

STATISTIQUE SPATIALE

Xavier Guyon
SAMOS - Université Paris 1

Février 2007
Conférence S.A.D.A.' 07 - Cotonou -Benin
(Statistique Appliquée pour le Développement en Afrique)

Table des matières

1 Géostatistique	1
1.1 Quelques rappels	1
1.2 Processus stationnaire, covariance	2
1.2.1 Définitions, exemples	2
1.2.2 Représentation spectrale d'une covariance	3
1.3 Processus intrinsèque, variogramme	3
1.3.1 Exemples de variogramme	4
1.3.2 Propriétés géométriques : continuité, différentiabilité	6
1.4 Prédiction à covariance connue : le krigeage	7
1.4.1 Le krigeage simple	7
1.4.2 Le krigeage universel	8
1.5 Estimation en géostatistique	9
1.5.1 Analyse variographique	9
1.5.2 Estimation empirique du variogramme	10
1.5.3 Validation d'un modèle de variogramme	12
1.6 Exercices	14
2 Auto-régression et corrélation spatiales	17
2.1 Auto-régression stationnaire	17
2.1.1 Moyenne glissante (<i>MA</i>) et <i>ARMA</i>	17
2.1.2 Auto-régression simultanées (<i>SAR</i>)	18
2.1.3 Auto-régression conditionnelle (<i>CAR</i>)	18
2.2 Modèle non-stationnaire sur S fini	20
2.2.1 Modèle <i>SAR</i>	20
2.2.2 Représentation markovienne <i>CAR</i>	21
2.2.3 Champ gaussien markovien.	21
2.2.4 <i>AR</i> en présence d'exogènes : modèle <i>SARX</i>	22
2.3 Auto-corrélation spatiale latticielle	23
2.3.1 L'indice de Moran	23
2.3.2 Test asymptotique d'indépendance spatiale	24
2.3.3 L'indice de Geary	24
2.3.4 Test de permutation d'indépendance spatiale	25
2.4 Statistique d'un champ du second ordre	27
2.4.1 Estimation d'un modèle stationnaire sur \mathbb{Z}^d	27
2.4.2 Estimation d'un modèle auto-régressif	29
2.4.3 Maximum par vraisemblance pour un champ gaussien centré	30
2.4.4 Estimation d'une régression spatiale.	30
2.5 Exercices	34
3 Champ de Markov sur réseau	37
3.1 Recollement de lois conditionnelles	37
3.2 Champ de Gibbs sur S	38
3.2.1 Potentiel d'interaction et spécification de Gibbs	38
3.2.2 Exemples de spécification de Gibbs	39
3.3 Champ de Markov et champ de Gibbs	41
3.3.1 Cliques d'un graphe, champ de Markov	41
3.3.2 Le théorème de Hammersley-Clifford	41
3.4 Les auto-modèles de Besag	41
3.4.1 Recollement de lois conditionnelles et auto-modèle	41
3.4.2 Exemples d'auto-modèles	42

3.5	Estimation d'un champ de Markov	44
3.5.1	Le maximum de vraisemblance	44
3.5.2	Pseudo-vraisemblance conditionnelle (PVC) de Besag	45
3.5.3	La méthode de codage	47
3.5.4	Précisions comparées du <i>MV</i> , <i>MPVC</i> et du <i>Codage</i>	49
3.6	Exercices	50
4	Processus ponctuels spatiaux	53
4.1	Définitions et notations	53
4.1.1	Espace exponentiel des configurations	53
4.1.2	Loi d'un processus ponctuel.	53
4.2	Exemples de processus ponctuels	54
4.2.1	Processus Binomial.	54
4.2.2	Processus Ponctuel de Poisson (PPP)	55
4.3	Densité d'un processus ponctuel	57
4.3.1	Définitions	57
4.3.2	Exemples	58
4.4	Caractéristiques du second ordre d'un PP	60
4.4.1	Moment et intensité d'ordre p	60
4.4.2	Mesure de Palm.	60
4.4.3	Deux distances au plus proche voisin.	61
4.4.4	Moment réduit d'ordre 2 de Ripley	61
4.5	Processus ponctuel de Markov	62
4.6	Statistique des processus ponctuels spatiaux	63
4.6.1	Estimation de l'intensité d'un PP	63
4.6.2	Estimation de caractéristiques au second ordre	63
4.6.3	Estimation par MC d'un modèle paramétrique	65
4.6.4	Test de Monte Carlo	65
4.6.5	Pseudo-vraisemblance d'un <i>PP</i>	67
4.7	Calcul de Monte Carlo d'une vraisemblance	68
4.8	Exercices	69
5	Simulation des modèles spatiaux	73
5.1	Chaînes de Markov	73
5.1.1	Construction d'un algorithme MCMC	74
5.2	Les deux algorithmes MCMC de base	74
5.2.1	Echantillonneur de Gibbs sur un espace produit	74
5.2.2	L'algorithme de Metropolis-Hastings (MH)	75
5.3	Simulation d'un champ de Markov sur un réseau	76
5.3.1	Les deux algorithmes de base	76
5.3.2	Exemples	76
5.4	Simulation d'un processus ponctuel	77
5.4.1	Simulation conditionnelle	77
5.4.2	Simulation inconditionnelle	78
5.5	Contrôle de la convergence des méthodes MCMC	78
5.6	Simulation d'un champ gaussien sur $S \subseteq \mathbb{R}^d$	80
5.6.1	Simulation d'un champ stationnaire	80
5.6.2	Simulation conditionnelle	83
5.7	Exercices	83

Introduction

La statistique spatiale étudie des phénomènes observés sur un ensemble spatial S . Les domaines d’application sont nombreux : sciences de l’environnement et de la terre, épidémiologie, agronomie, météorologie, économétrie, traitement d’image, etc. Cette diversité fait la richesse de ce domaine dont l’originalité, à la différence des séries temporelles, est d’utiliser des modélisations non-causales. Nous présentons ici les bases théoriques sur les modèles spatiaux et les méthodes statistiques associées.

Considérons un phénomène $X = (X_s, s \in S)$ où X est variable aléatoire indexée par un *ensemble spatial* S . La localisation d’un *site* $s \in S$ est en général géographique : station de mesure du cumul de précipitations sur un réseau météorologique d’une région $S \subset \mathbb{R}^2$, sites d’arbres fruitiers sur un verger $S \subset \mathbb{Z}^2$ et production, cantons et nombre de personnes affectées par une maladie sur le canton, points de carottage pour la mesure de densité en minerai de fer d’une zone de prospection $S \subset \mathbb{R}^3$, répartition d’arbres dans une forêt $S \subset \mathbb{R}^2$, lattice régulier des pixels $S = \{1, 2, \dots, n\}^2$ d’une image en niveaux de gris, etc.

Comme pour les séries temporelles, la statistique spatiale se différencie de la statistique classique par le fait que *les observations sont dépendantes*. Mais son originalité est de faire appel à des *modélisations non-causales*, l’espace S n’étant pas muni d’une relation d’ordre naturel et le “germe” d’un modèle spatial étant en général non-causal, s’exprimant en terme du voisinage de s “dans toutes les directions”. *En ce sens, la statistique spatiale, non-causale, se distingue des séries temporelles dont les modèles sont fondés sur la notion de passé.*

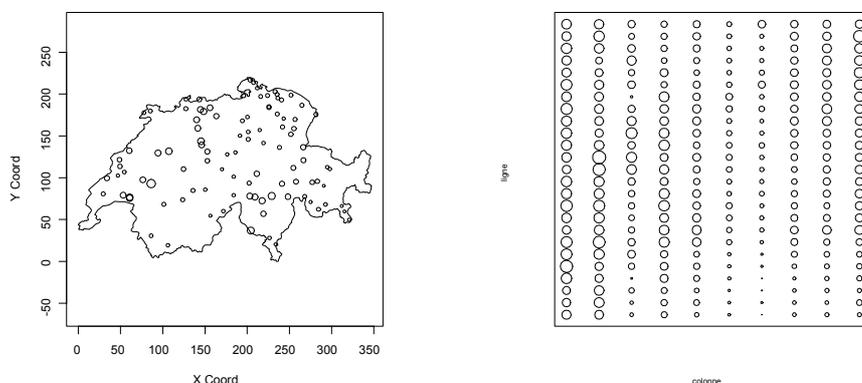


FIG. 1 – (a) Cumuls pluviométriques sur le réseau météorologique suisse le 8 mai 1986 (passage du nuage de Tchernobyl, données `sic.100` du package `geoR` du logiciel R [39]) ; (b) Porosité d’un sol (données `soil` du package `geoR`). Pour (a) et (b), les dimensions des symboles sont proportionnelles à la valeur X_s .

On distingue trois types de données spatiales suivant la nature de S : cette typologie structure ce document.

Les données géostatistiques (cf. Ch.1 et Fig.1) : S est un sous-espace continu de \mathbb{R}^2 (ou \mathbb{R}^d pour $d \geq 2$), le champ $X = \{X_s, s \in S\}$ est observé en n sites fixés $\{s_1, \dots, s_n\} \subset S$ et $X_s \in \mathbb{R}$ est à valeurs réelles. C’est le cas des données pluviométriques de la figure 1-a (réseau S irrégulier), ou encore des données de porosité d’un sol (cf. Fig. 1-b ; réseau S régulier). La géostatistique aborde l’identification et l’estimation du modèle, la séparation des variations à grande et à petite échelles, la construction de cartes de prédiction (ou krigeage) sur S .

Les données latticielles sur un réseau S discret (Ch.2 et 3) : $S \subset \mathbb{R}^2$ est *discret fixé*, X est observé sur S . Les sites s représentent en général des unités géographiques d’un réseau structuré à partir d’un graphe de voisinage \mathcal{G} (cf. Fig. 2-b) et X_s est une donnée “intégrée” sur la région s . Nous distinguerons : (i) les modèles *SAR* d’*Auto-Régressions Simultanées* réelles (Ch.2) et (ii) les *champs de Markov* à valeurs dans un espace d’état E général (Ch.3). Si en analyse d’image le lattice S est régulier, ce n’est pas le cas en économétrie ou épidémiologie spatiale (cf. Fig. 2-a et b). Les objectifs étudiés sont l’étude de la corrélation spatiale, la modélisation *SARX* à l’aide de covariables et l’estimation d’une régression spatiale.

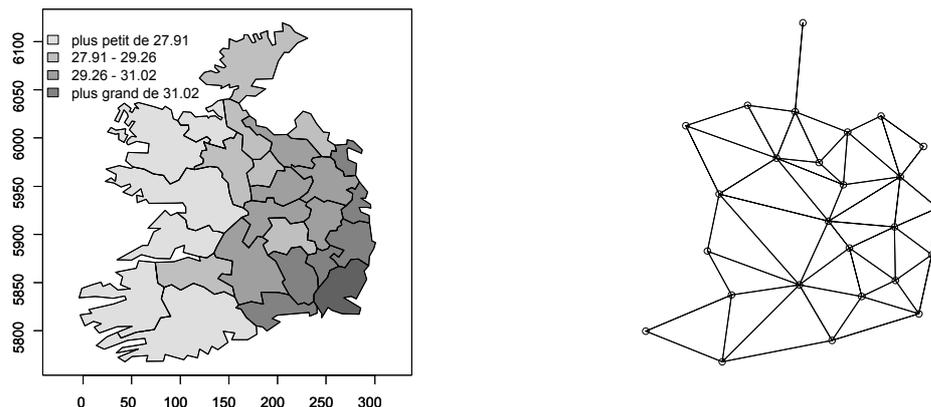


FIG. 2 – (a) Pourcentage d’individus du groupe sanguin A dans les 26 comtés de l’Irlande (données `eire` du package `spdep`); (b) graphe de voisinage entre les comtés.

Les processus ponctuels spatiaux (Ch. 4) : ici, c’est l’ensemble \mathbf{x} des sites où ont lieu les observations qui est la réalisation aléatoire d’un *processus ponctuel* (PP) X observé dans une fenêtre $S \subset \mathbb{R}^2$: $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, $x_i \in S$. Le nombre $n = n(\mathbf{x})$ de réalisations ponctuelles observées et leurs positions sont aléatoires. Le processus ponctuel X est *marqué* (PPM) si de plus on observe une *marque* $m(x_i)$ attachée à chaque x_i (par exemple le diamètre $m(x)$ de l’arbre observé en x). La figure 3-a donne la répartition des “centres” de cellules d’une section histologique observée au microscope; la figure 3-b donne la répartition d’aiguilles de pin dans une forêt ainsi que leur longueur. Une question centrale est de savoir si la répartition \mathbf{x} est plutôt régulière (comme à la figure 3-a) ou due au hasard (PP de Poisson) ou encore si elle présente des agrégats (PP de Neyman-Scott).

La distinction entre ces trois types de données n’est pas toujours nette. Par exemple, des données géostatistiques (des données ponctuelles) sur S conduisent à des données sur un réseau par intégration (comptage) sur les composantes d’une partition de S . A l’inverse, des données latticielles sur un réseau à maille très fine peuvent être considérées comme des données géostatistiques; si ces données sont binaires $\{0, 1\}$ avec une très faible probabilité de l’état 1, elles peuvent être interprétées comme des données ponctuelles.

Le dernier chapitre (Ch. 5) présente les deux méthodes de simulation MCMC (*Monte Carlo Markov Chain*) des modèles spatiaux : l’échantillonneur de Gibbs et l’algorithme de Métropolis. Ces méthodes s’avèrent nécessaire car on verra que les méthodes classiques sont inopérantes pour un modèle spatial. Les méthodes de Monte Carlo permettent de construire des procédures statistiques lorsque la loi d’une statistique n’est pas accessible et sont utiles pour valider un modèle.

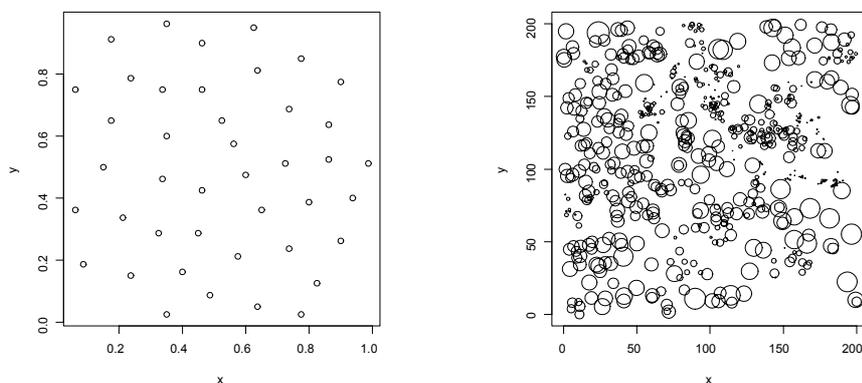


FIG. 3 – (a) Les 42 centres de cellules d’une section histologique observée au microscope (données `cells` du package `spatstat`); (b) Positions et tailles de 584 aiguilles de pin dans un sous-bois (données `longleaf` du package `spatstat`).

Pour en savoir davantage sur la statistique spatiale, on pourra se reporter à l'aide en ligne du logiciel R, bien documentée et régulièrement mise à jour, particulièrement à ses *packages spatiaux* `geoR`, `gstat`, `RandomFields`, `spatial`, `spatstat`, `spdep`. R est disponible gratuitement sur site internet du Comprehensive R Archive Network (CRAN) : <http://cran.r-project.org/>. Plusieurs manuels sont distribués avec R (cf. "R pour les débutants" d'Emmanuel Paradis) ainsi que d'autres projets de développement de programmes (cf. <http://sal.uiuc.edu/csiss/R>)

Voici quelques ouvrages de références sur le sujet :

- N. CRESSIE [1993], *Statistics for Spatial Data*; Wiley.
J.J. DROESBEKE, M. LEJEUNE & G. SAPORTA (EDITEURS) [2006], *Analyse statistique des données spatiales*; Eds. Technip.
P. DIGGLE [2003], *Statistical Analysis of Spatial Point Patterns*; Oxford University Press.
P. DIGGLE P. & P.J. RIBEIRO [2006], *Model-based Geostatistics*, Springer Series in Statistics.
X. GUYON [1995], *Random fields on a network : modeling, statistics and applications*; Springer.
MØLLER, J. & WAAGEPETERSEN, R.P. [2004] *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall.
B. D. RIPLEY [1981], *Spatial Statistics*, Wiley.
O. SCHABENBERGER ET C.A. GOTWAY [2005], *Statistical methods for spatial data analysis*, Chapman et Hall.
H. WACKERNAGEL [1998], *Multivariate Geostatistics : An Introduction with Applications*. Springer-Verlag.

Ces notes ont été préparées alors que nous achevons avec Carlo Gaetan, professeur à l'Université Ca' Foscari de Venise, la préparation d'un livre intitulé : "Modélisation et Statistique Spatiales". La rédaction de ce cours a évidemment beaucoup profité de toute l'expérience "spatiale" de Carlo : je tiens ici à l'en remercier très vivement.

Paris, Février 2007

Chapitre 1

Géostatistique

Soit S un ensemble spatial, $S \subset \mathbb{R}^2$ ou $S \subset \mathbb{R}^3$. Un champ aléatoire X sur S et à valeur dans E est la donnée d'une collection $X = \{X_s, s \in S\}$ de variables aléatoires (v.a.) indexées par S et à valeurs dans E . Ce chapitre est consacré à l'étude des champs du second ordre, c'est-à-dire des champs à *valeurs réelles* ($E = \mathbb{R}$) de *variance finie* en tout site $s \in S$. On étudiera également la classe plus générale des champs à accroissements stationnaires au second ordre ou modèle intrinsèque.

L'approche *géostatistique* modélise X par sa covariance (modèle au second ordre) ou par son variogramme (modèle intrinsèque). Développée initialement pour la modélisation et la prévision des réserves minières, la géostatistique est utilisée aujourd'hui dans des domaines très variés tels que les sciences de la terre, l'environnement, l'épidémiologie ([51];[12];[20]). Un objectif central est la prévision, ou krigeage, de X_s aux sites non observés.

1.1 Quelques rappels

Un processus aléatoire réel $X = \{X_s, s \in S\}$ est une fonction mesurable $X : S \rightarrow \mathbb{R}$ à valeur réelle. La loi du processus est caractérisée par ses lois jointes finies dimensionnelles $(X_{s_1}, X_{s_2}, \dots, X_{s_n})$ pour toute partie finie $\{s_1, s_2, \dots, s_n\}$ de S .

Définition 1.1 X est un processus au second ordre (on note $X \in L^2$) si pour tout $s \in S$, $E(X_s^2) < \infty$. La moyenne est la fonction $m : S \rightarrow \mathbb{R}$ définie par $m(s) = E(X_s)$, la covariance la fonction $c : S \times S \rightarrow \mathbb{R}$ définie par $c(s, t) = Cov(X_s, X_t)$.

X est centré si $m(s) = 0$ en tout s . La propriété caractéristique d'une covariance est d'être *semi-définie positive* (s.d.p.) :

$$\forall m \geq 1, a \in \mathbb{R}^m \text{ et } \{s_1, s_2, \dots, s_m\} \subseteq S : \sum_{i,j=1,m} a_i a_j c(s_i, s_j) \geq 0.$$

Cette positivité résulte de l'égalité :

$$Var\left(\sum_{i=1,m} a_i X_{s_i}\right) = \sum_{i,j=1,m} a_i a_j c(s_i, s_j).$$

La covariance est *définie positive* (d.p.) si de plus $\sum_{i,j=1,m} a_i a_j c(s_i, s_j) > 0$ pour $a \neq 0$.

Les processus gaussiens constituent une sous-classe importante des processus de L^2 .

Définition 1.2 X est un processus gaussien sur S si, pour toute partie finie $\Lambda \subset S$ et toute suite réelle $a = (a_s, s \in \Lambda)$, $\sum_{s \in \Lambda} a_s X_s$ est une variable gaussienne.

Si la covariance Σ_Λ de $X_\Lambda = (X_s, s \in \Lambda)$ est inversible, X_Λ admet pour densité :

$$f_\Lambda(x_\Lambda) = (2\pi)^{-\#(\Lambda)/2} (\det \Sigma_\Lambda)^{-1/2} \exp\left\{-\frac{1}{2} {}^t(x_\Lambda - m_\Lambda) \Gamma_\Lambda^{-1} (x_\Lambda - m_\Lambda)\right\}$$

Ici, $\#(U)$ dénote le cardinal de U et $m_\Lambda = E(X_\Lambda)$. A toute moyenne m et covariance c d.p. est associé un unique champ gaussien de moyenne m et de covariance c .

Exemple 1.1 *Mouvement brownien sur \mathbb{R}^+ ; convolution discrète et continue.*

• X est un *mouvement brownien* sur $S = \mathbb{R}^+$: $X_0 = 0$, $X_s \sim \mathcal{N}(0, s)$, $s > 0$ et les accroissements $X(\cdot, t] = X_t - X_s$, $t > s \geq 0$ sont indépendants pour des intervalles disjoints. La covariance de X vaut : $c(s, t) = \min\{s, t\}$.

•• $S = \mathbb{Z}^2$, $c = (c(s, t), (s, t) \in S)$ est une suite réelle ne comportant qu'un nombre fini de termes non-nuls. Si ε est une suite de variables de L^2 , la convolution discrète suivante définit un processus de L^2 dit en moyenne mobile :

$$X_{s,t} = \sum_S c(s-u, t-v) \varepsilon_{u,v}$$

••• La version continue est la suivante : soit W un processus de Wiener sur \mathbb{R}^2 (les accroissements $W(du, dv)$ sont gaussiens, centrés, indépendants, de variance $du \times dv$). Si $c : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une fonction de carré intégrable ($\sigma^2 = \int_{\mathbb{R}^2} c^2(u, v) dudv < \infty$), l'intégrale suivante définit un processus gaussien (stationnaire) de variance σ^2 :

$$X_{s,t} = \int_{\mathbb{R}^2} g(s-u, t-v) W(du, dv)$$

1.2 Processus stationnaire, covariance

Soit X un champ sur $S \subset \mathbb{R}^d$ de moyenne m et de covariance c où S est un *sous-groupe additif* de \mathbb{R}^d (par exemple $S = \mathbb{R}^d$ ou $S = \mathbb{Z}^d$).

1.2.1 Définitions, exemples

Définition 1.3 X est un champ stationnaire au second ordre sur S si X est de moyenne constante et de covariance invariante par translation :

$$\forall s, t \in S : E(X_s) = m \quad \text{et} \quad c(s, t) = \text{Cov}(X_s, X_t) = C(t - s).$$

où $C(\cdot)$ est la fonction de covariance stationnaire. L'invariance par translation se traduit par : $\forall s, t, h \in S : \text{Cov}(X_{s+h}, X_{t+h}) = C(s - t)$. La fonction de corrélation $\rho(\cdot)$ est $h \mapsto \rho(h) = C(h)/C(0)$.

Proposition 1.1 (1) $\forall h \in S$, $|C(h)| \leq C(0) = \text{Var}(X_s) = \sigma_X^2$.

(2) $\forall n \geq 1$, $a \in \mathbb{R}^n$ et $\{t_1, t_2, \dots, t_m\} \subseteq S : \sum_{i,j=1,m} a_i a_j C(t_i - t_j) \geq 0$.

(3) Si C_1, C_2 sont des covariances stationnaires, les fonctions suivantes le sont aussi :

(i) $C(h) = a_1 C_1(h) + a_2 C_2(h)$, a_1 et $a_2 \geq 0$;

(ii) $C(h) = C_1(ah)$, $a > 0$; (iii) $C(h) = C_1(h)C_2(h)$.

Deux notions encadrent la notion de stationnarité dans L^2 : l'une, plus faible, étudiée au §1.3, est celle de processus à accroissements stationnaires ou *modèle intrinsèque* ; l'autre, plus forte, est la stationnarité au sens strict : X est *strictement stationnaire* si pour tout $k \in \mathbb{N}$, $(t_1, t_2, \dots, t_k) \in S^k$ et $\forall h \in S$, la loi de $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ ne dépend pas de h . Si X est stationnaire au sens strict et si $X \in L^2$, alors X est stationnaire de L^2 . L'inverse n'est pas vrai en général, mais les deux notions coïncident si X est gaussien.

Exemple 1.2 *Bruit Blanc Fort (BBF), BB faible (BBf), BB gaussien (BBG)*

X est un *Bruit Blanc Fort* si les v.a. $\{X_s, s \in S\}$ sont centrées, indépendantes et identiquement distribuées (i.i.d.). X est un *Bruit Blanc faible* si les variables $\{X_s, s \in S\}$ sont centrées, $\text{Var}(X_s) = \sigma^2 < \infty$, décorrélées : $\forall s \neq t : \text{Cov}(X_s, X_t) = 0$. Un BBF sur S est strictement stationnaire ; un BBf sur S est stationnaire de L^2 . X est un *BB Gaussien* si les v.a. $\{X_s, s \in S\}$ sont i.i.d. gaussiennes centrées.

Le champ X est *isotropique* si la covariance entre X_s et X_t ne dépend que de la distance euclidienne $\|s - t\|$ entre s et t : un champ isotropique est stationnaire ; de plus :

$$\text{pour tout } t, s \in S : C(s - t) = r(\|s - t\|).$$

avec r la fonction de covariance isotropique.

L'isotropie impose des restrictions sur la covariance si $d \geq 2$. Par exemple, si X est centré isotropique sur \mathbb{R}^d , et si on considère $(d + 1)$ points à distances mutuelles h ,

$$E\left\{ \sum_{i=1, d+1} X_{s_i}^2 \right\} = (d + 1)[r(0) + dr(h)].$$

On en déduit pour la corrélation : pour tout h , $\rho(h) \geq -1/d$, $\rho(h) \geq -0.5$ si $d = 2$. Ripley [43] établit des minoration plus fines : $\rho(h) \geq -0.403$ sur \mathbb{R}^2 et $\rho(h) \geq -0.218$ sur \mathbb{R}^3 .

1.2.2 Représentation spectrale d'une covariance

La théorie de Fourier et le théorème de Bochner mettent en bijection une covariance stationnaire C avec sa mesure spectrale F : il est donc équivalent de caractériser un modèle stationnaire de L^2 par C ou par F ([55]; [12]).

- Si $S = \mathbb{R}^d$, on associe à C la mesure spectrale F sur les boréliens de \mathbb{R}^d , ≥ 0 , bornée, vérifiant :

$$C(h) = \int_{\mathbb{R}^d} e^{i\langle u, h \rangle} F(du), \text{ où } \langle u, h \rangle = \sum_{i=1}^d u_i h_i \quad (1.1)$$

Si F admet une densité f , f est appelée la *densité spectrale* de X ; f est paire de carré intégrable.

Si X est isotropique de covariance $C(h) = C_0(\|h\|_2)$, C_0 admet la représentation,

$$C_0(t) = \int_{]0, \infty[} Y_d(ut) \Phi(du) \text{ où } Y_d(v) = \left(\frac{2}{v}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) \mathcal{J}_{(d-2)/2}(v), \quad (1.2)$$

$\mathcal{J}_\kappa(\cdot)$ étant la fonction de Bessel de première espèce d'ordre κ et μ une mesure positive bornée : ce résultat dit que la variété des covariances isotropiques r sur \mathbb{R}^d est la même que celle des mesures positives et bornées μ sur $]0, +\infty[$.

Cette identification permet de montrer que la fonction

$$C(h) = C_0(\|h\|) \text{ où } C_0(t) = b \exp(-a|t|), \quad a, b > 0,$$

est une covariance sur \mathbb{R}^d (dite *exponentielle*) pour toute dimension d .

- Si $S = \mathbb{Z}^d$, on associe à C une mesure spectrale F sur les boréliens de \mathbb{T}^d ($\mathbb{T} = [0, 2\pi[$), ≥ 0 , bornée, vérifiant :

$$C(h) = \int_{\mathbb{T}^d} e^{i\langle u, h \rangle} F(du) \quad (1.3)$$

F est à densité f si $\sum_{h \in \mathbb{Z}^d} C(h)^2 < \infty$: alors, $f \in L^2(\mathbb{T}^d)$ et le théorème d'inversion de Fourier donne :

$$f(u) = (2\pi)^{-d} \sum_{h \in \mathbb{Z}^d} C(h) e^{-i\langle u, h \rangle} \quad (1.4)$$

Si $\sum_{h \in \mathbb{Z}^d} |C(h)| < \infty$, la convergence est uniforme et f est continue. Plus la différentiabilité de f est grande, plus la convergence de C vers 0 à l'infini est rapide : si $f \in \mathcal{C}^k(\mathbb{T}^d)$ pour un multi-indice $k = (k_1, \dots, k_d) \in \mathbb{N}^d$, $\limsup_{h \rightarrow \infty} h^k |C(h)| < \infty$ où $h = (h_1, h_2, \dots, h_d) \rightarrow \infty$ signifie qu'au moins un $h_i \rightarrow \infty$ et $h^k = h_1^{k_1} \times \dots \times h_d^{k_d}$. En particulier, si f est infiniment dérivable, $C \rightarrow 0$ exponentiellement vite : c'est le cas des modèles *ARMA* (cf. §2.1.1) dont le spectre f est rationnel.

1.3 Processus intrinsèque, variogramme

La stationnarité dans L^2 n'est pas toujours satisfaite : par exemple si $X_s = Y_s + Z$ où Y est stationnaire de L^2 mais $Z \notin L^2$; ou encore si $X \in L^2$ n'est pas stationnaire (i.e. le mouvement brownien). Une façon d'affaiblir l'hypothèse de stationnarité L^2 est de considérer le processus :

$$I^{(h)} = \{I_s^{(h)} = X_{s+h} - X_s, s \in S\},$$

des h -accroissements de X , à h fixé, ces accroissements pouvant être stationnaires en s dans L^2 sans que X le soit ou que X ait une variance finie.

Définition 1.4 X est un processus intrinsèque si X est de moyenne constante et si pour tout $h \in S$, le processus $I^{(h)} = \{I_s^{(h)} = X_s - X_{s+h} : s \in S\}$ est stationnaire au second ordre. Le variogramme de X est alors défini par :

$$2\gamma(h) = \text{Var}(X_{s+h} - X_s)^2$$

$\gamma(\cdot)$ est le *semi-variogramme* de X .

Par exemple, le mouvement brownien sur \mathbb{R} est intrinsèque, non-stationnaire, de variogramme $|h|$ non-borné.

Remarquons que tout processus stationnaire au second ordre est intrinsèque. On a :

Proposition 1.2 1. $2\gamma(h) = 2\gamma(-h)$, $2\gamma(h) \geq 0$ et $2\gamma(0) = 0$.

2. Un variogramme est conditionnellement défini négatif (c.d.n.) :

$$\forall a \in \mathbb{R}^n \text{ t.q. } \sum_{i=1,n} a_i = 0 \text{ et } \{s_1, \dots, s_n\} \subseteq S : \sum_{i,j=1,n} a_i a_j \gamma(s_i - s_j) \leq 0.$$

3. Si γ est borné au voisinage de 0, $\exists a$ et $b \geq 0$ t.q., pour tout x , $2\gamma(x) \leq a\|x\|^2 + b$.

Un variogramme $2\gamma(\cdot)$ n'est pas nécessairement borné (considérer $|h|$ pour le mouvement brownien) mais croît au plus comme $\|h\|^2$. Un exemple de croissance quadratique $\gamma(t) = \sigma_1^2 t^2$ est celle du variogramme de $X_t = Z_0 + tZ_1$, $t \in \mathbb{R}$, où Z_0 et Z_1 sont centrées, indépendantes, $0 < \sigma_1^2 = \text{Var}(Z_1) < \infty$.

Variogramme d'un processus stationnaire.

Si X est stationnaire de covariance C , X est intrinsèque de variogramme

$$2\gamma(h) = 2(C(0) - C(h)) \quad (1.5)$$

Il suffit de considérer le mouvement brownien pour voir que la réciproque est fautive. Cependant, si le variogramme de X est borné, alors $X_t = Z_t + Y$ pour un processus Z stationnaire de L^2 (Matheron, [34]).

Si $C(h) \rightarrow 0$ lorsque $\|h\| \rightarrow \infty$, le semi-variogramme $\gamma(h)$ présentera un *palier* au niveau $C(0) = \text{Var}(X)$ pour $h \rightarrow \infty$. La *portée* du variogramme (resp. la *portée pratique*) est la distance à partir de laquelle le variogramme atteint son palier (resp. 95% de la valeur du palier), cf. Fig. 1.1.

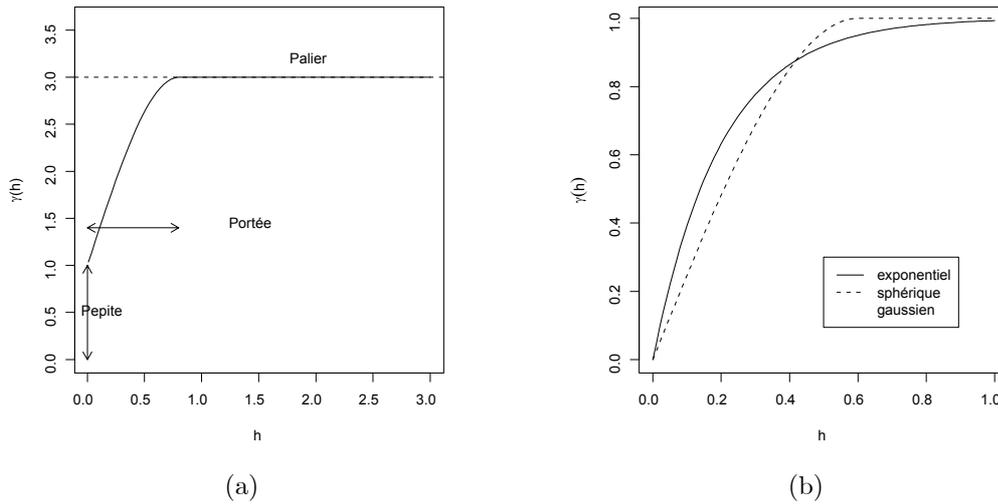


FIG. 1.1 – (a) Semi-variogramme d'un modèle intrinsèque avec composante pépitique; (b) modèles de variogrammes de même portée et de même palier.

1.3.1 Exemples de variogramme

Voici quelques exemples de variogrammes isotropiques sur \mathbb{R}^d , le paramètre $a > 0$ étant lié à la portée du variogramme ([51]; [12]). Les 5 premiers sont bornés, associés à une covariance $C(h) = \sigma^2 - \gamma(h)$, où σ^2 est la variance stationnaire associée.

- **Pépitique** : $\gamma(h; \sigma^2) = \sigma^2$ si $h > 0$, $\gamma(0) = 0$, associé à un BBf.
- **Exponentiel** : $\gamma(h; a, \sigma^2) = \sigma^2 \{1 - \exp(-\|h\|/a)\}$.
- **Sphérique** ($d \leq 3$) : $\gamma(h; a, \sigma^2) = \sigma^2 \{1.5\|h\|/a - 0.5(\|h\|/a)^3\}$ si $\|h\| \leq a$, $= \sigma^2$ sinon.
- **Gaussien** : $\gamma(h; a, \sigma^2) = \sigma^2(1 - \exp(-(\|h\|/a)^2))$.
- **Matérn** : $\gamma(h; a, \sigma^2, \nu) = \sigma^2 \{1 - 2^{1-\nu}(\|h\|/a)^\nu \mathcal{K}_\nu(\|h\|/a) / \Gamma(\nu)\}$. $\mathcal{K}_\nu(\cdot)$ est la fonction de Bessel modifiée de 2-ème espèce de paramètre $\nu > 0$ ([55]; [46]).
- **Puissance** : $\gamma(h; b, c) = b\|h\|^\alpha$, $\alpha \leq 2$.

L'interprétation du modèle sphérique est la suivante : le volume de l'intersection de deux sphères de \mathbb{R}^3 de diamètre a et de centres distants de d est $|S_a| \{1 - 1.5d/a + 0.5(d/a)^3\}$ si $d \leq a$ et 0 sinon, $|S_a|$ étant le volume de la sphère de rayon a : un processus réalisant la covariance sphérique est $X_s = N(S_a(s))$ comptant le nombre de points d'un processus ponctuel de Poisson homogène (cf. §4.2.2) de densité b dans la boule $S_a(s)$ centrée en s et de rayon a .

L'intérêt du modèle de Matérn réside dans son paramètre ν qui contrôle la régularité du variogramme en 0 ($\gamma(h) \sim h^{2\nu}$ à l'origine si $\nu \notin \mathbb{N}$, et $\gamma(h) \sim h^{2\nu} \log h$ sinon) ainsi que la régularité L^2 du processus (cf. §1.3.2 et Fig. 1.2) : $\nu = 1/2$ redonne le variogramme exponentiel, continu mais non dérivable en 0, $\nu = \infty$ le variogramme gaussien indéfiniment dérivable ; si $\nu > m$ le champ de Matérn est m fois dérivable.

Le variogramme *Puissance* est associé à un phénomène spatial *autosimilaire*, γ invariant par changement d'échelle puisque $\gamma(ch) = bc^\alpha \gamma(h)$ pour tout h et tout c .

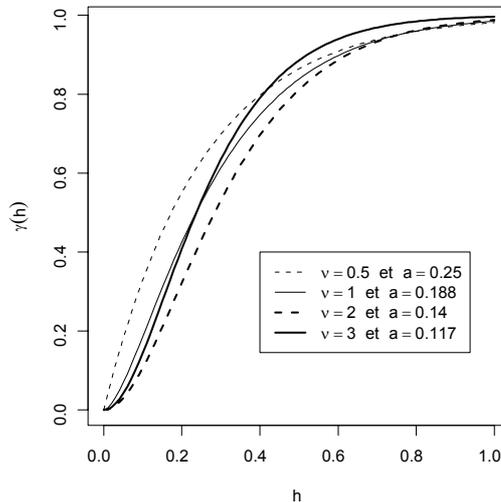


FIG. 1.2 – Semi-variogrammes de Matérn de même portée pour différents ν .

Chacun des modèles précédents peut être étendu par combinaison linéaire positive, en particulier en ajoutant à tout variogramme un effet de pépite.

Anisotropies. Si \vec{e} est une direction de \mathbb{R}^d , $\|\vec{e}\| = 1$, le variogramme directionnel d'un champ intrinsèque dans la direction \vec{e} est défini par

$$2\gamma(h) = \text{Var}(X_{s+h\vec{e}} - X_s) \quad \text{pour } h \in \mathbb{R}.$$

Il y a *anisotropie* du variogramme si deux variogrammes directionnels au moins diffèrent.

On distingue deux types d'anisotropie : l'anisotropie géométrique est associée à la déformation linéaire d'un modèle isotropique ; l'anisotropie de support correspond à une stratification du variogramme sur plusieurs sous-espaces de \mathbb{R}^d ([12] ; [45]).

• Un variogramme γ sur \mathbb{R}^d présente une *anisotropie géométrique* si γ est la déformation A -linéaire d'un variogramme isotropique γ_0 , c'est-à-dire si :

$$\gamma(h) = \gamma_0(\|Ah\|) = \gamma_0(\sqrt{{}^t h Q h})$$

où $Q = {}^t A A$. Un tel variogramme garde les mêmes niveaux de paliers dans toutes les directions (cf. Fig. 1.3-a) mais les portées diffèrent selon les directions : une représentation en "rose des vents" directionnelle permet de visualiser les différentes portées directionnelles.

En notant \tilde{h}_k les coordonnées de h dans la base propre orthonormée de Q associée aux valeurs propres $(\lambda_k, k = 1, \dots, d)$, $\gamma(h) = \gamma_0(\sum_{k=1}^d \lambda_k \tilde{h}_k^2)$. Par exemple, si A est la rotation autour de l'origine de \mathbb{R}^2 d'angle ϕ suivie de l'homothétie de rapport $0 \leq e \leq 1$ sur le nouvel axe des y , les portées décrivent une "rose des vents" elliptique d'excentricité e dans la nouvelle base. La figure 1.3-a illustre cette anisotropie dans \mathbb{R}^2 si γ_0 est un modèle sphérique de paramètres $a = 30$ et $\sigma^2 = 5$ et la déformation A est de paramètres $\phi = 45^\circ$ et $e = 0.3$.

•• On parle d'*anisotropie de support* si le variogramme $h \rightarrow \gamma(h)$, après un éventuel changement de coordonnées, ne dépend que de certaines coordonnées de h : par exemple, si $\mathbb{R}^d = E_1 \oplus E_2$ où $\dim(E_1) = d_1$ et si γ_0 est un variogramme isotropique sur \mathbb{R}^{d_1} , $\gamma(h) = \gamma_0(h_1)$ si $h = h_1 + h_2$, $h_1 \in E_1$, $h_2 \in E_2$; le palier (éventuellement la portée) de γ dépendra alors de la direction (cf. Fig. 1.3-b). On parle d'*anisotropie zonale* ou *anisotropie stratifiée*, si γ est la somme de différentes composantes présentant des anisotropies de support. Par exemple,

$$\gamma(h) = \gamma_1(\sqrt{h_1^2 + h_2^2}) + \gamma_2(|h_2|)$$

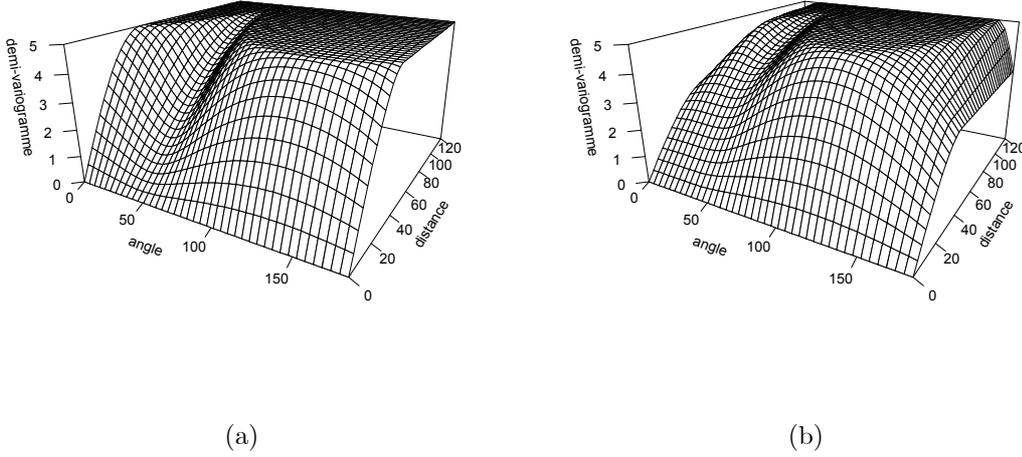


FIG. 1.3 – (a) Anisotropie géométrique et (b) anisotropie zonale.

présente un palier de niveau $\sigma_1^2 + \sigma_2^2$ dans la direction $(0, 1)$ et un palier de niveau σ_1^2 dans la direction $(1, 0)$ si σ_i^2 sont les paliers de γ_i , $i = 1, 2$. On peut étendre ce modèle en autorisant une anisotropie géométrique pour la première composante.

Plus généralement, une anisotropie s'obtiendra en combinant différents types d'anisotropies. La figure 1.3-b donne un tel exemple avec γ_1 le modèle sphérique de paramètres $a_1 = 30$ et $\sigma_1^2 = 2$ et γ_2 le modèle sphérique d'anisotropie géométrique de paramètres $a_2 = 30$, $\sigma_2^2 = 3$ d'une part et $\phi = 45^\circ$ et $e = 0.3$ d'autre part.

Décomposition en composantes spatiales. Si X est la somme de processus intrinsèques (resp. stationnaires de L^2) indépendants, il admet le variogramme (resp. la covariance) *gigogne*,

$$2\gamma(h) = \sum 2\gamma_j(h) \text{ ou } C(h) = \sum C_j(h).$$

Ce modèle est associé à des composantes spatiales agissant à des échelles différentes avec des paliers différents. Statistiquement, une composante à petite échelle ne pourra être identifiée que si la maille d'échantillonnage est assez fine et une composante à grande échelle que si le diamètre du domaine échantillonné est assez important. Cressie [16] suggère la décomposition,

$$X_s = m(s) + W_s + \eta_s + \varepsilon_s,$$

où $m(s)$ est une dérive déterministe régulière représentant la variation à grande échelle, W_s est une composante assez "lisse" modélisée par un processus intrinsèque de grande portée, η_s est une composante à micro-échelle indépendante de W_s et ε_s , l'erreur de mesure, est la composante pépitique.

1.3.2 Propriétés géométriques : continuité, différentiabilité

Considérons la convergence des v.a.r. dans L^2 : $X = \{X_s, s \in S\}$ est continu en moyenne quadratique (m.q.) en $s \in S$ si, $\forall s_n \rightarrow s$, $E(X_{s_n} - X_s)^2 \rightarrow 0$. La proposition suivante caractérise la continuité de X en m.q..

Proposition 1.3 Si $X \in L^2$ est un processus centré, X est continu en m.q. ssi sa covariance $C(s, t)$ est continue sur la diagonale de S^2 .

Examinons la différentiabilité dans L^2 si $S \subset \mathbb{R}$ (différentiabilité directionnelle si $S \subset \mathbb{R}^d$) : X sur $S \subset \mathbb{R}$ est différentiable en m.q. en s si

$$\exists \dot{X}_s \text{ t.q. : si } s_n \rightarrow s, \text{ alors } \frac{X_{s_n} - X_s}{s_n - s} \rightarrow \dot{X}_s \text{ dans } L^2.$$

Proposition 1.4 Soit $X \in L^2$, centré. Si $\frac{\partial^2}{\partial s \partial t} C(s, t)$ existe et est continue partout, alors X est différentiable en m.q. et $Cov(\dot{X}_s, \dot{X}_t) = \frac{\partial^2}{\partial s \partial t} C(s, t)$.

On déduit des propositions précédentes qu'un processus intrinsèque est continu en m.q. si son variogramme γ est continu en $h = 0$. De même, si la dérivée seconde $\gamma''(0)$ existe, X est différentiable en m.q.. Un processus stationnaire de L^2 sera continu (resp. dérivable) en m.q. si sa covariance est continue à l'origine (resp. $C''(0)$ existe). La figure 1.4 donne une idée de la régularité de X pour 4 types de variogrammes :

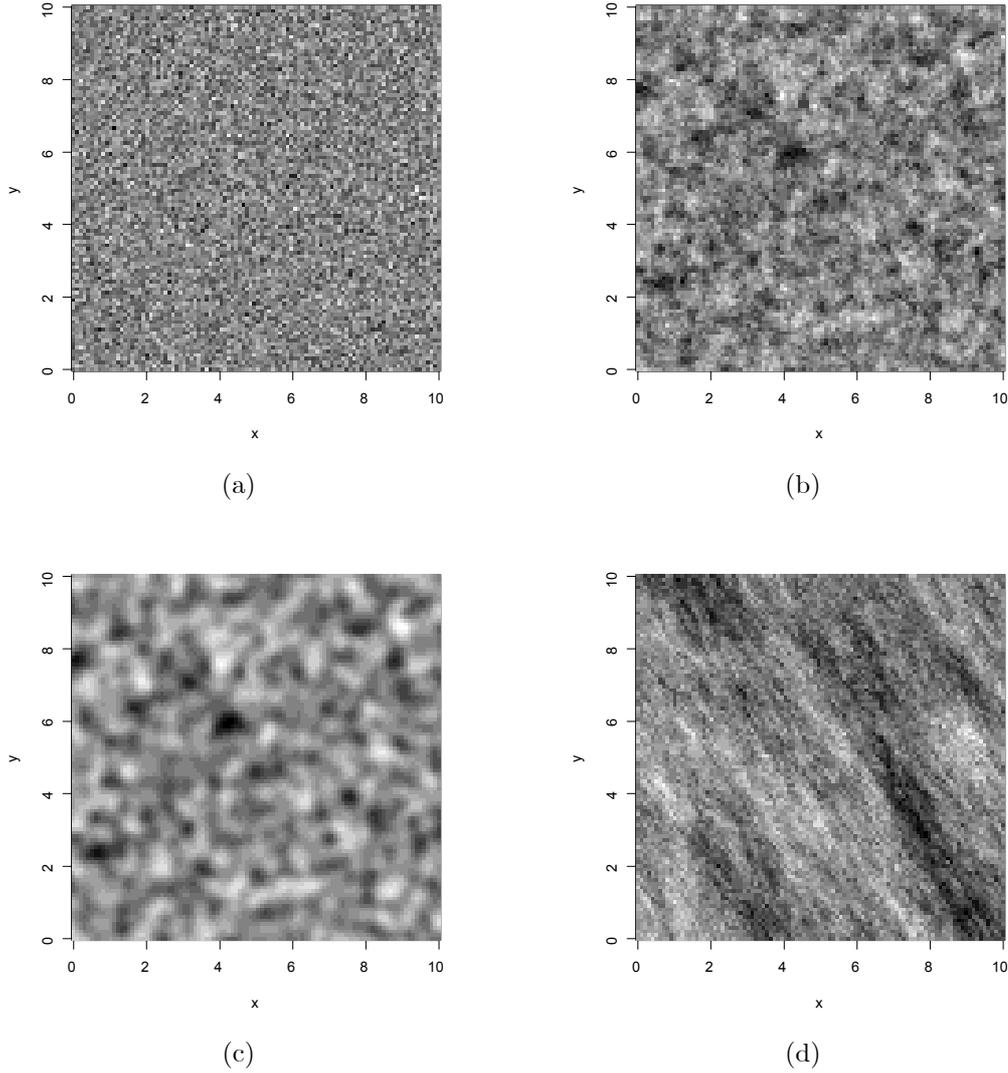


FIG. 1.4 – Quatre réalisations gaussiennes pour différents variogrammes : (a) pépitique, (b) exponentiel isotropique, (c) gaussien isotropique, (d) gaussien avec forte anisotropie.

- (a) γ , pépitique, n'est pas continue en 0 : X n'est pas continu en m.q., les trajectoires sont très irrégulières.
- (b) γ est linéaire à l'origine : $\gamma(h) = a + b \|h\| + o(\|h\|)$ est continue en 0 mais non dérivable en 0 ; X est continu en m.q. mais non dérivable.
- (c) γ est parabolique en 0 : $\gamma(h) = a + b \|h\|^2 + o(\|h\|^2)$ est de classe \mathcal{C}^2 à l'origine, les trajectoires sont continues et dérivables en m.q. ; (d) γ est parabolique en 0 mais avec anisotropie.

Remarquons que si γ est indéfiniment dérivable, X l'est aussi : en effet, dans ce cas [46],

$$X_t = \lim_{L^2} \sum_{k=0}^n \frac{t^k}{k!} X_0^{(k)}$$

De plus, X est purement déterministe car il suffit de connaître X dans un (petit) voisinage de 0 (on peut alors calculer toutes les dérivées de X en 0) pour connaître X partout. Cette remarque peut conduire à écarter un variogramme analytique (i.e. le variogramme gaussien) si on doute de ce déterminisme et/ou de l'hyper-régularité de X .

1.4 Prédiction à covariance connue : le krigeage

1.4.1 Le krigeage simple

Etant données n observations $\{X_{s_1}, \dots, X_{s_n}\}$ de X sur S , on veut prédire linéairement X_{s_0} en un nouveau site s_0 . La dénomination *krigeage*, qui recouvre ce problème, est due à Matheron, en référence aux travaux de *Krige*, un ingénieur minier sud-africain.

Supposons que la covariance (le variogramme) de X soit connue (dans la pratique, elle devra être préalablement estimée (cf. §1.5)). Notons : $X_0 = X_{s_0}$, $\mathbf{x} = (X_{s_1}, \dots, X_{s_n})$, $\Sigma = Cov(\mathbf{x})$, $\sigma_0^2 = Var(X_0)$ et $c = Cov(\mathbf{x}, X_0) \in \mathbb{R}^n$ et considérons un prédicteur linéaire :

$$\hat{X}_0 = \sum_{i=1, n} \lambda_i X_i = {}^t \lambda \mathbf{x} \quad (1.6)$$

Le critère de choix retenu pour la prévision est celui qui minimise le carré moyen de l'erreur $e_0 = X_0 - \hat{X}_0$

$$EQM(s_0) = E\{(X_0 - \hat{X}_0)^2\} \quad (1.7)$$

Supposons d'abord que $m(\cdot)$ soit connue, ou encore que X soit centré.

Proposition 1.5 Krigeage simple : la meilleure prédiction linéaire de X_0 est :

$$\hat{X}_0 = {}^t c \Sigma^{-1} \mathbf{x}. \quad (1.8)$$

La variance de l'erreur de prédiction vaut : $\tau^2(s_0) = \sigma_0^2 - {}^t c \Sigma^{-1} c$.

\hat{X}_0 est la prédiction BLUP (Best Linear Unbiased Prediction).

1.4.2 Le krigeage universel

Souvent, la moyenne de X est inconnue non-constante : X suit par exemple un modèle de régression spatiale. Si la régression est linéaire, la prédiction se fera par krigeage universel.

Le modèle de régression spatiale

On parle de régression spatiale si X est la somme d'une partie déterministe $m(\cdot)$ dite la dérive, tendance ou moyenne de X , et de ε , un champ de résidus centrés mais corrélés entre eux :

$$X_s = m(s) + \varepsilon_s$$

Il y a de nombreuses façons de modéliser $m(\cdot)$. Par exemple :

- $m(s) = \sum_{l=1, p} \beta_l f_l(s)$ est dans l'espace *linéaire* engendré par des fonctions connues f_l . Si $\{f_l\}$ est une base de polynômes, l'espace engendré est invariant vis-à-vis de l'origine des coordonnées. Par exemple, si $s = (x, y) \in \mathbb{R}^2$, on peut prendre $\{f_l\} = \{1, x, y, xy, x^2, y^2\}$.
- $m(s) = \sum_{l=1, p} \alpha_l Y_s^{(l)}$ s'exprime à partir d'*exogènes* observables.
- $m(\cdot)$ est une combinaison de ces deux formes.

