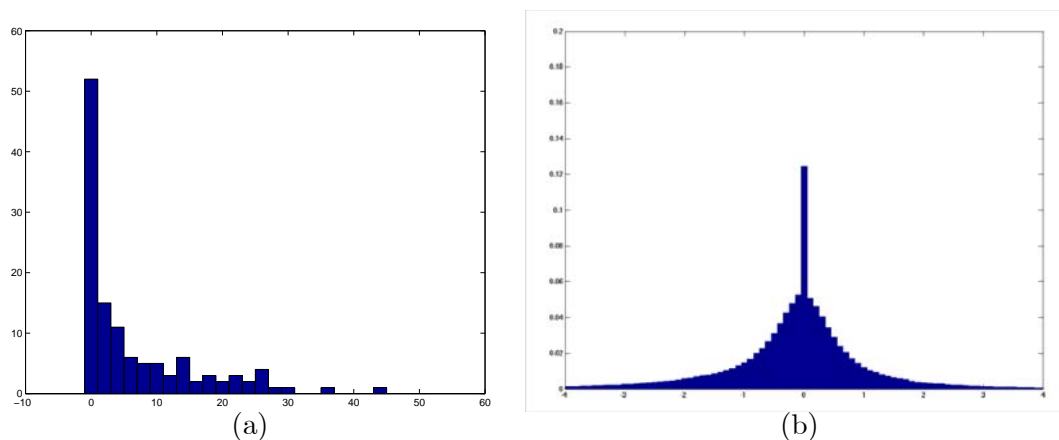


FIG 1.2 - (a) Distribution de 123 jours de pluie, et (b) Vitesse de mouvement renormalisée d'un pixel.

Lorsque les données sont enregistrées sur un réseau spatial S , elles font habituellement l'objet d'une étude via un modèle hiérarchique où le premier niveau considère par exemple l'absence/présence de pluie et le second niveau étudie la distribution conditionnelle de la pluie lorsque la hauteur des précipitations est positive (Cf par exemple [1]). La généralisation des auto-modèles multivariés définis précédemment en 1.3 nous permet une modélisation directe (non hiérarchique), souple, et propice tant à la simulation (méthode MCMC) qu'à l'estimation de modèle (par pseudo-vraisemblance conditionnelle).

Variable aléatoire à états mixtes dans $E = \{0\} \cup (0, \infty)$

Nous définissons une variable aléatoire à états mixtes de la façon suivante ([Har6-9-10]) dans le cas où l'espace d'état est $E = \{0\} \cup (0, \infty)$. On pose $X = 0$ avec probabilité $\gamma \in [0, 1]$, et sinon, avec probabilité $1 - \gamma$, X a une distribution continue sur $(0, \infty)$, de densité de probabilité g .

Formellement, l'espace d'état mesurable (E, \mathcal{E}) est muni d'une mesure de référence mixte

$$m(dx) = \delta_0(dx) + \lambda(dx), \quad (1.14)$$

où δ_0 est la mesure de Dirac en 0 et λ est la mesure de Lebesgue sur \mathbb{R} . Notons $\delta^*(x) = 1 - \mathbf{1}_{\{0\}}(x)$; la variable mixte X admet la densité f par rapport à m ,

$$f(x) = \gamma \mathbf{1}_{\{0\}}(x) + (1 - \gamma) \delta^*(x) g(x), \quad x \in E. \quad (1.15)$$

Dans l'idée de proposer une modélisation spatiale basée sur des auto-modèles, on considère des variables mixtes pour lesquelles la densité g est dans une famille exponentielle, possiblement multi-paramètre, de dimension ℓ :

$$g(x) = g_\xi(x) = H(\xi) L(x) \exp\{\langle \xi, T(x) \rangle\}, \quad \xi \in \mathbb{R}^\ell, \quad T(x) \in \mathbb{R}^\ell, \quad (1.16)$$

T et ξ désignant respectivement la statistique exhaustive et le paramètre naturel.

Alors on peut montrer que la densité f de X est aussi dans une famille exponentielle, de dimension $l + 1$:

$$\begin{aligned} f(x) &= f_\theta(x) = \gamma \mathbf{1}_{\{0\}}(x) + (1 - \gamma) \delta^*(x) g_\xi(x) \\ &= \exp \left[\delta^*(x) \ln \frac{(1 - \gamma) H(\xi)}{\gamma} + \langle \xi, T(x) \delta^*(x) \rangle + \ln \gamma + \delta^*(x) \ln L(x) \right] \\ &= H'(\theta) L'(x) \exp\{\langle \theta, B(x) \rangle\}, \end{aligned} \quad (1.17)$$

avec $H'(\theta) = \gamma$, $L'(x) = \exp\{\delta^*(x) \ln L(x)\}$; le paramètre canonique et la statistique exhaustive sont :

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \ln \frac{(1 - \gamma) H(\xi)}{\gamma} \\ \xi \end{pmatrix}, \quad B(x) = \begin{pmatrix} \delta^*(x) \\ T(x) \delta^*(x) \end{pmatrix}, \quad x \in E. \quad (1.18)$$

On retrouve les paramètres d'origine du modèle via la correspondance

$$\xi = \theta_2, \quad \gamma = \frac{H(\theta_2)}{H(\theta_2) + e^{\theta_1}}.$$

Donnons quelques exemples :

Exemple 1. Distribution exponentielle mixte ; on prend $g_\lambda(x) = \lambda e^{-\lambda x}$ avec $\lambda > 0$. Alors $\theta = (\ln \frac{(1 - \gamma)\lambda}{\gamma}, \lambda)^\top$ et $B(x) = (\delta^*(x), -x)^\top$.

Exemple 2. Distribution gaussienne positive mixte ; g est la densité du module d'une variable gaussienne, $g_\sigma(x) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$; on a

$$f_\theta(x) = \exp\{\langle \theta, B(x) \rangle + \ln \gamma\} \quad (1.19)$$

avec $\theta = (\ln \frac{2(1 - \gamma)}{\gamma\sigma\sqrt{2\pi}}, \frac{1}{2\sigma^2})^\top$ et $B(x) = (\delta^*(x), -x^2)^\top$.

Généralisation pour des états mixtes multiples ([Har10])

Si l'espace d'état considéré dans les applications est souvent $E = \{0\} \cup (0, \infty)$, il peut être plus général avec plusieurs "atomes" et une partie de \mathbb{R}^p . Nous généralisons dans [Har10] la définition des variables à états mixtes.

Soit $F = \{e_1, \dots, e_M\}$ un ensemble fini de M éléments et G un sous ensemble de \mathbb{R}^p . Soient $\mathbf{q} = (q_1, \dots, q_M)$ une loi de probabilité sur F et g une densité de probabilité sur G . X est une variable à états mixtes si, avec probabilité $\gamma \in [0, 1]$, X prend ses valeurs dans F avec la loi de probabilité \mathbf{q} ; sinon, avec probabilité $1 - \gamma$, X est à valeurs dans G avec g pour densité.

Bien que la nature de F puisse être arbitraire, nous considérons pour la suite $F \subset \mathbb{R}^p$ pour faciliter les développements; l'espace d'états de X est $E = \{e_1, \dots, e_M\} \cup G$, $e_i \in \mathbb{R}^p \setminus G$. On munit alors E d'une structure borélienne et de la mesure mixte $m(dx) = \sum_{i=1}^M \delta_{e_i}(dx) + \lambda(dx)$, où δ_e est la mesure de Dirac en e . X est donc une variable mixte de densité par rapport à m :

$$f(x) = \gamma \mathbf{1}_F(x) \sum_{i=1}^M q_i \mathbf{1}_{\{e_i\}}(x) + (1 - \gamma) \mathbf{1}_G(x) g(x), \quad x \in E. \quad (1.20)$$

Toujours avec l'idée de définir ultérieurement un auto-modèle pour des données spatiales, rajoutons l'hypothèse que g est dans une famille exponentielle de dimension l , et que la loi de probabilité \mathbf{q} est positive.

Ecrivant $\sum_{i=1}^M q_i \mathbf{1}_{\{e_i\}}(x) = \exp \sum_{i=1}^M \mathbf{1}_{\{e_i\}}(x) (k_i + \ln q_M)$, avec $k_i = \ln \frac{q_i}{q_M}$, $i = 1, \dots, M$, la densité f de X prend la forme suivante :

$$\begin{aligned} f(x) &= f_\theta(x) \\ &= \exp \left[\mathbf{1}_F(x) \left\{ \sum_{i=1}^M \mathbf{1}_{\{e_i\}}(x) (k_i + \ln q_M) + \ln \gamma \right\} + \mathbf{1}_G(x) \{ \ln[(1 - \gamma)H(\xi)L(x)] + \langle \xi, T(x) \rangle \} \right] \\ &= \exp \left[\mathbf{1}_G(x) \ln \frac{(1 - \gamma)H(\xi)}{\gamma q_M} + \langle \xi, T(x) \mathbf{1}_G(x) \rangle + \sum_{i=1}^{M-1} k_i \mathbf{1}_{\{e_i\}}(x) + \ln(\gamma q_M) + \mathbf{1}_G(x) \ln L(x) \right] \\ &= H'(\theta) L'(x) \exp \langle \theta, B(x) \rangle, \end{aligned} \quad (1.21)$$

où $H'(\theta) = \ln(\gamma q_M)$, et $L'(x) = L(x) \mathbf{1}_G(x)$.

Autrement dit, f_θ est dans une famille exponentielle de dimension $\ell + M$, de paramètre naturel et statistique exhaustive

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{M-1} \\ \theta_M \\ \theta_{M+1} \end{pmatrix} = \begin{pmatrix} k_1 \\ \vdots \\ k_{M-1} \\ \ln \frac{(1-\gamma)H(\xi)}{\gamma q_M} \\ \xi \end{pmatrix}, \quad B(x) = \begin{pmatrix} \mathbf{1}_{\{e_1\}}(x) \\ \vdots \\ \mathbf{1}_{\{e_{M-1}\}}(x) \\ \mathbf{1}_G(x) \\ T(x) \mathbf{1}_G(x) \end{pmatrix}, \quad x \in E. \quad (1.22)$$

Notons que θ_{M+1} et $T(x) \mathbf{1}_G(x)$ sont de dimension l et que par définition $B(e_M) = 0$ et

$L'(e_M) = 1$. De plus, les équations suivantes permettent de remonter aux paramètres originaux.

$$\begin{aligned} k_i &= \theta_i, \quad 1 \leq i < M, \\ \xi &= \theta_{M+1}, \\ q_i &= \frac{e^{k_i}}{e^{k_1} + \dots + e^{k_M}}, \quad 1 \leq i \leq M, \\ \gamma &= \frac{H(\xi)}{H(\xi) + q_M e^{\theta_1}}. \end{aligned}$$

Cette généralisation n'est pas purement théorique mais trouve des applications; prenons l'exemple d'une variable à états mixtes et censurée. Soit Z une variable exponentielle de paramètre λ censurée en $K > 0$ de densité de probabilité $g(z) = \lambda e^{-\lambda z} \mathbf{1}_{(0,K)}(z) + e^{-\lambda K} \delta_K(z)$. On définit alors la variable à états mixtes X qui vaut 0 avec probabilité α et qui a la distribution de Z sinon, avec probabilité $1 - \alpha$. Dans ce cas, X peut être vue comme une variable à états mixtes générale sur l'espace d'états $E = \{0, K\} \cup (0, K)$ et avec

$$\begin{aligned} \gamma &= \alpha + (1 - \alpha)e^{-\lambda K}, \\ \mathbf{q} &= \frac{1}{\alpha + (1 - \alpha)e^{-\lambda K}} (\alpha, (1 - \alpha)e^{-\lambda K}), \\ g\lambda(x) &= \frac{\lambda}{1 - e^{-\lambda K}} e^{-\lambda x}, \quad x \in (0, K). \end{aligned}$$

D'après ce qui précède, la densité de X est dans une famille exponentielle de dimension 3 de paramètre naturel et de statistique exhaustive donnés par

$$\theta = \begin{pmatrix} \ln \frac{\alpha}{1 - \alpha} + \lambda K \\ \ln \lambda + \lambda K \\ \lambda \end{pmatrix}, \quad B(x) = \begin{pmatrix} \mathbf{1}_{\{0\}}(x) \\ \mathbf{1}_{(0,K)}(x) \\ -x \mathbf{1}_{(0,K)}(x) \end{pmatrix}, \quad x \in E.$$

1.4.2 Auto-modèles à états mixtes

Revenons au cas de données spatiales caractérisées par une structure mixte; tout est en place pour construire proprement un auto-modèle pour ces données. Il suffit d'appliquer le théorème 6 sur les auto-modèles multi-paramètres au cas de distributions conditionnelles dans une famille exponentielle de type (1.21), la configuration de référence étant $\tau = (e_M, \dots, e_M)$, $e_M \in F$.

Nous nous intéressons alors à quelques caractéristiques de ces modèles. Tout d'abord, concernant les caractéristiques de compétition ou de coopération spatiale entre sites voisins, il appartient de redéfinir la coopération spatiale; dans le cas standard, on dit qu'il y a coopération (resp. compétition) spatiale si en chaque site i , l'espérance conditionnelle $E[X_i | x^i]$ est non décroissante (resp. non croissante) en chaque valeur voisine x_j , $j \in \partial i$ et croissante (resp. décroissante) en au moins une. Nous adaptons cette définition au cas où l'espace d'état est $E^S = (F \cup G)^S$, en remplaçant l'espérance conditionnelle par la fonction $R(x^i) = E[X_i \mathbf{1}_G(X_i) | x^i]$. Si $E = \{0\} \cup (0, \infty)$, on montre que les définitions classique et généralisée coïncident.

Nous détaillons ensuite dans [Har10] l'exemple des auto-modèles dont les lois conditionnelles sont exponentielles mixtes. La simplicité de ces lois permet d'investiguer les propriétés

de ce type de modèle, et c'est une distribution utile dans les applications ; la proposition 1 de [Har10] donne les conditions assurant l'existence de tels modèles. Dans le cas d'un auto-modèle à lois conditionnelles exponentielles, on sait que les conditions d'intégrabilité sur l'énergie impliquent une compétition spatiale entre sites voisins, ce qui est souvent contraire à la modélisation souhaitée. Il est facile de voir que dans le cas des lois exponentielles mixtes, on a encore cet inconvénient. Nous proposons deux alternatives pour résoudre cette limitation, en tronquant ou en censurant les observations en une valeur K donnée. Le modèle tronqué est défini pour l'espace d'état $E = \{0\} \cup (0, K]$ et utilise la densité $g_\lambda(x) = H(\lambda)e^{-\lambda x} \mathbf{1}_{(0, K]}(x)$, avec $H(\lambda) = \frac{\lambda}{1-e^{-\lambda K}}$ sur $(0, K]$, tandis que le modèle censuré avec $E = \{0, K\} \cup (0, K)$ utilise la densité $g(z) = \lambda e^{-\lambda z} \mathbf{1}_{(0, K)}(z) + e^{-\lambda K} \delta_K(z)$. Dans chaque alternative, nous donnons les conditions d'existence des auto-modèles mixtes associés, ainsi que les conditions assurant un comportement de coopération ou de compétition spatiale ([Har10] propositions 2 et 3).

Application ([Har6])

Ce travail réalisé en collaboration avec J.-F. Yao et des chercheurs de l'IRISA a consisté à utiliser les auto-modèles multi-paramètres pour la modélisation de données mixtes issues de mesures de mouvement dans une séquence vidéo et évaluer leurs performances.

Plus précisément, si $p = (x, y)$ est un point de l'image, notons $I(q, t)$ l'intensité du pixel q au temps t , $\nabla I(q, t)$ le gradient spatial de cette intensité et $v_n(q, t) = \frac{I(q, t) - I(q, t+1)}{\|\nabla I(q, t)\|}$ les résidus de flot. On considère alors la mesure de mouvement résiduel définie comme une moyenne pondérée locale des magnitudes des résidus de flot :

$$v_{res}(p, t) = \frac{\sum_{q \in \mathcal{F}(p)} \|\nabla I(q, t)\|^2 \cdot |v_n(q, t)|}{\max(\eta^2, \sum_{q \in \mathcal{F}(p)} \|\nabla I(q, t)\|^2)} \quad (1.23)$$

où $\mathcal{F}(p)$ est une fenêtre locale centrée sur le pixel p (typiquement une fenêtre 3×3), et η^2 une constante liée au bruit. Des travaux antérieurs ([21], [35]) ont montré la bonne aptitude de cette mesure (1.23) à la détection ou la reconnaissance de mouvement.

Dans [Har6], on considère des séquences vidéos "naturelles" comme des vidéos de feuillages, d'herbe, rivière, mer... Les histogrammes des $\{v_{res}\}$ sur ces séquences montrent clairement des états de type mixte avec un pic en zéro. Par exemple, la figure 1.3 ci-dessous représente (a) une image de feuillages tirée d'une séquence, (b) la texture de mouvement associée $\{v_{res}\}$ et (c) l'histogramme de celle-ci. On ne s'intéresse pas ici à la modélisation de l'intensité dans le temps mais aux mesures mêmes de mouvement, d'où le terme de "texture" de mouvement.

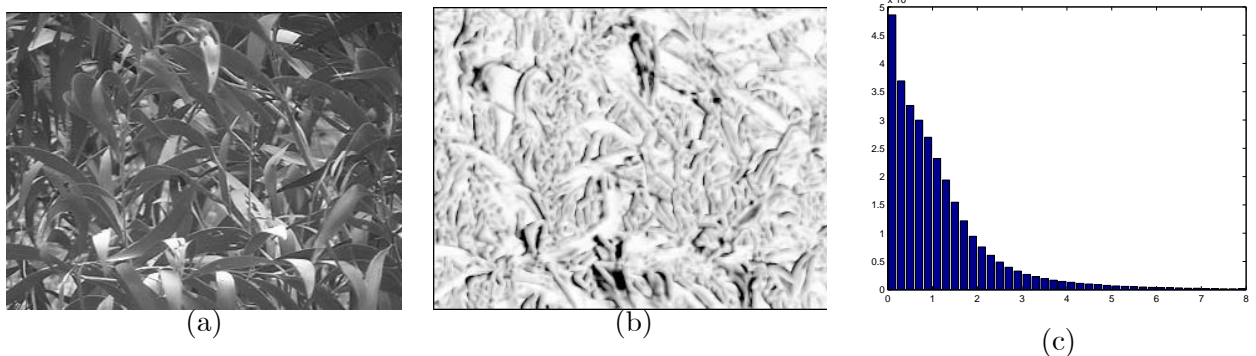


FIG 1.3 - Vidéo de feuillages : (a) une image de la séquence, (b) une texture de mouvement $\{v_{res}\}$, (c) histogramme de cette texture.

