SYLVIA RICHARDSON

Processus spatialement dépendants : convergence vers la normalité, tests d'association et applications, 1989

ORSAY
n° d'ordre : 3559

UNIVERSITE PARIS SUD

Centre d'Orsay

T H E S E

De Doctorat d'Etat Es Sciences Mathématiques

présentée pour obtenir le grade de

DOCTEUR ES-SCIENCES

par

Sylvia RICHARDSON

Sujet : PROCESSUS SPATIALEMENT DEPENDANTS : CONVERGENCE VERS LA NORMALITE, TESTS D'ASSOCIATION ET APPLICATIONS.

Soutenue le 19 Octobre 1989 devant la Commission d'examen composée de :

MM.	DACUNHA-CASTELLE	Didier,	Président
	BRETAGNOLLE	Jean	
	HUBER	Catherine	
	LELLOUCH	Joseph	
	THOUVENOT	Jean-Paul	

ABSTRACT

The first part of this thesis is concerned with the study of ergodic properties of stopping time transformations, T^v where T is an automorphism and v a transformation with values in Z. The demonstrations use cutting and stacking methods. Results when T is Bernoulli have been published in Z. Warhs. Verv. (vol 48, 1979, with K.M. Wilkinson) and density results when T is ergodic in the Proc. London Math. Soc (vol 43, 1981).

The second part of this thesis is composed of several papers, in probability and in statistics, about spatial processes and the testing of association. A first paper, published in J.A.P. (vol 18, 1981 with D. Hémon) derives the asymptotic variance of the correlation coefficient between two independent lattice processes. The speed of convergence in the Central Limit Theorem for weakly dependent processes in Z^d is the subject of the second paper (published in Z. Warhs. Verw (vol 66, 1984, with X. Guyon). The speed is dependent on the dimension d, on moments conditions and on mixing conditions.

The development of tests of association between spatial processes having a controlled Type I error form the statistical part of this thesis. These tests were proposed in a paper published in Biometrics (vol 45, 1989, with P. Clifford and D. Hémon). Their power and an extension to the testing of partial correlations was further studied (to appear in Statistics in Medecine). The applications discussed concern the field of geographical correlations in epidemiology (Int. J. Epidemiol vol 16, 1987, with I. Stücker and D. Hémon).

Keywords : Ergodic properties of stopping times, spatial processes, Central Limit Theorem, Speed of convergence, Correlation coefficient, Tests of association, Effective degrees of freedom, geographical epidemiology.

A.M.S. code : 28 D 05, 60 F 05, 60 G 60, 62 E 25, 62 H 20, 62 M 30, 62 P 10, 92 A 15.

AVANT-PROPOS

et

REMERCIEMENTS

Mes recherches, commencées dans le département de Mathématiques de l'Université de Nottingham, ont d'abord porté sur un sujet de théorie ergodique suggéré par K.M. Wilkinson, mon directeur de Ph.D. J'ai étudié les propriétés ergodiques des temps d'arrêt auxquels ont été imposées des conditions telles que ces transformations soient des automorphismes d'un espace de Lebesgue continu. Ces temps d'arrêt peuvent aussi être considérés comme des temps d'arrêt probabilistes qui sont définis sur une suite de variables, mais la classe des temps d'arrêt probabilistes est strictement plus large.

Je me suis ensuite intéressée à d'autres types de processus, les processus spatialement dépendants et plus particulièrement à l'étude de tests d'association. Mes recherches ont été menées à la fois du point de vue probabiliste et du point de vue statistique. Elles ont aussi comporté un domaine d'application, les études de corrélations écologiques. Ces recherches ont été faites dans le Laboratoire de Statistiques Médicales de l'Université de Paris V et dans l'Unité 170 de l'INSERM en collaboration avec D. Hémon (U.170), X. Guyon (Université de Paris I) et P. Clifford (Université d'Oxford).

Je tiens tout d'abord à remercier Didier Dacunha-Castelle, Professeur, qui me fait l'honneur de présider le jury de cette thèse ainsi que Jean Bretagnolle, Professeur, qui a si gentiment accepté d'en assurer la direction.

Je dois remercier tout particulièrement mes amis et collègues avec qui j'ai signé des articles :

Keith Wilkinson qui m'a introduit à la théorie ergodique ;

Xavier Guyon qui m'a fait comprendre les beautés et les difficultés des processus dans Z^2 ;

Peter Clifford dont la rigueur et les intuitions statistiques m'ont beaucoup apporté ;

Denis Hémon qui, dès le départ, a su me montrer l'intérêt et les différents aspects de ce sujet et m'a soutenu activement tout au long de mes recherches.

Qu'ils trouvent ici l'expression de ma reconnaissance.

Je remercie chaleureusement Catherine Huber, Professeur, et Joseph Lellouch, Directeur de recherches, des encouragements qu'il m'ont toujours prodigués et d'avoir bien voulu faire partie du jury.

Je voudrais aussi remercier mes deux rapporteurs pour la commission des thèses, Monsieur J.P. Thouvenot, Directeur de recherches, et Monsieur E. Bolthausen, Professeur ; Monsieur J.P. Thouvenot a également eu la gentillesse de faire partie du jury.

Il me faut encore remercier les membres du Laboratoire de Statistiques Médicales de l'Université de Paris V pour le soutien qu'ils m'ont apporté ; ainsi que les chercheurs et techniciens de l'Unité 170 de l'INSERM qui ont beaucoup contribué à faire aboutir ce travail.

Ces remerciements ne seraient pas complets si je ne disais pas que c'est à Philippe Lazar, Directeur Général de l'INSERM, que je dois l'intérêt qui ne m'a pas quitté pour l'épidémiologie géographique. Qu'il trouve ici l'expression de ma gratitude.

TABLE DES MATIERES

Première partie : Processus ergodiques : études des propriétés ergodiques des temps d'arrêt.

Les définitions et les résultats classiques en théorie ergodique utilisés dans cette partie peuvent se trouver dans les ouvrages de Halmos (5) et Walters (9).

Les temps d'arrêt considérés sont de la forme τ^v où τ est un automorphisme (c'est-à-dire une transformation bijective préservant la mesure) d'un espace de Lebesgue (Ω,B,μ) et v une transformation mesurable à valeur dans N, telle que les ensembles $\tau^v\{v = n\}$ forment une partition de Ω. Dans ce cas la transformation τ^v est-elle même un automorphisme. Cette classe de transformations inclut les temps de retour $\tau^v{}_D$ à un ensemble D et la restriction de $\tau^v{}_D$ à l'espace de Lebesgue (D,B_D,μ_D) coïncide avec τ_D l'automorphisme induit.

Ces temps d'arrêt ont été introduits par Neveu (6) qui en a donné une décomposition en termes de produit de temps de retour associés à une suite décroissante d'ensemble $\{D_i\}$. Quand l'espérance de v, E(v), est finie, on peut montrer qu'elle est égale à un entier k qui coïncide avec le nombre d'ensembles $\{D_i\}$, i = 0..., k-1 dans la décomposition de Neveu. L'entropie de τ^v est reliée à celle de τ via l'espérance de v.

De nombreux auteurs se sont intéressés aux propriétés des transformations induites : Friedman (3), Friedman et Ornstein (4), Ornstein et Smorodinski (7). Nous avons cherché à voir comment leurs résultats pouvaient se généraliser aux temps d'arrêt ; une première différence étant que τ^v n'est pas nécessairement ergodique quand τ l'est, bien que τ_D le soit également.

Mes recherches se sont portées essentiellement sur deux aspects. Dans un premier temps, j'ai étudié les temps d'arrêt liés à une transformation τ linéaire par morceaux, définie sur [0,1] par une "tour". Dans un deuxième temps, la transformation τ est seulement supposée ergodique.

1. Temps d'arrêts et "tours"

Une tour, définie sur [0,1], est un ensemble de colonnes composées d'intervalles disjoints de même mesure, empilés les uns sur les autres. A l'intérieur des colonnes, une application partielle bijective est définie de manière à ce que chaque intervalle soit linéairement appliqué sur l'intervalle situé directement au-dessus. Au sommet de chaque colonne, la transformation est définie par récurrence en utilisant la méthode du "cutting and stacking". A l'étape n, la tour existant T_n est remplacée par une nouvelle tour T_{n+1} constituée par une copie réduite de moitié de T_n : $1/2T_n$,à la base (cutting) et par une copie de T_n au dessus de chaque colonne de $1/2T_n$ (stacking). Ainsi à l'étape n, la transformation est étendue linéairement par morceaux à la moitié du sommet. Comme la mesure du sommet tend vers zéro quand n tend vers l'infini, la transformation est éventuellement définie sur l'ensemble des intervalles qui composent la tour. Elle préserve la mesure et elle est ergodique.

Cette classe de transformations, introduite par Chacon (1), a été largement étudiée (voir Friedman (2)). Friedman (3) et Shields (8) ont montré que les transformations définies par des tours sont markoviennes (Markov shifts). Il est facile de poser des conditions telles qu'elles soient, de plus, mélangeantes. Par isomorphisme, cette classe inclut donc les transformations de Bernoulli (Bernoulli shifts) d'entropie finie quelconque (Friedman et Ornstein (4)).

Quand τ est une transformation définie par une tour T, on peut suivre explicitement l'action de τ^v dans la tour quand les intervalles de T sont des sous-ensembles de la partition $E_i = D_{i-1} - D_i$, $0 \leq i \leq k$, associée au temps d'arrêt v, $E(v) = k$. Pour certains intervalles (au plus k) dans chaque colonne, la transformation de τ^v n'est pas définie. Ceci conduit à décomposer la tour T en une union de k tours $\{T_i\}$ disjointes. A l'intérieur de chaque T_i, les intervalles sont transformés par τ^v. L'action de τ^v sur les sommets des colonnes peut être identifiée grâce à des fonctions associées à v, définies sur la partition des intervalles de la tour. Ces fonctions caractérisent de manière unique un ensemble de permutations qui indiquent quelle tour placer au dessus de quelle colonne dans la procédure de "cutting and stacking".

Cette construction géométrique, que nous avons nommée : "tower blocks", peut être étudiée indépendamment des temps d'arrêts. La transformation qui lui est associée est markovienne et on peut donner des conditions explicites pour qu'elle soit ergodique ou de Bernoulli. On en déduit que si τ est de Bernoulli d'entropie finie, l'ensemble des τ^v qui sont de Bernoulli est dense (par rapport à la métrique uniforme) dans la famille des transformations $\{\tau^v\}$. Ceci généralise un théorème de Friedman (3) pour les transformations induites.

Ces résultats faisaient partie de mon doctorat (Ph.D) que j'ai soutenu en 1978 à l'Université de Nottingham (Directeur K.M. Wilkinson); ils ont été publiés dans Z. Wahrscheinlichkeitstheorie verw.(liste des travaux (1)).

2. Approximation des temps d'arrêt

Dans un deuxième temps, en supposant seulement τ ergodique, j'ai montré comment on peut approcher une transformation τ^v quelconque par une transformation $\tau^{v'}$ de même entropie et ergodique sur l'ensemble $\{v' \neq 0\}$. A nouveau, ces résultats sont fondés sur une construction explicite par induction de la transformation $\tau^{v'}$. L'étape inductive est fournie par le théorème de Rohlin, qui donne une représentation géométrique d'une transformation ergodique τ quelconque à l'aide de τ-colonnes. Le théorème de Rohlin a aussi été utilisé par Friedman et Ornstein (4) dans leur construction d'une transformation mélangeante. Les fonctions associées à v qui ont été introduites dans la construction précédente sont à nouveau sollicitées et le temps d'arrêt v' est défini comme la limite d'une suite de temps d'arrêt $v^{(r)}$. A chaque étape le temps d'arrêt $v^{(r)}$ est légèrement modifié afin que la transformation limite soit ergodique. Une fois l'ergodicité obtenue, en utilisant des résultats connus sur les transformations induites (4,7), on peut en déduire des résultats analogues concernant la densité des τ^v mélangeantes ou possédant la propriété de Kolmogorov.

Ces résultats font également partie de mon Ph.D et ont été publiés dans les Proceedings of the London Mathematical Society (liste des travaux (2)).

Deuxième partie : Processus spatialement dépendants : convergence vers la normalité, tests d'association et applications.

Etudier ou tester l'association entre deux variables aléatoires est un problème classique en statistique. Pearson en 1896 introduisait déjà une mesure d'association : le coefficient de corrélation empirique. Dans le travail que je vais présenter, je me suis intéressée au cas où les variables sont observées en différents points d'un espace géographique et sont des réalisations de processus stochastiques pour lesquels une dépendance entre des variables indexées par des points géographiquement proches existe.

Cette dépendance ou autocorrélation spatiale peut avoir des origines diverses. Il peut s'agir d'une propriété intrinsèque au processus lui-même, comme l'existence d'interactions entre les différents sites, de phénomènes de diffusion ou de contagion. Cette autocorrélation peut également résulter indirectement de l'influence sur la variable étudiée d'un certain nombre d'autres facteurs aléatoires ou déterministes qui varient à une échelle spatialement plus large que celle du réseau de points considérés. C'est le cas de variables définies comme une moyenne sur une unité géographique de variables dichotomiques observées par individu. Nos exemples d'applications, choisis en épidémiologie, se situent dans ce cadre d'autocorrélation crée indirectement.

Cette dépendance spatiale a de multiples conséquences en particulier sur les conditions de convergence vers la normalité dans le Théorème de la Limite Centrale ainsi que sur la vitesse de cette convergence.

Du point de vue des tests d'association, les méthodes statistiques classiques fondées sur la corrélation ou la régression ne sont pas applicables directement au cas de variables spatialement dépendantes et il est nécessaire de les modifier. Les conséquences du fait de ne pas prendre en compte l'autocorrélation dans les modèles de régression ont été passées en revue par Johnston (13) pour les séries chronologiques et Cliff et Ord (5) pour les séries spatiales.

Certains auteurs ont proposé des mesures d'association qui impliquent une transformation des données par filtrage pour réduire l'autocorrélation avant d'appliquer des techniques standards : ce sont les travaux de Haugh (10) développés dans le cadre des séries chronologiques. La généralisation de cette méthode à des données spatiales irrégulières soulève des difficultés.

Mes travaux ont considéré deux aspects : l'un probabiliste et l'autre statistique. J'ai aussi développé des applications en épidémiologie.

1. Approche probabiliste

1.1 Variance asymptotique du coefficient de corrélation

Il s'agit d'une étude sur la variance asymptotique du coefficient de corrélation entre deux processus mutuellement indépendants, chacun étant autocorrelé spatialement. Dans le cadre des séries chronologiques, la variance et les covariances des coefficients de corrélation croisés ont été étudiés par Bartlett (2,3).

La variance asymptotique est établie pour des processus gaussiens sur un lattice et une suite croissante $\{D_n\}$ de domaines de Z^2 de forme quelconque, de cardinal $|D_n|$.

Supposons que $\{X(u)\}$ et $\{Y(u)\}$, $u \in Z^2$, soient deux processus gaussiens stationnaires, d'espérance nulle et de variance finie, σ^2_X et σ^2_Y . Supposons de plus que les autocovariances de X et Y : $C_X(u,v) = E\,(X(u)\,X(v))$ et $C_Y(u,v) = E\,(Y(u)\,Y(v))$, $u,v \in Z^2$, soient à décroissance exponentielle.

Soit r_{XY}, le coefficient de corrélation empirique entre $\{X(u)\}$ et $\{Y(u)\}$, $u \in D_n$. Alors :

$$\lim_{n \to \infty} |D_n|\ \mathrm{var}\,(r_{XY}) = \sum_{v \in Z^2} C_X(0,v)\, C_Y(0,v)/\sigma^2_X\, \sigma^2_Y$$

où 0 est l'origine et la suite $\{D_n\}$ peut être de forme quelconque du moment qu'asymptotiquement son bord δD_n soit négligeable par rapport à D_n. Cette condition sur la suite $\{D_n\}$ est souvent retrouvée quand les processus spatiaux stationnaires sont étudiés.

Dans le cas de processus markoviens conditionnels ou autorégressifs simultanés des plus proches voisins, une expression explicite de la variance en fonction des paramètres des processus est donnée. Ces calculs numériques illustrent bien l'influence considérable de l'autocorrélation sur la variance du coefficient de corrélation.

Ces travaux ont fait l'objet de deux publications dans le Journal of Applied Probability et la Revue de Statistique Appliquée en collaboration avec D. Hémon., (liste des travaux (3), (4)).

1.2 Normalité asymptotique de r_{XY} sous des hypothèses de mélange

Les résultats énoncés dans ce paragraphe ne sont pas publiés. On peut les considérer comme une annexe aux publications des paragraphes 1.1 et 1.3.

Dans le cadre de variables stationnaires dépendantes sur Z^d, $d \geq 1$, les théorèmes de la limite centrale (T.L.C.) sont habituellement obtenus par l'imposition de conditions de mélange qui assurent en quelque sorte une faible dépendance pour des σ-algèbres engendrées par des variables éloignées, (Ibragimov et Linnick (12) et Hall et Heyde (8) pour $d = 1$, Bolthausen (4) pour $d \geq 1$).

En particulier le champ $\{X(u)\}$ est défini comme α-mélangeant (mélange fort) si pour toutes parties A et B de Z^d, $\sup |P(E \cap F) - P(E)\ P(F)| \leq \alpha\ (d\ (A,B))$

$E \in \mathcal{F}_A, F \in \mathcal{F}_B$

avec $\alpha(m) \dashrightarrow 0$ quand $m \dashrightarrow +\infty$, $\mathcal{F}_A$ et $\mathcal{F}_B$ étant les σ-algèbres engendrées par X sur A et B et d la distance du sup.

Pour les champs gaussiens stationnaires, cette propriété de mélange fort est liée à la décroissance des covariances qui peut être facilement vérifiée quand on connait la forme de la densité spectrale. Cette propriété peut aussi être vérifiée dans le cas de champs gaussiens markoviens non nécessairement stationnaires (Kunsch (14), Guyon (7)).

Une autre notion de mélange, introduite par Dobrushin (6), fait intervenir les parties A et B par l'intermédiaire de leurs cardinaux ou d'une fonction de ceux-ci. Elle est particulièrement bien adaptée aux champs de Gibbs (Kunsch (15), Guyon (7)) pour

lesquels on ne sait pas vérifier le mélange fort. Bolthausen considère un coefficient de mélange $\alpha_2(m)$ où $|A| \leq 2$.

Une classe également intéressante de processus dépendants est celle des processus linéaires (ou moyennes mobiles infinies) qui ont souvent été étudiés pour les séries chronologiques (Anderson (1), Rosenblatt (17)). Pour ceux-ci, un T.L.C. peut être montré directement en imposant des conditions sur les coefficients de la moyenne mobile. Les processus linéaires ne sont pas nécessairement fortement mélangeants (Withers (23)).

Ces différentes notions peuvent être utilisées pour formuler des conditions entraînant la normalité asymptotique de r_{XY}. A titre d'exemple nous allons en formuler deux.

Soit $\{X(u)\}$ et $\{Y(u)\}$, deux processus stationnaires sur Z^2 centrés et de variance finie, σ^2_X et σ^2_Y.

Soit $\{D_n\}$ une suite croissante de domaines de Z^2, $U\, D_n = Z^2$ $\lim_{n \to \infty} |\delta D_n| / |D_n| = 0$.

Les conditions alternatives suivantes sont considérées :

(i) $\{X(u)\}$, $\{Y(u)\}$ ε $L_{2+\delta}$, pour un $\delta>0$, sont mutuellement indépendants et mélangeants avec coefficients de mélange $\alpha_2^X(m)$ et $\alpha_2^Y(m)$ satisfaisant

$$\sum_{m=1}^{\infty} m\, \alpha_2^X(m)^{\delta/2+\delta} < \infty \text{ et } \sum_{m=1}^{\infty} m\, \alpha_2^Y(m)^{\delta/2+\delta} < \infty \quad ,$$

(ii) $\{X(u)\}$, $\{Y(u)\}$ sont linéaires : $X(u) = \sum_{k \varepsilon Z^2} g_k\, Z(u-k)$, $Y(u) = \sum_{l \varepsilon Z^2} h_l\, W(u-l)$

avec $Z(u)$ et $W(u)$ i.i.d. de moyenne nulle et de variance finie, $\sum |g_k| < \infty$, $\sum |h_l| < \infty$ et les σ-algèbres engendrées par les $\{Z(u)\}$ et les $\{W(u)\}$ sont indépendantes.

Sous l'une ou l'autre de ces conditions, on montre alors que : $\Sigma = \sum_{v \varepsilon Z^2} C_X(0,v)\, C_Y(0,v)$

est absolument convergente et que si $\Sigma > 0$, $|D_n|^{1/2}\, r_{XY}$ converge en distribution quand $n \to \infty$ vers une distribution normale, d'espérance nulle et de variance $\Sigma / \sigma^2_X\, \sigma^2_Y$.