Si X est observé en n sites $s_i \in S$, le modèle s'écrit :

$$X_{s_i} = {}^t z_{s_i} \delta + \varepsilon_{s_i}, \quad i = 1, \dots, n \quad (1.9)$$

où $z_{s_i} \in \mathbb{R}^p$ est observée, $\delta \in \mathbb{R}^p$ est inconnu et $\varepsilon = (\varepsilon_s)$ est un résidu spatialement corrélé. Notant $U(n) = {}^t (U_{s_1}, \dots, U_{s_n})$, ${}^t Z_n = (z_{s_1}, \dots, z_{s_n})$, le modèle (1.9) s'écrit matriciellement :

$$X(n) = Z_n \delta + \varepsilon(n), \text{ avec} \quad (1.10)$$

$$E(\varepsilon(n)) = 0 \text{ et } Cov(\varepsilon(n)) = \Sigma. \quad (1.11)$$

Le krigeage universel

Supposons que X suive le modèle (1.9) : connaissant z_0 , on cherche la prédiction linéaire sans biais de $X_{s_0} = X_0$ (c.a.d. vérifiant ${}^t \lambda Z = {}^t z_0$) minimisant (1.7).

Proposition 1.6 Krigeage universel : la meilleure prédiction linéaire sans biais de X_0 est :

$$\hat{X}_0 = \{ {}^t c \Sigma^{-1} + {}^t (z_0 - {}^t Z \Sigma^{-1} c) ({}^t Z \Sigma^{-1} Z)^{-1} {}^t Z \Sigma^{-1} \} \mathbf{x} \quad (1.12)$$

La variance de l'erreur de prédiction est

$$\tau^2(s_0) = \sigma_0^2 - {}^t c \Sigma^{-1} c + {}^t (z_0 - {}^t Z \Sigma^{-1} c) ({}^t Z \Sigma^{-1} Z)^{-1} (z_0 - {}^t Z \Sigma^{-1} c) \quad (1.13)$$

L'interprétation de la formule de krigeage universel est la suivante :

$$\hat{X}_0 = {}^t z_0 \hat{\delta} + c \Sigma^{-1} (x - Z \hat{\delta})$$

où $\hat{\delta} = ({}^t Z \Sigma^{-1} Z)^{-1} {}^t Z \Sigma^{-1} \mathbf{x}$ est l'estimateur Moindres Carrés Généralisés (MCG) de δ : \hat{X}_0 est la somme de l'estimation des MCG de $E(X_0) = {}^t z_0 \delta$ et du krigeage simple des résidus estimés $\hat{\varepsilon} = (\mathbf{x} - Z \hat{\delta})$.

Le *krigeage ordinaire* correspond à la situation $X_s = m + \varepsilon_s$ avec m inconnu.

On peut également donner les formules du krigeage en termes de variogramme si ε est intrinsèque car la stationnarité n'est pas requise dans l'obtention des résultats (1.12) et (1.13).

Propriétés de la surface de krigage

Le krigage est un *interpolateur exact* : $\hat{X}_{s_0} = X_{s_0}$ si s_0 est un site d'observation. En effet si $s_0 = s_i$ et si c est la i -ème colonne de Σ , alors $\Sigma^{-1}c = {}^t e_i$ où e_i est le i -ème vecteur de la base canonique de \mathbb{R}^n et ${}^t Z \Sigma^{-1} = {}^t z_0$.

La régularité de la covariance (du variogramme) en 0 détermine la *régularité de la surface de krigage*, en particulier aux points d'échantillonnage (cf. Fig. 1.5). Considérons l'exemple d'un modèle de moyenne constante inconnue m :

1. pour le modèle pépitique, si $s_0 \neq s_i$, alors $\Sigma^{-1}c = 0$, $\hat{m} = \sum_{i=1}^n X_{s_i}/n$ et la prédiction est la moyenne arithmétique si $s_0 \neq s_i$ avec des pics $\hat{X}_{s_0} = X_{s_i}$ si $s_0 = s_i$; plus généralement, si $C(h)$ est discontinue en 0, $s \mapsto \hat{X}_s$ est discontinue aux points d'échantillonnage ;
2. si $C(h)$ est linéaire à l'origine, $s \mapsto \hat{X}_s$ est continue partout mais non dérivable aux points d'échantillonnage ;
3. si $C(h)$ est parabolique en 0, $s \mapsto \hat{X}_s$ est continue et dérivable partout.

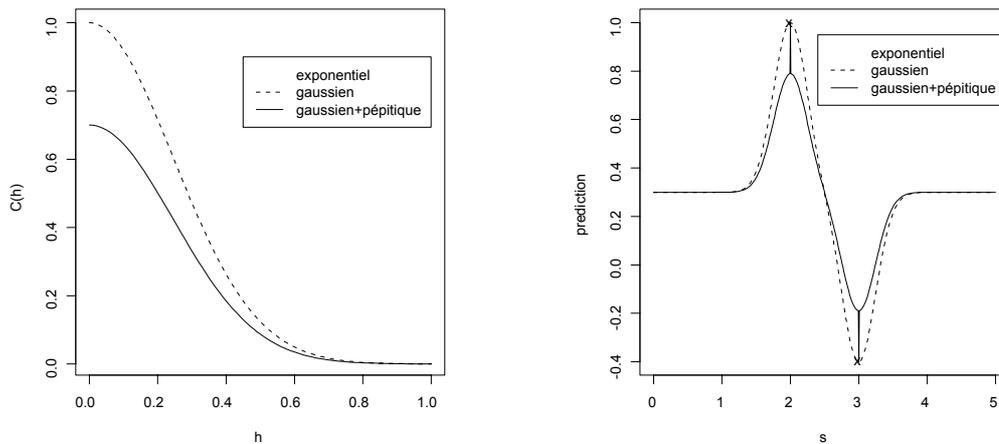


FIG. 1.5 – *A droite* : régularité de la surface de krigage (ici $S = \mathbb{R}$, $X_2 = 1$, $X_3 = -0.4$) en fonction de la régularité de la covariogramme en 0. *A gauche* : (i) exponentielle $C(h) = \exp(-|h|/0.2)$, (ii) gaussienne $C(h) = \exp(-0.6|h|^2/\sqrt{3})$, (iii) gaussien + effet pépitique $C(h) = \mathbf{1}_0(h) + 0.7 \exp(-.6\|h\|^2/\sqrt{3})(1 - \mathbf{1}_0(h))$.

1.5 Estimation en géostatistique

Considérons X , un champ intrinsèque sur $S \subset \mathbb{R}^d$ de moyenne constante et de variogramme $2\gamma(h) = E(X_{s+h} - X_s)^2$, X étant observé en n sites $\mathcal{O} = \{s_1, s_2, \dots, s_n\}$ de S : $X(n) = {}^t(X_{s_1}, \dots, X_{s_n})$.

1.5.1 Analyse variographique

Supposons dans un premier temps que γ soit isotropique. Le nuage variographique ou *nuée variographique*, est la donnée des $n(n-1)/2$ points de $(\mathbb{R}^+)^2$:

$$\mathcal{N}_{\mathcal{O}} = \{(\|s_i - s_j\|, (X_{s_i} - X_{s_j})^2/2), 1 \leq i < j \leq n\}$$

$(X_{s_i} - X_{s_j})^2/2$ estimant sans biais $\gamma(s_i - s_j)$, ce nuage est la bonne représentation de $h \mapsto 2\gamma(h)$ mais des couples (s_i, s_j) de sites avec un grand carré peuvent apparaître à côté d'autres couples à la même distance mais de petit carré (cf. Fig. 1.7-a) : à l'état brut, ce nuage ne suffit pas pour bien analyser les caractéristiques du variogramme telles que la portée, le seuil ou encore la présence d'un effet de pépite. Pour corriger cela, on effectue un lissage du nuage que l'on superpose au nuage de points ; ce lissage est obtenu soit par moyenne locale, soit, plus généralement, en utilisant un noyau de convolution.

A priori, le variogramme 2γ n'étant pas isotropique, il est conseillé de représenter les nuées variographiques dans plusieurs directions à partir des vecteurs $s_i - s_j$ "proches" de cette direction. On détectera ainsi empiriquement une éventuelle anisotropie du variogramme. En général, on retient les 4 orientations cardinales S, SE, E et NE avec une tolérance angulaire de ± 22.5 degrés autour de chacune.

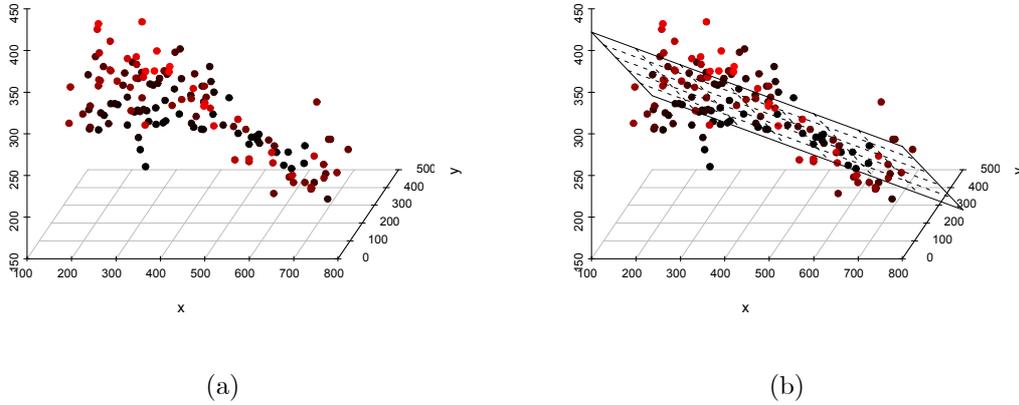


FIG. 1.6 – (a) Données de précipitations sur 143 stations du réseau météorologique de l’Etat du Parana, Brésil (données `parana` de `geoR`); (b) estimation d’une tendance lineaire $m(s)$.

Exemple 1.3 Pluies dans l’Etat du Parana (Diggle et Ribeiro)

Ces données donnent les hauteurs de pluie moyenne X_s sur la période mai-juin pour 143 stations météorologiques de l’Etat du Parana (Brésil).

La nuée variographique isotrope calculée sur les données brutes X étant très bruitée, on effectue un lissage par noyau gaussien avec un paramètre de lissage (ici l’écart-type) égal à 100 (cf. Fig. 1.7-a). Ce lissage, calculé dans les 4 directions cardinales, révèle une forte anisotropie (cf. Fig. 1.7-b). Cette anisotropie peut être due à une non-stationnarité en moyenne : la figure 1.6 invite au modèle de surface de réponse affine $E(X_s) = m(s) = \beta_0 + \beta_1 x + \beta_2 y$, $s = (x, y) \in \mathbb{R}^2$, pour lequel les résidus estimés par MCO restent anisotropiques (cf. Fig. 1.7-c). Si on avait proposé une surface de réponse d’ordre 2, $m(s) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2$, l’examen des nuées variographiques inviteraient à retenir un modèle de résidus proche d’un bruit blanc (cf. Fig. 1.7-d).

1.5.2 Estimation empirique du variogramme

Notons $\#A$ le cardinal de A . L’estimateur empirique du variogramme est

$$2\hat{\gamma}_n(h) = \frac{1}{\#N(h)} \sum_{s_i, s_j \in N(h)} (X_{s_i} - X_{s_j})^2, \quad h \in \mathbb{R}^d \quad (1.14)$$

où $N(h) = \{(s_i, s_j) : h - \Delta \leq s_i - s_j \leq h + \Delta; i, j = 1, \dots, n\}$ est une classe approximante à tolérance Δ de $h \in \mathbb{R}^d$. En pratique, on choisit Δ de telle sorte que chaque classe comporte au moins 30 couples de points.

Si X est stationnaire de L^2 , la covariance est estimée empiriquement par

$$\hat{C}_n(h) = \frac{1}{\#N(h)} \sum_{s_i, s_j \in N(h)} (X_{s_i} - \bar{X})(X_{s_j} - \bar{X}), \quad h \in \mathbb{R}^d \quad (1.15)$$

où $\bar{X} = n^{-1} \sum_{i=1}^n X_{s_i}$ estime sans biais la moyenne μ .

L’avantage de $2\hat{\gamma}_n(h)$ en comparaison de $\hat{C}_n(h)$ est de *ne pas demander l’estimation préalable de la moyenne* μ . De plus, sous hypothèse de stationnarité intrinsèque, $2\hat{\gamma}_n(h)$ estime sans biais $2\gamma(h)$. La proposition suivante donne sa loi sous hypothèse gaussienne, $2\hat{\gamma}_n(h)$ s’écrivant comme une forme quadratique $2\hat{\gamma}_n(h) = {}^t X A(h) X$ où $A(h)$ est une matrice symétrique s.d.p. $n \times n$ et de rang $\leq N(h)$; comme $A(h)\mathbf{1} = 0$, on peut supposer $\mu = 0$.

Proposition 1.7 *Si $X \sim \mathcal{N}_n(0, \Sigma)$, alors $\hat{\gamma}(h) \sim \sum_{i=1}^{N(h)} \lambda_i(h) \chi_{i1}^2$ où les $\lambda_i(h)$ sont les $N(h)$ valeurs propres non-nulles de $A(h)\Sigma$ et les χ_{i1}^2 sont des χ^2 i.i.d. à 1-ddl. En particulier, $E(\hat{\gamma}(h)) = \text{Trace}(A(h)\Sigma)$ et $\text{Var}(\hat{\gamma}(h)) = 2 \times \text{Trace}\{(A(h)\Sigma)^2\}$.*

Ce résultat permettra de calculer les pondérations pour l’estimation paramétrique par MC pondérés (cf (1.18)).

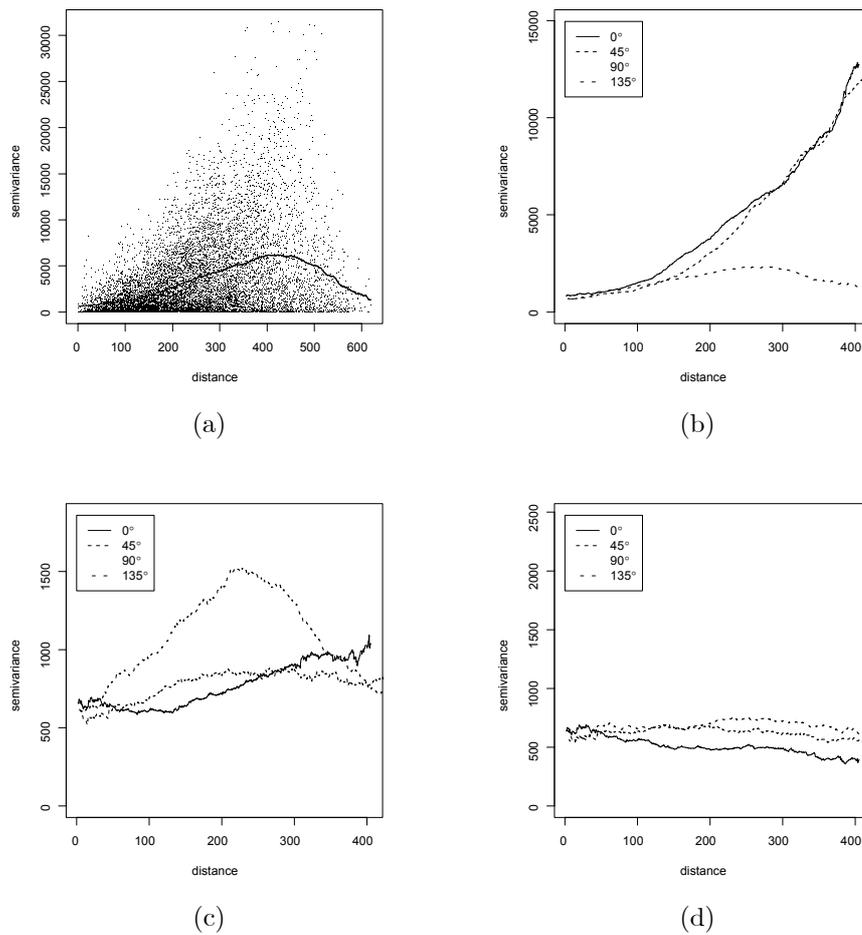


FIG. 1.7 – Analyse variographique des précipitations du Parana : (a) nuage variographique sur les données brutes et variogramme isotrope lissé ; (b) variogrammes lissés dans quatre directions ; (c) variogrammes lissés des résidus d'un modèle affine ; (d) variogrammes lissés des résidus d'un modèle quadratique.

L'estimation $\hat{\gamma}(h)$ étant peu robuste aux valeurs élevées de $X_{s_i} - X_{s_j}$, Cressie et Hawkins [16] ont proposé l'estimateur robustifié,

$$2\bar{\gamma}_n(h) = \left\{ 0.457 + \frac{0.494}{\#N(h)} \right\}^{-1} \left\{ \frac{1}{\#N(h)} \sum_{(s_i, s_j) \in N(h)} |X_{s_i} - X_{s_j}|^{\frac{1}{2}} \right\}^4$$

Ils montrent que le biais et la variance de $\bar{\gamma}$ sont moindres que ceux de $\hat{\gamma}$.

Estimation paramétrique d'un modèle de variogramme $\gamma(\cdot; \theta)$

Les modèles de variogramme $\gamma(\cdot; \theta)$ présentés au §1.3.1 sont paramétriques. L'estimation de θ par moindres carrés ordinaires (MCO) est une valeur

$$\hat{\theta}_{MCO} = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1, k} (\hat{\gamma}_n(h_i) - \gamma(h_i; \theta))^2 \quad (1.16)$$

où k est le nombre de classes $\mathcal{H} = \{h_1, h_2, \dots, h_k\}$ retenues pour l'estimation empirique de γ . On peut aussi utiliser $\bar{\gamma}_n(h_i)$ à la place de $\hat{\gamma}_n(h_i)$.

La méthode n'est pas efficace car les $\hat{\gamma}_n(h_i)$ ne sont ni indépendants ni de même variance. On préférera l'estimation par moindres carrés généralisés (MCG)

$$\hat{\theta}_{MCG} = \operatorname{argmin}_{\theta \in \Theta} {}^t(\hat{\gamma}_n - g(\theta))V_n^{-1}(\hat{\gamma}_n - g(\theta)) \quad (1.17)$$

avec $\hat{\gamma}_n = {}^t(\hat{\gamma}_n(h_1), \dots, \hat{\gamma}_n(h_k))$, $g(\theta) = {}^t(\gamma(h_1, \theta), \dots, \gamma(h_k, \theta))$ et $V_n = \operatorname{Cov}(\hat{\gamma}_n)$.

Le calcul de V_n se révélant souvent difficile, un compromis consiste à utiliser l'estimateur des MC pondérés (MCP) pour les poids $\operatorname{var}(\hat{\gamma}_n(h_i)) \simeq 2\gamma^2(h_i, \theta)/\#N(h_i)$ (cf. Proposition 1.7),

$$\hat{\theta}_{MCP} = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1, k} \frac{\#N(h_i)}{\gamma^2(h_i; \theta)} (\hat{\gamma}_n(h_i) - \gamma(h_i; \theta))^2 \quad (1.18)$$

Maximum de vraisemblance

Si X est gaussien, stationnaire de moyenne μ connue, alors $X(n) \sim \mathcal{N}_n(\mu, \Sigma(\theta))$ a pour log-vraisemblance,

$$l_n(\theta) = -\frac{1}{2} \{ \log |\Sigma(\theta)| + {}^t(X(n) - \mu)\Sigma^{-1}(\theta)(X(n) - \mu) \} \quad (1.19)$$

Pour maximiser (1.19), on itère le calcul de $|\Sigma(\theta)|$ et de $\Sigma^{-1}(\theta)$. Si $\hat{\theta}$ est l'estimation du MV de θ , celle du variogramme est $2\gamma(h, \hat{\theta})$. Les résultats relatifs au MV sont présentés dans le cadre général des régressions spatiales aux § 2.4.3.

1.5.3 Validation d'un modèle de variogramme

Validation croisée

La *validation croisée* consiste à éliminer à tour de rôle chaque observation x_i et à la prévoir par krigeage (\tilde{x}_i) sur la base des autres observations, sans réestimer le modèle de variogramme : le critère de validation croisée est celui de l'erreur quadratique normalisée moyenne,

$$EQNM = \frac{1}{n} \sum_{i=1, n} \frac{(x_{s_i} - \tilde{x}_{s_i})^2}{\tilde{\sigma}_{s_i}^2}$$

où $\tilde{\sigma}_{s_i}^2$ est la variance de krigeage. Si le modèle de variogramme est valide et bien estimé, on obtiendra une valeur $EQNM$ proche de 1. En première approximation, les erreurs normalisées $(X_{s_i} - \tilde{X}_{s_i})/\tilde{\sigma}_{s_i}$ sont $\mathcal{N}(0, 1)$ et on jugera de l'adéquation à partir d'un intervalle de confiance d'un χ_n^2 , proche d'une normale $\mathcal{N}(n, 2n)$ si n est grand : on retient le modèle au niveau (approximatif) de 95% si :

$$|EQNM - 1| \leq 1.96 \sqrt{\frac{2}{n}}$$

Validation par méthode de Monte Carlo : Bootstrap paramétrique

On peut également valider un modèle par la méthode de *Bootstrap paramétrique* qui suit. Soit $\hat{\theta}$ un estimateur du paramètre θ du modèle $\gamma(\cdot, \theta)$ à valider. On réalise m simulations $(X^{(j)}, j = 1, \dots, m)$ i.i.d. d'un champ intrinsèque sur $S = \{s_1, s_2, \dots, s_n\}$ de variogramme $\gamma(\cdot; \hat{\theta})$. Pour chaque simulation j , on calcule les $\{\hat{\gamma}_n^{(j)}(h_i), i = 1, k\}$ empiriques sur $\mathcal{H} = \{h_1, h_2, \dots, h_k\}$ ainsi que leurs enveloppes inférieures et supérieures, $\hat{\gamma}_{inf}(h_i) = \min_{j=1, m} \hat{\gamma}_n^{(j)}(h_i)$ et $\hat{\gamma}_{sup}(h_i) = \max_{j=1, m} \hat{\gamma}_n^{(j)}(h_i)$, $i = 1, k$. Si on accorde le même poids à chaque simulation, ces deux enveloppes définissent un intervalle de confiance empirique de niveau $\geq (1 - \frac{2}{m+1})$:

$$\Pr(\hat{\gamma}_n^{(j)}(h_i) < \hat{\gamma}_{inf}(h_i)) = \Pr(\hat{\gamma}_n^{(j)}(h_i) > \hat{\gamma}_{sup}(h_i)) \leq \frac{1}{m+1}$$

avec égalité à $(1 - \frac{2}{m+1})$ si les valeurs $\hat{\gamma}_n^{(j)}(h_i)$ sont toutes différentes. On représentera alors graphiquement les fonctions $h_i \rightarrow \hat{\gamma}_{inf}(h_i)$ et $h_i \rightarrow \hat{\gamma}_{sup}(h_i)$, $i = 1, k$: si les estimations empiriques sur les données initiales $\hat{\gamma}_n(h_i)$ sont dans la bande de confiance $\{[\hat{\gamma}_{inf}(h_i), \hat{\gamma}_{sup}(h_i)], i = 1, \dots, k\}$, on conclura raisonnablement que X est un processus intrinsèque de variogramme $\theta \mapsto \gamma(\cdot, \theta)$ (cf. la figure 1.9-b) : sinon, on rejettera le modèle $\gamma(\cdot, \theta)$, l'examen des causes de ce rejet pouvant servir à spécifier un modèle alternatif.

Exemple 1.4 *Nuage radioactif de Tchernobyl et pluie journalière en Suisse.*

La pluie journalière étant un bon indicateur des effets des retombées radioactives, on dispose des cumuls pluviométriques provenant du réseau météorologique suisse relevés le 8 mai 1986, jour où le nuage de Tchernobyl se déplaçait au-dessus de l'Europe. Ces données ont fait l'objet d'un concours où les chercheurs étaient invités à proposer leurs modèles et méthodes de prévision : 100 données étant disponibles (données `sic.100` dans le package `geoR`) sur les 467 du réseau Suisse (cf Fig. 1.8), le but était de prévoir au mieux les 367 non-disponibles.

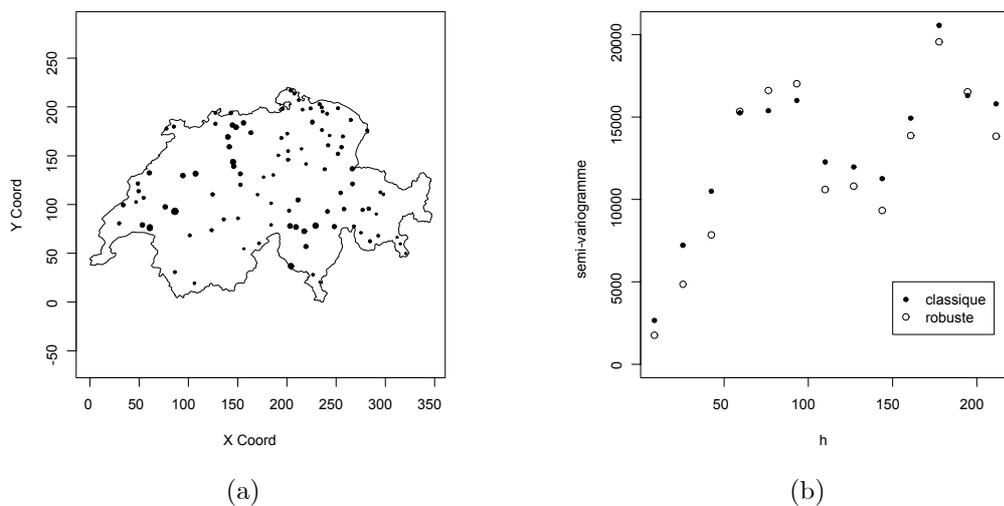


FIG. 1.8 – (a) Données de pluie pour 100 stations météorologiques suisses : les dimensions des symboles sont proportionnelles à l'intensité de la pluie ; (b) les 2 estimations non-paramétriques du semi-variogramme pour 13 classes de distance.

Si les estimateurs (1.14) et (1.15) ne garantissent pas la condition conditionnellement définie négative que doit satisfaire un variogramme, ils permettent néanmoins d'estimer un modèle paramétrique (§1.3.1) : le tableau 1.1 présente les estimations obtenues en utilisant `geoR` pour le variogramme de Matérn.

Méthode	\hat{a}	$\hat{\sigma}^2$	$\hat{\nu}$
MCO	17.20	15135.53	1.21
MCC	18.19	15000.57	1.00
MV	13.40	13664.45	1.31

TAB. 1.1 – Estimations paramétriques du semi-variogramme de Matérn pour les données de pluie.

La figure 1.9 compare les semi-variogrammes obtenus par MCO, MCP et MV. Les EQNM des MCO, MCP et MV valent respectivement 1.37, 1.01 et 1.09, ce qui valide le modèle de Matérn et l'estimation par MCP.

La figure 1.10 donne la carte des prévisions de pluie ainsi que celle de leurs écarts-types.

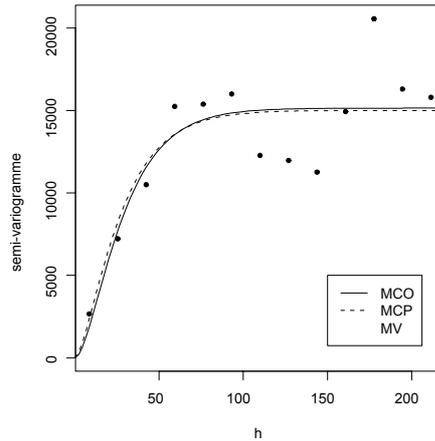


FIG. 1.9 – Les 3 estimations paramétriques du semi-variogramme.

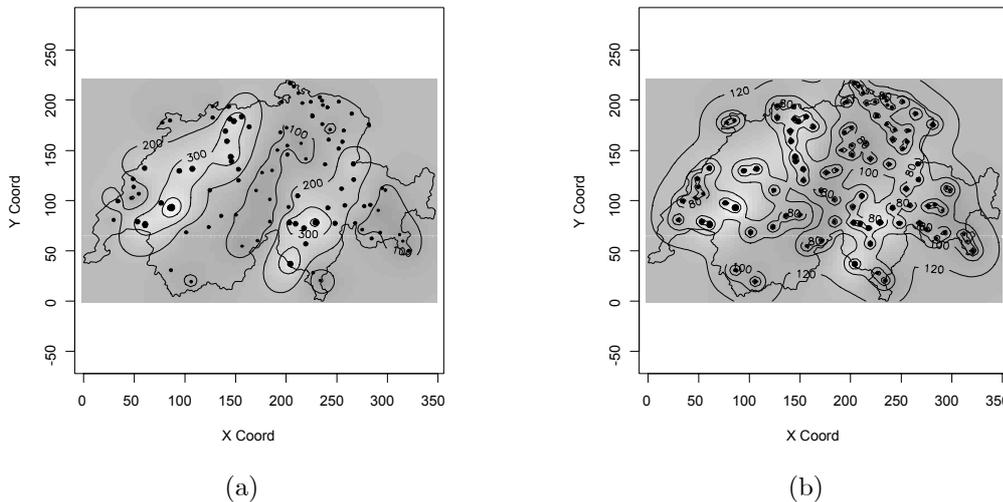


FIG. 1.10 – Krigeage sur les données de pluie : (a) prévisions \hat{X}_s et (b) écarts-types de l'erreur de prédiction.

1.6 Exercices

Exercice 1.1 *Corrélation spatiale et variance d'une moyenne.*

$X = \{X_s, s \in \mathbb{Z}^d\}$ est un champ stationnaire de moyenne m et de covariance $C(h) = \sigma^2 \rho^{\|h\|_1}$.

(1) Pour $d = 1$ et $\bar{X} = \frac{1}{9} \sum_{t=1,9} X_t$, montrez que :

$$V_1(\rho) = \text{Var}\{\bar{X}\} = \frac{\sigma^2}{81} \left\{ 9 + 2 \sum_{k=1,8} (9 - k) \rho^k \right\}.$$

(2) Pour $d = 2$ et $\bar{X} = \frac{1}{9} \sum_{s,t=1,3} X_{s,t}$, montrez que :

$$V_2(\rho) = \text{Var}(\bar{X}) = \frac{\sigma^2}{81} \{ 9 + 24\rho + 28\rho^2 + 16\rho^3 + 4\rho^4 \}.$$

Comparez $V_1(0)$, $V_1(\rho)$ et $V_2(\rho)$. Vérifiez que pour $\rho = 1/2$, ces 3 valeurs sont dans les rapports 9, 15.76 et 30.25.

Exercice 1.2 *Condition suffisante de s.d.p. d'une matrice.*

(1) C est s.d.p. si C est diagonalement dominante : $\forall i, C_{ii} \geq \sum_{j:j \neq i} |C_{ij}|$.

(2) Écrire cette condition pour : (i) $C(i, j) = \rho^{|i-j|}$ sur \mathbb{Z} ; (ii) $C(i, j) = 1$ si $i = j$, ρ si $\|i - j\|_\infty = 1$ et 0 sinon, sur \mathbb{Z}^d . Montrer que cette condition n'est pas nécessaire.

Exercice 1.3 Produit de covariances, extension d'une covariance.

(1) Si $C_k(\cdot)$, $k = 1, d$ sont des covariances stationnaires sur \mathbb{R}^1 , montrer que $C(h) = \prod_{k=1,d} C_k(h_k)$ est une covariance sur \mathbb{R}^d si $h = (h_1, \dots, h_d)$.

(3) Démontrer que $C_0(h) = (1 - \frac{|h|}{\sqrt{2}})1_{\{|h| \leq \sqrt{2}\}}$ est une covariance sur \mathbb{R} . Vérifier que pour les choix $s_{ij} = (i, j)$ et $a_{ij} = (-1)^{i+j}$:

$$\sum_{(i,j),(k,l) \in \{1,2,\dots,7\}^2} a_{ij} a_{kl} C_0(\|s_{ij} - s_{kl}\|) < 0.$$

En déduire que $C_0(\|h\|)$ n'est pas une covariance sur \mathbb{R}^2 .

Exercice 1.4 Propriété de Markov du variogramme exponentiel sur \mathbb{R}^1 .

X est un processus stationnaire de covariance $C(t) = \sigma^2 \rho^{|t|}$ observé en n sites $\{s_1 < s_2 < \dots < s_n\} \subset \mathbb{R}^1$. Démontrer que si $s_0 < s_1$, le krigeage de X en s_0 est $\hat{X}_{s_0} = \rho^{s_1 - s_0} X_{s_1}$ et que si $s_k < s_0 < s_{k+1}$, \hat{X}_{s_0} dépend uniquement de X_{s_k} et $X_{s_{k+1}}$.

Exercice 1.5 Pour $X = (X_t, t \in \mathbb{R})$ stationnaire de covariance C , on observe $X_0 = -1$ et $X_1 = 1$. Montrer que le krigeage simple vaut $\hat{X}_s = \{C(s-1) - C(s)\} / \{C(0) - C(1)\}$ et la variance de l'erreur de prédiction $\tau^2(s)$

$$\tau^2(s) = C(0) \left(1 - \frac{(C(s) + C(s-1))^2}{C(0)^2 - C(1)^2} \right) + 2 \frac{C(s)C(s-1)}{C(0) - C(1)}$$

Tracer $s \mapsto \hat{X}_s$ et $s \mapsto \tau^2(s)$ pour $s \in [-3, 3]$ et la covariance de Matérn avec $C(0) = 1$, $a = 1$, $\nu = 1/2$ (resp. $\nu = 3/2$). Commenter les graphiques.

Exercice 1.6 Modèles à corrélation nulle à distance > 1 .

$X = (X_t, t \in \mathbb{Z})$ est un processus stationnaire de corrélation ρ à distance 1 et 0 sinon. On note $X(n) = {}^t(X_1, X_2, \dots, X_n)$ et $\Sigma_n = \text{Cov}(X(n))$

- (1) A quelle condition sur ρ , Σ_3 est s.d.p. ? Même question pour Σ_4 .
- (2) Quelle est la fonction de corrélation du MA(1) sur \mathbb{Z} : $X_t = \varepsilon_t + a\varepsilon_{t-1}$ où ε est un BBF ? En déduire que $|\rho| \leq 1/2$ assure que, pour tout n , Σ_n est s.d.p..
- (3) Déterminer le krigeage de X_0 sur X_1, X_2 . Même question si : (i) $E(X_t) = m$ est inconnue ; (ii) $E(X_t) = at$.
- (4) Mêmes questions pour $X = (X_{s,t}, (s,t) \in \mathbb{Z}^2)$, un champ stationnaire corrélé à ρ à distance 1 et à 0 sinon.

Exercice 1.7 Soit $C(t)$ une covariance stationnaire. Démontrer que :

$$|C(s+h) - C(s)| \leq \sqrt{2C(0)} \sqrt{C(0) - C(h)}$$

Indication : $C(s) - C(t) = E\{X(h) [X(s+h) - X(t+h)]\}$.

En déduire qu'une covariance stationnaire continue à l'origine est uniformément continue.

Exercice 1.8 Concentration en métaux toxiques dans le sol. Le jeu de données *meuse* (près du fleuve "La Meuse") du package *gstat* donne les concentrations dans le sol de différents métaux ainsi que des variables expliquant le type de sol et son utilisation. La variable à expliquer est le **zinc**.

1. Proposer différentes estimations empiriques du variogramme de **zinc** pour différentes directions. Y a-t-il anisotropie ?
2. Proposer des modèles de variogramme avec éventuelle anisotropie géométrique. Les estimer. Comparer les résultats du MV et des MCG par validation croisée.
3. Proposer une modélisation alternative par régression spatiale avec comme covariables les coordonnées spatiales. L'anisotropie disparaît-elle ? Proposer un modèle pour le variogramme des résidus et estimer le modèle de régression spatiale.
4. Dessiner la carte de la concentration en zinc obtenue par krigeage.
5. Proposer des modèles de régression spatiale incorporant des variables liées à l'utilisation du sol. Estimer ces régressions par MV.

Chapitre 2

Auto-régression et corrélation spatiales

Une *Auto-régression (AR)* spatiale est un modèle de *données réelles* défini sur un *ensemble discret spatial* S . L'ensemble S est régulier (i.e. $S \subset \mathbb{Z}^2$) ou non, sa géométrie étant associée à un graphe orienté d'influence \mathcal{R} . La corrélation spatiale d'une auto-régression découlera du modèle retenu. Ces modèles explicatifs sont bien adaptés à des mesures agrégées par unités spatiales : par exemple, en épidémiologie, X_s est le nombre de personnes touchées par une maladie donnée dans un canton s d'un département S ; en agronomie, X_s est le rendement d'une parcelle de culture s .

Après la présentation de modèles stationnaires sur $S = \mathbb{Z}^d$, nous décrirons deux modèles très utilisés en statistique spatiale, à savoir les modèles *AR-simultanés SAR* et les modèles *AR conditionnels CAR*, et la statistique de ces modèles. Nous définirons également les indices de Moran et de Geary qui permettent de tester si des données spatiales (sur un réseau irrégulier) sont indépendantes ou non.

2.1 Auto-régression stationnaire

2.1.1 Moyenne glissante (MA) et ARMA

Modèle MA

Un modèle $MA(\infty)$ est un processus linéaire,

$$X_t = \sum_{s \in \mathbb{Z}^d} c_s \eta_{t-s},$$

associé à une suite $c = (c_s)$ de l^2 ($\sum_{\mathbb{Z}^d} c_s^2 < \infty$) et à un $\text{BBf}(\sigma_\eta^2)$ $\eta : X$ est la *moyenne mobile* (Moving Average) infinie du bruit η pour les poids c . On parle de *modèle MA* si le support $M = \{s \in \mathbb{Z}^d : c_s \neq 0\}$ est fini. La covariance et la densité spectrale d'un MA sur \mathbb{Z}^d valent respectivement :

$$r(h) = \sigma_\eta^2 \sum_{t \in \mathbb{Z}^d} c_t c_{t+h} \text{ et } f(u) = \frac{\sigma_\eta^2}{(2\pi)^d} \left| \sum_{t \in \mathbb{Z}^d} c_t \exp i \langle u, t \rangle \right|^2 .$$

Modèle ARMA

Soient $P(z)$ et $Q(z)$ deux polynômes de la variable complexe multidimensionnelle $z \in \mathbb{C}^d$,

$$P(z) = 1 - \sum_{s \in R} a_s z^s \text{ et } Q(z) = 1 + \sum_{s \in M} c_s z^s$$

R (resp. M), le support *AR* (resp. le support *MA*) sont deux parties finies de \mathbb{Z}^d ne contenant pas l'origine, $z^s = z_1^{s_1} \cdots z_d^{s_d}$ si $s = (s_1, s_2, \dots, s_d)$. Notons B l'opérateur de retard, $B^s X_t = X_{t-s}$: le modèle défini par,

$$\text{Si } t \in \mathbb{Z}^d : P(B)X_t = Q(B)\eta_t, \text{ ou encore} \tag{2.1}$$

$$X_t = \sum_{s \in R} a_s X_{t-s} + \eta_t + \sum_{s \in M} c_s \eta_{t-s}. \tag{2.2}$$

existe si P ne s'annule pas sur le tore \mathbb{T}^d ($\mathbb{T} = \{\xi \in \mathbb{C}, |\xi| = 1\}$). X est stationnaire de densité spectrale,

$$f(u) = \frac{\sigma^2}{(2\pi)^d} \left| \frac{Q}{P}(e^{iu}) \right|^2 \text{ où } e^{iu} = (e^{iu_1}, \dots, e^{iu_d}).$$

Un modèle MA correspond à $P \equiv 1$; un modèle AR à $Q \equiv 1$.

Comme pour une série temporelle ($d = 1$), un champ à densité spectrale continue peut être approché dans L^2 par un $ARMA$, les fractions rationnelles étant denses dans l'espace de fonctions continues sur le tore \mathbb{T}^d . La densité spectrale d'un $ARMA$ étant rationnelle, sa covariance décroît exponentiellement vers 0 à l'infini. Comme pour une série temporelle, les covariances vérifient à partir d'un certain rang des équations de récurrences linéaires de Yule-Walker. Sur \mathbb{Z} , ces équations se résolvent analytiquement et elles constituent un outil d'identification des portées R et M et d'estimation des paramètres a et c . Ce n'est plus le cas si $d \geq 2$: c'est une des raisons qui rend difficile l'utilisation de modèles $ARMA$ sur \mathbb{Z}^d pour $d \geq 2$. Une autre raison de fond est qu'une modélisation $ARMA$ "bilatérale" n'admet pas de représentation unilatérale (ou causale) finie pour l'ordre lexicographique si $d \geq 2$.

Deux représentations AR , les SAR et les CAR , sont très populaires.

2.1.2 Auto-régression simultanées (SAR)

Un modèle stationnaire SAR (*Simultaneous AR*) relatif au BBf η s'écrit :

$$\text{Si } t \in S : X_t = \sum_{s \in R} a_s X_{t-s} + \eta_t. \quad (2.3)$$

X_t est la somme pondérée des valeurs X_u aux R -voisins de t , bruitée par η_t . (2.3) est un système d'équations AR simultanées au sens de l'économétrie : les $\{X_{t-s}, s \in R\}$ sont des covariables endogènes "retardées spatialement" qui influencent X_t , un site u influençant sur t si $t - u \in R$: cette relation définit le graphe orienté \mathcal{R} du modèle SAR .

Exemple 2.1 Exemples de SAR .

Modèle semi-causal Espace \times Temps : $s \in \mathbb{Z}$ repère l'espace, $t \in \mathbb{N}$ le temps ; une dynamique markovienne en t et locale dans l'espace s est

$$\forall t \in \mathbb{N} \text{ et } s \in \mathbb{Z} : X_{s,t} = \alpha X_{s,t-1} + \beta(X_{s-1,t} + X_{s+1,t}) + \varepsilon_{s,t}.$$

La liaison temporelle $(s, t-1) \rightarrow (s, t)$ est orientée, les liaisons spatiales instantanées $(s, t) \longleftrightarrow (s \mp 1, t)$ ne le sont pas. La représentation causale lexicographique de ce SAR est infinie, la prévision de X_{00} faisant intervenir X aux sites $\{(-1, 0), (-2, 0), (k, -1) \text{ pour } k \geq -1\}$.

Modèle SAR isotropique aux 4-ppv sur \mathbb{Z}^2 :

$$\forall (s, t) \in \mathbb{Z}^2 : X_{s,t} = \alpha(X_{s-1,t} + X_{s+1,t} + X_{s,t-1} + X_{s,t+1}) + \varepsilon_{s,t}.$$

existe ssi, $\forall \lambda, \mu \in [0, 2\pi[$, $P(\lambda, \mu) = 1 - 2\alpha(\cos \lambda + \cos \mu) \neq 0$, c.a.d. si $|\alpha| < 1/4$. La d.s. de X vaut $f(\lambda, \mu) = \sigma_\varepsilon^2 P(\lambda, \mu)^{-2}$. Ici, \mathcal{R} est symétrique.

Modèle $SAR(1)$ factorisant : le SAR

$$X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} - \alpha\beta X_{s-1,t-1} + \varepsilon_{s,t}, \quad |\alpha| \text{ et } |\beta| < 1 \quad (2.4)$$

a pour covariance : $r(s, t) = \sigma_\varepsilon^2 (1 - \alpha^2)^{-1} (1 - \beta^2)^{-1} \alpha^{|s|} \beta^{|t|}$.

Les modèles SAR sont très utilisés pour leur simplicité de paramétrisation. Cependant, il faudra veiller à deux problèmes :

1. Sans contrainte, un SAR peut ne pas être identifiable (un modèle paramétrique $\mathcal{M}(\theta)$ est *identifiable* si les lois qu'il définit pour deux valeurs différentes de θ sont différentes). Par exemple, sur \mathbb{Z} , les représentations SAR : (i) $X_t = aX_{t-1} + bX_{t+1} + \eta_t$, $t \in \mathbb{Z}^1$, $a \neq b$, $|a|, |b| < 1/2$, (ii) $X_t = bX_{t-1} + aX_{t+1} + \eta_t^*$ et (iii) $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t$, sont identiques pour des choix convenables de a_1, a_2 et des BBf η, η^* et ε (identifier les 3 d.s.). (i) est identifiable sous la contrainte $a < b$.
2. Comme pour des équations simultanées de l'économétrie, l'estimation d'un SAR par la méthode des moindres carrés ordinaires (MCO) sur les résidus n'est pas consistante (cf. Prop. 2.3).

2.1.3 Auto-régression conditionnelle (CAR)

Un modèle CAR donne la représentation en espérance conditionnelle, ou prédiction, linéaire de X_s sur les autres X_t .

Soit L une partie finie et symétrique de \mathbb{Z}^d , $0 \notin L$ et L^+ le demi-espace L positif pour l'ordre lexicographique sur \mathbb{Z}^d .

Définition 2.1 *Un CAR(L) stationnaire de L^2 s'écrit :*

$$X_t = \sum_{s \in L} c_s X_{t-s} + e_t,$$

avec, pour tout $s \in L^+$, $c_s = c_{-s}$, où e est un résidu conditionnel centré vérifiant : $\forall s \neq t, \text{Cov}(e_t, X_s) = 0$.

La non-corrélation entre X_s et e_t pour $s \neq t$ traduit que $\sum_{s \in L} c_s X_{t-s}$ est la meilleure prédiction linéaire dans L^2 de X_t sur les $\{X_s, s \neq t\}$: X est L -markovien au sens de la prédiction linéaire. Si X est gaussien, c'est la meilleure prédiction : X est un champ gaussien L -markovien. Plusieurs aspects différencient les CAR des SAR :

1. Un CAR est soumis à des contraintes paramétriques : L symétrique et $c_s = c_{-s}$.
2. Les résidus conditionnels e forment un *bruit coloré* :

$$\text{Cov}(e_t, e_{t+s}) = \begin{cases} \sigma_e^2 & \text{si } s = 0, \\ -\sigma_e^2 c_s & \text{si } s \in L \\ 0 & \text{sinon.} \end{cases}$$

3. Les résidus e_t sont décorrélés des X_s si $s \neq t$.
4. La densité spectrale d'un $CAR(L)$ est

$$f_X(u) = \frac{\sigma_e^2}{(2\pi)^d} \frac{1}{(1 - 2 \sum_{s \in L^+} c_s \cos\langle s, u \rangle)}$$

Là où la densité spectrale d'un CAR est proportionnelle à $P(u)^{-1}$, celle du SAR "associé" (le résidu conditionnel e est remplacé par un BBf ε) est proportionnelle à $|P(u)|^{-2}$.

Comme pour un SAR , les équations de Yule-Walker sur la covariance d'un CAR s'obtiennent en multipliant par X_s l'équation définissant X_t , puis en prenant l'espérance : par exemple, pour le CAR isotropique aux 4-ppv sur \mathbb{Z}^2 , ces équations s'écrivent :

$$\forall s, r(s) = \sigma_e^2 \times \delta_0(s) + a \sum_{t: \|t-s\|_1=1} r(t).$$

Trois raisons peuvent justifier l'utilisation d'un modèle CAR :

- une représentation CAR est intrinsèque, celle de la meilleure prédiction linéaire de X_t sur les toutes autres valeurs $\{X_s, s \neq t\}$;
- l'estimation d'un modèle CAR par MCO est consistante (cf. Prop. 2.3);
- la famille des CAR contient celle des SAR , strictement si $d \geq 2$.

Cependant, le voisinage de dépendance d'un CAR et sa paramétrisation sont plus complexes que celles d'un SAR . C'est la raison qui fait que dans la pratique les SAR sont plus utilisés que les CAR .

Proposition 2.1 *Modèles SAR et modèles CAR*

1. *Tout SAR est un CAR. Sur \mathbb{Z} , les deux classes coïncident.*
2. *Si $d \geq 2$, la famille CAR est plus grande que celle des SAR.*

Démonstration :

(1) Le SAR : $X_t = \sum_{s \in R} a_s X_{t-s} + \varepsilon_t$, ou $P(B)X_t = \varepsilon_t$, admet la représentation $CAR(L)$: $C(B)X_t = e_t$ où,

$$L = R - R, c_0 = 1, \text{ et si } s \neq 0, \sigma_e^2 \sum a_v a_{v+s} = \sigma_e^2 c_s$$

Il suffit de développer $|P(e^{i\lambda})|^2 = |1 - \sum_{s \in R} a_s e^{i\langle \lambda, s \rangle}|^2$ et d'identifier les d.s.,

$$f(u) = \frac{\sigma_e^2}{(2\pi)^d |P(e^{iu})|^2} = \frac{\sigma_e^2}{(2\pi)^d C(e^{iu})}, \text{ avec } c_0 = 1$$

L'identité des deux classes si $d = 1$ résulte du théorème de Fejer qui dit que tout polynôme trigonométrique $P(e^{i\lambda})$ d'une seule variable qui est ≥ 0 est le module au carré d'un polynôme trigonométrique : $\exists Q$ t.q. $P(e^{i\lambda}) = |Q(e^{i\lambda})|^2$.

(2) Sur \mathbb{Z}^2 le CAR , $X_t = a \sum_{s: \|s-t\|_1=1} X_s + e_t$, $a \neq 0$, n'admet pas de représentation SAR . En effet, $f_X^{-1}(\lambda_1, \lambda_2) = c(1 - 2a(\cos \lambda_1 + \cos \lambda_2))$: si un SAR admettait f_X pour d.s., son support R contiendrait les points $(1, 0)$ et $(0, 1)$ et donc un terme $\neq 0$ en $\cos(\lambda_1 - \lambda_2)$ devrait apparaître dans f_X , ce qui n'est pas le cas. \square

Un processus MA de support fini est à covariance de portée bornée. Si $d = 1$, le théorème de Fejer assure que la réciproque est vraie : un processus sur \mathbb{Z} de covariance de portée bornée est une MA . Ceci n'est plus vrai si $d \geq 2$.

Exemple 2.2 Exemples de correspondance SAR \rightarrow CAR

1. L'AR causal (cf. Fig. 2.1-a) :

$$X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} + \varepsilon_{s,t},$$

est un CAR(L) avec $L^+ = \{(1,0), (0,1), (-1,1)\}$, $c_{1,0} = \alpha\kappa^2$, $c_{0,1} = \beta\kappa^2$, $c_{-1,1} = -\alpha\beta\kappa^2$, $\sigma_\varepsilon^2 = \sigma_e^2\kappa^2$ où $\kappa^2 = (1 + \alpha^2 + \beta^2)$. Ici et dans les deux exemples suivant, κ^2 mesure le facteur de gain dans l'erreur de prédiction de X en utilisant la représentation CAR au lieu de la représentation SAR.

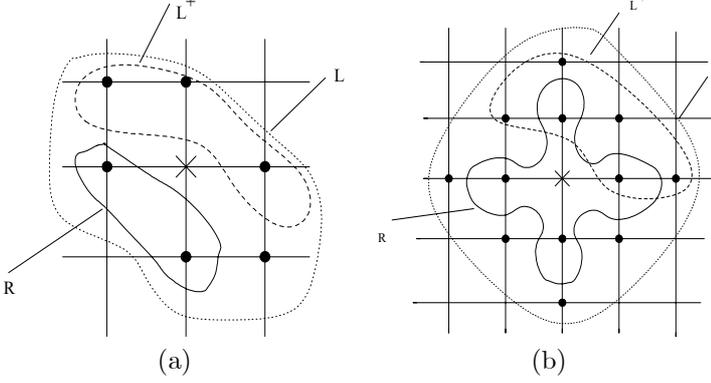


FIG. 2.1 – (a) Support $R = \{(1,0), (0,1)\}$ du modèle SAR causal et support L du CAR associé; (b) support $R = \{(1,0), (0,1), (-1,0), (0,-1)\}$ du modèle SAR non-causal et support L du CAR associé.

2. Le SAR non-causal,

$$X_{st} = a(X_{s-1,t} + X_{s+1,t}) + b(X_{s,t-1}, X_{s,t+1}) + \varepsilon_{s,t},$$

est le CAR(L) (cf. Fig. 2.1-b) où $L^+ = \{(1,0), (2,0), (-1,1), (0,1), (1,1), (0,2)\}$, $c_{1,0} = 2a\kappa^2$, $c_{0,1} = 2b\kappa^2$, $c_{2,0} = 2a^2\kappa^2$, $c_{0,2} = 2b^2\kappa^2$, $c_{-1,1} = -2ab\kappa^2$, $\sigma_\varepsilon^2 = \kappa^2\sigma_e^2$, et $\kappa^2 = 1 + 2a^2 + 2b^2$.

3. Le SAR factorisant :

$$X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} - \alpha\beta X_{s-1,t-1} + \varepsilon_{s,t}, |\alpha| \text{ et } |\beta| < 1,$$

est un CAR aux 8-ppv, $\sigma_\varepsilon^2 = \kappa^2\sigma_e^2$ où $\kappa^2 = (1 + \alpha^2)(1 + \beta^2)$, $c_{1,0} = \alpha(1 + \alpha^2)^{-1}$, $c_{0,1} = \beta(1 + \beta^2)^{-1}$, $c_{1,1} = c_{-1,1} = -c_{1,0} \times c_{0,1}$.