On propose d'utiliser la distribution mixte positive gaussienne (1.19) de l'exemple 2, ce choix résultant de l'observation des différents histogrammes.

Nous donnons des conditions sur les paramètres pour que le modèle soit bien défini puis précisons davantage ce modèle pour nous situer dans le cadre de l'application envisagée sur les textures de mouvement. Nous faisons alors les hypothèses suivantes : lattice régulier rectangulaire avec le système de voisinage des 4 plus proches voisins, invariance par translation (mais on autorise l'anisotropie entre les directions verticales et horizontales), coopération spatiale. On aboutit au modèle suivant, décrit par son énergie :

$$Q(x) = \sum_{i \in S} \{a\delta^*(x_i) - bx_i^2\} + c_1 \sum_{\langle i,j \rangle_H} \delta^*(x_i)\delta^*(x_j) + c_2 \sum_{\langle i,j \rangle_V} \delta^*(x_i)\delta^*(x_j).$$

Ce modèle comporte 4 paramètres $\phi = (a, b, c_1, c_2)$ et est bien défini si $b > 0$. Il est isotropique si $c_1 = c_2$. Une étude par simulations est menée pour comprendre le rôle de chacun des paramètres sur le nombre de valeurs nulles, la granularité de la texture, ses orientations. L'estimation de ces paramètres par pseudo-vraisemblance conditionnelle a de bonnes performances.

Revenant aux données réelles de textures de mouvement, on évalue ensuite la capacité du modèle à refléter des caractéristiques fondamentales de textures, comme la stationnarité spatiale, l'isotropie ou l'anisotropie. Ce premier modèle simple s'est montré très satisfaisant et prometteur et apporte une réelle contribution à la compréhension de ce type de données.

Ce travail a permis de mettre en évidence l'intérêt de ces modèles pour la modélisation de données spatiales de nature mixte. Les propriétés statistiques sont examinées dans un cadre spatio-temporel.

1.4.3 Dynamique temporelle d'auto-modèles à états mixtes ([Har13])

Toujours dans l'étude de la modélisation de données spatiales à états mixtes, rajoutons la dimension temporelle. Nous proposons une dynamique spatio-temporelle pour des champs markoviens à espace d'états mixte. Pour cela, on utilise le modèle CMCM de chaîne de Markov de champs de Markov que l'on l'adapte au cadre des auto-modèles à états mixtes.

Cette nouvelle perspective engendre de nouvelles études théoriques liées à l'asymptotique dans le temps ; après avoir posé le modèle général, nous étudions une dynamique auto-exponentielle mixte puis une dynamique auto-normale mixte ; pour chaque modèle, nous regardons ses conditions d'admissibilité ainsi que ses propriétés statistiques, notamment l'ergodicité.

Nous présentons une application sur données réelles pour la reconnaissance de textures de mouvement, et de tracking (suivi automatique d'un déplacement). Une première modélisation avec voisinage spatial temporel (au temps $t-1$) mais pas de voisins instantanés est très prometteuse. Ce travail est réalisé en collaboration avec T. Crivelli, de l'Université de Buenos Aires.

On considère l'espace d'états mixte $E = \{0\} \cup \mathbb{R}^*$ avec les quatre conditions suivantes ; on note $X(t-1) = x$ pour la configuration au temps $t-1$ et $X(t) = y$ pour celle au temps t .

[B1] $X = \{X(t), t \in \mathbb{N}^*\}$ est une chaîne de Markov homogène sur E^S .

Conditionnellement au passé et presque sûrement en x , la transition admet une densité positive

$$P(x, y) = Z^{-1}(x) \exp Q(y | x) . \tag{1.24}$$

[B2] Hypothèse de “pairwise only dependence” :

$$Q(y | x) = \sum_{i \in S} G_i(y_i | x) + \sum_{\{i,j\}} G_{ij}(y_i, y_j | x),$$

avec $G_i(0 | x) = G_{ij}(0, y_j | x) = G_{ij}(y_i, 0 | x) = 0$ pour tout $i, j \in S$.

[B3] Pour tout i , conditionnellement à $(X^i(t) = y^i, X(t-1) = x)$, la distribution de $X_i(t)$ est dans une famille exponentielle à états mixtes (1.17),

$$f_i(y_i | y^i, x) = H'_i(\theta_i(y^i, x)) L'_i(y_i) \exp\langle \theta_i(y^i, x), B_i(y_i) \rangle \quad (1.25)$$

avec $\theta_i(y^i, x) \in \mathbb{R}^d$, $B_i(y_i) \in \mathbb{R}^d$ et les conditions d'identifiabilité $B_i(0) = 0$ et $L'_i(0) = 1$ pour tout $i \in S$.

[B4] Pour tout $i \in S$, $\text{Span} \{B_i(y_i), y_i \in E\} = \mathbb{R}^d$.

Proposition 7 ([Har13] Proposition 1) *Supposons que les conditions [B1] à [B4] soient vérifiées ; alors, conditionnellement à $X(t-1) = x$, il existe une famille $\{\alpha_i(x) : i \in S\}$ de vecteurs de dimension d et une famille de matrices $\{\beta_{ij}(x) : i, j \in S, i \neq j\}$ de taille $d \times d$ vérifiant $\beta_{ji}(x) = \beta_{ij}^T(x)$ telles que*

$$\theta_i(y^i, x) = \alpha_i(x) + \sum_{j:j \neq i} \beta_{ij}(x) B_j(y_j), \quad i \in S. \quad (1.26)$$

Et les potentiels s'écrivent :

$$\begin{aligned} G_i(y_i | x) &= \langle \alpha_i(x), B_i(y_i) \rangle + C_i(y_i) \\ G_{ij}(y_i, y_j | x) &= B_i(y_i)^T \beta_{ij}(x) B_j(y_j). \end{aligned}$$

Le résultat ci-dessus donne un cadre très général pour une grande variété de modèles possibles ; les fonctions α and β décrivent le choix des dépendances passées et instantanées et leur forme est libre, soumise au seul impératif de l'intégrabilité de $\exp Q(y | x)$.

On examine ensuite le cas d'une dynamique auto-exponentielle mixte et on donne les conditions pour que la chaîne soit bien définie et ergodique ([Har13] Proposition 2). Puis nous nous intéressons au cas gaussien, et à nouveau nous donnons les conditions sous lesquelles le modèle est bien défini et ergodique ([Har13] Proposition 3). Dans les deux cas, l'ergodicité s'obtient comme dans [Har5], en utilisant un critère de Lyapounov.

Dans un but applicatif, nous spécifions davantage le cadre gaussien en choisissant les variances conditionnelles indépendantes des voisins, $\sigma_i^2(y^i, x) = \sigma_i^2$, et en éliminant les interactions entre voisins faisant intervenir des carrés. Alors les moyennes et variances des lois gaussiennes conditionnelles sont respectivement

$$\begin{aligned} m_i(y^i, x) &= \frac{1}{2b_i} \left[c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l + \sum_{j \in \partial i} k_{ij} y_j \right], \\ \text{et } \sigma_i^2(y^i, x) &= \sigma_i^2 = \frac{1}{2b_i}. \end{aligned}$$

Ainsi, la moyenne est une combinaison linéaire des voisins instantanés et du passé, ce qui est souvent une bonne hypothèse de modélisation. Par ailleurs, la probabilité de prendre la valeur zéro est $\gamma_i(y^i, x) = \left[1 + \sigma_i \sqrt{2\pi} \exp \left\{ a_i + \sum_{l \in \partial i^-} \alpha_{li} \delta^*(x_l) + \sum_{j: \{i,j\}} e_{ij} \delta^*(y_j) + \frac{m_i^2(y^i, x)}{2\sigma_i^2} \right\} \right]^{-1}$.

Finalement, l'énergie conditionnelle au passé est donnée par

$$Q(y | x) = \sum_{i \in S} \left((a_i + \sum_{l \in \partial i^-} \alpha_{li} \delta^*(x_l)) \delta^*(y_i) - b_i y_i^2 + (c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l) y_i \right) \quad (1.27)$$

$$+ \sum_{(i,j): \{i,j\}} (e_{ij} \delta^*(y_i) \delta^*(y_j) + k_{ij} y_i y_j) .$$

Ce modèle est bien défini si chaque variance est strictement positive et si l'énergie (conditionnelle) est intégrable ; on montre ([Har13] Proposition 4) que ceci est assuré par la condition suivante : pour tout $i \in S$, $b_i > \frac{1}{2} \sum_{j: j \neq i} |k_{ij}|$. Cette condition très simple est libre des autres paramètres $(a_i, c_i, \alpha_{li}, \varepsilon_{li}, e_{ij})$ du modèle. On trouve dans cette même proposition une condition assurant l'ergodicité.

Application

Nous nous intéressons à nouveau aux textures de mouvement. Nous reprenons les notations de [Har6] mais modifions $v_n(i, t)$ et $v_{res}(i, t)$ respectivement par les vecteurs $v(i, t) = -\frac{I(i, t) - I(i, t-1)}{\|\nabla I(i, t)\|} \frac{\nabla I(i, t)}{\|\nabla I(i, t)\|}$ et

$$\bar{v}(i, t) = \frac{\sum_{j \in \mathcal{F}(i)} v(j, t) \|\nabla I(j, t)\|^2}{\max(\eta^2, \sum_{j \in \mathcal{F}(i)} \|\nabla I(j, t)\|^2)} .$$

Nous cherchons à modéliser le mouvement de l'observation du site i au temps t donné par

$$x_i(t) = \bar{v}(i, t) \times \frac{\nabla I(i, t)^T}{\|\nabla I(i, t)\|} .$$

Nous renvoyons le lecteur à [14] pour plus de détails sur la motivation de la construction de $x_i(t)$; notons cependant que $x_i(t)$ est réel (et plus positif comme dans [Har6]).

Nous nous intéressons encore aux textures de mouvement issues de scènes naturelles ; nous représentons Figure 1.4 deux exemples de séquences traitées ; dans la première colonne à gauche, à titre illustratif, (a) est une image instantanée extraite de la séquence vidéo traitée ; au milieu, (b) représente la texture de mouvement associée $\{x_i(t), i \in S\}$, et la dernière colonne (c) donne les histogrammes de ces textures. La première série provient d'une séquence vidéo de jet de vapeur et celle du bas d'une vidéo d'herbe.

Examinant les histogrammes, nous choisissons un modèle mixte auto-normal. Précisément, $X(t) = \{X_i(t), i \in S\}$ est la texture de mouvement au temps t sur le lattice S des pixels, et chaque $X_i(t)$ est distribué suivant une loi gaussienne à états mixtes, conditionnellement au passé.

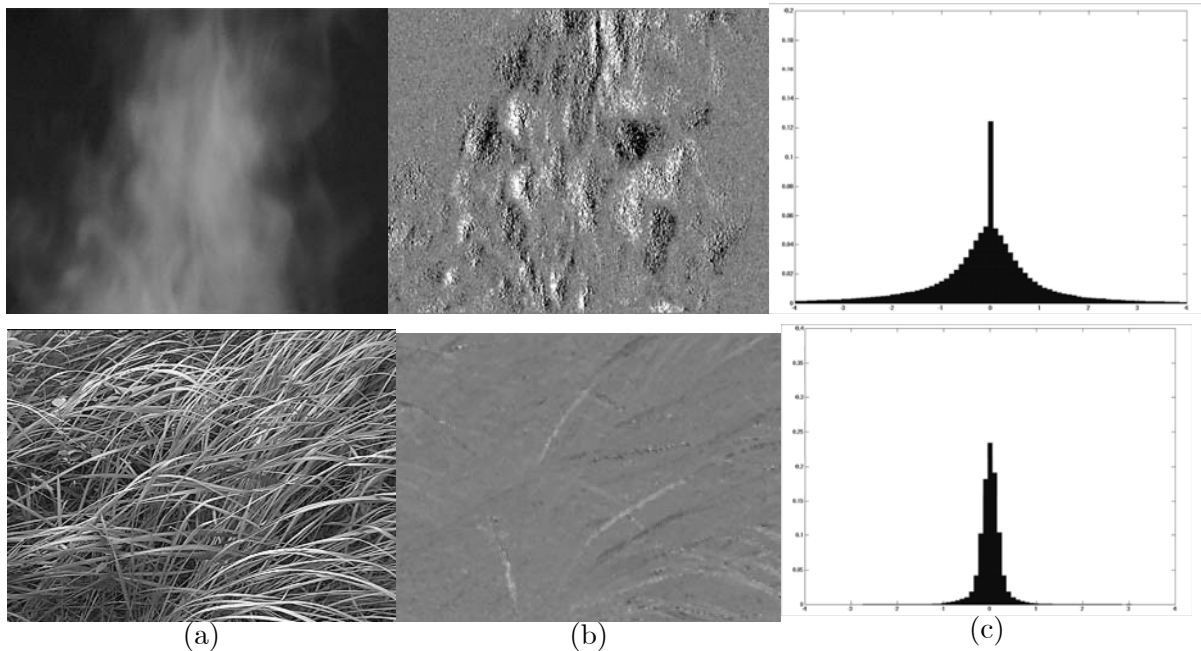


FIG 1.4 - Vidéos de vapeur et d'herbe : (a) images de la séquence, (b) textures de mouvement $\{x_i\}$, (c) histogrammes de ces textures.

Nous choisissons le système de voisinage suivant : on suppose l'indépendance spatiale conditionnelle au temps t , ce qui signifie que nous n'avons pas de voisinage instantané ∂i (mais pour autant, $x_i(t)$ et $x_j(t)$ restent corrélés) ; le voisinage ∂i^- du passé d'un site i au temps $t - 1$ comprend 9 points, ses 8 plus proches voisins plus lui même (Cf Figure 1.5). On suppose le modèle invariant par translation et spatialement symétrique, mais a priori anisotropique dans les quatre directions (verticale, horizontale, diagonale et anti diagonale).

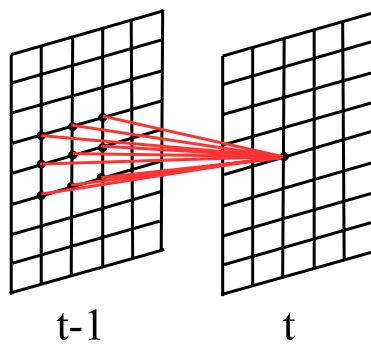


FIG 1.5 - Voisinage d'un site du lattice

Sous ces hypothèses, la loi conditionnelle gaussienne à états mixtes en un site i a pour

moyenne, variance, et probabilité de valeur nulle,

$$m_i(y^i, x) = m_i(x) = \frac{1}{2b} \left[c + \sum_{l \in \partial i^-} \varepsilon_l x_l \right], \quad \sigma_i^2(y^i, x) = \sigma^2 = \frac{1}{2b}, \quad (1.28)$$

$$\gamma_i(y^i, x) = \gamma_i(x) = \left[1 + \sigma \sqrt{2\pi} \exp \left\{ a + \sum_{l \in \partial i^-} \alpha_l \delta^*(x_l) + \frac{m_i^2(x)}{2\sigma^2} \right\} \right]^{-1}, \quad (1.29)$$

et l'énergie conditionnelle $Q(y | x)$ s'écrit :

$$Q(y | x) = \sum_{i \in S} \left(a + \sum_{l \in \partial i^-} \alpha_l \delta^*(x_l) \right) \delta^*(y_i) - b y_i^2 + \left(c + \sum_{l \in \partial i^-} \varepsilon_l x_l \right) y_i. \quad (1.30)$$

D'après ce qui précède pour le modèle général (1.27), cette énergie est admissible si $b > 0$. Le modèle comporte 13 paramètres $\phi = (a, b, c, \{\varepsilon_l\}, \{\alpha_l\}, l \in \partial i^-)$, avec $\{\varepsilon_l\} = \{\varepsilon_i, \varepsilon_{iV}, \varepsilon_{iH}, \varepsilon_{iD}, \varepsilon_{iAD}\}$ correspondant respectivement au site i et ses voisins verticaux, horizontaux, diagonaux et anti diagonaux au temps $t - 1$ (et idem pour les $\{\alpha_l\}$). Ces paramètres sont estimés par la méthode de maximum de pseudo-vraisemblance conditionnelle. Formellement, la condition permettant d'établir l'ergodicité de ce modèle, et donc la consistance est $\exists \eta, \forall i \in S, \sup_x m_i(x) \leq \eta$, ce qui suppose alors de modifier théoriquement la loi gaussienne pour en prendre une version tronquée ou censurée.

Ce modèle a été mis en oeuvre dans [15] pour faire de la reconnaissance de textures de mouvement; dans un premier temps, on "apprend" une texture de mouvement via l'estimation de ses paramètres; ceci permet ensuite de faire de la classification de textures, en utilisant la divergence (conditionnelle) de Kullback-Leibler pour mesurer la distance entre deux densités. L'application considérée dans la thèse de Tomas Crivelli ([16]) est le "tracking" ou suivi d'un objet particulier en mouvement dans la séquence vidéo; la méthode, qui a été comparée avec d'autres sur plusieurs types de séquences vidéo (flamme d'un feu avec panache de fumée, flamme de torchère ou groupe de personnes traversant une rue), s'est révélée plus performante.

1.5 Perspectives

Plusieurs perspectives s'offrent pour compléter ce travail. Sur les auto-modèles, il est intéressant de savoir si l'extension générale des auto-modèles multivariés est compatible avec des extensions telles que la levée de la condition de positivité, ou la "multiway dependence" où les interactions portent au delà des paires.

Dans le cas des auto-modèles mixtes, il reste à étudier les propriétés d'estimation et de validation de modèles : normalité asymptotique de la méthode d'estimation par pseudo vraisemblance conditionnelle, identification par pseudo vraisemblance conditionnelle pénalisée (de type AIC) et test de validation de modèle. Enfin, on peut aussi s'intéresser à l'intégration de covariables explicatives dans une modélisation auto-markovienne.

Rappelons que beaucoup d'auto-modèles présentent des propriétés de compétition spatiale, impliquant un tronquage ou une censure des données, ces procédures autorisant tant la coopération que la compétition spatiale. Ce paramètre de tronquage n'est pas estimé avec les paramètres

du modèle, mais préalablement fixé, ce qui pose la question de son choix par le modélisateur. Kaiser et Cressie ([30]) s'intéressent à ce problème dans le contexte des auto-modèles de Poisson ; il serait intéressant dans le cas des données mixtes de proposer des méthodes pour choisir ce paramètre. On peut également se demander quel est son effet sur la corrélation spatiale.