La démonstration sous la condition (i) repose sur la vérification des hypothèses du T.L.C de Bolthausen (4). La convergence des variances empiriques vers σ^2_X et σ^2_Y est

à nouveau utilisé. La convergence des autres termes vers 0 étant assurée par les conditions de sommabilité des coefficients $\{g_k\}$ ou $\{h_l\}$.

Pour d=1 Anderson (1) démontre un T.L.C. sous des conditions semblables à (ii). Ces conditions ont été affaiblies par la suite par Hannan (9) et Hall et Heyde (8) qui utilisent alors une approche de martingale. Cette approche ne se généralise pas naturellement à Z^d,

$d > 1$.

1.3 Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants

Il est aussi intéressant d'étudier la vitesse de convergence dans le T.L.C. pour des variables faiblement dépendantes indexées par Z^d. Ceci a été fait dans le cadre général d'un champ $\{X(u)\}$ centré, non nécessairement stationnaire, vérifiant $\sup_{u\in Z^2} \|X(u)\|_{2+\delta} < \infty$, pour un $\delta > 0$,

et d'une suite strictement croissante de domaines à laquelle sont associés les suites $(S_n),(\sigma_n^2)$:

$$S_n = \sum_{u\in D_n} X(u) \quad , \sigma_n^2 = \mathrm{Var}\ S_n \ .$$

Ni la stationnarité, ni la forme des domaines D_n n'intervient plus explicitement contrairement à la conjecture de Prakasa Rao (16). Il faut seulement supposer que quand D_n croît, on apporte toujours du "nouveau" au sens de la variance, d'où une condition :

$$\liminf \sigma_n^2 |D_n|^{-1} = \alpha > 0.$$

Soit Δ_n, la distance entre S_n renormalisée et une normale réduite de fonction de répartition Φ:

$$\Delta_n = \sup_x |P(S_n/\sigma_n \le x) - \Phi(x)| \ .$$

La technique utilisée par Tikhomirov (21) pour d = 1 a été généralisée. Elle consiste à évaluer la distance entre $f_n(t)$, la fonction caractéristique de S_n/σ_n , et $\exp(-t^2/2)$ et à utiliser ensuite l'inégalité de Berry-Esseen. Pour ce faire, d/dt $f_n(t)$ est développée, conduisant à une équation différentielle qui fait intervenir - t $f_n(t)$ et des termes de reste dont le contrôle est lié à la vitesse de convergence. Dans certains termes du reste, la dimension **d** apparaît explicitement dans les majorations qu'on effectue sur la covariance de sommes partielles,

majorations qui utilisent les inégalités classiques du mélange (cf (8)). La vitesse de convergence va donc dépendre à la fois de de la dimension **d,** de conditions sur les moments et de conditions sur le mélange.

Cas d'un champ m-dépendant (c.a.d. $\alpha(k) = 0$ pour $k \geq m$)

a) si $0 < \delta < 1$, $\Delta_n = O(\sigma_n^{-\delta})$.

b) si $\delta \geq 1$, $\Delta_n = O((\text{Log } \sigma_n)^{(d-1)/2} \sigma_n^{-1})$

Quand $\delta < 1$, la vitesse optimale est atteinte (Sergin (18)). Dans le cas de variables i.i.d., cela correspond à une vitesse en $n^{-\delta/2}$.

Quand $\delta \geq 1$ et $d = 1$, on a également la vitesse optimale en σ_n^{-1} ; mais quand $d > 1$, il y a apparition d'un terme en $(\text{Log } \sigma_n)^{(d-1)/2}$ lié à l'ordre du développement de d/dt $f_n(t)$. La question reste ouverte de savoir si cette vitesse peut être améliorée sans conditions de moments plus fortes (Heinrich (11)).

Cas d'un champ α-mélangeant, α à décroissance exponentielle

a) $\Delta_n = O((\text{Log } \sigma_n)^{d[1+\delta]} \sigma_n^{-\delta})$, $0 < \delta < 1$

$\Delta_n = O((\text{Log } \sigma_n)^{2d} \sigma_n^{-1})$, $\delta > 1$, $d \geq 2$

$\Delta_n = O((\text{Log } \sigma_n)^{5/2} \sigma_n^{-1})$, $\delta > 1$, $d = 1$

b) si de plus X est dans $L^{4+\delta}$, $\delta > 0$, alors $\Delta_n = O((\text{Log } \sigma_n)^d \sigma_n^{-1})$

La vitesse a aussi été étudiée pour un mélange à décroissance puissance.

En utilisant une autre technique et un mélange de type Dobrushin, Takahata (19) obtient une vitesse optimale pour des champs m-dépendants sous une condition L^8, tandis que pour des champs α-mélangeants à décroissance exponentielle et sous une condition $L^{8+\delta}$, $\delta > 0$, cette vitesse optimale est ralentie pour le facteur $(\text{Log } \sigma_n)^d$ comme dans le cas (b) qui impose une condition $L^{4+\delta}$, $\delta > 0$ mais un mélange plus fort.

De même que dans §1.2, les résultats formulés peuvent être appliqués en particulier à l'étude de r_{XY} sous l'hypothèse nulle d'indépendance entre X et Y.

Ces résultats ont fait l'objet d'une publication dans Z. Wahrscheinlichkeitstheorie verw. en collaboration avec X. Guyon (liste des travaux (5)).

2. Tests d'association

Plusieurs approches - qui se trouvent résumées dans les comptes rendus d'un colloque (liste des travaux (6)) - peuvent être envisagées pour tester l'association entre des variables autocorrelées. Je me suis intéressée au développement d'un test fondé sur le coefficient de corrélation empirique **r** qui, bien que paramétrique, ne nécessite pas le choix et l'estimation de modèles paramétriques spécifiques (comme dans les modèles de régression avec paramétrisation spatiale de la matrice de variance-covariance des erreurs).

Les publications présentées dans cette partie constituent des étapes successives du développement de ce test : élaboration de la méthode, étude du risque de première espèce et de la puissance, extension au test du lien conditionnel.

Du point de vue statistique, la variance asymptotique du coefficient de corrélation **r** trouvée au §1.1 ne peut-être utilisée directement pour modifier une procédure de test car elle fait intervenir les autocorrélations pour tous les décalages et ne tient pas compte de l'estimation des moyennes et des variances.

2.1 Approximation de la variance de r_{XY}

Une bonne estimation de cette variance pour un domaine A de taille finie a été développée. Cette estimation est basée sur une approximation au premier ordre de la variance dans le cas de variables gaussiennes autocorrelées, $X \sim N(\mu_X, \Sigma_X)$, $Y \sim N(\mu_Y, \Sigma_Y)$ indépendantes, définies sur un ensemble A de N sites.

Dénotons par Σ_ξ et Σ_η les matrices de variance-covariance des vecteurs d'éléments

$\xi_\alpha = X_\alpha - \bar{X}$, $\eta_\alpha = Y_\alpha - \bar{Y}$, $\alpha \in A$ où $\bar{X} = \sum_{\alpha \in A} X_\alpha / N$, et $\bar{Y} = \sum_{\alpha \in A} Y_\alpha / N$.

Soit r_{XY}, le coefficient de corrélation empirique, $r_{XY} = s_{XY}/s_X s_Y$, $s_{XY} = N^{-1} \Sigma(X_\alpha - \bar{X})(Y_\alpha - \bar{Y})$, $s_X^2 = N^{-1} \Sigma(X_\alpha - \bar{X})^2$ et σ^2_r sa variance.

Au premier ordre :

$$\sigma^2_r = \frac{tr\,(\Sigma_\xi\,\Sigma_\eta)}{tr\,(\Sigma_\xi)\,tr\,(\Sigma_\eta)} \qquad (1)$$

Dans le cas particulier où : (i) Σ_ξ et Σ_η commutent, (ii) Σ_η est proportionelle à une matrice idempotente, (iii) les valeurs propres de Σ_ξ sont nulles si les valeurs propres associées de Σ_η sont aussi nulles, l'approximation (1) est exacte. En effet dans ce cas, r_{XY} a pour densité $f_M(r) = (1-r^2)^{1/2(M-4)} / B(1/2, 1/2(M-2))$, $|r| \leq 1$ où B est la fonction beta et M=1+ rang (Σ_η).

2.2 Tests d'association modifiés

L'approximation (1) conduit à un estimateur de σ^2_r sous l'hypothèse qu'une partition $\{S_k\}$ k=1,...K de AxA puisse être définie de telle sorte que la covariance de X et de Y soit constante dans chaque élément S_k. Quand les points de A sont répartis irrégulièrement dans le plan, les strates peuvent être indexées par une fonction discrétisée de la distance.

En définissant $\hat{C}_X(k) = \sum_{S_k} (X_\alpha - \bar{X})(X_\beta - \bar{X})/N_k$

où $N_k = |S_k|$, on est conduit à l'estimateur de σ^2_r :

$$\hat{\sigma}^2_r = \frac{\sum_k N_k \hat{C}_X(k) \hat{C}_Y(k)}{N^2 s^2_X s^2_Y} \quad . \qquad (2)$$

a) Tests modifiés

Des tests d'association modifiés pour tenir compte de l'autocorrélation peuvent alors être proposés en considérant soit une covariance renormalisée :

$W = N s_{XY}(\Sigma N_k \hat{C}_X(k) \hat{C}_Y(k))^{-1/2}$, que l'on teste suivant une N(0,1) en se basant sur un T.L.C, soit un test modifié de r_{XY} : $t_{\hat{M}-2}$, fondé sur une modification, $\hat{M}$, des degrés de liberté du test sur la corrélation :

$$\hat{M} = \hat{\sigma}^{-2}_r + 1 \quad .$$

Précisément r_{XY} est testé suivant la densité $f_{\hat{M}}(r)$ définie au §2.1.

b) Risques de première espèce

Une première étude par simulations portant à la fois sur des processus à valeurs dans un lattice ou dans un réseau irrégulier montre que le risque d'erreur α de ces tests modifiés reste alors proche de la valeur nominale. Dans cette étude, les performances des statistiques W et $t_{\hat{M}-2}$ sont équivalentes.

Ces résultats ont fait l'objet d'un article écrit en collaboration avec P. Clifford et D. Hémon, paru dans Biometrics (liste des travaux (7)).

Comme les statistiques W et $t_{\hat{M}-2}$ ne sont pas fondées sur les mêmes approximations des distributions, les performances des tests W et $t_{\hat{M}-2}$ ont ensuite été comparées pour des échantillons de faible taille. On constate alors une supériorité du test $t_{\hat{M}-2}$ sur W.

Il a d'autre part été vérifié que les risques de première espèce restaient stables si différentes partitions isotropes $\{S_k\}$ de AxA étaient définies. Ce risque ne s'accroît que si trop peu de strates sont considérées.

Dans cette étude par simulations il a aussi été montré que le risque de première espèce du test non paramétrique d'association proposé par Tjøstheim (22) n'était pas bien contrôlé et augmentait avec une autocorrélation positive.

2.3 Puissance des test modifiés

La puissance des tests modifiés est étudiée sous une hypothèse alternative de modèle linéaire entre X et Y : $H_1 : Y = aX + W$, $X \sim N(\mu_X, \Sigma_X)$, $W \sim N(\mu_W, \Sigma_W)$ et X et W indépendantes. Soit Σ_θ la matrice de variance-covariance du vecteur d'éléments $W_\alpha - \bar{W}$, $\alpha \in A$.

Il est difficile de calculer la puissance des tests modifiés car l'on ne connaît pas exactement leurs lois sous H_1. Cette puissance peut être évaluée par simulations. Par ailleurs il peut être intéressant de calculer la puissance $\pi_T(s_{XY})$ d'un test de la covariance s_{XY} où l'estimateur $N^{-2} \sum N_k \hat{C}_X(k) \hat{C}_Y(k)$ de sa variance est remplacé par son évaluation théorique sous H1 :

$$N^{-2} \text{ tr } (\Sigma_\xi \Sigma_\eta) = N^{-2} [a^2 \text{ tr}(\Sigma_\xi^2) + \text{tr}(\Sigma_\xi \Sigma_\theta)].$$

Pour pouvoir comparer les puissances observées par simulations à une référence, la puissance $\pi_{N*}(r)$ du test standard de la corrélation basé sur un nombre N* de points compatible avec la variance empirique observée de r_{XY} a aussi été calculée. Des tableaux indiquant $\pi_T(s_{XY})$, $\pi_{N*}(r)$ et les puissances observées de W et $t_{\hat{M}-2}$ par simulations sur un domaine irrégulier sont donnés dans (8).

On constate que dans l'ensemble toutes les puissances calculées ou observées sont proches. La différence ne devient notable que dans quelques cas de fortes autocorrélations pour l'un ou l'autre des processus.

Ainsi on peut dire que la puissance "théorique" $\pi_T(s_{XY})$ donne une bonne indication dans la plupart des cas rencontrés. Les tests modifiés ne présentent pas de perte de puissance notable par rapport à un test classique fondé sur un nombre d'observations N* "équivalent". *L'étude de la puissance par simulation et de certains aspects complémentaires ont fait l'objet d'une communication écrite (liste des travaux (8)).*

2.4 Tests modifiés des corrélations partielles

Il est souvent intéressant de pouvoir tester le lien conditionnel entre deux variables après conditionnement sur une ou plusieurs autres variables, ce qui revient à tester la corrélation partielle. Les méthodes développées précédemment s'étendent aisément au test de la corrélation partielle si l'on considère un cadre gaussien et les distributions conditionnelles appropriées.

En pratique les tests modifiés W et $t_{\hat{M}-2}$ seront calculés sur les résidus, obtenus par les moindres carrés, des régressions linéaires sur les variables de conditionnement. Une étude par simulations a permis de vérifier la bonne performance de cette méthode.
Ces résultats vont paraître dans Statistics in Medecine (liste des travaux (9)).

3. Applications à l'étude des corrélations écologiques en épidémiologie

Les études où les variables considérées sont moyennées sur des groupes (par exemple les taux de mortalité par département) sont appelées des études de corrélations écologiques. Ces études sont sujettes à différentes sources de biais qu'il faut prendre en compte pour pouvoir en donner une interprétation valable.

Dans un premier temps, je me suis attachée à discuter ces biais quantitativement en comparant la forme des relations dose-effet individuelles ou agrégées et l'effet du niveau d'agrégation. Dans une deuxième partie, j'ai comparé sur des exemples les risques relatifs estimés par des études écologiques et des études individuelles. On peut montrer que les relations dose-effet données par les études écologiques et les études individuelles sont d'autant plus proches que le facteur de risque étudié est prépondérant dans l'ensemble des facteurs de risque relevants, que les fractions de populations concernées sont bien identifiées et que la relation dose-effet individuelle est proche de la linéarité.

Ces réflexions ont fait l'objet d'une publication dans l'International Journal of Epidemiology en collaboration avec I. Stücker et D. Hémon (liste des travaux (10)).

D'autre part, j'ai appliqué à différents problèmes épidémiologiques les méthodes proposées. Les résultats trouvés sont en bon accord avec certains facteurs de risque isolés dans des enquêtes individuelles et illustrent l'intérêt des méthodes statistiques proposées.

Ces résultats sont présentés dans la dernière partie de l'article sur le test ajusté de la corrélation partielle (liste des travaux (9)).

4. Conclusion

Dans cette deuxième partie, j'ai résumé un ensemble de travaux centrés sur l'étude de tests d'association entre des variables spatialement dépendantes. Ceci m'a conduit d'une part à m'intéresser à certaines techniques probabilistes, utiles dans le cadre de variables faiblement dépendantes pour étudier la consistance ou la convergence vers la normalité de certaines statistiques. D'autre part, j'ai été amenée à proposer un test d'association modifié fondé sur le coefficient de corrélation empirique. Le coefficient de corrélation est une mesure d'association qui, bien que paramétrique, ne nécessite pas l'identification et l'estimation de modèles particuliers. Le risque de première espèce et la puissance des tests modifiés ont été étudiés, ainsi que leur généralisation au test du lien conditionnel. J'ai montré l'intérêt d'utiliser ces test dans des applications épidémiologiques sur les corrélations écologiques.

Il serait intéressant de développer d'autres mesures d'association, en particulier non paramétriques. Un index d'association non paramétrique a été proposé par Tjøstheim sans que le risque d'erreur de la méthode proposée n'ait été bien contrôlé. Une approche fondée sur les permutations serait possible si on prenait soin de faire une inférence conditionnelle à la préservation d'une structure spatiale.

Références de la première partie

(1) Chacon, R.V. A geometric construction of measure preserving transformations, Proc. Fifth Berkeley Sym. Math. Stat. Prob., University of California Press, 1967, Vol II, part 2, 335-360.

(2) Friedman, N.A. Introduction to Ergodic Theory. New York : van Nostrand 1970

(3) Friedman, N.A. Bernoulli shifts induce Bernoulli shifts. Advances in Math. 10, 39-48 (1970)

(4) Friedman, N.A., Ornstein, D.S. Ergodic transformations induce mixing transformations. Advances in Math. 10, 147-163 (1973)

(5) Halmos, P.R. Lectures on Ergodic Theory. New-York - Chelsea 1956.

(6) Neveu J. Temps d'arrêt d'un système dynamique. Z. Wahrscheinlichkeitstheorie verw. Gebiete 13, 81-94 (1969).

(7) Ornstein, D.S., Smorodinski, M. Ergodic flows of positive entropy can be time changed to become K.flows. Israel J. Math. 26, 75-83 (1977).

(8) Shields, P. Cutting and independent stacking of intervals. Math. Systems Theory 7, 1-4 (1973).

(9) Walters, P. An Introduction to Ergodic Theory. Graduate Text in Mathematics. Vol 79 New-York, Springer Verlag 1981.

Références de la deuxième partie

(1) Anderson, T.W. The statistical analysis of time series. New York : Wiley (1971).

(2) Bartlett, M.S. Some aspects of the time-correlation problem in regard to tests of significance. Journal of the Royal Statistical Society, 98, 536-543 (1935).

(3) Bartlett, M.S. An Introduction to Stochastic Processes, 2nd edition Cambridge University Press 1966.

(4) Bolthausen, E. On the central limit theorem for stationary random fields. Ann. Probability 10, 1047-1050 (1982).

(5) Cliff, A.D. and Ord, J.K. Model building and the analysis of spatial pattern in human geography. Journal of the Royal Statistical Society B, 37, 297-348 (1975).

(6) Dobrushin, R.L. The description of a random field by its conditional distribution and its regularity condition. Theory Probability Appl. 13, 197-227 (1968).

(7) Guyon, X. Estimation par pseudo-vraisemblance conditionnelle pour les champs : étude asymptotique et application aux champs markoviens.
Proceedings of the 6th Franco-Belgian Meeting of Statisticians, November 1985.
Edited by F. Droesbeke. Travaux et Recherches Vol 11, p. 15-62.
Publications des Facultés universitaires St Louis, Bruxelles.

(8) Hall, P., Heyde, C.C. Martingale limit theory and its application. New York : Academic Press 1980.

(9) Hannan, E.J. Multiple Times Series.
New York : Wiley 1970.

(10) Haugh, L.D. and Box, G.E.P. Identification of dynamic regression (distributed lag) models connecting two time series.
J. Amer. Statist. Assoc. 72, 121-130 (1977).

(11) Heinrich, L. Asymptotic expansions in the Central Limit Theorem for a special class of m-dependent random fields.
Mathematische Nachrichten 83-106 (1987).

(12) Ibragimov, I.A., Linnik, Yu.V. Independent and stationary sequences of random variables. Groningen : Wolters-Nordhoff 1971.

(13) Johnston, J. Econometric Methods, 2nd edition.
New York : McGrawHill 1972.

(14) Künsch, H.R. Reelwertige Zufallsfelder anf einem Gitter : Interpolation, variations sprinzip und Stat. Analyse, Thèse E.T.H. de Zurich, 1980.

(15) Künsch, H.R. Decay of correlations under Dobrushin's uniqueness condition and its applications.
Comm. Math. Phys. 84, 207-222 (1982).

(16) Prakasa Rao, B.L. A non uniform estimates of the rate of convergence in the central limit theorem for m-dependent random fields.
Z. Wahrscheinlichkeitstheorie verw. Gebiete 58, 247-256 (1981).

(17) Rosenblatt, M. Stationary sequences and random fields.
Boston : Birkhäuser 1985.

(18) Shergin, V.V. An estimate of the remainder term in the central limit theorem for m-dependent random variables.
Theory Probability Appl. 24, 782-796 (1979).

(19) Takahata, H. On the rates in the central limit theorem for weakly dependent random fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete 64, 445-456 (1983).

(20) Tempelman, A.A. Ergodic theorems for dynamical systems.
Trans. Moscow Math. Soc. 26, 94-132 (1972).

(21) Tikhomirov, A.N. On the converge rate in the centra limit theorem for weakly dependent random variables. Theory Probability Appl. 25, 790-809 (1980).