2.2 Modèle non-stationnaire sur S fini

2.2.1 Modèle SAR

Un champ réel sur $S = \{1, 2, \dots, n\}$ est une v.a. vectorielle $X \in \mathbb{R}^n$. Considérons $\varepsilon = (\varepsilon_t, t \in S)$ un bruit centré de L^2 : une représentation AR sur la base de ε est

$$X_t = \sum_{s \in S: s \neq t} a_{t,s} X_s + \varepsilon_t;$$

Soit A la matrice $n \times n$ définissant l'AR : $\forall s, t \in S$, $A_{s,s} = 1$, $A_{t,s} = -a_{t,s}$ si $t \neq s$. L'auto-régression,

$$AX = \varepsilon,$$

est bien définie si A est inversible. Si $\Gamma = Cov(\varepsilon)$, $\Sigma = Cov(X) = A^{-1}\Gamma^t(A^{-1})$.

Choisissons pour ε un BBf de variance 1 ($\Gamma = I_n$), et notons $<$ un ordre total d'énumération des points de S . Si X est centré, de covariance Σ inversible, alors X admet une unique représentation AR causale relative à ε et à l'ordre $<$; cette représentation causale est associée à la matrice A triangulaire inférieure de la factorisation de Cholesky $\Sigma = {}^tAA$. Puisque Σ , comme A , dépend de $n(n+1)/2$ paramètres, le modèle AR causal est bien identifié. Les représentations AR équivalentes s'écrivent $\tilde{A}X = \eta$, où, pour P orthogonale, $\eta = P\varepsilon$ (η est encore un BBf de variance 1) et $\tilde{A} = PA$.

2.2.2 Représentation markovienne CAR

Une représentation CAR s'écrit en terme d'espérance conditionnelle (de prédiction) linéaire :

$$\forall t \in S : X_t = \sum_{s \in S: s \neq t} c_{t,s} X_s + e_t \text{ avec} \quad (2.5)$$

$$\text{Var}(e_t) = \sigma_t^2 > 0, \text{Cov}(X_t, e_s) = 0 \text{ si } t \neq s. \quad (2.6)$$

Notons C la matrice de coefficients $C_{s,s} = 0$ et $C_{t,s} = c_{t,s}$ si $s \neq t$, D la matrice diagonale de coefficients $D_{t,t} = \sigma_t^2$. Les équations de Yule-Walker,

$$\Sigma = (I - C)^{-1} D$$

s'obtiennent en multipliant (2.5) par X_s , $s \in S$, puis en prenant l'espérance. Si $\Sigma^{-1} = Q$ existe, les paramètres CAR vérifient donc :

$$D^{-1}(I - C) \text{ est symétrique : } \forall t, s \in S : c_{t,s} \sigma_s^2 = c_{s,t} \sigma_t^2 \quad (2.7)$$

Un algorithme d'estimation d'un CAR devra donc intégrer ces contraintes ; si X est stationnaire, on reparamétrera en $c_{t-s} = c_{t,s} = c_{s-t}$ pour $t \neq s$. Sous hypothèse gaussienne, le modèle (2.5) est complètement spécifié.

2.2.3 Champ gaussien markovien.

Soit X un champ gaussien sur $S : X \sim \mathcal{N}_n(\mu, \Sigma)$. Supposons que S soit muni d'un graphe \mathcal{G} symétrique et sans boucle, $\langle s, t \rangle$ signifiant que t et s sont voisins pour \mathcal{G} . On dira que X est un champ \mathcal{G} -markovien si, notant $Q = (q_{s,t}) = \Sigma^{-1}$, alors $q_{st} = 0$ sauf si $\langle s, t \rangle$. Dans ce cas, la loi conditionnelle $\mathcal{L}(X_t | X_s, s \neq t)$ suit le modèle CAR :

$$\forall t \in S : X_t - \mu_t = -q_{t,t}^{-1} \sum_{s: \langle t, s \rangle} q_{t,s} (X_s - \mu_s) + e_t, \text{Var}(e_t) = q_{t,t}^{-1} \quad (2.8)$$

Comme on le verra au §3.2.2, ces modèles sont des modèles de Gibbs.

Graphe markovien \mathcal{G} d'un SAR.

Le SAR gaussien $AX = \varepsilon$ où $\varepsilon \sim \text{BBG}(\sigma^2)$ existe si A^{-1} existe. Si $\Sigma^{-1} = Q = \sigma^{-2}({}^t A A)$, et si \mathcal{R} est le graphe de la représentation SAR : $\langle t, s \rangle_{\mathcal{R}} \iff a_{t,s} \neq 0$, la représentation CAR (2.8) associée vérifie :

1. Les coefficients du CAR sont : $c_{t,s} = -q_{t,s}/q_{t,t}$ où $q_{t,s} = \sum_{l \in S} a_{l,t} a_{l,s}$;
2. Le graphe \mathcal{G} de la représentation markovienne CAR de X est :

$$\langle t, s \rangle_{\mathcal{G}} \iff \begin{cases} \text{soit } \langle t, s \rangle_{\mathcal{R}}, \text{ soit } \langle s, t \rangle_{\mathcal{R}} \\ \text{soit } \exists l \in S \text{ t.q. } \langle l, t \rangle_{\mathcal{R}} \text{ et } \langle l, s \rangle_{\mathcal{R}} \end{cases}$$

\mathcal{G} est *non-orienté* de portée "double" de celle de \mathcal{R} (cf. Fig. 2.2).

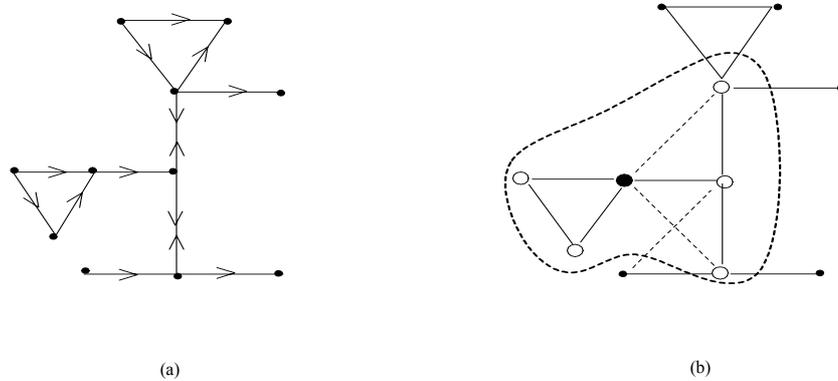


FIG. 2.2 – (a) Graphe orienté \mathcal{R} d'un SAR; (b) CAR associé, graphe \mathcal{G} (nouvelles liaisons en pointillé) et voisinage conditionnel (o) du point (•).

Exemple 2.3 Représentation CAR d'un SAR aux plus proches voisins.

Soit $W = (w_{t,s})_{t,s \in S}$ une matrice de poids connus mesurant l'influence de s sur t , avec, pour tout t , $w_{t,t} = 0$. Par exemple, W est la matrice de contiguïté spatiale constituée de 1 si s a une influence sur t , de 0 sinon. Une modélisation classique est le *SAR* à un paramètre ρ :

$$\begin{aligned} X_t &= \rho \sum_{s:s \neq t} w_{t,s} X_s + \varepsilon_t, t \in S, \text{ soit :} \\ X &= \rho W X + \varepsilon, \end{aligned}$$

où ε est un *BBf*(σ_ε^2). X est défini dès que $A = I - \rho W$ est inversible. Le *CAR* associé est,

$$\begin{aligned} X_t &= -q_{t,t}^{-1} \left(\sum_{s:s \neq t} q_{s,t} X_s \right) + e_t, \text{ avec } \text{Var}(e_t) = q_{t,t}^{-1} \text{ et} \\ Q &= \Sigma_X^{-1} = \sigma_\varepsilon^{-2} \{I - \rho(W + {}^t W) + \rho^2 {}^t W W\}. \end{aligned}$$

Notons : $D = I + \rho^2 \text{Diag}({}^t W W)$, $[M]$ la matrice nulle sur la diagonale et égale à M ailleurs. Alors $\text{Var}(e_t) = q_{t,t}^{-1} = \sigma_\varepsilon^2 d_{t,t}^{-1}$ et la prédiction linéaire optimale de X est :

$$\hat{X} = D^{-1} \{ \rho({}^t W + W) - \rho^2 [{}^t W W] \} X.$$

2.2.4 AR en présence d'exogènes : modèle *SARX*

Supposons que $X = (X_1, X_2, \dots, X_n)$ soit la variable endogène observée sur $S = \{1, 2, \dots, n\}$ et que Z est la matrice $n \times p$ de conditions exogènes observables $z_i \in \mathbb{R}^p$, $i = 1, n$. Les modèles *SARX*, très utilisés en économétrie spatiale, sont des modèles mixtes spatiaux de régression-Auto-régression (Anselin, [3]).

SARX général.

Un modèle *SARX* (X pour eXogène) propose une même structure *SAR* agissant séparément sur l'endogène X et l'exogène Z . Nous limitant aux *SAR* à un paramètre associé à une matrice de poids W , ces modèles s'écrivent :

$$X = \rho W X + Z\beta + WZ\gamma + \varepsilon, \rho \in \mathbb{R}, \beta, \gamma \in \mathbb{R}^p, \quad (2.9)$$

où ε est un *BBf*(σ^2). L'existence de (2.9) est assurée dès que $(I - \rho W)$ est inversible.

Les covariables expliquant X_s sont de trois types : l'endogène "retardée" spatialement $X_{\partial s}$, l'exogène "brute" $Z_s \in \mathbb{R}^p$ et l'exogène "retardée" spatialement $(WZ)_s$. On notera que les covariables Z et WZ sont autonomes dans leurs influences sur X . L'espérance et la covariance inverse de X valent :

$$E(X) = (I - \rho W)^{-1} (Z\beta + WZ\gamma) \text{ et } \sigma^2 \Sigma^{-1} = (I - \rho {}^t W)(I - \rho W). \quad (2.10)$$

La vraisemblance de X s'en déduit facilement dans le cas gaussien.

Deux sous-modèles sont classiquement utilisés.

Sous-modèle à facteurs communs.

Encore appelé *modèle de Durbin spatial*, c'est le sous-modèle *SAR* sur la variable recentrée,

$$(I - \rho W)(X - Z\beta) = \varepsilon$$

C'est aussi le modèle (2.9) sous la contrainte : $\gamma = -\rho\beta$. Ici, $E(X) = Z\beta$ et la variable recentrée suit un *SAR*(ρ, W). La covariance spatiale de X est identique à celle (2.10) du *SARX*.

Modèle à décalage spatial.

C'est le modèle *SARX* sans la covariable exogène WZ de voisinage ($\gamma = 0$) :

$$X = \rho W X + Z\beta + \varepsilon.$$

D'espérance $E(X) = (I - \rho W)^{-1} Z\beta$, ce modèle garde la structure de covariance (2.10) d'un *SARX*.

Si X est gaussien, on estimera ces différents modèles par maximum de vraisemblance.

Exemple 2.4 *Part du produit agricole brut consommée localement.*

Cliff et Ord ([13]; données `eire` du package `spdep`) ont analysé la variabilité spatiale de la part X du produit agricole brut consommée sur place pour les 26 comtés S de l'Irlande. Ces pourcentages sont mis en rapport avec un indice Y mesurant l'accessibilité routière du comté (cf. Fig 2.3-b). Une étude graphique préliminaire montre qu'un modèle affine pour la moyenne,

$$X_s = \beta_0 + \beta_1 Y_s + \varepsilon_s, \quad s \in S \quad (2.11)$$

est raisonnable. L'analyse des résidus de (2.11) estimés par moindres carrés ordinaires révèle une importante corrélation spatiale. Celle-ci, jointe à la covariable Y , va être modélisée par un *SARX* à l'aide d'une matrice de contiguïté $W = (w_{t,s})_{t,s \in S}$.

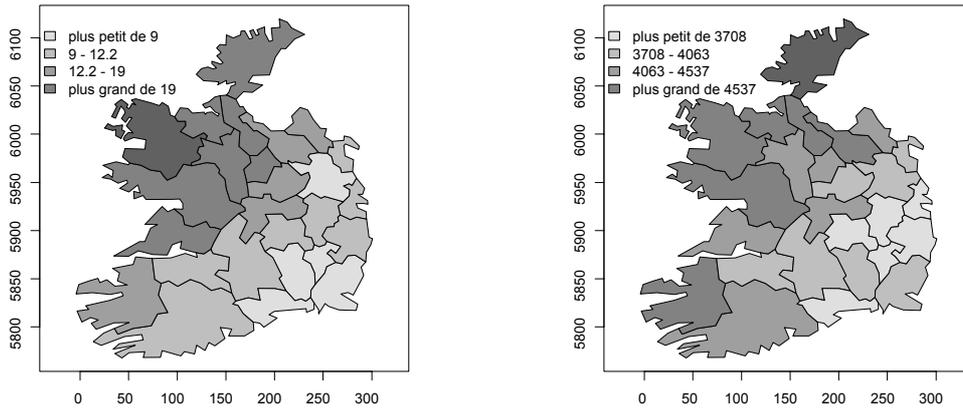


FIG. 2.3 – (a) Part X du produit agricole brut consommée sur place dans les 26 comtés de l'Irlande (données `eire` du package `spdep`); (b) indice Y d'accessibilité routière.

Pour le modèle à décalage spatial, $(X_t, t \in \partial S)$ est une exogène au même titre que Y_s , ε étant un BBG(σ^2) :

$$X_s = \beta_0 + \beta_1 Y_s + \gamma \sum_{t \in S} w_{s,t} X_t + \varepsilon_s. \quad (2.12)$$

Un modèle plus général à résidus *SAR* est $X_s = \beta_0 + \beta_1 Y_s + \gamma \sum_{t \in S} w_{s,t} X_t + \varepsilon_s$ où $\varepsilon_s = \lambda \sum_{t \in S} w_{s,t} \varepsilon_t + \eta_s$.

2.3 Auto-corrélation spatiale latticielle

On considère X un champ réel sur un réseau discret S , non nécessairement régulier, muni d'un *graphe d'influence* \mathcal{R} , (i, j) signifiant que i a une influence sur j , $j \neq i$. D'autre part on se donne une *matrice de poids* positifs bornés $W = \{w_{ij}, (i, j) \in \mathcal{R}\}$ quantifiant ce degré d'influence, avec $w_{ii} = 0$ et $w_{ij} = 0$ si $(i, j) \notin \mathcal{R}$, par exemple $w_{ij} = g(\|i - j\|)$ où g est positive décroissante. Suivant le problème étudié, w_{ij} peut intégrer d'autres facteurs exogènes tels que la part $i(j)$ de la frontière de i due à j voisine (environnement, écologie, épidémiologie), un indicateur des communications entre i et j (économie), etc. Comme une corrélation, l'autocorrélation spatiale mesure la ressemblance entre voisins. La littérature propose deux indices classiques.

2.3.1 L'indice de Moran

Soit X un processus de variance finie observé sur $D_n \subset S$, une partie de cardinal n . Supposons dans un premier temps que les X_i soient centrés, $\sigma_i^2 = \text{Var}(X_i)$. Une W -mesure de l'auto-covariance spatiale aux proches voisins est donnée par

$$C_n = \sum_{i,j \in D_n} w_{i,j} X_i X_j$$

Normalisée, cette auto-covariance fournit une auto-corrélation.

Sous l'hypothèse (H_0) : "les X_i sont indépendants",

$$\text{Var}(C_n) = \sum_{i,j \in D_n} (w_{ij}^2 + w_{ij} w_{ji}) \sigma_i^2 \sigma_j^2.$$

Ainsi, $I_n = \{Var(C_n)\}^{-1/2}C_n$ est centrée réduite. Si les σ_i^2 sont connus, un test asymptotique de (H_0) s'obtient si C_n satisfait un théorème central limite. Si pour tout i , $\sigma_i^2 = a_i\sigma^2$ où les $a_i > 0$ sont connus, il suffira d'estimer σ^2 par $\hat{\sigma}^2 = n^{-1} \sum_{i \in D_n} a_i^{-1} X_i^2$ et de considérer l'indice C_n renormalisé par son écart type estimé.

Si pour tout i , $E(X_i) = \mu$ et $Var(X_i) = \sigma^2$ sont estimés respectivement par $\bar{X} = n^{-1} \sum_{i \in D_n} X_i$ et $\hat{\sigma}^2 = (n-1)^{-1} \sum_{i \in D_n} (X_i - \bar{X})^2$, l'indice de Moran est

$$I_n^M = \frac{(n-1) \sum_{i,j \in D_n} w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\{\sum_{i,j \in D_n} w_{ij}\} \sum_{i \in D_n} (X_i - \bar{X})^2} \quad (2.13)$$

Par exemple, si X est stationnaire sur $S = \{1, 2, \dots, n\} \subset \mathbb{Z}$ et si W est la matrice de contiguité aux 2-ppv, $I_n^M = \hat{\rho}(1)$ est l'estimateur de la corrélation à distance 1.

Posant $s_{0n} = \sum_{i,j \in D_n} w_{ij}$, $s_{1n} = \sum_{i,j \in D_n} (w_{ij}^2 + w_{ij}w_{ji})$, on a

$$I_n^M = \frac{s_{1n}^{1/2}}{s_{0n}} \{\widehat{Var}(C_n)\}^{-1/2} C_n = \frac{s_{1n}^{1/2}}{s_{0n}} \hat{I}_n.$$

Comme s_{0n} et s_{1n} sont de l'ordre n , et que, sous (H_0) , $E(C_n) = 0$ et $\hat{\sigma}^2 \xrightarrow{\text{Pr}} \sigma^2$, on a,

$$\text{sous } (H_0) : E(I_n^M) = o(1) \quad \text{et} \quad Var(I_n^M) = \frac{s_{1n}}{s_{0n}^2} (1 + o(1)).$$

I_n^M est d'autant plus grand que des valeurs semblables apparaissent en des sites voisins, d'autant plus petit que des valeurs dissemblables apparaissent en des sites voisins. Dans le premier cas on parle d'agrégation, de coopération spatiale ou d'auto-corrélation positive, dans le deuxième, de répulsion, de compétition spatiale ou d'auto-corrélation négative. Le cas d'observations indépendantes correspond à une valeur de I_n^M proche de 0.

2.3.2 Test asymptotique d'indépendance spatiale

Soit (D_n) une suite strictement croissante de sous-ensembles finis de $S \subset \mathbb{R}^d$, W une matrice de poids bornés connus sur le graphe $\mathcal{R} \subset S^2$, de portée R et à voisinages $\partial i = \{j : j \text{ voisins de } i\}$ uniformément bornés,

$$\begin{aligned} W &= (w_{ij}, i, j \in S) \text{ avec } w_{ii} = 0 \text{ et } w_{ij} = 0 \text{ si } \|i - j\| > R, \\ \exists M &< \infty \text{ t.q. } \forall i, j : |w_{ij}| \leq M \text{ et } \#\partial i \leq M. \end{aligned}$$

On a le résultat de normalité asymptotique suivant (qui n'exige pas que X soit gaussien) :

Proposition 2.2 *Normalité de l'indice de Moran sous (H_0)*

Supposons que :

$$\exists \delta > 0 \text{ t.q. } \sup_{i \in S} E(|X_i|^{4+2\delta}) < \infty \text{ et } \liminf_n \frac{s_{1n}}{s_{0n}} > 0.$$

Alors, sous (H_0) , l'indice de Moran est asymptotiquement gaussien :

$$\frac{s_{0n}}{\sqrt{s_{1n}}} I_n^M \xrightarrow{\text{loi}} \mathcal{N}(0, 1)$$

Le cas d'un champ X gaussien.

Si les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, le résultat précédent peut être précisé car I_n^M est le quotient de deux formes quadratiques gaussiennes ([13]) :

$$\text{sous } (H_0) : E(I_n^M) = -\frac{1}{n} \quad \text{et} \quad \text{var}(I_n^M) = \frac{n^2 s_{1n} - n s_{2n} + 3 s_{0n}^2}{(n^2 - 1) s_{0n}^2} - \frac{1}{(n-1)^2}$$

où $s_{2n} = \sum_{i \in D_n} (w_i + w_{\cdot i})^2$, $w_i = \sum_{i \in D_n} w_{ij}$ et $w_{\cdot j} = \sum_{j \in D_n} w_{ij}$.

L'indice de Moran peut aussi être construit pour des données ordinales ou pour des données binaires [13].

2.3.3 L'indice de Geary

L'indice de Geary

$$I_n^G = \frac{(n-1) \sum_{i,j \in D_n} w_{ij} (X_{s_i} - X_{s_j})^2}{2 \{\sum_{i,j \in D_n} w_{ij}\} \sum_{i \in D_n} (X_i - \bar{X})^2}$$

mesure la dépendance spatiale de façon similaire à celle d'un variogramme : I_n^G est d'autant plus petit que des valeurs semblables apparaissent en des sites voisins. I_n^G est sensible aux différences importantes pour des couples

de points voisins là où I_M est sensible aux données extrêmes de X . Sous les conditions de la proposition précédente et sous (H_0) , I_n^G est asymptotiquement gaussien :

$$\sqrt{s_{0n}(2s_{1n} + s_{2n})^{-1}}(I_n^G - 1) \sim \mathcal{N}(0, 1).$$

Si de plus X est gaussien,

$$E(I_n^G) = 1, \quad \text{var}(I_n^G) = \frac{(2s_{1n} + s_{2n})(n-1) - 4s_{2n}}{2(n+1)s_{0n}}.$$

2.3.4 Test de permutation d'indépendance spatiale

D'une façon générale, la loi permutationnelle d'une statistique réelle $I(X)$ de $X = (X_i, i = 1, n)$, conditionnellement aux n valeurs observées $(x_i, i = 1, n)$, est la loi uniforme sur l'ensemble des valeurs $I_\sigma = I(x_\sigma)$ de I pour les $n!$ permutations σ de $\{1, 2, \dots, n\}$. Pour tester l'hypothèse d'indépendance (H_0) : "les X_i sont *i.i.d.*", sans connaître la loi des X_i , on utilisera la loi permutationnelle de l'indice de Moran (ou de l'indice de Geary). La justification du test est que, sous (H_0) , une permutation des $\{X_i\}$ ne change pas la loi globale de X :

$$(X_i, i = 1, n) \sim (X_{\sigma(i)}, i = 1, n).$$

Le niveau de significativité bilatéral (resp. unilatéral) associé est :

$$p_a = \frac{1}{n!} \sum_{\sigma} \mathbf{1}\{|I_{\sigma}| > a\} \quad (\text{resp. } p_a^* = \frac{1}{n!} \sum_{\sigma} \mathbf{1}\{I_{\sigma} > a\}).$$

Lorsque l'énumération de toutes les permutations n'est pas possible, on recourt à la méthode de Monte Carlo en choisissant au hasard, pour m assez grand, m permutations $\{\sigma_1, \sigma_2, \dots, \sigma_m\}$ pour lesquelles on calcule les valeurs I_{σ} et les seuils de Monte Carlo associés p_a^{MC} et p_a^{MC*} .

L'avantage d'une méthode permutationnelle est de fournir un test non asymptotique sans hypothèse de loi sur X et /ou quand la loi asymptotique de I_n^M n'est pas connue.

Si les n valeurs observées $(x_i, i = 1, n)$ sont différentes, l'espérance \mathbb{E}_P de l'indice de Moran I_n^M sous la loi permutationnelle vaut, sous (H_0) :

$$\mathbb{E}_P(I_n^M) = -\frac{1}{n-1}.$$

Exemple 2.5 *Pourcentage du groupe sanguin A dans 26 comtés d'Irlande.*

La figure 2.4 représente le pourcentage des individus de groupe sanguin A dans les 26 comtés de l'Irlande (cf. [13] et le jeu de données `eire` dans le package `spdep`). La figure 2.4-a montre une ressemblance claire entre voisins, deux comtés étant voisins s'ils ont une frontière commune (cf. Fig. 2.4-b). On calcule (cf. Tab. 2.1) les indices de Moran et Geary $t^a = (I^a - E(I^a))/\sqrt{\text{var}(I^a)}$, $a = M, G$ (cf. `spdep`) pour cette relation de voisinage et $w_{ij} = 1/|\partial i|$ pour $(i, j) \in \mathcal{R}$, 0 sinon, ainsi que les niveaux de significativité p^a (gaussien) et p_{MC}^a (pour la loi permutationnelle avec $m = 1000$). Les résultats confirment cette dépendance spatiale.

	Index	t^a	p^a	p_{MC}^a
Moran	0.554	4.663	0	0.001
Geary	0.380	-4.547	0	0.001

TAB. 2.1 – Taux du groupe sanguin A : indices de Moran et de Geary et statistiques associées.

D'autres tests de permutation d'indépendance spatiale.

Des statistiques autres que l'indice de Moran ou l'indice de Geary peuvent être retenues pour tester l'indépendance spatiale. Suivant Peyrard et altri [37], considérons un champ $X = \{X_{i,j}, (i, j) \in S\}$ observé sur la grille régulière $S = \{(i, j) : i = 1, \dots, I; j = 1, \dots, J\}$, X repérant une variable quantitative ou qualitative ordonnée mesurant le degré de gravité d'une maladie d'une espèce végétale plantée dans un verger. Bien que le réseau S soit régulier, la maille n'est pas nécessairement carrée. Si la distance entre les lignes i est plus importante que celles entre les colonnes j , on pourra retenir comme statistique d'évaluation de la dépendance spatiale le variogramme à distance d le long de la ligne, estimé par :

$$\Gamma(d) = \hat{\gamma}(0, d) = \frac{1}{I(J-1)} \sum_{i=1}^I \sum_{j=1}^{J-d} (X_{i,j} - X_{i,j+d})^2, \quad d \geq 1$$

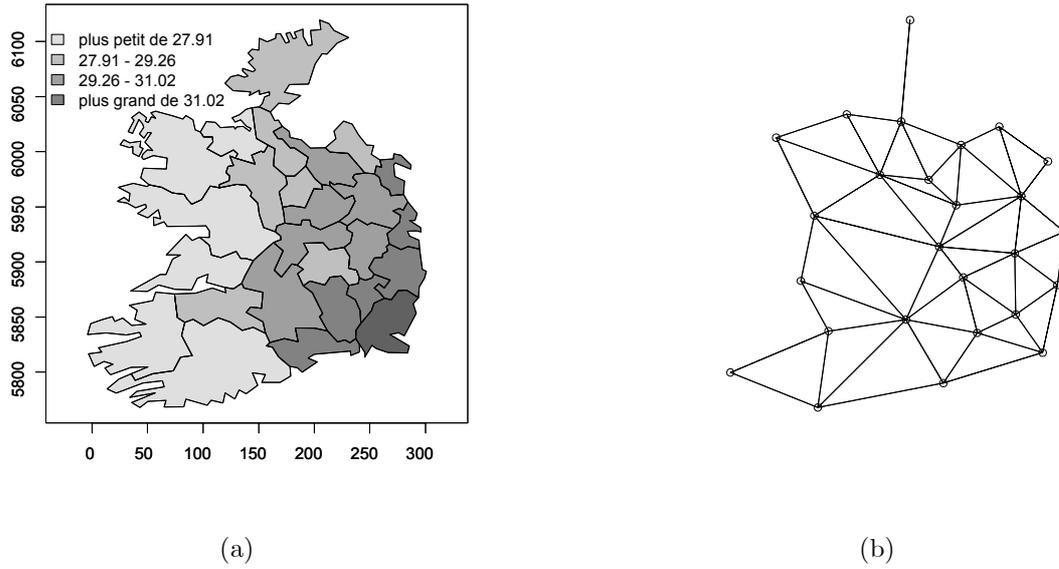


FIG. 2.4 – (a) Pourcentage de la population présentant le groupe sanguin A dans les 26 comtés d'Irlande; (b) graphe d'influence \mathcal{R} (symétrique) sur les 26 comtés.

Pour tester l'*indépendance totale* (H_0^G) entre tous les sites, on considérera les permutations Σ sur l'ensemble S de tous les sites et les valeurs associées,

$$\Gamma_{\Sigma}(d) = \frac{1}{I(J-1)} \sum_{i=1}^I \sum_{j=1}^{J-d} (X_{\Sigma(i,j)} - X_{\Sigma(i,j+d)})^2, \quad d \geq 1$$

Le test de (H_0^G) peut être réalisé pour toute distance d en construisant une bande de confiance pour $\gamma(0, d)$ avec des quantiles prédéterminés sur la base de m permutations globales Σ choisies au hasard (cf. Fig. 2.6-a) : si $\hat{\gamma}(0, d)$ sort significativement de cette bande, on rejettera (H_0^G).

Pour tester l'*indépendance entre les lignes* i , (H_0^L), on retiendra comme indicateurs les variogrammes à distance d le long des colonnes et on les comparera à ceux qui ont été calculés pour des permutations σ portant sur les seuls indices de ligne i :

$$\Delta(d) = \hat{\gamma}(d, 0) = \frac{1}{J(I-1)} \sum_{i=1}^{I-d} \sum_{j=1, J} (X_{i+d, j} - X_{i, j})^2, \quad d \geq 1$$

$$\hat{\gamma}_{\sigma}(d, 0) = \frac{1}{J(I-1)} \sum_{i=1}^{I-d} \sum_{j=1, J} (X_{\sigma(i+d, j)} - X_{\sigma(i, j)})^2, \quad d \geq 1$$

Le test de (H_0^L) est alors construit sur la base de m permutations de lignes σ choisies au hasard (cf. Fig. 2.6-b). D'autres choix de permutations sont possibles pour tester cette hypothèse.

Exemple 2.6 Dépérissement de la lavande

[37] examinent d'autres méthodes de permutations appliquées à d'autres tests (homogénéité spatiale, deux champs indépendants, indépendances conditionnelles) afin d'explorer des données sur une grille régulière. Considérons l'une des études.

Une parcelle comporte $I = 11$ lignes de $J = 80$ plants de lavande (colonnes). Deux plants consécutifs d'une même ligne sont espacés de 20 cm alors que les lignes de culture sont espacées de 1 mètre. La parcelle est atteinte de dépérissement (le mycoplasme) et on note de 0 (sain, pixel noir) à 5 (fortement atteint, pixel blanc) l'état sanitaire de chaque plant (cf. Fig. 2.5).

Les résultats des tests sont donnés pour $m = 200$ permutations et des niveaux de confiance 2.5% et 97.5%. Pour le test d'indépendance globale (H_0^G) (cf. Fig. 2.6-a), $\Gamma(d)$ sort toujours de la bande de confiance construite : il n'y a pas indépendance globale. Pour le deuxième test (H_0^L) (cf. Fig. 2.6-b), $\Delta(d)$ sort de la bande de confiance pour $d = 1, 6$ et 7 , se plaçant à sa limite pour $d = 2, 3$ et 5 : on rejettera encore l'indépendance entre les lignes.

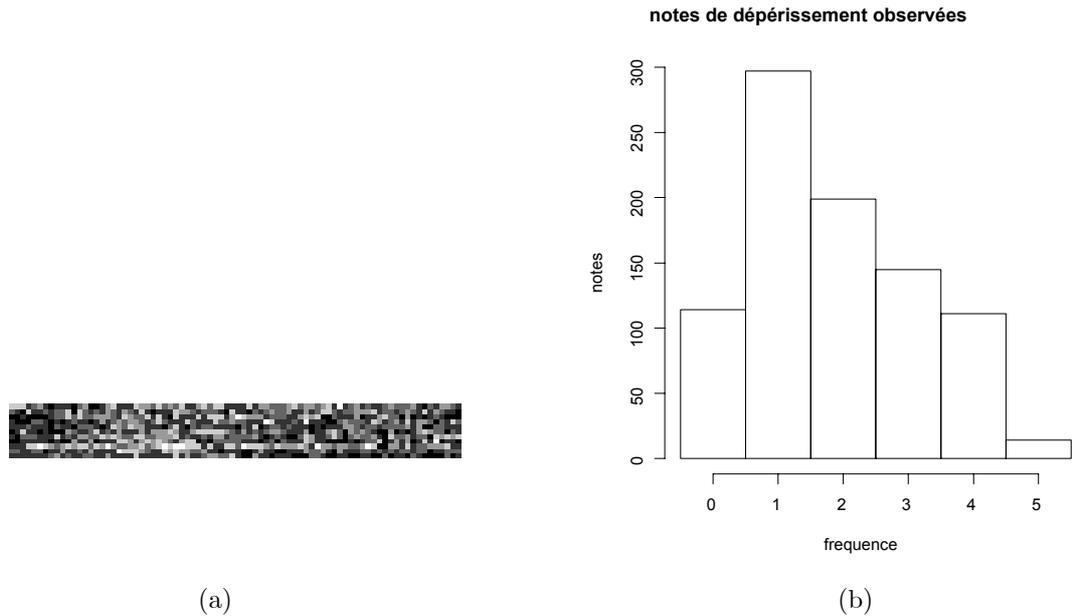


FIG. 2.5 – (a) Répartition spatiale des notes (de 1 à 5) de dépérissement de la lavande cultivée du plateau de Sault ; (b) histogramme du nombre de plants ayant une note de dépérissement. Source : [37].

2.4 Statistique d'un champ du second ordre

Supposons que X soit un champ réel du second ordre sur un réseau discret S . Nous examinerons d'abord la statistique d'un champ stationnaire sur $S = \mathbb{Z}^d$, puis celle d'un champ AR et enfin le cas d'une régression $X = \mathcal{Z}\delta + \varepsilon$ où le résidu ε est un champ spatial du second ordre. Des résultats asymptotiques sont présentés si X est gaussien.

2.4.1 Estimation d'un modèle stationnaire sur \mathbb{Z}^d

Soit X un champ réel sur \mathbb{Z}^d , centré et stationnaire dans L^2 . Pour simplifier, on supposera que l'observation $X(n)$ de X est faite sur le cube $D_n = \{1, 2, \dots, n\}^d$; le nombre d'observations est $N = n^d = \#(D_n)$. Les résultats présentés ici généralisent ceux connus des séries chronologiques à la dimension $d \geq 2$. Cependant, une différence apparaît concernant l'importance des *effets de bord* pour un processus spatial.

Covariance empirique et rabotage des données.

La covariance empirique en $k \in \mathbb{Z}^d$ vaut :

$$\hat{C}_n(k) = \frac{1}{N} \sum_{i, i+k \in D_n} X_i X_{i+k}, \quad k \in \mathbb{Z}^d. \quad (2.14)$$

Le choix d'une normalisation commune par N^{-1} pour tous les k fait de $\hat{C}_n(\cdot)$ une fonction s.d.p., de support $\Delta(n) = D_n - D_n = \{i - j : i, j \in D_n\}$.

En statistique spatiale, les *effets de bord* de D_n augmentent avec la dimension d du réseau : en effet, à $N = n^d$ nombre fixé d'observations, le pourcentage de points au bord de D_n , en $dN^{-1/d}$, augmente avec d . Pour les covariances empiriques, un calcul simple montre que :

$$\lim_n \sqrt{N} E(\hat{C}_n(k) - C(k)) = \begin{cases} 0 & \text{si } d = 1 \\ -\{|k_1| + |k_2|\} & \text{si } d = 2 \\ +\infty & \text{si } d > 2 \text{ et } k \neq 0 \end{cases}.$$

Sur \mathbb{Z} , l'effet de bord, en N^{-1} , est sans conséquence sur le biais asymptotique de $\sqrt{N}(\hat{C}_n(k) - C(k))$; sur \mathbb{Z}^2 , il est en $N^{-1/2}$ et devient significatif, du même ordre que la vitesse $N^{-1/2}$; pour $d \geq 3$, ce biais est dominant.

Pour éliminer ce biais, une solution consisterait à remplacer la normalisation en N^{-1} par $N(k)^{-1}$ où $N(k)$ est le nombre de couples (X_i, X_{i+k}) intervenant dans (2.14); mais on perd alors la propriété de s.d.p. de la covariance estimée.

Pour remédier à ces inconvénients, Tuckey [49] définit le *rabotage* des données $X^w(n)$ au bord de D_n (*tapered data*) en utilisant un rabot w . Pour $d = 1, 2, 3$ et pour un rabotage convenable, ces estimateurs ne présentent pas

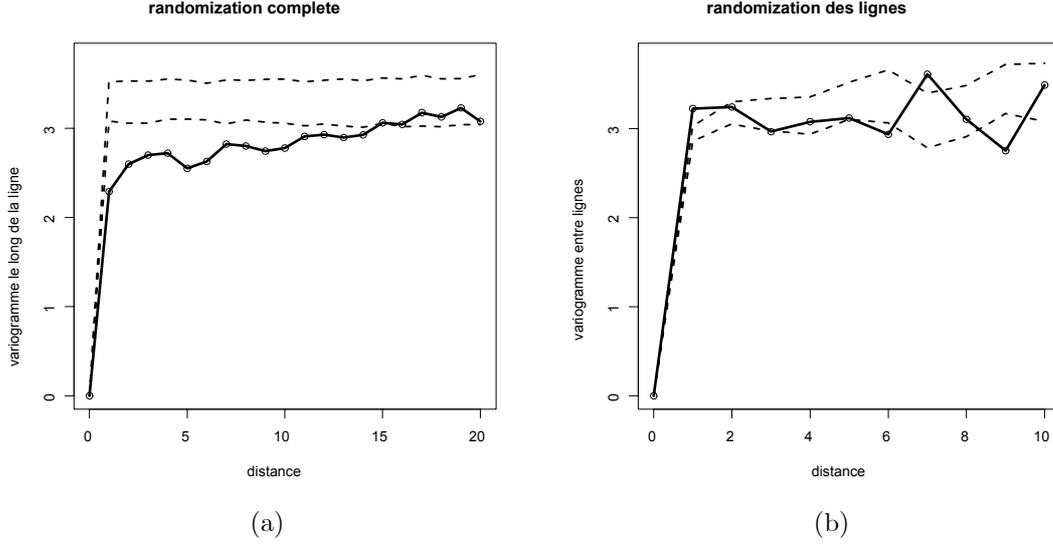


FIG. 2.6 – (a) Test d’indépendance totale par permutation ; (b) Test d’indépendance entre lignes. Les courbes en pointillé correspondent aux limites de confiance à 2.5% et 97.5%.

les défauts soulignés précédemment. Le rabotage améliore aussi l’analyse statistique car il diminue le poids des observations de bord qui souvent s’écartent du modèle.

Définissons ce rabotage. Soit $w : [0, 1] \rightarrow [0, 1]$, $w(0) = 0$, $w(1) = 1$, une fonction croissante de classe \mathcal{C}^2 (le profil de rabotage). Le w -rabot rabotant $100(1 - \rho)\%$ des données de bord est défini, pour $0 \leq \rho \leq 1$, par :

$$h(u) = \begin{cases} w\left(\frac{2u}{\rho}\right) & \text{si } 0 \leq u \leq \frac{\rho}{2} \\ 1 & \text{si } \frac{\rho}{2} \leq u \leq \frac{1}{2} \end{cases}$$

et $h(u) = h(1 - u)$ pour $1/2 \leq u \leq 1$. Par exemple, le rabot de Tuckey-Hanning est $w(u) = 2^{-1}(1 - \cos \pi u)$. Les données rabotées $X^w(n)$ sont :

$$X_i^w = a_n(i)X_i, \text{ où } a_n(i) = \prod_{k=1}^d h\left(\frac{i_k - 0.5}{n}\right), \quad i \in D_n.$$

La covariance empirique rabotée \widehat{C}_n^w s’obtient alors à partir des données rabotées $X^w(n)$: pour $d = 1, 2, 3$ et un choix $\rho_n = o(n^{-1/4})$, le biais d’un estimateur “raboté” n’est plus contribuant ([18] ; [28]).

Pseudo-vraisemblance “gaussienne”.

Supposons que X soit paramétré via sa densité spectrale f_θ , où $\theta \in \Theta$, un compact de \mathbb{R}^k . Le spectrogramme raboté $I_n^w(\lambda) = 2\pi^{-d} \sum_{k=-n+1}^{n-1} \widehat{C}_n^w(k) e^{i\langle \lambda, k \rangle}$ n’est autre que l’estimation de $f_\theta(\lambda)$ de coefficients de Fourier estimés par les covariances empiriques rabotées. Si X est gaussien, Whittle [52] a montré qu’une bonne approximation de la log-vraisemblance de $X(n)$ est $-2U_n(\theta)$ où

$$U_n(\alpha) = \frac{1}{(2\pi)^d} \int_{T^d} \left\{ \log f_\alpha(\lambda) + \frac{I_n^w(\lambda)}{f_\alpha(\lambda)} \right\} d\lambda, \quad T = [0, 2\pi[$$

Si X n’est pas gaussien, la minimisation de U_n conduit encore à une bonne estimation (non efficace) de θ sous des conditions générales assez faibles : $-2U_n(\theta)$ est appelée la pseudo-vraisemblance “gaussienne” de Whittle.

Soit $\widehat{\theta}_n = \text{Arg min}_{\alpha \in \Theta} U(\alpha)$. L’étude asymptotique de $\widehat{\theta}_n$ requiert l’existence des moments d’ordre 4 du processus ainsi que des conditions de faible dépendance pour le champ X . Pour simplifier, supposons que X soit gaussien et notons :

$$\Gamma(\theta) = (2\pi)^{-d} \int_{T^d} \{(\log f)_\theta^{(1)t} (\log f)_\theta^{(1)}\}(\lambda) d\lambda$$

$$e(h) = \left[\int_0^1 h^4(u) du \right] \left[\int_0^1 h^2(u) du \right]^{-2}$$

- (W1) : Il existe $0 < m \leq M < \infty$ t.q. $m \leq f_\theta \leq M$.
- (W2) : θ est identifiable, c.a.d. $\theta \mapsto f_\theta(\cdot)$ est injective.

- (W3) : $f_\theta(\lambda)$ est analytique en λ ; $f_\theta^{(2)}$ existe et est continue en (θ, λ) .

Théorème 2.1 Si X gaussien vérifie (W), $\hat{\theta}_n \xrightarrow{\text{Pr}} \theta$. De plus, si $\Gamma^{-1}(\theta)$ existe et si $d \leq 3$,

$$n^{\frac{d}{2}}(\hat{\theta}_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}_p(0, e^d(h)\Gamma^{-1}(\theta))$$

Les hypothèses (W) sont vérifiées pour un ARMA identifiable. Le résultat est vrai sans rabotage ($e(h) = 1$) pour $d = 1$. Pour les dimensions $d = 2$ et 3 , on peut choisir un h_n -rabotage tel que $e(h_n) \downarrow 1$.

Le test d'une sous-hypothèse $(H_0) : C(\theta) = 0$, où $C : \mathbb{R}^k \rightarrow \mathbb{R}^l$ est régulière de classe \mathcal{C}^2 , s'obtient en constatant que $C(\hat{\theta}_n)$ est asymptotiquement gaussien et centré sous (H_0) . Si X n'est pas gaussien mais faiblement dépendant, il y a toujours convergence gaussienne, la variance asymptotique faisant intervenir les moments d'ordre 4 de X [28].

2.4.2 Estimation d'un modèle auto-régressif

Un champ réel centré de L^2 observé sur $S = \{1, 2, \dots, n\}$ est un vecteur aléatoire centré $X = X(n) = {}^t(X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ de covariance $\Sigma = \text{cov}(X)$. On va examiner les deux situations classiques d'AR :

1. L'auto-régression simultanée (SAR)

$$X_i = \sum_{j \in S: j \neq i} a_{i,j} X_j + \varepsilon_i,$$

où ε est un BB de variance σ^2 . Si A est la matrice $A_{i,i} = 0$ et $A_{i,j} = a_{i,j}$ si $i \neq j$, $i, j \in S$, alors $X = AX + \varepsilon$ est bien défini dès que $(I - A)$ est inversible, de covariance $\Sigma = \sigma^2 \{ {}^t(I - A)(I - A) \}^{-1}$. Il faudra s'assurer de l'identifiabilité des paramètres.

2. L'auto-régression conditionnelle (CAR)

$$X_i = \sum_{j: j \neq i} c_{i,j} X_j + e_i,$$

où $\text{Var}(e_i) = \sigma_i^2 > 0$ et $\text{Cov}(X_i, e_j) = 0$ si $i \neq j$; si C est la matrice $C_{i,i} = 0$ et $C_{i,j} = c_{i,j}$ si $i \neq j$, $i, j \in S$, et D la matrice diagonale de coefficients $D_{i,i} = \sigma_i^2$, $(I - C)X = e$ et $\Sigma = D(I - C)^{-1}$. Il faudra imposer les contraintes : $\forall i \neq j$, $c_{i,j}\sigma_j^2 = c_{j,i}\sigma_i^2$. Si $\forall i$, $\sigma_i^2 = \sigma^2$ (resp. si X est stationnaire), ces contraintes sont : $\forall s, t$, $c_{t,s} = c_{s,t}$ (resp. $c_{t-s} = c_{s-t}$).

Un SAR (resp. un CAR) est un modèle linéaire en A (resp. en C). Un SAR est un CAR, avec $C = {}^tA + A + {}^tAA$, sans linéarité en A . Deux méthodes d'estimation sont envisageables :

1. le maximum de vraisemblance si X est gaussien, sera présenté au §2.4.4 dans le contexte plus général d'une régression spatiale. Si le résidus est un AR, sa covariance prendra une forme paramétrique spécifique. Si le modèle n'est pas gaussien, cette fonctionnelle "gaussienne" restera une bonne fonctionnelle d'estimation.
2. l'estimation par Moindres Carrés Ordinaires (MCO), naturelle puisque qu'un AR est un modèle linéaire en $\theta = A$ pour un SAR, en $\theta = C$ pour un CAR. Les MCO minimisent $\|\varepsilon(\theta)\|^2$ pour un SAR, $\|e(\theta)\|^2$ pour un CAR. Un avantage des MCO est de fournir un estimateur explicite facile à calculer.

Cependant, l'estimation MCO d'un SAR n'est pas consistante !

Proposition 2.3 L'estimation des MCO d'un modèle SAR est non-consistante. Celle d'un CAR est consistante.

Démonstration : C'est un résultat classique pour un modèle $Y = \mathcal{X}\theta + \varepsilon$ où les ε sont corrélés aux régresseurs \mathcal{X} (i.e. équations simultanées en économétrie). Pour s'en convaincre, considérons le SAR sur \mathbb{Z}^1 , $X_t = a(X_{t-1} + X_{t+1}) + \varepsilon_t$, $|a| < 1/2$, observé sur $\{1, 2, \dots, n\}$: si r est la covariance de X , l'estimation des MCO de a est

$$\hat{a}_n = \frac{\sum_{t=2, n} X_t(X_{t-1} + X_{t+1})}{\sum_{t=2, n} (X_{t-1} + X_{t+1})^2}$$

et $\hat{a}_n \rightarrow [r_1 - a(r_0 + r_2)] / (r_0 + r_2) \neq a$ si $a \neq 0$.

La consistance des MCO pour un CAR résulte du fait que les résidus conditionnels e_i sont décorrélés des co-variables explicatives $\mathcal{X}(i) = \{X_j, j \neq i\}$. □

Pour un CAR gaussien, l'estimation des MCO coïncide avec le maximum de la pseudo-vraisemblance conditionnelle de X (cf. §3.5.2).

2.4.3 Maximum par vraisemblance pour un champ gaussien centré

Si X est un vecteur gaussien centré de \mathbb{R}^n et si $Cov(X) = \Sigma(\theta)$, l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ maximise

$$l_n(\theta) = -\frac{1}{2} \left\{ \log |\Sigma(\theta)| + \frac{1}{2} {}^t X \Sigma^{-1}(\theta) X \right\} \quad (2.15)$$

Le calcul de $\hat{\theta}_n$ nécessitant une méthode d'optimisation itérative, la difficulté réside dans le calcul de $|\Sigma(\theta)|$ et de ${}^t X \Sigma^{-1}(\theta) X$. Pour une AR , on dispose de la forme de Σ^{-1} et de ${}^t X \Sigma^{-1}(\theta) X$:

1. pour un SAR :

$${}^t X \Sigma^{-1}(\theta) X = \sigma^{-2} \|(I - A)X\|^2.$$

L'estimation du MV est $\hat{\sigma}^2 = \sigma^2(\hat{\theta}) = n^{-1} \|(I - A(\hat{\theta}))X\|^2$ où $\hat{\theta}$ minimise $-2n^{-1} \log |I - A(\theta)| + \log(\sigma^2(\theta))$.

2. Pour un CAR de variance résiduelle constante σ_e^2 ,

$${}^t X \Sigma^{-1}(\theta) X = \sigma_e^{-2} \{ {}^t X (I - C) X \}.$$

L'estimation du MV est $\hat{\sigma}_e^2 = \sigma_e^2(\hat{\theta}) = n^{-1} {}^t X (I - A(\hat{\theta})) X$ où $\hat{\theta}$ minimise $-2n^{-1} \log |I - A(\theta)| + \log(\sigma_e^2(\theta))$.

Le calcul itératif de $|\Sigma(\theta)|$ est simple pour certains modèles : par exemple, pour le SAR à un paramètre, $(I - \rho W)X = \varepsilon$, où W est une matrice symétrique telle que $I - \rho W$ est inversible, et ε est un BB,

$$|\Sigma(\theta)| = \sigma^{2n} |I - \rho W|^{-2} \quad \text{et} \quad |I - \rho W| = \prod_{i=1}^n (1 - \rho w_i),$$

où $w_1 \leq w_2 \leq \dots \leq w_n$ sont les valeurs propres de W , valeurs qui ne sont pas recalculées au cours des itérations. La covariance estimée $\Sigma(\hat{\rho})$ sera d.p. si : $\hat{\rho} < w_n^{-1}$ si $0 \leq w_1$, $\hat{\rho} > w_1^{-1}$ si $w_n \leq 0$, $w_1^{-1} < \hat{\rho} < w_n^{-1}$ si $w_1 < 0 < w_n$.

Pour le CAR ,

$$(I - \rho W)X = e \quad \text{et} \quad |\Sigma(\theta)| = \sigma_e^{2n} |I - \rho W|^{-1}.$$

2.4.4 Estimation d'une régression spatiale.

On s'intéresse ici à l'estimation conjointe des paramètres de la moyenne et de ceux du résidu ε . Une régression spatiale (1.9) s'écrit, pour $\varepsilon = (\varepsilon_s)$ un champ du second ordre,

$$X = Z\delta + \varepsilon \quad (2.16)$$

Z , la matrice $n \times p$ des n conditions exogènes $z_i \in \mathbb{R}^p$, est observée, $\delta \in \mathbb{R}^p$ est le paramètre au premier ordre à estimer. Des exemples de modèle pour la moyenne sont :

1. $E(X_{x,y})$ au site $(x, y) \in \mathbb{R}^2$ est un polynôme de degré deux en (x, y) ($p = 6$) ;
2. $E(X_{i,j}) = \mu + \alpha_i + \beta_j$ est le modèle additif d'analyse de la variance à deux facteurs, $(i, j) \in \{1, 2, \dots, I\} \times \{1, 2, \dots, J\}$ ($p = I + J - 1$) ;
3. pour z_i une variable exogène observable, $E(X_i) = {}^t g(z_i) \delta$ où $g : E \rightarrow \mathbb{R}^p$ est connue ;
4. $E(X_i)$ est un modèle d'analyse de covariance combinant un modèle de régression et un modèle d'analyse de la variance.

Estimation de la régression par MCO

L'estimation de δ par MCO des résidus ε (2.16) est :

$$\tilde{\delta} = ({}^t Z Z)^{-1} {}^t Z X$$

Cet estimateur est sans biais, de variance, si $\Sigma = Var(\varepsilon)$,

$$Var(\tilde{\delta}) = \Delta = ({}^t Z Z)^{-1} {}^t Z \Sigma Z ({}^t Z Z)^{-1}$$

Si X est gaussien, $\tilde{\delta} \sim \mathcal{N}_q(\delta, \Delta)$. Il faut noter :

1. $\tilde{\delta}$ n'est pas efficace (c'est l'estimateur des MCG qui l'est, cf. (2.18)) ;
2. si, pour tout i , $Var(\varepsilon_i) = \sigma^2$, l'estimateur habituel de σ^2 basé sur les résidus estimés $\tilde{\varepsilon} = X - Z\tilde{\delta}$ est biaisé ;
3. la statistique habituelle du test F d'une sous-hypothèse sur δ déduite de cette estimation ne suit pas en général une loi de Fisher.

4. l'intérêt des MCO est d'être un bon estimateur n'exigeant pas de connaître la structure spatiale Σ : c'est un *bon estimateur* pour démarrer une procédure itérative d'estimation (MCQG (cf. §2.4.4; MV pour un champ de Gibbs, cf. §4.7) .

Si l'estimation $\tilde{\delta}$ des MCO est consistante, les MCO permettent de construire un algorithme simple d'estimation de la covariance Σ , qui, à son tour, permet de rendre "réalisable" la procédure des MCG. Donnons des conditions qui assurent la consistance des MCO en même temps qu'elles précisent la vitesse de consistance de cette estimation. Si A est une matrice symétrique s.d.p., on note $\lambda_M(A)$ (resp. $\lambda_m(A)$) la plus grande (resp. la plus petite) valeur propre de A . Faisant dépendre de n , le nombre d'observations, l'écriture des matrices Z de (2.16) et $\Sigma = cov(\varepsilon)$, on a la propriété suivante :

Proposition 2.4 *L'estimateur $\tilde{\delta}_n$ des MCO est consistant sous les deux conditions suivantes :*

- (i) $\lambda_M(\Sigma_n)$ est uniformément bornée en n ;
- (ii) $\lambda_m({}^t Z_n Z_n) \rightarrow \infty$ si $n \rightarrow \infty$ ou encore $({}^t Z_n Z_n)^{-1} \rightarrow 0$.