Bibliographie

- [1] Ailliot P., Thompson C., Thomson P. (2009) Space time modeling of precipitation using a hidden Markov model and censored Gaussian distributions, *J. Roy. Statist. Soc. C* Vol. 58, no3, pp. 405-426.
- [2] Amemiya T. (1985) *Advanced econometrics*, Blackwell.
- [3] Arnold B.C. and Press J. (1989) Compatible conditional distributions. *JASA*, **84**, 405, 152-156.
- [4] Arnold B.C. and Strauss D. (1988) Bivariate distribution with exponential conditionals. *JASA*, **83**, 402, 522-527.
- [5] Arnold B.C., Castillo E. and Sarabia J.M. (1991) Conditionally specified distributions. *L.N.S.* **74**, Springer.
- [6] Arnold B. C., Castillo E., Sarabia J.M. (1999) *Conditional Specification of Statistical Models*. Springer-Verlag, New York.
- [7] Allcroft, D. J. and Glasbey C. A. (2003) A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. *JRSS C* 52 (4), 487–498.
- [8] Bayomog S. (1996) Estimation of a Markov field dynamic. *Prepub. SAMOS*.
- [9] Bayomog S. (1994) *Modélisation et analyse des données spatio-temporelles*. PhD Thesis, Univ. d'Orsay, France.
- [10] Besag J. (1974) Spatial interactions and the statistical analysis of lattice systems. *JRSS B* 148, 1–36.
- [11] Besag J. (1977) Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika* **64**, no. 3, 616–618.
- [12] Cressie N. (1991) *Statistics for spatial Data*. New York, Wiley.
- [13] Cressie N. and Lele S. (1992) New models for Markov random fields. *Adv. Appl. Probab.* **29**, 877-884
- [14] Crivelli T., Cernuschi-Frias B., Bouthemy P., Yao J.F. (2006) Mixed-state Markov random fields for motion texture modelling and segmentation. *In proc. of the IEEE Int. Conf. on Image processing, ICIP'06*, 1857–1860.
- [15] Crivelli T., Bouthemy P., Cernuschi-Frias B., Yao J.F. (2009) Learning mixed-state Markov models for statistical motion texture tracking. *2nd IEEE International Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA09)*. Kyoto, Japan. September 2009.
- [16] Crivelli T. (2010) Mixed state Markov models for image motion analysis. *Ph. D thesis, Université Rennes 1, France*.

- [17] Chadoeuf J., Nandris D., Geiger J.P., Nicole M., Piarrat J.C. (1992) Modélisation spatio temporelle d'une épidémie par un processus de Gibbs : estimation et tests. *Biometrics*, **48**, 1165-1175.
- [18] Comets F. and Janzura M. (1998) A central limit theorem for conditionally centered random fields with an application to Markov random fields. *J. Appl. Probab.* **35**, 608–621.
- [19] Duflo M. (1990) *Méthodes récursives aléatoires*. Masson.
- [20] Duflo M. (1997) *Random iterative models*. Springer.
- [21] Fablet R, Bouthemy P. (2003) Motion recognition using non parametric image motion models estimated from temporal and multiscale cooccurrence statistics. *IEEE Trans. on pattern Analysis and machine intelligence* 25(122), 1619-1624.
- [22] Golub G.H., Van Loan C.F. (1996) *Matrix computations*. The Johns Hopkins University Press.
- [23] Guyon X. (1995) *Random Fields on a Network : Modeling, Statistics, and Applications*. Springer-Verlag, New York.
- [24] Guyon X. and Künsch H.R. (1992) Asymptotic comparison of estimators in the Ising model. *L.N.S.* 74, Springer, 177-198.
- [25] Georgii H.O. (1988) *Gibbs measures and phase transitions*. De Gruyter.
- [26] Geyer C.J. and Thompson E.A. (1992) Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *JRSS,B* **54**, 657-699.
- [27] Hall P. and Heyde C.C. (1980) *Martingale limit theory and its applications*. Acad. Press.
- [28] Huang and Ogata Y. (2002) Generalized pseudo-likelihood estimates for Markov random fields on lattice. *Ann. Inst. Statist. Math.* **54**, (1), 1-18.
- [29] Jensen J.L. and Künsch H.R. (1994) On asymptotic normality of pseudo-likelihood estimate for pairwise interaction processes. *Ann. Inst. Statist. Math.*, **46**, 475-486.
- [30] Kaiser M.S., Cressie N. (1997) Modeling of Poisson variables with positive spatial dependence. *Statistics and probability letters* 35, 423-432.
- [31] Kaiser M.S. and Cressie N. (2000) The construction of multivariate distributions from Markov random fields. *J. Multivariate Anal.* **73**, 199-220.
- [32] Kaiser M.S., Cressie N. and Lee J. (2002) Spatial mixture models based on exponential family conditional distributions. *Statistica Sinica* **12**, 449-474
- [33] Lee J., Kaiser M.S., Cressie N. (2001) Multiway dependence in exponential family conditional distributions. *J. Multi. Anal.* **79**, 171-90
- [34] Odobez J.-M., Bouthemy P. (1995) Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation Vol. 6, No 4, 348-365*.
- [35] Odobez J.-M., Bouthemy P. (1997) Separation of moving regions from background in an image sequence acquired with a mobile camera. *Chapt. 8, pp. 283-311, Video Data Compression for Multimedia computing (Eds. H.H. Li, S. Sun and H. Derin), Kluwer Academic Publisher*.
- [36] Prum B. (1986) *Processus sur un réseau et mesure de Gibbs. Applications*. Masson.
- [37] Senoussi R. (1990) Statistique asymptotique presque-sûre de modèles statistiques convexes. *Ann. Inst. Henri Poincaré* **26**, 19-44.

- [38] Sinai Ya. G. (1982) *Theory of Phase Transitions : Rigorous Results*. Pergamon Press, Oxford.
- [39] Whittle P. (1963) Stochastic processes in several dimensions. *Bull. Inst. Statist. Inst.* **40**, 974-994.
- [40] Younes L. (1988) Estimation and Annealing for Gibbsian fields. *Ann. I.H.P.* **2**, 269-294.
- [41] Zucchini W. and Guttorp P. (1991) A hidden Markov model for space time precipitation. *Water resour. Res.* **27**, 1917-1923.
- [42] MacDonald I. and Zucchini W. (1997) *Hidden Markov and other models for discrete valued time series*. Monographs on statistic and applied probability. Chapman and Hall.

Chapitre 2

Adoption de standards et diffusion de technologies

2.1 Introduction

L'objet de cette étude vient d'un problème présenté par un collègue économiste de l'Université Paris 1 : la modélisation des mécanismes conduisant à l'adoption de certains standards technologiques au détriment d'autres. L'étude de la diffusion d'un processus technologique, de la dynamique de formation d'une opinion ou d'un choix collectif d'individus placés sur un ensemble de sites S constitue un élément important de l'économie spatiale, citons par exemple [1], [2], [5], [19]... Des agents placés sur un réseau choisissent entre plusieurs standards suivant une règle contextuelle. Nous entendons par standard aussi bien un produit, qu'un nouveau service ou une nouvelle technologie. Diverses modélisations ont été proposées : exemple chaîne de Markov (par ex. [9]), automate cellulaire ([7]), champs de Gibbs (par ex. [12], [25]), modèle d'apprentissage (réseaux neuronaux), jeux évolutionnistes (par ex. [20]), algorithmes génétiques... Nous travaillons ici sur la modélisation markovienne spatio-temporelle par potentiels de Gibbs.

Décrivons la situation spatiale : n consommateurs ou agents sont placés aléatoirement sur un lattice $S = \{1, 2, \dots, n\}$, muni d'un graphe de voisinage \mathcal{G} . Pour simplifier, on considère que chaque individu a le choix entre deux standards $E = \{0, 1\}$. La procédure d'adoption de standards est séquentielle et locale : chaque individu fait son choix selon une stratégie indépendante de ses voisins, ou suit une stratégie de préférence locale basée sur l'observation des choix antérieurs de ses voisins. L'observation finale disponible est $X = \{X_i, i \in S\}$.

Un "balayage" de S est un choix séquentiel progressif pour tous les sites de S suivant un ordre donné par une permutation de $[1, 2, \dots, n]$. Une fois le premier balayage de S effectué, la procédure peut être itérée. Une différence importante avec le premier balayage est la suivante ; à partir du deuxième balayage, tous les sites voisins d'un site sont déjà occupés. Le cadre de balayages itératifs est connu, nous le rappelons brièvement.

Réaffectations itératives : la dynamique de l'échantillonneur de Gibbs

Supposons les balayages séquentiels et notons σ_k le $k^{\text{ième}}$ balayage, $x = X(k-1)$ et $y = X(k)$ les configurations avant et après le $k^{\text{ième}}$ balayage. Alors, $X = (X(k))_{k \geq 2}$ est une chaîne de

Markov sur $\{0, 1\}^S$ de transitions données par

$$P_{\sigma_k}(x, y) = \prod_{i=1, n} p_i(y_i | y_1, y_2, \dots, y_{i-1}, x_{i+1}, \dots, x_n).$$

Dans le cas où les balayages diffèrent, X est une chaîne inhomogène et la loi de $X(k)$ est $X(k) \sim \mu P_{\sigma_1} P_{\sigma_2} P_{\sigma_3} \dots P_{\sigma_k}$ où μ est la loi initiale de $X(0)$. Mais si les parcours de visites sont les mêmes à chaque balayage, $\sigma_k \equiv \sigma$ pour tout $k \geq 1$, la chaîne est alors homogène de transition $P = P_\sigma$, et $X(k) \sim \mu P^k$. Le résultat suivant établit l'ergodicité de la chaîne (Cf par exemple [13] ou [17] ou [Har4] Proposition 2).

Proposition 8 *Supposons que $\delta = \inf\{p_i(x_i | x^i), x \in \{0, 1\}^S\} > 0$.*

- (1) *Dans le cas homogène, la chaîne X est ergodique de loi limite π , l'unique loi invariante.*
- (2) *Dans le cas inhomogène, supposons qu'il existe une loi π invariante pour toute permutation σ , ($\forall \sigma, \pi P_\sigma = \pi$). Alors X est ergodique convergeant vers π .*

Une application importante de ce résultat est l'échantillonneur de Gibbs [13]. Considérons une distribution π sur E^S satisfaisant la condition de positivité

$$\forall x \in E^S, \pi(x) > 0,$$

et notons $\pi_i(x_i | x^i)$ ses lois conditionnelles en chaque site $i \in S$. L'algorithme de l'échantillonneur de Gibbs de la loi π consiste à prendre $p_i(x_i | x^i) = \pi_i(x_i | x^i)$; en effet, puisque la condition de positivité sur π implique la stricte positivité des π_i , on a $\delta > 0$. De plus, π étant invariante pour chaque π_i , elle l'est encore pour chaque P_σ . La chaîne X est donc ergodique de loi limite π , et on voit ici que l'ordre dans lequel les balayages sont effectués n'a pas de conséquence. En pratique, il suffit donc d'itérer un nombre suffisant de balayages basés sur le seul calcul des lois conditionnelles π_i pour simuler un processus de loi π .

Exemple : *Echantillonneur d'un champ de Markov*

Considérons le tore bidimensionnel $S = T^2$ muni de la relation aux 4 p.p.v., X étant à valeurs $\{0, 1\}$. Supposons que π soit donnée par

$$\pi(x) = Z^{-1} \exp\left\{ \alpha \sum_{i \in S} x_i + \beta \sum_{\langle i, j \rangle} x_i x_j \right\}. \quad (2.1)$$

Les deux standards sont équiprobables si $\alpha + 2\beta = 0$, et si $\alpha + 2\beta > 0$, 1 est plus probable que 0. La loi conditionnelle au site i est $\pi_i(x_i | x^i) = \frac{\exp x_i (\alpha + \beta \sum_{j \in \partial i} x_j)}{1 + \exp(\alpha + \beta \sum_{j \in \partial i} x_j)}$.

Choix synchrones ou parallélisation massive

Supposons que dans le passage de $X(k-1) = x$ à $X(k) = y$, les n choix soient synchrones. La nouvelle transition s'écrit

$$Q(x, y) = \prod_{i=1, n} p_i(y_i | x^i).$$

Q reste ergodique, mais sa loi limite μ , unique loi stationnaire solution de $\mu Q = \mu$, diffère de celle obtenue pour un balayage séquentiel ([24]). En général, on ne dispose pas d'une forme analytique explicite pour μ .

Si nous itérons les balayages, qu'ils soient séquentiels ou synchrones, et si les choix de standard sont effectués suivant une règle d'affectation locale, ces modèles présentent une corrélation spatiale non nulle. Nous proposons dans [Har4] des tests de non "coordination" spatiale pour discriminer les stratégies de choix réalisés en toute indépendance ou sous l'influence des voisins.

Nous étudions dans [Har8] et [Har11] le cas d'un balayage unique dans le cadre où les stratégies de choix sont locales; nous sommes alors dans un contexte non ergodique et l'étude est de nature empirique; nous nous intéressons à l'effet de la présence initiale d'un standard sur le marché (avant apparition des concurrents) sur la configuration finale; pour caractériser celle-ci, nous examinons la fréquence finale de standards, l'évolution de la corrélation spatiale et proposons des indicateurs d'agrégation spatiale.

2.2 Tests de répartition spatiale ([Har4])

Revenons aux stratégies de choix. Celui-ci s'opère-t-il sous l'influence des voisins ou est-il effectué en toute indépendance? Nous proposons des tests de l'hypothèse nulle (H_0) d'absence de "coordination" spatiale. Lorsque le modèle spatial est paramétrique, la non coordination spatiale correspond à la nullité de certains paramètres et on peut obtenir des tests dérivant d'estimateurs standards et de leurs propriétés asymptotiques. Nous nous situons dans un cadre non paramétrique, et les statistiques de tests que nous proposons sont basées sur l'estimation empirique de la corrélation aux plus proches voisins et sur l'étude de sa loi asymptotique sous (H_0), l'asymptotique portant sur la taille de la fenêtre d'observation. Nous comparons empiriquement ces tests en examinant leurs puissances sous deux alternatives spécifiées pour X .

2.2.1 Modèles sans coordination spatiale

En l'absence de coordination spatiale, nous définissons le modèle pour X de la façon suivante.

1. On choisit une distribution π pour une variable $Y = (Y_i, i = 1, \dots, n) \in \{0, 1\}^n$. Cette distribution π dépend de $(2^n - 1)$ paramètres, les probabilités $\pi_A^Y = P(Y_A = 1_A, Y^A = 0^A)$, où $A \in \mathcal{P}_n$, l'ensemble des parties de S , avec la contrainte $\sum_A \pi_A^Y = 1$.

2. On choisit une permutation σ de S qui affecte y_i au site $\sigma(i)$, selon une loi μ sur Ξ_n , l'ensemble des permutations de S . On considère alors $X = (X_i, i = 1, \dots, n)$ avec $X_i = Y_{\sigma^{-1}(i)}$, et la distribution de X dépend donc de π et de μ .

Posons $S_Y = \sum_{i=1}^n Y_i = S_X$, $p = \frac{1}{n}E[S_Y]$ et définissons les probabilités suivantes :

$$\pi(k) = P(S_Y = k) = \sum_{A \in \mathcal{P}_n, |A|=k} \pi_A^Y, \quad p_k = \frac{\pi(k)}{\binom{n}{k}}.$$

Le résultat suivant donne la distribution des variables X_i :

Proposition 9 ([Har4] Proposition 1) *Supposons que σ est uniforme sur Ξ_n . Alors :*

(1) *Pour toute partie B de S , $\pi_B^X = p_{|B|}$. En particulier, pour toute permutation δ de Ξ_n , les variables $(X_{\delta(1)}, X_{\delta(2)}, \dots, X_{\delta(n)})$ ont la même loi.*

(2) *Les variables X_i sont des variables de Bernoulli $B(p)$ de covariance $Cov(X_i, X_j) = \frac{Var S_Y}{n(n-1)} + \frac{E[S_Y]^2}{n^2(n-1)} - \frac{E[S_Y]}{n(n-1)}$.*

(3) *Conditionnellement à $S_Y = b$, X est une variable uniforme concentrée sur les parties de S de taille b et vérifie $E[X_i] = \frac{b}{n}$ et $Cov(X_i, X_j) = -\frac{b(1-\frac{b}{n})}{(n-1)}$.*

(3) signifie que conditionnellement à $S_Y = b$ (b agents ont choisi le standard 1), les variables X_i sont de même loi, qui est indépendante de la permutation σ et de la distribution initiale des Y_i . De plus, la corrélation des X_i devient petite dès que n est grand.

2.2.2 Modèles avec coordination spatiale

Si maintenant le choix d'un standard en un site dépend des choix passés dans le voisinage de ce site, il va apparaître une dépendance spatiale, positive s'il y a coopération, négative s'il y a compétition.

Le modèle de choix séquentiel en une étape

Une fois le placement σ des individus fixé, le modèle est le suivant ;

1. Au site $\sigma(1)$, on choisit $X_{\sigma(1)} = Y_1 \sim \mathcal{B}(p_1)$.

Supposons réalisés les choix y_1, y_2, \dots, y_{i-1} , $i \leq n$.

2. Au site $\sigma(i)$, on choisit X suivant une règle contextuelle $X_{\sigma(i)} = Y_i \sim \mathcal{B}(p_i(y^{i-}, \sigma))$ où $p_i(y^{i-}, \sigma)$ s'exprime en fonction des choix antérieurs des voisins de $\sigma(i)$.

Prenons deux exemples simples de telles règles de choix.

Modèle d'Ising progressif :

Considérons le tore d -dimensionnel $S = \{1, 2, \dots, N\}^d$, avec $n = N^d$ sites, muni de la relation aux $2d$ -p.p.v. (avec les conditions naturelles de raccordements aux bords). Notons $n_i(y) = \sum y_j$, la somme étant étendue aux voisins du site $\sigma(i)$ qui ont déjà choisi, c'est-à-dire aux $j < i$ tels que $\langle \sigma(i), \sigma(j) \rangle$. La loi de Y_i est donnée, pour α, β deux paramètres réels, par

$$p_i(y_i | y^{i-}) = \frac{\exp y_i \{\alpha + \beta n_i(y)\}}{1 + \exp \{\alpha + \beta n_i(y)\}}, \quad y_i \in \{0, 1\}.$$

Le paramètre β contrôle la coordination spatiale : $\beta > 0$ favorise la coopération alors que $\beta < 0$ traduit une compétition.