(22) Tjøstheim, D. A measure of association for spatial variables.
Biometrika 65, 109-114 (1978).

(23) Withers, C.S. Conditions for linear processes to be strong mixing.
Z. Wahrscheinlichkeitstheorie verw. Gebiete 57, 477-480 (1981).

LISTE DES TRAVAUX

(1) S.T. Richardson, K.M. Wilkinson
Stopping time transformations and towers.
Z. Wahrscheinlichkeitstheorie verw. Gebiete 48. 259-284, (1979)

(2) S.T. Richardson
Stopping times for measure-preserving transformations : some approximation results.
Proceedings of the London Mathematical Society (3) 43 : 273-294 (1981).

(3) S.T. Richardson, D. Hémon
On the variance of the sample correlation between two independent lattice processes.
J. Appl. Prob. 18, 943-948 (1981).

(4) S.T. Richardson, D. Hémon
Autocorrélation spatiale : ses conséquences sur la corrélation empirique de deux processus spatiaux.
Revue de Statistique Appliquée, 30 : 41-51 (1982).

(5) X. Guyon, S.T. Richardson
Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants.
Z. Wahrscheinlichkeitstheorie verw. Gebiete, 66 : 297-314 (1984).

(6) S.T. Richardson
Testing the association between two spatial processes : a review of different approaches
Spatial processes and spatial time series analysis. Proceedings of the 6th Franco-Belgian Meeting of Statisticians, November 1985.
Edited by F. Droesbeke. Travaux et Recherches Vol 11, p. 193-200.
Publications des Facultés universitaires St Louis, Bruxelles.

(7) P. Clifford, S.T. Richardson, D. Hémon
Assessing the significance of the correlation between two spatial processes.
Biometrics, 45 : 123-134 (1989).

(8) S.T. Richardson, P. Clifford
Testing association between spatial processes.
A paraître dans Proceedings of the A.M.S conference on Spatial Statistics and Imaging (June 1988).

(9) S.T. Richardson
A method for testing the significance of geographical correlations with application to industrial lung cancer in France.
A paraître dans Statistics in Medecine.

(10) S.T. Richardson, I. Stücker, D. Hémon
Comparison of relative risks obtained in ecological and individual studies : some methodological considerations.
International Journal of Epidemiology, Vol 16, n°1, 111-120 (1987).

PREMIERE PARTIE

PROCESSUS ERGODIQUES : ETUDES DES PROPRIETES ERGODIQUES DES TEMPS D'ARRET

STOPPING TIME TRANSFORMATIONS AND TOWERS

Z. Wahrscheinlichkeitstheorie verw. Gebiete
48, 259–284 (1979)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Stopping Time Transformations and Towers

S.T. Richardson[1] and K.M. Wilkinson[2]

[1] Department of Statistics, University of Warwick, Coventry CV4 7AL, England
[2] Department of Mathematics, The University of Nottingham, University Park, Nottingham NG7 2RD England

§ 1. Introduction

In this paper we shall study properties of τ^ν where τ is an automorphism of a Lebesgue space and ν is a non-negative integer valued measurable function. We shall always assume that ν satisfies the conditions, given by Neveu [4] and discussed in § 4, which ensure that τ^ν is again an automorphism. Such a ν is called a stopping time and τ^ν a stopping time transformation. The set of all stopping time transformations derived from a particular automorphism τ is equivalent to the positive part of the full group of τ.

We shall restrict attention to that class of automorphisms which are defined by means of a tower construction. This class contains many interesting transformations including Bernoulli shifts of any prescribed entropy. For an automorphism in this class, using a theorem of Neveu, we can follow the action of τ^ν in the tower and this leads to an analysis of τ^ν by means of a more general contruction which we call a tower block construction.

This new construction appears to be of interest in its own right and so because of this and also for reasons of clarity we shall commence by introducing the construction and discuss properties of the transformation so defined. In §§ 4 and 5 we show that in general τ^ν, where τ is defined by means of a tower construction, can be represented by a tower block construction. We are then able to use the results of § 3 to obtain density results for τ^ν, our main results being firstly that the set of τ^ν which are ergodic and secondly the set of τ^ν which are Bernoulli are both dense in the set of stopping time transformations derived from the same tower transformation τ. As a consequence of this last result we have that if τ is Bernoulli the set of τ^ν which are also Bernoulli is dense, thus generalising a result of Friedman for the induced transformation [2]. Saleski [6] has also investigated properties of a stopping time transformation where τ is Bernoulli, particularly in the case where ν has expected value 2.

Throughout the paper we take $\Omega = [0, 1]$, $\mathcal{B}$ the set of Borel subsets of Ω and μ Lebesgue measure on $\mathcal{B}$.

0044-3719/79/0048/0259/$05.20

§ 2. Tower Block Construction

Our initial definitions follow Friedman ([1], § 6 and [2]). A column C is an ordered set of disjoint left-closed, right-open subintervals of Ω.

$$C=(L_i\colon 1\leqq i\leqq h).$$

The sets L_i, $1\leqq i\leqq h$, are referred to as column levels and they are assumed to each have the same measure. C has base $B(C)=L_1$, top $A(C)=L_h$, width $w(C)=\mu(L_1)$ and height $h(C)=h$. We let

$$C^*=\bigcup_{i=1}^{h} L_i$$

and define columns C_1, C_2 to be disjoint if C_1^* and C_2^* are disjoint sets. We denote by τ_C the one-to-one measure preserving mapping from C^*-L_h to C^*-L_1 which maps L_i linearly onto L_{i+1}, $1\leqq i\leqq h-1$.

A tower T is an ordered set of disjoint columns.

$$T=(C_i\colon 1\leqq i\leqq m).$$

T has base $B(T)=\bigcup_{i=1}^{m} B(C_i)$, top $A(T)=\bigcup_{i=1}^{m} A(C_i)$ and width $w(T)=\sum_{i=1}^{m} w(C_i)=\mu(B(T))$. We let

$$T^*=\bigcup_{i=1}^{m} C_i^*$$

and define towers T_1, T_2 to be disjoint if T_1^* and T_2^* are disjoint sets. τ_T is the mapping from $T^*-A(T)$ to $T^*-B(T)$ consisting of the mappings τ_{C_i} in each column C_i, $1\leqq i\leqq m$.

If $C_1=(L_{1i}\colon 1\leqq i\leqq h_1)$ and $C_2=(L_{2i}\colon 1\leqq i\leqq h_2)$ are two disjoint columns with $w(C_1)=w(C_2)$ we define

$$C_1*C_2=(L_{11}, L_{12}, \ldots, L_{1h_1}, L_{21}, L_{22}, \ldots, L_{2h_2}),$$

so C_1*C_2 is the column obtained by placing C_2 above C_1.

If $T_1=(C_{1i}\colon 1\leqq i\leqq m_1)$ and $T_2=(C_{2i}\colon 1\leqq i\leqq m_2)$ are disjoint towers we define

$$T_1+T_2=(C_{11}, C_{12}, \ldots, C_{1m_1}, C_{21}, C_{22}, \ldots, C_{2m_2}),$$

so T_1+T_2 is the tower obtained by placing T_2 at the side of T_1.

Suppose $\alpha_i\geqq 0$, $0\leqq i\leqq r$, and $\sum_{i=0}^{r}\alpha_i=1$. If $L=[a, b)$ we define

$$\alpha_j L=\left[a+(b-a)\sum_{i=0}^{j-1}\alpha_i,\ a+(b-a)\sum_{i=0}^{j}\alpha_i\right), \qquad 0\leqq j\leqq r.$$

Then for a column $C=(L_i\colon 1\leqq i\leqq h)$ we define

$$\alpha_j C=(\alpha_j L_i\colon 1\leqq i\leqq h), \qquad 0\leqq j\leqq r.$$

and for a tower $T=(C_i\colon 1\leqq i\leqq m)$ we define

$$\alpha_j T=(\alpha_j C_i\colon 1\leqq i\leqq m),\qquad 0\leqq j\leqq r.$$

Note that $w(\alpha_j C)=\alpha_j w(C)$ and $w(\alpha_j T)=\alpha_j w(T)$.

If C is a column and $T=(C_i\colon 1\leqq i\leqq m)$ is a tower such that C^* and T^* are disjoint and $w(C)=w(T)$ we define a tower $C*T$ by

$$C*T=(\alpha_{i-1}C*C_i,\ 1\leqq i\leqq m)$$

where $\alpha_{i-1}=w(C_i)/w(C)$, $1\leqq i\leqq m$. So $C*T$ is obtained by placing the tower T above C. If $T_1=(C_i\colon 1\leqq i\leqq m)$ and T_2 are disjoint towers with $w(T_1)=w(T_2)$ we define a tower T_1*T_2 by

$$T_1*T_2=\sum_{j=1}^{m} C_j*\alpha_{j-1}T_2$$

where $\alpha_{j-1}=w(C_j)/w(T_2)$, $1\leqq j\leqq m$. So T_1*T_2 is obtained by placing a copy of T_2 above each column in T_1.

If $\alpha_0=\alpha_1=1/2$ we adopt the notation

$$(\tfrac{1}{2}L)_i=\alpha_i L,\qquad (\tfrac{1}{2}C)_i=\alpha_i C,\qquad (\tfrac{1}{2}T)_i=\alpha_i T$$

$i=0, 1$, where L, C, T are, respectively, a column level, a column and a tower. For any tower T we define a tower $S(T)$ by

$$S(T)=(\tfrac{1}{2}T)_0*(\tfrac{1}{2}T)_1.$$

So $S(T)$ consists of a copy of T in the base and a further copy of T above each column in the base copy. The base copy of T in $S(T)$ will be referred to as the copy of rank 0 and, if T has m columns, there are m copies of T of rank 1 in $S(T)$, one above each column in the rank 0 copy.

Letting $S^0(T)=T$ we define $S^u(T)$, $u\geqq 1$, inductively by

$$S^{u+1}(T)=S(S^u(T)).$$

$S^u(T)$, $u\geqq 0$, consists of one rank 0 copy of T, m rank 1 copies of T, m^2 rank 2 copies of T, ..., and m^{2^u-1} copies of T of rank 2^u-1 and each copy of T of rank i, $1\leqq i\leqq 2^u-1$ is immediately above a column in a rank $i-1$ copy of T. $S^u(T)$, $u\geqq 0$, also contains m^{2^u} columns.

If the ith column in a tower T has the form

$$\alpha_{i_1}^1 c_1*\alpha_{i_2}^2 c_2*\ldots*\alpha_{i_l}^l c_l$$

where $\alpha_s^j\geqq 0$, $\sum_{s=0}^{r_j}\alpha_s^j=1$, $0\leqq i_j\leqq r_j$, $1\leqq j\leqq l$, and c_j, $1\leqq j\leqq l$, are columns, it will sometimes be unimportant what the values of $\alpha_{i_j}^j$, $1\leqq j\leqq l$, are. In such a case we write

$$C_i(T)=[c_1, c_2, \ldots, c_l]$$

and take this to mean that the ith column of T consists of a copy of c_1 with a copy of c_2 above it, etc.

In particular, if $T=(C_i,\ 1\leqq i\leqq m)$, then the m^2 columns of $S(T)$ are given by

$$C_{(i_0-1)m+i_1}(S(T))=[C_{i_0}, C_{i_1}], \qquad 1\leqq i_j\leqq m,\ j=0, 1,$$

and more generally, for $u\geqq 1$, the m^{2^u} columns of $S^u(T)$ are given by

$$C_{(i_0-1)m^{2^u-1}+(i_1-1)m^{2^u-2}+\ldots+(i_{2^u-2}-1)m+i_{2^u-1}}(S^u(T))$$
$$=[C_{i_0}, C_{i_1}, \ldots, C_{i_{2^u-1}}], \qquad 1\leqq i_j\leqq m,\ 0\leqq j\leqq 2^u-1.$$

Note that $\tau_{S(T)}$ agrees with τ_T on $T^*-A(T)$ and that, in addition, $\tau_{S(T)}$ is defined on half of $A(T)$. In fact the set on which $\tau_{S(T)}$ is undefined is $A(S(T))$ which has measure $\frac{1}{2}w(T)$. Similarly $\tau_{S^u(T)}$, $u\geqq 2$, agrees with $\tau_{S^v(T)}$ on $T^*-A(S^v(T))$, $0\leqq v\leqq u-1$ and

$$\mu(A(S^u(T)))=\frac{1}{2^u}w(T).$$

Hence we may define a mapping $\tau(T)$: $T^*\to T^*$ by

$$\tau(T)=\lim_{u\to\infty}\tau_{S^u(T)}.$$

$\tau(T)$ is invertible, preserves Lebesgue measure and is ergodic ([1]).

We now wish to define a more general construction for which the resulting transformation need not be ergodic. We shall use this construction to investigate properties of stopping time transformations. We start with an ordered set of towers

$$\mathrm{I\!I}=(T_i:\ 1\leqq i\leqq k)$$

each with m columns,

$$T_i=(C_{ij}:\ 1\leqq j\leqq m)$$

and $w(C_{i_1,j})=w(C_{i_2,j})$, $1\leqq i_1, i_2\leqq k$, $1\leqq j\leqq m$. $\mathrm{I\!I}$ is called a tower block and has base

$$B(\mathrm{I\!I})=\bigcup_{i=1}^{k}B(T_i), \qquad \text{top } A(\mathrm{I\!I})=\bigcup_{i=1}^{k}A(T_i)$$

and width $w(\mathrm{I\!I})=\mu(B(\mathrm{I\!I}))=kw(T_1)$. We also put

$$\mathrm{I\!I}^*=\bigcup_{i=1}^{k}T_i^*.$$

Let $\Sigma=(\sigma_i:\ 1\leqq i\leqq m)$ where each σ_i, $1\leqq i\leqq m$, is a permutation of $\{1, 2, \ldots, k\}$ and define $P(\Sigma, \mathrm{I\!I})$ to be the tower block $(P(\Sigma, T_i):\ 1\leqq i\leqq k)$ where

$$P(\Sigma, T_i)=\sum_{j=1}^{m}(\tfrac{1}{2}C_{ij})_0 * \chi_{j-1}(\tfrac{1}{2}T_{\sigma_i(i)})_1, \qquad 1\leqq i\leqq k,$$

and

$$\alpha_{j-1} = w(C_{ij})/w(T_i), \qquad 1 \leqq j \leqq m.$$

So $P(\Sigma, T_i)$ has a copy of T_i in the base and then a copy of $T_{\sigma_j(i)}$ above the jth column of T_i. In fact

$$C_{(i_0-1)m+i_1}(P(\Sigma, T_i)) = [C_{i,i_0}, C_{\sigma_{i_0}(i),i_1}], \qquad 1 \leqq i_j \leqq m,\ j = 0, 1.$$

We shall let $\tau_{\mathrm{I\!I}}$ be the transformation defined on $\mathrm{I\!I}^* - A(\mathrm{I\!I})$ which is τ_{T_i} on each $T_i^* - A(T_i)$, $1 \leqq i \leqq k$. Clearly $\tau_{P(\Sigma, \mathrm{I\!I})}$ agrees with $\tau_{\mathrm{I\!I}}$ on $\mathrm{I\!I}^* - A(\mathrm{I\!I})$ and is also defined on half of $\mathrm{I\!I}^* - A(\mathrm{I\!I})$.

We now wish to extend the definition of $\tau_{\mathrm{I\!I}}$ by defining an inductive tower block construction. Let $\mathrm{I\!I}$ and Σ be as above. For each $u \geqq 0$ we define an ordered collection of permutations of $\{1, 2, \ldots, k\}$,

$$\Sigma_u = \{\sigma_i^u : 1 \leqq i \leqq m^{2^u}\},$$

by $\Sigma_0 = \Sigma$ (i.e. $\sigma_i^0 = \sigma_i$, $1 \leqq i \leqq m$) and for $u \geqq 0$, $1 \leqq j \leqq m^{2^u}$, $1 \leqq l \leqq m^{2^u}$,

$$\sigma^{u+1}_{(j-1)m^{2^u}+l}(i) = \sigma_l^u(\sigma_j^u(i)). \tag{2.1}$$

for each $1 \leqq i \leqq k$. We then define $P^0(T_i) = T_i$, $1 \leqq i \leqq k$, and for each $u \geqq 0$ we inductively define

$$P^{u+1}(T_i) = P(\Sigma_u, P^u(T_i)), \qquad 1 \leqq i \leqq k,$$

and define $P^u(\mathrm{I\!I})$, $u \geqq 0$, to be the tower block

$$(P^u(T_i) : 1 \leqq i \leqq k).$$

Note that $P^u(T_i)$, $1 \leqq i \leqq k$, $u \geqq 0$, consists of a rank 0 copy of T_i, above each of the m columns of T_i there is a copy of one of the towers T_i, $1 \leqq i \leqq k$, above each of the m^2 columns in these rank 1 copies there is a copy of one of the towers T_i, $1 \leqq i \leqq k, \ldots$ and at the top of $P^u(T_i)$ there are m^{2^u-1} copies of some of the towers T_i, $1 \leqq i \leqq k$, of rank $2^u - 1$, one above each column in the copies of rank $2^u - 2$. Because of (2.1) for each $u \geqq 0$, $1 \leqq i \leqq k$,

$$\begin{aligned} &C_{(i_0-1)m^{2^u-1}+(i_1-1)m^{2^u-2}+\cdots+(i_{2^u-2}-1)m+i_{2^u-1}} P^u(T_i) \\ &\quad = [C_{ii_0}, C_{\sigma_{i_0}(i),i_1}, C_{\sigma_{i_1}(\sigma_{i_0}(i)),i_2}, \ldots, (\sigma_{i_{2^u-3}}(\ldots(\sigma_{i_1}(\sigma_{i_0}(i)))\ldots)), i_{2^u-1} \\ &\qquad C_{\sigma_{i_{2^u-2}}}], \\ &\qquad 1 \leqq i_j \leqq m, \ 0 \leqq j \leqq 2^u - 1. \end{aligned} \tag{2.2}$$

So (2.1) ensures that above every copy of the column C_{ij} of rank $0, 1, \ldots, 2^u - 2$ in $P^u(\mathrm{I\!I})$, $u \geqq 1$, we have a copy of the tower $T_{\sigma_j(i)}$. Note also that in $P^u(\mathrm{I\!I})$, $u \geqq 0$, there are m^l copies of each T_i, $1 \leqq i \leqq k$, of rank l, $1 \leqq l \leqq 2^u - 1$ distributed among the various $P^u(T_i)$, $1 \leqq i \leqq k$.

As $\tau_{P^u(\mathrm{I\!I})}$, $u \geqq 1$, agrees with each of $\tau_{P^v(\mathrm{I\!I})}$ on $\mathrm{I\!I}^* - A(P^v(\mathrm{I\!I}))$, $0 \leqq v \leqq u-1$, and as

$$\mu(A(P^u(\mathrm{I\!I}))) = \frac{1}{2^u} w(\mathrm{I\!I}), \qquad u \geqq 0,$$

we can define a mapping $\tau(\Sigma, \mathbb{T}): \mathbb{T}^* \to \mathbb{T}^*$ by

$$\tau(\Sigma, \mathbb{T}) = \lim_{u \to \infty} \tau_{P^u(\mathbb{T})}.$$

So $\tau(\Sigma, \mathbb{T})$ maps a column level in $\mathbb{T}$ which is a subset of $\mathbb{T}^* - A(\mathbb{T})$ linearly onto the column level immediately above and maps $A(C_{ij})$, $1 \leqq j \leqq k$, $1 \leqq j \leqq m$, onto $B(T_{\sigma_j(i)})$.

The above construction in the case $k = 1$ yields

$$P^u(\mathbb{T}) = S^u(T_1), \quad u \geqq 0,$$

where $S^u(T_1)$, $u \geqq 0$, is as defined in Friedman's construction. Note that $\tau(\Sigma, \mathbb{T})$ is invertible and preserves Lebesgue measure. As an example of $\tau(\Sigma, \mathbb{T})$ which is not ergodic take $k \geqq 2$ and each σ_i, $1 \leqq i \leqq m$, to be the identity permutation. In this case

$$P^u(T_i) = S^u(T_i)$$

for each $u \geqq 0$, $1 \leqq i \leqq k$ and so $\tau(\Sigma, \mathbb{T})$ leaves each set T_i^*, $1 \leqq i \leqq k$, invariant.

§ 3. Mixing Properties of $\tau(\Sigma, \mathbb{T})$

Throughout this section we take $\tau = \tau(\Sigma, \mathbb{T})$ where Σ and $\mathbb{T}$ are as defined in § 2. We shall first show that τ is a Markov shift, thus generalising a result of Friedman [2] and Shields [7].