Commentaires :

1. La condition (ii) porte sur le dispositif expérimental exogène (Z_n). Si par exemple :

$$\frac{1}{n}({}^t Z_n Z_n) \longrightarrow Q, \quad (2.17)$$

où Q est d.p., (ii) est satisfaite et la vitesse de convergence, en variance, est en n^{-1} . Un dispositif expérimental qui réalise (2.17) est, si $n = m \times r$, Z_n la répétition de m dispositifs R_0 de dimension $r \times q$ de rang plein q ($Q = ({}^t R_0 R_0)/m$).

2. La condition (i) porte sur le modèle spatial Σ . Une condition suffisante assurant (i) est que la covariance γ_X de X vérifie, avec ou sans stationnarité de X :

$$\sup_{i \in S} \left\{ \sum_{j \in S} |\gamma_X(i, j)| \right\} < \infty$$

Cette condition est satisfaite pour une large classe de processus.

Moindres Carrés Quasi Généralisés (MCQG)

Considérons le modèle (2.16) et supposons que $\Sigma = Cov(X)$ soit connue. L'estimation de δ , linéaire, sans biais et de moindre variance est celle des Moindres Carrés Généralisés (MCG) :

$$\hat{\delta}_{MCG} = ({}^t Z \Sigma^{-1} Z)^{-1} {}^t Z \Sigma^{-1} X. \quad (2.18)$$

Si X est gaussien, les MCG coïncident avec le maximum de vraisemblance.

Pendant Σ est en général inconnue et X n'est pas toujours gaussien. La stratégie est alors la suivante : on commence par estimer δ par MCO ; puis, si $\Sigma = \Sigma(\theta)$ est de forme paramétrique connue, on estimera (δ, θ) par moindres carrés quasi généralisés (MCQG) en utilisant l'algorithme à bascule suivant :

1. estimer δ par MCO : $\tilde{\delta} = ({}^t Z Z)^{-1} {}^t Z X$.
2. calculer les résidus des MCO : $\tilde{\varepsilon} = X - Z \tilde{\delta}$
3. sur la base de $\tilde{\varepsilon}$, estimer $\tilde{\theta}$ (pour $\Sigma(\theta)$ ou $2\gamma(\theta)$) par un critère (1.16-1.17).
4. estimer Σ par $\tilde{\Sigma} = \Sigma(\tilde{\theta})$ puis δ par MCG($\tilde{\Sigma}$).

Les étapes 2-4 peuvent être itérées jusqu'à la convergence numérique des estimations. Sous certaines conditions techniques [2], on peut montrer que si l'estimation $\tilde{\delta}$ des MCO de δ est convergente à la vitesse standard \sqrt{n} , les MCG et les MCQG sont asymptotiquement équivalents : les MCQG sont donc de bons estimateurs. Étudions le cas d'une régression gaussienne.

Maximum de vraisemblance pour une régression gaussienne

Supposons que $X \sim \mathcal{N}_n(Z\delta, \Sigma(\theta))$. Si $\Sigma(\theta) = \sigma^2 Q(\tau)$, $\theta = (\sigma^2, \tau)$, l'opposée $-l(\delta, \theta)$ de la log-vraisemblance de $X(n)$ s'écrit, à une constante près :

$$\begin{aligned} -2l(\delta, \theta) &= \log |\Sigma(\theta)| + {}^t (X - Z\delta) \Sigma^{-1}(\theta) (X - Z\delta) \\ &= n \log \sigma^2 + \log |Q(\tau)| + {}^t (X - Z\delta) Q^{-1}(\tau) (X - Z\delta) / \sigma^2 \end{aligned}$$

L'estimation du MV à τ connu est donc :

$$\hat{\delta} = ({}^t Z Q^{-1}(\tau) Z)^{-1} {}^t Z Q^{-1}(\tau) X, \quad (2.19)$$

$$\hat{\sigma}^2 = (X - Z\hat{\delta})Q^{-1}(\tau)(X - Z\hat{\delta})/n. \quad (2.20)$$

Le profil $l^*(\tau)$ en τ de la log-vraisemblance s'obtient en remplaçant (δ, σ^2) par leur estimation dans $l(\delta, (\sigma^2, \tau))$:

$$l^*(\tau) = -\frac{n}{2} \{ \log[{}^t X Q^{-1}(\tau)(I - P(\tau))X] + \log |Q(\tau)| + n(1 - \log n) \} \quad (2.21)$$

où $P(\tau) = Z({}^t Z Q^{-1}(\tau) Z)^{-1} {}^t Z Q^{-1}(\tau)$ est la Q^{-1} -projection orthogonale sur l'espace engendré par les colonnes de Z . On estime alors τ par $\hat{\tau}$ en maximisant (2.21), puis δ et σ^2 par (2.19) et (2.20) avec $\hat{\tau}$.

Mardia et Marshall ([33]; cf. également [16], p. 484) montrent sous certaines conditions que l'estimateur $\hat{\eta} = (\hat{\delta}, \hat{\theta})$ du MV de η est consistant et asymptotiquement gaussien. Ces conditions se simplifient si X est stationnaire et observé sur un sous-domaine $S \subset \mathbb{R}^d$ dont les points sont espacés de $a > 0 : \forall i \neq j \in S, \|i - j\| \geq a$. Reste à identifier l'information de Fisher de $(\hat{\delta}, \hat{\theta})$, $J(\delta, \theta) = E_{\delta, \theta} \{-l^{(2)}(\delta, \theta)\}$.

Sous les conditions assurant la normalité asymptotique,

$$J^{1/2}(\delta, \theta) \{ (\hat{\delta}, \hat{\theta}) - (\delta, \theta) \} \xrightarrow{loi} \mathcal{N}_{p+q}(0, I_{p+q}) \text{ si } n \rightarrow \infty \text{ où } J(\delta, \theta) = \begin{pmatrix} J_\delta & 0 \\ 0 & J_\theta \end{pmatrix}.$$

avec $J_\delta = {}^t X \Sigma^{-1}(\theta) X$ et $(J_\theta)_{ij} = (\frac{1}{2} \text{tr}(-\Sigma^j \Sigma_i))_{ij}$ où $\Sigma_i = \partial \Sigma / \partial \theta_i$, $\Sigma^j = \partial \Sigma^{-1} / \partial \theta_j$.

On fera les remarques suivantes :

1. puisque $\Sigma^j = -\Sigma^{-1} \Sigma_j \Sigma^{-1}$ et $\Sigma_i = -\Sigma \Sigma^i \Sigma$, $(J_\theta)_{ij} = 2^{-1} \text{tr}(\Sigma^{-1} \Sigma_j \Sigma^{-1} \Sigma_i) = 2^{-1} \text{tr}(\Sigma^j \Sigma \Sigma^i \Sigma) = -2^{-1} \text{tr}(\Sigma^i \Sigma_j)$; on utilisera la première expression de J_θ si $\Sigma(\theta)$ est de forme paramétrique connue (modèle de covariance ou de variogramme), la deuxième si c'est $\Sigma^{-1}(\theta)$ qui est de forme paramétrique connue (modèle *SAR* ou *CAR*);
2. $\hat{\delta}$ et $\hat{\theta}$ sont asymptotiquement indépendants;
3. $J(\delta, \theta)$ ne dépend que de θ : dans la pratique, on estime $J(\delta, \theta)$ soit en remplaçant θ par $\hat{\theta}$ ou en utilisant la matrice d'information observée $-l^{(2)}(\hat{\delta}, \hat{\theta})$.

Ce résultat de normalité asymptotique permet de construire tests et régions de confiance pour (δ, θ) de façon jointe ou séparée : par exemple, la région de confiance approchée au niveau α pour δ est $\{(\hat{\delta} - \delta) J_\delta^{-1} (\hat{\delta} - \delta) \leq q(p; \alpha)\}$ où $q(p; \alpha)$ est le α -quantile d'un χ_p^2 .

Exemple 2.7 *Pourcentage du groupe sanguin (suite).*

Le test de Moran a indiqué une dépendance spatiale pour la variable "pourcentage". Cette dépendance peut provenir d'autres variables expliquant ce pourcentage : par exemple, on a observé que le pourcentage est plus élevé dans les régions de colonisation anglo-normande. Pour cela, on va considérer deux variables explicatives, **towns** le nombre de villes par surface unitaire, et **pale**, qualitative, qui indique si le comté était sous contrôle anglo-normand ou non.

La régression du pourcentage sur ces deux variables pour des résidus i.i.d. (modèle (a)) montre que seule **pale** est significative (cf. Tab. 2.2, modèle (a)).

Coefficients	Modèles	
	(a)	(b)
Intercept	27.573 (0.545)	28.232 (1.066)
towns	-0.360 (2.967)	-
pale	4.342 (1.085)	2.434 (0.764)
ρ	-	0.684 (0.148)

TAB. 2.2 – Résultats de l'estimation par MV de régressions spatiales sur les données **eire** : (a) modèle avec deux covariables **towns** et **pale** et erreurs i.i.d. gaussiennes; (b) modèle avec **pale** comme seule covariable et erreurs *SAR*.

Mais la corrélation spatiale subsiste : en effet, l'indice de Moran des résidus est significatif. On propose alors un modèle (b) de régression avec **pale** comme seule covariable et des résidus *SAR*, $r = \rho W r + \varepsilon$. Les résultats du MV (cf. Tab. 2.2, modèle (b)) montre que le paramètre ρ de corrélation spatiale est toujours significatif.

Régression	A		B	
Modèle	χ^2	Press	χ^2	Press
<i>MCO</i>	××	1.90×10^{-6}	××	0.96×10^{-6}
<i>CAR</i>	$\chi_1^2 = 25.6$	1.44×10^{-6}	$\chi_1^2 = 2.1$	0.94×10^{-6}
<i>SAR</i>	$\chi_1^2 = 32.0$	1.37×10^{-6}	$\chi_1^2 = 8.2$	0.87×10^{-6}
Disque	$\chi_1^2 = 9.7$	1.40×10^{-6}	$\chi_1^2 = 9.4$	0.88×10^{-6}
Expo+pépité	$\chi_2^2 = 30.4$	1.34×10^{-6}	$\chi_2^2 = 8.4$	0.86×10^{-6}
Matérn	$\chi_2^2 = 25.5$	1.41×10^{-6}	$\chi_2^2 = 8.6$	0.86×10^{-6}

TAB. 2.3 – Gain d’ajustement par MV (χ^2) et de prédiction (Press) pour 5 modèles de résidus et les régressions (A) et (B) en comparaison des *MCO*.

Exemple 2.8 *Implantations industrielles, consommation de tabac et cancer du poumon.*

Cet exemple reprend l’étude épidémiologique [40] de Richarson, Guihenneuc et Lasserre sur la mortalité $Y = (Y_i)$ par cancer du poumon et ses liens avec l’implantation des industries métallurgiques (*métal*), mécaniques (*méca*) et textiles (*textile*) d’une part et la consommation de cigarettes (*tabac*) d’autre part.

Si des données épidémiologiques Y présentent une auto-corrélation spatiale, l’interprétation n’est pas la même si on a à faire à une maladie infectieuse (phénomène de diffusion spatiale locale) ou à une maladie chronique, comme c’est le cas ici. Dans ce cas, la corrélation trouve (en partie) son origine dans des covariables de risques présentant elles-mêmes des dépendances spatiales. Si certaines de ces variables sont observables, on pourra les introduire dans la régression en veillant à maintenir un modèle “paramétriquement” parcimonieux. Mais souvent elles ne le sont pas et cette corrélation est due à d’autres facteurs pris en charge par la covariable endogène $Y_{\partial i}$.

Y , la variable à expliquer, est le taux de mortalité standardisé du cancer du poumon (nombre de décès par cancer/nombre d’habitants) chez les hommes de 35 à 74 ans sur les deux années 1968-1969. Ce taux est mesuré sur 83 départements français (données INSERM-CépicDc IFR69). Les ventes de cigarettes (SEITA) retenues sont celles de 1953, la remontée de 15 ans vers le passé permettant de mesurer l’influence de la consommation de tabac sur le cancer du poumon. La régression étudiée est,

$$Y = X\beta + u, \quad Y \text{ et } u \in \mathbb{R}^{83}$$

notée (A) quand on prend comme covariables les 3 variables industrielles et (B) si on y ajoute la *tabac*. Cinq modèles spatiaux (paramètre θ) sont proposés pour le résidu : les deux premiers sont du type *AR* sur réseau, les trois suivants, géostatistiques, sont associés à une structure paramétrique isotropique de la covariance spatiale $\Sigma = Cov(u)$:

1. Le *CAR* gaussien $u_i = c \sum_{j \in \partial i} w_{ij} x_j + e_i$, $Var(e_i) = \sigma_e^2$ pour W la matrice (symétrique) de 0 et 1 de contiguïté spatiale des 83 départements ;
2. Le *SAR* gaussien $u_i = b \sum_{j \in \partial i} w_{ij}^* x_j + \varepsilon_i$, $Var(\varepsilon_i) = \sigma_\varepsilon^2$ où W^* est la matrice W où chaque ligne a été normalisée à 1 ;
3. Le modèle disque : $\Sigma_{i,j} = \sigma^2 f_a(\|i - j\|_2)$;
4. Le modèle de Matérn de paramètres $(\sigma^2; a, \nu)$ (cf. §1.3.1) ;
5. Le modèle exponentiel avec effet pépitique : si $i \neq j$,

$$\Sigma_{i,j} = \begin{cases} \sigma^2 \gamma \exp\{-\lambda \|i - j\|_2\}, \gamma \leq 1 \text{ si } i \neq j \\ \sigma^2 \text{ sinon.} \end{cases}$$

Le modèle “disque” est l’équivalent en dimension 2 du modèle sphérique, la sphère de rayon a étant remplacée par le disque de rayon a . Notons que ces modèles ne sont pas emboîtés ; les trois premiers sont de dimension 2, les deux derniers de dimension 3.

L’indice de Moran calculé pour la matrice de poids normalisés W^* montre que les variables sont spatialement corrélées (0.53 pour Y , 0.58 pour *tabac*, 0.41 pour *métal* et 0.26 pour *textile*) sauf pour la variable *méca* (0.12). De même, les résidus des *MCO* des régressions (A) et (B) sont significativement corrélés.

• La table (2.3) donne le gain en log-vraisemblance entre les *MCO* (résidus u i.i.d) et chacun des modèles de résidu pour (A) et (B). Ce gain est également évalué par le critère de prédiction *Press* (Predictive Sum of Squares), somme des carrés des résidus conditionnels (ϵ_i) du modèle estimé, où $\epsilon_i = u_i - \sum_{j:j \neq i} \frac{\sigma^{ij}}{\sigma^{ii}} u_j$ où $(\sigma^{ij})^{-1} = \widehat{\Sigma}(\widehat{\beta}, \widehat{\theta})$.

Le quantile d’un χ_1^2 à 95% valant 3.841 (resp. 5.911 pour un χ_2^2), les 5 modélisations spatiales améliorent significativement le modèle sauf *CAR*-(B). L’amélioration de la prédiction est de l’ordre de 30% pour (A) et de (10%) pour (B), la prise en compte de la variable *tabac* faisant, comme attendu, décroître la statistique du χ^2 et le critère *Press*.

Régression	A	A	B	B
Modèle	paramètres	e.t.	paramètres	e.t.
CAR	$\hat{c} = 0.175$	0.005	$\hat{c} = 0.077$	0.057
SAR	$\hat{b} = 0.613$	0.103	$\hat{b} = 0.353$	0.138
disque	$\hat{a} = 94.31$	5.66	$\hat{a} = 35.73$	4.92
expo+pépite	$\hat{\gamma} = 0.745$	0.091	$\hat{\gamma} = 0.554$	0.265
	$\hat{\lambda} = 0.0035$	0.0018	$\hat{\lambda} = 0.012$	0.008
Matérn	$\hat{\nu} = 0.305$		$\hat{\nu} = 0.228$	
	$\hat{a} = 112.36$	16.41	$\hat{a} = 75.19$	23.74

TAB. 2.4 – Estimation du MV des paramètres des 5 modèles spatiaux, sous (A) et sous (B) (e.t. pour écart-type).

Régression et modèle de u	Métallurgie	Mécanique	Textile
(i) : (A) et MCO	2.46 (0.39)	1.88 (0.63)	1.12 (0.53)
(ii) : (A) et SAR	1.39 (0.42)	2.11 (0.55)	0.38 (0.49)
(iii) : (A) et Matérn	1.35 (0.43)	1.90 (0.57)	0.38 (0.50)
(i) : (B) et MCO	1.50 (0.27)	1.29 (0.46)	0.84 (0.38)
(ii) : (B) et SAR	1.11 (0.32)	1.37 (0.45)	0.61 (0.39)
(iii) : (B) et Matérn	1.05 (0.32)	1.24 (0.44)	0.62 (0.38)

TAB. 2.5 – Estimations (écarts types) des pentes des régressions (A) et (B) sous trois modèles de résidus : (i) indépendance, (ii) SAR et (iii) Matérn.

•• La table (2.4) donne les estimateurs du MV des paramètres spatiaux et leurs écarts types. Tous les paramètres sont significatifs pour (A) et montrent une dépendance plus faible pour (B). La valeur estimée $\hat{c} = 0,175$ du CAR-(A) est proche de la valeur maximum autorisée $c_{\max} = 0.179$, inverse de la plus grande valeur propre de W ; cette “proximité” doit être relativisée car la corrélation aux proches voisins reste modérée même quand c est proche de c_{\max} . Pour le modèle CAR-(B), c n’est plus significatif. Pour le modèle SAR associé à W^* , b s’interprète comme le poids d’influence du voisinage. Le paramètre $2\hat{a} \simeq 188$ km du modèle disque représente la distance à partir de laquelle la corrélation s’annule. Le paramètre $\hat{\nu} = 0.305$ du modèle de Matérn indique une décroissance de type linéaire de la covariance au voisinage de 0.

••• La table (2.5) donne les estimations des régressions (A) et (B) pour trois modèles de résidus, (i) indépendants (MCO), (ii) SAR et (iii) covariance de Matérn. On constate que la modélisation spatiale a une forte influence sur l’estimation des paramètres mais pas sur leur précision : par exemple, pour (A) et *métal*, la modélisation spatiale réduit de moitié la pente ($\hat{\beta}_{MCO}/\hat{\beta}_{SAR} = 1.77$) et la statistique t associée ($t_{\beta_{MCO}}/t_{\beta_{SAR}} = 1.95$) ; d’autre part, elle a rendu non-significative la variable *textile*. Comme on pouvait s’y attendre, la prise en compte de la variable *tabac* diminue significativement l’ensemble des coefficients et leurs écarts types.

En conclusion, la prise en compte de la structure spatiale du résidu influence significativement l’estimation des régressions (A) et (B), les variables *métal* et *méca* étant significatives mais pas la variable *textile*.

2.5 Exercices

Exercice 2.1 Trois prévisions pour un AR factorisant sur \mathbb{Z}^2 .

Soit X un champ gaussien stationnaire centré sur \mathbb{Z}^2 , de covariance $C(h) = \sigma^2 \rho^{\|h\|_1}$. On considère les 3 prédictions de X_0 :

- (i) par 0 ($= E(X_0)$) ;
- (ii) la prévision causale optimale à partir de $X_{1,0}$, $X_{0,1}$ et $X_{1,1}$;
- (iii) la prévision optimale à partir des $\{X_t, t \neq 0\}$.

Explicitiez les deux derniers prédicteurs. Vérifier que pour $\rho = 1/2$, les variances d’erreurs de prédiction sont dans les rapports 1, 0.5625 et 0.36.

Exercice 2.2 Expliciter les représentations CAR des SAR suivants :

$$X_{s,t} = aX_{s-1,t} + bX_{s,t-1} + cX_{s+1,t-1} + \varepsilon_{s,t}$$

$$X_{s,t} = a(X_{s-1,t} + X_{s+1,t}) + b(X_{s,t-1} + X_{s,t+1}) + c(X_{s-1,t-1} + X_{s+1,t+1}) + \varepsilon_{s,t}.$$

Exercice 2.3 Deux représentations SAR équivalentes.

On considère les deux SAR sur le tore à 3 points, $S = \{1, 2, 3\}$:

$$\begin{aligned} X_i &= aX_{i+1} + \varepsilon_i, \quad i = 1, 3 \quad (X_4 \equiv X_1), \quad \varepsilon = BBf(\sigma^2), \text{ et} \\ X_i &= b(X_{i-1} + X_{i+1}) + \eta_i, \quad i = 1, 3 \quad (X_0 \equiv X_3 \text{ et } X_4 \equiv X_1) \text{ et } \eta = BBf(\tau^2). \end{aligned}$$

Expliciter les covariances inverses des deux modèles. Préciser les valeurs de (a, σ^2) et (b, τ^2) conduisant au même modèle.

Exercice 2.4 Identifier sous forme d'une série la variance $\text{Var}(X_0)$ du modèle CAR isotropique aux 4-ppv sur \mathbb{Z}^2 de paramètre a , $|a| < 1/4$. Déterminer la corrélation à distance 1 et tracer la courbe de corrélation $a \rightarrow \rho(a)$. Même question en dimension $d = 1$, $|a| < 1/2$.

Exercice 2.5 Comportement d'un CAR isotropique à la frontière.

(1) On considère le CAR isotropique aux 4-ppv sur \mathbb{Z}^2 de paramètre $a = \frac{1}{4} - \varepsilon$ avec $\varepsilon \downarrow 0$. Quelle équation vérifie $\rho_\varepsilon = \rho_X(1, 0)$? Montrer que $1 - \rho_\varepsilon \sim -\frac{\pi}{2}(\log \varepsilon)^{-1}$. Pour quelle valeur de a obtient-on $\rho = 0.99$?

(2) Même étude pour le modèle isotropique aux 2-ppv sur \mathbb{Z}^1 .

Exercice 2.6 Modèle gaussien échangeable sur $S = \{1, 2, \dots, n\}$.

On considère $X \in \mathbb{R}^n$ suivant le modèle SAR : $X = \alpha JX + \varepsilon$, où ε est un BB gaussien et J la matrice $n \times n$ de coefficients $J_{ij} = 1$ si $i \neq j$, $J_{ii} = 0$ sinon. On notera que l'ensemble $\{aI + bJ, a, b \in \mathbb{R}\}$ est stable par multiplication et par inversion dès que $aI + bJ$ est inversible.

(1) A quelle condition sur α le modèle est-il bien défini ? Montrer que $\text{Cov}(X) = r_0I + r_1J$. Identifier r_0 et r_1 .

(2) Vérifier que X est le CAR : $X = \beta JX + e$. Identifier β et $\text{Cov}(e)$.

Exercice 2.7 Choisir X un SAR gaussien non-stationnaire sur S fini et expliciter sa représentation CAR : graphe, lois conditionnelles, loi globale, prédiction de X_i à partir des autres observations.

Exercice 2.8 Loi limite de l'indice de Moran.

Déterminer la loi asymptotique de l'indice de Moran sous l'hypothèse d'indépendance pour des poids $w_{i,j} = 1$ si (i, j) , $w_{i,j} = 0$ sinon et pour les graphes suivants :

1. $S = \mathbb{Z}^2$ et la relation aux 4-p.p.v. (resp. aux 8-p.p.v.).
2. S est le réseau triangulaire régulier de \mathbb{R}^2 et la relation aux 6-p.p.v.
3. $S = \mathbb{Z}^d$ et la relation aux 2d-p.p.v..

Exercice 2.9 Test de factorisation d'une covariance.

Soit X un CAR gaussien, centré, stationnaire et aux 8-ppv sur \mathbb{Z}^2 .

1. Expliciter ce modèle et le sous-modèle (F) à covariance factorisante.
2. X est observé sur le carré de côté n . Écrire les équations du MV et du PMV pour chaque modèle. Tester (F). Déterminer sous (F) la loi asymptotique de $\Delta = n \times \{(\hat{r}_{00}\hat{r}_{11} - \hat{r}_{10}\hat{r}_{01})^2 + (\hat{r}_{00}\hat{r}_{1,-1} - \hat{r}_{-1,0}\hat{r}_{01})^2\}$.

Exercice 2.10 Régression spatiale gaussienne.

On considère le modèle de régression spatiale gaussienne $Z = X\beta + \delta$, Z et $\delta \in \mathbb{R}^n$, $\beta \in \mathbb{R}^p$, où δ est le modèle SAR : $(I - \varphi W)\delta = \varepsilon$ pour $\varepsilon \sim BBG(\sigma^2)$. Identifier la matrice d'information $J_{p+2}(\theta)$ de $\theta = (\beta, \sigma^2, \delta)$. Même question si δ est un modèle CAR.

Exercice 2.11 Plan d'expérience sur un champ aléatoire.

p traitements d'effets moyens $\mu = {}^t(\mu_1, \mu_2, \dots, \mu_p)$ sont appliqués, chacun r fois, sur les $n = r \times p$ parcelles d'un champ $S = \{1, 2, \dots, n\}$. On observe Y , de moyenne $D\mu$ où D est la matrice $n \times p$ des placements des traitements (D vérifie : ${}^tDD = rI_p$). On suppose que le processus des résidus $X = Y - D\mu$ est un modèle CAR spatial : $X = \beta WX + e$, où $W_{ij} = 1$ si $|i - j| = 1$, et $W_{ij} = 0$ sinon.

1. Estimer μ par MCO. En déduire une estimation \hat{X} des résidus.
2. Estimer β par MC conditionnel sur la base de \hat{X} .
3. En déduire l'estimation des MCG($\hat{\beta}$) de μ . Interpréter.

Chapitre 3

Champ de Markov sur réseau

Soit $X = (X_i, i \in S)$ un champ aléatoire sur un ensemble discret de sites $S = \{1, 2, \dots, n\}$ et à valeurs dans l'espace d'états $\Omega = E^S$. Le réseau S peut être un réseau régulier, i.e. $S \subseteq \mathbb{Z}^d$, ou non. L'espace d'états E est général. Si π est la (densité de la) loi de X , nous nous intéressons ici à la description de π à partir des lois conditionnelles, ce qui est utile chaque fois qu'un modèle est connu à partir de ses comportements locaux. En particulier, nous allons répondre à la question : *étant donnée une famille $\{\nu_i(\cdot | x^i), i \in S, x^i = (x_j, j \neq i)\}$ de lois sur E dépendant de la réalisation x^i de x ailleurs qu'en i , à quelles conditions ces lois $\{\nu_i\}$ sont-elles les lois conditionnelles $\pi_i(\cdot | x^i)$ d'une loi jointe π ?* On pourra consulter sur ce sujet le livre d'Arnold, Castillo et Sarabia [4].

Répondre à cette question permet de décrire une loi jointe π à partir de comportements conditionnels $\nu_i(\cdot | x^i)$ en chaque site. Si $\nu_i(\cdot | x^i)$ est locale, ne dépendant que de $x_{\partial i}$, la configuration de x au voisinage de i , la complexité du modèle s'en trouvera fortement réduite : dans ce cas, on dit que X est un *champ de Markov*.

Sans condition, des lois conditionnelles $\{\nu_i\}$ ne se recollent pas. Les *spécifications de Gibbs* définissent des lois conditionnelles qui se recollent sans condition. L'importance des modèles de Gibbs est renforcée par le théorème de Hammersley-Clifford qui dit qu'un champ de Markov est un champ de Gibbs avec des potentiels locaux. Les *auto-modèles markoviens* de Besag forment une sous-classe de champs de Markov particulièrement simple et utile en statistique spatiale. Nous étudierons les trois méthodes d'estimation d'un champ de Markov : maximum de vraisemblance (MV), maximum de pseudo-vraisemblance conditionnelle (PVC) et méthode de codage.

Contrairement à une modélisation gaussienne (ou du second ordre) pour laquelle $E = \mathbb{R}$ ou \mathbb{R}^p , l'espace d'états E d'un champ de Gibbs-Markov est général, quantitatif, qualitatif, voire mixte. Par exemple, on prendra :

- $E = \mathbb{R}^+$ pour un champ exponentiel ou un champ Gamma (hauteurs de pluies) ;
- $E = \mathbb{N}$ pour un champ poissonnien (comptage de malades) ;
- $E = \{a_0, a_1, \dots, a_{m-1}\}$ pour un champ catégoriel (variétés végétales), $E = \{0, 1\}$ pour un champ binaire (présence/absence) ;
- $E = \Lambda \times \mathbb{R}^p$ pour un champ croisant un label $\lambda \in \Lambda$ catégoriel et une modalité quantitative multivariée (i.e. en télédétection multispectrale),
- $E = \{0\} \cup]0, +\infty[$ pour un état mixte (i.e., en pluviométrie, $X = 0$ s'il ne pleut pas, $X > 0$ sinon),
- $E = \mathbb{R}$ ou \mathbb{R}^p pour un champ gaussien.

On supposera que l'espace d'états E est muni d'une mesure de référence $\lambda > 0$: si $E \subset \mathbb{R}^p$, on prendra λ la mesure de Lebesgue ; pour E discret, λ est la mesure de comptage. Les définitions et les résultats de ce chapitre s'étendent sans difficulté au cas où les espaces d'états E_i dépendent du site i .

3.1 Recollement de lois conditionnelles

Sans condition, des lois conditionnelles $\{\nu_i\}$ ne se recollent pas. Il y a plusieurs façons de comprendre cela : l'exercice (3.3) utilise un argument de dimension paramétrique ; on vérifiera aussi que deux lois conditionnelles gaussiennes $(X_1 | X_2 = x_2) \sim \mathcal{N}(\rho_1 x_2, 1)$ et $(X_2 | X_1 = x_1) \sim \mathcal{N}(\rho_2 x_1, 1)$ ne se recollent pas si $\rho_1 \neq \rho_2$.

Par contre, pour $E = \{0, 1\}$, un exemple de lois se recollant automatiquement est, si $\beta_{ij} = \beta_{ji}$ pour tout $i \neq j$,

$$\nu_i(x_i | x^i) = \frac{\exp x_i(\alpha_i + \sum_{j \neq i} \beta_{ij} x_j)}{1 + \exp(\alpha_i + \sum_j \beta_{ij} x_j)}, i \in S.$$

La loi jointe associée,

$$\pi(x) = C \exp\left\{\sum_i \alpha_i x_i + \sum_{i < j} \beta_{ij} x_i x_j\right\},$$

est appelée *champ de Markov auto-logistique*. Cette propriété d'auto-modèle sera présentée au §3.4 dans un contexte général.

La détermination des conditions de recollement de lois conditionnelles est une question importante [4]. Fixons $x^* = (x_1^*, x_2^*, \dots, x_n^*)$ un état de référence de E^S : si π admet $\nu = \{\nu_i\}$ comme lois conditionnelles, alors

$$\begin{aligned} \frac{\pi(x_1, \dots, x_n)}{\pi(x_1^*, \dots, x_n^*)} &= \prod_{i=0}^{n-1} \frac{\pi(x_1^*, \dots, x_i^*, x_{i+1}, x_{i+2}, \dots, x_n)}{\pi(x_1^*, \dots, x_i^*, x_{i+1}^*, x_{i+2}, \dots, x_n)} \\ &= \prod_{i=0}^{n-1} \frac{\nu_i(x_{i+1} \mid x_1^*, \dots, x_i^*, x_{i+2}, \dots, x_n)}{\nu_i(x_{i+1}^* \mid x_1^*, \dots, x_i^*, x_{i+2}, \dots, x_n)} \end{aligned}$$

Ainsi, si les lois ν_i se recollent en π , π peut être reconstruite à partir des ν_i . Mais cette reconstruction doit être invariante, d'une part par permutations sur les coordonnées $(1, 2, \dots, n)$, d'autre part vis à vis du choix de l'état de référence x^* : ces invariances traduisent exactement les contraintes sur $\{\nu_i\}$ qui assurent la cohérence des lois conditionnelles.

3.2 Champ de Gibbs sur S

Supposons que S soit dénombrable, i.e. $S = \mathbb{Z}^d$. Notons $\mathcal{S} = \mathcal{P}_F(S)$ la famille des parties finies de S , $y_A = (y_i, i \in A)$ la configuration de y sur A , $y^A = (y_i, i \notin A)$ celle à l'extérieur de A , $y^i = y^{\{i\}}$ celle en dehors du site i , $\Omega_A = \{y_A, y \in \Omega\}$ et $\Omega^A = \{y^A, y \in \Omega\}$.

3.2.1 Potentiel d'interaction et spécification de Gibbs

Un champ de Gibbs est associé à une famille $\pi(\Phi)$ de lois conditionnelles définies à partir de potentiels d'interaction Φ .

Définition 3.1 *Potentiel, énergie admissible, spécification de Gibbs*

(1) *Un potentiel d'interaction est une famille $\Phi = \{\Phi_A, A \in \mathcal{S}\}$ d'applications mesurables $\Phi_A : \Omega_A \rightarrow \mathbb{R}$ t.q., pour tout $\Lambda \in \mathcal{S}$, la somme suivante existe,*

$$U_\Lambda^\Phi(y) = \sum_{A \in \mathcal{S}: A \cap \Lambda \neq \emptyset} \Phi_A(y) \quad (3.1)$$

U_Λ^Φ est appelée l'énergie du potentiel Φ sur la partie Λ .

(2) *Le potentiel Φ est admissible si pour tout $\Lambda \in \mathcal{S}$ et tout $y^\Lambda \in \Omega^\Lambda$,*

$$Z_\Lambda^\Phi(y^\Lambda) = \int_{\Omega_\Lambda} \exp U_\Lambda^\Phi(y_\Lambda, y^\Lambda) dy_\Lambda < +\infty$$

(3) *Si Φ est admissible, la spécification de Gibbs $\pi(\Phi)$ associée à Φ est la famille de lois conditionnelles :*

$$\pi_\Lambda^\Phi(y_\Lambda \mid y^\Lambda) = \{Z_\Lambda^\Phi(y^\Lambda)\}^{-1} \exp U_\Lambda^\Phi(y), \Lambda \in \mathcal{S} \text{ et } y^\Lambda \in \Omega^\Lambda. \quad (3.2)$$

L'énergie $U_\Lambda^\Phi(y)$ n'est autre, à une constante près, que la log-densité de $\pi_\Lambda^\Phi(\cdot \mid y^\Lambda)$. La sommabilité (3.1) est assurée si (S, d) est un espace métrique et si Φ est de portée bornée : $\exists R > 0$ tel que $\Phi_A \equiv 0$ si le diamètre de A dépasse R (seul un nombre fini de potentiels $\Phi_A \neq 0$ contribue à U_Λ^Φ). Cette admissibilité fait de chaque $\pi_\Lambda^\Phi(\cdot \mid y^\Lambda)$ une loi de probabilité.

Spécification conditionnelle $\pi(\Phi)$ et lois globales $\mathcal{G}(\Phi)$.

Notons $\mathcal{G}(\Phi)$ la famille des lois μ sur Ω de lois conditionnelles $\pi(\Phi)$.

1. Si S est fini, $\pi(y) = \pi_S(y)$ et $\mathcal{G}(\Phi) = \{\pi\}$.
2. Si S est infini, l'existence et/ou l'unicité d'une loi globale de spécification $\pi(\Phi)$ n'est pas assurée sans hypothèses particulières. Répondre à cette question difficile est l'objet de la mécanique statistique [24], sujet que nous n'aborderons pas ici.

Identifiabilité du potentiel Φ .

Sans contrainte, Φ n'est pas identifiable : par exemple, pour tout c réel, si Φ_A est modifié en $\tilde{\Phi}_A = \Phi_A + c$, $\pi(\Phi) \equiv \pi(\tilde{\Phi})$. Une façon de rendre les potentiels identifiables est la suivante : notons 0 un état de référence de E ; alors, les contraintes suivantes rendent Φ identifiable :

$$\forall A \neq \emptyset, \Phi_A(y) = 0 \text{ si pour un } i \in A, y_i = 0 \quad (3.3)$$

En effet, la formule d'inversion de Moëbius,

$$\Phi_A(x_A) = \sum_{V \subseteq A} (-1)^{\#(A \setminus V)} U(x_A, \mathbf{0}^A) \quad (3.4)$$

associe univoquement à l'énergie U des potentiels qui vérifient (3.3). Par exemple, pour $S = \{1, 2\}$, cette décomposition s'écrit $U - U(0, 0) = \Phi_1 + \Phi_2 + \Phi_{12}$ pour $\Phi_1(x) = U(x, 0) - U(0, 0)$, $\Phi_2(y) = U(0, y) - U(0, 0)$ et $\Phi_{1,2}(x, y) = U(x, y) - U(x, 0) - U(0, y) + U(0, 0)$. Pour S fini, les Φ_A sont les A -interactions dans la décomposition d'analyse de la variance de U .

3.2.2 Exemples de spécification de Gibbs

Modéliser un champ à partir d'un potentiel $\Phi = \{\Phi_A, A \in \mathcal{C}\}$ requiert l'expertise du spécialiste. Les paramètres du modèle sont : (1) les parties \mathcal{C} indexant Φ ; (2) les potentiels Φ_A pour $A \in \mathcal{C}$.

Si $\Phi_A(y) = \langle \theta, \phi_A(y) \rangle$ où $\phi_A \in \mathbb{R}^p$ est connue, π_Λ appartient à la *famille exponentielle*

$$\pi_\Lambda(y) = Z^{-1}(\theta) \exp\langle \theta, T_\Lambda(y) \rangle \text{ où } T_\Lambda(y) = \sum_{A: A \cap \Lambda \neq \emptyset} \phi_A(y).$$

Disposer d'expressions explicites pour les lois conditionnelles permet d'utiliser des algorithmes markoviens MCMC de simulations (cf. Ch 5) tel l'échantillonneur de Gibbs (cf. §5.3.1) et/ou de recourir à l'estimation par pseudo-vraisemblance conditionnelle (*PVC*; cf. §3.5.2), facile à mettre en oeuvre là où le calcul du maximum de vraisemblance est plus difficile (cf §3.5.2).

Modèles d'Ising sur $S \subset \mathbb{Z}^2$.

Ils modélisent les configurations de spins $E = \{-1, +1\}$ sur S .

(1) *Modèle isotropique aux 4-plus proches voisins (ppv) sur \mathbb{Z}^2 .*

Les seuls potentiels non-nuls sont les potentiels de singletons $\Phi_{\{i\}}(y) = \alpha y_i$, $i \in S$ et les potentiels de paires $\Phi_{\{i,j\}}(y) = \beta y_i y_j$ si $\|i - j\|_1 = 1$, encore noté $\langle i, j \rangle$. La spécification est d'énergie $U_\Lambda(y_\Lambda | y^\Lambda) = \alpha \sum_{i \in \Lambda} y_i + \beta \sum_{i \in \Lambda, j \in S: \langle i, j \rangle} y_i y_j$, les lois conditionnelles s'écrivant :

$$\pi_\Lambda(y_\Lambda | y^\Lambda) = Z^{-1}(\alpha, \beta; y^\Lambda) \exp U_\Lambda(y_\Lambda; y^\Lambda)$$

La constante $Z^{-1}(\alpha, \beta; y^\Lambda) = \sum_{y_\Lambda \in E^\Lambda} \exp U_\Lambda(y_\Lambda; y^\Lambda)$, somme de $2^{|\Lambda|}$ termes, est incalculable pour Λ grand. La loi conditionnelle en i dépend des 4-ppv :

$$\pi_i(y_i | y^i) = \frac{\exp y_i (\alpha + \beta v_i(y))}{2 \operatorname{ch}(\alpha + \beta v_i(y))}, \text{ avec } v_i(y) = \sum_{j: \langle i, j \rangle} y_j.$$

Il existe toujours au moins une loi globale de spécification $\pi(\Phi)$; pour $\alpha = 0$, elle est unique si $\beta < \beta_c = \log(1 + \sqrt{2})/4$ [24].

(2) *$S \subset \mathbb{Z}^2$ fini* : pour le même potentiel, le modèle global π existe toujours et il est unique. Le modèle sur le *tore* $S = T^2 = \{1, 2, \dots, m\}^2$ est obtenu en considérant la relation de voisinage modulo m : $\langle i, j \rangle \iff |i_1 - j_1| + |i_2 - j_2| \equiv 1 \pmod{m}$.

(3) On peut choisir $E^* = \{0, 1\}$ comme espace d'état (en écologie, présence/absence d'une espèce; en épidémiologie, état sain/malade). La bijection $y_i = 2x_i - 1$ de E^* sur $E = \{-1, +1\}$ et la correspondance $a = 2\alpha - 8\beta$ et $b = 4\beta$ associe au modèle d'Ising en $y \in E$ le modèle en $x \in E^*$:

$$\pi_\Lambda(x_\Lambda | x^\Lambda) = Z^{-1}(a, b, x^\Lambda) \exp\{a \sum_{i \in \Lambda} x_i + b \sum_{i \in \Lambda, j \in S: \langle i, j \rangle} x_i x_j\}$$

Généralisations du modèle d'Ising.

Il existe de nombreuses généralisations pouvant couvrir un large spectre de situations réelles.

1. On peut introduire une *anisotropie* en prenant un paramètre β_H pour les voisins horizontaux et β_V pour les voisins verticaux.
2. Un modèle *non-stationnaire* a pour potentiels $\Phi_{\{i\}}(y) = \alpha_i y_i$, $\Phi_{\{i,j\}}(y) = \beta_{\{i,j\}} y_i y_j$, les $\alpha_i, \beta_{\{i,j\}}$ dépendant de i et de (i, j) et/ou s'explicitant à partir de poids connus ou de conditions exogènes observables.
3. La *relation de voisinage* $\langle i, j \rangle$ peut être associée à un graphe \mathcal{G} général, symétrique et sans boucle : i.e., sur \mathbb{Z}^2 , le graphe aux 8 ppv est défini par $\langle i, j \rangle$ si $\|i - j\|_\infty \leq 1$.
4. On peut envisager des potentiels au-delà des paires : par exemple, pour un champ de Gibbs aux 8-ppv, on introduit des potentiels de triplets $\{i, j, k\}$ et de quadruplets $\{i, j, k, l\}$ pour des sites mutuellement voisins.

5. Le nombre d'états, qualitatifs ou quantitatifs, peut être augmenté : variétés végétales observées, niveaux de gris d'une image.

Les simulations des figures 3.1 illustrent la grande variété de textures que ces modèles génèrent ainsi que leur bonne adéquation pour la synthèse de motifs réels. (cf. Winkler [53] et le logiciel de simulation **AntsInFields**).

Modèle à nombre fini d'états.

Le modèle d'Ising peut être étendu à un espace à K états qualitatifs $E = \{a_0, a_1, \dots, a_{K-1}\}$ (des couleurs, des textures, des qualités, des niveaux de gris) en prenant $\Phi_{\{i\}}(y) = \alpha_k$ si $y_i = a_k$ et $\Phi_{\{i,j\}}(y) = \beta_{k,l} = \beta_{l,k}$ si $\{y_i, y_j\} = \{a_k, a_l\}$ pour i et j voisins. Ce modèle (dit de Potts) est utilisé en traitement d'images. Si S est fini, l'énergie vaut

$$U(y) = \sum_k \alpha_k n_k + \sum_{k < l} \beta_{kl} n_{kl} \quad (3.5)$$

où n_k est le nombre de sites de modalité a_k , n_{kl} le nombre de sites voisins de modalités $\{a_k, a_l\}$. Plus α_k est important, plus a_k est probable ; quant à β_{kl} , il contrôle la vraisemblance de la configuration $\{a_k, a_l\}$ en des sites voisins : i.e. on interdira $\{a_k, a_l\}$ en des sites voisins en prenant $-\beta_{kl} > 0$ et grand.

Tels quels, les paramètres ne sont pas identifiables : si a_0 est l'état de référence, les contraintes $\alpha_0 = 0$ et $\forall k : \beta_{0,k} = 0$ rendent Φ identifiable. S'il y a échangeabilité des K modalités, c.a.d. si $\forall k \neq l, \alpha_k \equiv \alpha, \beta_{kl} \equiv \beta$, le modèle dépend du seul paramètre d'interaction β , de loi

$$\pi(y) = Z^{-1} \exp\{-\beta n(y)\} \quad (3.6)$$

où $n(y)$ est le nombre de paires de sites voisins de même modalité. En restauration d'image ([23] ; [11] ; [53]) lorsque π est choisie comme loi a priori sur l'objet à reconstruire, β est un paramètre de régularisation : plus β sera grand négatif, plus les plages de modalités constantes qui seront reconstruites seront régulières.

Spécification gaussienne sur $S = \{1, 2, \dots, n\}$.

Une loi gaussienne $Y = (Y_i, i \in S) \sim \mathcal{N}_n(\mu, \Sigma)$ est un champ de Gibbs si $\Sigma^{-1} = Q$ existe. L'énergie $U(y) = \frac{1}{2} {}^t(y - \mu)Q^{-1}(y - \mu)$ admet pour potentiels

$$\Phi_{\{i\}}(y) = y_i \left\{ \sum q_{ij} \mu_j \right\} - \frac{1}{2} q_{ii} y_i^2, \quad \Phi_{\{i,j\}} = -q_{ij} y_i y_j \text{ si } i \neq j$$

On obtient les lois $\mathcal{L}_A(Y_A | y^A)$ à partir de l'énergie conditionnelle sur A . Y est \mathcal{G} -markovien si, $\forall i \neq j, q_{ij} \neq 0 \iff \langle i, j \rangle$ est une arête du graphe \mathcal{G} .

Famille exponentielle invariante par translation.

Pour $S = \mathbb{Z}^d$ et $\Omega = E^{\mathbb{Z}^d}$, on choisit p potentiels de base $\Phi_k : E^{V_k} \rightarrow \mathbb{R}$ mesurables et bornés, où $V_k \subset \mathbb{Z}^d$ est finie, $k = 1, p, \theta = {}^t(\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$. Notons

$$\Phi_V(y) = \Phi(y) = \sum_{k=1,p} \theta_k \Phi_k(y).$$

A partir de ce potentiel, on engendre les potentiels translatés de i , $\Phi_{V+i}(y) = \Phi_V(\tau_i(y))$ où $\tau_i(y)$ la i -translatée de y :

$$\forall j, (\tau_i(y))_j = y_{i+j}.$$

La spécification suivante appartient à une famille exponentielle invariante par translation,

$$\pi_{\Lambda}^{\theta}(y_{\Lambda} | y^{\Lambda}) = Z_{\Lambda, \theta}^{-1}(y^{\Lambda}) \exp\left\{ \sum_{i \in \mathbb{Z}^d: \{i+V\} \cap \Lambda \neq \emptyset} \Phi_{V+i}(y) \right\} \quad (3.7)$$

Par exemple, le modèle d'Ising invariant par translation sur \mathbb{Z}^2 et aux 4-ppv est associé aux potentiels, $\Phi_0(y) = y_0$ ($V_0 = \{0\}$), $\Phi_H(y) = y_{0,0}y_{1,0}$ ($V_{Hal} = \{(0,0), (0,1)\}$) et $\Phi_V(y) = y_{0,0}y_{1,0}$ ($V_{Val} = \{(0,0), (1,0)\}$). On a, par exemple : $\Phi_{\{i,j\}}(y) = y_{i,j}$ et $\Phi_{H+(i,j)}(y) = y_{i,j}y_{i+1,j}$. Le modèle est isotropique si $\theta_H = \theta_V$, de potentiels, Φ_0 et $\Phi_{iso}(y) = y_{0,0}(y_{1,0} + y_{1,0})$.

Modèle hiérarchique sur un espace produit $E = \Lambda \times \mathbb{R}^p$.

En télédétection, l'observation est par exemple $Y = (X, Z)$ où $X_i \in \Lambda$ est un label de texture et $Z_i \in \mathbb{R}^p$ une mesure multispectrale. Un modèle hiérarchique s'obtient en modélisant X par un modèle de Potts, puis, conditionnellement à X , Z par une texture de niveau de gris gaussienne multispectrale (cf. Ex. 3.2).

3.3 Champ de Markov et champ de Gibbs

L'intérêt des modèles de Gibbs est double : (1) ils définissent simplement des spécifications conditionnelles ; (2) ce sont des champs de Markov. Définissons cette notion.

3.3.1 Cliques d'un graphe, champ de Markov

On suppose que $S = \{1, 2, \dots, n\}$ est muni d'un graphe de voisinage \mathcal{G} *symétrique* et sans boucle. Deux sites $i \neq j$ sont voisins (noté $\langle i, j \rangle$) si (i, j) est une arête de \mathcal{G} . La frontière de voisinage de A est

$$\partial A = \{i \in S, i \notin A : \exists j \in A, \langle i, j \rangle\}.$$

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

Définition 3.2 *Champ de Markov et cliques d'un graphe*

(1) Y est un champ de Markov sur S pour le graphe \mathcal{G} si, $\forall A \subset S$ et $y^A \in \Omega^A$, la loi de Y sur A conditionnelle à y^A ne dépend que de $y_{\partial A}$:

$$\forall A \subset S \text{ et } y^A \in \Omega^A : \pi_A(y_A | y^A) = \pi_A(y_A | y_{\partial A}).$$

(2) Une partie C non-vide de S est une clique du graphe \mathcal{G} si soit C est un singleton, soit tous les éléments de C sont deux à deux voisins pour \mathcal{G} . L'ensemble des cliques de \mathcal{G} est noté $\mathcal{C}(\mathcal{G})$.

Par exemple, pour le graphe de voisinage aux 4-ppv sur \mathbb{Z}^2 , les cliques sont constituées des singletons $\{i\}$ et des sous-ensembles $\{i, j\}$ avec $\|i - j\|_1 = 1$. Pour la relation aux 8-ppv, il faudra ajouter tous les triplets $\{i, j, k\}$ et quadruplets $\{i, j, k, l\}$ de sites u, v à distance $\|u - v\|_\infty \leq 1$.

A une famille \mathcal{C} de parties de S contenant tous les singletons, on associe le graphe de voisinage $\mathcal{G}(\mathcal{C})$ ainsi défini : $i \neq j$ sont voisins de $\mathcal{G}(\mathcal{C})$ si $\exists C \in \mathcal{C}$ t.q. $\{i, j\} \subset C$. La propriété suivante identifie un champ de Markov à un champ de Gibbs.

3.3.2 Le théorème de Hammersley-Clifford

Proposition 3.1 *Le théorème de Hammersley-Clifford (Besag, 1974).*

(1) Soit π un \mathcal{G} -champ de Markov sur E , vérifiant la condition de positivité : $\forall A \subset S$ et $y \in E^S$, $\pi_A(y_A | y^A) > 0$. Alors π est un champ de Gibbs de potentiel $\Phi = \{\Phi_A, A \in \mathcal{C}\}$ où \mathcal{C} est l'ensemble des cliques du graphe \mathcal{G} .

(2) Réciproquement, un champ de Gibbs de potentiels $\Phi = \{\Phi_A, A \in \mathcal{C}\}$ est un champ de Markov pour le graphe de voisinage $\mathcal{G}(\mathcal{C})$.

L'importance du point (1) est de limiter les potentiels aux seules cliques C du graphe \mathcal{G} .

3.4 Les auto-modèles de Besag

Ces modèles à états dans $E \subseteq \mathbb{R}$ sont caractérisés par des densités conditionnelles appartenant à une famille exponentielle [7]. Enonçons le résultat qui est à la base de leur définition.

3.4.1 Recollement de lois conditionnelles et auto-modèle

Soit π un champ de Markov sur $S = \{1, 2, \dots, n\}$ avec des potentiels ne chargeant que les cliques d'au plus deux points,

$$\pi(x) = C \exp\left\{\sum_S \Phi_i(x_i) + \sum_{\{i,j\}} \Phi_{ij}(x_i, x_j)\right\}, \quad (3.8)$$

vérifiant : $\forall x \in E^n$, $\pi(x) > 0$ et les contraintes d'idenfiabilité (3.3). Notons 0 l'état de référence de E .

Théorème 3.1 *Supposons de plus que chaque loi conditionnelle de π appartienne à une famille exponentielle :*

$$\forall i \in S : \log \pi_i(x_i | x^i) = A_i(x^i)B_i(x_i) + C_i(x_i) + D_i(x^i) \quad (3.9)$$

où $B_i(0) = C_i(0) = 0$. Alors :

(a) $\forall i, j \in S, i \neq j, \exists \alpha_i$ et $\beta_{ij} = \beta_{ji}$ tels que

$$A_i(x^i) = \alpha_i + \sum_{j \neq i} \beta_{ij} B_j(x_j) \quad (3.10)$$

$$\Phi_i(x_i) = \alpha_i B_i(x_i) + C_i(x_i) \text{ et } \Phi_{ij}(x_i, x_j) = \beta_{ij} B_i(x_i) B_j(x_j) \quad (3.11)$$

(b) Réciproquement, des lois conditionnelles vérifiant (3.9) et (3.10) se recollent en une loi jointe (3.8) est un champ de Markov de potentiels (3.11).

Une classe importante de champs de Markov est celle des champs de Gibbs à valeurs dans $E \subseteq \mathbb{R}$ avec des cliques d'au plus 2 points et des potentiels de paires quadratiques $\Phi_{\{i,j\}}(y_i, y_j) = \beta_{ij}y_iy_j$.

Définition 3.3 *Auto-modèle markovien de Besag*

Y est un auto-modèle si Y est à valeurs dans \mathbb{R} et si sa loi π s'écrit :

$$\pi(y) = Z^{-1} \exp\left\{\sum_{i \in S} \Phi_i(y_i) + \sum_{\langle i,j \rangle} \beta_{ij}y_iy_j\right\}, \text{ avec, } \forall i, j, \beta_{ij} = \beta_{ji} \quad (3.12)$$

Le théorème précédent permet d'affirmer qu'une famille de lois conditionnelles $\nu = \{\nu_i(\cdot|y^i), i \in S\}$ vérifiant (3.9-3.10) et telle que, $\forall i \in S, B_i(y_i) = y_i$, se recolle en un auto-modèle de Markov de loi π donnée par (3.12) dès que, $\forall i \neq j, \beta_{ij} = \beta_{ji}$. Présentons les principaux auto-modèles.