Modèle d'affectation majoritaire :

Dans le modèle précédent, on pose d tel que $\alpha + d\beta = 0$ et on fait tendre β vers l'infini. On choisit avec certitude $y_i = 1$ (resp. $y_i = 0$) si $n_i(y) > d$ (resp. $n_i(y) < d$), et on choisit $\{0, 1\}$ avec les mêmes probabilités sinon. L'écriture analytique de la loi jointe de Y est toujours possible,

$$\pi(y) = \frac{1}{n!} \sum_{\sigma \in \Xi_n} \prod_{i=1, n} p_i(y_i | y^{i-}, \sigma)$$

mais elle est tributaire de la complexité géométrique due aux permutations σ . Ceci explique l'impossibilité d'exprimer analytiquement les caractéristiques de π , telle la corrélation aux p.p.v.. Par contre, celles-ci peuvent s'obtenir facilement par simulation ([Har8-11]).

2.2.3 Tests de non-coordination spatiale : le cas d'un lattice régulier aux ν plus proches voisins

Considérons $S = \{1, 2, \dots, n\}$ un lattice régulier avec le système de voisinage des ν plus proches voisins. On supposera dans tout ce paragraphe que l'on est en situation de non-coordination spatiale comme décrit précédemment et que les variables Y_i sont indépendantes. Notons :

(B) le modèle de Bernoulli *i.i.d.* de probabilité π . Puisque les Y_i sont *i.i.d.* de loi $\mathcal{B}(\pi)$, les variables X_i sont également *i.i.d.* $\mathcal{B}(\pi)$,

(HG) le modèle conditionnel à $b = \sum_i X_i$.

Une statistique naturelle pour tester l'hypothèse (H_0) de non-coordination spatiale est construite à partir de l'estimateur C_n de la covariance aux p.p.v., qui s'exprime à partir de la variable M_n :

$$M_n = \frac{2}{n\nu} \sum_{\langle i,j \rangle} X_i X_j.$$

Pour les X_i , il y a non coordination spatiale si $X_i = Y_{\sigma^{-1}(i)}$ est "indépendant" de $X_j = Y_{\sigma^{-1}(j)}$; dans le cas (B), cela correspond à la non corrélation spatiale i.e. $Cov(X_i, X_j) = 0$; mais dans le cas (HG), cette corrélation n'est pas nulle. Résumons l'expression de la statistique dans les trois situations que nous allons étudier.

• (B) et π connue

$$C_n = M_n - \pi^2$$

•• (B) et π inconnue

Si π est inconnue, π^2 est estimée sans biais par $\frac{n}{n-1}(\bar{X}_n^2 - \frac{1}{n}\bar{X}_n)$. La statistique remplaçant C_n est alors

$$\Gamma_n = M_n - \frac{n}{n-1}(\bar{X}_n^2 - \frac{1}{n}\bar{X}_n)$$

••• (HG)

$$C_n = M_n - \frac{b(b-1)}{n(n-1)}.$$

Le résultat ci-dessous donne les lois asymptotiques pour les trois cas.

Proposition 10 *Test d'absence de coordination spatiale* ([Har4] Propositions 3,4,5).

(1) Dans le cadre du modèle (B) avec π connue,

$$\sqrt{\frac{\nu n}{2}} C_n \longrightarrow \mathcal{N}(0, \pi^2(1-\pi)[1+(2\nu-1)\pi]).$$

(2) Dans le cadre du modèle (B) avec π inconnue,

$$\sqrt{\frac{\nu n}{2}} \Gamma_n \longrightarrow \mathcal{N}(0, \pi^2(1-\pi)^2).$$

(3) Dans le cadre du modèle conditionnel (HG). Supposons (S) : $\frac{b}{n} \xrightarrow[n \rightarrow \infty]{} \pi$, alors

$$\sqrt{\frac{\nu n}{2}} \left(M_n - \frac{\bar{X}_n(n\bar{X}_n - 1)}{(n-1)} \right) \longrightarrow \mathcal{N}(0, \pi^2(1-\pi)^2).$$

Dans le résultat (3), l'hypothèse (S) revient à dire qu'il y a stabilisation de la proportion des standards.

2.2.4 Autres tests et étude empirique des puissances

D'autres statistiques de test de non-coordination spatiale sont envisageables, portant par exemple sur une autre évaluation de la corrélation spatiale. Entre deux tests, on privilégiera celui de plus grande puissance. Nous envisageons deux types d'alternatives pour X , et calculons empiriquement par simulation la puissance des tests proposés au niveau 5%.

Alternative 1. Test $(H_0) = (B)$ contre $(H_1) = (H_\beta)$: X est un modèle d'Ising avec $\alpha + 2\beta = 0$.

Sous l'hypothèse alternative, les lois conditionnelles s'écrivent

$$p_i(x_i | \beta, n_i(x)) = \frac{\exp x_i \{-2\beta + \beta n_i(x)\}}{1 + \exp\{-2\beta + \beta n_i(x)\}}$$

L'hypothèse nulle correspond à $\beta = 0$ dans (H_β) . Notons que la corrélation n'est pas explicite.

Nous avons simulé des lattices pour des valeurs de β variant de 0 à 1.3; lors de chaque simulation, nous calculons d'une part la corrélation empirique à distance 1 et d'autre part les statistiques de test, ceci dans les deux contextes où π est connue ou inconnue.

En ce qui concerne la corrélation, à β fixé, on observe que la corrélation augmente avec la taille du lattice.

La figure 2.1 donne les courbes de puissance empiriques sur la base de 500 simulations d'un lattice de taille 10×10 au niveau 5%, pour les 5 statistiques de test suivantes :

$$\begin{aligned} \text{Modèle } (B), \pi \text{ connue : } & C4 = \frac{\sqrt{2n}(M_n - \pi^2)}{\sqrt{\pi^2(1-\pi)(1+7\pi)}} \text{ et } C8 = \frac{\sqrt{4n}(M_n^* - \pi^2)}{\sqrt{\pi^2(1-\pi)(1+15\pi)}} \\ \text{Modèle } (B), \pi \text{ inconnue : } & G4 = \frac{\sqrt{2n}(M_n - \widehat{\pi}^2)}{\widehat{\pi}(1-\widehat{\pi})} \text{ et } G8 = \frac{\sqrt{4n}(M_n^* - \widehat{\pi}^2)}{\widehat{\pi}(1-\widehat{\pi})} \\ \text{Modèle } (B), \pi \text{ inconnue : } & G2 = \frac{\sqrt{n}(M_n^{**} - \pi^2)}{\widehat{\pi}(1-\widehat{\pi})}, \text{ avec } M_n^{**} = \frac{1}{n} \sum_{s,s',t,|s-s'|=1} X_{s,t} X_{s',t}. \end{aligned}$$

On observe que les tests fondés sur le calcul de la corrélation aux 4 p.p.v. sont meilleurs que ceux utilisant 2 ou 8 voisins, et cela dans les deux situations π connue ou non : les tests basés sur les 2 p.p.v. n'utilisent pas assez d'observations; dans le cas du système de voisinage aux 8 p.p.v., la corrélation diagonale étant plus petite que celle à distance 1, cela introduit un biais négatif sur $G8$ ou $C8$.

On observe par ailleurs que les tests basés sur les statistiques lorsque π est inconnue sont plus puissants que ceux utilisant la vraie valeur de π lorsque π est connue.

Alternative 2. $(H_0) : (HG)$ contre $(H_1) : X$ est un modèle d'Ising avec $\alpha + 2\beta = 0$ et conditionnel à $S_X = \frac{n}{2}$.

Nous itérons le même travail que dans l'alternative 1, les simulations étant ici réalisées en utilisant la dynamique de Metropolis d'échange de spins conditionnelle à $S_X = \frac{n}{2}$ ([15], §6.2.1).

Comme précédemment, nous constatons que la corrélation à distance 1 augmente avec la taille du lattice. Les courbes de puissance sont tracées pour les deux statistiques :

$$\text{Modèle } (HG) : C4 = \frac{\sqrt{2n}(M_n - \frac{n-2}{4(n-1)})}{\sqrt{A_4 + B_4}} \text{ et } C8 = \frac{\sqrt{4n}(M_n^* - \frac{n-2}{4(n-1)})}{\sqrt{A_8 + B_8}}$$

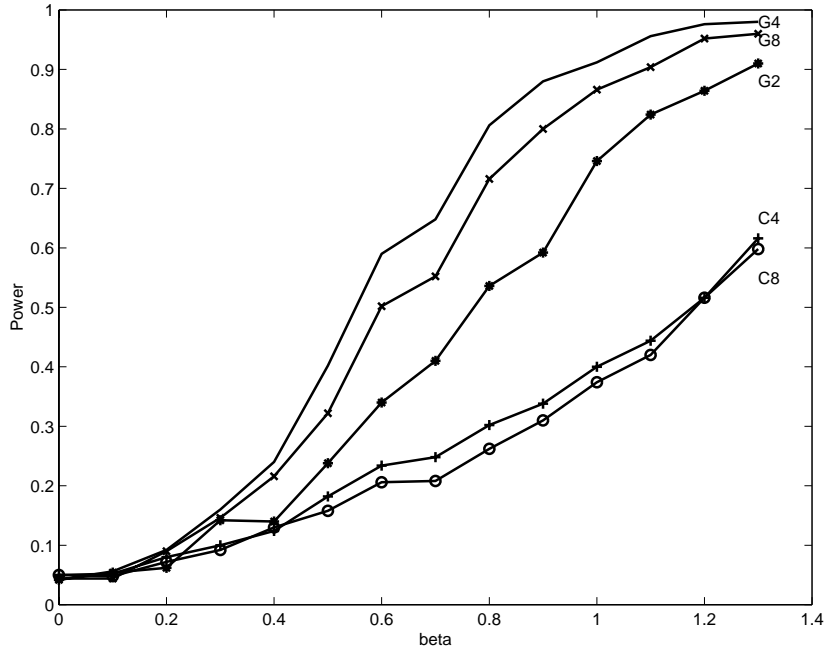


FIG. 2.1 – Modèle (B). Puissances des statistiques aux 2 ppv ($G2$), 4 ppv ($C4$, $G4$), et 8 ppv ($C8$, $G8$).

avec A_4 , B_4 , (resp. A_8 , B_8) les constantes données dans la proposition 10 pour $\nu = 4$ (resp. $\nu = 8$) et $b = \frac{\nu}{2}$. On constate également, comme dans le cadre de l'alternative 1, que les tests basés sur les 4 p.p.v. sont plus puissants que ceux basés sur les 8 p.p.v., et que la puissance augmente avec la taille du lattice.

Cas de plus de deux standards

Supposons que nous ayons $r + 1 \geq 2$ standards $\{a_0, a_1, \dots, a_r\}$ en compétition. En un site i , la réalisation sera codée par la variable multidimensionnelle $X_i \in \{0, 1\}^r$ de \mathbf{R}^r

$$X_i(l) = \begin{cases} 1 & \text{si } X_i = a_l \\ 0 & \text{sinon} \end{cases}, l = 1, \dots, r.$$

On utilise des techniques similaires à celles développées antérieurement ; par exemple pour un modèle de Bernoulli, on obtient un test de non coordination spatiale.

2.3 Répartition spatiale avec présence initiale d'un standard ([Har8, Har11])

Toujours dans le cadre d'un choix unique entre deux standards, on cherche à décrire la configuration finale dans l'hypothèse où la procédure de choix débute après affectation initiale d'une proportion τ de l'un des standards en compétition ($0 \leq \tau < 1$), les sites préalablement affectés étant choisis régulièrement ou aléatoirement sur S . On parcourt ensuite aléatoirement

les $(1-\tau)n$ sites restant et chaque nouvel agent fait son choix de standard de façon locale suivant une règle définie.

Nous considérons le système de voisinage des 4 plus proches voisins avec les trois règles d'affectation suivantes :

1. *Le choix majoritaire fort* : l'agent choisit le standard qui est majoritaire parmi ses quatre voisins. En cas d'égalité, ou lorsque l'ensemble des sites voisins est non occupé, il choisit A (resp. B) avec la probabilité π (resp. $1-\pi$).
2. *Le choix majoritaire faible* : si le nombre de sites voisins occupés est inférieur ou égal à deux, l'agent choisit A (resp. B) avec la probabilité π (resp. $1-\pi$). A partir de trois voisins occupés, l'agent suit la règle de choix majoritaire.
3. *Le choix probabiliste de type Ising* : si les 4 sites voisins de i sont non occupés, l'agent choisit A (resp. B) avec la probabilité π (resp. $1-\pi$). Dans le cas contraire, notant $y_{\partial i}$ la configuration des quatre plus proches voisins de i , il choisit A avec la probabilité

$$\pi_i(A | y_{\partial i}) = \frac{\exp\{\beta \sum_{j \in \partial i} y_j\}}{\exp\{\beta \sum_{j \in \partial i} y_j\} + \exp\{-\beta \sum_{j \in \partial i} y_j\}} = 1 - \pi_i(B | y_{\partial i})$$

Pour une loi de la géométrie de remplissage donnée, la répétition de simulations permet d'obtenir une loi empirique pour le pourcentage final de standards A et les corrélations spatiales. Nous nous intéressons ensuite à décrire la configuration finale via des mesures de l'agrégation spatiale; à cet effet, nous donnons un indicateur d'agrégation et trois paramètres mesurant la connexité. Ces caractéristiques dégagent à la fois l'influence de la règle d'affectation et celle du taux de contamination initiale sur la répartition spatiale finale des deux standards. Pour chaque indicateur, nous mesurons ses propriétés liées à l'agrégation spatiale en comparant sa valeur avec celle obtenue pour un champ aléatoire "au hasard" comportant le même nombre final de standards A, mais répartis uniformément. Les indicateurs proposés peuvent être utiles pour construire des tests (de permutation) d'indépendance spatiale.

Dans cette étude, nous avons choisi $\pi = 0.5$ dans les règles majoritaires et considéré trois valeurs du paramètre β de coordination spatiale pour la règle de type Ising : $\beta = 0.5$ qui correspond à une faible coordination spatiale, puis $\beta = 1$ qui correspond à un début de formation d'amas, et enfin $\beta = 3$ où la formation d'agrégats (clusters) est plus prononcée. Pour chaque valeur de τ variant de 2% à 99%, et pour chacun des modèles d'affectation, nous réalisons 400 simulations sur le tore carré de taille 64×64 .

La fréquence finale du standard A

La figure 2.2 représente l'évolution de la fréquence finale de standards A en fonction de τ pour les cinq règles d'affectation. On voit sans surprise que la croissance de la courbe est d'autant plus importante que la règle force la coopération locale. Mais l'intérêt de ces courbes est de montrer que l'effet différentiel de la contamination initiale est plus important pour des petites valeurs du taux. En effet, par exemple, pour la règle d'affectation majoritaire forte avec un taux initial de 20%, le standard A sera présent à 90% dans la configuration finale; en comparaison, si les 80% des sites initialement non affectés s'étaient répartis équitablement entre A et B, la proportion finale de A aurait été de $\frac{80}{2}\% + 20\% = 60\%$, soit une différence entre les deux situations de 30%. Mais dans le cas d'un taux initial de 10%, la différence serait de 23%.

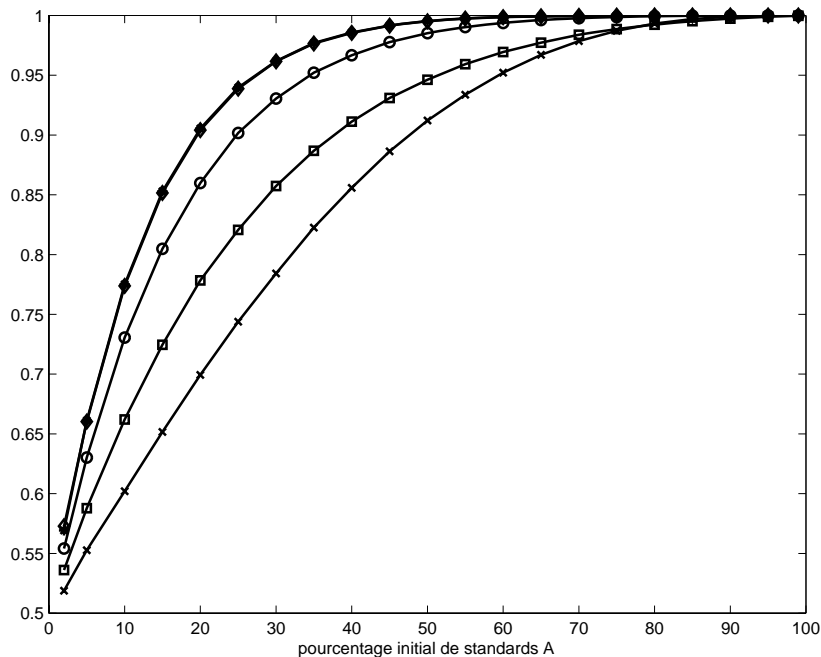


FIG. 2.2 – Fréquence finale de standards A suivant la règle d’affectation : x : faible ; □ : Ising $\beta = 0.5$; o : Ising $\beta = 1$; ◇ : Ising $\beta = 3$ et forte (confondues).

La corrélation spatiale

Nous calculons la corrélation à distance 1, à distance 2 et à distance $\sqrt{2}$, notées respectivement ρ_1 , $\rho_{\sqrt{2}}$, et ρ_2 , et enfin celle aux 8 plus proches voisins ρ_{8v} , et ce pour chaque règle d’affectation et nous examinons leur évolution en fonction de τ . Quelle que soit la règle, et pour chaque corrélation, celle-ci est positive et décroît vers zéro lorsque τ augmente. On constate une hiérarchie entre les différentes règles ; pour un taux initial de standards inférieur à 50%, la corrélation spatiale est plus importante et décroît plus vite vers zéro pour les règles qui forcent davantage le choix du standard A. A titre d’exemple, la figure 2.3 représente l’évolution de la corrélation à distance 1 en fonction de τ .

Une autre façon d’appréhender l’effet spatial d’une contamination initiale et d’une dynamique de choix locaux sur la configuration finale pour une règle donnée, est de comparer cette configuration (C) avec celle qui résulterait d’une répartition aléatoire uniforme (C_0) pour un même nombre final de standards A. Le champ (C_0) est d’allure et de caractéristiques distinctes de (C). Par exemple, nous représentons figure 2.5 deux champs (C) et (C_0) obtenus pour la règle d’affectation Ising avec $\beta = 0.5$, et le même taux initial $\tau = 0.05$; ces deux champs ont la même proportion de standards A, 57.79%, mais on distingue visuellement les deux configurations l’une de l’autre, celle résultant d’un choix uniforme étant caractérisée par une corrélation spatiale moins élevée. En effet, dans le cas des champs (C_0), la corrélation est quasi nulle ($|\rho_1| < 2.10^{-3}$), quelles que soient la règle d’affectation et la valeur de τ . Il est donc clair que la corrélation spatiale constitue un bon critère de différenciation des champs issus d’une procédure de balayage avec règle d’adoption locale, une corrélation positive supérieure à 0.002 correspondant à un champ de type (C).