Let $\mathscr{L}$ be the partition of $\mathbb{T}^*$ consisting of the column levels in $\mathbb{T}$ and let $\mathscr{L}_n = \bigvee_{i=0}^{n-1} \tau^i \mathscr{L}$. If $L \in \mathscr{L}$, $n \geqq 1$, we shall let

$$\mathscr{L}_n(L) = \{M \in \mathscr{L}_n : M \subset L\}.$$

A partition $\mathscr{L}$ is a Markov partition for a transformation τ with invariant measure μ if we can find $p(L, M) \geqq 0$, $L, M \in \mathscr{L}$, satisfying

(3.1) For each $M \in \mathscr{L}$, $\mu(M) = \sum_{L \in \mathscr{L}} \mu(L)\, p(L, M)$.

(3.2) For each $L \in \mathscr{L}$, $\sum_{M \in \mathscr{L}} p(L, M) = 1$, and

(3.3) For each $D \in \mathscr{L}_n(L)$, $L \in \mathscr{L}$ and $M \in \mathscr{L}$,

$$\mu(\tau D \cap M) = p(L, M)\, \mu(D).$$

We first show that an atom of $\mathscr{L}_n$, $n \geqq 1$, can be approximated by the union of certain column levels in $P^u(\mathbb{T})$, $u \geqq 0$.

Lemma 3.1. *Let $L \in \mathscr{L}$ and $M \in \mathscr{L}_n(L)$, $n \geqq 1$. Then for each $u \geqq 0$ we can write*

$$M = D_u \cup E_u$$

where D_u is a union of copies of L in $P^u(\mathbb{T})$ and $E_u \subset \bigcup_{j=0}^{n-1} \tau^j(B(P^u(\mathbb{T})))$.

Proof. Let $\mathscr{L}=(L_i\colon 1\leqq i\leqq s)$. Each column level in $P^u(\mathrm{I\!I})$ is a subset of an atom of $\mathscr{L}$ and if F is a column level of $P^u(\mathrm{I\!I})$ we write $l(F)=i$ if $F\subset L_i$, $1\leqq i\leqq s$, i.e. $l(F)=i$ if and only if F is a copy of L_i. Now for each i, $1\leqq i\leqq s$, there are $2^u m^{2^u-1}$ column levels F of $P^u(\mathrm{I\!I})$ with $l(F)=i$. Let

$$M=L_{i_1}\cap\tau L_{i_2}\cap\ldots\cap\tau^{n-1}L_{i_n}$$

and let F_j, $1\leqq j\leqq 2^u m^{2^u-1}$, be the copies of L_{i_1} in $P^u(\mathrm{I\!I})$. Define J_1 to be the set of those j, $1\leqq j\leqq 2^u m^{2^u-1}$, for which the column of $P^u(\mathrm{I\!I})$ containing F_j has no more than $n-2$ column levels beneath F_j. For $j\in J_1^c$ we let $F_{j,1}=F_j$ and for $1\leqq p\leqq n-1$ we let $F_{j,p+1}$ denote the column level immediately below $F_{j,p}$. Define

$$J_2=\{j\in J_1^c\colon l(F_{j,p})=i_p,\ 1\leqq p\leqq n\}$$

and

$$J_3=\{j\in J_1^c\colon l(F_{j,p})\neq i_p \text{ for at least one } p,\ 1\leqq p\leqq n\}$$

and let

$$D_u=\bigcup_{j\in J_2}F_j,\qquad E_u=M-D_u.$$

As $E_u\subset\bigcup_{j\in J_1}F_j\subset\bigcup_{i=0}^{n-1}\tau^i B(P^u(\mathrm{I\!I}))$, we have the result.

Corollary 3.2. *Given $\varepsilon>0$ we can find $U=U(n,\varepsilon)$ such that $\mu(E_u)<\varepsilon$ for $u\geqq U$.*

Proof. $\mu\left(\bigcup_{i=0}^{n-1}\tau^i B(P^u(\mathrm{I\!I}))\right)=\dfrac{(n-1)}{2^u}w(\mathrm{I\!I})$.

Lemma 3.3. $\mathscr{L}$ is a Markov partition for τ.

Proof. Define for $L, M\in\mathscr{L}$

$$p(L,M)=\begin{cases}1 & L\subset A(\mathrm{I\!I}),\ M=\tau L,\\ 0 & L\subset A(\mathrm{I\!I}),\ M\neq\tau L,\\ \dfrac{\mu(M)}{w(T_1)} & L=A(C_{ij}),\ M\subset B(T_{\sigma_j(i)}),\ 1\leqq i\leqq k,\ 1\leqq j\leqq m,\\ 0 & L=A(C_{ij}),\ M\not\subset B(T_{\sigma_j(i)}),\ 1\leqq i\leqq k,\ 1\leqq j\leqq m.\end{cases}$$

We must check that this definition satisfies (3.1), (3.2), (3.3).

Let $M\in\mathscr{L}$, $M\not\subset B(\mathrm{I\!I})$ and let $L_1=\tau^{-1}M$. Then

$$p(L,M)=\begin{cases}1 & L=L_1\\ 0 & L\neq L_1\end{cases}$$

and so

$$\sum_{L\in\mathscr{L}}\mu(L)\,p(L,M)=\mu(L_1)=\mu(M)$$

and (3.1) is true in this case. If now $M \subset B(T_i)$, $1 \leqq i \leqq k$, then

$$p(L, M) = \begin{cases} \dfrac{\mu(M)}{w(T_1)}, & L = A(C_{lj}) \quad \text{with } \sigma_j(l) = i \\ 0 & \text{otherwise} \end{cases}$$

and so

$$\sum_{L \in \mathscr{L}} \mu(L) p(L, M) = \sum_{\{(j,\, l):\, \sigma_j(l) = i\}} \mu(A(C_{lj})) \frac{\mu(M)}{w(T_1)} = \mu(M)$$

as

$$\sum_{\{(j,\, l):\, \sigma_j(l) = i\}} \mu(A(C_{lj})) = \sum_{j=1}^{m} \mu(A(C_{\sigma_j^{-1}(i),\, j})) = w(T_1).$$

So (3.1) holds.

Now if $L \not\subset A(\mathbb{T})$,

$$p(L, M) = \begin{cases} 1 & M = \tau L \\ 0 & M \neq \tau L \end{cases}$$

and so

$$\sum_{M \in \mathscr{L}} p(L, M) = 1$$

and if $L = A(C_{ij})$, $1 \leqq i \leqq k$, $1 \leqq j \leqq m$,

$$p(L, M) = \begin{cases} \dfrac{\mu(M)}{w(T_1)}, & M \subset B(T_{\sigma_j(i)}), \\ 0, & M \not\subset B(T_{\sigma_j(i)}). \end{cases}$$

and so

$$\sum_{M \in \mathscr{L}} p(L, M) = \sum_{M \subset B(T_{\sigma_j(i)})} \frac{\mu(M)}{w(T_1)} = 1.$$

So (3.2) holds.

Now suppose $D \in \mathscr{L}_n(L)$, $L \not\subset A(\mathbb{T})$ and let $M_1 = \tau L$. Clearly

$$\mu(\tau D \cap M) = 0$$

for $M \neq M_1$. Recall that by Corollary 3.2 we can write for $u \geqq U(n, \varepsilon)$

$$D = D_u \cup E_u$$

where D_u is a union of copies of L in $P^u(\mathbb{T})$ and $\mu(E_u) < \varepsilon$. As $L \not\subset A(\mathbb{T})$ no copy of L in $P^u(\mathbb{T})$ is a subset of $A(P^u(\mathbb{T}))$. Hence the column level in $P^u(\mathbb{T})$ immediately above each copy of L where $L \subset D_u$ will be a copy of M_1. Hence

$$\mu(\tau D_u \cap M_1) = \mu(D_u)$$

and so

$$\mu(D)-\varepsilon<\mu(\tau D_u\cap M_1)\leqq\mu(\tau D\cap M_1)\leqq\mu(D).$$

As ε is arbitrary we have

$$\mu(\tau D\cap M_1)=\mu(D)$$

as required for this choice of L since $p(L, M_1)=1$.

Now suppose $D\in\mathscr{L}_n(L)$ and $L=A(C_{ij})$, $1\leqq i\leqq k$, $1\leqq j\leqq m$. As τ maps every copy of $A(C_{ij})$ linearly to the base of a copy of $T_{\sigma_j(i)}$ we have

$$\mu(\tau D\cap M)=0$$

for $M\notin B(T_{\sigma_j(i)})$. We write, as before,

$$D=D_u\cup E_u$$

and let

$$D_u=D_u^1\cup F_u$$

where D_u^1 is a union of copies of L of rank less than 2^u-1 and F_u is a union of copies of L of rank 2^u. If u is large enough we have, for $\varepsilon>0$,

$$\mu(E_u\cup F_u)<\varepsilon.$$

Now, if $M\subset B(T_{\sigma_j(i)})$, we have

$$\mu(\tau D_u^1\cap M)=\frac{\mu(M)}{w(T_1)}\mu(D_u^1)$$

and so

$$(\mu(D)-\varepsilon)\frac{\mu(M)}{w(T_1)}<\mu(\tau D_u^1\cap M)\leqq\mu(\tau D\cap M).$$

But

$$\mu(\tau D\cap M)=\mu(\tau D_u^1\cap M)+\mu(\tau(E_u\cup F_u)\cap M)<\frac{\mu(M)}{w(T_1)}+\varepsilon.$$

As ε is arbitrary we have

$$\mu(\tau D\cap M)=\frac{\mu(M)}{w(T_1)}\mu(D)$$

as required for this choice of L, M as $p(L, M)=\dfrac{\mu(M)}{w(T_1)}$. Hence (3.3) holds and the proof of the lemma is complete.

For $u\geqq 1$, let $\mathscr{L}^u$ denote the partition of $\mathbb{T}^*$ consisting of the column levels in $P^u(\mathbb{T})$. Analogous to Lemma 3.3 we have

Lemma 3.4. *For each $u\geqq 1$, $\mathscr{L}^u$ is a Markov partition for τ.*

Proof. Define for $L, M \in \mathscr{L}^u$

$$p_u(L, M) = \begin{cases} 1 & L \triangleleft A(P^u(\mathbb{T})),\ M = \tau L \\ 0 & L \triangleleft A(P^u(\mathbb{T})),\ M \neq \tau L \\ \dfrac{\mu(M)}{w(P^u(T_1))} & L = A(C^u_{il}),\quad M \subset B(T_{\sigma^u_i(i)}) \\ 0 & L = A(C^u_{il}),\quad M \not\subset B(T_{\sigma^u_i(i)}) \end{cases}$$

where C^u_{il} is the lth column of $P^u(T_i)$. That this definition satisfies (3.1), (3.2), (3.3) may be checked in a way analogous to the proof of Lemma 3.3.

Let $\mathscr{A}_0$ be the σ-algebra $\bigvee_{-\infty}^{\infty} \tau^i \mathscr{L}$ and for $u \geqq 1$ let $\mathscr{A}_u$ be the σ-algebra $\bigvee_{-\infty}^{\infty} \tau^i \mathscr{L}^u$.

Corollary 3.5. *τ is a Markov shift on each $\mathscr{A}_u$, $u \geqq 0$.*

τ is ergodic on $\mathscr{A}_0$ if for each $L, M \in \mathscr{L}$ we can find $h \in \mathbb{Z}$ such that

(3.4) $\quad \mu(\tau^h L \cap M) > 0.$

Now if L is a column level in C_{ij} and M is a column level in C_{pq}, $1 \leqq i, p \leqq k$, $1 \leqq j, q \leqq m$ and $(i, j) = (p, q)$ then (3.4) holds with h the difference in the heights of L and M in the column C_{ij}. If $(i, j) \neq (p, q)$ then (3.4) holds if and only if for some $u \geqq 1$ there is a column in $P^u(\mathbb{T})$ having a copy of C_{ij} in the base and a copy of C_{pq} in the same column or vice versa. Now by (2.2) a typical column of $P^u(T_i)$ with a copy of C_{ij} in the case has the form

$$[C_{ij}, C_{\sigma_j(i), i_1}, C_{\sigma_{i_1}(\sigma_j(i)), i_2}, \ldots, C_{\sigma_{i_{2^u-2}}(\sigma_{i_{2^u-3}}(\ldots(\sigma_{i_1}(\sigma_j(i)))\ldots)), i_{2^u-1}}]$$

where $1 \leqq i_t \leqq m$, $1 \leqq t \leqq 2^u - 1$. Hence we have

Lemma 3.6. *τ is ergodic on $\mathscr{A}_0$ if and only if for each $1 \leqq i, p \leqq k$, $1 \leqq j, q \leqq m$, at least one of the following holds*

(i) $(i, j) = (p, q)$,

(ii) $\sigma_j(i) = p \quad$ *or* $\quad \sigma_q(p) = i$,

(iii) *for some $u \geqq 2$ we can find $1 \leqq s \leqq 2^u - 2$ and $1 \leqq i_t \leqq m$, $1 \leqq t \leqq s$ such that*

$$\sigma_{i_s}(\sigma_{i_{s-1}}(\ldots(\sigma_{i_1}(\sigma_j(i)))\ldots)) = p$$

or

$$\sigma_{i_s}(\sigma_{i_{s-1}}(\ldots(\sigma_{i_1}(\sigma_q(p)))\ldots)) = i.$$

Corollary 3.7. *If for some j_1, $1 \leqq j_1 \leqq m$, σ_{j_1} is the permutation*

$$\sigma_{j_1}(i) = i + 1 \pmod{k}$$

then τ is ergodic on $\mathscr{A}_0$.

Proof. If $(i,j) \neq (p,q)$ and $\sigma_j(i) \neq p$, suppose that

$$p = \sigma_j(i) + s \pmod{k},$$

then $\sigma_{j_1}^{(s)}(\sigma_j(i)) = p$, where $\sigma_{j_1}^{(s)}$ denotes the sth power of σ_{j_1}. In a similar way we can prove

Lemma 3.8. *Suppose $u \geqq 0$. If for some j_1, $1 \leqq j_1 \leqq m^{2^u}$, $\sigma_{j_1}^u$ is the permutation*

$$\sigma_{j_1}^u(i) = i + 1 \pmod{k}$$

then τ is ergodic on $\mathscr{A}_u$.

Theorem 3.9. *τ is ergodic on $\mathscr{B}(\mathbb{T}^*)$, the Borel subsets of $\mathbb{T}^*$, if for some $1 \leqq j_1, j_2 \leqq m$, $\sigma_{j_1}, \sigma_{j_2}$ are defined by*

$$\sigma_{j_1}(i) = i$$
$$\sigma_{j_2}(i) = i + 1 \pmod{k}.$$

Proof. Note that for each $u \geqq 0$, $l \in \{1, 2, \ldots, k\}$

$$\sigma^u_{(j_1-1)m^{2^u-1}+(j_1-1)m^{2^u-2}+\cdots+(j_1-1)m+j_1}(l) = \sigma_{j_1}^{(2^u)}(l) = l$$

and

$$\sigma^u_{(j_1-1)m^{2^u-1}+(j_1-1)m^{2^u-2}+\cdots+(j_1-1)m+j_2}(l) = \sigma_{j_2}(\sigma_{j_1}^{(2^u-1)}(l)) = l + 1 \pmod{k},$$

where $\sigma_{j_1}^{(2^u)}$ and $\sigma_{j_1}^{(2^u-1)}$ denote the 2^uth and (2^u-1)th powers of σ_{j_1} respectively. Hence, by Lemma 3.8, τ is ergodic on each $\mathscr{A}_u$, $u \geqq 0$. As $\mathscr{B}(\mathbb{T}^*)$ is the σ-algebra generated by $\{\mathscr{A}_u : u \geqq 0\}$, τ is ergodic on $\mathscr{B}(\mathbb{T}^*)$.

We now investigate the conditions under which τ is mixing. As τ is a Markov Shift on each $\mathscr{A}_u$, $u \geqq 0$, τ will be mixing on $\mathscr{A}_u$ if it is ergodic and aperiodic. In addition to conditions on the permutations we have to impose a condition on heights of certain columns of $\mathbb{T}$, as was the case for Friedman's construction, to obtain the aperiodicity of τ.

Lemma 3.10. *If for some j_1, j_2, j_3, $1 \leqq j_i \leqq m$, $i = 1, 2, 3$, and some i, $1 \leqq i \leqq k$, we have*

(i) $\sigma_{j_1}(l) = l + 1 \pmod{k}$, $\quad l \in \{1, 2, \ldots, k\}$,

(ii) $\sigma_{j_2}(l) = \sigma_{j_3}(l) = l$, $\quad l \in \{1, 2, \ldots, k\}$,

(iii) *if h is the height of C_{ij_2} then C_{ij_3} has height $h+1$,*

then τ is mixing on $\mathscr{A}_0$.

Proof. τ is ergodic on $\mathscr{A}_0$ by Lemma 3.7. Moreover

$$C_{(j_2-1)m^3+(j_2-1)m^2+(j_2-1)m+p}(P^2(T_i)) = [C_{ij_2}, C_{ij_2}, C_{ij_2}, C_{ip}], \quad 1 \leqq p \leqq m$$

and

$$C_{(j_2-1)m^3+(j_3-1)m^2+(j_2-1)m+q}(P^2(T_i))$$
$$=[C_{ij_2}, C_{ij_3}, C_{ij_2}, C_{iq}], \quad 1 \leqq q \leqq m,$$

and so for any column level L in C_{ij_2} we have

$$\mu(\tau^{2h} L \cap L) > 0 \quad \text{and} \quad \mu(\tau^{2h+1} L \cap L) > 0$$

which implies that τ is aperiodic and therefore mixing.

In a similar way we can show

Lemma 3.11. *Suppose* $u \geqq 0$. *If for some* j_1, j_2, j_3, $1 \leqq j_i \leqq m^{2^u}$, $i = 1, 2, 3$, *and some* i, $1 \leqq i \leqq k$,

(i) $\sigma_{j_1}^u(l) = l+1 \pmod k$, $\quad l \in \{1, 2, \ldots, k\}$,

(ii) $\sigma_{j_2}^u(l) = \sigma_{j_3}^u(l) = l$. $\quad l \in \{1, 2, \ldots, k\}$,

(iii) *the heights of the* j_2th *and* j_3th *columns of* $P^u(T_i)$ *differ by* 1,

then τ *is mixing on* $\mathscr{A}_u$.

However it turns out that if the conditions of Lemma 3.10 are satisfied by $\mathbb{T}$ and Σ then the conditions of Lemma 3.11 are satisfied by $P^u(\mathbb{T})$ and Σ_u for each $u \geqq 0$.

Theorem 3.12. If $\mathbb{T}$ and Σ satisfy the conditions of Lemma 3.10 then τ is mixing on each $\mathscr{A}_u$, $u \geqq 0$.

Proof. Note that

$$C_{(j_2-1)m^{2^u-1}+(j_2-1)m^{2^u-2}+\ldots+(j_2-1)m+j_2}(P^u(T_i)) = [C_{ij_2}, C_{ij_2}, \ldots, C_{ij_2}, C_{ij_2}]$$

and

$$C_{(j_2-1)m^{2^u-1}+(j_2-1)m^{2^u-2}+\ldots+(j_2-1)m+j_3}(P^u(T_i)) = [C_{ij_2}, C_{ij_2}, \ldots, C_{ij_2}, C_{ij_3}]$$

have heights differing by 1 and also for $l \in \{1, 2, \ldots, k\}$

$$\sigma^u_{(j_2-1)m^{2^u-1}+(j_2-1)m^{2^u-2}+\ldots+(j_2-1)m+j_2}(l) = \sigma_{j_2}^{(2^u)}(l) = l$$

and

$$\sigma^u_{(j_2-1)m^{2^u-1}+(j_2-1)m^{2^u-2}+\ldots+(j_2-1)m+j_3}(l) = \sigma_{j_3}(\sigma_{j_2}^{(2^u-1)}(l)) = l,$$

where $\sigma_{j_2}^{(2^u)}$ and $\sigma_{j_2}^{(2^u-1)}$ represent the 2^uth and (2^u-1)th powers of σ_{j_2} respectively. Moreover for $1 \leqq l \leqq k$

$$\sigma^u_{(j_2-1)m^{2^u-1}+(j_2-1)m^{2^u-2}+\ldots+(j_2-1)m+j_1}(l) = \sigma_{j_1}(\sigma_{j_2}^{(2^u-1)}(l)) = l+1 \pmod k.$$

Hence the conditions of Lemma 3.11 hold and so τ is mixing on each $\mathscr{A}_u$, $u \geqq 0$.

Corollary 3.13. *If* $\mathbb{T}$ *and* Σ *satisfy the conditions of Lemma* 3.10, τ *is a Bernoulli shift on* $\mathscr{A}_u$, $u \geqq 0$.

Proof. In [3] it is shown that a mixing Markov shift is weak Bernoulli and hence a Bernoulli shift.