Les auto-modèles peuvent également être définis si $E \subset \mathbb{R}^d$ ([10]).

3.4.2 Exemples d'auto-modèles

Auto-modèle logistique : $E = \{0, 1\}$.

C'est le modèle d'énergie $U(y) = \sum_i \alpha_i y_i + \sum_{\langle i,j \rangle} \beta_{ij} y_i y_j$ avec, $\forall i \neq j, \beta_{ij} = \beta_{ji}$. La loi conditionnelle en i ,

$$\pi_i(y_i|y^i) = \frac{\exp y_i \{\alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j\}}{1 + \exp\{\alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j\}}$$

est un modèle Logit de paramètre $\theta_i(y_i) = \{\alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j\}$. Ces lois conditionnelles se recollent en la loi jointe π d'énergie U .

Auto-modèle binomial : $E = \{0, 1, 2, \dots, N\}$.

Considérons des lois conditionnelles binomiales $\pi_i(\cdot|y_i) \sim \mathcal{B}in(N, \theta_i(y_i))$: si

$$A_i(y^i) = \log \frac{\theta_i(y_i)}{1 - \theta_i(y_i)} = \alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j \text{ et } B_i(y_i) = y_i,$$

avec pour tout $i \neq j, \beta_{ij} = \beta_{ji}$, ces lois se recollent en une loi jointe π d'énergie $U(y) = \sum_i (\alpha_i y_i + \log \binom{n}{y_i}) + \sum_{\langle i,j \rangle} \beta_{ij} y_i y_j$. Le paramètre binomial conditionnel suit le modèle Logistique aux ppv :

$$\theta_i(y_i) = \frac{\exp\{\alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j\}}{1 + \exp\{\alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j\}} = [1 + \exp -\{\alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j\}]^{-1}.$$

L'espace d'états étant fini, le modèle est toujours admissible, pouvant modéliser autant la compétition ($\beta_{ij} > 0$) que la coopération spatiale ($\beta_{ij} < 0$). Le modèle auto-binomial peut être défini sur un espace $\Omega = \prod_{i \in S} E_i$ où les effectifs $\#E_i = N_i < \infty$ sont finis mais différents.

Auto-modèle poissonnien : $E = \mathbb{N}$.

Si $\forall i \neq j, \beta_{ij} = \beta_{ji} \leq 0$, les lois de Poisson $\pi_i(\cdot|y^i) \sim \mathcal{P}(\lambda_i(y^i))$, $i \in S$, où

$$\log \lambda_i(y^i) = A_i(y^i) = \alpha_i + \sum_{j:\langle i,j \rangle} \beta_{ij} y_j$$

se recollent en une loi jointe d'énergie $U(y) = \sum_i (\alpha_i y_i + \log(y_i!)) + \sum_{j:\langle i,j \rangle} \beta_{ij} y_i y_j$. U est admissible ssi : $\forall i \neq j, \beta_{ij} \leq 0$; si c'est le cas, $\exp U(y) \leq \prod_{i \in S} \exp(\alpha_i y_i)/(y_i!)$ et $\sum_{\mathbb{N}^S} \exp U(y)$ converge; sinon, i.e. si $\beta_{1,2} > 0$, on a pour y_1 et y_2 assez grands,

$$\exp U(y_1, y_2, 0, \dots, 0) = \frac{\exp\{\alpha_1 y_1 + \alpha_2 y_2 + \beta_{1,2} y_1 y_2\}}{y_1! y_2!} \geq \frac{\exp\{(1/2)\beta_{1,2} y_1 y_2\}}{y_1! y_2!}$$

et le terme minorant est celui d'une série divergente.

La condition d'admissibilité ($\beta_{ij} \leq 0$) induisant une *compétition* entre sites voisins est une limitation de la modélisation auto-poissonnienne. On peut autoriser la *coopération* en bornant E à une valeur $K < \infty$, $E = \{0, 1, \dots, K\}$, i.e. en considérant les variables $Z_i = \inf\{Y_i, K\}$ censurées à droite ou encore en retenant les lois de Poisson conditionnelles à $\{Y \leq K\}$. Le nouvel espace d'état étant fini, ces modèles seront toujours admissibles, permettant de recouvrir autant des situations de compétition que des situations de coopération.

Auto-modèle exponentiel : $E =]0, +\infty[$.

Ce sont des modèles d'énergie $U(y) = -\left\{\sum_i \alpha_i y_i + \sum_{\langle i,j \rangle} \beta_{ij} y_i y_j\right\}$: U est admissible si pour tout i, j , $\alpha_i > 0$ et $\beta_{ij} = \beta_{ji} \geq 0$. Les lois conditionnelles sont des lois exponentielles, $\pi_i(\cdot | y^i) \sim \mathcal{Exp}(\mu_i(y^i))$, de paramètres,

$$\mu_i(y^i) = \left\{ \alpha_i + \sum_{j: \langle i,j \rangle} \beta_{ij} y_j \right\}.$$

Les contraintes $\beta_{ij} \geq 0$ traduisent une compétition entre i et j . On pourra autoriser la coopération en tronquant E à $[0, K]$.

Une généralisation de ce modèle consiste à faire intervenir des cliques de 2 points et plus [4] :

$$\pi(y) = Z^{-1} \exp\left\{-\sum_{A \in \mathcal{C}} \beta_A \left(\prod_{l \in A} y_l\right)\right\}$$

l'énergie étant admissible si pour tout $i \in S$: $\beta_{\{i\}} > 0$ et tout $A \subset S$, $\beta_A \geq 0$ si $\#A \geq 2$. Les lois conditionnelles en un site sont encore des lois exponentielles.

Auto-modèle gaussien : $E = \mathbb{R}$.

Une loi gaussienne $Y \sim \mathcal{N}_n(\mu, \Sigma)$ sur $S = \{1, 2, \dots, n\}$ de covariance inversible $\Sigma^{-1} = Q$ est un modèle de Gibbs d'énergie $U(y) = -(1/2)^t (y - \mu) Q (y - \mu)$ et de potentiels de singletons et de paires :

$$\begin{aligned} \Phi_i(y_i) &= -\frac{1}{2} q_{ii} y_i^2 + \alpha_i y_i, \text{ où } \alpha_i = \sum_j q_{ij} \mu_j \text{ et} \\ \Phi_{ij}(y_i, y_j) &= -q_{ij} y_i y_j \end{aligned}$$

D'énergie conditionnelle $U_i(y_i | y^i) = \Phi_i(y_i) + \sum_{j \neq i} \Phi_{ij}(y_i, y_j)$, $\mathcal{L}_i(Y_i | y^i)$ est normale de moyenne $\mu_i(y^i) = -q_{ii}^{-1} \sum_{j \neq i} q_{ij} (y_j - \mu_j)$ et de variance q_{ii}^{-1} . La contrainte d'admissibilité est que Q soit symétrique et d.p..

Auto-modèle avec covariables.

Sans contraintes, un modèle non-stationnaire est de trop grande dimension pour être utile. On peut réduire la dimension en modélisant les paramètres $\theta = \{(\alpha_i, \beta_{ij}), i \neq j\}$ en utilisant des covariables observables $x = (x_i, i \in S)$ et/ou des poids $\{(a_i), (w_{ij})\}$ connus, w étant symétrique. Par exemple, le choix de $\beta_{ij} = \delta w_{ij}$, $x_i \in \mathbb{R}^p$ et $\alpha_i = \sum_{j=1, p} \gamma_j a_i x_{i,j}$ pour le modèle auto-logistique donne un modèle à $p+1$ paramètres,

$$\theta_i(y_i, x_i) = \sum_{j=1}^p \gamma_j a_i x_{i,j} + \delta v_i \text{ où } v_i = \sum_{j: \langle i,j \rangle} w_{ij} y_j$$

Exemple 3.1 *Modélisation de la mortalité par cancer dans la région de Valence (Espagne).*

L'objet de l'étude de Ferrándiz et altri [22] est de voir si la concentration en nitrate dans l'eau potable a un effet ou non sur la mortalité par cancer (vessie, colon, prostate et estomac), la région de Valence, d'intense activité agricole, concentrant d'importantes quantités de fertilisants. La région est divisée en 263 cantons, i étant le chef-lieu de canton et Y_i agrégeant, pour un type de cancer et sur la période 1975-1980, le nombre de décès du canton i . Deux covariables sont retenues : x_{1i} , le pourcentage de la population de plus de 40 ans, et x_{2i} , la concentration en nitrate dans l'eau potable distribuée.

Classiquement, l'analyse statistique des $Y = (Y_i, i \in S)$, variables de comptage, utilise un modèle log-linéaire de régression poissonnienne incorporant les variables explicatives appropriées $x_i = (x_{i1}, x_{i2})$,

$$(Y_i | y^i, x_i) \sim \mathcal{P}(\lambda_i) \text{ où } \log(\lambda_i) = \alpha_i + \sum_{k=1}^p \beta_k x_{ik}.$$

Ce modèle log-linéaire postule l'indépendance des comptages, ce qui est mis en doute.

Un auto-modèle de Poisson permet de lever cette hypothèse : il préserve l'explication de Y_i à partir des covariables x mais autorise la dépendance des Y_i , les variables $y_{\partial i}$ voisines étant des auto-covariables influençant Y_i . Ces modèles permettront de détecter s'il y a ou non dépendance spatiale entre les $\{Y_i\}$ et/ou si d'autres covariables facteurs de risque présentant une dépendance spatiale ont été oubliées dans le modèle. L'auto-modèle poissonnien retenu est le suivant :

$$(Y_i | y^i, x_i) \sim \mathcal{P}(\lambda_i) \text{ où } \log(\lambda_i) = \alpha_i + \sum_{k=1,2} \beta_k x_{ik} + \sum_{j: \langle i,j \rangle} \gamma_{i,j} y_j \quad (3.13)$$

Les paramètres (α_i) captent les propensions à la maladie propres à chaque canton i , les paramètres (β_k) l'influence des covariables (x_k) alors que les paramètres $(\gamma_{i,j}, j \in \partial i, i \in S)$ résument les dépendances entre sites voisins.

Il y a deux interprétations concernant ces paramètres γ :

1. ils mesurent une influence réelle directe entre variables y voisines, comme c'est naturel pour un phénomène de contagion (ce n'est pas le cas ici) ;
2. ils captent et résument d'autres facteurs de risque non présents dans le modèle et présentant une structure spatiale : au site i , ces effets cachés sont alors pris en charge par les covariables observables $y_{\partial i}$. Tester que $\gamma \neq 0$ signifie que ces facteurs sont significatifs. En général, il y a confusion entre ces deux origines.

Pour que le modèle soit admissible, les paramètres γ du modèle doivent être ≤ 0 , des valeurs positives pouvant être acceptées à condition de tronquer les variables de réponse y_i , ce qui est raisonnable car y_i ne peut dépasser la population u_i du canton i .

Le modèle (3.13) tel quel est inutilisable car il y a plus de paramètres que d'observations. D'autre part, il faut identifier la relation de voisinage spatial. Ferrandiz et altri [22] proposent la chose suivante : si u_i est la population du canton i , si d_{ij} est la distance entre les chefs lieux de cantons i et j , la relation de voisinage $\langle i, j \rangle$ et les paramètres γ_{ij} vont se déduire des indices de proximité (a_{ij}) suivants :

$$\langle i, j \rangle \text{ si } a_{ij} = \frac{\sqrt{u_i \times u_j}}{d_{ij}} > a \text{ et } \gamma_{ij} = \gamma \times a_{ij}$$

$a > 0$ est une valeur préalablement calibrée.

Quant aux α_i , ils sont modélisés sur la base d'un seul paramètre α :

$$\alpha_i = \alpha + \log(u_i),$$

L'interprétation du coefficient 1 devant $\log(u_i)$ est que l'on suppose que le nombre moyen λ_i de décès est proportionnel à l'effectif u_i de la population du canton i . On obtient ainsi un modèle à 4 paramètres, $\theta = (\alpha, (\beta_1, \beta_2), \gamma)$, $(Y_i | y^i, x_i) \sim \mathcal{P}(\lambda_i)$ où

$$\log(\lambda_i) = \alpha + \log(u_i) + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma \sum_{j : \langle i, j \rangle} a_{i,j} y_j \quad (3.14)$$

Si $\gamma = 0$, on retrouve la régression poissonnienne log-linéaire de constante α , de covariables (x_1, x_2) et de compensateur $\log(u)$.

3.5 Estimation d'un champ de Markov

Soit X un champ de Gibbs-Markov défini sur un ensemble de sites $S \subset \mathbb{R}^d$ muni d'un graphe de voisinage \mathcal{G} , X étant à valeurs dans $\Omega = E^S$. Soit $\theta \in \Theta \subset \mathbb{R}^p$ et $\Phi_\theta = \{\Phi_{A,\theta}, A \in \mathcal{S}\}$ la famille des potentiels définissant la spécification conditionnelle π_θ de X , $\theta \in \Theta$, un compact de \mathbb{R}^k , $\mathcal{G}(\pi_\theta)$ l'ensemble des lois sur Ω de spécification π_θ . On supposera que \mathcal{G} ne dépend pas de θ et que X est observé sur $D_n \cup \partial D_n \subset S$, où ∂D_n est la frontière de voisinage de D_n . On va présenter trois procédures d'estimation : le maximum de vraisemblance (*MV*), le maximum de pseudo-vraisemblance conditionnelle (*PVC*) et l'estimation par *C*-codage.

3.5.1 Le maximum de vraisemblance

Si X est de loi $P_\theta \in \mathcal{G}(\pi_\theta)$, la loi de X sur D_n conditionnelle à $x(\partial D_n)$ est

$$\pi_n(x; \theta) = Z_n^{-1}(x_{\partial D_n}; \theta) \exp\{U_n(x; \theta)\}$$

où $U_n(x; \theta) = \sum_{A: A \cap D_n \neq \emptyset} \Phi_A(x; \theta)$. Si D_n est grand, la constante de normalisation

$$Z_n^{-1}(x_{\partial D_n}; \theta) = \int_{E^{D_n}} \exp\{U_n(x; \theta)\} dx_{D_n}$$

est en général incalculable, ce qui rend plus difficile l'utilisation du *MV*. Nous présenterons au §4.7 une méthode MCMC de calcul approché de $Z_n(x_{\partial D_n}; \theta)$ pour un processus de Gibbs, qu'il soit latticiel ou ponctuel.

Propriété asymptotique du maximum de vraisemblance.

On suppose que X est un champ de Gibbs sur $S = \mathbb{Z}^d$ de spécification appartenant à une famille exponentielle de potentiels mesurables, bornés et *invariants par translation*, de vraisemblance,

$$\pi_n(x; \theta) = Z_n^{-1}(x_{\partial D_n}; \theta) \exp \sum_{k=1,p} \theta_k \left\{ \sum_{i \in D_n, i+A_k \cap D_n \neq \emptyset} \Phi_{A_k}(\tau_i(x)) \right\} \quad (3.15)$$

où $(\tau_i(x))_j = x_{i+j}$ est la configuration x traduite de i et $\Phi = \{\Phi_{A_k}, k = 1, p\}$ sont p potentiels continus. Alors, l'estimation du *MV* est convergente si $D_n \uparrow \mathbb{Z}^d$ [14]. Plus généralement, si P_θ est caractérisée par ses spécifications et si P_θ est faiblement dépendante, alors l'estimateur du *MV* est asymptotiquement gaussien ([28]; cf. (3.17)).

3.5.2 Pseudo-vraisemblance conditionnelle (PVC) de Besag

La pseudo-vraisemblance conditionnelle (PVC) d'un champ de Markov X est le produit pour les $i \in D_n$ des densités conditionnelles en i ; son logarithme vaut

$$l_n^{PV}(\theta) = d_n^{-1} \sum_{i \in D_n} \log \pi_i(x_i | x_{\partial i}, \theta).$$

L'estimation du maximum de PVC est une valeur maximisant cette PVC. Cette méthode, proposée par Besag [7], est simple à mettre en oeuvre car elle évite le calcul de la constante de normalisation d'une vraisemblance jointe. La perte d'efficacité par rapport au MV est réduite si la dépendance du champ n'est pas trop importante (cf. §3.5.4). Cette estimation est une bonne valeur initiale dans une procédure numérique itérative que requiert des méthodes numériquement plus complexe telle le MV d'un champ de Gibbs (cf. §4.7). Spécifions cette méthode dans deux contextes particuliers.

Estimation d'une texture binaire. X est une texture à états $\{0, 1\}$ observée sur le carré $D_n = \{1, n\}^2 \subset \mathbb{Z}^2$, markovienne de potentiels $\Phi_{\{i,j\}}^k(x) = \beta_k x_i x_j$ si $i - j = \pm k$ où $k \in L \subset \mathbb{Z}^2$, une partie finie et symétrique. La pseudo-vraisemblance vaut

$$L_n^{PV} = \prod_{i \in D_n} \frac{\exp x_i (\beta_0 + \sum_{k \neq 0} \beta_k x_{i+k})}{1 + \exp(\beta_0 + \sum_{k \neq 0} \beta_k x_{i+k})}.$$

La figure 3.1 donne 3 exemples de textures réelles qui ont été modélisées par des champs de Markov, puis

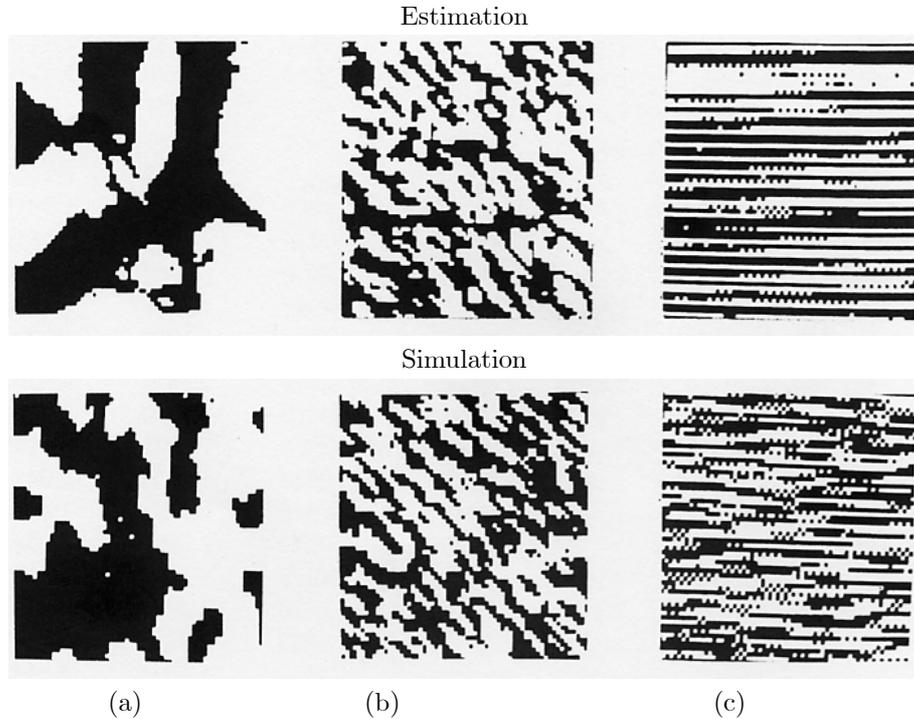


FIG. 3.1 – Textures binaires réelles 64×64 , estimation et simulation : (a) caillou, (b) liège, (c) rideau (source [17]).

estimées par maximum de PVC, et enfin simulées par échantillonneur de Gibbs. L'adéquation des simulations avec les textures originales montre l'intérêt de la modélisation par champ de Markov ainsi que la bonne qualité de l'estimation par PVC.

CAR gaussien à un paramètre. Soit le CAR : $X = \beta W X + e$ où, $\forall i \in D_n$, $Var(e_i) = \kappa$ et W est une matrice symétrique connue à diagonale nulle. Le modèle existe si $C = I - \beta W$ est d.p., soit encore $\beta \lambda < 1$ pour toute valeur propre de W . En situation gaussienne, l'estimation du PMV est identique à celle des MCO sur les résidus conditionnels. On obtient [8] :

$$\hat{\beta}_n = \frac{{}^t X W X}{{}^t X W^2 X} \text{ et } \hat{\kappa} = n^{-1} \{ {}^t X X - ({}^t X W X)^2 / {}^t X W X \}$$

Asymptotiquement, $E(\hat{\beta}_n) \sim \beta$, $Var(\hat{\beta}_n) = 2tr(W^2) / \{tr(C^{-1}W^2)\}^2$, $E(\hat{\kappa}) \sim \kappa$ et $Var(\hat{\kappa}) = 2\kappa^2 tr(C^2) / n^2$.

Nous allons présenter trois résultats asymptotiques pour l'estimation du maximum de *PVC* : le premier de consistance de l'estimation ; les deux suivants de normalité asymptotique lorsque le modèle de Gibbs est à potentiels invariants par translation, bornés ou gaussiens.

Asymptotique de l'estimation du *PMV*.

Le cas d'une spécification invariante par translation.

Supposons que X soit un champ de Markov de spécification invariante par translation sur $S = \mathbb{Z}^d$ de loi $P_\theta \in \mathcal{G}(\pi_\theta)$ où π_θ appartient une famille exponentielle (3.7). Définissons les deux matrices de pseudo-information :

$$\begin{aligned} I_n(\theta) &= \sum_{i \in D_n} \sum_{j \in \partial i \cap D_n} Y_i(\theta) {}^t Y_j(\theta) \text{ où } Y_i(\theta) = \{\log \pi_i(x_i/x_{\partial i}; \theta)\}_\theta^{(1)}, \\ J_n(\theta) &= \sum_{i \in D_n} Z_i(\theta) \text{ où } Z_i(\theta) = -\{\log \pi(x_i/x_{\partial i}; \theta)\}_{\theta^2}^{(2)}. \end{aligned} \quad (3.16)$$

Proposition 3.2 *Si π_θ appartient à la famille exponentielle (3.7) de potentiels bornés et invariants par translation. et si $\theta \mapsto \pi_0(\cdot | \cdot; \theta)$ est identifiable en θ , on a (Comets et Janzura, [15]) :*

$$\{I_n(\hat{\theta}_n)\}^{-1/2} J_n(\hat{\theta}_n) \{\hat{\theta}_n - \theta\} \xrightarrow{\text{loi}} \mathcal{N}_p(0, I_p)$$

Ce résultat ne suppose ni l'unicité de P_θ dans $\mathcal{G}(\pi_\theta)$, ni sa stationnarité, ni son ergodicité, ni sa faible dépendance : seule l'invariance par translation du potentiel est supposée. La paramétrisation de $\pi_0(\cdot | \cdot; \theta)$ est identifiable si elle vérifie :

$$\text{si } \theta \neq \theta', \exists x_0 \text{ et } x_{\partial 0} \text{ tels que } \pi_0(x_0 | x_{\partial 0}; \theta) \neq \pi_0(x_0 | x_{\partial 0}; \theta')$$

Le cas d'une spécification sur un réseau irrégulier.

En environnement non-stationnaire, le réseau S est *irrégulier*. Soit \mathcal{G} un graphe sur S et X un champ markovien sur (S, \mathcal{G}) à valeurs dans $\Omega = \prod_{i \in S} E_i$, la loi P_θ de X étant de spécification π_θ , où $\theta \in \Theta$, un compact de \mathbb{R}^p . Supposons que X est observé sur une suite croissante D_n et sur ∂D_n , $d_n = \#\{D_n\}$. On dit que $C \subset S$ est un *sous-ensemble de codage* (resp. de *codage fort*) de S si :

$$\begin{aligned} \text{Codage} &: \forall i, j \in C, i \neq j, \text{ alors } j \notin \partial i; \\ \text{Codage fort} &: \forall i, j \in C, i \neq j, \text{ alors } \partial i \cap \partial j = \emptyset. \end{aligned}$$

Par exemple, les cases noires (resp. blanches) d'un dammier d'échec définissent un ensemble de codage pour la relation au ppp sur $\{1, 2, \dots, 8\}^2$.

Nous allons donner trois conditions assurant la consistance de l'estimation par *PVC*. Notons π_i et $\pi_{\partial i}$ les spécifications de X en i et sur ∂i ,

$$m_i(\theta, \alpha; x_{\partial i}) = -E_\theta \left\{ \log \frac{\pi_i(X_i | x_{\partial i}, \alpha)}{\pi_i(X_i | x_{\partial i}, \theta)} \right\} \geq 0$$

et posons :

1. *Sur le graphe (S, \mathcal{G}) :* il existe C un sous-ensemble de codage réunion disjointe de K sous-ensembles de codage fort $\{C_k\}$ tels que, posant $A_n = A \cap D_n$, $c_n = \#\{C_n\}$, $c_{n,1} = \#\{C_{1,n}\}$,
 - (i) $\liminf_n c_{1,n}/c_n > 0$ et $\liminf_n c_n/d_n > 0$;
 - (ii) $\chi = \prod_{i \in \partial i} E_i$ est un espace fixe si $i \in C_1$.
2. *Minoration de π_θ :* $\exists c > 0$ t.q., $\forall i \in C_1, x_i, x_{\partial i}, x_{\partial \partial i}$ et $\alpha \in \Theta : \pi_i(x_i | x_{\partial i}, \alpha)$ et $\pi_{\partial i}(x_{\partial i} | x_{\partial \partial i}, \alpha) \geq c$.
3. *Identifiabilité en θ de π_θ :* pour $z \in \chi$, posons $m(\theta, \alpha; z) = \inf_{i \in C_1} m_i(\theta, \alpha; z)$; alors, si $\theta \neq \alpha$,

$$\int_\chi m_i(\theta, \alpha; z) \lambda(dz) \neq 0.$$

Sous ces conditions, l'estimateur $\hat{\theta}_n$ du maximum de *PVC* de X sur D_n est convergent [28]. Ce résultat reste vrai pour l'estimateur de codage (cf. §3.5.3). La première condition porte sur la géométrie du réseau. On vérifie la deuxième si la spécification appartient à une famille exponentielle de paramètre θ , si l'espace d'état est compact et si les potentiels sont continus. La dernière condition assure l'identifiabilité paramétrique du modèle.

Normalité asymptotique (MV et PVC) si X est faiblement dépendant.

Soient $\{\Phi_{A,\theta}\}$ les potentiels définissant X , $\Phi_{A,\theta}^{(l)}$, leurs dérivées d'ordre $l = 0, 1, 2$ en θ . Supposons que les $\{\Phi_{A,\theta}^{(l)}\}$ soient uniformément bornés, que X est faiblement dépendant (cf. [28] ; p. 191) et que la suite (D_n) est à bords asymptotiquement négligeables : $(\#D_n/\#D_n \rightarrow 0)$. Dans ces conditions, on a la normalité asymptotique pour le MV comme pour l'estimation par PVC ([28], §5.3.2), les matrices de variances étant ainsi approchées par :

- Pour le MV, la variance inverse de $\hat{\theta}_n^{(MV)}$ est l'information de Fisher $I_n(\theta)$:

$$I_n(\theta) = \sum_{i,j \in D_n} Cov_\theta(\bar{\Phi}_i^{(1)}(\theta), \bar{\Phi}_j^{(1)}(\theta)) \text{ où } \bar{\Phi}_i(\theta) = \sum_{A:i \in A} \frac{\Phi_{A,\theta}}{\#(A)} \quad (3.17)$$

- Pour la PVC, la variance de $\hat{\theta}_n^{(PVC)}$ dépendra des pseudo-informations de Fisher I_n^* et J_n^* :

$$J_n^*(\theta) = \sum_{i,j \in D_n} Cov_\theta(H_i(\theta), H_j(\theta)) \text{ et } I_n^*(\theta) = \sum_{i,j \in D_n} E_\theta\{Var_\theta^{X_{\partial i}}(H_i(\theta))\},$$

où $H_i(\theta) = H_i(x_i; x_{\partial i}, \theta) = \sum_{A:i \in A} \Phi_{A,\theta}(x)$ est l'énergie conditionnelle en i . La variance asymptotique de l'estimation par PVC est approchée par

$$Var(\hat{\theta}_n^{(PVC)}) \simeq (I_n^*(\theta))^{-1} J_n^*(\theta) (I_n^*(\theta))^*.$$

Les matrices I_n , I_n^* et J_n^* peuvent être approchées par méthode de Monte Carlo sous $\hat{\theta}_n$.

3.5.3 La méthode de codage

Soit X un champ de Markov sur (S, \mathcal{G}) . Rappelons que $C \subset S$ est un *sous-ensemble de codage* de (S, \mathcal{G}) si deux points distincts de C ne sont jamais \mathcal{G} -voisins. Dans ce cas, les *variables conditionnelles* $\{(X_i | x_{\partial i}), i \in C\}$ sont *indépendantes*. La pseudo-vraisemblance de codage l_n^C sur $C_n = C \cap D_n$ est le produit sur les sites de C_n des densités conditionnelles, de logarithme,

$$l_n^C(\theta) = \sum_{C_n} \log \pi_i(x_i | x_{\partial i}, \theta)$$

l_n^C est une log-vraisemblance de variables indépendantes non identiquement distribuées (i.n.i.d.). Par rapport à l'estimation par PVC, on perd l'information apportée par les sites $i \in D_n \setminus C_n$, mais on garde l'expression d'une vraisemblance globale, ce qui permet de maintenir les résultats asymptotiques de la théorie du MV pour des variables i.n.i.d. en situation de sous-ergodicité.

Proposition 3.3 ([27];[28]) *Supposons que π_θ appartienne à une famille exponentielle et notons $I_n^C(\theta)$ la matrice de C-information conditionnelle,*

$$I_n^C(\theta) = \sum_{i \in C_n} Var\{\log \pi_i(x_i | x_{\partial i}; \theta)\}_\theta^{(1)}$$

Alors, si $\exists I(\theta)$, déterministe et d.p. telle que : $\liminf_n \frac{1}{c_n} I_n^C(\theta) \geq I(\theta)$, on a :

$$I_n(\hat{\theta}_{n,C})^{\frac{1}{2}} \{\hat{\theta}_{n,C} - \theta\} \xrightarrow{loi} \mathcal{N}_p(0, I_p)$$

L'information de Fisher $I_n(\hat{\theta}_{n,C})$ est approchée par la dérivée seconde $-(l_n^C)^{(2)}(\theta)$ prise en $\hat{\theta}_{n,C}$ ou par la somme sur C_n des variances empiriques calculées par méthode de Monte Carlo sous $\hat{\theta}_{n,C}$.

Plusieurs choix sont possibles pour l'ensemble de codage C mais on préférera un choix où C est maximal : par exemple, \mathbb{Z}^2 pour la relation aux 4-ppv est partitionné en deux codages maximaux $C^+ = \{(i, j) : i + j \text{ pair}\}$ et $C^- = \{(i, j) : i + j \text{ impair}\}$. A chaque codage C correspondra une estimation $\hat{\theta}_{n,C}$, mais ces estimations sont dépendantes.

Test du χ^2 de codage. Si $\theta = r(\varphi)$, $\varphi \in \mathbb{R}^q$, est une sous-hypothèse (H_0) de rang q et de classe \mathcal{C}^2 , alors

$$\text{Sous } (H_0) : 2\{l_n^C(\hat{\theta}_{n,C}) - l_n^C(\hat{\varphi}_{n,C})\} \xrightarrow{loi} \chi_{p-q}^2.$$

Exemple 3.2 *Mortalités par cancer dans la région de Valence (suite).*

Le modèle (3.14) : $(Y_i | y^i, x_i) \sim \mathcal{P}(\lambda_i(x_i, y_{\partial i}))$, $i = 1, n$ ($n = 263$ cantons), est à 4 paramètres, $\theta = (\alpha, (\beta_1, \beta_2), \gamma)$, s'écrivant, pour la covariable endogène de voisinage $a_i(y_{\partial i}) = \sum_{j : \langle i, j \rangle} a_{i,j} y_j$:

$$\log(\lambda_i) = \alpha + \log(u_i) + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma \times a_i(y_{\partial i}) \quad (3.18)$$

Le paramètre a a été calibré à $a = 0.025$, valeur pour laquelle l'estimation des paramètres présente une bonne stabilité.

MV : scores de Monte Carlo.

La loi jointe auto-poissonnienne existe si $\gamma \leq 0$, de log-densité :

$$\begin{aligned} l(\theta, y) &= -\log c(\theta) + \langle \theta, t(y) \rangle + d(y, u) \text{ avec} \\ c(\theta) &= \sum_{y \in N^n} \exp\left\{-\sum_{i=1,n} \log(y_i!) + \langle \theta, t(y) \rangle + d(y, u)\right\} \end{aligned}$$

pour la statistique exhaustive de \mathbb{R}^4 : $t_1(y) = \sum_{i=1,n} y_i$, $t_k(y) = \sum_{i=1,n} x_{i,k} y_i$ ($k = 2, 3$) et $t_4(y) = \sum_{i=1,n} y_i a_i(y_{\partial i})$. La méthode des scores de Fisher s'écrit alors,

$$\begin{aligned} \theta^{(l+1)} &= \theta^{(l)} + I(\theta^{(l)})^{-1} \nabla l(\theta^{(l)}, y) \text{ où} \\ I(\theta) &= \text{Var}_{\theta}(t(Y)), \quad \nabla l(\theta, y) = t(y) - E_{\theta}(t(Y)) \end{aligned}$$

sont respectivement l'information de Fisher et le gradient de la log-densité pris au point θ .

La difficulté dans l'implémentation de cet algorithme est que calcul de l'espérance et de la variance sous θ . La méthode des *Scores de Monte Carlo* consiste à estimer empiriquement $E_{\theta}(t(Y))$ et $\text{Var}_{\theta}(t(Y))$ sur la base d'un m -échantillon $\{y^{(h)}, h = 1, m\}$ de Y sous θ . Y étant un champ de Markov, cette simulation s'obtient par échantillonneur de Gibbs (cf. §5.2.1).

Une alternative numérique est de calculer directement par méthode MCMC la vraisemblance $l(\theta, y)$ puis d'utiliser un algorithme itératif d'optimisation de cette fonction (cf. §4.7; [25]).

Estimation par PVC.

La log-PVC $l_{pvc}(\theta, y)$ se calcule facilement. Le modèle auto-poissonnien étant log-linéaire, la maximisation de $l_{pvc}(\theta, y)$ s'obtiendra en utilisant un logiciel dédié aux modèles linéaires généralisés (MLG). Si cet algorithme fournit le bon estimateur $\hat{\theta}_n^{pvc}$, les variances proposées ne sont pas les bonnes, les variables conditionnelles $\{(Y_i | y^i, x_i), i = 1, n\}$ n'étant pas indépendantes, contrairement à l'hypothèse classique d'un MLG. Si on avait utilisé une PVC de C -codage, alors les variances calculées seraient les bonnes, les $\{(Y_i | y^C, x_i), i \in C\}$ étant elles, et conditionnellement aux covariables y^C , indépendantes.

Les résultats.

Les optimisations ont été effectuées sans contraintes sur γ : pour les 4 types de cancer, Ferrandiz et altri [22] trouvent $\hat{\gamma} < 0$. L'interprétation est que les covariables $a_i(y_{\partial i}) = \sum_{j : \langle i, j \rangle} a_{i,j} y_j$ prennent en charge des facteurs de *risques spatiaux "compétitifs"* absents du modèle.

Le tableau ci-dessous donne les résultats pour les *cancers de la prostate* par deux méthodes (MV et PVC) et 4 modèles, "constant" (α), "AR-poisson" ((α, γ)), "Régression" ($(\alpha, (\beta_1, \beta_2))$) et le modèle de Régression-Auto-régression (3.14) à 4 paramètres ($\alpha, (\beta_1, \beta_2), \gamma$) :

Modèle	α	γ	β_1	β_2	χ^2	$st(\chi^2)$
Constant	-7.3	*****	*****	*****	483.1	5.95
AR-Poisson : PVC	-7.76	-6.28 10 ⁻⁹	*****	*****	384.2	3.11
MV	-7.76	-6.52 10 ⁻⁹	*****	*****	363.2	2.65
Régression	-8.91	*****	2.96	-1.96 10 ⁻³	323.2	1.83
Complet : PVC	-8.776	-2.83 10 ⁻⁹	2.69	-2.15 10 ⁻³	333.0	2.12
MV	-8771	-2.80 10 ⁻³	2.67	-2.17 10 ⁻³	309.1	1.36

Pour les modèles avec composante AR, le MV est préférable à la PVC. Un indicateur de qualité du modèle est donné par la distance du χ^2 entre les observations y_i et leurs prédictions $\hat{y}_i = \lambda_i(\hat{\theta}; x_i, y_{\partial i})$ pour la distance renormalisée :

$$\chi^2 = \sum_{i=1,n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}$$

Si le modèle est bon, χ^2 suit grossièrement (il y a dépendance des y_i) un χ_n^2 . On rejettera le modèle au niveau 5% si $\chi^2 \geq 1.96(n + \sqrt{2n}) \sim 300.6$: pour cette règle, seule la régression-AR est acceptable. L'examen séparé des résidus $\frac{(y_i - \hat{y}_i)}{\sqrt{\hat{y}_i}}$ permet de détecter des erreurs de spécification du modèle.

Une autre procédure de validation est obtenue par Bootstrap paramétrique : pour un modèle donné et l'estimation $\hat{\theta}$ de son paramètre, on simule un m -échantillon $\{y^{(h)}(x, \hat{\theta}), h = 1, m\}$ de Y sous θ et sous x . Pour chaque

simulation $y^{(h)}$, on estime θ puis on calcule les prédictions $\{\hat{y}_i^{(l)}\}$ sur la base du modèle $\hat{\theta}^{(h)}$ ainsi que les valeurs associées $\{\chi^{2(l)}\}$: la statistique $st(\chi^2)$ est le quantile empirique au niveau 95% de la distribution centrée réduite de l'échantillon $\{\chi^{2(l)}, h = 1, 100\}$: si cette valeur est ≤ 1.96 , on validera le modèle.

Exemple 3.3 *Statistique d'un CAR gaussien : modélisation d'un essai à blanc.*

Les données de Mercer et Hall [35] mesurent les rendements en blé d'un essai à blanc (pas de traitement) sur un champ rectangulaire découpé en 20×25 parcelles de même taille $2.5 \text{ m} \times 3.3 \text{ m}$. Une analyse graphique préalable suggère qu'il n'y a pas d'effet ligne (cf. Fig. 3.2-b) . On étudie alors la régression spatiale avec le seul

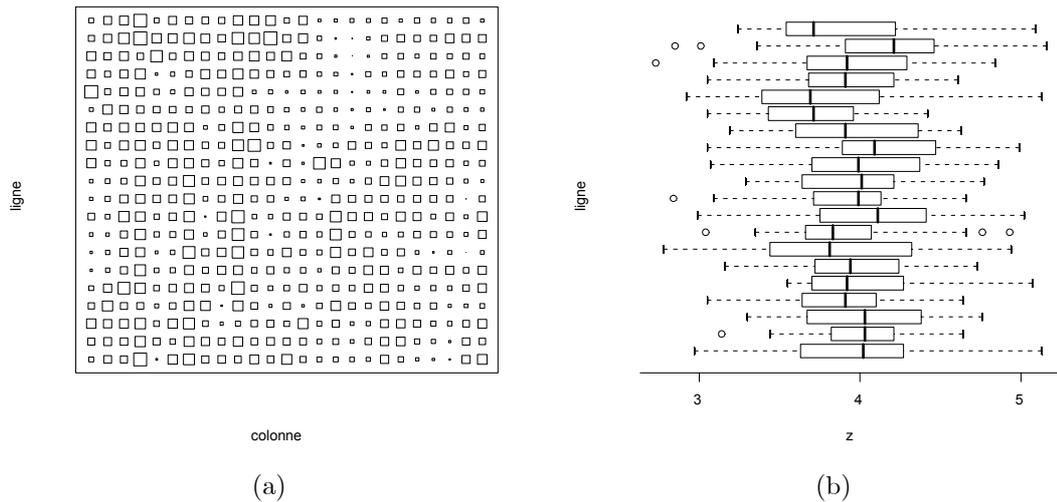


FIG. 3.2 – (a) Données de Mercer et Hall : rendements de blé sur un champ de 20×25 parcelles de même taille ; les dimensions des symboles sont proportionnelles au rendement. (b) Boîtes a moustaches par lignes.

effet colonne et un résidu $CAR(L)$ stationnaire,

$$X_{i,j} = \beta_j + \varepsilon_{i,j} \text{ et } \varepsilon_t = \sum_{s \in L(h)} c_s \varepsilon_{t+s} + e_t, \text{ Var}(e_t) = \sigma^2$$

pour des voisinages $L(h) = \{(k, l) : 0 < |k| + |l| \leq h\}$, $h = 0, 1$. Le tableau 3.1 présente les estimations du CAR par codage et par MV. Pour $L(1)$, on a retenu le codage $C^+ = \{(i, j) : i + j \text{ pair}\}$.

On teste alors $(H_0) : L(0)$, “les résidus sont indépendants”, dans $(H_1) = L_{iso}(1)$, “les résidus sont isotropiques aux 4-ppv”. Les deux tests de différence de contraste, celui de codage et comme celui du MV, de statistique $T_n^a = 2\{l_n^a(\hat{\theta}_{n,a}) - l_n^a(\hat{\varphi}_{n,C})\}$, rejettent l'indépendance (H_0) . Les tests correspondants de $(H_1) : L_{iso}(1)$ dans $(H'_1) = L(1)$, “le résidu est un CAR aux 4-ppv”, concluent à l'isotropie.

Méthode d'estimation paramètres	Codage			MV		
	\hat{c}_{10}	\hat{c}_{01}	l_n^C	\hat{c}_{10}	\hat{c}_{01}	l_n
$L(0)$: indépendance	-	-	-104.092	-	-	488.046
$L_{iso}(1)$: isotropie $c_{01} = c_{10}$	0.196	0.196	-69.352	0.205	0.205	525.872
$L(1)$: modèle aux 4-ppv	0.212	0.181	-69.172	0.233	0.177	526.351

TAB. 3.1 – Données de Mercer et Hall : estimation par codage et par MV.

3.5.4 Précisions comparées du MV, MPVC et du Codage

Intuitivement, on perd de l'information en passant du MV au MPVC et du MPVC au Codage. La justification théorique de cette intuition n'est possible que si on sait calculer les variances asymptotiques pour chacune des méthodes. Un exemple est celui d'un champ X gaussien markovien, isotropique aux 2-d ppv sur \mathbb{Z}^d [8] :

$$X_t = \beta \sum_{s: \|s-t\|_1=1} X_s + e_t$$

Si $|\beta| < 1/2d$, X est stationnaire, ergodique. Notant ρ_1 la corrélation à distance 1 et si X est observé sur le cube à n points, les précisions du MV , PMV et du codage sont,

$$\begin{aligned}\lim(n \times \text{Var} \widehat{\beta}_{MV}) &= \frac{1}{2} (2\pi)^{-d} \int_{T^d} \left\{ \frac{\sum_1^d \cos \lambda_i}{1 - 2\beta \sum_1^d \cos \lambda_i} \right\}^2 d\lambda_1 \cdots d\lambda_d \\ \lim(n \times \text{Var} \widehat{\beta}_{PMV}) &= \frac{2\beta^2(1 - 2d\rho_1)^2}{2d\rho_1^2} \text{ si } \beta \neq 0, = 1/d \text{ sinon} \\ \lim(n \times \text{Var} \widehat{\beta}_C) &= \frac{\beta(1 - 2d\beta\rho_1)}{d\rho_1} \text{ si } \beta \neq 0, = 1/d \text{ sinon}\end{aligned}$$

Pour $\beta = 0$, MV , PMV et codage ont la même efficacité. Le tableau 3.2 donne, pour $d = 2$, les efficacités relatives $e_1 = MV/PMV$ et $e_2 = MV/Codage$ ainsi que la corrélation ρ_1 aux ppv pour différentes valeurs de β .

4β	0.0	0.1	0.2	0.3	0.4	0.6	0.8	0.9	0.95	0.99
ρ_1	0.0	0.03	0.05	0.08	0.11	0.17	0.27	0.35	0.60	0.73
e_1	1.00	0.99	0.97	0.92	0.86	0.68	0.42	0.25	0.15	0.04
e_2	1.00	1.00	0.99	0.97	0.95	0.87	0.71	0.56	0.42	0.19

TAB. 3.2 – Efficacités relatives $e_1 = MV/Codage$, $e_2 = MV/PMV$ et corrélation $\rho_1(\beta)$ aux ppv pour un champ gaussien isotropique aux 4-ppv sur \mathbb{Z}^2 .

3.6 Exercices

Exercice 3.1 Modèle de Gibbs sur un réseau triangulaire plan

(a) Identifier le modèle stationnaire aux 6 plus proches voisins sur le réseau triangulaire plan à 3 états $E = \{a, b, c\}$: cliques, potentiels, dimension paramétrique, loi conditionnelle en un site, sur une partie. Expliciter le modèle si $E \subset \mathbb{R}$.

(b) Même question mais : (i) le modèle est isotropique ; (ii) le modèle a au plus des potentiels de paires ; (iii) les potentiels Φ_A sont invariants par permutation sur A .

(c) Mêmes questions pour un modèle à deux états.

Exercice 3.2 Champ de Markov pour la segmentation de texture.

Sur $S = \{1, 2, \dots, n\}^2$, on modélise un champ (X, Λ) à valeurs dans $E = \mathbb{R} \times \{1, 2, \dots, K\}$ de la façon hiérarchique suivante :

(i) Λ , qui repère la texture, est le modèle de Potts échangeable aux 8-ppv (3.6) ;

(ii-1) $(X_i \mid \lambda_i = k) \sim \mathcal{N}(0, \sigma_k^2)$ est une texture de rugosité ;

(ii-2) $(X_i \mid \lambda_i = k) \sim \mathcal{N}(\mu_k, \sigma^2)$ est une texture de niveau de gris ;

(ii-3) sur une plage de label constant k , X est un CAR gaussien isotropique aux 4-ppv de paramètre (α_k, σ_e^2) , $0 \leq \alpha_k < 1/4$ (texture de covariation).

Expliciter les différents modèles (cliques, potentiels). On observe X : déterminer les lois $(\Lambda \mid X)$, $(\Lambda_i \mid \Lambda^i, X)$. Utilisant l'échantillonneur de Gibbs (cf. §5.2), simuler $(\Lambda \mid X)$.

Exercice 3.3 Contraintes de recollement.

$S = \{1, 2, \dots, n\}$, $E = \{0, 1, 2, \dots, K-1\}$ et $\mathcal{F} = \{\nu_i(x_i \mid x^i), i \in S\}$ est une famille non-contrainte de lois conditionnelles.

(a) Quelle est la dimension paramétrique de \mathcal{F} ? Quel est le nombre de contraintes à imposer si on veut que \mathcal{F} s'identifie aux lois conditionnelles d'une loi jointe π sur E^S ?

(b) Pour $n = K = 2$, écrire la contrainte que doivent vérifier les 4 paramètres des 4 lois conditionnelles afin que celles-ci se recollent.

(c) On considère le noyau aux 2-ppv sur \mathbb{Z} à états dans $E = \{0, 1, \dots, K-1\}$:

$$Q(y \mid x, z) = P(X_0 = y \mid X_{-1} = x, X_{+1} = z)$$

Quelle est la dimension paramétrique de Q ? A quelles conditions ces lois conditionnelles sont-elles celles d'un champ de Markov aux 2-ppv ?

Exercice 3.4 Modèle causal et écriture bilatérale associée.

(a) Soit Y une chaîne de Markov sur $E = \{-1, +1\}$ de transitions $p = P(Y_i = 1/Y_{i-1} = 1)$ et $q = P(Y_i = -1/Y_{i-1} = -1)$. Démontrer que Y est un champ de Markov bilatéral aux 2-ppv. En déduire que le noyau conditionnel bilatéral s'écrit,

$$Q(y|x, z) = Z^{-1}(x, z) \exp\{x(\alpha + \beta(y + z))\}$$

où $\alpha = (1/2) \log(p/q)$ et $\beta = (1/4) \log\{pq/[(1-p)(1-q)]\}$. Interpréter les situations : $\alpha = 0$; $\beta = 0$.

(b) Sur \mathbb{Z} et pour $E = \{0, 1\}$, donner la représentation bilatérale markovienne du processus de Markov homogène de mémoire 2 et de transition : $P(Y_i = 1/Y_{i-2} = a, Y_{i-1} = b) = p(a, b)$. Un champ de Markov aux 4-ppv sur \mathbb{Z} est-il en général une chaîne de Markov de mémoire 2 ?

(c) On considère sur $S = \{1, 2, \dots, n\}$ le modèle d'énergie :

$$U(y) = a \sum_{i=1}^n y_i + b \sum_{i=1}^{n-1} y_i y_{i+1} + c \sum_{i=1}^{n-2} y_i y_{i+2}.$$

On partitionne S en $I \cup P$, où I est l'ensemble des indices impairs, P celui des indices pairs. Vérifier que (Y_I/y_P) est une chaîne de Markov.

(d) On considère sur \mathbb{Z}^2 un champ Y binaire causal pour l'ordre lexicographique, de loi conditionnelle en (i, j) dépendant des seuls sites $(i-1, j)$ et $(i, j-1)$ du passé. Donner la modélisation bilatérale de Y .

Exercice 3.5 Chaîne et champ de Markov échantillonnés.

(a) Une chaîne $Y = (Y_1, Y_2, \dots, Y_n)$ de loi initiale $\nu \sim Y_1$ et de transitions $\{q_i, i = 1, n-1\}$ a pour loi : $P_\nu(y) = \nu(y_1) \prod_{i=1, n-1} q_i(y_i, y_{i+1})$. Montrer que Y observée un instant sur deux est une chaîne de Markov. Expliciter sa loi (transition, potentiels) si Y est homogène à états dans $E = \{0, 1\}$.

(b) Y , un champ de Markov aux 4-ppv sur le tore $S = \{1, 2, \dots, n\}^2$, est observé sur $S^+ = \{(i, j) \in S : i + j \text{ est paire}\}$. Montrer que Y_{S^+} est un champ de Markov de cliques maximales $\{C_{ij} = \{(i, j), (i+2, j), (i+1, j+1), (i+1, j-1)\}, (i, j) \in S^+\}$. Déterminer la loi de Y_{S^+} si Y est stationnaire et $E = \{0, 1\}$.

(c) Montrer que la propriété de Markov du (b) se perd : (i) si Y est markovien aux 8-ppv; (ii) si on échantillonne Y sur $S_2 = \{(i, j) \in S : i \text{ et } j \text{ pairs}\}$.

Exercice 3.6 Estimation d'un modèle CAR aux 4-ppv.

X est un CAR gaussien centré, stationnaire, aux 4-ppv, de paramètre $\theta = (\alpha, \beta, \sigma^2)$, où σ^2 est la variance résiduelle conditionnelle, $|\alpha| + |\beta| < 1/2$. X est observé sur le carré de côté n .

1. Préciser la loi asymptotique de l'estimateur du MV (resp. du MPV) de θ si les données sont convenablement rabotées au bord. Tester l'isotropie.
2. Même question si $Y = X + \varepsilon$ où ε est un BBG(σ^2) indépendant de X .

Exercice 3.7 Quelques tests du χ^2 de codage.

X est un champ de Markov de spécification invariante par translation sur $\{0, 1, \dots, n-1\}^2 \subset \mathbb{Z}^2$.

1. Décrire le test d'isotropie si X est le modèle d'Ising aux 4-p.p.v..
2. On considère le modèle isotropique (3.5) à K états $E = \{1, 2, \dots, K\}$ aux 4-ppv et le sous-modèle (E) échangeable (3.6). Décrire le test d'échangeabilité (E).
3. Soit X un champ V -markovien à K états, de spécification invariante par translation, où $V = \partial 0 \subset \mathbb{Z}^2$ est finie et symétrique, $0 \notin V$.
 - (a) Ecrire le modèle général de X (faire un choix de V et de K).
 - (b) Décrire les sous-modèles : (a) les cliques ont au plus 2 points; (b) le modèle est isotropique; (c) : (a) \cap (b); (d) le modèle est auto-binomial sur $E = \{0, 1, \dots, K-1\}$.
 - (c) Proposer des tests de ces sous-modèles dans le modèle général.

Exercice 3.8 Indépendance de 2 caractères spatiaux.

$(U_i, V_i) \in \{0, 1\}^2$ sont 2 caractères binaires observés sur $\{1, 2, \dots, n\}^2$. Le modèle de (U, V) est markovien isotropique aux 4-p.p.v. et invariant par translation.

1. Vérifier que le modèle général est à 12 paramètres, de potentiels :
 - (i) de singletons : $\phi_1(u, v) = \alpha u + \beta v + \gamma uv$;
 - (ii) de paires :

$$\phi_2((u, v), (w, t)) = \delta_1 uw + \delta_2 vt + \delta_3 ut + \delta_4 vw + \delta_5 uwt + \delta_6 vwt + \delta_7 uvw + \delta_8 uvt + \delta_9 uvwt.$$

2. Spécifier et tester le sous-modèle d'indépendance de U et V .

3. Soit (ω) le sous-modèle où, dans ϕ_2 , seuls δ_1 et δ_2 sont non-nuls. Déterminer les lois conditionnelles $\pi_i(u_i, v_i \mid \cdot)$ et $\nu_i(u_i \mid \cdot)$, le conditionnement portant sur toutes les autres observations de u et de v . Construire le test du χ^2 de codage de la sous-hypothèse d'indépendance de U et de V sur la base des lois conditionnelles $\nu_i(\cdot \mid \cdot)$.