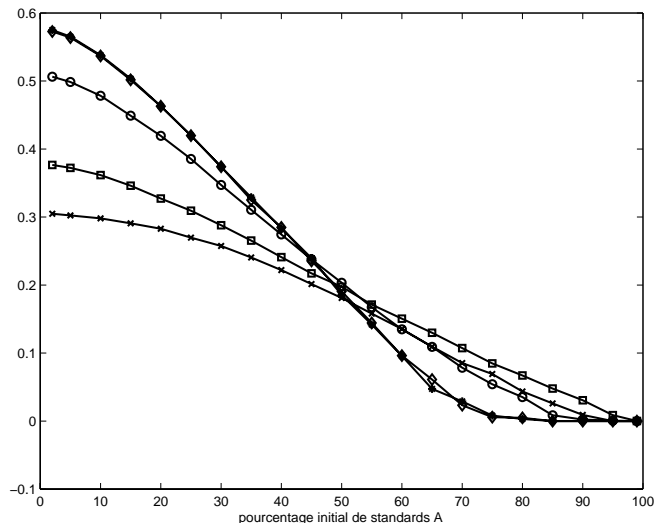


FIG. 2.3 – Corrélation à distance 1 suivant les règles d’affectation (courbes Ising $\beta = 3$ et forte quasi confondues) : x : faible; \square : Ising $\beta = 0.5$; o : Ising $\beta = 1$; \diamond : Ising $\beta = 3$; * : forte

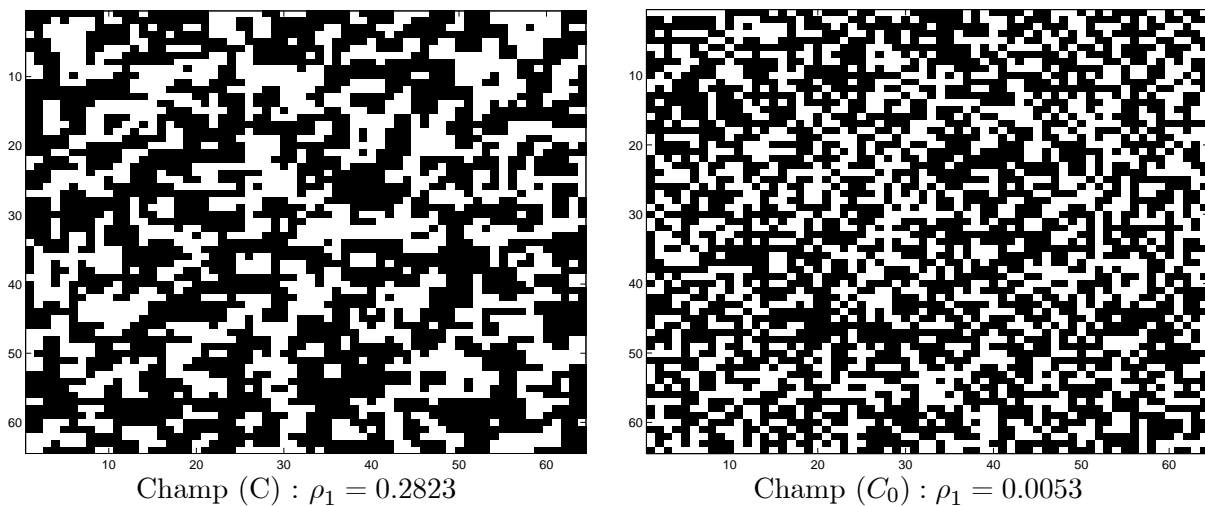


FIG. 2.5 - Champs (C) issu d’un Ising $\beta = 0.5$ et $\tau = 0.05$, et (C_0) uniforme, $f_A = 57.79\%$.

Mesures d’agrégation spatiale

Lorsque nous examinons des réalisations de champs (C) et les comparons à celles de champs (C_0), une différence essentielle est la formation d’agrégats ou clusters pour les champs (C) (cf. l’exemple ci-dessus). Nous proposons plusieurs indicateurs pour évaluer cette différenciation spatiale.

L’*indicateur d’agrégation IA* est le nombre d’arêtes joignant des sites voisins de même standard A, rapporté au nombre total de sites au standard A.

Nous calculons également trois indicateurs de connexité : *ncc* le nombre de composantes connexes de sites A, *mcc* la taille moyenne de ces composantes, et *maxcc* la taille de la plus

grande composante connexe. Nous étudions l'évolution de ces indicateurs pour les différentes règles d'affectation et les différentes valeurs de τ pour les champs (C) et (C_0).

Pour l'exemple Ising $\beta = 0.5$, $\tau = 0.05$ ($f_A = 57.79\%$), les indicateurs sont les suivants :

Champ (C)	Champ (C_0)
$\rho_1 = 0.2823, IA = 1.3520,$	$\rho_1 = 0.0053, IA = 1.1019,$
$ncc = 59, mcc = 38.08,$	$ncc = 185, mcc = 12.15,$
$maxcc = 2132$	$maxcc = 268$

Hormis $maxcc$, ces indicateurs caractérisent bien l'organisation spatiale des champs issus d'une procédure de choix en un balayage si on sait que le taux initial de contamination est peu élevé. En revanche, dès que ce taux initial est relativement important (20 à 50% selon les indicateurs et les règles), ils ne sont plus pertinents ; quant à $maxcc$, il ne discrimine les types de champs que lorsque le taux est faible (par exemple inférieur à 10% pour la règle d'Ising avec $\beta = 1$).

2.4 Perspectives

Reprenant les paradigmes d'adoption et de diffusion de standards en économie (Cf. par exemple [19]), nous souhaitons formaliser l'interprétation "chaîne de Markov" d'un certain nombre de dynamiques. Une dynamique est caractérisée par sa transition. La discrétisation en temps et en espace permettra une approximation de diffusion à partir de laquelle il sera plus facile de décrire les caractéristiques analytiques de la loi stationnaire et qui devrait permettre d'interpréter certains comportements de la dynamique observés empiriquement.

Bibliographie

- [1] Arthur W.B. (1989) Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, 99, 116-131.
- [2] Banerjee A.V. (1992) A simple model of herd behaviour. *The quarterly journal of economics*, Vol. CVII.
- [3] Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems. *JRSS B*, 36, 192-236.
- [4] Besag J., Moran P.A.P. (1975) On the estimation and testing of spatial interaction for Gaussian lattice processes. *Biometrika*, 62, 3, 555-562.
- [5] Bikhchandani S., Hirshleifer D., Welch I. (1998) Learning from the behaviour of others : conformity, fads, and informational cascades. *Journal of economic perspectives*, Vol 12 n°3, 151-170.
- [6] Bolthausen E. (1982) On the C.L.T. for stationnary mixing random fields. *Ann. Proba.*, vol. 10, n° 4, 1047-1050.
- [7] Chopard B., Dupuis A., Masselot A., Luthi P. (2002) Cellular automata and lattice Boltzmann techniques : an approach to model and simulate complex systems. *Adv. Complex systems*, 5 Vol 2-3, 103-246.
- [8] Cliff A.D., and Ord J.K. (1981) *Spatial autocorrelation (2nd ed)*. Pion, Londres.
- [9] Cox J.T. (1989) Coalescing random walks and voter model consensus times on the torus Z^d . *Ann. Proba.*, Vol. 17, n°4, 1333-1366.
- [10] David P.A., Foray D. (1993) Percolation structures, Markov random fields and the economics of EDI standard diffusion. in *Pogorel (ed.), Global telecommunications strategies and technological changes*, North-Holland.
- [11] Feller W. (1957) *An introduction to Probability theory and its applications*, J. Wiley, 2^{ème} Edition (Tome 1).
- [12] Galam S. (1997) Rational group decision making : a random field Ising model at T=0. *Physica A*, 238, 66-80.
- [13] Geman D. et Geman S. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE-PAMI*, 6, 721-741.
- [14] Geman D. (1990) Random fields and inverse problem in imaging. *Ecole de Saint Flour, Lect. Notes in Maths. n°1427*, Springer.
- [15] Guyon X. (1995) *Random fields on a network : modeling, statistics and applications*. Springer.

- [16] Isaacson D.L., Madsen R.Q. (1976) *Markov Chain : Theory and Application*, Wiley
- [17] Kemeny J., Snell J.L. (1960) *Finite Markov Chain*, Van Nostrand.
- [18] Malarz K., Antosiewicz W., Kapinska J., Kulakowski K., B.Tadic B. (2007) Avalanches in complex spin networks. *Physica A*, 373C, 785-795.
- [19] Orléan A. (1998) Informational influences and the ambivalence of imitation, in *Lesourne and Orléan (Eds.), Advances in self organization and evolutionary economics, Economica*, 39-56.
- [20] Phan D., Beugnard A. (2001) Moduleco, a multi-agent modular framework, for the simulation of network effects and population dynamics in social sciences, market & organisations. *8ème Rencontre Internationale ACSEG (Approches Connexionnistes en Sciences Economiques et de Gestion)*, Université de Rennes, IGR 22-23 novembre.
- [21] Prum B. (1986) *Processus sur un réseau et mesure de Gibbs*. Masson.
- [22] Robert C. (1996) *Méthodes de simulation en Statistiques*. Economica.
- [23] Thébaud G., Peyrard N., Dallot S., Calonnec A, Labonne G. (2005) Investigating disease spread between two assessment dates with permutation tests on a lattice. *Analytical and theoretical plant pathology*, vol 95 no12, 1453-1461.
- [24] Trouvé A. (1988) Problèmes de convergence et d'ergodicité pour les algorithmes de recuit parallélisés. *CRAS*, t. 307, série 1, 161-164.
- [25] Weisbuch G., Stauffer D. (2000) Hits and flops dynamics. Economic dynamics from the physics point of view. *Physica A*. n°3 – 4, 563 – 576.

Chapitre 3

Calcul de constante de normalisation pour un modèle de Gibbs

3.1 Introduction

Pour des modèles de Gibbs, il est bien connu que le calcul des lois marginales ainsi que celui de la constante de normalisation C sont en général impossible. Pour une distribution π de Gibbs discrète, cela résulte de la sommation sur un grand nombre d'éléments, rapidement infaisable. Par exemple, considérons un modèle d'Ising sur une grille 10×10 , la constante de normalisation se calcule par sommation sur 2^{100} termes. Ce problème a un impact direct et important sur de nombreuses applications ; par exemple, il rend impossible l'estimation par maximum de la vraisemblance. Des recherches récentes ont été menées pour contourner le problème ; par exemple, en statistique spatiale, la vraisemblance est remplacée par la pseudo-vraisemblance conditionnelle ([2]). Un autre contournement classique consiste à évaluer la constante de normalisation par méthodes de Monte Carlo (Cf par exemple [11]). Cependant, il est parfois possible de calculer la vraie valeur de C avec un algorithme efficace (Cf par exemple [8]).

Si l'on considère π , un processus de Gibbs sur un espace d'états fini, l'idée générale est d'envisager π comme un champ de Markov dont il est aisé de manipuler les lois conditionnelles. Cette approche permet d'établir une récurrence sur les lois marginales et le calcul de la constante de normalisation.

Nous avons retrouvé quelques résultats établis récemment ([14]), mais sous une forme différente. De plus, nous proposons une nouvelle récurrence, toujours basée sur les probabilités conditionnelles de π , aboutissant à un calcul explicite et facile à implémenter de la constante de normalisation. La performance numérique de cet algorithme est évaluée pour quelques exemples.

Nous généralisons les résultats au cas de processus de Gibbs spatiaux et calculons notamment la constante de normalisation C pour π un modèle d'Ising spatial. La méthode s'applique bien pour des lattices dont une dimension reste petite ; en pratique, pour un lattice de taille $m \times T$, T n'est pas limitée, mais m doit rester inférieure ou égale à 10 ou 12.

3.2 Récursions pour une distribution de Gibbs temporelle ([Har12])

Soient $T > 0$ un entier positif, E un espace d'états finis à N éléments, et $Z(T) = (Z_1, Z_2, \dots, Z_T)$ une variable aléatoire temporelle dont la distribution π s'écrit sous une forme factorisante sur E^T (Cf Reeves and Pettitt [14]) : pour $z(T) = (z_1, z_2, \dots, z_T)$,

$$\pi(z(T)) = C^{-1} \exp \sum_{s=1, T-1} h_s(z_s, z_{s+1}) = C^{-1} \prod_{s=1, T-1} H_s(z_s, z_{s+1}) \quad (3.1)$$

Donnons quelques notations : dans tout ce qui suit, les H_s , $s = 1, T-1$, sont des matrices de taille $N \times N$ de composantes $H_s(u, v) = \exp h_s(u, v)$. Pour deux telles matrices $N \times N$ H et G , pour $F = (F(v), v \in E)$, un vecteur ligne de taille N , et pour un vecteur colonne $B = (B(v), v \in E)$ de taille N , on note

$$HG(u, v) = \sum_{w \in E} H(u, w)G(w, v), \quad FH(v) = \sum_{u \in E} F(u)H(u, v) \text{ et } HB(u) = \sum_{v \in E} H(u, v)B(v).$$

On note également $u_s^t = (u_s, u_{s+1}, \dots, u_t)$ avec $1 \leq s < t \leq T$, π_S les lois marginales de π sur $S \subseteq \mathcal{T} = \{1, 2, \dots, T\}$, et $\pi_s^t = \pi_{\{s, \dots, t\}}$.

On peut interpréter π comme une distribution de Gibbs d'énergie

$$U_T(z(T)) = \sum_{s=1, T-1} h_s(z_s, z_{s+1})$$

associée aux potentiels de paires $(h_s)_{s=1, T-1}$. En particulier, π est un champ de Markov bilatéral aux 2 plus proches voisins :

$$\pi(z_t \mid z_s, s \neq t) = \frac{H_{t-1}(z_{t-1}, z_t)H_t(z_t, z_{t+1})}{H_{t-1}H_t(z_{t-1}, z_{t+1})} = \pi(z_t \mid z_{t-1}, z_{t+1}).$$

$Z(T)$ est aussi une chaîne de Markov d'ordre 1 avec

$$\pi(z_t \mid z_s, s \leq t-1) = \frac{H_{t-1}(z_{t-1}, z_t) \sum_{u_{t+1}^T} H_t(z_t, u_{t+1}) \prod_{s=t+1}^{T-1} H_s(u_s, u_{s+1})}{\sum_{u_t^T} H_{t-1}(z_{t-1}, u_t) \prod_{s=t}^{T-1} H_s(u_s, u_{s+1})} = \pi(z_t \mid z_{t-1}).$$

3.2.1 Récursions sur les marginales et calcul exact de la constante de normalisation

Posons $B_T = \mathbf{1}$ le vecteur colonne de taille N de composantes égales à 1. En marginalisant sur z_T , on obtient :

$$\pi_1^{T-1}(z_1, z_2, \dots, z_{T-1}) = C^{-1} \left\{ \prod_{s=1, T-2} H_s(z_s, z_{s+1}) \right\} (H_{T-1} B_T)(z_{T-1}).$$

On définit la suite de vecteurs (B_t) par $B_{t-1} = H_{t-1} B_t$, et on a alors pour $t = T$ à 2 :

$$\begin{aligned} \pi_1^t(z_1, z_2, \dots, z_t) &= C^{-1} \prod_{s=1, t-1} H_s(z_s, z_{s+1}) (H_t B_{t+1})(z_t) \\ &= C^{-1} \prod_{s=1, t-1} H_s(z_s, z_{s+1}) (H_t \cdots H_{T-1} B_T)(z_t). \end{aligned} \quad (3.2)$$

On descend ainsi jusqu'à $\pi_{\{1\}}$ et on obtient la constante de normalisation en posant $C = C \times \sum_{z_1 \in E} \pi_{\{1\}}(z_1)$:

Proposition 11 ([14], et [Har12] Proposition 3.2)

$$C = B_T^T \left\{ \prod_{s=1, T-1} H_s \right\} B_T. \quad (3.3)$$

Si les potentiels sont invariants dans le temps, avec $H_s = H$, on a simplement

$$C = \mathbf{1}^T (H)^{T-2} H_{T-1} \mathbf{1}.$$

On retrouve les résultats de Liu ([8]), ou Reeves and Pettitt ([14]). Cependant, notre formule est intéressante car elle permet de calculer la constante via un simple produit de matrices. En particulier, dans le cas des potentiels invariants dans le temps, si on peut diagonaliser la matrice H , alors le calcul s'effectue quelle que soit la dimension temporelle T .

Inversement, si on s'intéresse maintenant aux récursions "forward", en définissant $F_1 = B_T^T$ et $F_t = F_{t-1} H_{t-1}$ pour $t \geq 2$, on obtient les lois marginales de (z_t, \dots, z_T) :

$$\begin{aligned} \pi_t^T(z_t, z_{t+1}, \dots, z_T) &= C^{-1} \left\{ (F_{t-1} H_{t-1})(z_t) \prod_{s=t, T-1} H_s(z_s, z_{s+1}) \right\} \\ &= C^{-1} \left\{ (F_1 H_1 H_2 \dots H_{t-1})(z_t) \prod_{s=t, T-1} H_s(z_s, z_{s+1}) \right\}. \end{aligned}$$

Plus généralement, le résultat suivant donne la formule pour n'importe quelle marginale :

Proposition 12 1 - La loi marginale de π sur $S = \{s_1, s_2, \dots, s_q\}$ avec $1 = s_1 < s_2 < \dots < s_{q-1} < s_q = T$ est

$$\pi_S(z_1, z_{s_2}, \dots, z_{s_{q-1}}, z_T) = C^{-1} \prod_{i=1, q-1} \left(\prod_{s=s_i}^{s_{i+1}-1} H_s \right) (z_{s_i}, z_{s_{i+1}}), \quad (3.4)$$

2 - La loi marginale sur $S_{1,T} = S \setminus \{1, T\}$ est obtenue en remplaçant le premier produit $(\prod_1^{s_2-1} H_s)(z_1, z_{s_2})$ par $\mathbf{1}^T (\prod_1^{s_2-1} H_s)(z_{s_2})$ et le dernier $(\prod_{s_{q-1}}^{s_q-1} H_s)(z_{s_{q-1}}, z_T)$ par $((\prod_{s_{q-1}}^{s_q-1} H_s) \mathbf{1})(z_{s_{q-1}})$.