Theorem 3.14. *If* $\mathbb{T}$ *and* Σ *satisfy the conditions of Lemma* 3.10 *then* τ *is Bernoulli on* $\mathscr{B}(\mathbb{T}^*)$.

Proof. By Corollary 3.13, τ is Bernoulli on each of the σ-algebras $\mathscr{A}_u$, $u \geqq 0$, which form an increasing sequence. Hence by a result of Ornstein [5] τ is Bernoulli on $\mathscr{B}(\mathbb{T}^*)$, the σ-algebra generated by $\mathscr{A}_u$, $u \geqq 0$.

§ 4. Stopping Time Transformations

Let τ be an automorphism of $(\Omega, \mathscr{B}, \mu)$. If v is a measurable mapping from $(\Omega, \mathscr{B})$ to the natural numbers $\mathbb{N}$ then the transformation τ^v defined by

$$\tau^v \omega = \tau^n \omega, \qquad \omega \in \{v = n\},\ n \in \mathbb{N},$$

is an automorphism of $(\Omega, \mathscr{B}, \mu)$ if and only if the sets $\tau^n\{v = n\}$, $n \in \mathbb{N}$, form a partition of Ω. If this is the case v is called a stopping time for $(\Omega, \mathscr{B}, \mu, \tau)$ and τ^v is called a stopping time transformation. Neveu [4] established

Theorem 4.1. (i) *If* $\Omega \supset D_0 \supset D_1 \ldots$ *is a sequence of measurable sets such that either* $D_{k-1} \neq \emptyset$ *and* $D_k = \emptyset$ *for some* $k \geqq 1$ *or* $\bigcap_{n=0}^{\infty} D_n = \emptyset$ *then there is a stopping time* v *of* $(\Omega, \mathscr{B}, \mu, \tau)$ *such that*

$$(4.1) \qquad \tau^v \omega = \begin{cases} \omega & \omega \in E_0 = D_0^c \\ \tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{n-1}}(\omega))\ldots)) & \omega \in E_n,\ n \geqq 1 \end{cases}$$

where $E_n = D_{n-1} - D_n$, $n \geqq 1$, *and for any* $D \in \mathscr{B}$, τ_D *represents the induced transformation of* τ *on* D.

(ii) *Corresponding to any stopping time* v *of* $(\Omega, \mathscr{B}, \mu, \tau)$ *there is a sequence of measurable sets* $\Omega \supset D_0 \supset D_1 \ldots$ *as in* (i) *such that* τ^v *has the form* (4.1).

(iii) *If* v *is a stopping time for* $(\Omega, \mathscr{B}, \mu, \tau)$ *and* τ *is ergodic then* $\int v\, d\mu$ *is either a non-negative integer or* $+\infty$. *If* $\int v\, d\mu = k < \infty$ *then the sets* D_n, $n \geqq 0$, *defined in* (ii) *satisfy* $D_{k-1} \neq \emptyset$, $D_k = \emptyset$ *whereas if* $\int v\, d\mu = +\infty$ *we have* $D_n \neq \emptyset$, $n \geqq 0$ *and* $\bigcap_{n=0}^{\infty} D_n = \emptyset$.

In the sequel we shall assume that $\int v\, d\mu = k < \infty$, noting that if $\int v\, d\mu = +\infty$ and $\Omega \supset D_0 \supset D_1 \ldots$ are the sets defined by Theorem 4.1(ii) the transformation

$$\tau^{v_k} = \begin{cases} \omega & \omega \in D_k \\ \tau^v \omega & \omega \in D_k^c \end{cases}$$

differs from τ^v on a set of small measure for large k. As τ^v is the identity transformation on $\{v = 0\}$, for our study of mixing properties of τ^v we shall restrict attention to the behaviour of τ^v on $D_0 = \bigcup_{n=1}^{k} E_n$. Note that the sets E_n,

$0 \leqq n \leqq k$, form a partition of Ω. We shall only consider those transformations τ for which there is a tower T satisfying $T^* = \Omega$ and $\tau = \tau(T)$ and we shall assume that our stopping time is chosen so that each of the sets E_n, $0 \leqq n \leqq k$, defined in Theorem 4.1 is a union of column levels in T. If a column level L in T is a subset of E_n, $0 \leqq n \leqq k$, we say that L has name n and write $n(L) = n$. We shall also assume that in each column of T there is at least one column level with name k. Note that if for some tower T and stopping time ν these conditions are not satisfied then, as the towers $S^u(T)$, $u \geqq 1$, have smaller and smaller column widths, we can modify ν on as small a subset of Ω as we please to have these conditions satisfied by $S^u(T)$ for some $u \geqq 1$ and a slightly modified stopping time.

As we restrict attention to the action of τ^ν on $\bigcup_{n=1}^{k} E_n$ we delete from T any column level with name 0 and denote by $\hat{T}$ the resulting tower. Note that the deleted set is invariant under τ^ν and that $\widehat{S^u(T)} = S^u(\hat{T})$, $u \geqq 1$.

Recall that in the definition of τ, τ maps each column level in T except tops of columns, linearly onto another column level within the same column of T. If $C = (L_i : 1 \leqq i \leqq h)$ is a column in $\hat{T}$, in order to define τ^ν on a level L_i with name n, by Theorem 4.1, we first have to define $\tau_{D_{n-1}}$ on L_i, then, if $n \geqq 2$, define $\tau_{D_{n-2}}$ on $\tau_{D_{n-1}}(L_i)$, etc. As $D_{n-1} = \bigcup_{l=n}^{k} E_l$, $1 \leqq n \leqq k$, $\tau_{D_{n-1}}(L_i)$ is another column level within C if and only if

$$\{j : i+1 \leqq j \leqq h,\ n(L_j) \geqq n\}$$

is non-empty, in which case $\tau_{D_{n-1}}$ maps L_i linearly onto L_{i_1} where

$$i_1 = \min\{j : i+1 \leqq j \leqq h,\ n(L_j) \geqq n\}.$$

Then, if $n \geqq 2$, $\tau_{D_{n-2}}(L_{i_1})$ is another column level within C if and only if

$$\{j : i_1 + 1 \leqq j \leqq h,\ n(L_j) \geqq n-1\}$$

is non-empty, in which case $\tau_{D_{n-2}}$ maps L_{i_1} linearly onto L_{i_2} where

$$i_2 = \min\{j : i_1 + 1 \leqq j \leqq h,\ n(L_j) \geqq n-1\}.$$

Clearly there may be column levels other than $A(C)$, i.e. some L_i, $1 \leqq i \leqq h-1$, for which, if $n(L_i) = n$, some or all of the sets

$$\tau_{D_{n-l}}(\tau_{D_{n-l+1}}(\ldots(\tau_{D_{n-1}}(L_i))\ldots)), \qquad 1 \leqq l \leqq n,$$

are not column levels within C. In our tower block construction for τ^ν we shall be inerested in the number of these sets which are not column levels within C. For a column level L_i, $1 \leqq i \leqq h$, with name n we shall define $u(L_i) = 0$ if each of

$$\tau_{D_{n-l}}(\tau_{D_{n-l+1}}(\ldots(\tau_{D_{n-1}}(L_i))\ldots)), \qquad 1 \leqq l \leqq n,$$

is a column level within C and $u(L_i)=u\geqq 1$ if each of

$$\tau_{D_{n-l}}(\tau_{D_{n-l-1}}(\ldots(\tau_{D_{n-1}}(L_i))\ldots)), \qquad 1\leqq l\leqq n-u,$$

is a column level within C but

$$\tau_{D_{u-1}}(\tau_{D_u}(\ldots(\tau_{D_{n-1}}(L_i))\ldots)).$$

is not a column level in $C(u(L_i)=n$ meaning $\tau_{D_{n-1}}(L_i)$ is not a column level in $C)$. So $u(L_i)$ counts the number of steps in the iterative definition of τ^{ν} on L_i given by Theorem 4.1 which are left undefined within the column containing L_i.

Equivalently, if $n(L_i)=n$, $u(L_i)=0$ if the set

$$\{i_1, i_2, \ldots, i_n: i<i_1<i_2<\ldots<i_n\leqq h,\ n(L_{i_j})\geqq n-j+1,\ 1\leqq j\leqq n\}$$

is non-empty, $u(L_i)=u$, with $1\leqq u\leqq n-1$ if the sets

$$\{i_1, i_2, \ldots, i_s: i<i_1<i_2<\ldots<i_s\leqq h,\ n(L_{i_j})\geqq n-j+1,\ 1\leqq j\leqq s\}$$

are non-empty for $s\leqq n-u$ and empty for $s\geqq n-u+1$ and $u(L_i)=n$ if the set

$$\{j: i<j\leqq h,\ n(L_j)\geqq n\}$$

is empty.

Note that for any column level L in $\hat{T}$ with $n(L)=n$ we have $0\leqq u(L)\leqq n$. Any column level L in $\hat{T}$ with $u(L)\geqq 1$ is called a ν-top. Note that a column level L is not a ν-top if and only if there is another column level L' in the same column as L and above L such that $\tau^{\nu}(L)=L'$. If C is a column in $\hat{T}$ we let $A_1(C), A_2(C), \ldots, A_l(C)$ denote the ν-tops in C where we number from the top of C, so $A_1(C)=A(C)$, $A_2(C)$ is the first ν-top in C below $A_1(C)$, etc. We now study properties of the ν-tops in $\hat{T}$ and, in particular, we shall show that each column in $\hat{T}$ may be assumed to contain exactly k ν-tops.

Lemma 4.2. *For each column C in $\hat{T}$, $n(A_i(C))\geqq i$, $1\leqq i\leqq l$.*

Proof. Clearly $n(A_1(C))\geqq 1$. Suppose $n(A_i(C))\geqq i$, $1\leqq i\leqq j$. Then if $n(A_{j+1}(C))\leqq j$ we have $u(A_{j+1}(C))=0$ and so $A_{j+1}(C)$ is not a ν-top.

Corollary 4.3. *There are at most k ν-tops in each column of $\hat{T}$.*

Proof. For any column level L in $\hat{T}$ we have $n(L)\leqq k$.

Lemma 4.4. *If C is a column in $\hat{T}$ and L is a column level of C lying between $A_i(C)$ and $A_{i+1}(C)$, $1\leqq i\leqq l-1$, then $n(L)\leqq i$.*

Proof. The result is clearly true for $i=1$ since $A_2(C)$ must be the first column level in C (counting from the top of C) with name greater than or equal to 2. Now suppose the result is true for $1\leqq i\leqq j$. If $j\leqq l-2$, let L be the first column level below $A_{j-1}(C)$ with $n(L)=n\geqq j+2$. We show that L is a ν-top and so $L=A_{j-2}(C)$ and the result is true for $1\leqq i\leqq j+1$.

Now $\tau_{D_{n-1}}(L)$ is a column level in C if and only if there is a column level L' above L with $n(L')\geqq n$ and so by the inductive hypothesis and the definition of L.

L can only be one of $A_i(C)$, $1 \leqq i \leqq j+1$. Similarly each of the sets

$$\tau_{D_{n-t}}(\tau_{D_{n-t+1}}(\ldots(\tau_{D_{n-1}}(L))\ldots)), \qquad 1 \leqq t \leqq n,$$

can be a column level in C if and only if it is equal to one of $A_i(C)$, $1 \leqq i \leqq j+1$. Thus we see that $u(L) \geqq 1$ and so L is a v-top.

Corollary 4.5. *If $n(A_j(C)) = n$ and $u(A_j(C)) = u$, then each of*

$$\tau_{D_{n-i}}(\tau_{D_{n-i+1}}(\ldots(\tau_{D_{n-1}}(A_j(C))\ldots)), \qquad 1 \leqq i \leqq n-u,$$

is a v-top.

Corollary 4.6. *Suppose $\hat{T} = (C_j\colon 1 \leqq j \leqq m)$ and $S(\hat{T}) = (C'_j\colon 1 \leqq j \leqq m^2)$. If l_j is the number of v-tops in C_j, $1 \leqq j \leqq m$, and l'_j is the number of v-tops in C'_j, $1 \leqq j \leqq m^2$, then if $l_j \leqq k-1$,*

$$l'_{(i-1)m+j} \geqq l_j + 1$$

for each $1 \leqq i \leqq m$.

Proof. Recall that for $1 \leqq i \leqq m$

$$C_{(i-1)m+j}(S(\hat{T})) = [C_i, C_j].$$

The copies of $A_n(C_j)$, $1 \leqq n \leqq l_j$, will again be v-tops in $C_{(i-1)m+j}$. Moreover if L is the highest column level in C_i with $n(L) \geqq l_j + 1$, which exists since we have assumed that each column in T contains a column level with name k, the copy of L in the base of $C_{(i-1)m+j}$ will be a v-top in $C_{(i-1)m+j}$.

Corollary 4.7. *For any tower T each column in $S^{k-1}(\hat{T})$ contains exactly k v-tops.*

Because of this last corollary we shall assume in the sequel that T is chosen so that each column in $\hat{T}$ contains exactly k v-tops. If $\hat{T} = (C_j\colon 1 \leqq j \leqq m)$ we shall let

$$A^v(\hat{T}) = \bigcup_{j=1}^{m} \bigcup_{n=1}^{k} A_n(C_j).$$

Lemma 4.8. *If C is a column in $\hat{T}$ with v-tops $A_i(C)$, $1 \leqq i \leqq k$, then $i \neq j$ implies*

$$u(A_i(C)) \neq u(A_j(C)).$$

Proof. Suppose $i \neq j$ and $u(A_i(C)) = u(A_j(C)) = u$, say, and let $n(A_i(C)) = n_i$, $n(A_j(C)) = n_j$. If

$$L_i = \begin{cases} \tau_{D_u}(\tau_{D_{u-1}}(\ldots(\tau_{D_{n_i-1}}(A_i(C)))\ldots)), & u \leqq n_i - 1 \\ A_i(C), & u = n_i \end{cases}$$

and

$$L_j = \begin{cases} \tau_{D_u}(\tau_{D_{u-1}}(\ldots\tau_{D_{n_j-1}}(A_j(C)))\ldots)), & u \leqq n_j - 1 \\ A_j(C), & u = n_j \end{cases}$$

then $L_i \neq L_j$ and we may as well assume that L_j is above L_i in C. We shall see that whatever the values of u, n_i, n_j, then $\tau_{D_{u-1}}(L_i)$ is a column level in C, contradicting the assumption that $u(A_i(C))=u$. If $u=n_j$ the result follows from the fact that $n(L_j)=u$, whereas if $u \leqq n_j - 1$ the result follows from the fact that $n(L_j) \geqq u+1$.

§ 5. The Tower Block Construction for τ^v

In this section se shall show that if $\tau=\tau(T)$ and v is a stopping time for $(\Omega, \mathscr{B}, \mu, \tau)$ with $\int v\, d\mu = k$ then we can find a tower block $\mathbb{T}$ and a set of permutations Σ such that $\tau^v = \tau(\Sigma, \mathbb{T})$.

Let C be a column in $\hat{T}$ and let $L_n(C)$ be the first column level (counting from the base) which satisfies $n(L_n(C)) \geqq n$. Then define $B_1(C)=L_1(C)$ and for $n \geqq 2$

$$B_n(C)=\tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{n-2}}(L_n(C)))\ldots)),$$

where we assume that T is chosen so that each $B_n(C)$, $1 \leqq n \leqq k$, is a column level in C. (If this is not the case for T it will be for $S^u(T)$ for some $u \geqq 1$.)

Now for each $B_n(C)$, $1 \leqq n \leqq k$, there is $i_n \geqq 0$ such that $(\tau^v)^{i_n} B_n(C)$ is a v-top in C and in the column

$$D_n(C)=\{(\tau^v)^j B_n(C): 0 \leqq j \leqq i_n\}$$

τ^v maps each column level except the v-top $(\tau^v)^{i_n} B_n(C)$ linearly onto the column level immediately above it. Since τ^v is invertible we have

$$\bigcup_{n=1}^{k} D_n(C)^* = C^*.$$

Now if $\hat{T}=(C_j: 1 \leqq j \leqq m)$ we define for $1 \leqq n \leqq k$

$$T_n=(C_{nj}: 1 \leqq j \leqq m)$$

where $C_{nj}=D_n(C_j)$ and we let

$$\mathbb{T}=(T_n: 1 \leqq n \leqq k).$$

Note that $\mathbb{T}^* = \hat{T}^*$. We also clearly have

Lemma 5.1. (i) $A(\mathbb{T})=A^v(\hat{T})$.

(ii) $\tau_{\mathbb{T}}(\omega)=\tau^v(\omega)$, $\quad \omega \in \mathbb{T}^* - A(\mathbb{T})$.

We shall define $\Sigma=\{\sigma_j: 1 \leqq j \leqq m\}$ by $\sigma_j(n)=u(A(C_{nj}))$. Because of Lemma 4.8, each σ_j, $1 \leqq j \leqq m$, is a permutation of $\{1, 2, \ldots, k\}$.

Lemma 5.2. *For each* $u \geqq 0$,

(i) $A(P^u(\mathbb{T}))=A^v(S^u(\hat{T}))$.

(ii) $\tau_{P^u(\mathbb{T})}(\omega)=\tau^v(\omega)$, $\quad \omega \in \mathbb{T}^* - A(P^u(\mathbb{T}))$.

Proof. The result is true for $u=0$ by Lemma 5.1. Suppose the result is true for $0 \leqq u \leqq v$ and let

$$P^v(T_i)=(C_{ij}: 1 \leqq j \leqq m^{2v}),$$
$$S^v(\hat{T}) =(C_j: 1 \leqq j \leqq m^{2v}).$$

Then

$$A(P^{v+1}(\mathbb{T})) = \bigcup_{n=1}^{k} A(P^{v+1}(T_n))$$
$$= \bigcup_{j=1}^{m^{2v}} \bigcup_{n=1}^{k} A(\alpha_{j-1}(\tfrac{1}{2}(P^v(T_{\sigma_j^v(n)}))_1))$$

where $\alpha_{j-1} = w(C_{ij})/w(P^v(T_i))$

$$= \bigcup_{j=1}^{m^{2v}} \bigcup_{n=1}^{k} A(\alpha_{j-1}(\tfrac{1}{2}(P^v(T_n))_1))$$
$$= \bigcup_{n=1}^{k} A(\tfrac{1}{2}(P^v(T_n))_1)$$
$$= A^v(\tfrac{1}{2}(S^v(\hat{T}))_1) = A^v(S^{v-1}(\hat{T})).$$

So (i) holds for $u=v+1$. To check that (ii) holds for $u=v+1$, it is sufficient to check that τ^v agrees with $\tau_{P^{v-1}(\mathbb{T})}$ on $A(P^v(\mathbb{T}))-A(P^{v-1}(\mathbb{T}))$.

Now for each $1 \leqq i \leqq k$, $1 \leqq j \leqq m^{2v}$, $A(C_{ij})$ is a v-top in C_j in $S^v(\hat{T})$. Suppose that

$$u(A(C_{ij}))=u, \qquad n(A(C_{ij}))=n.$$

So each of

$$\tau_{D_{n-l}}(\tau_{D_{n-l+1}}(\ldots(\tau_{D_{n-1}}(A(C_{ij})))\ldots)), \qquad 1 \leqq l \leqq n-u$$

is a column level in $S^v(\hat{T})$ above $A(C_{ij})$ but

$$\tau_{D_{u-1}}(\tau_{D_u}(\ldots(\tau_{D_{n-1}}(A(C_{ij})))\ldots))$$

is not a column level in $S^v(\hat{T})$. Now looking at $S^{v+1}(\hat{T})$ we have

$$\tau_{D_{u-1}}(\tau_{D_u}(\ldots(\tau_{D_{n-1}}((\tfrac{1}{2}A(C_{ij}))_0))\ldots))$$
$$= \bigcup_{l=1}^{m^{2v}} \alpha_{j-1}(\tfrac{1}{2}L_u(C_l))_1$$

where $\alpha_{j-1} = w(C_j)/w(S^v(\hat{T}))$, and so

$$\tau^v((\tfrac{1}{2}A(C_{ij}))_0) = \bigcup_{l=1}^{m^{2v}} \alpha_{j-1}(\tfrac{1}{2}B_u(C_l))_1$$
$$= \bigcup_{l=1}^{m^{2v}} \alpha_{j-1}(\tfrac{1}{2}B_{\sigma_j^v(i)}(C_l))_1$$
$$= B(\alpha_{j-1}(\tfrac{1}{2}P^v(T_{\sigma_j^v(i)}))_1)$$

and so τ^v agrees with $\tau_{P^v(\mathbb{T})}$ on each $A(C_{ij})$, $1 \leqq i \leqq k$, $1 \leqq j \leqq m^{2^v}$ and so

$$\tau_{P^{v+1}(\mathbb{T})}(\omega) = \tau^v(\omega)$$

for $\omega \in \mathbb{T}^* - A(P^{v+1}(\mathbb{T}))$ and so, by induction, the lemma is proved.