Chapitre 4

Processus ponctuels spatiaux

Ce chapitre étudie la situation où c'est la répartition spatiale $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ des points de $S \subset \mathbb{R}^2$ où ont lieu les observations qui est aléatoire. On dit que \mathbf{x} est la réalisation d'un processus ponctuel (PP) X sur S . La figure 4.1 donne quatre exemples : (a) est une répartition "au hasard", aucune structure spatiale n'apparaît ; la répartition (b) est plus régulière, chaque point se préservant un espace propre autour de lui alors que pour (c) les points ont tendance à s'agréger. L'exemple (d) montre en plus des variables attachées aux points de la répartition.

Notre objectif est de donner une approche heuristique de ces modèles de processus ponctuels définis sur un fermé S de \mathbb{R}^d . Certaines notions requièrent des justifications techniques non abordées ici [47]. On notera \mathcal{B}_d (resp. $\mathcal{B}(S)$, $\mathcal{B}_b(S)$) l'ensemble des boréliens de \mathbb{R}^d (resp. des boréliens de S , des boréliens bornés de S), ν_d (ν s'il n'y a pas d'ambiguïté) la mesure de Lebesgue sur \mathcal{B}_d .

4.1 Définitions et notations

La réalisation \mathbf{x} d'un PP X sur S est la donnée d'une collection de points

$$\mathbf{x} = \{x_i, i \in I\}, x_i \in S.$$

4.1.1 Espace exponentiel des configurations

Une réalisation \mathbf{x} est dite *localement finie* si elle n'a pas de points d'accumulation dans \mathbb{R}^d ; dans ce cas, le nombre $n(\mathbf{x})$ de points de \mathbf{x} est fini si S est fermée et bornée, dénombrable sinon. Un PP est *simple* s'il n'y a pas de répétition de points : si X est simple, \mathbf{x} coïncide avec un sous-ensemble $\{x_1, x_2, \dots, x_n\} \subseteq S$ si $n = n(\mathbf{x})$. Tous les PP que nous considérerons ici sont simples et à configuration presque sûrement localement finie.

L'espace des configurations E d'un PP sur S s'identifie avec la réunion des espaces E_n des configurations à n points sur S , $n \geq 0$: $E = \bigcup_{n \geq 0} E_n$, l'*espace exponentiel* des configurations, est muni de la tribu \mathcal{E} qui rend mesurables les variables de comptage $N(A) : E \rightarrow \mathbb{N}$, $N(A)$ comptant le nombre de points de X dans A , $A \in \mathcal{B}_b(S)$. Des exemples d'événements associés à un PP sont : "il y a au plus 50 points dans la configuration \mathbf{x} " ; "les points de \mathbf{x} sont distants d'au moins r , $r > 0$ donné" ; "0 est un point de \mathbf{x} ", "il n'y a aucun point dans A ".

On identifiera les événements $F \in \mathcal{E}_n$ avec les parties $F \subset S^n$ invariantes par permutations $\sigma \in \Xi_n$ des coordonnées, c'est-à-dire, notant $\mathbf{x}_\sigma = (x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$ si $x = (x_1, x_2, \dots, x_n)$, les parties $F \subset S^n$ telles que :

$$\text{pour toute permutation } \sigma \in \Xi_n : \mathbf{x}_\sigma \in F \iff \mathbf{x} \in F$$

Définition 4.1 *Un processus ponctuel X sur S est la donnée d'une collection aléatoire $X = \{X_i\} \subset S$ localement finie t.q., $\forall A \in \mathcal{B}_b(S)$, $N_X(A) : E \rightarrow \mathbb{N}$, le nombre de points de X dans A , est une variable aléatoire.*

4.1.2 Loi d'un processus ponctuel.

La loi d'un PP est une probabilité P sur (E, \mathcal{E}) . Cette loi est caractérisée par les lois des v.a. $N(A)$ où A parcourt une famille génératrice de $\mathcal{B}_b(S)$: la distribution finie dimensionnelle d'un processus ponctuel X est la donnée pour tout $m \geq 1$ et tout m -uplet (A_1, A_2, \dots, A_m) de $\mathcal{B}_b(S)$ des distributions sur \mathbb{N}^m de $(N(A_1), N(A_2), \dots, N(A_m))$.

Un PP sur \mathbb{R}^d est *stationnaire* si, $\forall x \in \mathbb{R}^d$, la loi du PP translaté $X_x = \{X_i + x\}$ est la même que celle de X ; le PP est *isotropique* si de plus la loi de ρX , X après ρ -rotation, est identique à celle de X pour toute rotation ρ .

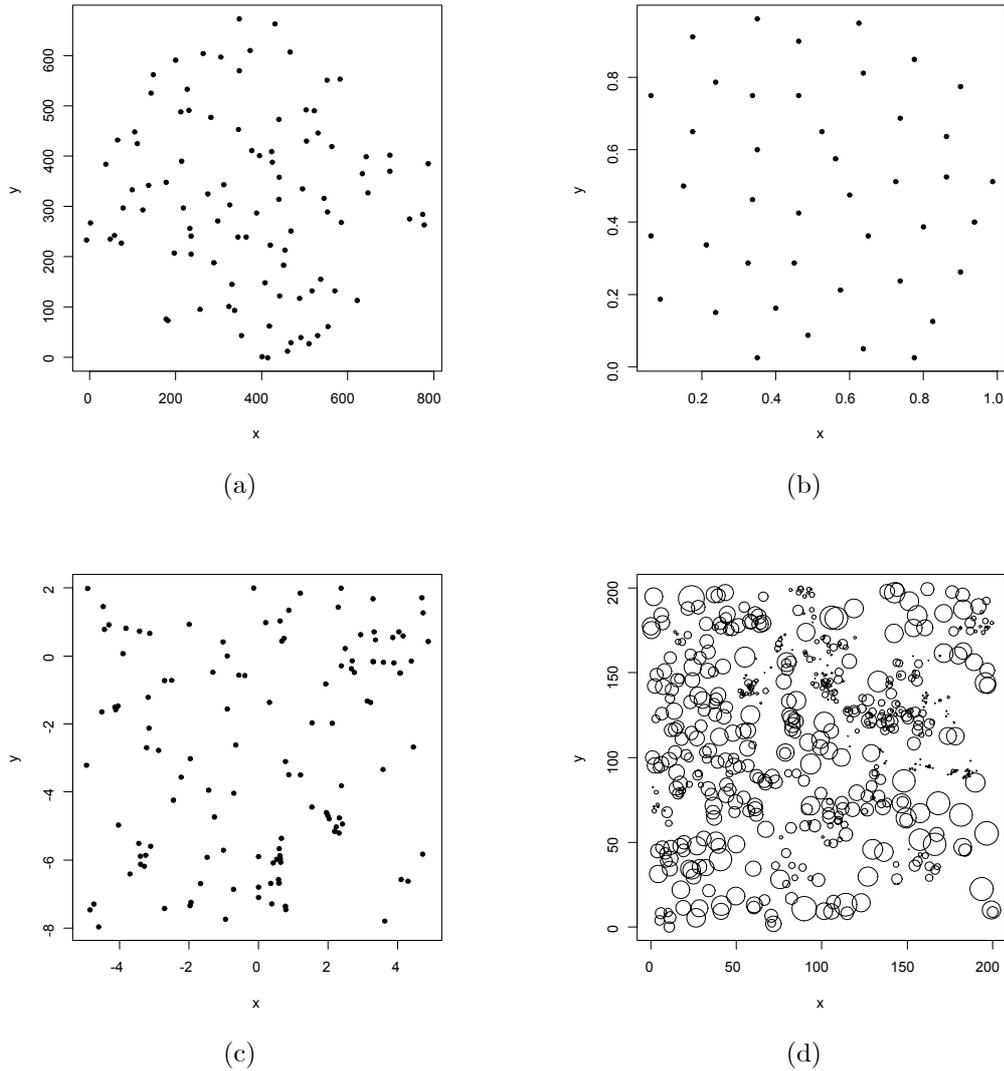


FIG. 4.1 – Exemples de répartition ponctuelle : (a) 97 fourmilières (données `ants` de `spatstat`) ; (b) 42 centres de cellules d’une section histologique observée au microscope (données `cells` de `spatstat`) ; (c) 126 pins d’une forêt finlandaise (données `finpines` de `spatstat`) ; (d) Positions et tailles de 584 aiguilles de pins dans un sous-bois (données `longleaf` de `spatstat`).

Processus ponctuel marqué. Soit S un fermé de \mathbb{R}^d et K un espace métrique complet séparable. Un *processus ponctuel marqué* (PPM) (X, M) sur $S \times K$ est un *PP* sur $S \times K$ tel que X soit un *PP* sur S .

Des exemples d’espaces de marques sont : $K = \{m_1, m_2, \dots, m_K\}$ (K types de points), $K = \mathbb{R}^+$ (la marque est un disque centré en x de rayon $r \geq 0$; (cf. Fig. 4.1-(d)), $K = [0, 2\pi[\times \mathbb{R}^+$ (la marque est un segment centré en x , d’orientation $\theta \in [0, 2\pi[$ et de longueur $l \geq 0$).

4.2 Exemples de processus ponctuels

4.2.1 Processus Binomial.

Soit $S \subset \mathbb{R}^d$ bornée de volume $\nu(S) > 0$. Un *PP binomial* à n points est constitué de n points i.i.d. uniformes sur S : si $\{A_1, A_2, \dots, A_k\}$ est une partition borélienne de S ,

$$(N(A_1), N(A_2), \dots, N(A_k)) \sim \mathcal{M}(n; q_1, q_2, \dots, q_k)$$

la loi multinomiale de paramètres n et $q_i = \nu(A_i)/\nu(S)$, $i = 1, k$.

4.2.2 Processus Ponctuel de Poisson (PPP)

PPP homogène.

Un *processus ponctuel de Poisson* (PPP) homogène d'intensité $\lambda > 0$ sur un borélien S borné est caractérisé par les deux propriétés suivantes :

1. $N(S)$ suit une loi de Poisson de paramètre $\lambda\nu(S)$;
2. Si $N(S) = n$, les n points sont i.i.d. uniformes sur S ,

$$P(N(S) = n) = e^{-\lambda\nu(S)} \frac{(\lambda\nu(S))^n}{n!}, \text{ et } g_n(\{x_1, x_2, \dots, x_n\}) = |S|^{-n}.$$

Une deuxième caractérisation d'un PPP(λ) est la suivante :

- 1*. Pour tout borélien A , $N(A) \sim \mathcal{P}(\lambda\nu(A))$;
- 2*. Si (A_i) sont des boréliens disjoints, les v.a. $N(A_i)$ sont indépendantes.

FIG. 4.2 – (a) Réalisation d'un PPP homogène d'intensité $\lambda = 100$ sur $S = [0, 1]^2$; (b) réalisation du PP à noyau dur (4.1) avec $\beta = 200$, $\gamma = 0$ et $r = 0.05$.

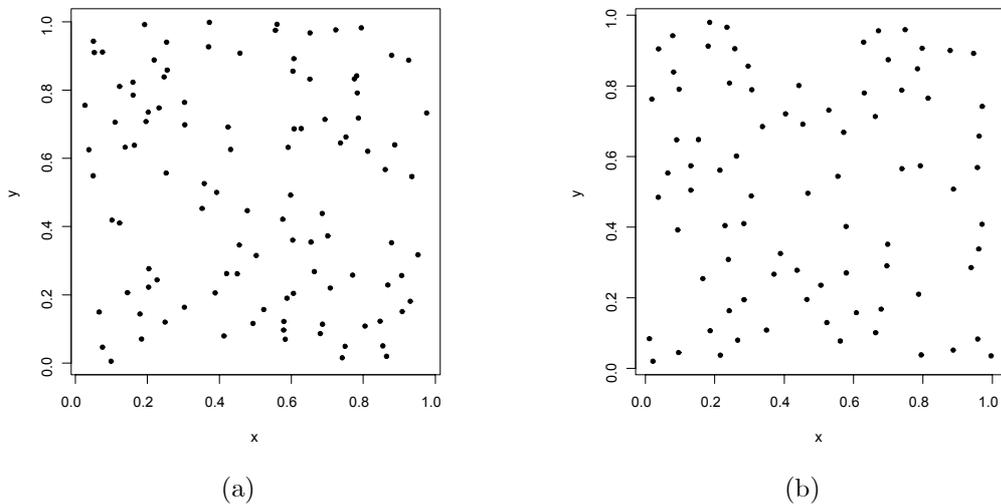


FIG. 4.3 – (a) Réalisation d'un PPP homogène d'intensité $\lambda = 100$ sur $S = [0, 1]^2$; (b) réalisation d'un PP à noyau dur (4.1) sur $S = [0, 1]^2$ avec $\beta = 200$, $\gamma = 0$ et $r = 0.05$.

Un PPP homogène modélise une répartition spatiale homogène et “au hasard” (cf. Fig. 4.3-a), hypothèse notée CSR par les anglo-saxons (“Complete Spatial Randomness”). D'autres distributions, tel le modèle à noyau dur (hard core model), sont *plus régulières*, chaque point préservant un certain espace vital autour de lui (cf. Fig. 4.3-b et §4.3.2). D'autres au contraire sont *moins régulières*, tel le modèle Neyman-Scott où des descendants s'agrègent autour de points parents. Ces modèles vont être définis au §4.3 à partir de leur densité inconditionnelle.

PP de Poisson inhomogène.

Soit $\lambda(\cdot)$ une *mesure d'intensité* sur $\mathcal{B}_b(S)$ positive et bornée sur les compacts.

Définition 4.2 X est un PPP inhomogène d'intensité $\lambda(\cdot)$ si :

- (i) $\forall A \in \mathcal{B}_b(S)$, $N(A) \sim \mathcal{P}(\lambda(A))$;
- (ii) Si $A, B \in \mathcal{B}_b(S)$ sont disjointes, $N(A)$ et $N(B)$ sont indépendantes.

La mesure $\lambda(\cdot)$ module localement le nombre moyen de points tout en préservant hasard et indépendance dans la répartition spatiale. On notera PPP(λ) un tel PPP. Pour un PPP homogène d'intensité δ , $\lambda(\cdot) = \delta \times \nu(\cdot)$ où $\nu(\cdot)$ est la mesure de Lebesgue.

La simulation d'un PPP inhomogène (cf. Fig. 4.4) est réalisée par la méthode suivante de rejet ou d'*effacement de points* : si la densité $\lambda(\cdot)$ est bornée par $c < \infty$, on réalise l'algorithme :

1. **Simuler** $X_h = \{x_i\}$, un PPP homogène sur S d'intensité c .
2. **Effacer indépendamment les x_i** avec la proba $(1 - \lambda(x_i)/c)$.

Le processus X après effacements est un PPP(λ) : en effet, considérant dx le volume infinitésimal autour de x , $P(N_X(dx) = 1) = P(N_{X_h}(dx) = 1 \text{ et } x \text{ non-effacé}) = \lambda(x)dx$.

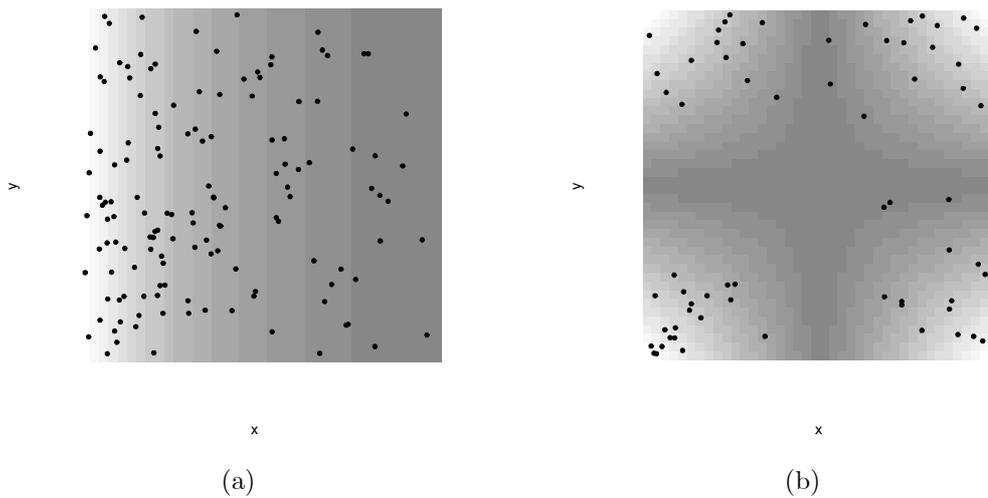


FIG. 4.4 – Réalisation d'un PPP inhomogène sur $[0, 1]^2$ d'intensité (a) $\lambda(x, y) = 400e^{-3x}$ ($E(N(S)) = 126.70$); (b) $\lambda(x, y) = 800|0.5 - x||0.5 - y|$ ($E(N(S)) = 50$).

Plus régulier : le modèle à noyau dur.

Une façon de “régulariser” une configuration spatiale est d'interdire les points trop proches, ce qui est bien adapté chaque fois que l'individu i placé en x nécessite un espace propre $B(x_i, r)$ (répartition d'arbres dans une forêt, d'animaux sur un territoire). Ces modèles seront définis au §4.3.2 à partir de leur densité inconditionnelle.

Des agrégats : le PP de Neyman-Scott.

Plaçons-nous dans le contexte suivant d'une dynamique de population :

1. X , qui repère la position des parents, est un PPP(λ);
2. à la génération suivante, chaque parent x_i engendre des descendants Y_{x_i} en nombre K_{x_i} et en position Y_{x_i} autour de x_i , variables i.i.d..

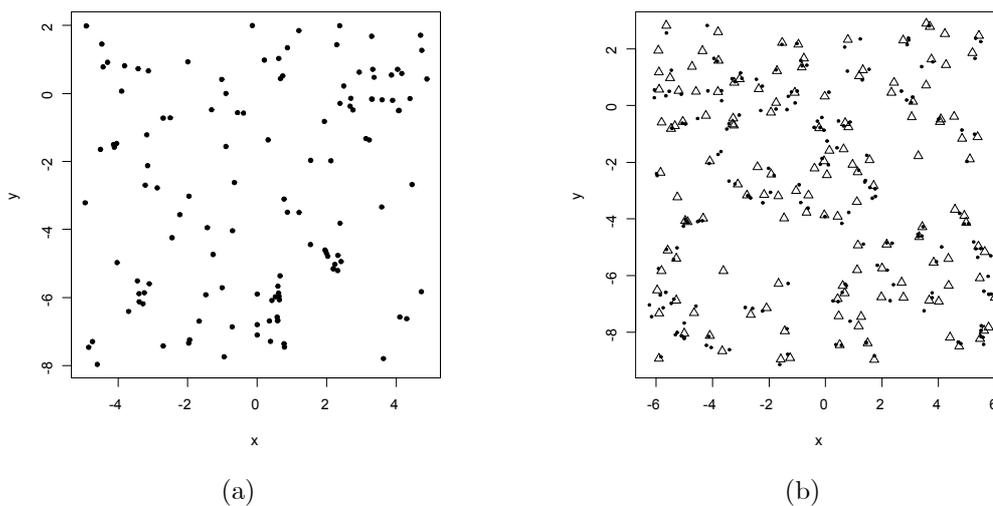


FIG. 4.5 – (a) Données réelles : localisations de 126 pins dans une forêt finlandaise (données `finpines` de `spatstat`); (b) simulation d'un processus de Neyman-Scott. Le nombre de descendants K suit une loi de Poisson de moyenne μ , et la position d'un descendant (\bullet) autour d'un parent (\square) suit une loi $\mathcal{N}_2(0, \sigma^2 I_2)$. Les paramètres $\hat{\mu} = 0.672$, $\hat{\lambda} = 1.875$, $\hat{\sigma}^2 = 0.00944$, sont estimés par MC sur la fonction K à partir des données (a).

Le processus de Neyman-Scott est la superposition $D = \cup_{x_i \in X} Y_{x_i}$ de tous les descendants (cf. Fig. 4.5); il présente des agrégats autour des parents. Les généralisations sont multiples : autre choix de répartition spatiale

des parents X , dépendances entre les descendants (compétitions), lois de descendances non-identiques (variabilité de fertilité parentale), etc....

PP de Cox ou PP de Poisson doublement stochastique.

C'est un PPP d'intensité $\lambda(\cdot)$ réalisation d'un processus stochastique $Z(\cdot)$: si $Z = \{Z_s \geq 0, s \in S\}$ vérifie pour tout B borné, $\int_B Z_s ds < \infty$ p.s., la distribution de X conditionnelle à Z est un PPP d'intensité $\Lambda(B) = \int_B Z_s ds$, $B \subseteq S$. Ce processus est aussi appelé *processus de Cox* modulé par Z .

Le processus de Cox résulte d'un contexte où l'environnement (l'intensité) est aléatoire, réalisation d'un processus aléatoire. [?] définissent un PP de Cox log-gaussien en prenant pour $\lambda(\cdot)$ un modèle log-linéaire avec un effet aléatoire gaussien,

$$\log \lambda(s) = {}^t z(s)\beta + \Psi(s),$$

où $\Psi = (\Psi(s))$ est un champ gaussien centré. La corrélation de Ψ contrôlera la dépendance spatiale du PP.

Processus ponctuel marqué.

Soit $B(x, r) \subset S$ la boule fermée de centre x et de rayon $r > 0$. Un exemple de PPM est la donnée de centres $X = \{x_i\} \subset \mathbb{R}^2$, réalisations d'un PPP(λ) et de marques $B(x_i, r_i)$ en x_i de rayons suivant des lois exponentielles $\mathcal{Exp}(r^{-1})$ i.i.d., $r > 0$, indépendantes de X . En morphologie mathématique, l'ensemble aléatoire $\mathcal{X} = \cup_{x_i \in X} B(x_i, r_i)$ est appelé un *processus booléen*. Pour un *processus de fibres*, on choisira des marques m_i attachées à x_i qui sont curvilignes, par exemple un segment centré en x_i , de longueur $l_i \sim \mathcal{Exp}(l^{-1})$ et d'orientation θ_i uniforme sur $[0, 2\pi[$ indépendante de l_i , les (θ_i, l_i) étant i.i.d. et indépendantes de X .

Un PP multivarié $X = (X(1), X(2), \dots, X(M))$ peut être vu comme un PPM à nombre fini de marques : pour chaque m , $X(m)$ est le sous-ensemble de S qui repère les positions de l'espèce m . X s'identifie au processus PPM $\tilde{X} = \cup_{m=1}^M X(m)$ superposition des $X(m)$. La figure 4.1-d donne un exemple de réalisation d'un PPM.

4.3 Densité d'un processus ponctuel

4.3.1 Définitions

Rappelons que l'on identifie un événement de E_n portant sur des réalisations à $n(\mathbf{x}) = n$ points par les sous-ensembles $F \subset S^n$ invariants par les permutations des coordonnées. On va d'abord définir la densité d'un PP conditionnelle à la réalisation de n points, puis la densité inconditionnelle d'un PP.

Densité conditionnelle à $n(\mathbf{x}) = n$ sur S

Notons Ξ_n l'ensemble des partitions de $\{1, 2, \dots, n\}$. Une densité conditionnelle à $n(\mathbf{x}) = n$ points est une fonction $f_n : S^n \rightarrow \mathbb{R}^+$ invariante par les permutations de Ξ_n et d'intégrale sur S^n égale à 1 :

$$\forall \sigma \in \Xi_n : f_n(x_1, x_2, \dots, x_n) = f_n(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) \geq 0 \text{ et } \int_{S^n} f_n(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1$$

On a, pour tout événement F de E_n :

$$P(X \in F \mid n(x) = n) = \int_F f_n(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Exemples. Un processus binomial à n points sur S est à densité uniforme :

$$f_n(x_1, x_2, \dots, x_n) = \frac{1}{\nu(S)^n} \mathbf{1}\{(x_1, x_2, \dots, x_n) \in S^n\}.$$

Un PPP($\lambda(\cdot)$) d'intensité $\lambda(\cdot)$ sur S borné, de mesure $\lambda(S) > 0$ finie, admet pour densité :

$$f_n(x_1, x_2, \dots, x_n) = \frac{1}{\lambda(S)^n} \mathbf{1}\{(x_1, x_2, \dots, x_n) \in S^n\} \lambda(x_1) \lambda(x_2) \dots \lambda(x_n).$$

Densité inconditionnelle par rapport au PPP(1)

La densité en \mathbf{x} d'un PP X sur S par rapport à celle de Y , un PPP(1), au point de réalisation $(n_S = n; \mathbf{x} = (x_1, x_2, \dots, x_n))$ de $\mathbb{N} \times S^n$ est, par définition :

$$f(\mathbf{x}) = \frac{P(n_S(X) = n; X \in dx_1 \times dx_2 \times \dots \times dx_n)}{P(n_S(Y) = n; Y \in dx_1 \times dx_2 \times \dots \times dx_n)}$$

Puisque $P(n_S(Y) = n) = \frac{e^{-\nu(S)}}{n!} \nu(S)^n$ pour un PPP(1), on a :

$$P(n_X(\mathbf{x}) = n; dx_1 \times dx_2 \times \cdots \times dx_n) = \frac{e^{-\nu(S)}}{n!} f(\mathbf{x}) dx_1 \times dx_2 \times \cdots \times dx_n$$

Ainsi :

$$p_n = P(N_S(X) = n) = \frac{e^{-\nu(S)}}{n!} \int_{S^n} f(x_1, x_2, \dots, x_n) dx_1 \times dx_2 \times \cdots \times dx_n$$

Pour être une densité, $f \geq 0$ doit satisfaire deux conditions :

- (i) $\forall n \geq 1, p_n = \frac{e^{-\nu(S)}}{n!} \int_{S^n} f(x_1, \dots, x_n) dx_1 \cdots dx_n < \infty$;
- (ii) $\int f(\mathbf{x}) d\mathbf{x} = \sum_{n \geq 0} p_n = 1$.

La première condition est une condition d'admissibilité; la deuxième une condition de normalisation, généralement impossible à vérifier du fait des intégrales multiples à calculer. Pour cette raison, une densité f sera définie à un facteur près,

$$f(\mathbf{x}) = cg(\mathbf{x}),$$

où g est admissible, c assurant la normalisation.

Comme pour un champ de Markov latticiel, la simulation d'un PP de densité f (resp. l'estimation d'un PP par PVC) n'exige pas de connaître c pour simuler X (resp. pour estimer θ). Nous donnons au §4.7 une procédure de Monte Carlo permettant de calculer la constante c .

4.3.2 Exemples

Processus de Poisson homogène PPP(λ).

C'est le PP de densité $f(\mathbf{x}) = e^{\{(1-\lambda)\nu(S)\}} \lambda^n$ si $n = n(\mathbf{x})$. On a :

- (i) $p_n = \exp(-\lambda\nu(S)) (\lambda\nu(S))^n / n!$;
- (ii) les n points sont i.i.d. uniformes sur S .

Processus de Poisson inhomogène d'intensité $\lambda : S \rightarrow \mathbb{R}^+$.

C'est le PP de densité $f(\mathbf{x}) = \exp\{\nu(S) - \lambda(S)\} \prod_{x_i \in \mathbf{x}} \lambda(x_i)^n$ si $n = n(\mathbf{x})$.

Famille exponentielle de PP de Gibbs.

Pour θ et $U(\mathbf{x}) \in \mathbb{R}^p$, une famille exponentielle de densité de PP est

$$f_\theta(\mathbf{x}) = c(\theta) \exp\langle \theta, U(\mathbf{x}) \rangle.$$

Si $U(x)$ dérive d'un potentiel Φ , $U = \sum_C \Phi_C$ (cf §4.5), on dit que X est un PP de Gibbs. Donnons un exemple.

PP de Strauss, PP à noyau dur.

Soit $r > 0$ un rayon fixé. La densité d'un PP de Strauss correspond au choix $\theta = (a, b)$, $U_1(x) = n(\mathbf{x})$ et $U_2(\mathbf{x}) = \sum_{i < j} \mathbf{1}(\|x_i - x_j\| \leq r) \doteq s(\mathbf{x})$. U_2 , associé aux potentiels de paires $\Phi_{\{x_i, x_j\}}(x) = \mathbf{1}(\|x_i - x_j\| \leq r)$, compte le nombre de paires " r -voisines" de \mathbf{x} . Notant $\beta = e^a$ et $\gamma = e^b$, la densité du PP de Strauss est

$$\text{Strauss : } f_\theta(\mathbf{x}) = c\beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}. \quad (4.1)$$

Le PPP homogène d'intensité β correspond à $\gamma = 1$; $\gamma = 0$ correspond au modèle à noyau dur, de densité :

$$r\text{-Noyau dur : } f_\theta(\mathbf{x}) = c\beta^{n(\mathbf{x})} \mathbf{1}\{\forall i \neq j, \|x_i - x_j\| > r\}$$

Le modèle à noyau dur exclut toute configuration avec des sites à distance $\leq r$.

Décrivons les distributions conditionnelles (cf. Fig. 4.6-a-b) et inconditionnelles (cf. Fig. 4.6-c-d) du PP de Strauss.

Conditionnellement à $n(\mathbf{x}) = n$, $f_\theta(x) \sim \gamma^{s(\mathbf{x})}$:

- Si $\gamma < 1$, le processus X est plus régulier qu'un PP binomial, d'autant plus que γ est proche de 0;
- $\gamma = 1$ correspond au processus binomial à n points.
- Si $\gamma > 1$, X est moins régulier qu'un PP binomial, d'autant que γ est grand : des agrégats apparaissent.

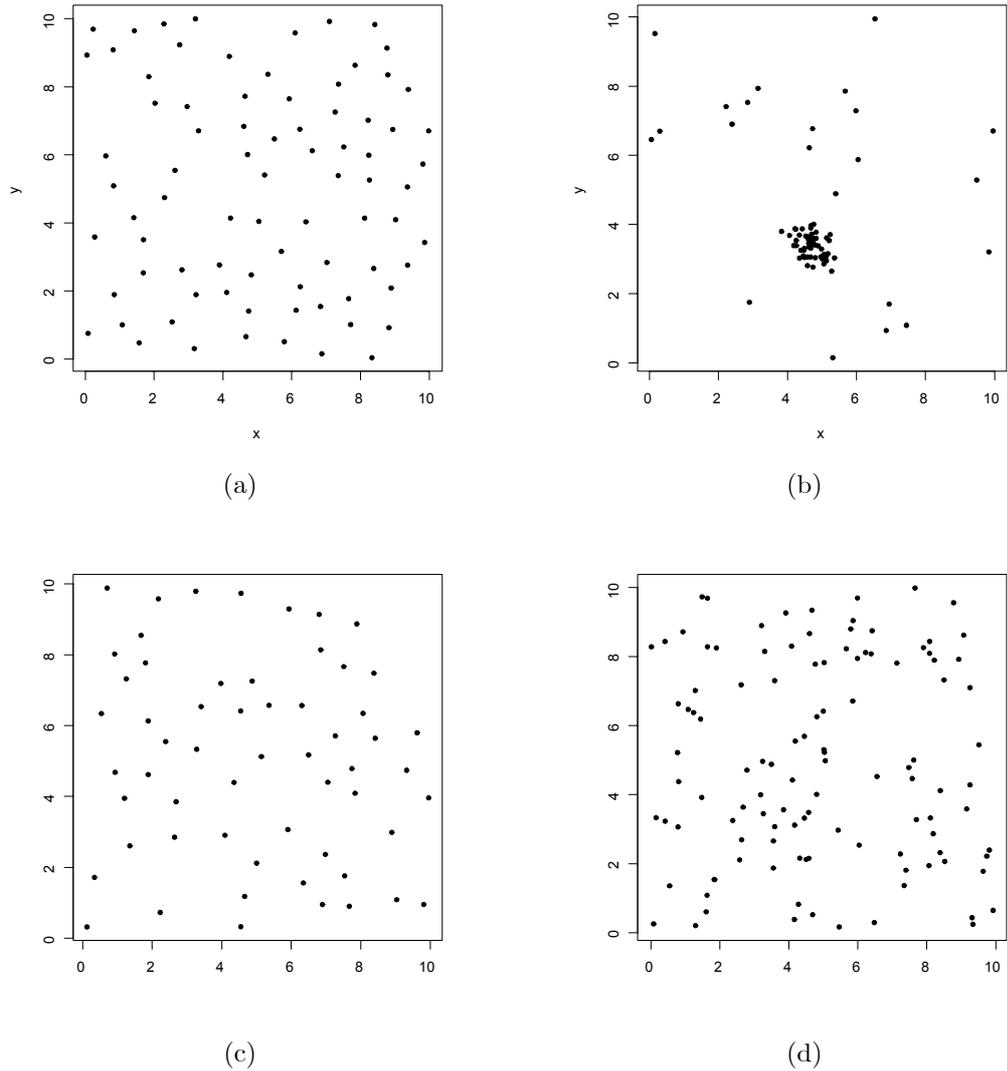


FIG. 4.6 – Réalisation d'un PP de Strauss sur $S = [0, 10]^2$, $r = 0.7$: conditionnel à $n = 80$, (a) $\gamma = 0$, (b) $\gamma = 1.5$; inconditionnel avec $\beta = 2$, (c) $\gamma = 0$ (noyau dur), (d) $\gamma = 0.8$.

Inconditionnellement, f_θ n'est admissible que si $\gamma \leq 1$.

Plusieurs généralisations du PP de Strauss sont possibles :

(i) en faisant intervenir les *triplets* de points de \mathbf{x} deux à deux distants de moins de r : si $t(\mathbf{x})$ compte ces triplets, $f_\theta(\mathbf{x}) = c(\theta)\beta^{n(\mathbf{x})}\gamma^{s(\mathbf{x})}\delta^{t(\mathbf{x})}$ est admissible ssi $\{\gamma \text{ et } \delta \leq 1\}$ ou si $\{\delta < 1\}$;

(ii) en saturant la statistique $m_{x_i}(\mathbf{x}) = \sum_{\xi \in \mathbf{x}: \xi \neq x_i} \mathbf{1}_{\|x_i - \xi\| \leq r}$ à une valeur $d < \infty$: si $g(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} \min\{d, m_{x_i}(\mathbf{x})\}$, $f_\theta(\mathbf{x}) = c(\theta)\beta^{n(\mathbf{x})}\gamma^{g(\mathbf{x})}$ est une densité toujours admissible (PP de saturation de Geyer);

(iii) en modélisant le *potentiel* de paires par une fonction en escalier $\phi(x_i, x_j) = \gamma_k$ si $d(x_i, x_j) \in (r_{k-1}, r_k]$, $k = 1, p$, 0 sinon, pour un choix prédéterminé de seuils $0 = r_0 < r_1 < r_2 < \dots < r_p$, $p \geq 1$. Si $s_k(\mathbf{x})$ est le nombre de couples de \mathbf{x} à distance dans $(r_{k-1}, r_k]$, $k = 1, \dots, p$,

$$f_\theta(\mathbf{x}; \beta, \gamma_1, \dots, \gamma_p) = c(\theta)\beta^{n(\mathbf{x})} \prod_{k=1}^p \gamma_k^{s_k(\mathbf{x})}$$

est admissible ssi $\gamma_1 \leq 1$. Cette densité appartient à la famille exponentielle associée à $U(\mathbf{x}) = (n(\mathbf{x}), s_1(\mathbf{x}), \dots, s_p(\mathbf{x}))$ et $\theta = (\theta_1 = \log \beta, \theta_2 = \log \gamma_1, \dots, \theta_{p+1} = \log \gamma_p)$.

(iv) Les modèles à *interaction d'aire* et les modèles à *interaction de connexité* (Baddeley et van Lieshout, [50]) ont pour densité $f_\theta(\mathbf{x}) = c\beta^{n(\mathbf{x})}\gamma^{h(\mathbf{x})}$ où h est définie de la façon suivante. Soit $r > 0$, $B(\mathbf{x}) = \{\cup_{x_i} B(x_i, r/2)\} \cap S$, où $B(x, R)$ est la boule de centre x de rayon R , $a(\mathbf{x})$ l'aire de $B(\mathbf{x})$ et $c(\mathbf{x})$ le nombre de ses composantes connexes (cf. Fig. 4.7). Le PP à *interaction d'aire* correspond à $h(\mathbf{x}) = a(\mathbf{x})$, celui à *interaction de connexité* à $h(\mathbf{x}) = c(\mathbf{x})$. S étant bornée et $r > 0$, a et c sont uniformément bornées et les deux densités sont admissibles sans contraintes

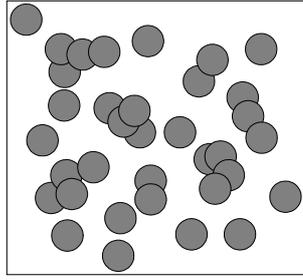


FIG. 4.7 – Une réalisation de l'ensemble $B(\mathbf{x})$ permettant de définir l'aire $a(\mathbf{x})$ et le nombre de composantes connexes $c(\mathbf{x})$ (ici $c(\mathbf{x}) = 23$).

sur les paramètres (β, γ) . Pour chacun des deux modèles, la répartition spatiale sera plus (resp. moins) régulière si $\gamma < 1$ (resp. $\gamma > 1$), $\gamma = 1$ correspondant à un PPP(β). Les calculs numériques de $a(\mathbf{x})$ et de $c(\mathbf{x})$ nécessiteront des techniques de discrétisation appropriées.

4.4 Caractéristiques du second ordre d'un PP

La notion de moment adaptée aux PP est celle de *mesure de moment d'ordre p*. Les moments d'ordre 1 et 2 sont à la base des procédures d'estimation.

4.4.1 Moment et intensité d'ordre p

Soit X un PP sur S bornée et (B_i) p boréliens bornés de S . Les produits $B_1 \times \dots \times B_p$ engendrant $\mathcal{B}(S^p)$, on définit la *mesure de moment d'ordre p* de X par,

$$\mu_p(B_1 \times \dots \times B_p) = E(N(B_1) \dots N(B_p)).$$

Moment et intensité d'ordre 1.

Pour $p = 1$, on obtient la *mesure d'intensité* de X :

$$\text{Mesure d'intensité : } \lambda(B) = E(N(B)) = \mu_1(B).$$

$\lambda(dx)$ est la probabilité qu'il y ait un point de X dans le volume infinitésimal dx autour de x . Si $\mu_1(dx) = \rho_1(x)dx$, ρ_1 est la *densité d'intensité* de X au premier ordre. Si X est stationnaire, λ étant invariante par translation, $\lambda(B) = \lambda\nu(B)$: λ est l'intensité de X , nombre moyen de points de X par volume unité.

Moment et intensité d'ordre 2.

Le *moment factoriel* d'ordre 2 de $B_1 \times B_2$ est :

$$\text{Moment factoriel d'ordre 2 : } \alpha_2(B_1 \times B_2) = \mu_2(B_1 \times B_2) - \lambda(B_1 \cap B_2).$$

Si $x \neq y$, $\alpha_2(dx \times dy)$ est la probabilité que X présente un point dans le volume dx et un point dans le volume dy . Si α_2 est à densité $\rho_2(x, y)$, ρ_2 est la *densité d'intensité d'ordre deux* de X . Si X est stationnaire (resp. isotropique), $\rho_2(x, y)$ ne dépendra que de $x - y$ (resp. que de $\|x - y\|$).

Covariance entre les comptages $N(A_1)$ et $N(A_2)$

μ_1 et α_2 permettent de calculer la covariance des variables de comptage $N(A_1)$ et $N(A_2)$:

$$\text{Cov}(N(A_1), N(A_2)) = \alpha_2(A_1 \times A_2) + \mu_1(A_1 \cap A_2) - \mu_1(A_1)\mu_1(A_2)$$

4.4.2 Mesure de Palm.

Soit x un point de S , X un PP sur S de loi P . La probabilité de Palm P_x de X au point x est la loi de X conditionnelle à la présence d'un point de X en x :

$$\forall F \in \mathcal{E} : P_x(F) = P(F \mid \{x \in X\}).$$

La mesure de Palm P_x permet de définir les statistiques conditionnelles à la présence d'un point de X en x [47] : par exemple la distance au plus proche voisin de $x \in X$, $d(x, X) = \inf\{y \in X \text{ et } y \neq x \mid x \in X\}$, le nombre de points de la configuration dans la boule $B(x, r)$ sachant que $x \in X$.

La difficulté pour définir P_x tient au fait que l'événement conditionnant $\{x \in X\}$ est de probabilité nulle : l'approche heuristique consiste à conditionner par l'événement $\{X \cap B(x, \varepsilon) \neq \emptyset\}$, de probabilité > 0 si $\varepsilon > 0$ et si la densité d'ordre un de X est > 0 , puis de voir si, pour F un événement "extérieur" à x , la probabilité conditionnelle :

$$P_{x,\varepsilon}(F) = \frac{P(F \cap \{X \cap B(x, \varepsilon) \neq \emptyset\})}{P(\{X \cap B(x, \varepsilon) \neq \emptyset\})}$$

admet une limite lorsque $\varepsilon \rightarrow 0$. Si la limite existe, c'est la mesure P_x . Par exemple, si X un PPP, les deux événements F et $\{X \cap B(x, \varepsilon) \neq \emptyset\}$ du numérateur sont indépendants pour ε petit et donc : $\forall x, P_{x,\varepsilon} = P$, la mesure de Palm P_x est constante en x : pour un PPP, conditionner ou non par la présence d'un point de X en x est sans conséquence sur la probabilité des événements F extérieurs à x .

4.4.3 Deux distances au plus proche voisin.

La distribution de la distance d'un point $x \in X$ au plus proche voisin de X est définie par :

$$G_x(r) = P_x(d(x, X \setminus \{x\}) \leq r), \quad r \geq 0. \quad (4.2)$$

Si X est stationnaire, $G_x = G$ est indépendante de x .

Cette distribution doit être distinguée de la distance d'un point **courant** $x \in \mathbb{R}^d$ (non nécessairement dans X) au plus proche voisin de X :

$$F_x(r) = P(d(x, X \setminus \{x\}) \leq r), \quad r \geq 0. \quad (4.3)$$

Un indicateur du caractère poissonnien de X est

$$J(r) = \frac{1 - G(r)}{1 - F(r)}.$$

$J > 1$, $J = 1$ et $J < 1$ indiquent respectivement plus, autant ou moins de régularité qu'un PPP.

Pour un PPP(λ) homogène sur \mathbb{R}^2 ,

$$G(r) = F(r) = 1 - \exp\{-\lambda\pi r^2\}$$

est d'espérance $(2\sqrt{\lambda})^{-1}$ et de variance $\lambda^{-1}(\pi^{-1} - 0.25)$.

4.4.4 Moment réduit d'ordre 2 de Ripley

Supposons que X soit isotropique sur \mathbb{R}^d d'intensité λ . Un bon indicateur du second ordre de la répartition des points de X est la fonction K définie par :

$$K(h) = \frac{1}{\lambda} E_x[N(B(x, h) \setminus \{x\})], \quad h \geq 0, \quad (4.4)$$

où E_x est l'espérance sachant que $x \in X$. On a deux interprétation de $K(h)$:

- $\lambda K(h)$ est le nombre moyen de points de X dans la boule $B(x, h)$, x non compté et sachant que $x \in X$;
- $\lambda^2 K(h)$ est le nombre moyen de paires de points $x_i \neq x_j$ de X à distance $\leq h$ où $x_i \in U$, un ensemble de surface unité.

Notant $\lambda_2(\|x - y\|) = \rho_2(x, y)$ la densité isotropique d'ordre 2 de X et $b_d = \pi^{d/2} \Gamma(1 + d/2)^{-1}$ le volume de la sphère unité de \mathbb{R}^d , on a :

$$K(h) = \frac{d \times b_d}{\lambda^2} \int_0^h u^{d-1} \rho_2(u) du. \quad (4.5)$$

Le calcul analytique de K est possible pour certains modèles :

1. *PPP homogène* sur \mathbb{R}^d : $K(h) = b_d \times h^d$ ($K(h) = \pi h^2$ si $d = 2$).
2. *PP à noyau dur* : de densité $c\beta^{n(x)} \mathbf{1}\{\forall i \neq j, \|x_i - x_j\| > r\}$ par rapport au PPP(λ),

$$K(h) = \frac{2\lambda\pi}{\exp(-\lambda\pi r^2)} \int_0^h u k(u) du,$$

où $k(u) = \exp(-\lambda V(u, r))$ si $u \geq r$, 0 sinon, $V(u, r)$ étant le volume de l'union de deux boules de rayon r dont les centres sont distants u .

3. *PP de Neyman-Scott* sur \mathbb{R}^2 : la position des parents suivant un PPP(λ), chaque parent engendrant un nombre aléatoire N de descendants de lois de dispersion autour d'un parent isotropique, la distance D au parent ayant pour fonction de répartition $\Phi(h) = P(D \leq h)$,

$$K(h) = \pi h^2 + \frac{E(N(N-1))\Phi(h)}{\lambda E(N)},$$

Un deuxième indicateur du caractère poissonnien.

La fonction

$$L(h) = \left(\frac{K(h)}{b_d} \right)^{1/d}$$

est un deuxième indicateur permettant d'évaluer le caractère poissonnien homogène d'une répartition spatiale sur \mathbb{R}^d : $L(h) = h$ si X est un PPP homogène ; une fonction L concave indiquera des répartitions avec agrégats ; une fonction L convexe indiquera des répartitions plus régulières que celles d'un PPP. L'estimation de L découlera naturellement de celle de K .

4.5 Processus ponctuel de Markov

Cette notion a été introduite par Ripley et Kelly [41]. Soit X un PP de densité f par rapport à un PPP(λ), où λ est une mesure à densité positive de masse finie sur les ensembles bornés. Soit $u \sim v$ une relation de voisinage symétrique sur S (i.e. $u \sim v$ si $\|u - v\| \leq r$). La frontière de voisinage de $A \subset S$ est $\partial A = \{v \in S : \exists u \in A \text{ t.q. } u \sim v\}$ et on note $\partial\{u\} = \partial u$ si $u \in S$.

Définition 4.3 X de densité f est markovien pour la relation \sim si, pour tout \mathbf{x} , on a :

1. $f(\mathbf{x}) > 0$ implique $f(\mathbf{y}) > 0$ si $\mathbf{y} \subset \mathbf{x}$: f est héréditaire.
2. si $f(\mathbf{x}) > 0$, alors $\lambda(u; \mathbf{x}) = f(\mathbf{x} \cup \{u\})/f(\mathbf{x})$ ne dépend que de u et de $\partial u \cap \mathbf{x}$.

La première condition traduit que si \mathbf{x} est réalisable, toute sous-configuration $\mathbf{y} \subset \mathbf{x}$ l'est aussi ? Ainsi, si $\mathbf{y} = \mathbf{x} \cup \{u\}$ est réalisable, $\lambda(u; \mathbf{x})$, l'intensité conditionnelle de Papangelou, est bien définie, densité de probabilité qu'il y ait un point en u sachant que \mathbf{x} est réalisée ailleurs. La deuxième dit que cette intensité conditionnelle ne dépend que de $\partial u \cap \mathbf{x}$, les points de \mathbf{x} voisins de u .

Des exemples de PP de Markov sont : (i) Le PPP d'intensité $\lambda(\cdot)$ dont l'intensité conditionnelle vaut $\lambda(u; \mathbf{x}) \equiv \lambda(u)$ pour toute relation de voisinage ; (ii) le PP de Strauss, avec $\lambda(u; \mathbf{x}) = \beta \exp\{\log \gamma \times \sum_i \mathbf{1}\{\|x_i - u\| \leq r\}\}$ pour la relation de r -voisinage : $u \sim v$ si $\|u - v\| \leq r$.

La propriété de Markov locale en u s'étend à tout borélien A de S : si X est markovien, la loi de $X \cap A$ conditionnelle à $X \cap A^c$ ne dépend de $X \cap \partial A \cap A^c$, la configuration de X sur $\partial A \cap A^c$. Comme pour un champ de Markov sur un réseau (cf 3.3.2), on dispose d'un théorème de Hammersley-Clifford caractérisant la densité d'un PP de Markov en terme de potentiels définis sur les cliques d'un graphe. Une clique pour \sim est une configuration $\mathbf{x} = (x_i)$ t.q. $\forall i \neq j, x_i \sim x_j$, avec la convention que les singletons sont aussi des cliques. Notons \mathcal{C} la famille des cliques de (S, \sim) .

Proposition 4.1 Un PP de densité f est markovien pour la relation \sim si et seulement si $\exists \Phi : E \rightarrow (\mathbb{R}^+)^*$ mesurable t.q.

$$f(\mathbf{x}) = \prod_{\mathbf{y} \subset \mathbf{x}, \mathbf{y} \in \mathcal{C}} \Phi(\mathbf{y}) = \exp \sum_{\mathbf{y} \subset \mathbf{x}, \mathbf{y} \in \mathcal{C}} \phi(\mathbf{y}).$$

$\phi = \log \Phi$ est le potentiel d'interaction de Gibbs : un exemple de densité de PP à interactions de paires est $f(\mathbf{x}) = \alpha \prod_i \beta(x_i) \prod_{x_i \sim x_j, i < j} \gamma(x_i, x_j)$. Le processus de Strauss correspond aux potentiels $\beta(x_i) = \beta$ et $\gamma(x_i, x_j) = \gamma \mathbf{1}\{\|x_i - x_j\| \leq r\}$.

Propriété de Markov pour un PP marqué.

La définition de la propriété de Markov et le théorème de Hammersley-Clifford restent inchangés si $Y = (X, M)$ est un PPM sur $S \times K$ pour \sim , une relation de voisinage symétrique sur $S \times K$. Si (X, M) est à marques indépendantes et si X est markovien pour une relation de voisinage \sim sur S , (X, M) est un PPM markovien pour la relation $(x, m) \sim (y, o) \iff x \sim y$ sur S . Un exemple de modèle isotrope à interactions de paires et à nombre fini de marques $M = \{1, 2, \dots, K\}$ est donné par la densité en $\mathbf{y} = \{(x_i, m_i)\}$:

$$f(\mathbf{y}) = \alpha \prod_i \beta_{m_i} \prod_{i < j} \gamma_{m_i, m_j}(\|x_i - x_j\|).$$

Si $\gamma_{kl}(d) \equiv 1$ pour $d > r_{k,l}$, où les réels $r_{k,l} > 0$, $k, l \in K$, $k \neq l$, sont fixés, les conditions (1) et (2) sont vérifiées pour la relation de voisinage : $(x, m) \sim (x', m') \iff \|x - x'\| \leq r_{m, m'}$: Y est un PPM markovien.

4.6 Statistique des processus ponctuels spatiaux

4.6.1 Estimation de l'intensité d'un PP

Intensité homogène

Soit X un PP observé sur une partie A compacte de \mathbb{R}^d , $\mathbf{x}_A = \{x_1, \dots, x_n\}$ sa réalisation à $n = n(A)$ points. Si X est stationnaire, $E(N(A)) = \lambda\nu(A)$ et l'intensité λ est estimée par

$$\hat{\lambda} = \frac{n(A)}{\nu(A)}. \quad (4.6)$$

$\hat{\lambda}$ estime sans biais λ . Si X est un PPP homogène, $\hat{\lambda}$ est l'estimateur du MV.

Un PP est *ergodique* si la probabilité des événements invariants par translation est soit 0, soit 1 ([47]; B est invariant si $\mathbf{x}_{-x} \in B$ pour tous $x \in S$ et $\mathbf{x} \in B$). Cette propriété entraîne que pour toute fonction $f : E \rightarrow \mathbb{R}$ mesurable et bornée, la moyenne empirique des translatés de la configuration \mathbf{x} ,

$$\frac{1}{nm} \sum_{j=1}^m \sum_{i=1}^n f(\mathbf{x}_{-(i,j)}) \rightarrow E(f(X)).$$

Un PPP homogène, le PP de Neyman-Scott dérivant d'un PPP homogène sont des exemples de PP ergodiques.

Si X est ergodique et si (A_n) est une suite croissante de convexes compacts t.q. $A_n \rightarrow \mathbb{R}^d$, $\hat{\lambda}_n = n(A_n)/\nu(A_n) \rightarrow \lambda$. En particulier, l'estimateur de l'intensité est consistant si X est PPP homogène.

Estimation d'une intensité inhomogène

Estimation non-paramétrique.

L'estimation non-paramétrique de $x \mapsto \lambda(x)$ est similaire à celle d'une densité de probabilité. Diggle [19] propose l'estimateur :

$$\hat{\lambda}_\sigma(x) = \frac{1}{K_\sigma(x)} \sum_{i=1}^n \frac{1}{\sigma^d} k\left(\frac{x - x_i}{\sigma}\right)$$

où $k : \mathbb{R}^d \rightarrow \mathbb{R}^+$ est un noyau symétrique d'intégrale égale à 1, $\sigma > 0$ est un paramètre de lissage et $K_\sigma(x) = \int_A \sigma^{-d} k\{(x - \xi)/\sigma\} \nu(d\xi)$ est une correction pour les effets du bord. Le choix de σ est important et il est conseillé d'essayer différentes valeurs. A l'inverse, le choix de $k(\cdot)$ l'est moins, un choix classique est le noyau gaussien $k(x) = (2\pi)^{-d/2} e^{-\|x\|^2/2}$ ou le noyau d'Epanechnikov $k(x) = c(d)\mathbf{1}(\|x\| < 1)(1 - \|x\|^2)$.

Exemple 4.1 Intensité de la localisation d'érables dans une forêt

La figure 4.8 donne la localisation de 514 érables d'une forêt du Michigan (donnée `lansing` du `package spatstat`). Ces données, recalées dans un carré unitaire, montrent une intensité non-constante. Cette intensité est estimée avec un noyau gaussien et pour 3 valeurs du paramètre de lissage, $\sigma = 0.01, 0.05$ et 0.1 .