Performances numériques

Nous avons calculé la constante de normalisation C pour deux exemples.

Le premier est celui d'un modèle auto-logistique avec des potentiels de singletons et de paires indépendants du temps $\theta_t(z_t) = \alpha z_t$, $t = 1, T$ et $\Psi_t(z_t, z_{t+1}) = \beta z_t z_{t+1}$ pour $t \leq T-1$. Ici, $h_t(z_t, z_{t+1}) = \theta_t(z_t) + \Psi_t(z_t, z_{t+1})$ pour $t = 1, T-2$ et le potentiel terminal est $h_{T-1}(z_{T-1}, z_T) = \theta_{T-1}(z_{T-1}) + \Psi_{T-1}(z_{T-1}, z_T) + \theta_T(z_T)$.

Nous calculons la constante $C = \mathbf{1}^T (H)^{T-2} H_{T-1} \mathbf{1}$ pour des valeurs croissantes de T jusqu'à $T = 700$; le calcul est instantané dans chaque cas, c'est-à-dire que le "computing time" rendu par le logiciel (matlab), exprimé en secondes, est nul.

Le second exemple est celui d'un modèle d'Ising temporel bivarié. L'espace d'états est $E = \{0, 1\}^2$ et on considère le modèle de potentiels invariants $h_s = h$,

$$h((x_1, y_1), (x_2, y_2)) = \alpha x_1 + \beta y_1 + \gamma x_1 y_1 + \alpha x_2 + \beta y_2 + \gamma x_2 y_2 + \delta(x_1 x_2 + y_1 y_2) .$$

Nous avons calculé la constante de deux façons, en calculant directement la puissance H^{T-2} , ou bien en utilisant une diagonalisation de H et l'expression $C = \mathbf{1}^T P D^{T-2} P^{-1} H_{T-1} \mathbf{1}$. Les temps de calculs sont instantanés pour les deux méthodes, la diagonalisation n'apporte rien puisque ici la taille de H est petite, 4×4 .

3.2.2 Extension aux processus “ r -factorisables”

Reeves et Pettitt ([14]) considèrent des distributions plus générales qui se factorisent en potentiels de portée r : $\pi(z(T)) = \prod_{s=1}^{T-r} H_s(z_s, z_{s+1}, \dots, z_{s+r})$. On retrouve les résultats précédents sur les objets (*) définis sur $E^* = E^{r+1}$ comme suit : si H est une fonction sur E^* , on définit H^* sur $E^* \times E^*$ par :

$$H^*(u, v) = H(u_2^{r+1}, v_{r+2}) \prod_{i=1}^r \mathbf{1}(u_{i+1} = v_i). \quad (3.5)$$

Notamment, la constante s'exprime comme $C = \mathbf{1}^T (\prod_{s=1}^{T-r} H_s^*) \mathbf{1}$.

3.2.3 Récursion sur les marginales conditionnelle au futur

On propose un nouvel algorithme pour calculer récursivement les marginales, en utilisant un conditionnement sur le futur.

Réécrivons la distribution générale (3.1) en explicitant les potentiels de singletons et de paires :

$$h_s(z_s, z_{s+1}) = \theta_s(z_s) + \Psi_s(z_s, z_{s+1}) \text{ pour } s = 1, T$$

avec la convention $\Psi_T \equiv 0$. C'est à dire qu'on a encore pour $t = 1, T$, et $z_1^t = (z_1 z_2, \dots, z_t)$:

$$U_t(z_1^t) = \sum_{s=1, t} \theta_s(z_s) + \sum_{s=1, t-1} \Psi_s(z_s, z_{s+1}) .$$

On définit la *contribution* de π conditionnelle au futur (z_{t+1}, \dots, z_T) de la façon suivante. Tout d'abord, il est clair que $\pi(z_1, z_2, \dots, z_t \mid z_{t+1}, \dots, z_T) = \pi(z_1, z_2, \dots, z_t \mid z_{t+1})$. Alors, pour $t < T$,

$$\pi(z_1^t \mid z_{t+1}) = C_t^{-1}(z_{t+1}) \exp U_t^*(z_1^t; z_{t+1}),$$

où $U_t^*(z_1^t; z_{t+1}) = U_t(z_1^t) + \Psi_t(z_t, z_{t+1})$ est l'énergie conditionnelle au futur et $C_{t+1}(z_{t+1}) = \sum_{u_1^t \in E^t} \exp \{U_t^*(u_1^t; z_{t+1})\}$.

On définit alors pour $1 \leq t \leq T$,

$$\gamma_t(z_1^t; u) = \exp U_t^*(z_1^t; u)$$

la contribution de π_1^t conditionnelle au futur $z_{t+1} = u$, puis $\Gamma_t(z_1^t)$ le vecteur ligne de taille N de ces contributions de π à l'instant t conditionnelles au futur

$$\Gamma_t(z_1^t)(u) = \gamma_t(z_1^t; u) \quad ; \quad u \in E .$$

Si $t = T$, $\Gamma_T(z(T))$ est le vecteur constant de composantes $\gamma_T(z(T)) = \exp U_T(z(T))$.

Posons alors pour $1 \leq t \leq T$, $H_t(u, v) = \exp\{\theta_t(u) + \Psi_t(u, v)\}$, $u, v \in E$; on définit la suite de vecteurs colonnes ($D_t, t = T, 2$) par $D_T = (1, 0, \dots, 0)^T$, et $D_{t-1} = H_t D_t$ for $t \leq T$. On obtient le résultat suivant.

Proposition 13 ([Har12] Propositions 3.1 et 3.2)

1 - Formule de récurrence sur les contributions conditionnelles au futur.

Pour $2 \leq t \leq T$,

$$\sum_{z_t \in E} \Gamma_t(z_1^{t-1}, z_t) = \Gamma_{t-1}(z_1^{t-1}) H_t.$$

2 - Expression des lois marginales.

Pour $1 \leq t \leq T$,

$$\pi_1^t(z_1^t) = C^{-1} \times \Gamma_t(z_1^t) D_t. \quad (3.6)$$

De la même façon que précédemment, ce résultat s'étend pour des processus dont la distribution s'exprime comme un produit de potentiels de portée r .

3.2.4 Cas d'un processus de Gibbs spatial

L'idée est de considérer le champ spatial comme un processus temporel multidimensionnel.

Soit $Z = (Z_s, s = (t, i) \in \mathcal{S})$ un champ spatial sur $\mathcal{S} = \mathcal{T} \times \mathcal{I}$, où $\mathcal{T} = \{1, 2, \dots, T\}$ et $\mathcal{I} = \{1, 2, \dots, m\}$, et avec $Z_s \in F$. On considère alors le processus temporel $Z_t = (Z_{(t,i)}, i \in \mathcal{I})$, avec $Z_t \in E = F^m$, sur lequel on a les résultats précédents.

Supposons que Z est un processus de Gibbs de loi π caractérisée par une énergie U . On réécrit cette énergie en fonction de potentiels liés à la dimension "temporelle"; si la plus grande hauteur de ces nouveaux potentiels est r , alors (Z_t) est un champ de Markov aux $2r$ plus proches voisins mais aussi un processus de Markov de mémoire r ; $Y_t = (Z_t, Z_{t+1}, \dots, Z_{t+r})$, $t = 1, T - r$, est une chaîne de Markov sur $E^* = E^r$ pour laquelle on a les résultats sur les lois marginales (3.6) et le calcul de la constante de normalisation (3.3).

Exemple : le modèle d'Ising

On considère $Z = (Z_{(t,i)}, (t, i) \in \mathcal{S})$ un champ markovien sur le lattice $\mathcal{S} = \mathcal{T} \times \mathcal{I} = \{1, 2, \dots, T\} \times \{1, 2, \dots, m\}$ muni du système de voisinage aux 4 plus proches voisins, avec $Z_{(t,i)} \in F = \{-1, +1\}$. La loi jointe de Z est caractérisée par les potentiels de paires et de singletons suivants :

$$\begin{aligned} \Phi_{t,i}(z) &= \alpha z_{(t,i)} \quad \text{pour } (t, i) \in \mathcal{S}, \\ \Phi_{\{(t,i),(t,i+1)\}}(z) &= \beta z_{(t,i)} z_{(t,i+1)} \quad \text{pour } 1 \leq i \leq m - 1, \\ \Phi_{\{(t,i),(t+1,i)\}}(z) &= \delta z_{(t,i)} z_{(t+1,i)} \quad \text{pour } 2 \leq t \leq T. \end{aligned}$$

On réécrit alors l'énergie en fonction des nouveaux potentiels donnés par

$$\begin{aligned}\theta_t(z_t) &= \theta(z_t) = \alpha \sum_{i=1,m} z_{(t,i)} + \beta \sum_{i=1,m-1} z_{(t,i)} z_{(t,i+1)}, \\ \Psi_t(z_t, z_{t+1}) &= \Psi(z_t, z_{t+1}) = \delta \sum_{i=1,m} z_{(t,i)} z_{(t+1,i)}, \quad 1 \leq t \leq T-1.\end{aligned}$$

Nous avons calculé la constante de normalisation pour ce modèle avec $m = 10$ et des valeurs croissantes de T jusque $T = 100$. Nous utilisons deux méthodes : par calcul des puissances H^{T-2} de H , ou bien en utilisant une diagonalisation de H . Il s'avère que les temps de calcul sont plus longs (d'environ 30 secondes) dans la méthode utilisant la diagonalisation. En effet, le temps utilisé pour la procédure de diagonalisation est trop important et on perd le bénéfice de son usage. On obtient rapidement le calcul de C , par exemple en moins de 20 secondes pour un lattice de taille 10×100 .

Perspectives

Il reste à définir de bonnes approximations pour le calcul des marginales et de la constante pour des grands lattices ; si le lattice est de taille $T \times M$, avec $M \leq T$, une idée est de le découper dans le sens de M en petits morceaux de taille $T \times m$ sur lesquels on peut mener des calculs exacts, puis de "recoller" les résultats.

Bibliographie

- [1] Bartolucci F., Besag J. (2002) A recursive algorithm for Markov random fields, *Biometrika* 89, 3 : 724-730.
- [2] Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems, *J.R.S.S. B*, 36 : 192-236.
- [3] Friel, A.N. Pettitt, R. Reeves, and E. Wit. (2009) Bayesian inference in hidden Markov random fields for binary data defined on large lattices. *Journal of Computational and Graphical Statistics*. vol 18, pp. 243-261.
- [4] Friel N., Rue H. (2007) Recursive computing and simulation-free inference for general factorizable models. *Biometrika*. vol. 94, pp. 661-672.
- [5] Guyon X. (1995) *Random fields on a network : modeling, statistics and applications*, Springer.
- [6] Kindermann R., Snell J.L. (1980) *Markov random fields and their applications*, Contemp. Maths.
- [7] Lauritzen S. L. (1996) *Graphical models*, Oxford Science Publications.
- [8] Liu J. (2001) *Monte Carlo strategies in scientific computing*, Springer.
- [9] Lovinson G. (2006) A matrix-valued Bernoulli distribution, *J. of Mult. Anal.*, 97 : 1573-1585.
- [10] Mcgrory C.A., Titterington M., Reeves R., Pettitt A.N. (2009) Variational Bayes for estimating the parameters of a hidden Potts model. *Statistics and computing*. Vol 19 329-340.
- [11] Moeller J., Pettitt A.N., Reeves R., Berthelsen K.K. (2006) An efficient Markov chain Monte Carlo method for distributions with intractable normalizing constants, *Biometrika* 93, 2 : 451-458.
- [12] Nandhini G., Sangaranarayanan M.V. (2009) Partition function of nearest neighbour Ising models : some new insights. *J. Chem. Sci.*, Vol 121, No5, 595-599.
- [13] Pettitt A.N., Friel N., Reeves R. (2003) Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice, *J.R.S.S. B*. 65, Part 1 : 235-246.
- [14] Reeves R., Pettitt A.N. (2004) Efficient recursions for general factorisable models, *Biometrika* 91, 3 : 751-757.
- [15] Tjelmeland H., Austad H.M. (2010) Exact and approximate recursive calculations for binary Markov random fields defined on graphs. *Technical report, Statistics 2/2010, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.*

Chapitre 4

Quelques résultats en statistique paramétrique

4.1 Estimation par minimum de contraste en situation non ergodique

4.1.1 Processus de contraste fort ([Har0, Har1])

Si P_θ , $\theta \in \Theta$, est un modèle paramétré pour la variable $Y = (Y_i, i \in S)$ où S est infini dénombrable ($S = \mathbb{Z}^d$ par exemple), la méthode d'estimation par minimum de contraste pour θ consiste à choisir une valeur $\hat{\theta}$ minimisant une fonctionnelle $\alpha \rightarrow U_n(\alpha, Y(n))$ où $Y(n) = \{Y_i, i \in D_n\}$, D_n étant une suite de domaines d'observations tendant vers S . Cette fonctionnelle U_n est par exemple, l'opposée de la log-vraisemblance, la pseudo-vraisemblance conditionnelle de Besag pour un champ de Markov, le contraste de Whittle pour une série chronologique, la somme des carrés résiduels pour un modèle conditionnel. Les propriétés asymptotiques de cette méthode sont étudiées dans Dacunha-Castelle et Duflo ([11]) dans le cas ergodique. Dans le cadre de ma thèse, j'ai proposé une généralisation à la définition et aux propriétés des processus de contraste de ([11]). D'une part, la notion de contraste fort, pour laquelle les conditions habituelles en P_θ -probabilité sont remplacées par des conditions P_θ -presque sûre, permettent d'obtenir un résultat de consistance presque sûre de l'estimateur. D'autre part, sous une condition de sous-ergodicité, nous obtenons la consistance et la normalité asymptotique dans la situation non ergodique. A partir de ces résultats, on bâtit un test asymptotique d'une sous-hypothèse par différence de contraste qui suit la loi d'une somme pondérée de Chi-deux à un degré de liberté et indépendantes. ([Har1-2]).

Nous considérons le modèle paramétrique $(\Omega, \mathcal{A}, (P_\alpha)_{\alpha \in \Theta})$ où Θ est un compact de \mathbb{R}^k d'intérieur $\hat{\Theta}$ non vide. On note la filtration $\mathcal{F} = \{\mathcal{F}_n, n \in \mathbb{N}\}$ avec $\mathcal{F}_n = \sigma(Y_i, i \in D_n)$.

Une *fonction de contraste* relative à θ est une fonction $\alpha \rightarrow K(\theta, \alpha)$ de Θ dans \mathbb{R}^+ ayant un minimum unique en $\alpha = \theta$ ([11]).

Un *processus de contraste fort* relatif à θ et à K est une fonction U de $\Omega \times \mathbb{N} \times \Theta$ dans \mathbb{R} , \mathcal{F} -adaptée, vérifiant :

$$\liminf_{n \rightarrow \infty} (U_n(\alpha) - U_n(\theta)) \geq K(\theta, \alpha) \quad P_\theta - p.s. \quad (4.1)$$

La condition initiale de ([11]) sur le processus de contraste U est

$$\lim_{n \rightarrow \infty} (U_n(\alpha) - U_n(\theta)) = K(\theta, \alpha) \quad \text{en } P_\theta \text{ - probabilité.} \quad (4.2)$$

Cette condition est vérifiée pour des modèles stationnaires et ergodiques dès que la paramétrisation est propre ; c'est pourquoi on parle de *condition ergodique*, alors que (4.1) sera dite *condition de sous-ergodicité*.

Soit $\hat{\theta}_n = \text{Arg inf}_{\alpha \in \Theta} U_n(\alpha)$ l'estimateur du minimum de contraste. Les résultats suivants établissent la consistance forte puis la normalité asymptotique de cet estimateur ([Har0], [Har1], [21]).

Posons les deux conditions suivantes :

(C₁) : $\alpha \rightarrow U_n(\alpha, \omega)$ et $\alpha \rightarrow K(\theta, \alpha)$ sont continues.

(C₂) : Notant W_n le module de continuité de U_n ,

$$W_n(\eta) = \sup \{|U_n(\alpha) - U_n(\beta)|, |\alpha - \beta| < \eta, \alpha \text{ et } \beta \in \Theta\},$$

il existe une suite (ε_k) qui tend vers zéro telle que $P_\theta[\limsup_{n \rightarrow \infty} W_n(1/k) < \varepsilon_k] = 1$.

Théorème 14 *Consistance forte ([Har1], [Har0])*

Sous (C₁) et (C₂), $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta$ P_θ -presque sûrement.

Remarque : Si (4.1) et (C₂) sont affaiblies respectivement en $\liminf_{n \rightarrow \infty} (U_n(\alpha) - U_n(\theta)) \geq K(\theta, \alpha)$ en P_θ -probabilité et $\lim_{n \rightarrow \infty} P_\theta[W_n(1/k) < \varepsilon_k] = 1$, alors on a $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta$ en P_θ -probabilité.

Corollaire 15 *Contraste additif ([Har1], [Har0])*

Soit $U_n = \sum_{k=1}^K a_{n,k} U_{n,k}$, $a_{n,k} \geq 0$ et supposons que

(1) les processus $U_{n,1}, \dots, U_{n,K}$ satisfont (C₂) et sont continus en α ,

(2) il existe un k_0 tel que U_{n,k_0} est un processus de contraste fort relatif à une fonction de contraste $K_0(\theta, \alpha)$ continue en θ ,

(3) pour ce même k_0 , il existe $a_0 = \liminf_{n \rightarrow \infty} a_{n,k_0} > 0$,

(4) pour tout $k \neq k_0$, $\liminf_{n \rightarrow \infty} (U_{n,k}(\alpha) - U_{n,k}(\theta)) \geq 0$.

Alors l'estimateur de minimum de contraste associé au processus U_n converge presque sûrement vers θ .

Le corollaire est utile dès que le processus de contraste se décompose en une somme de sous-contrastes dont l'un au moins se traite bien. En particulier, on appliquera ce résultat pour le contraste de la pseudo-vraisemblance conditionnelle de Besag ([3]) en le décomposant sur des sous-ensembles de codage.

Le cas des processus de contrastes convexes est traitée par Senoussi ([31]) sous la condition d'ergodicité (4.2) ; dans ce cas, on n'a pas besoin de condition sur le module de continuité.

Présentons rapidement quelques applications.