Hence we have

Theorem 5.3. $\tau^v = \tau(\Sigma, \mathbb{T})$.

And the following theorem follows from the results of § 3.

Theorem 5.4. (i) τ^v *is ergodic on* $\{v \neq 0\}$ *if for some* $1 \leqq j_1, j_2 \leqq m$, $\sigma_{j_1}, \sigma_{j_2}$ *are defined by*

$$\sigma_{j_1}(i) = i, \qquad \sigma_{j_2}(i) = i+1 \pmod{k}, \quad i \in \{1, 2, \ldots, k\}.$$

(ii) τ^v *is Bernoulli on* $\{v \neq 0\}$ *if for some* j_1, j_2, j_3, $1 \leqq j_i \leqq m$, $i = 1, 2, 3$, *and some* i, $1 \leqq i \leqq k$ *we have*

(a) $\sigma_{j_1}(l) = l+1 \pmod{k}$, $\quad l \in \{1, 2, \ldots, k\}$,

(b) $\sigma_{j_2}(l) = \sigma_{j_3}(l) = l$, $\quad l \in \{1, 2, \ldots, k\}$,

(c) *If* $T_i = (C_{ij}, 1 \leqq j \leqq m)$ *then the heights of* C_{ij_2} *and* C_{ij_3} *differ by one.*

§ 6. Density Results

In order to investigate density results we need to investigate to what extent a stopping time needs to be modified in order to satisfy the conditions for τ^v to be ergodic or Bernoulli as outlined in Theorem 5.4. These conditions are mainly concerned with the permutations Σ which are in turn determined by the names of column levels in T. So in this section we shall assign names between 0 and $k \geqq 1$ to column levels in T and take v to be the corresponding stopping time of $(\Omega, \mathscr{B}, \mu, \tau(T))$ as defined by Theorem 4.1(i). It will be convenient when assigning names to column levels to insist that if v is the stopping time which arises then the k v-tops in that column are the top k levels in that column.

Lemma 6.1. *Let* A_i, $1 \leqq i \leqq k$, *be the top* k *levels (numbered from the top) in a column* C *of* $\hat{T}$. *A necessary and sufficient condition for each* A_i, $1 \leqq i \leqq k$, *to be a* v-top *is*

$$n(A_i) \in \{i, i+1, \ldots, k\}, \qquad 1 \leqq i \leqq k.$$

Proof. That this condition is necessary is shown in Lemma 4.2. If $n(A_i) = n_i \geqq i$, then not all of

$$\tau_{D_{n_i - j}}(\tau_{D_{n_i - j+1}}(\ldots(\tau_{D_{n_i-1}}(A_i))\ldots)), \qquad 0 \leqq j \leqq n_i$$

can be column levels in C as there are only $i-1$ column levels above A_i in C.

We shall define an n-sequence of order k to be a collection $\{n_i : 1 \leqq i \leqq k\}$ satisfying $n_i \in \{i, i+1, \ldots, k\}$, $1 \leqq i \leqq k$. Thus if we give the ith level (counting from

the top) of a column in $\hat{T}$ the name n_i, $1 \leqq i \leqq k$, where $\{n_i: 1 \leqq i \leqq k\}$ is an n-sequence of order k, then these k column levels are v-tops. Note that the number of different n-sequences of order k is $k!$. If $\{n_i: 1 \leqq i \leqq k\}$ is an n-sequence of order k then each of

$$\{l, n_1+1, n_2+1, \ldots, n_k+1\}, \qquad 1 \leqq l \leqq k+1,$$

is an n-sequence of order $k+1$ and in fact we can obtain each of the $(k+1)!$ n-sequences of order $k+1$ from the n-sequences of order k by this method.

A collection $\{u_i: 1 \leqq i \leqq k\}$ is called a u-sequence of order k if it is a permutation of $\{1, 2, \ldots, k\}$. Corresponding to each n-sequence of order k we define a u-sequence of order k by $u_1 = n_1$ and for $j \geqq 2$, $u_j = u$, $1 \leqq u \leqq n_j - 1$, if the sets

$$\{i_1, i_2, \ldots, i_s: 1 \leqq i_s < i_{s-1} < \ldots < i_1 < j,\ n_{i_t} \geqq n_j - t + 1,\ 1 \leqq t \leqq s\}$$

are non-empty for $s \leqq n_j - u$ and empty for $s \geqq n_j - u + 1$ and $u_j = n_j$ if the set

$$\{i: 1 \leqq i < j,\ n_i \geqq n_j\}$$

is empty.

Equivalently, if A_i, $1 \leqq i \leqq k$, are the top k levels (counting from the top) of a column in $\hat{T}$ with $n(A_i) = n_i$, then $u_i = u(A_i)$. That the collection $\{u_i: 1 \leqq i \leqq k\}$ is a u-sequence of order k follows from Lemma 4.8.

Lemma 6.2. *Let $\{n_i: 1 \leqq i \leqq k\}$ be a n-sequence of order k and $\{u_i: 1 \leqq i \leqq k\}$ the corresponding u-sequence. If $\{n'_i: 1 \leqq i \leqq k+1\}$ is the n-sequence of order $k+1$ defined for some l, $1 \leqq l \leqq k+1$, by*

$$n'_1 = l$$
$$n'_{j+1} = n_j + 1 \qquad 1 \leqq j \leqq k$$

then the corresponding u-sequence of order $k+1$, $\{u'_i: 1 \leqq i \leqq k+1\}$ is defined by

$$u'_1 = l$$

and for $1 \leqq j \leqq k$,

$$u'_{j+1} \begin{cases} u_j & 1 \leqq u_j \leqq l-1 \\ u_j + 1 & l \leqq u_j \leqq k. \end{cases}$$

Proof. $u'_1 = l$ by definition. Now for $2 \leqq j \leqq k+1$, $1 \leqq s \leqq n'_j - 1$,

$$\begin{aligned} &\{i_1, i_2, \ldots, i_s: 2 \leqq i_s < i_{s-1} < \ldots < i_1 < j,\ n'_{i_t} \geqq n'_j - t + 1,\ 1 \leqq t \leqq s\} \\ &\quad = \{j_1 + 1, j_2 + 1, \ldots, j_s + 1: 1 \leqq j_s < j_{s-1} < \ldots < j_1 < j - 1, \\ &\qquad n_{j_t} \geqq n_{j-1} - t + 1,\ 1 \leqq t \leqq s\}. \end{aligned}$$

The set on the right hand side is non-empty for $s \leqq n_{j-1} - u_{j-1}$ and empty for $s \geqq n_{j-1} - u_{j-1} + 1$. Since $n'_1 = l$, if $1 \leqq u_{j-1} \leqq l-1$ and $s = n_{j-1} - u_{j-1}$,

$$\{i_1, i_2, \ldots, i_s, i_{s-1}: 1 \leqq i_{s-1} < i_s < \ldots < i_1 < j,\ n'_{i_t} \geqq n'_j - t + 1,\ 1 \leqq t \leqq s+1\}$$
$$= \{1\} \cup \{i_1, i_2, \ldots, i_s: 2 \leqq i_s < \ldots < i_1 < j,\ n'_{i_t} \geqq n'_j - t + 1,\ 1 \leqq t \leqq s\}$$

and so $u'_j = u_{j-1}$ in this case, whereas if $l \leqq u_{j-1} \leqq k$ and $s = n_{j-1} - u_{j-1}$

$$\{i_1, i_2, \ldots, i_s, i_{s-1}: 1 \leqq i_{s-1} < i_s < \ldots < i_1 < j,\ n'_{i_t} \geqq n'_j - t + 1,\ 1 \leqq t \leqq s+1\}$$
$$= \{i_1, i_2, \ldots, i_s: 2 \leqq i_s < \ldots < i_1 < j,\ n'_{i_t} \geqq n'_j - t + 1,\ 1 \leqq t \leqq s\}$$

and so $u'_j = u_{j-1} + 1$.

Theorem 6.3. *The mapping which takes an n-sequence of order k to the corresponding u-sequence of order k is one-to-one.*

Proof. Note that there are $k!$ n-sequences of order k and $k!$ u-sequences of order k. Hence it is sufficient to check that each n-sequence of order k gives rise to a unique u-sequence of order k. We proceed by induction on k. Clearly the result is true for $k = 1$. Suppose to each n-sequence of order k there is associated a unique u-sequence of order k. Lemma 6.2 shows that the procedure we have used to proceed from n-sequences of order k to n-sequences of order $k+1$ will preserve the uniqueness of definition of u-sequences.

As a result of this theorem there is a unique n-sequence of order k which gives rise to any particular u-sequence of order k. As Σ depends on the value of u on v-tops we want to see how we must change the names in columns of $\hat{T}$ in order to obtain the permutations required for the resulting transformations to be ergodic or Bernoulli. Let $\hat{T} = (C_j: 1 \leqq j \leqq m)$ and for each $1 \leqq j \leqq m$ let $A_i(C_j)$, $1 \leqq i \leqq k$, represent the top k levels (counting from the top) of the column C_j.

Theorem 6.4. *Let ρ be a permutation of $\{1, 2, \ldots, k\}$. We can assign names to $A_i(C_j)$, $1 \leqq i \leqq k$, so that $\sigma_j = \rho$.*

Proof. Let $B_n(C_j)$, $1 \leqq n \leqq k$, be as defined in § 5 and first of all assign the name k to each $A_i(C_j)$, $1 \leqq i \leqq k$. This ensures that each $A_i(C_j)$, $1 \leqq i \leqq k$, is a v-top. Now for each n, $1 \leqq n \leqq k$, we can find $t(n) \in \{1, 2, \ldots, k\}$ and $i_n \geqq 0$ such that

$$(\tau^v)^{i_n} B_n(C_j) = A_{t(n)}(C_j).$$

Note that t is a permutation of $\{1, 2, \ldots, k\}$.

Now we shall assign a new name n_i to $A_i(C_j)$, $1 \leqq i \leqq k$, with $\{n_i: 1 \leqq i \leqq k\}$ a particular n-sequence of order k and we denote by v' the stopping time associated with the partition E'_i, $1 \leqq i \leqq k$, by Theorem 4.1, where E'_i is the union of column levels now having name i. Note that changing the names in this way does not affect the permutation t, i.e., we shall still have for each $1 \leqq n \leqq k$

$$(\tau^{v'})^{i_n} B_n(C_j) = A_{t(n)}(C_j).$$

for, if $i_n = 0$, this is clearly true and if $i_n \geqq 1$ and

$$n((\tau^{v'})^{i_n - 1} B_n(C_j)) = l$$

for some l with $1 \leqq l \leqq k$, then for some h, $1 \leqq h \leqq l$,

$$\tau_{D_h}(\tau_{D_{h+1}}(\ldots(\tau_{D_{l-1}}((\tau^{\nu'})^{i_n-1} B_n(C_j)))\ldots)) \notin \bigcup_{i=1}^{k} A_i(C_j)$$

and

$$\tau_{D_{h-1}}(\tau_{D_h}(\ldots(\tau_{D_{l-1}}((\tau^{\nu'})^{i_n-1} B_n(C_j)))\ldots)) \subset \bigcup_{i=1}^{k} A_i(C_j).$$

But because $n(A_k(C_j))=k$ we have

$$\tau_{D_{h-1}}(\tau_{D_h}(\ldots(\tau_{D_{l-1}}((\tau^{\nu'})^{i_n-1} B_n(C_j)))\ldots)) = A_k(C_j)$$

and so

$$\tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{l-1}}((\tau^{\nu'})^{i_n-1} B_n(C_j)))\ldots)) = A_{k-h+1}(C_j)$$

irrespective of the particular values of $n(A_i(C_j)) \in \{i, i+1, \ldots, k\}$, $1 \leqq i \leqq k$.

In order to have $\sigma_j = \rho$ the particular n-sequence of order k we choose is the one corresponding to the u-sequence of order k defined by

$$u_i = \rho(t^{-1}(i)), \qquad 1 \leqq i \leqq k.$$

Then, for $1 \leqq i \leqq k$, $\sigma_j(i) = u_{t(i)} = \rho(i)$, as required.

In discussing the density of stopping time transformations with specified properties we shall work with the uniform metric

$$d(\tau_1, \tau_2) = \mu\{\omega: \tau_1(\omega) \neq \tau_2(\omega)\}$$

where τ_1, τ_2 are both automorphisms of $(\Omega, \mathcal{B}, \mu)$. It will be of use to relate the proximity of two stopping time transformations with the proximity of the associated partitions which are defined by Neveu's Theorem.

Lemma 6.5. *Let ν and ν' be stopping times for $(\Omega, \mathcal{B}, \mu, \tau)$ with $\int \nu \, d\mu = \int \nu' \, d\mu = k \geqq 1$ and let E_n, $0 \leqq n \leqq k$, and E'_n, $0 \leqq n \leqq k$ be as defined for ν and ν' respectively by Theorem 4.1. For any $\varepsilon > 0$ there is $\eta(\varepsilon, k)$ such that*

$$\mu(E_n \triangle E'_n) < \eta(\varepsilon, k), \qquad 0 \leqq n \leqq k,$$

implies

$$d(\tau^{\nu}, \tau^{\nu'}) < \varepsilon.$$

Proof. The result follows from the fact that, for $1 \leqq n \leqq k$,

$$\mu(\{\omega \in E_n \cap E'_n: \tau^{\nu}(\omega) \neq \tau^{\nu'}(\omega)\}) \leqq \sum_{l=0}^{n-1} \mu(D_l \triangle D'_l)$$

where

$$D_l = \bigcup_{m=l+1}^{k} E_m \quad \text{and} \quad D'_l = \bigcup_{m=l+1}^{k} E'_m.$$

For the case $n=1$ we note that

$$\{\omega\in E_1\cap E_1' : \tau^{\nu}(\omega)=\tau^{\nu'}(\omega)\}=\{\omega\in E_1\cap E_1' : \tau_{D_0}(\omega)=\tau_{D_0'}(\omega)\}$$
$$=(E_1\cap E_1')\cap\tau_{D_0}^{-1}(D_0\cap D_0')\cap\tau_{D_0'}^{-1}(D_0\cap D_0').$$

Hence

$$\{\omega\in E_1\cap E_1' : \tau^{\nu}(\omega)\neq\tau^{\nu'}(\omega)\}=(E_1\cap E_1')\cap[\tau_{D_0}^{-1}(D_0\setminus D_0')\cup\tau_{D_0'}^{-1}(D_0'\setminus D_0)]$$

and so

$$\mu(\{\omega\in E_1\cap E_1' : \tau^{\nu}(\omega)\neq\tau^{\nu'}(\omega)\})\leqq\mu(D_0\setminus D_0')+\mu(D_0'\setminus D_0)=\mu(D_0\triangle D_0').$$

For the case $n=2$ we note that if

$$\omega\in E_2\cap E_2'\cap\tau_{D_1}^{-1}[(D_1\cap D_1')\cap\tau_{D_0}^{-1}(D_0\cap D_0')]\cap\tau_{D_1'}^{-1}[(D_1\cap D_1')\cap\tau_{D_0'}^{-1}(D_0\cap D_0')]$$

then

$$\tau_{D_1}(\omega)=\tau_{D_1'}(\omega)\quad\text{and}\quad\tau_{D_0}(\tau_{D_1}(\omega))=\tau_{D_0'}(\tau_{D_1'}(\omega)),\qquad\text{so}$$

$$\{\omega\in E_2\cap E_2' : \tau^{\nu}(\omega)=\tau^{\nu'}(\omega)\}$$
$$\supseteq E_2\cap E_2'\cap\tau_{D_1}^{-1}[(D_1\cap D_1')\cap\tau_{D_0}^{-1}(D_0\cap D_0')]$$
$$\cap\tau_{D_1'}^{-1}[(D_1\cap D_1')\cap\tau_{D_0'}^{-1}(D_0\cap D_0')]$$

or

$$\{\omega\in E_2\cap E_2' : \tau^{\nu}(\omega)\neq\tau^{\nu'}(\omega)\}$$
$$\subseteq E_2\cap E_2'\cap(\tau_{D_1}^{-1}[(D_1\setminus D_1')\cup\tau_{D_0}^{-1}(D_0\setminus D_0')]$$
$$\cup\tau_{D_1'}^{-1}[(D_1'\setminus D_1)\cup\tau_{D_0'}^{-1}(D_0'\setminus D_0)]).$$

Hence

$$\mu(\{\omega\in E_2\cap E_2' : \tau^{\nu}(\omega)\neq\tau^{\nu'}(\omega)\})$$
$$\leqq\mu(D_1\setminus D_1')+\mu(D_0\setminus D_0')+\mu(D_1'\setminus D_1)+\mu(D_0'\setminus D_0)$$
$$=\mu(D_0\triangle D_0')+\mu(D_1\triangle D_1').$$

The cases $n\geqq 3$ follow in an analogous way.

Now

$$d(\tau^{\nu},\tau^{\nu})=\mu(\{\omega : \tau^{\nu}(\omega)\neq\tau^{\nu'}(\omega)\})$$
$$\leqq\sum_{n=0}^{k}\mu(E_n\setminus E_n')+\sum_{n=1}^{k}\mu(\{\omega\in E_n\cap E_n' : \tau^{\nu}(\omega)\neq\tau^{\nu'}(\omega)\})$$
$$\leqq\sum_{n=0}^{k}\mu(E_n\triangle E_n')+\sum_{n=1}^{k}\sum_{l=0}^{n-1}\mu(D_l\triangle D_l')$$
$$\leqq\sum_{n=0}^{k}\mu(E_n\triangle E_n')+\sum_{n=1}^{k}\sum_{l=0}^{n-1}\sum_{m=l+1}^{k}\mu(E_m\triangle E_m')$$

which establishes the result.

Theorem 6.6. *Let $\tau=\tau(T)$. If v is a stopping time for $(\Omega, \mathscr{B}, \mu, \tau)$ with $\int v\,d\mu=k\geqq 1$ then for any $\varepsilon>0$ there is a stopping time v' with $\int v'\,d\mu=k$ such that $\tau^{v'}$ is ergodic on $\{v'\neq 0\}$ and*

$$d(\tau^{v}, \tau^{v'})<\varepsilon.$$

Proof. Let E_n, $0\leqq n\leqq k$, be as defined for v by Theorem 4.1. As the column levels in $S^v(T)$, $v\geqq 0$, generate $\mathscr{B}$ we can find $v\geqq 0$ such that

(i) There are disjoint sets E'_n, $0\leqq n\leqq k$, which are unions of column levels in $S^v(T)$ with $\mu(E_n \Delta E'_n)<\eta(\varepsilon, k)/2$.

(ii) $\mu(B(S^v(T)))<\eta(\varepsilon, k)/2k$.

In C_1 give each column level name n if it is a subset of E'_n, $0\leqq n\leqq k$, and then modify the top k column levels so that

$$\sigma_1(i)=i+1 \pmod k, \qquad i\in\{1, 2, \ldots, k\}.$$

In C_2 give each column level name n if it is a subset of E'_n, $0\leqq n\leqq k$, and then modify the top k column levels so that

$$\sigma_2(i)=i, \qquad i\in\{1, 2, \ldots, k\}.$$

For $3\leqq i\leqq m$, give each column level in C_i name n if it is a subset of E'_n and then assign the name k to each of the top k column levels. Now let F_n be the union of those column levels now having name n, $0\leqq n\leqq k$ and define v' so that

$$\tau^{v'}\omega=\begin{cases}\omega & \omega\in F_0\\ \tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{n-1}}(\omega))\ldots)). & \omega\in F_n,\ 1\leqq n\leqq k,\end{cases}$$

where $D_{n-1}=\bigcup_{l=n}^{k} F_l$, $1\leqq n\leqq k$. Note that because of Theorem 5.4(i)

$\tau^{v'}$ is ergodic on $\{v'\neq 0\}$.

Also

$$\mu(E'_n \Delta F_n)<k\mu(A(S^v(T)))<\mu(\varepsilon, k)/2$$

and so

$$\mu(E_n \Delta F_n)<\mu(E_n \Delta E'_n)+\mu(E'_n \Delta F_n)<\eta(\varepsilon, k).$$

Hence, by Lemma 6.5, we have

$$d(\tau^{v}, \tau^{v'})<\varepsilon.$$

Theorem 6.7. *Let $\tau=\tau(T)$. If v is a stopping time for $(\Omega, \mathscr{B}, \mu, \tau)$ with $\int v\,d\mu=k\geqq 1$ then for any $\varepsilon>0$ there is a stopping time v' with $\int v'\,d\mu=k$ such that $\tau^{v'}$ is Bernoulli on $\{v'\neq 0\}$ and*

$$d(\tau^{v}, \tau^{v'})<\varepsilon.$$

Proof. Let E_n, $0 \leqq n \leqq k$, be as defined for v by Theorem 4.1. As the column levels in $S^v(T)$, $v \geqq 0$, generate $\mathscr{B}$ we can find $v \geqq 0$ such that

(i) there is a partition F_n, $0 \leqq n \leqq k$, of Ω, with each F_n a union of column levels in $S^v(T)$ and

$$\mu(E_n \triangle F_n) < \eta(\varepsilon/2, k).$$

(ii) If $S^v(T) = (C_i : 1 \leqq i \leqq m)$ then $\mu(C_i^*) < \eta(\varepsilon/2, k)/4$, $1 \leqq i \leqq m$.