Estimation d'un modèle paramétrique.

Si X est un PPP inhomogène d'intensité $\lambda(\cdot; \theta)$ observé sur A . L'estimateur du MV de θ maximise la log-densité sur A ,

$$\log f_\theta(\mathbf{x}) = \nu(A) - \int_A \lambda(\xi; \theta) d\xi + \sum_{x_i \in \mathbf{x}} \log \lambda(x_i; \theta).$$

Cette maximisation s'obtient en utilisant un logiciel dédié à l'estimation des modèles linéaires généralisés (MLG; cf. §4.6.5) dès que $\lambda(\cdot; \theta)$ suit un modèle log-linéaire.

Si X est un PP inhomogène d'intensité $\lambda(\cdot; \theta)$, on utilise encore cette fonctionnelle de "pseudo-vraisemblance marginale" pour estimer θ .

4.6.2 Estimation de caractéristiques au second ordre

Estimation non-paramétrique de G , F et J .

Si X est un PP stationnaire, et $G(\cdot)$ la distribution de la distance d'un point $x \in X$ au ppv, $G(h) = P_x(d(x, X \setminus \{x\}) \leq h)$, $r \geq 0$. Soit n le nombre des points de X dans $A_{\ominus h} = \{x \in A : B(x, h) \subseteq A\}$ où $B(x, h)$ est la boule de centre x et de rayon h ($A_{\ominus h}$ est le h -intérieur de A), et h_i la distance d'un point observé $x_i \in x$ à son ppv, $i = 1, n$. Un estimateur non paramétrique de G est la fonction de répartition empirique

$$\hat{G}(h) = \frac{1}{n} \sum_{i=1, n} \mathbf{1}_{(0, h_i]}(h)$$

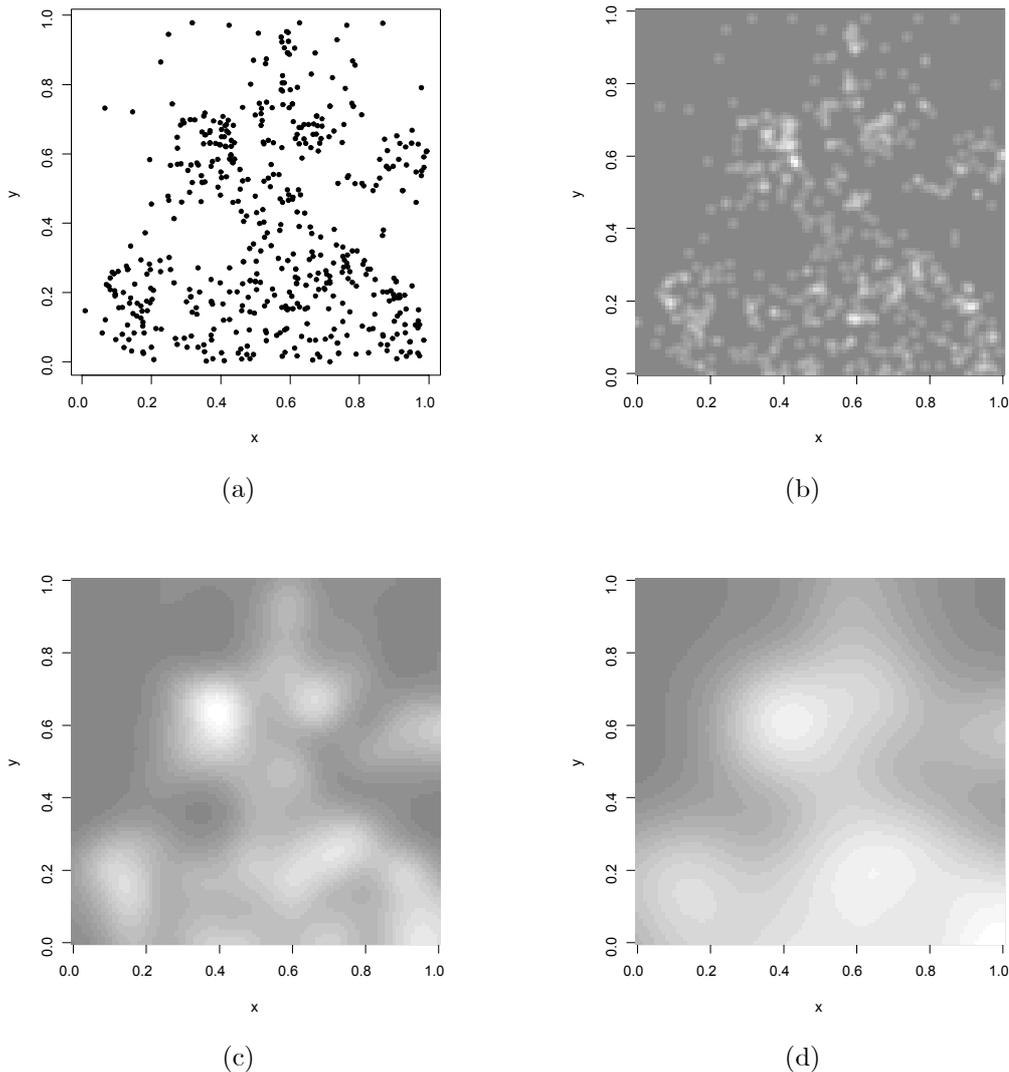


FIG. 4.8 – (a) localisations de 514 érables d’une forêt du Michigan. Estimations non-paramétriques de l’intensité avec : (b) $\sigma = 0.01$, (c) $\sigma = 0.05$ et (d) $\sigma = 0.1$.

Pour estimer $F(h) = P(d(x, X \setminus \{x\}) \leq h)$, $h \geq 0$, la distribution de la distance d’un point courant $x \in \mathbb{R}^d$ au ppv de X , on commence par choisir, indépendamment de X , une grille régulière sur $A_{\ominus h}$. Soit m le nombre des points de la grille et h_i^* la distance d’un point de la grille à son ppv dans X_A , $i = 1, \dots, m$. Un estimateur non-paramétrique de F est la fonction de répartition empirique

$$\hat{F}(h) = \frac{1}{m} \sum_{i=1, n} \mathbf{1}_{(0, h_i^*]}(h)$$

Un estimateur non-paramétrique de J est donc $\hat{J}(h) = \frac{1 - \hat{G}(h)}{1 - \hat{F}(h)}$ si $\hat{F}(h) < 1$. La figure 4.9 (première colonne) donne l’estimation de J pour les données **ants** de fourmières, **cells** de cellules et **finpines** de pins. L’hypothèse poissonnienne est acceptable pour la répartition des fourmières et rejetée pour les deux autres, avec plus de régularité pour les cellules et moins pour les pins.

Estimation de la fonction K de Ripley.

Commençons par proposer des estimateurs non-paramétriques de K . Si X est un PP homogène d’intensité λ , un estimateur naïf de $E = \lambda K$ est,

$$\hat{E}_0(h) = \frac{1}{n(A)} \sum_{i \neq j} \mathbf{1}_{\{\|x_i - x_j\| \leq h\}}.$$

\hat{E}_0 sous-estime K car un point au bord de A a moins de voisins dans A qu'un point intérieur. Pour un PP isotropique sur \mathbb{R}^2 , Ripley [43] propose la correction locale,

$$\hat{E}(h) = \frac{1}{n(A)} \sum_{i \neq j} w(x_i, x_j) \mathbf{1}_{\{\|x_i - x_j\| \leq h\}}$$

où $w(x_i, x_j)$ est l'inverse de la proportion du périmètre du cercle C_{ij} centré en x_i et passant par x_j qui se trouve dans A . Cet estimateur est non-biaisé si $h < h^*$, la plus grande distance t.q. un cercle centré en un point de A coupe A (par exemple, si $A = [0, 1]^2$, $h^* = \sqrt{2}$). Si $\hat{\lambda} = n(A)/\nu(A)$ est l'estimateur sans biais de l'intensité de X , on en déduit les estimations de K et de L :

$$\hat{K}(h) = \hat{\lambda}^{-1} \hat{E}(h) \text{ et } \hat{L}(h) = \pi^{-1/2} \hat{K}(h)^{1/2}$$

L'analyse des données **ants**, **cells** et **finpines** en utilisant \hat{L} (cf. Fig. 4.9, deuxième colonne) conduit aux mêmes conclusions que celles du paragraphe précédent.

4.6.3 Estimation par MC d'un modèle paramétrique

Une façon d'estimer un modèle paramétrique est d'utiliser une méthode de moindres carrés relative à une fonctionnelle de X . Par exemple, si X admet une densité f_θ qui permet le calcul analytique de K_θ , on peut estimer θ en minimisant, pour une puissance $c > 0$ et un choix de portées $\mathcal{H} = \{h_1, h_2, \dots, h_k\}$,

$$D^*(\theta) = \sum_{i=1, k} w_i \{\hat{K}(h_i)^c - K(h_i; \theta)^c\}^2 \quad (4.7)$$

pour un choix de poids w_i convenable. Diggle recommande de prendre $h_0 = 0.25$ si $A = [0, 1]^2$, $c = 0.5$ pour les PP réguliers et $c = 0.25$ pour les PP présentant des agrégats. Si la forme analytique de K_θ est inconnue, on en obtiendra une approximation de Monte Carlo $K_\theta^{MC} = \sum_{i=1}^m \hat{K}_{\theta_i}/m$ en simulant m réalisations indépendantes $x_A^{(i)}$ de X sous f_θ et en calculant \hat{K}_{θ_i} pour chaque i .

4.6.4 Test de Monte Carlo

Les lois des estimateurs précédents étant en général inconnues, on peut utiliser une procédure passe-partout par Bootstrap paramétrique pour tester une hypothèse ou valider un modèle. Décrivons en une.

Supposons que X soit un PPP homogène sur \mathbb{R}^2 d'intensité λ (hypothèse (H_0)) et qu'on veuille construire un intervalle de confiance pour $K(h)$ à h fixé. On commence par estimer λ par $\hat{\lambda}$, puis on simule m réalisations indépendantes $x_A^{(i)}$ de X sur A , un PPP($\hat{\lambda}$), et pour chaque $x_A^{(i)}$ on calcule $\hat{K}_i(h)$, $i = 1, \dots, m$. A partir de la distribution empirique des $\{\hat{K}_i(h), i = 1, \dots, m\}$, on approxime les quantiles de $\hat{K}(h)$. Pour tester (H_0) , il reste à les comparer à la statistique $\hat{K}_0(h)$ calculée à partir de l'observation x_A . Si le calcul de \hat{K} est coûteux, on choisira un m "petit", les enveloppes des \hat{K}_i , $\hat{K}_{inf}(h) = \min_i \hat{K}_i(h)$ et $\hat{K}_{sup}(h) = \max_i \hat{K}_i(h)$ définissant, sous (H_0) , l'intervalle de confiance

$$P(\hat{K}_0(h) < \hat{K}_{inf}(h)) = P(\hat{K}_0(h) > \hat{K}_{sup}(h)) \leq 1/(m+1)$$

avec égalité si les \hat{K}_i sont différents. On pourra aussi représenter les fonctions \hat{K}_{inf} , \hat{K}_0 , \hat{K}_{sup} et les comparer avec πh^2 : si \hat{K}_0 est comprise entre les deux enveloppes, on conclura raisonnablement que X est un PPP homogène.

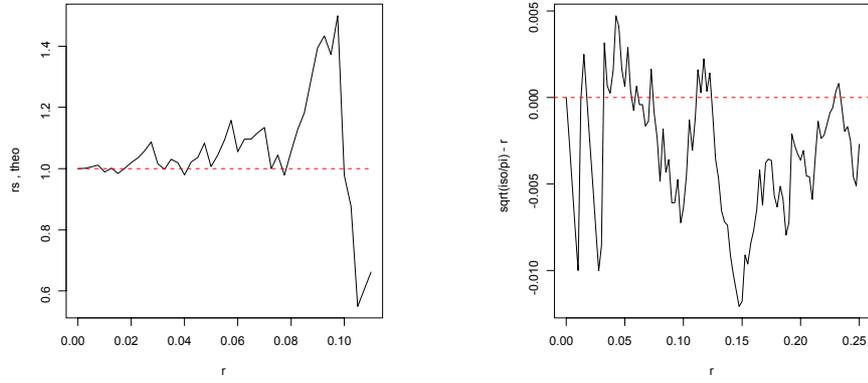
Ces procédures de Monte Carlo s'étendent facilement à la validation d'un modèle général.

Exemple 4.2 Données **finpines** : estimation de la fonction K_θ et validation de modèle

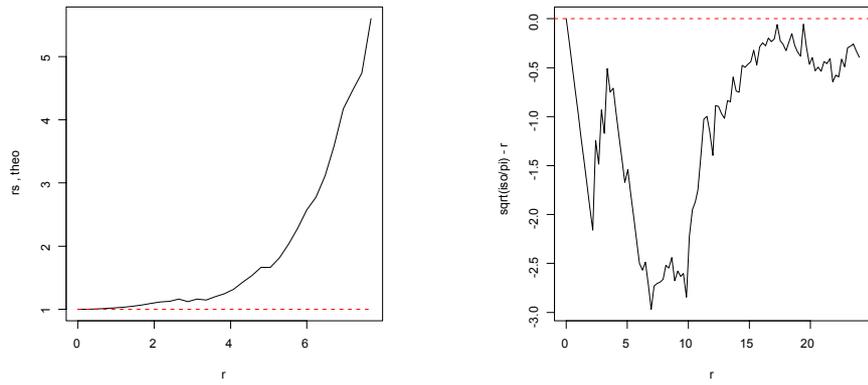
L'examen de la figure 4.9-c laisse penser que la répartition X présente des agrégats. On propose de modéliser X par un processus de Neyman-Scott dont la position des parents est un PPP(λ), le nombre de descendants une loi de Poisson de moyenne μ et la position d'un descendant autour d'un parent une loi $\mathcal{N}_2(0, \sigma^2 I_2)$. Les paramètres sont estimés en minimisant (4.7) avec $h_0 = 2.5$, $c = 0.25$, la somme (4.7) étant prise sur une grille de 180 points. Pour ce modèle,

$$K(h; \theta) = \pi h^2 + \theta_1^{-1} \{1 - \exp(-h^2/4\theta_2)\},$$

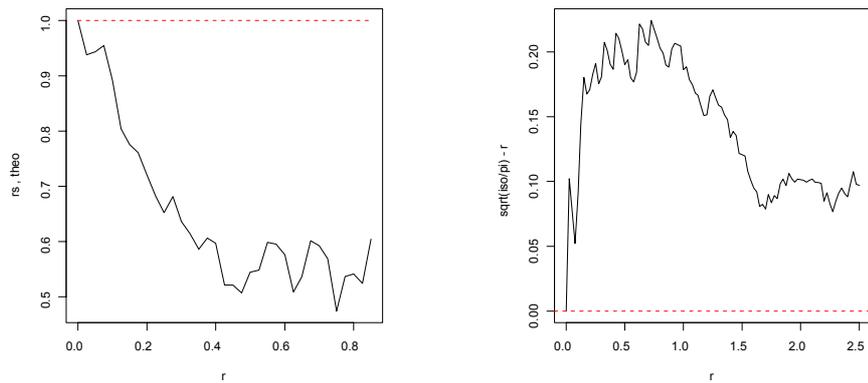
avec $\theta_1 = \mu$ et $\theta_2 = \sigma^2$; on obtient $\hat{\mu} = 1.875$, $\hat{\sigma}^2 = 0.00944$ et $\hat{\lambda} = n(\mathbf{x})/(\hat{\mu}|A|) = 0.672$. On valide le modèle en simulant 40 réalisations d'un PP de Neyman-Scott de paramètres $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\lambda}$: les enveloppes de confiance (cf. Fig. 4.10-b) montrent que ce choix de modèle est raisonnable, $\hat{K}_0(h)$ étant dans la bande de confiance pour $h > 0.2$. L'hypothèse d'un PPP homogène est fortement rejetée (cf. Fig. 4.10-a).



(a)



(b)



(c)

FIG. 4.9 – Deux résumés d'ordre 2 : à gauche $h \mapsto \hat{J}(h)$ et à droite $h \mapsto (\hat{L}(h) - h)$, pour trois répartitions spatiales (cf. Fig. 4.1-a-b-c). Si X est un PPP homogène, $L(h) - h \equiv 0$ et $J \equiv 1$.

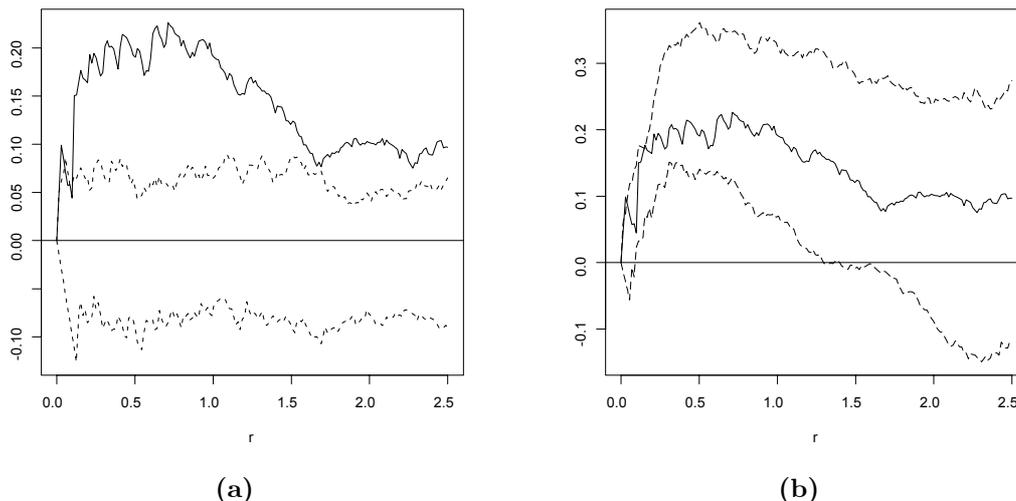


FIG. 4.10 – Estimation de $D(h) = L(h) - h$ pour les données **finpines** (cf. Fig. 4.1-c) : en ligne continue (a-b), l'estimation non-paramétrique de D ; (b) en pointillés, l'estimation paramétrique pour le modèle de N-S. Ces estimations sont ensuite comparées aux enveloppes inférieure et supérieure (en tirets) obtenues à partir de 40 simulations : (a) d'un PPP homogène de même intensité; (b) d'un processus de N-S aux paramètres estimés.

4.6.5 Pseudo-vraisemblance d'un PP

La notion de densité conditionnelle d'un PP en un site n'ayant pas de sens, il faut construire une nouvelle notion de pseudo-vraisemblance conditionnelle. Intuitivement, celle-ci peut se déduire de celle d'un champ sur un réseau de la façon suivante [43] :

1. pour une partition fine de S , on associe au PP le processus latticiel de comptage sur chaque élément de la partition ;
2. on définit alors la PVC de ce processus de comptage ;
3. on étudie la limite de cette PVC quand la partition tend vers 0.

La limite est identifiable si la densité f_θ de X est héréditaire et *stable* :

$$\exists c_\theta \text{ et } K_\theta > 0 \text{ t.q. : } \forall \mathbf{x} : f_\theta(\mathbf{x}) \leq c_\theta K_\theta^{n(\mathbf{x})}.$$

Jensen et Moller [31] montrent que si la suite des partitions devient de plus en plus fine, alors :

$$\lim_{i \rightarrow \infty} \prod_{j=1, m_i} f_\theta(\mathbf{x}_{A_{i,j}} | \mathbf{x}_{S \setminus A_{i,j}}) = \exp\{\lambda\nu(S) - \Lambda_\theta(S, \mathbf{x})\} \prod_{\xi \in \mathbf{x}} \lambda_\theta(\xi, \mathbf{x} \setminus \{\xi\}) \text{ où}$$

$$\lambda_\theta(\xi, \mathbf{x}) = \frac{f_\theta(\mathbf{x} \cup \xi)}{f_\theta(\mathbf{x})} \mathbf{1}\{f_\theta(\mathbf{x}) > 0\} \text{ et } \Lambda_\theta(A, \mathbf{x}) = \int_A \lambda_\theta(\xi, \mathbf{x}) d\xi$$

Définition 4.4 La PVC de X sur A est,

$$pl_A(\mathbf{x}, \theta) = \exp\{-\Lambda_\theta(A, \mathbf{x})\} \prod_{\xi \in \mathbf{x}} \lambda_\theta(\xi, \mathbf{x} \setminus \{\xi\}).$$

$\lambda_\theta(\xi, \mathbf{x})$, l'intensité conditionnelle, mesure la densité de probabilité d'avoir un point en ξ sachant que la configuration ailleurs est \mathbf{x} : la PVC est, au facteur multiplicatif $\exp\{-\Lambda_\theta(A, \mathbf{x})\}$ près, égale au produit sur les sites \mathbf{x} des intensités conditionnelles. Comme pour champ de Gibbs, la PVC évite le calcul de la constante de normalisation d'une densité jointe.

L'estimateur $\hat{\theta}_S^{pvc}$ de PVC sur S est une valeur qui maximise la PVC , $\theta \mapsto pl_S(\mathbf{x}, \theta)$ étant concave si f_θ appartient à une famille exponentielle, strictement si le modèle est identifiable.

Pour calculer $\Lambda_\theta(\mathbf{x}, S) = \int_S \lambda_\theta(\mathbf{x}, \xi) d\xi$, on approxime l'intégrale par,

$$\log pl_S(\mathbf{x}, \theta) \simeq \sum_{i=1, n(\mathbf{x})} \log \lambda_\theta(x_i, \mathbf{x} \setminus \{x_i\}) - \sum_{i=1, m} \lambda_\theta(u_j, \mathbf{x}) w_j \quad (4.8)$$

pour un choix $\{u_j, w_j \ j = 1, \dots, m\}$ convenable. Si l'ensemble des u_j contient \mathbf{x} , (4.8) se réécrit,

$$\log pl_S(\mathbf{x}, \theta) \simeq \sum_{j=1, m} (y_j \log \lambda_j^* - \lambda_j^*) w_j \quad (4.9)$$

où $\lambda_j^* = \lambda_\theta(u_j, \mathbf{x} \setminus \{u_j\})$ si $u_j \in \mathbf{x}$, $\lambda_j^* = \lambda_\theta(u_j, \mathbf{x})$ sinon, $y_j = \mathbb{1}[u_j \in \mathbf{x}]/w_j$. Le terme de droite de (4.9) est formellement équivalent à la log-vraisemblance, pondérée par des poids w_j , de variables de Poisson indépendantes et de moyenne λ_j^* , (4.9) étant maximisée en utilisant un logiciel d'estimation des modèles linéaires généralisés (cf. le package `spatstat`).

4.7 Calcul de Monte Carlo d'une vraisemblance

Les développements de ce paragraphe s'appliquent chaque fois que X est un champ de loi P_θ , $\theta \in \Theta \subset \mathbb{R}^d$ de densité π ,

$$\pi(x; \theta) = \frac{g(x; \theta)}{Z(\theta)} \text{ où } Z(\theta) = \int g(x; \theta) \mu(dx) < \infty.$$

la constante de normalisation $Z(\theta)$ étant analytiquement et numériquement incalculable. C'est le cas si X est un champ de Gibbs-Markov sur un réseau discret S ou encore si X est un PP de Markov sur une partie $S \subset \mathbb{R}^2$. Dans ce cas, l'estimation du MV est difficile comme le calcul de $Z(\theta)$. La méthode de Monte Carlo permet un calcul approché de $Z(\theta)$.

Soit θ une valeur courante du paramètre et $\psi \in \Theta$ une autre valeur fixée du paramètre. On estime alors le rapport :

$$\frac{Z(\theta)}{Z(\psi)} = E_\psi \left[\frac{g(X; \theta)}{g(X; \psi)} \right]$$

en utilisant la loi forte des grands nombres sous la loi $\pi(\cdot; \psi)$. Si on ne dispose pas d'une simulation exacte, on utilisera une méthode MCMC : si x est la réalisation de X , une approximation $l_N(\theta)$ de Monte Carlo (MC) du logarithme du rapport de vraisemblance $l(\theta) = \frac{f(x; \theta)}{f(x; \psi)} = \frac{Z(\psi)g(x; \theta)}{Z(\theta)g(x; \psi)}$ est obtenue en effectuant N simulations sous ψ ,

$$l_N(\theta) = \log \frac{g(x; \theta)}{g(x; \psi)} - \log \frac{1}{N} \sum_{j=1}^N \frac{g(X_j; \theta)}{g(X_j; \psi)}. \quad (4.10)$$

Cette approximation converge vers $l(\theta)$ si $N \rightarrow \infty$, l'échantillonnage par importance (importance sampling) disant que l'approximation sera d'autant meilleure que ψ est proche de θ . On obtient ainsi $l_N(\theta)$ et $\hat{\theta}_N$, l'approximation MC de $l(\theta)$ et de $\hat{\theta}$, l'estimateur du MV.

Pour une famille exponentielle, $\pi(x; \theta) = Z(\theta)^{-1} \exp\{t\theta v(x)\}$, (4.10) et sa dérivée s'écrivent :

$$l_N(\theta) = t(\theta - \psi)v(x) - \log \left[\sum_{j=1, N} \exp\{t(\theta - \psi)v(X_j)\} \right] - \log N.$$

$\hat{\theta}_N$, la valeur maximisant $l_N(\theta)$, dépend de l'observation x et des simulations X_1, \dots, X_N sous ψ . Pour N grand, si $\hat{\theta}$ est l'estimateur du MV exact, l'erreur de Monte Carlo $e_N = \sqrt{N}(\hat{\theta}_N - \hat{\theta})$ est approximativement normale [25]. Un pas de la méthode de Newton-Raphson s'écrit,

$$\theta_{k+1} = \theta_k - [l_N^{(2)}(\theta_k)]^{-1} \left\{ v(x) - \sum_{j=1, N} \frac{v(X_j)}{N} \right\},$$

et $Cov(\hat{\theta})$ est estimée par $-[l_N^{(2)}(\theta_k)]^{-1}$. L'approximation (4.10) étant d'autant meilleure que ψ est proche de θ , on initialisera la procédure itérative à $\hat{\theta}^{pvc}$, l'estimation du maximum de PVC de θ .

Exemple 4.3 Modélisation de la distribution spatiale d'une espèce végétale

On observe la présence ($X_s = 1$) ou l'absence ($X_s = 0$) en s de la grande laïche (une espèce végétale poussant en terrain marécageux) sur un réseau régulier de 24×24 parcelles (cf. Fig. 4.11). On considère deux modèles auto-logistiques

$$P(X_i = 1 | x_j, j \neq i) = \exp(\eta_i) / \{1 + \exp(\eta_i)\}$$

aux 4-ppv et aux 8-ppv, notant $v_1(x_{\partial i}) = \sum_{j: \|i-j\|=1} x_j$, $v_2(x_{\partial i}) = \sum_{j: \|i-j\|=\sqrt{2}} x_j$:

$$(I) \eta_i = \beta_0 + \beta_1 v_1(x_{\partial i}), \quad (II) \eta_i = \beta_0 + \beta_1 v_1(x_{\partial i}) + \beta_2 v_2(x_{\partial i}).$$

Le tableau 4.1 donne les estimations des deux modèles pour 3 méthodes : codage, maximum de PVC et MV. Pour le MV, on a utilisé l'algorithme de Newton - Raphson en deux pas initialisé en $\psi = \hat{\theta}_{MPV}$, la vraisemblance étant approchée en utilisant $N = 500$ simulations. Pour le codage et la PVC, les valeurs entre parenthèses sont les écarts-types obtenus en utilisant le modèle linéaire généralisé. Pour le codage, cet écart-type est le bon.

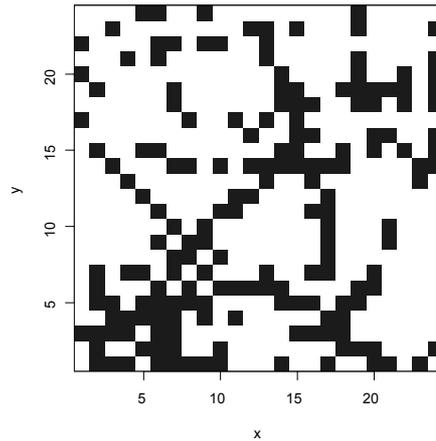


FIG. 4.11 – (a) Présence (■) ou absence (□) de grande laïche.

Modèles	β_0	β_1	β_2	Codage			
I	-1.486 (0.235)	0.497 (0.133)	-				
II	-2.093 (0.442)	0.477 (0.223)	0.391 (0.213)				
Pseudo - Vraisemblance							
I	-1.552 (0.172)	0.531 (0.098)	-				
II	-1.884 (0.206)	0.433 (0.102)	0.350 (0.106)				
Maximum de Vraisemblance							
I	-1.645 (0.196)	0.612 (0.129)	-				
II	-1.888 (0.229)	0.441 (0.137)	0.360 (0.152)				

TAB. 4.1 – Estimations (codage, PVC et MV) des modèles auto-logistiques aux 4 ou aux 8-ppv expliquant la répartition spatiale de la grande laïche.

Exemple 4.4 *Modèle expliquant la répartition de pins.*

Les données `swedishpines` du `package spatstat` représentent la répartition de 71 pins d'une forêt suédoise. Analysée par Ripley [42] pour le modèle de Strauss,

$$f_{\theta}(x) = c(\theta) \exp\{\theta_1 n(x) + \theta_2 s(x)\}, \text{ où } s(x) = \sum_{i < j} 1(\|x_i - x_j\| \leq r),$$

on complète l'étude en estimant le rayon d'interaction r par \hat{r} maximisant le profil $pl_A(r) = \max_{\theta} pl_A(x, \theta | r)$ de la pseudo-vraisemblance (cf. Fig. 4.12-a) : on trouve $\hat{r} = 7.5$. Le maximum de pseudo-vraisemblance est ensuite calculé en utilisant l'approximation (4.8). Pour le calcul de la vraisemblance, on a utilisé 100 simulations $X^{(i)}$, chacune obtenues par l'algorithme de Metropolis-Hastings arrêté après 10000 itérations (cf. §5.4.2). Finalement

Méthode	θ_1	θ_2
PVC	-3.937	-1.245
MV	-3.794 (0.247)	-1.266 (0.330)

TAB. 4.2 – Estimation du maximum de PVC et du MV du modèle de Strauss (pour $r = 7.5$) expliquant la localisations de 71 pins dans une forêt suédoise.

on a validé le modèle en simulant 40 réalisations de X sous $f_{\hat{\theta}_{MV}}$ (cf. Fig. 4.12-b).

4.8 Exercices

Exercice 4.1 *Simulation d'un point uniforme dans une boule.*

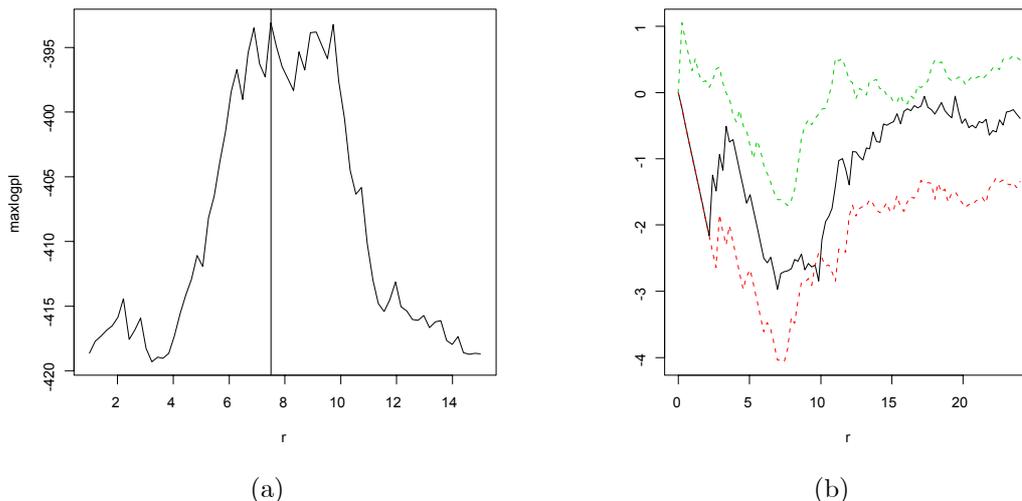


FIG. 4.12 – Estimation et validation du modèle de Strauss pour la répartition des 71 pins : (a) profil en r de la pseudo-vraisemblance ; (b) estimation non-paramétrique de $L(h) - h$ (ligne continue) comparée aux enveloppes inférieure et supérieure (en pointillés) obtenues à partir de 40 simulations d'un modèle de Strauss de paramètres $\hat{\theta}_1 = -3.794$, $\hat{\theta}_2 = -1.266$ et $r = 7.5$.

1. Soient U_1 et U_2 i.i.d. uniformes sur $[0, 1]$. Vérifier que le point de coordonnées polaires $(R, \Theta) = (\sqrt{U_1}, 2\pi U_2)$ est uniforme sur le disque $B_2(0, 1) \subset \mathbb{R}^2$. Quel est le gain de cette procédure “circulaire” par rapport à la méthode “cubique” consistant à ne retenir que les points (U_1, U_2) si $U_1^2 + U_2^2 \leq 1$?
2. Proposer une méthode de simulation “sphérique” d'un point uniforme dans la sphère $B_3(0, 1) \subset \mathbb{R}^3$. Quel est le gain de la méthode sphérique par rapport à la méthode cubique ?

Exercice 4.2 Vérifier que l'algorithme suivant simule une loi uniforme sur un borélien borné $W \subset \mathbb{R}^2$: inclure $W \subset R = \cup_{k=1}^K R_k$ dans une réunion finie de rectangles R_k d'intersections deux à deux de mesures nulles, puis :

1. choisir R_k avec la probabilité $\nu(R_k)/\nu(R)$.
2. simuler un point x uniforme sur R_k .
3. si $x \in W$, garder ce point. Sinon, retourner en 1.

Exercice 4.3 Soit X un processus de Cox sur S .

1. Montrer que X est surdispersé, c.a.d. : $\text{var}(N_X(B)) \geq E(N_X(B))$.
2. Calculer la loi de $N_X(B)$ pour $B \subset S$ si X est subordonné à $Z \equiv \xi$ où ξ suit une loi $\Gamma(\theta_1, \theta_2)$.

Exercice 4.4 Soit $X = \{X_i\}$ la réalisation d'un PPP(λ) homogène sur \mathbb{R}^1 , avec les conventions : $X_0 = \inf\{X_i, X_i \geq 0\}$ et $\forall i \in \mathbb{Z}, X_i \leq X_{i+1}$.

1. Montrer que les variables $X_i - X_{i-1}$ sont i.i.d. $\text{Exp}(\lambda)$.
2. Déterminer la loi des v.a. suivantes : (i) X_0 et X_1 ; (ii) $D_1 = \inf\{X_1 - X_0, X_0 - X_{-1}\}$, (iii) $D'_1 = \inf\{X_0, -X_{-1}\}$; (iv) $D_2 =$ distance au 2-ème ppv de X_0 .
3. Proposer des méthodes de simulation de X sur $[a, b]$.

Exercice 4.5 Soit X un PPP homogène sur \mathbb{R}^2 d'intensité λ .

1. Si $x_0 \in X$ et si $x_1 \in X$ est le point le plus proche de x_0 , déterminer la loi de $D_2 = \inf\{d(x_0, x), x \in X \setminus \{x_0, x_1\}\}$.
2. Même question si x_0 est un point courant de \mathbb{R}^2 .

Exercice 4.6 Fonction K du PP de Neyman-Scott.

Soit X un PP de Neyman-Scott sur \mathbb{R}^2 où la répartition des parents est un PPP(λ), le nombre de descendants d'un parent suit une loi de Poisson de moyenne μ et la position s d'un descendant autour d'un parent est $\mathcal{N}_2(0, \sigma^2 I_2)$. Montrer que $K(h) = \pi h^2 + \{1 - \exp(-h^2/4\sigma^2)\}/\lambda$.

Exercice 4.7 Processus de Bernoulli latticiel et PPP.

1. Si $S_\delta = (\delta\mathbb{Z})^d$ est le réseau d -dimensionnel de pas $\delta > 0$, le processus de Bernoulli $Y(\delta) = \{Y_i(\delta), i \in S_\delta\}$ de paramètre p est la donnée de variables i.i.d. de Bernoulli de paramètre p . Montrer que si $p = p_\delta = \lambda\delta^d$, le processus de Bernoulli converge vers un PPP(λ) si $\delta \rightarrow 0$.
2. Soit X un PPP(λ) homogène sur \mathbb{R}^d , et $Y_i = \mathbf{1}((X \cap [i, i+1]) \neq \emptyset)$, où pour $i \in \mathbb{Z}^d$, $[i, i+1[$ est le cube de côté 1 de base i . Montrer que Y est un processus de Bernoulli.
3. Soit X un PPP(λ) homogène sur \mathbb{R}^2 . On définit $Z = \{Z_i, i \in \mathbb{Z}^2\}$ où $Z_i = N(i + A)$ pour $A =]-1, 1]^2$. Calculer $E(Z_i)$ et $\text{cov}(Z_i, Z_j)$.

Exercice 4.8 Simuler les processus ponctuels suivants sur $S = [0, 1]^2$:

1. un PPP inhomogène d'intensité $\lambda(x, y) = 1 + 4xy$.
2. 20 points d'un processus à noyau dur pour $r = 0.05$.
3. le processus de Neyman-Scott où : (i) les parents x_i suivent un PPP(20); (ii) les lois du nombre de descendants et de dispersion autour des x_i sont i.i.d., respectivement $\text{Bin}(10, \frac{1}{2})$ et uniformes sur $B_2(x_i, 0.1)$.
4. un PPM $\{(x_i, m_i)\}$ où $X = (x_i)$ est un PPP(20), avec pour marque :
 - (a) une boule $B_2(x_i, r_i)$ de rayon $r_i \sim \mathcal{U}([0, 0.1])$;
 - (b) un segment centré en x_i , de longueur $\text{Exp}(10)$ et d'orientation uniforme.

Exercice 4.9 PVC pour un PP à noyau doux (soft core). Soit X un PP de densité

$$f_\omega(\mathbf{x}) = \alpha(\theta)\beta^{n(\mathbf{x})} \prod_{1 \leq i < j \leq n} \exp - \left(\frac{\sigma}{\|x_i - x_j\|} \right)^{2/\kappa}$$

sur $S = [0, 1]^2$, où $\theta = (\beta, \sigma)'$ avec $\beta > 0$, $0 \leq \sigma < \infty$, sont inconnues et $0 < \kappa < 1$ est connu, $\alpha(\theta)$ étant la constante de normalisation. Analyser l'influence des paramètres σ et κ sur la loi de la configuration spatiale. Écrire le modèle sous forme exponentielle et calculer sa pseudo-vraisemblance.

Exercice 4.10 MV pour le processus à noyau dur.

Soit X le PP à noyau dur sur $[0, 1]^2$ de densité $f_\theta(\mathbf{x}) = \alpha(\theta)h(\mathbf{x})\beta^{n(\mathbf{x})}$, où $h(\mathbf{x}) = \prod_{1 \leq i < j \leq n} \mathbf{1}\{\|x_i - x_j\| \geq \gamma\}$, $\theta = (\beta, \gamma)'$, $\beta > 0$, $\gamma > 0$. Montrer qu'on peut trouver une expression explicite pour l'EMV de γ mais que ce n'est pas possible pour β .

Exercice 4.11 Le jeu de données *lansing* du package *spatstat* donne la localisation spatiale de trois types de chêne : rouge, blanc et noir.

1. Estimer les caractéristiques au second ordre K et J pour les trois types de chênes.
2. Tester le caractère poissonnien des répartitions par une méthode Monte Carlo.

Exercice 4.12 Le jeu de données *spruce* du package *spatstat* donne la localisation et les diamètres des sapins d'une forêt saxonne. Nous examinerons seulement la variable de localisation.

1. Estimer le modèle de Strauss par la méthode du PMV en identifiant préliminairement le rayon d'interaction r .
2. Valider ce choix de modèle en simulant 40 réalisations du modèle estimé et en utilisant la fonction K
3. Un modèle avec un potentiel d'ordre deux modélisé par une fonction en escalier est-il préférable ?

Chapitre 5

Simulation des modèles spatiaux

La simulation d'une variable aléatoire est un outil numérique utile quand on ne dispose pas de solution analytique (ou numérique simple) à un problème donné : par exemple, le calcul de l'espérance d'une statistique réelle $g(X)$ de la v.a. X de loi π , $\pi(g) = \int_{\Omega} g(x)\pi(dx)$, est impraticable si X est à valeurs dans un espace de grande dimension $\Omega \subset \mathbb{R}^N$, ou encore si Ω est fini mais avec un très grand nombre de points. La méthode de Monte Carlo consiste alors à estimer $\pi(g)$ par

$$\bar{g}_n = \frac{1}{n} \sum_{i=1, n} g(X_i)$$

le long d'un n -échantillon $\{X_i, i = 1, \dots, n\}$ de la loi π : la loi forte des grands nombres dit alors que $\bar{g}_n \rightarrow \pi(g)$ si $n \rightarrow \infty$ et si $g(X)$ est π -intégrable ; de plus, si $Var_{\pi}\{g(X)\} < \infty$, le théorème central limite standard dit que cette convergence a lieu à la vitesse \sqrt{n} .

Si π peut être simulée par une méthode classique (inversion de la fonction de répartition, méthode d'acceptation-rejet, etc...), on dispose alors d'une méthode numérique simple permettant un calcul approché de l'espérance. Mais dès que la loi π est "compliquée", par exemple une loi de Gibbs, les méthodes de simulation classiques sont inopérantes. Des algorithmes par *dynamique d'une chaîne de Markov* $(X_n) \rightarrow \pi$ prennent le relais et permettent la simulation approchée d'une large classe de lois : ce sont les *méthodes de Monte Carlo par Chaîne de Markov*, ou MCMC (*Monte-Carlo Markov Chain*). Nous présenterons les deux algorithmes principaux :

1. l'échantillonneur de Gibbs, bien adapté aux champs de Gibbs sur réseau ;
2. l'algorithme de Metropolis.

Commençons par donner quelques rappels sur les chaînes de Markov, leurs propriétés de convergence et ce que doit vérifier un algorithme MCMC afin de permettre la simulation d'une loi π ([44];[56]).

5.1 Chaînes de Markov

Soit π une loi absolument continue par rapport à une mesure μ sur (Ω, \mathcal{E}) , la densité étant elle aussi notée π . Pour simuler π , on va construire une chaîne de Markov $X = (X_k, k \geq 0) \rightarrow \pi$ sur Ω de transition P : ainsi, pour k grand, X_k sera une simulation approchée de π .

La *transition* P d'une chaîne de Markov sur (Ω, \mathcal{E}) est une application $P : (\Omega, \mathcal{E}) \rightarrow [0, 1]$ telle que :

- (i) $\forall A \in \mathcal{E}$, $P(\cdot, A)$ est mesurable et
- (ii) $\forall x \in \Omega$, $P(x, \cdot)$ est une probabilité sur (Ω, \mathcal{E}) .

Si $X_0 \sim \nu_0$, la loi de (X_0, X_1, \dots, X_n) est

$$P_n(dx_0, dx_1, \dots, dx_n) = \nu(dx_0) \prod_{i=1, n} P(x_{i-1}, dx_i),$$

la loi ν_k de X_k étant $\nu_k(dx_k) = \int_{\Omega} \nu_0(dx) P^k(x, dx_k)$ où $P^k(x, \cdot)$ est la transition en k pas de la chaîne. En particulier, $(\nu P)(dy) = \int_{\Omega} \pi(dx) P(x, dy)$ est la loi de X_1 si ν est celle de X_0 .

Si Ω est fini à m états, une loi ν sur Ω est représentée par un vecteur ligne de \mathbb{R}^m , la transition P par la matrice $m \times m$ des probabilités,

$$P_{i,j} = \Pr(X_{n+1} = j \mid X_n = i).$$

La transition en k -pas est la puissance $k^{\text{ième}}$ de P , la loi de X_k étant νP^k . La convergence d'une chaîne de Markov est liée aux notions suivantes :

1. P est π -irréductible si, $\forall x \in \Omega$ et $A \in \mathcal{E}$ t.q. $\pi(A) > 0$, $\exists k = k(x, A)$ t.q. $P^k(x, A) > 0$.

2. Si Ω est fini, P est *primitive* si $\exists k > 0$ t.q., $\forall i, j, P^k(i, j) > 0$.
3. π est *invariante* pour P si $\pi P = \pi$.
4. P est *périodique* si $\exists d \geq 2$ et une partition $\{\Omega_1, \Omega_2, \dots, \Omega_d\}$ de Ω telle que $\forall x \in \Omega_i, \forall i, P(x, \Omega_{i+1}) = 1$, avec la convention $d + 1 \equiv 1$. Sinon, la chaîne est *apériodique*.

Si Ω est discret et si π est positive (noté $\pi > 0 : \forall i \in \Omega, \pi(\{i\}) > 0$), l'irréductibilité est identique à la communication entre tous les états : $\forall i, j \in \Omega, \exists k = k(i, j)$ t.q. $P^k(i, j) > 0$.

Si π est P invariante, π est une loi stationnaire puisque si $X_0 \sim \pi, X_k \sim \pi$ pour tout k et (X_k) est stationnaire.

Il est facile de vérifier que si (X_k) est une chaîne de transition P telle que, $\forall x \in \Omega$ et $\forall A \in \mathcal{E}, P^n(x, A) \rightarrow \pi(A)$, alors π est P -invariante, P est π irréductible et apériodique. La propriété remarquable à la base de la construction des méthodes MCMC est la réciproque de cette propriété. Définissons la *distance en variation totale* entre deux probabilités ν_1 et ν_2 :

$$\begin{aligned} \| \nu_1 - \nu_2 \|_{VT} &= \sup_{A \in \mathcal{E}} |\nu_1(A) - \nu_2(A)|, \\ \| \nu_1 - \nu_2 \|_{VT} &= \frac{1}{2} \sum_{i \in \Omega} |\nu_1(i) - \nu_2(i)| \text{ si } \Omega \text{ discret.} \end{aligned}$$

On a la propriété fondamentale de convergence des chaînes de Markov [48] :

Théorème 5.1 (i) Soit P une transition apériodique π -irréductible et π -invariante. Alors, π -p.s. en $x \in \Omega$, $\|P^k(x, \cdot) - \pi\|_{VT} \rightarrow 0$ si $k \rightarrow \infty$.

(ii) Si de plus, $\forall x \in \Omega, P(x, \cdot)$ est absolument continue par rapport à π , la convergence a lieu pour tout $x \in \Omega$.

(iii) Le cas Ω fini : si $\varepsilon = \inf_{i,j} P^k(i, j) > 0$ pour un $k > 0$, si $m = \#\Omega$ et si $[x]$ est la partie entière de x :

$$\|\nu P^n - \pi\|_{VT} \leq (1 - m\varepsilon)^{[n/k]} \quad (5.1)$$

5.1.1 Construction d'un algorithme MCMC

Pour construire un algorithme MCMC de simulation de π , il faut proposer une transition P (facile à simuler) qui soit π -invariante, π -irréductible et apériodique. La π -irréductibilité et l'apériodicité sont à vérifier au cas par cas. Une façon simple d'assurer la π -invariance est de proposer une transition qui soit π -réversible :

Définition 5.1 P est π -réversible si

$$\forall A, B \in \mathcal{E}, \int_A P(x, B) \pi(dx) = \int_B P(x, A) \pi(dx) \quad (5.2)$$

Si π et P sont à densité, la réversibilité s'écrit :

$$\begin{aligned} \forall x, y \in \Omega : \pi(x)p(x, y) &= \pi(y)p(y, x), \\ \forall i, j \in \Omega : \pi_i p_{ij} &= \pi_j p_{ji} \text{ si } \Omega \text{ est discret} \end{aligned}$$

Ainsi, si X est π -réversible, $(X_n, X_{n+1}) \sim (X_{n+1}, X_n)$ si $X_n \sim \pi$: la loi de la chaîne est invariante par retournement du temps.

D'autre part, si P est π -réversible, il est facile de voir que π est P -invariante.

5.2 Les deux algorithmes MCMC de base

Soit π une loi de densité $\pi(x)$ par rapport à μ . Si Ω est discret, on prendra pour μ la mesure de comptage. On supposera toujours vérifiée la condition de positivité : $\forall x \in \Omega, \pi(x) > 0$. Pour simuler π , on construit une chaîne π -irréductible et apériodique de transition P π -invariante.

5.2.1 Echantillonneur de Gibbs sur un espace produit

Soit $X = (X_i, i \in S), S = \{1, 2, \dots, n\}$, une variable de loi π à états dans $\Omega = \prod_{i \in S} E_i$. Pour tout $x = (x_1, x_2, \dots, x_n) \in \Omega$, on suppose que les lois conditionnelles $\pi_i(\cdot | x^i)$, où $x^i = (x_j, j \neq i)$ sont connues. La transition

$$P_i(x, y) = \pi_i(y_i | x^i) \mathbf{1}(x^i = y^i) \quad (5.3)$$

n'autorise de changement qu'au site i , à savoir $x_i \mapsto y_i$ suivant la probabilité $\pi_i(\cdot | x^i)$, x étant inchangée ailleurs.

Echantillonneur de Gibbs avec balayage systématique. On visite S séquentiellement dans l'ordre $1 \rightarrow 2 \rightarrow \dots \rightarrow (n-1) \rightarrow n$ (un balayage de S). A chaque pas, la valeur x_i en i est relaxée (tirée au hasard) suivant la loi π_i conditionnelle à l'état courant. La (densité de) transition de l'état $x = (x_1, x_2, \dots, x_n)$ à l'état $y = (y_1, y_2, \dots, y_n)$ après ce balayage de S est donnée par :

$$P_S(x, y) = \prod_{i=1}^n \pi_i(y_i \mid y_1, \dots, y_{i-1}, x_{i+1}, x_{i+2}, \dots, x_n).$$

Au i -ème pas du balayage, π_i est conditionnée par les $(i-1)$ valeurs y déjà relaxées et par les $(n-i)$ valeurs de x qui n'ont pas encore été relaxées.

Echantillonneur par balayage aléatoire. Soit $p = (p_1, p_2, \dots, p_n)$ une probabilité > 0 sur S : pour tout i , $p_i > 0$. A chaque pas, un site i est choisi au hasard suivant la loi p et la valeur en ce site est relaxée suivant la loi π_i conditionnelle à l'état courant. La transition pour un changement est,

$$P_A(x, y) = \sum_{i=1}^n p_i \pi_i(y_i \mid x^i) \mathbf{1}(x^i = y^i)$$

L'échantillonneur de Gibbs réalise la simulation de π [23] :

Théorème 5.2 *Supposons que $\pi > 0$. Alors, P_S et P_A sont π -irréductibles, apériodiques et de loi invariante π : $\forall \nu, \nu P^k \rightarrow \pi$.*

Construire P n'exige de connaître π qu'à un facteur multiplicatif près : en effet, si $\pi(x) = ce(x)$, les lois conditionnelles π_i ne dépendent pas de c . Cette remarque est importante car pour les champs de Gibbs, c n'est pas connu.

5.2.2 L'algorithme de Metropolis-Hastings (MH)

Proposé par Metropolis en 1953, l'algorithme a pris sa forme générale en 1970 avec Hastings [44]. Ici, l'espace d'état Ω n'est pas supposé être un produit.

Description de l'algorithme de MH.

On construit une transition P qui est π -réversible en 2 étapes :

- (i) on propose un changement $x \mapsto y$ selon une transition $Q(x, \cdot)$;
- (ii) on accepte ce changement avec la probabilité $a(x, y)$ où $a : \Omega \times \Omega \rightarrow]0, 1]$.

Notant $q(x, y)$ la densité de $Q(x, \cdot)$, la densité de $P(x, \cdot)$ s'écrit :

$$P(x, y) = a(x, y)q(x, y) + \mathbf{1}(x = y)[1 - \int_E a(x, z)q(x, z)dz] \quad (5.4)$$

Quant à la condition de π -réversibilité de P , elle s'écrit :

$$\forall x, y \in \Omega : \pi(x)q(x, y)a(x, y) = \pi(y)q(y, x)a(y, x) \quad (5.5)$$

Sous cette condition, q satisfait la condition de symétrie faible : $q(x, y) > 0 \Leftrightarrow q(y, x) > 0$. Dans ce cas, on définit le ratio de MH par,

$$r(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}.$$

Si q est symétrique, $r(x, y) = \pi(y)/\pi(x)$.

Pour que P soit irréductible, Q doit l'être, mais ce n'est pas suffisant : l'irréductibilité comme l'apériodicité de P devront être examinées au cas par cas.

Si Ω est fini, l'apériodicité est assurée sous l'une des conditions suivantes : (a) $\exists x_0$ t.q. $Q(x_0, x_0) > 0$; (b) $\exists(x_0, y_0)$ t.q. $r(x_0, y_0) < 1$; (c) Q est symétrique et π n'est pas la loi uniforme.

L'algorithme de Metropolis. La *dynamique de Metropolis* correspond au choix :

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right\}$$

Si q est symétrique, $a(x, y) = \min\{1, \pi(x)/\pi(y)\}$; l'algorithme de Metropolis s'écrit :

1. Soit x l'état initial : choisir y suivant $Q(x, \cdot)$
2. Si $\pi(y) \geq \pi(x)$, garder y . Revenir en 1.
3. Si $\pi(y) < \pi(x)$, tirer une loi uniforme U sur $[0, 1]$:
 - (a) si $U \leq p = \pi(y)/\pi(x)$, garder y .
 - (b) si $U > p$, garder la valeur initiale x .
4. Revenir en 1.