Variables indépendantes

Soit une suite de variables aléatoires indépendantes Y_1, Y_2, \dots . Ici Y_i est à valeurs dans Ω_i , espace mesurable, de loi $P_{\theta, i}$. Pour chaque $i \in \mathbb{N}$, soit $g_i : \Omega_i \times \Theta \mapsto \mathbb{R}$, mesurable en $Y_i \in \Omega_i$ et continue en θ , et soit $\{b_n, n \in \mathbb{N}\}$ une suite de réels positifs tendant vers $+\infty$.

On considère alors le contraste : $U_n(\alpha) = \frac{1}{b_n} \sum_{i=1}^n g_i(Y_i, \alpha)$.

Des choix classiques pour g_i sont $g_i(Y_i, \alpha) = -\log f_i(Y_i, \alpha)$ si f_i est la densité de Y_i (maximum de vraisemblance pour des variables indépendantes non identiquement distribuées), ou $g_i(Y_i, \alpha) = (Y_i - E_\alpha(Y_i))^2$ (moindres carrés ordinaires).

Contrastes du codage et de la pseudo-vraisemblance conditionnelle pour les modèles markoviens

S est ici un ensemble infini dénombrable de sites, muni d'un graphe symétrique définissant la relation de voisinage. Le champ markovien est spécifié par une famille de densités conditionnelles $\pi_A(y_A/y_{\partial A}, \alpha)$, $A \subset S$.

Soient B_n une suite strictement croissante de domaines d'observations et C un sous-ensemble infini de codage ; $C_n = B_n \cap C$. On étudie les deux contrastes introduits par Besag ([3]) :

Contraste du codage : $U_n^C(\alpha) = -\frac{1}{c_n} \sum_{i \in C_n} \log \pi_i(y_i | y_{\partial i}, \alpha)$ et $\hat{\theta}_n^C = \arg \min_{\Theta} U_n^C(\alpha)$.

Contraste de la pseudo-vraisemblance conditionnelle : $U_n(\alpha) = -\frac{1}{b_n} \sum_{B_n} \log \pi_i(y_i/y_{\partial i}, \alpha)$ et $\hat{\theta}_n = \arg \min_{\Theta} U_n(\alpha)$.

En s'appuyant sur le résultat précédent pour des observations conditionnellement indépendantes, on obtient la consistance de $\hat{\theta}_n^C$ et $\hat{\theta}_n$ dans un cadre assez général (Cf. [Har0] et [Har1]).

Variables dépendantes : Régression non linéaire à résidu ergodique.

On considère le modèle $y_t = f(x_t, \alpha) + \varepsilon_t$, $t = 1, \dots, T$, où le résidu est ergodique, et le contraste des moindres carrés $U_T(\alpha) = \frac{1}{T} \sum_{t=1}^T (y_t - f(x_t, \alpha))^2$.

Par exemple ε est gaussien de densité spectrale intégrable égale à $g(\lambda)|1 - e^{i\lambda}|^{-2d}$ avec $0 \leq d < 1/2$.

4.1.2 Normalité asymptotique ([Har0-1-2])

La vraie valeur du paramètre sera notée ici $\theta_0 \in \hat{\Theta}$ et $U_n^{(i)}(\theta_0)$ est la i ème dérivée partielle de U_n en θ_0 .

Pour deux matrices symétriques $p \times p$ A et B , on notera $A \geq B$ (resp. $A > B$) lorsque la différence $A - B$ est une matrice semi-définie positive (resp. définie positive). Pour une matrice A qui admet une diagonalisation $A = PDP^{-1}$, on note $A^{\frac{1}{2}} = PD^{\frac{1}{2}}P^{-1}$.

Posons les conditions suivantes :

(N_1) : Il existe un voisinage V de θ_0 sur lequel U_n est deux fois continûment différentiable et une variable aléatoire G P_{θ_0} -intégrable, tels que pour tous $1 \leq i, j \leq p$ et $\alpha \in V$, on ait $|\frac{\partial^2 U_n}{\partial \alpha_i \partial \alpha_j}(\alpha, \omega)| \leq G(\omega)$.

(N_2) : Il existe une suite $(I_{n,0})$ de matrices $p \times p$ et une matrice $I_0 > 0$ telles que

- (i) à partir d'un certain rang $I_{n,0} \geq I_0$
- (ii) $\lim_n \left(U_n^{(2)}(\theta_0) - I_{n,0} \right) = 0$ en P_{θ_0} -probabilité

- (N_3) : Il existe une suite (a_n) tendant vers l'infini et une matrice J_0 telles que
- (i) à partir d'un certain rang, $J_{n,0} = \text{Var} \left(\sqrt{a_n} U_n^{(1)}(\theta_0) \right)$ existe et vérifie $J_{n,0} \geq J_0$
 - (ii) $\sqrt{a_n} J_{n,0}^{-1/2} U_n^{(1)}(\theta_0) \xrightarrow{\mathcal{L}} N(0, I_p)$

Théorème 16 ([Har0]) *Supposons que l'estimateur du minimum de contraste θ_n de U_n soit consistant et que les hypothèses (N_1), (N_2) et (N_3) soient vérifiées. Alors*

$$\sqrt{a_n} J_{n,0}^{-1/2} I_{n,0} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, I_p)$$

Les conditions (N) ne sont pas d'un usage facile. Lorsque le contraste est de type additif, $U_n(\alpha) = \frac{1}{b_n} \sum_{i=1}^n g_i(Y_i, \alpha)$, et que le processus est faiblement dépendant, elles sont remplacées par des conditions portant sur la régularité des fonctions g et de mélange sur Y (Cf [Har0]).

Ces résultats permettent de construire un test statistique pour deux hypothèses emboîtées.

4.1.3 Test asymptotique de différence de contrastes ([Har0-1-2])

Considérons l'hypothèse (H) : “ $\theta \in \Theta$, ouvert de \mathbb{R}^p ”. Soit q un entier plus petit que p ; nous voulons alors tester la sous-hypothèse nulle (H_0) : “ $\theta = r(\alpha)$ ”, où r est une application de classe \mathcal{C}^2 d'un ouvert $\Lambda \subset \mathbb{R}^q$ dans Θ telle que

$$\theta_0 = r(\alpha_0) \text{ et } R_0 = \frac{\partial r}{\partial \alpha}(\alpha_0) \text{ est de rang } q.$$

Notons $\hat{\theta}_n$ et $\bar{\theta}_n = r(\hat{\alpha}_n)$ les estimateurs de minimum de contraste sous (H) et (H_0), et notons également $\bar{I}_n, \bar{J}_n, \bar{I}_0, \bar{J}_0$ les matrices apparaissant dans les conditions (N) relatives au contraste $U_n(r(\alpha))$ sous (H_0). Posons

$$A_n = J_{n,0}^{1/2} \left(I_{n,0}^{-1} - R_0 \bar{I}_n^{-1} R_0^T \right) J_{n,0}^{1/2} \quad (4.3)$$

A_n est une matrice symétrique positive de rang $p - q$.

Le test de sous-hypothèse est basé sur la statistique

$$T_n = 2a_n \left(U_n(\bar{\theta}_n) - U_n(\hat{\theta}_n) \right) .$$

Théorème 17 ([Har0])

Supposons que les estimateurs $\hat{\theta}_n$ et $\bar{\theta}_n$ soient convergents et que le contraste U_n vérifie, sous (H) et sous (H_0), les hypothèses (N_1) à (N_3). Alors, sous (H_0),

$$T_n \sim \sum_{i=1}^{p-q} \lambda_{i,n} \chi_{1,i}^2$$

où les $\chi_{1,i}^2$ sont des χ_1^2 indépendantes et les $\lambda_{i,n}$ sont les $p - q$ valeurs propres strictement positives de la matrice A_n définie par (4.3).

Remarques et exemples :

- *Cas stationnaire :*

Supposons que $\lim I_{n,0} = I_0$, $\lim \bar{I}_{n,0} = \bar{I}_0$, $\lim J_{n,0} = J_0$ et notons $A_0 = J_0^{1/2} (I_0^{-1} - R_0 \bar{I}_0^{-1} R_0^T) J_0^{1/2}$; alors sous l'hypothèse (H_0) on a $T_n \xrightarrow{\mathcal{L}} \sum_{i=1}^{p-q} \lambda_i \chi_1^2$ où les λ_i sont les $p - q$ valeurs propres de la matrice A_0 .

- *Cas $I_{n,0} = J_{n,0}$:*

Dans ce cas, la matrice $A_{n,0}$ est idempotente de rang $p - q$ ($\lambda_{i,n} = 1$) et la statistique T_n suit une loi de χ_{p-q}^2 ; on retrouve le résultat classique du test du rapport de vraisemblance mais dans une situation non nécessairement stationnaire puisqu'il suffit que pour chaque n , $I_{n,0} = J_{n,0}$. Ce sera le cas pour des observations indépendantes mais non identiquement distribuées avec le contraste de la vraisemblance ([Har0]), ou une chaîne de Markov non homogène dans le temps avec également le contraste de la vraisemblance ([Har2]), ou encore pour un champ de Markov avec le contraste du codage ([Har1]).

- Divers exemples sont étudiés dans [Har1] et [Har2]; le cadre du champ markovien est détaillé dans [Har1]. Nous examinons dans [Har2] le cas d'une chaîne de Markov inhomogène, puis celui d'un modèle d'Ising aux 4 plus proches voisins; enfin, nous considérons une série chronologique pour laquelle le paramètre est estimé par contraste de Whittle avec le test de vraisemblance.

- *Puissance du test :*

Nous avons évalué empiriquement la puissance du test pour deux exemples ([Har2]). Nous utilisons l'approximation $\sqrt{a_n} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, \tilde{I}_n \tilde{J}_n \tilde{I}_n^{-1})$ où $\tilde{I}_n = U_n^{(2)}(\hat{\theta}_n)$ et $\tilde{J}_n = a_n U_n^{(1)}(\hat{\theta}_n) U_n^{(1)}(\hat{\theta}_n)^T$. La fonction de répartition d'une somme pondérée de χ_1^2 n'ayant pas de forme explicite, elle est évaluée de façon numérique.

Nous considérons d'abord une chaîne de Markov homogène sur \mathbb{Z} , à deux états $\{-1, +1\}$; c'est encore un champ de Markov aux deux plus proches voisins de lois conditionnelles

$$P(Y_k = y \mid Y_{k-1} = x, Y_{k+1} = z) = \frac{\exp\{y(\alpha + \beta(x + z))\}}{2 \cosh(\alpha + \beta(x + z))}.$$

La chaîne Y est ergodique et exponentiellement mélangeante; on veut alors tester (H_0) : $\alpha = \beta = 0$; on envisage trois procédures de test basées sur les contrastes de codage (avec $C = \{2k, k \in \mathbb{Z}\}$), de pseudo-vraisemblance conditionnelle et de vraisemblance. Sous l'hypothèse (H_0) , les trois statistiques de test ont pour lois asymptotiques respectives

$$T_n^C \sim \chi_2^2, \quad T_n^{PVC} \sim \chi_{1,1}^2 + 2\chi_{1,2}^2, \quad TV \sim \chi_2^2. \quad (4.4)$$

On sait que la précision des estimateurs va en s'améliorant en allant du contraste du codage à celui de la vraisemblance en passant par la pseudo-vraisemblance ([22]); nous avons tracé les courbes de puissance empirique des trois tests sous l'alternative (H_1) : $\alpha = 0, \beta \neq 0$ correspondant à une loi stationnaire uniforme: les puissances sont similaires. En revanche pour une autre alternative (H_2) : $\alpha \neq 0, \beta = 0$ correspondant au cas i.i.d., on observe que la méthode de codage est inférieure à celle de pseudo-vraisemblance, elle-même inférieure à la vraisemblance.

Le second exemple est celui d'une série chronologique non gaussienne :

$$(1 + \alpha)Y_t + \frac{\beta}{2}(Y_{t-1} + Y_{t+1}) = \varepsilon_t$$

où $|\beta| < |1 + \alpha|$ et (ε_t) est un bruit blanc i.i.d. centré et non gaussien de variance σ^2 . Notons que Y est exponentiellement mélangeant. On utilise ici le contraste de Whittle défini par $U_n(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)} \right) d\lambda$, où f_θ désigne la densité spectrale et $I_n(\lambda)$ le périodogramme.

On teste alors $(H_0) : \alpha = \beta = 0$ correspondant au cas i.i.d.. La loi asymptotique de la statistique de différence de contraste est une somme pondérée de deux χ_1^2 indépendants, la pondération faisant intervenir le coefficient d'aplatissement du bruit. Nous comparons empiriquement la puissance de ce test avec celle du rapport de vraisemblance : les puissances restent très proches, le test de vraisemblance ayant une avance de l'ordre de 10% sur le test de Whittle aux points les plus différenciés.

4.2 Estimation sur le log-périodogramme “régularisé” ([Har0, Har3])

Parmi les méthodes d'estimation de processus stationnaires, la procédure de Whittle donne un estimateur consistant, asymptotiquement normal, et de plus dans le cas gaussien, asymptotiquement efficace. Cette procédure est encore valide dans le cas de processus stationnaires fortement dépendants (appelés encore à mémoire longue), pour lesquels la densité spectrale du processus n'est pas bornée.

Nous introduisons une procédure d'estimation basée sur le log-périodogramme ; cette méthode a été proposée à l'origine par Geweke et Porter-Hudak ([18]) puis Kashyap et Eom ([23]). Cette méthode est particulièrement intéressante pour des modèles faisant intervenir des paramètres en exposant ; c'est le cas des processus fractionnaires, par exemple les modèles FARIMA(p, d, q).

Le modèle général est le suivant :

$$X_t = A(B)\varepsilon_t \quad (4.5)$$

observé sur $t = 1, \dots, n$, où (ε_t) est un bruit blanc et $A(B) = \sum_{j \geq 0} a_j B^j$ avec $a_0 = 1$ et $\sum_{j \geq 0} a_j^2 < \infty$, B étant l'opérateur retard usuel ($BX_t = X_{t-1}$).

Notons $X_n(\lambda) = \sum_{t=1}^n X_t e^{it\lambda}$, $\varepsilon_n(\lambda) = \sum_{t=1}^n \varepsilon_t e^{it\lambda}$, et $I_{n,X}(\lambda) = \frac{1}{2\pi n} |X_n(\lambda)|^2$ et $I_{n,\varepsilon}(\lambda) = \frac{1}{2\pi n} |\varepsilon_n(\lambda)|^2$ les périodogrammes de X et ε .

Davis et Jones ([13]) et Brillinger ([6]) ont donné des résultats sur le comportement asymptotique du périodogramme dans le cas gaussien et à mémoire courte ; Geweke et Porter-Hudak ([18]) traitent le cas d'un FARIMA(p, d, q) avec $-\frac{1}{2} < d < 0$ (processus à courte portée) mais dans un cadre non gaussien, et Künsch ([27]) celui d'un processus fractionnaire non gaussien et à mémoire longue (Cf [Har0]).

Reprenons le modèle (4.5) : $X_t = A(\theta, B)\varepsilon_t$ dépend d'un paramètre $\theta \in \Theta \subset \mathbb{R}^p$, Θ compact.

Taniguchi ([33]) a comparé les estimateurs de Whittle et des moindres carrés sur le log-périodogramme dans le cas gaussien : tous deux sont asymptotiquement gaussiens ; l'estimateur basé sur le périodogramme n'est pas asymptotiquement efficace, au contraire de celui de Whittle, mais présente des propriétés de robustesse ; mais surtout, lorsque le paramètre du modèle est en exposant, il est solution d'une équation linéaire ce qui simplifie avantageusement les calculs.

Notons maintenant $J_{n,X} = |A(\theta, e^{i\lambda})|^2 I_{n,\varepsilon}(\lambda)$ l'approximation de $I_{n,X}(\lambda)$ et réécrivons cette égalité en en prenant le log ; on a alors

$$\log J_{n,X}(\lambda) = \log |A(\theta, e^{i\lambda})|^2 + \log I_{n,\varepsilon}(\lambda) . \quad (4.6)$$

On estime les paramètres du modèle en considérant la régression (4.6) aux fréquences $\lambda_k = \frac{2\pi k}{n}$, $k = 1$ à $\frac{n}{2}$. Si ε est un processus i.i.d. gaussien, l'estimateur du contraste des moindres carrés $U_n(\theta) = \frac{2}{n} \sum_{k=0}^{n/2} (\log J_{n,X}(\frac{k}{n}) - \log |A(\theta, e^{i\lambda})|^2 + \nu)^2$ (où ν est la constante d'Euler) est asymptotiquement normal, mais pas asymptotiquement efficace. C'est le principe de la démarche proposée par Kashyap et Eom ([23]) pour estimer les paramètres d et σ^2 d'un processus à longue portée $(1 - B)^d X_t = \sigma \varepsilon_t$ où (ε_t) est i.i.d. $N(0, 1)$ et $0 < d < 1/2$.

L'idée proposée dans [Har0] est de "régulariser" le périodogramme pour produire un estimateur efficace ; il s'agit de découper l'intervalle d'observation $[1, n]$ en m paquets de longueur K :

$$\left\{1, \dots, \frac{n}{2}\right\} = \bigcup_{l=0}^{\frac{m}{2}-1} \mathcal{I}_l \text{ avec } \mathcal{I}_l = \{lK + 1, \dots, (l+1)K\}$$

et on travaille avec les pseudo-périodogramme et périodogramme régularisés $\bar{J}_X = \frac{1}{K} \sum_{k \in \mathcal{I}_l} J_{n,X}(\frac{k}{n})$ et $\bar{I}_\varepsilon = \frac{1}{K} \sum_{k \in \mathcal{I}_l} I_{n,\varepsilon}(\frac{k}{n})$.

Sous des hypothèses de régularité de la densité spectrale de X et d'un "bon" découpage, on obtient un estimateur de la régression sur le log périodogramme régularisé asymptotiquement normal et efficace ; la variance asymptotique est celle de l'estimateur de Whittle calculé avec $J_{n,X}$.

L'ensemble de ces résultats a été présenté dans la thèse ([Har0]) et est résumé dans [Har3]. Ils s'appliquent à une grande classe de modèles, par exemple l'ensemble des processus MADE de Gonçalves ([19]) de type $\prod_j (1 - \alpha_j B)^{d_j} X_t = \varepsilon_t$.

Nous donnons également dans [Har0] quelques résultats lorsque le bruit ε n'est pas gaussien, en supposant notamment qu'il a moment d'ordre 4.