(iii) $\mu(B(S^v(T))) < \eta(\varepsilon/2, k)/2k$,

(iv) C_2 has height greater than k.

Let θ be the stopping time associated with the partition F_n, $0 \leqq n \leqq k$, by Theorem 4.1. In C_1 give each column level name n if it is a subset of F_n and then modify the top k column levels so that

$$\sigma_1(i) = i + 1 \pmod{k}, \qquad i \in \{1, 2, \ldots, k\}.$$

In C_2 give each column level except the top $k+1$ levels name 0, give the $(k+1)$st column level (counting from the top) name 1 and allocate names to the top k levels so as to give

$$\sigma_2(i) = i, \qquad i \in \{1, 2, \ldots, k\}.$$

In C_3 give each column level except the top k levels name 0 and give the top k levels names which give

$$\sigma_3(i) = i, \qquad i \in \{1, 2, \ldots, k\}.$$

For $4 \leqq j \leqq m$ give the top k column levels in C_j name k and any other column level name n if it is a subset of F_n. Now let E_n' be the union of column levels now having name n, $0 \leqq n \leqq k$, and define v' so that

$$\tau^{v'}\omega = \begin{cases} \omega & \omega \in E_0' \\ \tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{n-1}}(\omega))\ldots)), & \omega \in E_n', \ 1 \leqq n \leqq k, \end{cases}$$

where $D_{n-1} = \bigcup_{l=n}^{k} E_l'$, $1 \leqq n \leqq k$. Note that the second and third columns of T_1, as defined in § 5 have heights 2 and 1 respectively and so, by Theorem 5.4(ii) $\tau^{v'}$ is Bernoulli on $\{v' \neq 0\}$. Moreover, for $0 \leqq n \leqq k$,

$$\mu(F_n \triangle E_n') < \mu(C_2^*) + \mu(C_3^*) + k\mu(A(S^v(T))) < \eta(\varepsilon/2, k)$$

which implies that

$$d(\tau^{\theta}, \tau^{v'}) < \varepsilon/2.$$

Hence

$$d(\tau^{v}, \tau^{v'}) < \varepsilon.$$

Corollary 6.8. *If τ is a Bernoulli shift of finite entropy then the set of stopping time transformations τ^ν for which τ^ν is Bernoulli is dense in the set of stopping time transformations.*

Proof. Any Bernoulli shift of finite entropy is isomorphic to $\tau = \tau(T)$ for some T [2].

References

1. Friedman, N.A.: Introduction to Ergodic Theory. New York: van Nostrand 1970
2. Friedman, N.A.: Bernoulli shifts induce Bernoulli shifts. Advances in Math. **10**, 39–48 (1970)
3. Friedman, N.A., Orstein, D.S.: On isomorphism of weak Bernoulli transformations. Advances in Math. **5**, 365–394 (1970)
4. Neveu, J.: Temps d'arrêt d'un système dynamique. Z. Wahrscheinlichkeitstheorie verw. Gebiete **13**, 81–94 (1969)
5. Ornstein, D.S.: Bernoulli shifts with infinite entropy are isomorphic. Advances in Math. **5**, 339–348 (1970)
6. Saleski, A.: Stopping times for Bernoulli automorphisms. Pacific J. Math. **52**, 547–551 (1974)
7. Shields, P.: Cutting and Independent Stacking of Intervals. Math. Systems Theory **7**, 1–4 (1973)

Received Janurary 15, 1977; in revised form July 18, 1978

STOPPING TIMES FOR MEASURE-PRESERVING TRANSFORMATIONS : SOME APPROXIMATION RESULTS

STOPPING TIMES FOR MEASURE-PRESERVING TRANSFORMATIONS: SOME APPROXIMATION RESULTS

By S. T. RICHARDSON†

[Received 24 July 1979]

1. Introduction

In this paper we prove some approximation results within the class $\mathscr{S}(\tau)$ of stopping-time transformations derived from an ergodic automorphism τ of a Lebesgue space.

If ν is a non-negative integer-valued measurable function then τ^ν belongs to $\mathscr{S}(\tau)$ if and only if τ^ν is again an automorphism, and ν is then called a stopping time. Stopping-time transformations were first discussed by Neveu in [**6**]. $\mathscr{S}(\tau)$ includes induced transformations and coincides with the positive part of the full group of τ.

It is known, see [**1**] and [**2**], that $\mathscr{S}(\tau)$ is separable and complete with respect to the uniform metric and that the entropy of τ^ν is related to that of τ via the expected value of ν. In § 5 we approximate any stopping-time transformation τ^ν by a stopping-time transformation τ^{ν_1}, which has the same entropy and is ergodic on the set $\{\nu_1 \neq 0\}$. We then deduce density results for the set of mixing or Kolmogorov stopping-time transformations using results of [**4**] and [**7**]. These approximations are based upon a construction which is given in § 4. In a previous paper [**8**] we proved that when τ is a Bernoulli shift the set of τ^ν which are also Bernoulli shifts is dense. Throughout the paper we take $\Omega = [0, 1]$, $\mathscr{B}$ to be the set of Borel subsets of Ω. μ to be Lebesgue measure on $\mathscr{B}$, and τ to be an ergodic automorphism of $(\Omega, \mathscr{B}, \mu)$.

2. Preliminaries

Let ν be a measurable mapping form $(\Omega, \mathscr{B})$ to the natural numbers $\mathbb{N}$. Then the transformation τ^ν defined by

$$\tau^\nu(\omega) = \tau^n(\omega), \quad \omega \in \{\nu = n\}, \quad n \in \mathbb{N},$$

† The author wishes to thank her supervisor Dr K. M. Wilkinson for much help and encouragement in the preparation of this paper which is part of her Ph.D. thesis at the University of Nottingham.

5388.3.43

is an automorphism of $(\Omega, \mathcal{B}, \mu)$ if and only if the sets $\{\tau^n\{\nu = n\}: n \in \mathbb{N}\}$ form a partition of Ω. If this is the case, following Neveu, we call ν a *stopping time* for $(\Omega, \mathcal{B}, \mu, \tau)$ and τ^ν a *stopping-time transformation*. Neveu [6] established the following decomposition theorem:

THEOREM 1. (i) *If $\Omega \supset D_0 \supset D_1 \ldots$ is a sequence of measurable sets such that either $D_{k-1} \neq \emptyset$ and $D_k = \emptyset$ for some $k \geqslant 1$ or $\bigcap_{n=0}^{\infty} D_n = \emptyset$ then there is a stopping time ν of $(\Omega, \mathcal{B}, \mu, \tau)$ such that*

$$\tau^\nu\omega = \begin{cases} \omega, & \text{if } \omega \in E_0 = D_0{}^c, \\ \tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{n-1}}(\omega))\ldots)), & \text{if } \omega \in E_n \text{ for } n \geqslant 1, \end{cases} \tag{2.1}$$

where $E_n = D_{n-1} \setminus D_n$ for $n \geqslant 1$, and for any $D \in \mathcal{B}$, τ_D represents the induced transformation of τ on D.

(ii) *Corresponding to any stopping time ν of $(\Omega, \mathcal{B}, \mu, \tau)$ there is a sequence of measurable sets $\Omega \supset D_0 \supset D_1 \ldots$ as in* (i) *such that τ^ν has the form* (2.1).

(iii) *If ν is a stopping time for $(\Omega, \mathcal{B}, \mu, \tau)$ and τ is ergodic then $\int \nu \, d\mu$ is either a non-negative integer or $+\infty$. If $\int \nu \, d\mu = k < \infty$ then the sets D_n with $n \geqslant 0$, defined in* (ii) *satisfy $D_{k-1} \neq \emptyset$, $D_k = \emptyset$, whereas if $\int \nu \, d\mu = +\infty$ we have $D_n \neq \emptyset$, for $n \geqslant 0$, and $\bigcap_{n=0}^{\infty} D_n = \emptyset$.*

If we let $E_0 = D_0{}^c$ and $E_i = D_{i-1} \setminus D_i$, for $i \geqslant 1$, then, for a stopping time ν with $\int \nu \, d\mu = k < \infty$, the sets $\{E_i\}_{i=0}^k$ form a partition of Ω that we call the partition associated with ν. Conversely, given a partition $\{E_i\}_{i=0}^k$ of Ω, we say that the stopping time ν defined as in (2.1) with $D_0{}^c = E_0$ and $D_i = \bigcup_{j=i+1}^k E_j$, for $1 \leqslant i \leqslant k-1$, is the stopping time corresponding to the partition $\{E_i\}_{i=0}^k$.

Note that any stopping-time transformation τ^ν with $\int \nu \, d\mu = +\infty$ can be approximated by a stopping-time transformation τ^{ν_k} with $\int \nu_k \, d\mu = k < \infty$. Indeed if $\{D_i\}_{i=1}^{\infty}$ is the sequence of sets, decreasing to $\emptyset$, associated with ν by Theorem 1 (ii), we define

$$\tau^{\nu_k}(\omega) = \begin{cases} \omega, & \text{if } \omega \in D_k, \\ \tau^\nu(\omega), & \text{if } \omega \in D_k{}^c, \end{cases}$$

and we choose k large enough so that τ^ν and τ^{ν_k} differ on a set of arbitrary small measure.

The entropy of a stopping-time transformation is related to the number of sets in the decomposition as follows: if $\int \nu \, d\mu = k < \infty$, then

$$h(\tau^\nu) = kh(\tau). \tag{2.2}$$

The entropy of τ^ν was first studied by Belinskaya in [2], where (2.2) is proved under the additional hypothesis of ergodicity of τ^ν. By

decomposition of τ^ν into its ergodic component and use of a result of Rohlin's [9] relating the entropy of an automorphism to the entropy of its ergodic components, Belinskaya's result easily leads to (2.2).

The construction in §4 involves the geometric representation of an ergodic transformation via Rohlin's theorem. Let us recall some terminology, introduced principally in [3, 4, and 8].

Let $\{L_i\colon 0 \leqslant i < h\}$ be a collection of disjoint measurable sets. We say that $C = (L_i\colon 0 \leqslant i < h)$ is a *τ-column* if $\tau(L_i) = L_{i+1}$, for $0 \leqslant i < h-1$. The sets L_i, with $0 \leqslant i < h$, are the column levels of C; C has base $B(C) = L_0$, top $A(C) = L_{h-1}$, height $h(C) = h$, and width $w(C) = \mu(L_0)$. We let $C^* = \bigcup_{i=0}^{h-1} L_i$. Note that, even though τ is an automorphism in Ω, in a τ-column C, τ is not defined within C on $A(C)$. If $F \subset L_0$, the column $(\tau^i F\colon 0 \leqslant i < h)$ is called a *subcolumn* (or *copy*) of C with base F.

A *τ-tower* T is a finite set of disjoint columns $T = (C_j\colon 1 \leqslant j \leqslant m)$; T has base $B(T) = \bigcup_{j=1}^m B(C_j)$ and top $A(T) = \bigcup_{j=1}^m A(C_j)$. We let $T^* = \bigcup_{j=1}^m C_j^*$ and denote by $\mathscr{L}(T)$ the set of column levels of T. In the sequel, when we simply say column (respectively tower) we mean τ-column (respectively τ-tower).

We say that a column C is *pure* with respect to a partition $P = \{p_i\}_{i=1}^n$ if each level of C is a subset of an atom of P. Given any column C, we *purify* C with respect to P if we divide C into subcolumns that are pure with respect to P. This is done by taking the bases of these subcolumns equal to

$$B(C) \cap \bigcap_{j=0}^{h(C)-1} \tau^{-j} p_{i_j},$$

for all choices of $i_j \in \{1, \ldots, n\}$, $0 \leqslant j < h(C)$. We purify a tower with respect to P if we purify each column in the tower.

Let us now consider a fixed stopping time ν, with associated partition $\{E_i\}_{i=0}^k$, and a τ-column C, pure with respect to $\{E_i\}_{i=0}^k$. We can then label each column level of C with respect to $\{E_i\}_{i=0}^k$, that is, we define a function n on $\mathscr{L}(C)$, taking values in $\{0, \ldots, k\}$ by

$$n(L) = n \quad \text{if } L \subset E_n, \text{ for } 0 \leqslant n \leqslant k.$$

and say that L has *name* n. Levels with name 0 are invariant under τ^ν and will not be involved in the discussion that follows. The action of τ^ν within the column C can now be described by an appropriate sequence of names. Suppose that a column level L_i has name $n \geqslant 1$. Then $\tau^\nu(L_i)$ is another column level in C (we also say $\tau^\nu(L_i)$ is defined within C) if and only if there is a sequence of indices

$$i < i_1 < i_2 < \ldots < i_n < h$$

with $\varkappa(L_{i_t}) \geqslant n-t+1$ and $1 \leqslant t \leqslant n$. Further, $\tau^{\nu}(L_i) = L_{i_n}$ if each index i_t satisfies $i_t = \min\{j\colon i_{t-1}+1 \leqslant j < h,\ \varkappa(L_j) > n-t+1\}$ for $1 \leqslant t \leqslant n$.

Clearly there may be column levels other than $A(C)$, that is, some L_i, with $0 \leqslant i < h-1$, for which $\tau^{\nu}(L_i)$ is not defined within C. For a column level L_i with $\varkappa(L) = n$, we shall define $\omega(L_i) = u$, where $1 \leqslant u \leqslant n$, if the set

$$\{i_1, \ldots, i_s\colon i < i_1 < \ldots < i_s < h,\ \varkappa(L_{i_t}) \geqslant n-t+1, 1 \leqslant t \leqslant s\}$$

is non-empty for $s \leqslant n-u$ and empty for $s \geqslant n-u+1$, and $\omega(L_i) = 0$ if $\tau^{\nu}(L_i)$ is defined within the column. So $\omega(L_i)$ counts the number of steps in the iterative definition of τ^{ν} on L_i given by Theorem 1 which are left undefined within the column containing L_i.

In the sequel, we refer to [8] for properties of the functions $\varkappa$ and ω that we quote without proof.

Any column level L in C with $\omega(L) \geqslant 1$ is called a *ν-top* in C and it can be shown that there are at most k ν-tops in any column. Furthermore, any level L in C with $\omega(L) = 0$ is mapped by some power of τ^{ν} onto a unique ν-top.

If we now have two disjoint τ-columns C_0 and C_1 with $\tau(A(C_0)) = B(C_1)$, we denote by $C_0 * C_1$ the column defined by

$$(\tau^i B(C_0)\colon 0 \leqslant i < h(C_0)+h(C_1)).$$

Similarly, given two disjoint towers $T_1 = (C_{1j}\colon 1 \leqslant j \leqslant m_1)$ and $T_2 = (C_{2l}\colon 1 \leqslant l \leqslant m_2)$ with $\tau(A(T_1)) = B(T_2)$, we then define a tower with base $B(T_1)$ and top $A(T_2)$ denoted by $T_1 * T_2$. The bases of the columns of $T_1 * T_2$ are given by

$$\{B(C_{1j}) \cap \tau^{-h(C_{1j})} B(C_{2l})\colon 1 \leqslant j \leqslant m_1,\ 1 \leqslant l \leqslant m_2\},$$

so $T_1 * T_2$ has at most $m_1 m_2$ columns. If C_0 and C_1 are both pure with respect to $\{E_i\}_{i=0}^{k}$ and if L is a ν-top for C_0, then $\tau^{\nu}(L)$ might be defined within $C_0 * C_1$. Precisely, we define at most k levels in C_1: $\{B_i(C_1)\colon 1 \leqslant i \leqslant k\}$, called the ν-bases of C_1, by

$$\begin{cases} B_1(C_1) = L_1(C_1), \\ B_i(C_1) = \tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{i-2}}(L_i(C_1)))\ldots)), \quad \text{for } 2 \leqslant i \leqslant k, \end{cases}$$

if $\tau_{D_0}(\tau_{D_1}(\ldots(\tau_{D_{i-2}}(L_i(C_1)))\ldots))$ is a column level within C_1 and where $L_i(C_1)$ is the first column level of C_1 (counting from the base) which satisfies $\varkappa(L_i(C_1)) \geqslant i$. It follows that when L is a ν-top for C_0 with $\omega(L) = u \geqslant 1$ and when $B_u(C_1)$ exists, we have

$$\tau^{\nu}(L) = B_u(C_1). \tag{2.3}$$

Note that, for a column $C = (L_i: 0 \leqslant i < h)$, pure with respect to $\{E_i\}_{i=0}^{k}$, if there exists a sequence of indices $\{i_j\}_{j=0}^{k-1}$ with $0 \leqslant i_0 < i_1 < \ldots < i_{k-1} \leqslant h-1$ and $\kappa(L_{i_j}) \geqslant k-j$, for $0 \leqslant j \leqslant k-1$, then C possesses exactly k ν-bases. However, any level in C is the image under some power of τ^{ν} of a ν-basis.

Thus the name function characterizes the action of τ^{ν} within a column C_0; further, if the column C_0 is extended by considering $C_0 * C_1$, then τ^{ν} maps ν-tops of C_0 to ν-bases of C_1 and this action is characterized by the function α.

3. The α-blocks

The definitions concerning κ, α, ν-tops, and ν-bases given in the previous paragraph depend on ν only through the partition $\{E_i\}_{i=0}^{k}$ associated with ν. We could as well have stated these definitions with respect to a partition and taken ν to be the corresponding stopping time. Since ν and $\{E_i\}_{i=0}^{k}$ are interchangeable in this way, we simply refer to κ, α functions as defined with respect to a *name partition* $\{E_i\}_{i=0}^{k}$, and similarly for ν-tops and ν-bases.

In what follows it will be convenient to impose conditions on the name structure such that the k ν-tops in a column C, pure with respect to a name partition $\{E_i\}_{i=0}^{k}$, are the top k levels. We denote these top k levels by $D(C) = (A_j(C): 1 \leqslant j \leqslant k)$, numbering from the top so that $A_1(C) = A(C)$; if $T = (C_j: 1 \leqslant j \leqslant k)$ then we let $D(T) = (D(C_j): 1 \leqslant j \leqslant k)$.

A necessary and sufficient condition for each $A_i(C)$ to be a ν-top with respect to a name partition $\{E_i\}_{i=0}^{k}$ is

$$\kappa(A_i(C)) \in \{i, i+1, \ldots, k\}, \quad \text{for } 1 \leqslant i \leqslant k. \tag{3.1}$$

In this case $\{\alpha(A_i(C)): 1 \leqslant i \leqslant k\}$ is a permutation of $\{1, 2, \ldots, k\}$. Further, we note that if two name partitions $\{E_i\}_{i=0}^{k}$ and $\{E'_i\}_{i=0}^{k}$, with associated stopping times ν and ν' respectively, both satisfy (3.1) and coincide on $C \setminus D(C)$, then the sets of ν and ν'-tops coincide. If L is a level in C with $\alpha(C) = 0$ and $\tau^{\nu}(L) = A_j(C)$ for some j, where $1 \leqslant j \leqslant k$, then $\alpha'(L) = 0$ and $\tau^{\nu'}(L) = A_j(C)$.

We now define a specific structure of names, called a *α-block*, which we shall use later to control the action of τ^{ν}.

A tower $T = (C_j: 1 \leqslant j \leqslant k)$ is a α-block with respect to a name partition $\{E_i\}_{i=0}^{k}$ if

(i) $h(C_j) = k$, for $1 \leqslant j \leqslant k$ and $w(C_j) = w(C_l)$, for $1 \leqslant j, l \leqslant k$.

(ii) the columns of T are pure with respect to $\{E_i\}_{i=0}^{k}$,

(iii) $\alpha(A_i(C_j)) = i + j - 1 \pmod{k}$, for $1 \leqslant i, j \leqslant k$.

A u-block with $k = 3$ is illustrated below in Fig. 1. Above each level we have listed its name and the value of the u-function.