5.3 Simulation d'un champ de Markov sur un réseau

5.3.1 Les deux algorithmes de base

Soit X un champ de Markov sur $S = \{1, 2, \dots, n\}$, à états dans $\Omega = \prod_{i \in S} E_i$ (i.e. $\Omega = E^S$), de densité

$$\pi(x) = Z^{-1} \exp U(x), \text{ où } U(x) = \sum_{A \in \mathcal{C}} \Phi_A(x) \quad (5.6)$$

où \mathcal{C} est la famille des cliques du graphe de Markov. On peut simuler π :

1. soit par l'échantillonneur de Gibbs de lois conditionnelles

$$\pi_i(x_i | x^i) = Z_i^{-1}(x_{\partial i}) \exp\left\{ \sum_{A: A \ni i} \Phi_A(x) \right\};$$

2. soit par dynamique de Metropolis : si q est symétrique et $a^+ = \max\{a, 0\}$:

$$p(x, y) = q(x, y) \exp -\{U(x) - U(y)\}^+ \text{ pour } x \neq y.$$

5.3.2 Exemples

Modèle d'Ising isotropique aux 4-ppv.

$S = \{1, 2, \dots, n\}^2$, $E = \{-1, +1\}$; la loi jointe est $\pi(x) = Z^{-1} \exp\{\alpha \sum_i x_i + \beta \sum_{\langle i, j \rangle} x_i x_j\}$ et les lois conditionnelles, pour $v_i = \sum_{j \in \partial i} x_j$,

$$\pi_i(x_i | x^i) = \frac{\exp x_i(\alpha + \beta v_i)}{2 \operatorname{ch}(\alpha + \beta v_i)} \quad (5.7)$$

Dynamique de Metropolis par échange de spins.

On propose le changement $x \mapsto y$ suivant : choisir au hasard deux sites $i \neq j$, prendre y égale à x partout sauf en $\{i, j\}$ où les valeurs x_i et x_j sont permutées (on échange les spins). Pour un tel échange, $Q(x, y) = 2/[n^2(n^2 - 1)]$ et $Q(x, y) = 0$ sinon. Le calcul de $\Delta U(x, y) = U(y) - U(x)$ est local si X est markovien. Par exemple, pour le modèle (5.7),

$$\Delta U(x, y) = \begin{cases} \beta(x_j - x_i)(v_i - v_j) & \text{si } \|i - j\|_1 > 1 \\ \beta(x_j - x_i)(v_i - v_j) - \beta(x_j - x_i)^2 & \text{sinon} \end{cases}$$

Un pas de l'algorithme de Metropolis par échange de spins requiert :

1. le tirage d'une loi uniforme sur $\{1, 2, \dots, n\}^2$;
2. le calcul de $\Delta U(x, y)$: si $\Delta U(x, y) > 0$, on accepte y .
3. Sinon, un tirage $U \sim \mathcal{U}([0, 1])$, indépendant de 1. :
 - (a) si $U < \exp \Delta U(x, y)$, on permute les spins ;
 - (b) sinon, on reste en x .

Si $x(0)$ est la configuration initiale, l'algorithme évolue dans le sous-espace $\Omega_{x(0)}$ des configurations ayant le même nombre de spins $+1$ et -1 que $x(0)$. Q est irréductible sur Ω_0 puisque deux configurations de Ω_0 se correspondent par permutation et que toute permutation est un produit fini de transpositions. Si $x(0)$ n'est pas constante et si $\beta \neq 0$, π n'est pas constante et la chaîne est apériodique : l'algorithme est ergodique sur Ω_0 fournissant une simulation de la loi π restreinte à Ω_0 .

Simulation des auto-modèles.

Pour un auto-modèle, le produit des lois conditionnelles π_i étant absolument continu par rapport à π , l'échantillonneur de Gibbs converge vers π (cf. Fig. 5.1).

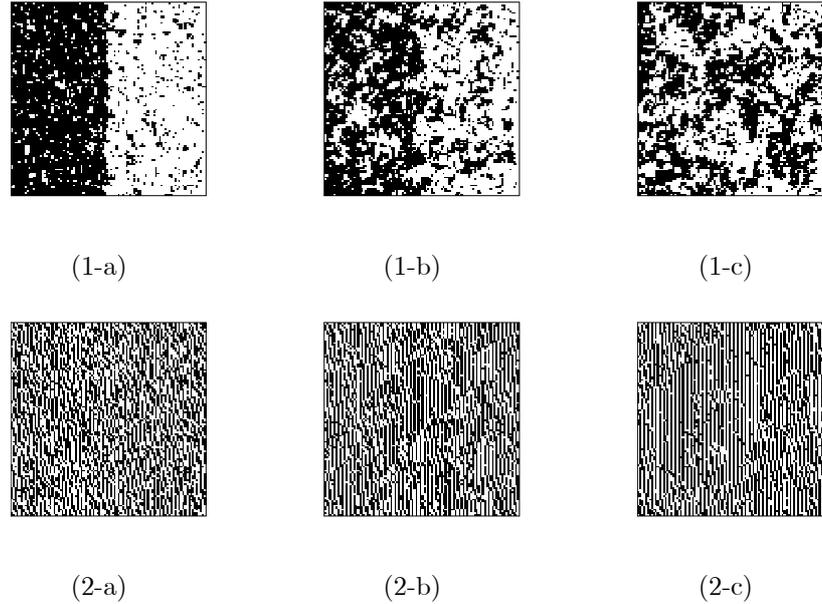


FIG. 5.1 – Simulation par échantillonneur de Gibbs de deux textures binaires auto-logistiques sur le réseau $\{1, 2, \dots, 100\}^2$. La configuration initiale est 1 sur la moitié gauche, 0 sur la moitié droite. Simulations : (a) après un balayage, (b) après 10 balayages, (c) après 1000 balayages. On a considéré : (1) le modèle isotropique aux 4-ppv avec $\alpha = -3$, $\beta = 1.5$; (2) le modèle aux 8-ppv avec $\alpha = -0.26$, $\beta_{0,1} = 2.1$, $\beta_{1,0} = -2.0$ et $\beta_{1,-1} = 0.13$. Les configurations 0 et 1 sont équiprobables pour chacun des deux modèles.

Auto modèle gaussien.

Soit $X \sim \mathcal{N}_n(m, \Sigma)$ et $Q = \Sigma^{-1}$. Pour le balayage $1 \mapsto 2 \mapsto \dots \mapsto n$, la i -ième relaxation conditionnelle à $z^i = (y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)$ se fait suivant la loi

$$\pi_i(x_i | z^i) \sim \mathcal{N}_1(-q_{ii}^{-1} \sum_{j:j \neq i} q_{ij} z_j, q_{ii}^{-1})$$

Si le champ est markovien, $\pi_i(x_i | x^i) = \pi_i(x_i | x_{\partial i})$.

Comparons cette procédure à celle utilisant la décomposition de Cholesky $\Sigma = T^t T$ de Σ : pour $S = \{1, 2, \dots, 100\}^2$, Σ est de dimension $10^4 \times 10^4$:

- si ε est un 10^4 -échantillon gaussien réduit, $X = T\varepsilon \sim \mathcal{N}_{10^4}(0, \Sigma)$. Le calcul de T est coûteux mais il n'est pas à répéter pour obtenir une autre réalisation de X et la simulation est exacte.
- l'échantillonneur de Gibbs exige environ 100 balayages, soit 100×10^4 simulations gaussiennes, sans rechercher la forme de Cholesky de Σ . Par contre, il faudra recommencer toutes les simulations pour obtenir une autre réalisation de X et la simulation sera approchée.

Espace d'états produit.

L'échantillonneur de Gibbs est bien adapté pour simuler des modèles à espace d'états $E = \Lambda \times \mathbb{R}^n$ produit, comme par exemple en télédétection où $\lambda \in \Lambda = \{1, 2, \dots, r\}$ est un label qualitatif (terre cultivée, forêt, eau, etc...) et $x \in \mathbb{R}^n$ une réponse multispectrale quantitative.

5.4 Simulation d'un processus ponctuel

Soit X un PP défini sur un borélien borné S de \mathbb{R}^d , $E = \sum_{n \geq 0} E_n$ l'espace exponentiel des configurations de X , E_n étant l'espace des configurations à n points. Pour simuler X , on va utiliser une dynamique de MH [36] dont l'implémentation est décrite dans l'aide en ligne du package `spatstat`.

Pour simplifier, on supposera que X admet une densité de la forme $f(\mathbf{x}) = c\beta^{n(\mathbf{x})}\gamma^{h(\mathbf{x})}$. Examinons dans un premier temps la simulation conditionnelle à un nombre fixé de points $n(\mathbf{x})$.

5.4.1 Simulation conditionnelle

Si $n(\mathbf{x}) = n$ est fixé, la densité en $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ est proportionnelle à $\gamma^{h(\mathbf{x})}$. Notant $(\mathbf{x} \cup \xi) \setminus \eta$ pour $(\mathbf{x} \cup \{\xi\}) \setminus \{\eta\}$, l'algorithme de MH propose :

1. de changer $\mathbf{x} \rightarrow \mathbf{y} = (\mathbf{x} \cup \xi) \setminus \eta$, $\eta \in \mathbf{x}$ et $\xi \in S$, avec la densité $q(\mathbf{x}, \mathbf{y}) = q(\mathbf{x}, \eta, \xi)$,
2. d'accepter \mathbf{y} avec la probabilité $a(\mathbf{x}, \mathbf{y}) = a(\mathbf{x}, \eta, \xi)$.

La probabilité d'acceptation $a(\mathbf{x}, \mathbf{y})$ assurant la π -réversibilité du noyau est,

$$a(\mathbf{x}, \mathbf{y}) = \min\{1, r(\mathbf{x}, \mathbf{y})\}, \text{ où } r(\mathbf{x}, \mathbf{y}) = \frac{f(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{f(\mathbf{x})q(\mathbf{x}, \mathbf{y})}.$$

Si le changement $\mathbf{x} \mapsto \mathbf{y}$ est obtenu en supprimant un point uniformément dans \mathbf{x} puis en faisant naître ξ uniformément dans S , l'algorithme est irréductible et apériodique, convergeant vers π .

5.4.2 Simulation inconditionnelle

Supposons que la densité f soit héréditaire : $\forall \mathbf{x} \in \Omega$ et $\xi \in S$, si $f(\mathbf{x} \cup \xi) > 0$, alors $f(\mathbf{x}) > 0$ (c'est le cas de $f(\mathbf{x}) = c\beta^{n(\mathbf{x})}\gamma^{h(\mathbf{x})}$). Autorisant une naissance ou une mort à chaque itération, les configurations circulent dans les différents espaces E_n . Si $\mathbf{x} \in E_n$, $n \geq 1$, un pas de l'algorithme est :

1. avec une probabilité $\alpha_{n,n+1}(\mathbf{x})$, on ajoute un point $\xi \in S$ choisi avec la densité $b(\mathbf{x}, \cdot)$ sur S ;
2. avec la probabilité $\alpha_{n,n-1}(\mathbf{x}) = 1 - \alpha_{n,n+1}(\mathbf{x})$, on supprime un point η de \mathbf{x} avec la probabilité $d(\mathbf{x}, \eta)$.

Pour $n = 0$, on reste dans la configuration vide avec la probabilité $\alpha_{0,0}(\emptyset) = 1 - \alpha_{0,1}(\emptyset)$.

Exemple : algorithme de Metropolis avec changements uniformes.

Pour l'algorithme de Metropolis, la probabilité d'acceptation est :

- $a(\mathbf{x}, \mathbf{x} \cup \xi) = \min\{1, r(\mathbf{x}, \mathbf{x} \cup \xi)\}$ pour la naissance $\mathbf{x} \rightarrow \mathbf{y} = \mathbf{x} \cup \xi$;
- $a(\mathbf{x}, \mathbf{x} \setminus \eta) = \min\{1, r(\mathbf{x}, \mathbf{x} \setminus \eta)\}$ pour la mort $\mathbf{x} \rightarrow \mathbf{y} = \mathbf{x} \setminus \eta$.

Retenons les choix uniformes :

$$\alpha_{n+1,n} = 1/2, \quad b(\mathbf{x}, \cdot) = \frac{1}{\nu(S)} \text{ et } d(\mathbf{x}, \eta) = \frac{1}{n} \text{ si } \mathbf{x} = \{x_1, x_2, \dots, x_n\} \text{ et } \eta \in \mathbf{x},$$

$$\text{alors : } r(\mathbf{x}, \mathbf{x} \cup \xi) = \frac{\nu(S)f(\mathbf{x} \cup \xi)}{n(\mathbf{x})f(\mathbf{x})}.$$

Proposition 5.1 *Pour ces choix, l'algorithme de Metropolis réalise la simulation de f .*

L'algorithme s'écrit :

1. avec la probabilité 1/2, on propose un point uniforme ξ , cette naissance étant retenue avec la probabilité $\inf\{1, r(\mathbf{x}, \mathbf{x} \cup \xi)\}$; sinon, la configuration \mathbf{x} est inchangée.
2. avec la probabilité 1/2, une mort a lieu en η , uniforme sur \mathbf{x} , et on retient cette mort avec la probabilité $\inf\{1, r(\mathbf{x} \setminus \eta, \mathbf{x})^{-1}\}$; sinon, la configuration \mathbf{x} est inchangée.

5.5 Contrôle de la convergence des méthodes MCMC

Le contrôle de la convergence $\|\nu P^k - \pi\| \rightarrow 0$ d'un algorithme MCMC est important car il devrait permettre de proposer une règle d'arrêt de l'algorithme. Si des résultats théoriques existent (i.e. théorème de Perron-Frobenius si Ω est fini), l'évaluation effective de ces vitesses est en général irréalisable car directement liée à la description précise du spectre des valeurs propres de P , ce qui est en général impossible. Une alternative consiste à utiliser un algorithme de simulation exacte proposé par Propp et Wilson [38].

Dans la pratique, il y a plusieurs aspects à considérer : choix de l'algorithme, choix de l'état initial, utilisation d'une seule chaîne ou de plusieurs chaînes indépendantes, choix du nombre n_0 d'itérations à partir duquel on pense que le régime stationnaire est atteint, écart type des simulations, sous-échantillonnage de la chaîne garantissant la presque indépendance. Sur ces sujets, le lecteur pourra consulter ([26]; [44]).

Contrôle de la convergence de la chaîne. Une méthode naturelle consiste à suivre l'évolution en k d'une statistique $h(X_k)$. Si π appartient à une famille exponentielle, on peut prendre pour h la statistique exhaustive de π . La figure 5.2 montre d'une part l'évolution de $n(X_k)$ et de $s(X_k)$ dans la simulation de MH d'un processus de Strauss et d'autre part celle de l'autocorrélation empirique temporelle pour chacune des statistiques : le choix de $n_0 = 300$ comme temps d'entrée en régime stationnaire semble raisonnable et un décalage temporel de $s = 100$ décorrèle approximativement $h(X_k)$ et $h(X_{k+s})$ (cf. Fig. 5.3).

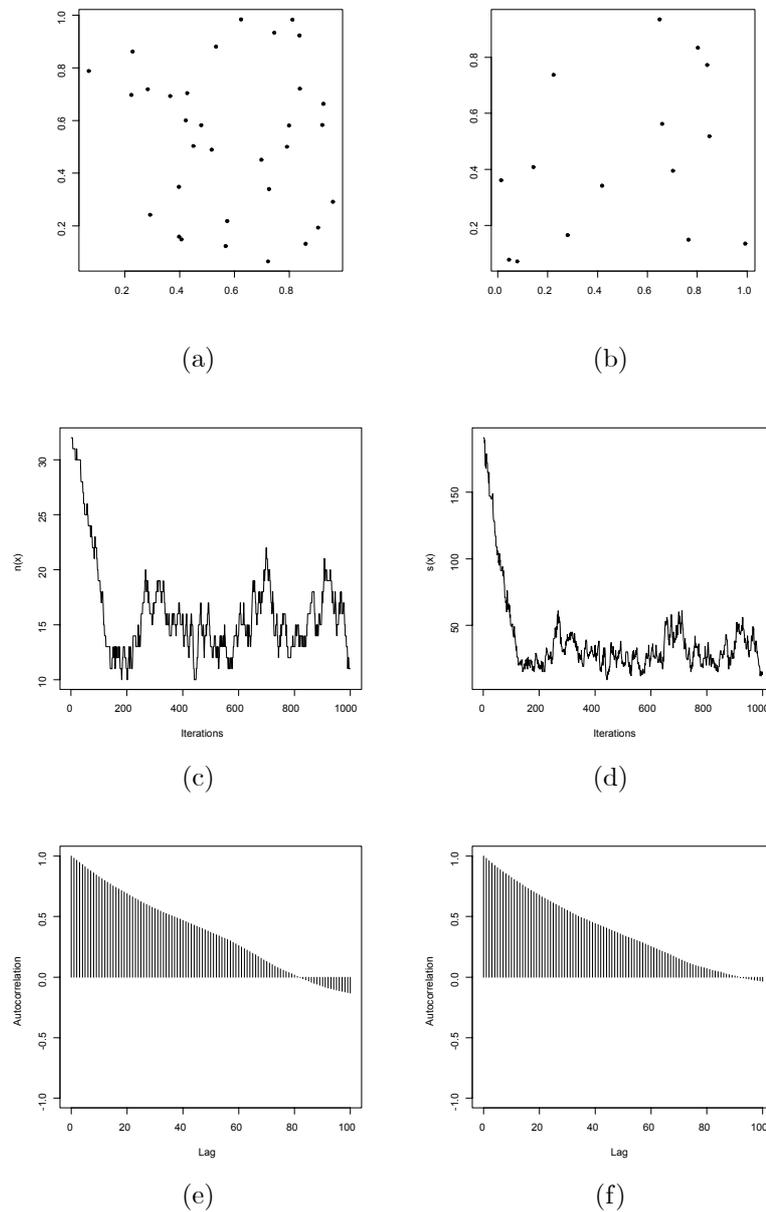


FIG. 5.2 – Simulation inconditionnelle d'un processus de Strauss X de paramètres $\beta = 40$, $\gamma = 0.8$, $r = 0.4$ sur $S = [0,1]^2$ par l'algorithme de Metropolis avec choix uniformes : (a) $X_0 \sim PPP(40)$, (b) X_{1000} . Contrôle de la convergence de l'algorithme avec les statistiques $n_S(X_k)$ (c) et $s(X_k)$ (d) au cours du temps et leurs fonctions d'auto-corrélations (e) et (f).

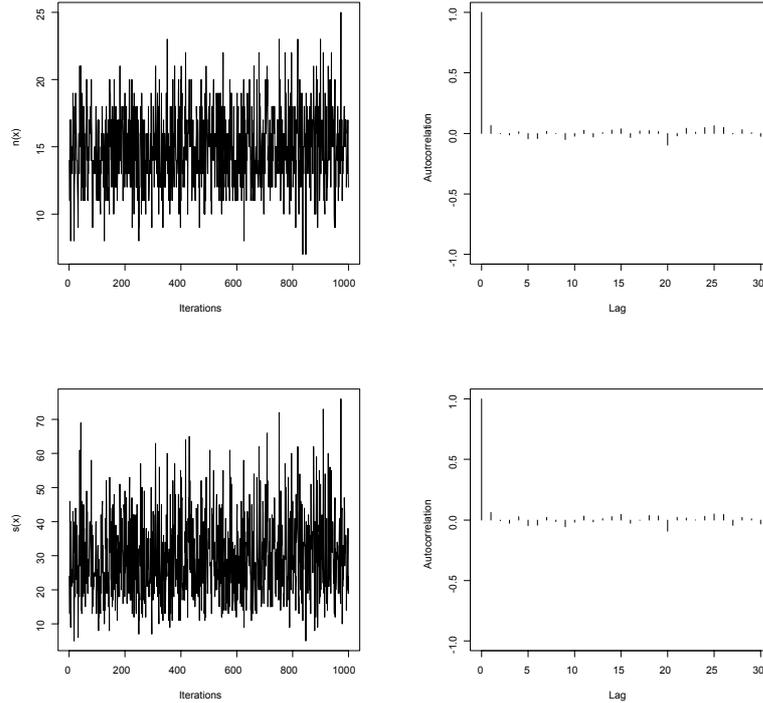


FIG. 5.3 – Sous-échantillonnage de la chaîne de simulation d’un processus de Strauss (cf. Fig. 5.2) : à gauche les statistiques $n_S(X_{sk})$ et $s(X_{sk})$, $k = 0, 1, 2, \dots$ avec $s = 100$; à droite les auto-corrélations correspondantes.

Une autre approche statistique consiste à dérouler m chaînes indépendantes en parallèle et initialisées en des états “dispersés” et à suivre les m trajectoires $\{h_k^{(i)} = h(X_k^{(i)}), k \geq 0\}$. Quand les chaînes atteignent leur régime stationnaire, leurs variances doivent être proches. Notons,

$$B = \frac{1}{m-1} \sum_{1,m} (\bar{h}^{(i)} - \bar{h})^2, W = \frac{1}{m-1} \sum_{1,m} S_i^2 \text{ où } S_i^2 = \frac{1}{n-1} \sum_{k=0, n-1} (h_k^{(i)} - \bar{h}^{(i)})^2.$$

On peut estimer $\text{var}_\pi(h)$ soit par $V = W + B$, soit par W : V est sans biais si $X_0^{(i)} \sim \pi$ mais surestime la variance sinon ; quant à W , il sous-estime la variance pour n fini car la chaîne n’a pas visité tous les états de X . Cependant les deux estimateurs convergent vers $\text{var}_\pi(h)$. On peut donc baser un contrôle de la convergence sur la statistique $R = V/W$: si les chaînes sont entrées en régime stationnaire, $R \cong 1$; sinon $R \gg 1$ et on doit poursuivre les itérations.

5.6 Simulation d’un champ gaussien sur $S \subseteq \mathbb{R}^d$

On veut simuler Y , un *champ gaussien* centré sur un sous-ensemble S de \mathbb{R}^d . Si $S = \{s_1, \dots, s_m\} \subseteq \mathbb{R}^d$ est fini, $Y = (Y_{s_1}, Y_{s_2}, \dots, Y_{s_m})$ est un vecteur gaussien : $\Sigma = \text{cov}(Y)$ étant s.d.p., $\exists T$ une matrice $m \times m$ t.q. $\Sigma = T^t T$, la décomposition de Choleski correspondant au choix où T est triangulaire inférieure. Alors, $Y = T\varepsilon \sim \mathcal{N}_m(0, \Sigma)$ si $\varepsilon \sim \mathcal{N}_m(0, I_m)$. La méthode est générale à condition de savoir calculer T , ce qui devient difficile si m est grand ; par contre, une fois T explicitée, il est facile d’obtenir d’autres simulations de Y à partir d’autres échantillons de ε . Une alternative, bien adaptée au cas gaussien markovien, consiste à utiliser l’échantillonneur de Gibbs.

D’autres méthodes permettent la simulation sur des sous-ensembles continus $S \subseteq \mathbb{R}^d$. Pour plus de détails, le lecteur se reportera au livre de Lantuéjoul [32].

5.6.1 Simulation d’un champ stationnaire

Supposons que l’on sache simuler sur $S \subseteq \mathbb{R}^d$ un champ X stationnaire de L^2 , non-nécessairement gaussien, centré, de variance 1 et de corrélation ρ . Soit $\{X^{(i)}, i \in \mathbb{N}\}$ une suite *i.i.d.* de tels champs et

$$Y_s^{(m)} = \frac{1}{\sqrt{m}} \sum_{i=1}^m X_s^{(i)}.$$

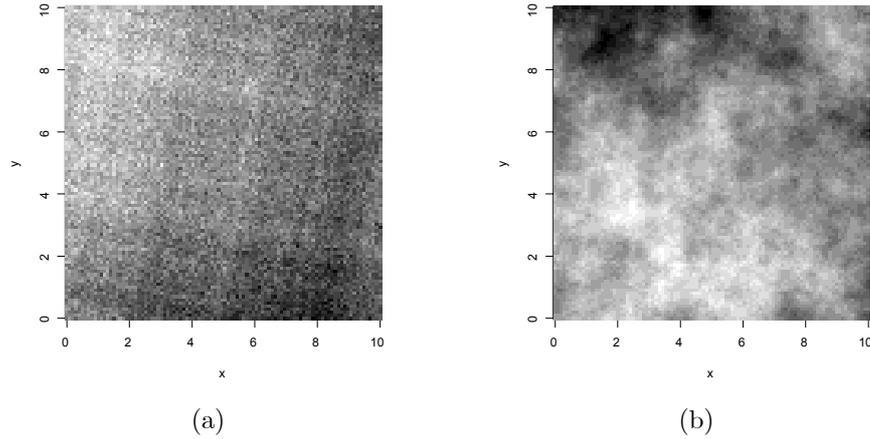


FIG. 5.4 – Deux réalisations d'un champ gaussien de covariance exponentielle obtenues par la méthode avec m bandes tournantes : (a) $m = 2$, (b) $m = 10$.

Pour m grand, $Y^{(m)}$ réalise approximativement un champ gaussien stationnaire centré de corrélation ρ . Reste à simuler le champ générateur X (pour la réalisation de ces simulations, cf. le package `RandomFields`).

La méthode spectrale. $C(h) = \int_{\mathbb{R}^d} e^{i\langle u, h \rangle} F(du)$ montre que F est une probabilité si $\text{Var}(X_s) = 1$. Si $V \sim F$ et $U \sim \mathcal{U}(0, 1)$ sont indépendantes, on considère $X_s = \sqrt{2} \cos(\langle V, s \rangle + 2\pi U)$: puisque $\int_0^1 \cos(\langle V, s \rangle + 2\pi u) du = 0$, X est centré de covariance $C(h)$:

$$C(h) = E(X_s X_{s+h}) = \int_{\mathbb{R}^d} \cos(\langle v, h \rangle) F(dv).$$

Ceci suggère l'algorithme :

1. **simuler** $u_1, \dots, u_m \sim \mathcal{U}(0, 1)$ et $v_1, \dots, v_m \sim F$, **tous indépendants**.
2. $\forall s \in S$, **retourner les valeurs** $Y_s^{(m)} = \sqrt{2} m^{-1} \sum_{i=1}^m \cos(\langle v_i, s \rangle + 2\pi u_i)$

Cette simulation est réalisable si on sait approcher la mesure spectrale F : c'est le cas si F est à support borné ou si F admet une densité à décroissance rapide à l'infini. Sinon, on peut utiliser la méthode des bandes tournantes.

La méthode des bandes tournantes. La méthode, proposée par Matheron, simule un processus isotropique sur \mathbb{R}^d à partir d'un processus stationnaire sur \mathbb{R}^1 . Soit $\mathcal{S}_d = \{s \in \mathbb{R}^d : \|s\| = 1\}$ la sphère de rayon 1 de \mathbb{R}^d , Z un processus sur \mathbb{R}^1 centré,

$$\text{cov}(Z_t, Z_{t'}) = C_Z(t - t'),$$

et V une direction uniforme sur \mathcal{S}_d . On pose $Y_s = Z_{\langle s, V \rangle}$, pour $s \in S$: Y_s est centré puisque $E(Z_{\langle s, V \rangle} | V = v) = E(Z_{\langle s, v \rangle}) = 0$, de covariance

$$C_Y(h) = \int_{\mathcal{S}_d} C_Z(\langle h, v \rangle) \tau(dv)$$

où τ est la loi uniforme sur \mathcal{S}_d . Considérons alors l'algorithme :

1. **simuler** m directions $v_1, \dots, v_m \sim \mathcal{U}(\mathcal{S}_d)$
2. **simuler** $z^{(1)}, \dots, z^{(m)}$, m processus indépendants de corrélations respectives $C_Z(\langle h, v_i \rangle)$, $i = 1, \dots, m$
3. **pour** $s \in S$ **retourner les valeurs** : $Y_s^{(m)} = m^{-1/2} \sum_{i=1}^m z_{\langle s, v_i \rangle}^{(i)}$.

La simulation de la figure 5.4 est la trace sur \mathbb{R}^2 d'une simulation par bandes tournantes sur \mathbb{R}^3 d'un champ de covariance exponentielle isotropique de paramètres $\sigma^2 = 4$ et $a = 10$. Pour choisir m , on contrôlera que les moments d'ordre 3 et 4 de $Y_s^{(m)}$ s'approchent de ceux d'une gaussienne.

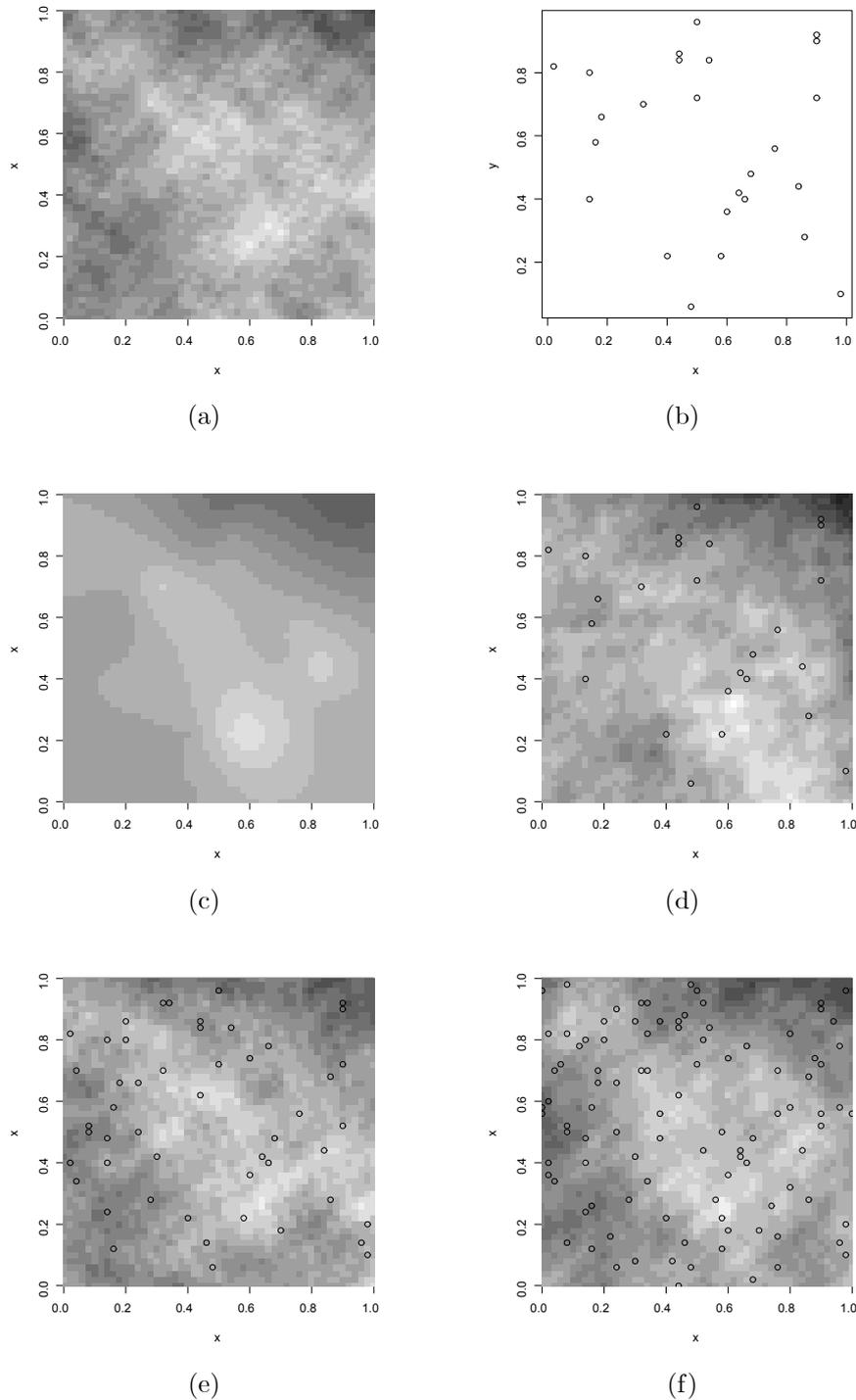


FIG. 5.5 – (a) Réalisation d'un processus gaussien centré de covariance $C(h) = \exp(-\|h\|)$; (b) 25 points échantillonnés; (c) krigeage étant donné les valeurs aux 25 points de (b); (d) simulation conditionnelle à ces 25 valeurs; (e) simulation conditionnelle à 50 valeurs; (f) simulation conditionnelle à 100 valeurs.

5.6.2 Simulation conditionnelle

Soit Y un champ gaussien centré observé sur une partie finie $C \subset \mathbb{R}^d$. On veut simuler Y sur $S \cup C \subseteq \mathbb{R}^d$ conditionnellement à $Y_C = y_C$ (cf. Fig. 5.5). Pour cela, on considère $X = \{X_s, s \in S\}$ une réalisation indépendante de ce champ, $\hat{X}_s, s \in S$, le krigeage simple sur les valeurs $\{X_c, c \in C\}$ et les décompositions :

$$X_s = \hat{X}_s + (X_s - \hat{X}_s), s \in S \text{ et } Y_s = \hat{Y}_s + (Y_s - \hat{Y}_s), s \in S$$

On propose l'algorithme :

1. calculer le prédicteur du krigeage simple \hat{Y}_s sachant Y_C ;
2. simuler X sur $S \cup C$ de même loi que Y ;
3. calculer le krigeage simple $\hat{X}_s, s \in S$, sachant X_C ;
4. retourner les valeurs $\tilde{Y}_s = \hat{Y}_s + (X_s - \hat{X}_s)$ pour $s \in D$.

\tilde{Y}_s réalise la simulation conditionnelle souhaitée. En effet, $\tilde{Y}_c = y_c + X_c - X_c = y_c$ si $s \in C$. D'autre part, $\{(X_s - \hat{X}_s), s \in S\}$ et $\{(Y_s - \hat{Y}_s), s \in S\}$ sont de même loi, celle du résidu conditionnel de Y en s n'étant autre que la loi recentrée de Y conditionnelle à y_C .

5.7 Exercices

Exercice 5.1 On considère le processus de Gibbs $Z_i = (X_i, Y_i) \in \{0, 1\}^2$ sur $S = \{1, 2, \dots, n\}^2$ d'énergie $\sum_{i \in S} \Phi_1(z_i) + \sum_{\langle i, j \rangle} \Phi_2(z_i, z_j)$ où, pour la relation aux 4-ppv,

$$\Phi_1(z_i) = \alpha x_i + \beta y_i + \gamma x_i y_i \text{ et } \Phi_2(z_i, z_j) = \delta x_i x_j + \eta y_i y_j.$$

1. Déterminer les lois $\pi_i(z_i | z^i), \pi_i^1(x_i | x^i, y), \pi_i^2(y_i | x, y^i)$.
2. Construire l'échantillonneur de Gibbs basé sur $\{\pi_i^1, \pi_i^2, i \in S\}$.

Exercice 5.2 Simulation des modèles de Matérn.

(1) Soit X un PPP(λ) homogène. Le modèle I de Matérn s'obtient en effaçant tous les couples de points à distance $\leq r$: $X^I = \{s \in X : \forall s' \in X, s' \neq s, \|s - s'\| > r\}$. Montrer que l'intensité du nouveau processus vaut $\lambda^* = \lambda \exp\{-\lambda \pi r^2\}$. Mettre en oeuvre la simulation d'un tel processus sur $[0, 1]^2$ avec $r = 0.02$ et $\lambda = 30$.

(2) Soit X un PPP(λ) homogène. A chaque point s de X on attache une marque continue indépendante $Y_s \sim h(\cdot)$. Le modèle II de Matérn s'obtient en effaçant un point s si $\exists s'$ a une distance $\leq r$ et si $Y_{s'} < Y_s$. Mettre en oeuvre la simulation d'un tel processus sur $[0, 1]^2$ avec $r = 0.02, \lambda = 30$ et h la distribution exponentielle de moyenne 4.

Exercice 5.3 On veut simuler la loi gaussienne bivariable sur S d'énergie U donnée par :

$$-U(x, y) = \sum_{i \in S} (x_i^2 + y_i^2) + \beta \sum_{\langle i, j \rangle} (x_i y_j + x_j y_i), |\beta| < 1/2.$$

Déterminer les lois conditionnelles $\pi_i(z_i | z^i), \pi_i^1(x_i | x^i, y)$ et $\pi_i^2(y_i | x, y^i)$. Proposer deux procédures de simulation.

Exercice 5.4 Soit Σ_n l'ensemble des permutations de $\{1, 2, \dots, n\}$ et π la loi uniforme sur Σ_n . Montrer que la transition P sur Σ_n qui permute au hasard deux indices i et j (avec l'éventualité $i = j$) est π -réversible, transitive et apériodique.

Exercice 5.5 Modèle à noyau dur sur un réseau discret.

Soit S un ensemble fini de sites muni d'un graphe de voisinage symétrique \mathcal{G} . Un modèle à noyau dur sur (S, \mathcal{G}) est un modèle binaire à états dans $E_0 = \{x \in \{0, 1\}^S : \text{si } \langle i, j \rangle, x_i x_j = 0\}, x_i = 0$ (resp. $x_i = 1$) représentant un site libre (resp. occupé). Ainsi, deux sites voisins de S ne peuvent simultanément être occupés. Soit π la loi uniforme sur E_0 et P la transition $x_n \mapsto x_{n+1}$ sur E_0 :

- (i) choisir au hasard uniforme un site i de S ;
- (ii) jeter une pièce non pipée : si pile et si les sites voisins de i sont à 0, alors $x_{n+1}(i) = 1$; sinon $x_{n+1}(i) = 0$;
- (iii) si $j \neq i, x_{n+1}(j) = x_n(j)$.

Vérifier que P est irréductible, apériodique, π -réversible. Application : si $S = \{1, 2, \dots, 10\}^2$ et si $\langle i, j \rangle$ est la relation de voisinage aux 4-ppv, évaluer le cardinal de E_0 .

Exercice 5.6 Réaliser la simulation par échange de spins du modèle d'Ising aux 4-ppv de paramètres $\alpha = 0$ et β sur le tore $\{1, 2, \dots, 64\}^2$, la configuration initiale étant équilibrée entre les deux spins $+1$ et -1 . Calculer la corrélation empirique $\rho(\beta)$ à distance 1 et construire la courbe empirique $\beta \mapsto \rho(\beta)$. Même étude en utilisant l'échantillonneur de Gibbs.

Exercice 5.7 *Simulation d'une texture à G-niveau de gris.*

Un Φ -modèle de texture à G niveaux de gris $\{0, 1, \dots, G-1\}$ a pour d'énergie $U(x) = \theta \sum_{\langle i, j \rangle} \Phi_d(x_i - x_j)$, où $\Phi_d(u) = \{1 + (u/d)^2\}^{-2}$. Pour ce modèle, les contrastes augmentent avec d et θ contrôle la corrélation spatiale. Simuler différentes textures par échantillonneur de Gibbs en faisant varier les paramètres θ et d ainsi que le nombre de niveaux de gris G .

Exercice 5.8 Mettre en oeuvre la dynamique de Métropolis pour simuler un processus de Strauss sur $[0, 1]^2$ à $n = 50$ points. Examiner les configurations simulées pour $\gamma = 0.01, 0.5, 1, 2, 10$ et $r = 0.05$.

Exercice 5.9 Soit $S = \{s_1, s_2, \dots\}$ un sous-ensemble discret de \mathbb{R}^2 . En utilisant le krigeage simple, mettre en oeuvre une procédure récursive de simulation sur S du processus gaussien centré de covariance $C(h) = e^{-\|h\|/4}$.

Bibliographie

- [1] ADLER, R.J. [1981], *The Geometry of Random Fields*, New York, Wiley.
- [2] AMEMIYA, T. [1985], *Advanced Econometric*. Oxford, Basil Blackwell.
- [3] ANSELIN L. [1988], *Spatial econometrics : Methods and models*, Kluwer.
- [4] ARNOLD, B.C., CASTILLO, E. & SARABIA, J.M. [1999] *Conditional Specification of Statistical Models*, New York, Springer.
- [5] BADDELEY, A.J. & TURNER, T.R. [2000], Practical maximum pseudolikelihood for spatial point patterns (with discussion), *Australian and New Zealand Journal of Statistics*, **42**, pp. 283-322.
- [6] BADDELEY, A.J. & T.R. TURNER [2005], **Spatstat** : an R package for analyzing spatial point patterns. *Journal of Statistical Software*, pp. 1-42.
- [7] BESAG, J. [1974], Spatial interaction and the statistical analysis of lattice systems, *JRSS B*, **36**, pp. 192-236.
- [8] BESAG, J. [1977], Efficiency of pseudo likelihood estimation for simple gaussian fields, *Biometrika*, **64**, pp. 616-618.
- [9] BESAG, J., & P.A.P. MORAN [1975], On the estimation and testing of spatial interaction for gaussian lattice processes, *Biometrika*, **62**, pp. 555-562.
- [10] BOUTHEMY P., HARDOUIN C., PIRIOU G. ET YAO J.F. [2006], Mixed-state Auto-models and motion texture modeling, *J. Math. Imaging Vision*, **25**, pp. 387-402.
- [11] CHALMOND, B. [2000], *Eléments de modélisation pour l'analyse d'image*. Springer.
- [12] CHILÈS, J.-P. & DELFINER, P. [1999], *Geostatistics*, New York, Wiley.
- [13] CLIFF, A.D. & ORD, J.K. [1981] *Spatial Processes : Models and Applications*, London, Pion.
- [14] COMETS, F. [1992], On consistency of a class of estimators for exponential families of Markov random fields on a lattice, *Annals of Statistics*, **20**, pp. 455-468.
- [15] COMETS, F. & JANZURA, M. [1998], A central limit theorem for conditionally centered random fields with an application to Markov fields, *Journal of Applied Probability*, **35**, pp. 608-621.
- [16] CRESSIE, N. [1993], *Statistics for Spatial Data*, New York, Wiley.
- [17] CROSS, G.R. & JAIN, A.K. [1983], Markov field texture models, *IEEE-PAMI*, **5**, pp. 155-169.
- [18] DALHAUS, R. & KÜNSCH, H. [1987], Edge effect and efficient parameter estimation for stationary random fields, *Biometrika*, **74**, 877-882.
- [19] DIGGLE, P.J. [2003], *Statistical Analysis of Spatial Point Patterns*, Oxford, Oxford University Press.
- [20] DIGGLE P. & RIBEIRO P.J. [2006], *Model-based Geostatistics*, Springer Series in Statistics, 2006.
- [21] J.J. DROESBEKE, M. LEJEUNE & G. SAPORTA (EDITEURS) [2006], *Analyse statistique des données spatiales*, Technip.
- [22] FERRÁNDIZ J., LÓPEZ A., LLOPIS A., MORALES M. & TEJERIZO M.L. [1995], Spatial interaction between neighbouring countries : cancer mortality data in Valencia (Spain), *Biometrics* **51**, pp. 665-678.
- [23] GEMAN D. & GEMAN S. [1984], Stochastic relaxation, Gibbs distributions and the bayesian restoration of images, *IEEE-PAMI*, **6**, 721-741.
- [24] GEORGII, H.O. [1988], *Gibbs measure and phase transitions*. De Gruyter.
- [25] GEYER, C.J. [1994], On the convergence of Monte Carlo maximum likelihood calculations, *JRSS B*, **56**, 261-274.
- [26] GILKS, W.R., S. RICHARDSON & D.J. SPIEGELHALTER [1996], *Markov Chain Monte Carlo in Practice*. London, Chapman & Hall.
- [27] GUYON, X. & HARDOUIN C. [1992], The Chi-2 difference of coding test for testing Markov random field hypothesis, *Lecture Notes in Statistics*, 165-176, Berlin, Springer.
- [28] GUYON, X. [1995], *Random Fields on a Network : Modeling, Statistics and Applications*, New York, Springer.

- [29] HASTINGS, W. [1970], Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97-109.
- [30] IHAKA R. & GENTLEMAN, R. [1996], R : a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, pp. 299-314.
- [31] JENSEN, J.L. & MØLLER, J. [1991], Pseudolikelihood for exponential family of spatial point processes, *Annals of Applied Probability*, **3**, pp. 445-461.
- [32] LANTUÉJOUL, C. [2002], *Geostatistical Simulation*, Berlin, Springer.
- [33] MARDIA, K. V. & MARSHALL, J. [1984], Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **71**, 289-295.
- [34] MATHERON, G. [1973], The intrinsic random function and their applications, *Advances in Applied Probability*, **5**, 439-468.
- [35] MERCER, W.B. & HALL, A.D. [1911], The experimental error of field trials, *J. Agricultural Sc.*, **4**, 107-132.
- [36] MØLLER, J. & WAAGEPETERSEN, R.P. [2004] *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall.
- [37] PEYRARD, N., CALONNEC, A., BONNOT, F. & CHADOEUF, J. [2005], Explorer un jeu de données sur grille par test de permutation, *Revue de Statistique Appliquée*, **LIII**, 59-78.
- [38] PROPP, J.G. & WILSON, D.B. [1996], Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms*, **9**, 223-252.
- [39] R DEVELOPMENT CORE TEAM [2005], *R : A language and environment for statistical computing*, Vienna, Austria, R Foundation for Statistical Computing.
- [40] RICHARSON, S., GUIHENNEUC, C. & LASSERRE, V. [1992], Spatial linear models with autocorrelated error structure, *The Statistician*, **41**, 539-557.
- [41] RIPLEY, B.D. & KELLY, F.P. [1977], Markov point processes, *J. London Math. Soc.*, **15**, pp. 188-192.
- [42] RIPLEY, B.D. [1981], *Spatial Statistics*. New York, Wiley.
- [43] RIPLEY, B.D. [1988], *Statistical Inference for Spatial Processes*. Cambridge, Cambridge University Press.
- [44] ROBERT, C. & CASELLA, G. [1999], *Monte-Carlo statistical methods*. Springer
- [45] SCHABENBERGER, O & GOTWAY, C.A. [2004], *Statistical methods for Spatial Data Analysis*, Chapman and Hall.
- [46] STEIN, M.L. [1999], *Interpolation of Spatial Data : Some Theory for Kriging*. New York, Springer.
- [47] STOYAN, D., KENDALL, W.S. & MECKE, J. [1995], *Stochastic Geometry and its Applications* 2nd ed. , New York, Wiley.
- [48] TIERNEY, L. [1994], Markov chains for exploring posterior distributions, *Ann. Statist*, **22**, 1701-1762.
- [49] TUCKEY, J.W. [1967], *Spectral Analysis Time Series*, New York, Wiley.
- [50] VAN LIESHOUT, M.N.M. [2000], *Markov Point Processes and their Applications*, London, Imperial College Press.
- [51] WACKERNAGEL, H. [2003] *Multivariate Geostatistics : An Introduction with Applications* New York, Springer.
- [52] WHITTLE, P. [1954], On stationary processes in the plane, *Biometrika*, **41**, 434-449.
- [53] WINKLER, G. [2003], *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*, New York, Springer.
- [54] WU, H. & F.W. HUFFER [1997], Modelling the ditribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics*, **4**, pp. 49-64.
- [55] YAGLOM, A.M. [1987], *Correlation Theory of Stationary and Related Random Functions. Volume I : Basic Results*, New York, Springer.
- [56] YCART, B. [2002], *Modèles et algorithmes markoviens*, Springer.

Index

- Algorithme de Metropolis, 75, 78
- Analyse variographique, 9
- Anisotropie, 5, 10
- AR (auto-régression), 18
 - conditionnelle (CAR), 18
 - estimation d'une, 29
 - factorisante, 18, 34
 - non-stationnaire, 20
 - simultanée (SAR), 18
- ARMA, 17
- Auto-modèle, 41
 - avec covariables, 43
 - binomial, 42
 - exponentiel, 43
 - gaussien, 43
 - logistique, 42
 - poissonnien, 42
- Bootstrap paramétrique, 13, 65
- Bruit blanc, 2
- Bruit coloré, 19
- CAR (AR conditionnelle), 18, 19
 - estimation d'un, 45
 - non-stationnaire, 21
- Chaîne de Markov, 51, 73
 - apériodique, 74
 - irréductible, 73
 - réversible, 74
- Champ de Gibbs, 38, 50, 58, 62, 68
- Champ de Markov, 37, 41
 - estimation d'un, 44
 - pour la segmentation, 50
 - simulation d'un, 76
- Champ gaussien markovien, 19, 21
- Clique d'un graphe, 41, 50
- Clique d'un PP, 62
- Codage d'un graphe, 47
- Corrélation spatiale, 23
- Covariance, 1
 - définie positive, 1
 - exponentielle, 3
 - isotropique, 2, 3
 - représentation spectrale, 3
 - stationnaire, 2
- Covariance empirique, 27
- Densité d'un PP, 57
- Densité spectrale, 3
- Distance au p.p.v. d'un PP, 61
- Echantillonneur de Gibbs, 74
 - aléatoire, 75
 - séquentiel, 75
- Effets de bord, 65
 - pour un PP, 63
 - sur un lattice, 27
- Espace exponentiel d'un PP, 53
- Estimation
 - d'un champ de Markov, 44
 - d'un champ gaussien, 30
 - d'un champ stationnaire, 27
 - d'un processus ponctuel, 63
 - d'une AR, 29
 - d'une régression spatiale, 30
 - par codage, 47
- Estimation d'un variogramme, 10
 - empirique, 10
 - paramétrique, 12
- Gibbs
 - Champ de, 38, 41
 - Echantillonneur de, 74
 - PP de, 58
- Graphe de Markov, 21, 41
- Identifiabilité, 18, 20, 46, 67
 - d'un potentiel, 38
- Indépendance spatiale, 23
 - Test (gaussien), 24
 - test de permutation, 25
- Indice d'autocorrelation
 - de Geary, 24
 - de Moran, 23
- Information de Fisher, 48, 68
 - d'une régression spatiale, 32
 - de l'estimation par codage, 47
 - MV d'un champ de Markov, 47
 - pseudo-information d'une PVC, 47
- Ising (modèle d'), 39
- Isotropie, 2, 18, 39, 49
- Krigeage, 7, 13
 - ordinaire, 8
 - simple, 8
 - universel, 8
- Méthode de Monte Carlo
 - pour un PP, 65
 - simulation par, 73
 - validation d'un variogramme, 12
- MA (moyenne glissante), 17, 19
- Maximum de vraisemblance, 1, 12
 - d'un champ gaussien, 30
 - pour un champ de Markov, 44
 - pour un PP, 68

- pour une régression spatiale, 31
- MCMC, 74
- Metropolis, 75
 - par échange de spins, 76, 84
 - simulation d'un PP, 78
- Modèle de Potts, 40, 50
- Moindres carrés, 12, 65
 - généralisés (MCG), 8, 31
 - ordinaires (MCO), 18, 31
 - pondérés, 10, 65
 - quasi généralisés (MCQG), 31
- Moment K d'un PP, 61, 64
- Monte Carlo, 13, 25, 65
 - calcul d'une vraisemblance, 68
 - simulation par (MCMC), 74
- Nuée variographique, 9
- Potentiel, 38
 - d'interaction, 50
 - d'un champ de Markov, 41
 - d'un PP, 58, 62
 - invariant par translation, 40
- PP de Cox, 57
- Processus
 - au second ordre, 1
 - continuité, 6
 - dérivabilité, 6
 - gaussien, 1
 - intrinsèque, 3
 - linéaire, 17
 - stationnaire, 2
- Processus ponctuel, 53
 - à noyau dur, 56, 58
 - binomial, 54
 - caractéristiques d'ordre 2, 60
 - de Gibbs, 58
 - de Neyman-Scott, 56, 70
 - de Strauss, 58
 - densité inconditionnelle d'un, 57
- Processus ponctuel de Markov, 62
 - intensité conditionnelle d'un, 62
- Processus ponctuel de Poisson, 55
 - densité, 58
 - doublement stochastique, 57
 - inhomogène, 55
 - intensité d'un, 55
- Processus ponctuel marqué, 54, 57
- Pseudo-vraisemblance
 - conditionnelle, 45
 - conditionnelle d'un PP, 67
 - d'un CAR gaussien, 45
 - gaussienne de Whittle, 28
 - marginale d'un PP, 63
- Régression spatiale, 8, 34
 - estimation, 30
- Rabotage des données, 27
- SAR (AR simultanée), 18, 19
 - aux plus proches voisins, 22
- SARX (SAR avec eXogène), 22
 - a décalage spatial, 22
 - modèle de Durbin, 22
- Semi définie positive, 1, 14
- Semi-variogramme, 3
- Simulation, 73
 - d'un auto-modèle, 76
 - d'un champ de Markov, 76
 - d'un champ gaussien, 80
 - d'un PP, 77
 - d'une texture, 45
 - de texture, 76
 - gaussienne conditionnelle, 83
- Spécification de Gibbs, 38, 39
 - gaussienne, 40
- Test de Monte Carlo
 - d'indépendance spatiale, 25
 - de validation de variogramme, 13
- Test de permutation, 25
- Texture, 45, 76
- Transition
 - d'une chaîne de Markov, 73
 - de l'échantillonneur de Gibbs, 74
 - de Metropolis, 75
 - réversible, 74
- Validation croisée, 12
- Validation d'un modèle
 - de PP, 65, 69
 - de variogramme, 12
 - par Bootstrap paramétrique, 13
- Variogramme, 3
 - anisotropique, 5
 - de Matérn, 4
 - exponentiel, 4
 - gaussien, 4
 - pépitique, 4
 - puissance, 4
 - sphérique, 4
- Variogramme empirique, 10