Le pseudo périodogramme $J_{n,X}$ n'étant pas observable, nous le remplaçons dans la pratique par $I_{n,X}$. Solo ([32]) a montré que la différence $\delta_n(\lambda) = I_{n,X}(\lambda) - J_{n,X}(\lambda)$ entre le vrai périodogramme et son approximation tend vers zéro, et ce à la vitesse $n^{1/4}$. L'idée abordée dans le travail avec F. Comte ([Har3] et preprints [8], [9], [10]) était de considérer la régression sur le vrai périodogramme régularisé \bar{I}_X en introduisant dans l'équation un terme d'erreur basé sur $\delta_n(\lambda)$. Nous avons également travaillé sur des extensions au cadre non gaussien. Ce gros travail issu de la thèse et de la collaboration avec F. Comte n'a pas donné lieu à publication en dehors de [Har3] mais les prépublications sont citées dans certains articles d'Econométrie et Statistique ([34]).

Bibliographie

- [1] Bayomog S. (1994) Modélisation et analyse de données spatio-temporelles. Thèse Université paris Sud.
- [2] Beran J. (1994) *Statistics for long memory processes*. Chapman and Hall.
- [3] Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems, *J.R.S.S. B*, 36 : 192-236.
- [4] Bolthausen E. (1982) On the central limit theorem for stationary mixing random fields. *Ann. Probab*, 10(4), 1047-1050.
- [5] Breiman L. (1968) *Probability*. Addison Wesley.
- [6] Brillinger D.R. (1981) *Time series, Data Analysis and theory*. Holden Day.
- [7] Brockwell P., Davis R. (1991) *Time series : theory and methods*. Springer-Verlag, New-York, 2nd Edition.
- [8] Comte F., Hardouin C. (1995) Regression on log-regularized periodogram under assumption on bounded spectral densities : the non fractional and the fractional cases. *Preprint CREST 9535*.
- [9] Comte F., Hardouin C. (1995) Regression on log-regularized periodogram for fractional models at low frequencies. *Preprint CREST 9536*.
- [10] Comte F., Hardouin C. (1995) Regression on log-regularized periodogram : comparison with Whittle estimator and study of the non Gaussian case. *Preprint CREST 9539*.
- [11] Dacunha-Castelle D. et Dufflo M. (1983) *Probabilités et statistiques 2*. Masson.
- [12] Dalhaus R. (1989) Efficient parameter estimation for Self-Similar Processes. *Annals of Statistics Vol. 17*, 1749-1766.
- [13] Davis H.T., Jones R.H. (1968) Estimation of the innovation variance of a stationary time series. *American Stat. Ass. Journal*.
- [14] Doukhan P. (1994) *Mixing : properties and examples*. Lecture Notes in Statist. 85, Springer-verlag, New-York.
- [15] Fox R., Taqqu M. (1986) Large sample properties of parameter estimates for strongly dependent stationary time series. *Annals of Stats. 14*, 5157-532.
- [16] Frigessi A., Piccioni M. (1990) Parameter estimation for two-dimensional Ising fields corrupted by noise. *Stoch. Proc and their Appl. 34*, 297-311.
- [17] Georgii H.O. (1988) *Gibbs measures and phase transitions*. De Gruyter, Berlin.
- [18] Geweke J., Porter-Hudak S. (1983) The estimation and application of long memory time series models. *JTSA Vol 4*, 221-238..

- [19] Gonçalves E. (1988) Processus fractionnaires. *Thèse, Université de Lille.*
- [20] Gray H.L., Zhang N.-F., Woodward W.A. (1989) On generalized fractional processes. *JTSA Vol 10, 233-257.*
- [21] Guyon X. (1995) *Random fields on a network : modeling, statistics and applications*, Springer.
- [22] Guyon X., Künsch H.R. (1992) Asymptotic comparison of estimators in the Ising model. *Stochastic models, statistical analysis and algorithms in image analysis (P. Barone and A. Frigessi Eds.), Lecture Notes in Statist. 74, Springer-verlag, New-York, 177-198.*
- [23] Kashyap R.L., Eom K.B. (1988) Estimation in long memory time series model. *JTSA Vol 9, 35-41.*
- [24] Hassler U. (1993) Regression of spectral estimators with fractionally integrated time series. *JTSA 14, 369-380.*
- [25] Hurvich C., Beltrao K. (1993) Asymptotics for the low-frequencies ordinates of the periodogram of a long memory time series. *JTSA Vol 14, 455-472.*
- [26] Issacson D.L., Madsen R.W. (1976) *Markov chains, theory and applications*. Wiley, New-York.
- [27] Künsch H. (1986) Discrimination between monotonic trends and long range dependence. *J. Appl. Prob. 23, 1025-1030.*
- [28] Priestley M.B. (1981) *Spectral analysis and time series*. Academic Press.
- [29] Robinson P.M. (1994) Semiparametric analysis of long-memory time series. *Annals of Stats. 22, 515-539.*
- [30] Robinson P.M. (1995) Log periodogram regression of time series with long range dependence. *Annals of Stats. 23, 1048-1072.*
- [31] Senoussi R. (1990) Statistique asymptotique presque sûre des modèles statistiques convexes. *Ann. Institut Henri Poincaré, 26, 19-44.*
- [32] Solo V. (1989) Asymptotics of quadratic forms for time series. *Preprint Dept Elec Comptr. Eng. John Hopkins University.*
- [33] Taniguchi M. (1979) On estimation of parameters of gaussian stationary processes. *J. Appl. Prob. 16, 575-591.*
- [34] Velasco C. (2000) Non-Gaussian Log-Periodogram Regression. *Econometric Theory Vol. 16, No. 1, . 44-79.*
- [35] Yao J.-F. (1990) Méthodes bayésiennes en segmentation d'image et estimation par rabotage des modèles spatiaux. *Thèse Université Paris Sud.*

Annexe A

Curriculum Vitae

Cécile HARDOUIN - CECCANTINI

Centre de recherche MODAL'X
Université Paris Ouest Nanterre La Défense
200 avenue de la République
92000 Nanterre

E-mail : cecile.hardouin@u-paris10.fr

Page web : http://www.u-paris10.fr/87676212/0/fiche___pagelibre/&RH=FR

Née le 20/02/1963

Mariée, cinq enfants

A.1 Synthèse de la carrière

Fonctions exercées

- Depuis 1992 Maître de conférences en mathématiques appliquées, section CNU 26, à l'Université Paris Ouest Nanterre la Défense (Paris X).
- 1991-1992 : ATER Université Paris 1
- 1988-1989 : professeur de mathématiques en classe de troisième
- 1987-1991 : Allocataire de recherche

Affiliations de Recherche

- Depuis 2010 Membre du centre de recherche MODAL'X,
Université Paris Ouest Nanterre La Défense.
http://www.u-paris10.fr/86618150/0/fiche_MODALX___pagelibre//

Membre associé de l'équipe SAMM^(*), Université Paris 1.
<http://samm.univ-paris1.fr/>

(*) Depuis le 1er janvier 2010, le SAMOS et l'équipe Marin Mersenne de l'Université Paris 1 constituent une nouvelle équipe nommée SAMM.

- 1991-2009 Membre du Laboratoire SAMOS, Université Paris 1.
- 1994-1998 Membre du centre de recherche MODAL'X , Université Paris 10.

Formation

- Octobre 1987 - Janvier 1992. Thèse de doctorat de Mathématiques Appliquées de l'Université Paris 7 sous la direction de Xavier Guyon.

Sujet : "Quelques résultats nouveaux en statistique des processus : contraste fort, régression à résidus à longue portée, estimation par log-périodogramme".

Jury : Mmes Laure Elie (présidente), Yvonne Girard, Dominique Picard, Mrs Jean-Marc Azais (rapporteur), Paul Doukhan, Xavier Guyon (directeur), Jose Leon (rapporteur).

- 1987 D.E.A de "Statistique et modèles mathématiques en économie" des Universités Paris 7 et Paris 1 mention AB.

A.2 Publications et travaux

Revue internationale à comité de lecture

- Hardouin C., Crivelli T. (2011) Mixed states spatio-temporal auto-models, *Statistics, Vol 45 Issue 1, Special issue : Statistics on dependent data, 87-100.*

- Hardouin C., Yao J.-F. (2008) Spatial modelling for mixed state observations, *Electronic Journal of Statistics Vol. 2, 213-233.*

- Hardouin C., Yao J.-F. (2008) Multi-parameter auto-models and their application, *Biometrika. 95(2), 335-349*

- Bouthemy P., Hardouin C., Piriou G., Yao J.-F. (2006) Mixed state auto-models and motion texture modeling, *Journal of Mathematical Imaging and Vision. Vol 25, n 3, 387-402.*

- Guyon X., Hardouin C. (2002) Markov chain Markov field dynamics : models and statistics, *Statistics 36, no. 4, 339-363.*

(Egalement référencé en 2001 par *Statistics 35, no. 4, 593-627*).

- Bayomog S., Guyon X., Hardouin C., Yao J.F. (1996) Test de différence de contrastes et somme pondérée de khi-deux, *The Canadian Journal of Statistics, Vol 24, n 1.*

Notes aux CRAS et assimilées

- Guyon X., Hardouin C (2010) Exact marginals and normalizing constant for Gibbs distributions, *C. R. Acad. Sci. Paris, Ser. I. 348. 199-201.*

- Hardouin C., Yao J.-F. (2007) Multi-parameter auto-models with applications to cooperative systems, *C. R. Acad. Sci. Paris, Ser. I 345.*

- Comte F., Hardouin C. (1997) Régression sur le log-périodogramme régularisé pour des processus gaussiens à densité spectrale bornée, *C. R. Acad. Sci. Paris, Ser. I. 325, no. 11, 1203-1206.*

Proceedings

- Hardouin C. (2007) Dumping influence on a non iterative dynamics, *Recent Advances in Stochastic Modelling and Data Analysis*, Editor C. H. Skiadas, World Scientific Publishing, 371-381.

- Hardouin C., Guyon X. (2001) Standards adoption dynamics and test for non spatial coordination, *Spatial statistics : methodological aspects and applications*, Lecture Notes in Statistics 159, Springer, 39-56.

- Guyon X., Hardouin C. (1992) The chi-square coding test for nested Markov random field hypotheses, *Stoch. Models, Stat. Methods and Algorithms in Image Analysis*. Lecture Notes in Statistics, 74, Springer, 165–176.

Revue nationale à comité de lecture

- Hardouin C. (2008) Influence d'une contamination initiale sur une dynamique spatiale non itérative, *Journal de la SFDS, Tome 149, no 4, 107-129*.

Travaux d'enseignement

- Carter L., Hardouin C. (2009) Use and misuse of quantitative and graphical information in Statistics. An Approach in Teaching. *Proceedings 10th Int. Conf. Mathematics Education into the 21st Century Project*.

Articles soumis et en préparation

- Guyon X., Hardouin C., Forward recursions and normalizing constant for Gibbs fields.
- Hardouin C., Some differences between sequential or synchronous dynamics for standard adoption.
- Guyon X., Hardouin C., Variogram identification by penalized least squares.

Travaux non publiés ou incomplètement publiés

- Comte F., Hardouin C. (1996) Log-regularized periodogram regression, *Prepub. SAMOS 65*.
- Comte F., Hardouin C. (1995) Regression on log-regularized periodogram for fractional models at low frequencies, *Prepub. SAMOS 46 et CREST 9536*.
- Comte F., Hardouin C. (1995) Regression on log-regularized periodogram under assumption of bounded spectral densities : the non fractional and the fractional cases, *Prepub. SAMOS 47 et CREST 9535*.
- Comte F., Hardouin C. (1995) Regression on log-regularized periodogram : Comparison with Whittle estimator and study of the non Gaussian case, *Prepub. CREST 9539*.
- Hardouin C. (1992) Quelques résultats nouveaux en statistique des processus : contraste fort, régression à résidus à longue portée, estimation par log-périodogramme. *Thèse Université Paris 7*.

A.3 Communications et conférences

Congrès internationaux avec actes

- “Adoption dynamics : sequential or synchronous modelling”, ASMDA 2011, Rome, Italie, juin 2011.
- “Mixed states spatio-temporal auto-models”, METMA V, International workshop on spatio-temporal modelling, St Jacques de Compostelle, juin 2010.
- “Recursions for marginals and normalizing constant for Gibbs processes”, SMTDA 2010, Chania, Crete, juin 2010.
- “Dumping influence on a non iterative dynamics” 12th International Conference on Applied stochastic Models and Data Analysis (ASMDA), Chania, Crete, mai 2007.
- “Multivariate auto-models, Application to mixed states data” 7th International Conference on Operations Research, La Havane, Cuba, Mars 2006.

- “Markovian automodels with mixed states”, International Symposium on Applied Stochastic Models and Data Analysis (ASMDA), Brest, Mai 2005.

Congrès nationaux

- “Modèle spatio-temporel pour des variables mixtes”, 41^{èmes} Journées de Statistique, Bordeaux (2009).
- “Non iterative standards adoption dynamics” Modèles et Apprentissages en Sciences Humaines et Sociales (MASHS), Créteil (2008).
- “Auto modèles markoviens mixtes”, aux 34^{èmes} Journées de Statistique, Bruxelles (2002).
- “Dynamique d’adoption de standards et tests de coordination spatiale”, aux 31^{èmes} Journées de Statistique, Grenoble (1999).
- “Estimation de modèles fractionnaires” aux 27^{èmes} Journées de Statistique, Jouy-en-Josas (1995).
- “Estimation par log-périodogramme. Application aux modèles fractionnaires” aux 26^{èmes} Journées de Statistique, Neuchatel (1994).
- “Régression du log-périodogramme” aux 24^{èmes} Journées de Statistique, Bruxelles (1992).
- “Modèles de contraste fort et statistique asymptotique pour des processus non ergodiques” aux 23^{èmes} Journées de Statistique, Strasbourg (1991).

Exposés lors de workshops, journées thématiques et séminaires

- “Markovian automodels with mixed states”, 5th French-Danish workshop on Spatial Statistics and Image Analysis in Biology. Grenoble (2004).
- “Markov Chain Markov Field dynamic : models and statistics” 2nd French-Danish workshop on Spatial Statistics and Image Analysis in Biology. Sandbjerg, Danemark (1998).
- Présentation de séminaires dans des universités ou des instituts de recherche français (AgroParisTech, Pau, Montpellier, Paris 7, Paris 1, Paris 5, Paris 6).
- Présentation d’articles ou de travaux personnels lors de journées thématiques (groupe spatial AgroParisTech, réseau MSTGA, institut de l’élevage).

A.4 Animation recherche et responsabilités collectives

Séjours

- Invitation à l’Université Ca’Foscari de Venise, par le professeur C. Gaetan, juin 2009.

Organisation de congrès et séminaires

- Organisation d’une session spatiale pour Applied Stochastic Models and Data Analysis (ASMDA 2011), Rome, Italie, juin 2011.
- Organisation locale des journées du réseau MSTGA, à l’Université Paris Ouest, Novembre 2010.
- Organisation d’une session spatiale pour Stochastic Modeling Techniques and Data Analysis International Conference (SMTDA 2010), Chania, Greece, juin 2010.
- Organisation de la session *Statistique spatiale et spatio-temporelle* aux Journées MAS, Rennes, août 2008.
- Organisation d’un colloque international de statistiques spatiales, le 27 avril 2007, à l’Université Paris 1.

<http://samos.univ-paris1.fr/spatialstat/>

- Organisation locale des deuxièmes journées du réseau MSTGA, à l'Université Paris 1, Novembre 2006.

- Participation à l'organisation locale des congrès organisés par le SAMOS (Limit Theorems and Applications Janvier 2008, Stochastic dynamics Juin 2007, WSOM 2005)

Réseaux

- Membre du réseau Modélisation Spatio-Temporelle sur Graphe et Approximation, organisé par le département Mathématiques et Informatique Appliquées de l'INRA depuis mai 2006 (<http://carlit.toulouse.inra.fr/MSTGA/>).

- Correspondante de ce réseau pour le RNSC et ISC-PIF (Réseau National des Systèmes Complexes et Institut des Systèmes Complexes de Paris Ile de France).

- Membre du groupe de travail en statistiques spatiales de l'AgroParistech (anciennement INA PG).

Expertise

- Membre élu du CNU section 26, depuis octobre 2007.

- Membre de comités de sélection des Université Paris Ouest Nanterre (1994-97, 2009), Paris Descartes (1998-2008, 2010), Paris 1 Panthéon Sorbonne (2000-08).

- Rapporteur pour les revues CSDA, C. R. Acad. Sci., Canadian Journal of Stats, IEEE Transactions on Image processing.

Activités administratives

- Membre de la commission d'attribution des services pour les mathématiques à l'Université Paris 10, de 2008 à ce jour (département de mathématiques et informatique).

- Rédaction d'une partie du rapport quadriennal du SAMOS, 2008.

- Rédaction de demandes de financement de projets au titre du BQR, 2006 et 2007.

- Membre du conseil de laboratoire de l'UMR MATISSE, représentant le centre de recherche SAMOS de 1999 à 2003.

- Responsable de la bibliothèque du SAMOS (devis, achats, inventaire) de 2000 à 2009 (environ 650 livres en 2009).

A.5 Enseignement et encadrement

Enseignements de licence

MASS	- TD de Mathématiques générales en 1L1 (algèbre, analyse). - CM et TD de Statistiques en L2 et L3.
MIAGE	- CM et TD de Statistiques en L3.
Sciences Economiques	- CM et TD de mathématiques en L1 (analyse, algèbre). - CM et TD de probabilités en L2.
Psychologie	- CM et TD de Statistiques en L1 (statistique descriptive). - Cours-TD de Statistique inférentielle en L3.
Ethnologie	- CM de Statistiques en L3.
AES	- Cours-TD de Statistiques en L1 et L2.

Enseignements de master

MASS	- TD de statistique mathématique en maîtrise. - TD de processus en maîtrise.
MIAGE	- CM et TD d'Analyse de données en M1 Utilisation d'Excel et de SAS.
Sciences Economiques	- CM de Statistiques en M1 Economie et Société - TD de préparation au Capes SES

Responsabilités pédagogiques

- Responsable de l'enseignement des statistiques en première année de Licence de Psychologie (environ 800 à 1000 étudiants) de 2003/2004 à 2008/2009.

- Préparation des étudiants à l'épreuve orale de mathématiques du CAPES de Sciences Economiques et Sociales. Supervision d'un groupe de discussion sur internet pour les étudiants et les enseignants.

Encadrement

- Tutorat d'étudiants en apprentissage dans la filière Miage depuis 2005-2006.

- Encadrement de mémoires d'étudiants en master 1 Miage.

Autres cours ou activités

- Correction de l'épreuve de "maths 2" du concours d'entrée à l'ESSEC depuis 1991.