$A_1(C_1)$	$u = 1$	$n = 1$	$A_1(C_2)$	$u = 2$	$n = 2$	$A_1(C_3)$	$u = 3$	$n = 3$
$A_2(C_1)$	$u = 2$	$n = 2$	$A_2(C_2)$	$u = 3$	$n = 3$	$A_2(C_3)$	$u = 1$	$n = 2$
$A_3(C_1)$	$u = 3$	$n = 3$	$A_3(C_2)$	$u = 1$	$n = 3$	$A_3(C_3)$	$u = 2$	$n = 3$
	C_1			C_2			C_3	

FIG. 1

Now suppose that there is no underlying name partition and that C is a τ-column of height k. Then we can define a name partition so as to convert C into a u-block. This is done by first dividing $A_1(C)$ into k disjoint subsets of equal measure, which we denote by $\{A_1(C)(j)\colon 1 \leqslant j \leqslant k\}$, and then defining

$$C_j = (\tau^i A_1(C)(j)\colon -(k-1) \leqslant i \leqslant 0).$$

Thus $T = (C_j\colon 1 \leqslant j \leqslant 1)$ satisfies condition (i) of the definition of a u-block. Secondly we assign names to column levels in each column C_j so that conditions (ii) and (iii) are satisfied: for $1 \leqslant j \leqslant k$, we let

$$n(A_i(C_j)) = \begin{cases} (j-1)+i, & \text{if } 1 \leqslant i \leqslant k-j+1. \\ i. & \text{if } k-j+1 < i \leqslant k. \end{cases}$$

In general, we start with an underlying name partition $\{E_i\}_{i=0}^k$ and a column C, with $h(C) \geqslant k$, pure with respect to $\{E_i\}_{i=0}^k$. We then, as above, subdivide C into k subcolumns of equal width, $\{C_j\colon 1 \leqslant j \leqslant k\}$ and modify the names on the top k levels of each C_j, *and only on these*, to yield a u-block. We call the resulting tower $U(C)$. Thus the new name partition $\{E_i\}_{i=0}^k$ defined is such that

(i) the tower $\{D(C_j)\colon 1 \leqslant j \leqslant k\}$ is a u-block with respect to $\{E_i'\}_{i=0}^k$,

(ii) $E_i \cap (\Omega \backslash D(C)^*) = E_i' \cap (\Omega \backslash D(C)^*)$, for $1 \leqslant i \leqslant k$.

We refer to $\{E_i'\}_{i=0}^k$ as the name partition corresponding to $\{E_i\}_{i=0}^k$ after formation of $U(C)$.

If $T = (C_l\colon 1 \leqslant l \leqslant m)$ is a tower with columns of equal height h, where $h \geqslant k$, which are pure with respect to a name partition $\{E_i\}_{i=0}^k$, then we let $U(T)$ denote the tower resulting after modification of each C_l to $U(C_l)$, with $1 \leqslant l \leqslant m$. Thus $U(T)$ has km columns and we refer to the new name partition as the partition corresponding to $\{E_i\}_{i=0}^k$ after formation of $U(T)$.

To outline the fundamental property of u-blocks, we consider the following situation. We have a column $C = C_0 * C_1$, where $h(C_0) \geqslant k$, $h(C_1) \geqslant k$, $w(C) > 0$, and both C_0 and C_1 are pure with respect to a

name partition $\{E_i\}_{i=0}^k$. We suppose that the k ν-bases of C_1, $\{B_l(C_1)\colon 1 \leqslant l \leqslant k\}$, exist. We form $U(C_0)$ and let $\{E_i'\}_{i=0}^k$ be the partition corresponding to $\{E_i\}_{i=0}^k$ after formation of $U(C_0)$. Note that C_1 is still pure with respect to $\{E_i'\}_{i=0}^k$. We adopt the following notation: if L is a column level in C of height h, then define

$$L(j) = \tau^{h-h(C_0)} A_1(C_0)(j), \quad \text{for } 1 \leqslant j \leqslant k.$$

Thus $L = \bigcup_{j=1}^k L(j)$. Let ν' be the stopping time associated with $\{E_i'\}_{i=0}^k$. Note that $\{B_l(C_1)\colon 1 \leqslant l \leqslant k\}$ are also the ν'-bases of C_1. We now show that the existence of a w-block at the top of C_0 controls 'uniformly' the intersection between images under $\tau^{\nu'}$ of levels in C_0 and levels in C_1.

PROPOSITION 1. *Let $C = C_0 * C_1$, $\{E_i'\}_{i=0}^k$, and ν' be as above. Let L be any level in C_0, and let L' be any level in C_1, both included in $E_0'^c$. Then there exists a unique $t > 0$ such that*

$$(\tau^{\nu'})^t L \cap L' \neq \emptyset,$$

and, for this value of t,

$$\mu((\tau^{\nu'})^t L \cap L') = w(C)/k.$$

Proof. We first show that for each i, with $1 \leqslant i \leqslant k$, k subsets of $A_i(C_0)$ are mapped by $\tau^{\nu'}$ to different ν'-bases. Precisely, by definition of a w-block, $w(A_i(C_0)(j)) = i + j - 1 \pmod k$ and thus, by (2.3),

$$\tau^{\nu'}(A_i(C_0)(j)) = B_{i+j-1}(C_1)(j),$$

where from now on, all indices are taken modulo k. For $1 \leqslant i, l \leqslant k$ fixed, there exists a unique $j = j(i, l)$ such that

$$i + j - 1 = l \pmod k,$$

and, for $j' \neq j$, $\tau^{\nu'}(A_i(C_0)(j')) \cap B_l(C_1) = \emptyset$. Hence, for any $1 \leqslant i, l \leqslant k$,

$$\mu(\tau^{\nu'}(A_i(C_0)) \cap B_l(C_1)) = \mu(B_l(C_1)(j)) = w(C)/k. \tag{3.2}$$

Now for a fixed j, with $1 \leqslant j \leqslant k$, $L(j)$ belongs to a subcolumn of C_0, pure with respect to $\{E_i'\}_{i=0}^k$, with $\{A_i(C_0)(j)\colon 1 \leqslant i \leqslant k\}$ as its ν'-tops. Hence there exists a unique $t_0 \in \mathbb{N}$, such that

$$(\tau^{\nu'})^{t_0} L(j) = A_i(C_0)(j), \tag{3.3}$$

for some i, with $1 \leqslant i \leqslant k$. Moreover, as remarked at the beginning of the paragraph, i is independent of j because, by definition of $U(C_0)$, for any j, with $1 \leqslant j \leqslant k$, the names of $\{A_i(C_0)(j)\colon 1 \leqslant i \leqslant k\}$ satisfy (3.1).

Hence (3.3) holds for any j, with $1 \leqslant j \leqslant k$, and

$$(\tau^{\nu'})^{t_0}L = \bigcup_{j=1}^{k} (\tau^{\nu'})^{t_0}L(j) = \bigcup_{j=1}^{k} A_i(C_0)(j) = A_i(C_0). \tag{3.4}$$

Finally, as C_1 is pure with respect to $\{E'_i\}_{i=0}^k$, there exists a unique l, with $1 \leqslant l \leqslant k$, and a unique $t_1 \in \mathbb{N}$ such that

$$(\tau^{\nu'})^{t_1}B_l(C_1) = L'. \tag{3.5}$$

Letting $t = t_0 + t_1 + 1$, we have, by (3.3), (3.4), and (3.5),

$$(\tau^{\nu'})^{t}L \cap L' = (\tau^{\nu'})^{t_1}[(\tau^{\nu'})A_i(C_0) \cap B_l(C_1)],$$

for unique $1 \leqslant i, l \leqslant k$, and the result follows from (3.2).

Note that Proposition 1 still holds if, instead of supposing that C_1 is pure with respect to $\{E_i\}_{i=0}^k$, we only suppose that the first $h(C_1) - k$ levels of C_1 are pure with respect to $\{E_i\}_{i=0}^k$ whilst $D(C_1)$ is partitioned by $\{E_i\}_{i=0}^k$ into any number of ω-blocks.

If we now have n disjoint τ-columns, satisfying $\tau(A(C_j)) = B(C_{j+1})$, where $0 \leqslant j < n$, we denote by $C = C_0 * C_1 * \ldots * C_n$ the column defined by $(\tau^i B(C_0)\colon 0 \leqslant i < \sum_{j=0}^{n} h(C_j))$. If $h(C_j) \geqslant k$ and each C_j, with $0 \leqslant j \leqslant n$, is pure with respect to a name partition $\{E_i\}_{i=0}^k$, we construct several ω-blocks by forming successively

$$\begin{aligned}
T_0 &= U(C_0),\\
T_1 &= U(T_0 * C_1),\\
&\vdots\\
T_l &= U(T_{l-1} * C_l), \quad \text{for } 1 \leqslant l \leqslant n-1,\\
&\vdots\\
T_n &= T_{n-1} * C_n.
\end{aligned}$$

By abuse of notation we sometimes denote T_n by $U(C_0 * C_1 * \ldots * C_n)$ and, when there is no possible confusion about the determination of the C_j $(0 \leqslant j \leqslant n)$, we simply say that we have formed independent ω-blocks within C. In this process, at each stage, we define a new name partition $\{E_i^{(l)}\}_{i=0}^k$, for $1 \leqslant l \leqslant n$, with $\{E_i^{(1)}\}_{i=0}^k$ the name partition corresponding to $\{E_i\}_{i=0}^k$ after formation of $U(C_0)$, and $\{E_i^{(l+1)}\}_{i=0}^k$ the name partition corresponding to $\{E_i^{(l)}\}_{i=0}^k$ after formation of $U(T_{l-1} * C_l)$, for $1 \leqslant l \leqslant n-1$. Note that the names of levels in $D(C_n)$ have not been modified. We refer to $\{E_i^{(n)}\}_{i=0}^k$ as the partition corresponding to $\{E_i\}_{i=0}^k$ after formation of $U(C_0 * \ldots * C_n)$. We can

describe this construction as follows. For $0 \leqslant l \leqslant n-1$, T^l consists of k^{l+1} columns, pure with respect to $\{E_i^{(l)}\}_{i=0}^k$, and $D(T_l)$ is a set of k^l $\mathcal{U}$-blocks. T_n consists of k^n columns, pure with respect to $\{E_i^{(n)}\}_{i=0}^k$, and $\{E_i^{(n)}\}_{i=0}^k$ coincides with $\{E_i\}_{i=0}^k$ on $\Omega\backslash(\bigcup_{l=0}^{n-1} D(C_l)^*)$. For an illustration of the formation of independent $\mathcal{U}$-blocks in the case where $k = 3$ and $n = 3$, see Fig. 2.

We now state a generalization of Proposition 1 which can be proved by induction.

PROPOSITION 2. *Let $C = C_0 * C_1 * \ldots * C_n$ be a column, pure with respect to a name partition $\{E_i\}_{i=0}^k$, with $w(C) > 0$ and $h(C_j) \geqslant k$, for $0 \leqslant j \leqslant h$. Suppose also that C_n possesses k ν-bases with respect to $\{E_i\}_{i=0}^k$. Let $\{E_i'\}_{i=0}^k$ be the partition corresponding to $\{E_i\}_{i=0}^k$ after formation of $U(C_0 * \ldots * C_n)$, and let ν' be the stopping time associated with $\{E_i'\}_{i=0}^k$. Then, for any level L in C_{i_0} and level L' in C_{i_1}, where $0 \leqslant i_0 < i_1 \leqslant n$ and both levels are included in $E_0'^c$, there exist t_j, with $1 \leqslant j \leqslant k^{i_1 - i_0 - 1}$, such that*

$$\mu([\bigcup_j (\tau^{\nu'})^{t_j} L] \cap L') = \frac{w(C)}{k},$$

the union being taken over j ranging from 1 to $k^{i_1 - i_0 - 1}$.

In the sequel, we adopt the notation

$$C = [C_0, C_1, \ldots, C_n]$$

if $C = c_0 * c_1 * \ldots * c_n$, where c_i is a copy of C_i for each i. Forming independent $\mathcal{U}$-blocks within C is then taken to mean forming $U(c_0 * \ldots * c_n)$. We conclude this section by quoting two results that we shall use in the next section. Suppose that $P = \{p_i\}_{i \in I}$ is a partition of Ω, with I a finite set of indices, and that σ is an ergodic automorphism. The following theorem is referred to as the strong Rohlin theorem.

THEOREM 2. *Given n, a positive integer, and $\varepsilon > 0$, there exists a σ-column C of height n and base B such that*

(i) $\mu(C^*) > 1 - \varepsilon$.

(ii) $d(P/B) = d(P)$,

where

$$d(P) = \{\mu(p_i)\}_{i \in I} \quad \text{and} \quad d(P/B) = \left\{\frac{\mu(p_i \cap B)}{\mu(B)}\right\}_{i \in I}.$$

For a set $A \in \bigvee_{i=0}^{m-1} \sigma^{-i} P$, with $m \geqslant 1$ and

$$A = p_{i_0} \cap \sigma^{-1} p_{i_1} \cap \ldots \cap \sigma^{-(m-1)} p_{i_{m-1}},$$

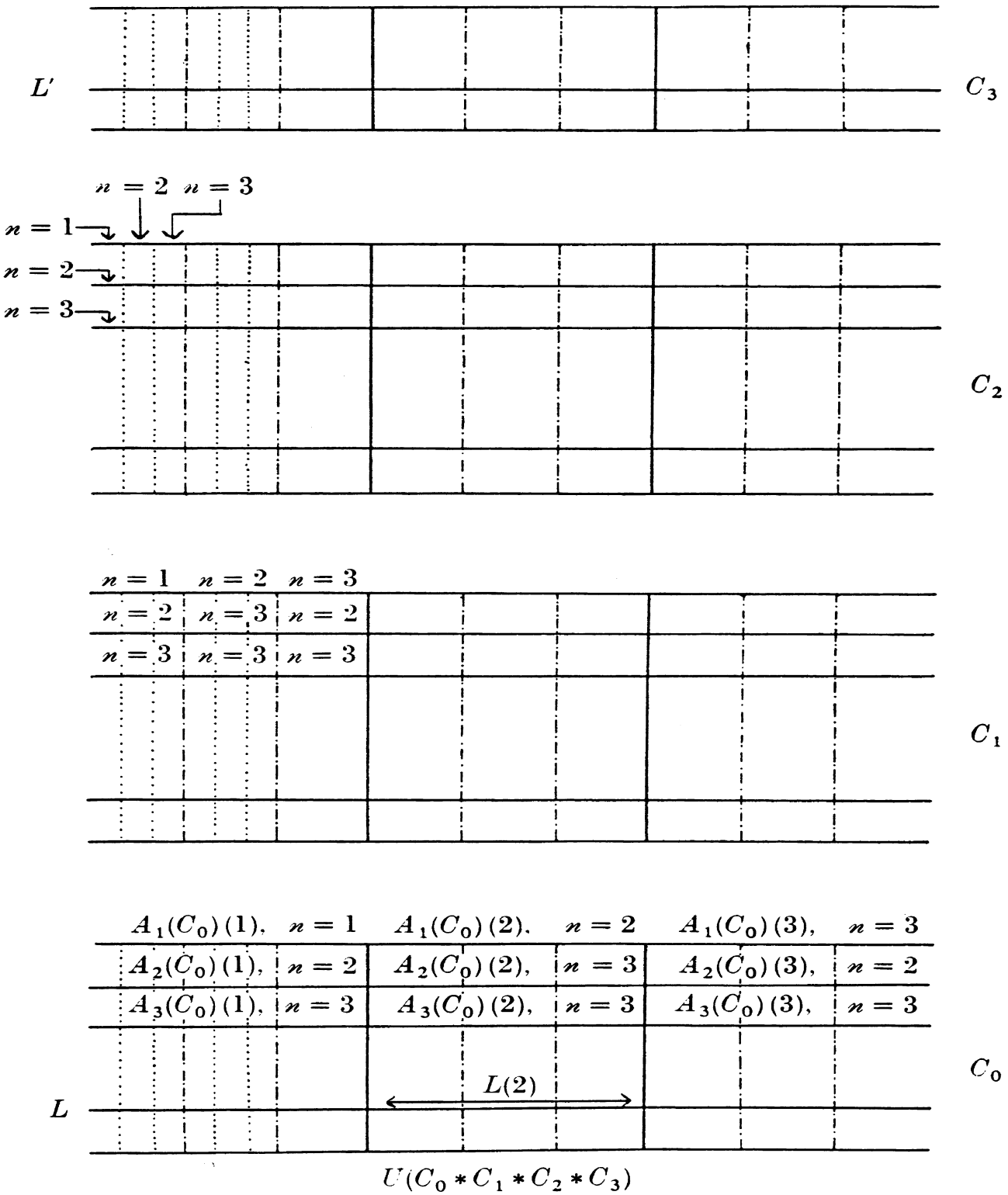


FIG. 2. Forming independent α-blocks in $C = C_0 * C_1 * C_2 * C_3$.

we call the sequence $(i_0, \ldots, i_{m-1})$ the *P-m-name* of A with respect to σ. As a consequence of the ergodic theorem, the following holds: for any $\varepsilon > 0$, there exists $N > 0$ such that for $m \geqslant N$, there is a collection $\mathscr{E}_m$ of sets in $\bigvee_{i=0}^{m-1} \sigma^{-i}P$ of total measure at least $1-\varepsilon$ such that for all $p_i \in P$ and $A \in \mathscr{E}_m$,

$$|f_A(i, m) - \mu(p_i)| < \varepsilon,$$

where $f_A(i, m)$ is the relative frequency of the index i in the P-m-name of A with respect to σ.

4. Construction

Throughout this section ν is a given stopping time for $(\Omega, \mathscr{B}, \mu, \tau)$ with $\int \nu \, d\mu = k \geqslant 1$ and associated partition $\{E_i\}_{i=0}^{k}$.

We construct a stopping time ν', with $\int \nu' d\mu = k$, for which the automorphism $\tau^{\nu'}$, whilst remaining close to τ^{ν}, is ergodic on the set $\{\nu' \neq 0\}$. This involves constructing a sequence of τ-towers $\{T_r\}_{r \geqslant 1}$ and of stopping times $\{\nu^{(r)}\}_{r \geqslant 1}$. As in [4] where Friedman and Ornstein induced a mixing transformation, the recurrence step is provided by a Rohlin's theorem, though here we need to use the strong form of Rohlin's theorem (Theorem 2). The stopping time ν' is defined as the limit of the sequence $\{\nu^{(r)}\}_{r \geqslant 1}$, where at the rth stage we modify the existing name partition by introducing independent α-blocks and take $\nu^{(r)}$ to be the stopping time associated with this modified partition.

We shall use the following notation. $\{S_r\}_{r \geqslant 0}$ is a decreasing sequence of sets formed inductively in the construction, and $\tau_r = \tau_{S_r}$, the induced transformation on S_r. For each $r \in \mathbb{N}$, the dyadic intervals of order r induce on S_r a partition denoted by Q_r, that is

$$Q_r = \{S_r \cap I_{r,j} \colon 1 \leqslant j \leqslant 2^r\},$$

where $I_{r,j} = [(j-1)2^{-r}, j2^{-r})$, for $1 \leqslant j \leqslant 2^r$ and $r \geqslant 1$. We shall purify the towers T_r, where $S_r = T_r^*$, with respect to Q_r in order that the set of levels of the successive towers generates the Borel σ-algebra.

The construction involves a decreasing sequence of positive real numbers $\{\eta_r\}_{r \geqslant 1}$, where $\eta_1 < 1$ is given and $\{\eta_r\}_{r \geqslant 2}$ are determined inductively. Later, in §5, further conditions will be imposed on the sequence $\{\eta_r\}_{r \geqslant 1}$ to prove approximation results.

The first step of the induction process constructs a τ-tower T_1, pure with respect to $\{E_i\}_{i=0}^{k}$ and such that the top k levels in each column have, after modification if necessary, name k.

Let $S_0 = \Omega$. Choose a positive integer $t(1)$ satisfying

$$t(1) > 2k/\eta_1. \tag{4.1}$$

and apply Rohlin's theorem to the ergodic transformation $\tau_0 = \tau$ to find $F_1 \subset S_0$ which has $t(1)$ disjoint images under τ_0 and is such that

$$\mu\left(\bigcup_{i=0}^{t(1)-1} \tau_0{}^i F_1\right) > \mu(S_0) - \eta_1. \tag{4.2}$$

Next we purify the τ_0-column $(\tau_0{}^i F_i\colon 0 \leqslant i < t(1))$ with respect to the partition $\{E_i\}_{i=0}^k$, and so obtain a tower T_1 where each column level $L \in \mathscr{L}(T_1)$ is given a unique name by

$$\nu(L) = i \quad \text{if and only if} \quad L \subset E_i, \quad \text{for } 0 \leqslant i \leqslant k.$$

Suppose that T_1 has m_1 columns, say, $T_1 = (C_{1j}\colon 1 \leqslant j \leqslant m_1)$, with bases $B_{1j} = B(C_{1j})$. Let w_1 be the width of the smallest column of T_1 and let $H_1 = h(C_{1j}) = t(1)$, where $1 \leqslant j \leqslant m_1$. Note that $H_1 \geqslant k$. We now change to k the names of all levels in $D(T_1)$. This defines a new name function ν_1 on $\mathscr{L}(T_1)$. Precisely, for $L \in T_1 \setminus D(T_1)$, we have $\nu_1(L) = \nu(L)$, whilst for $L \in D(T_1)$, $\nu_1(L) = k$. Letting $S_1 = T_1^*$, we define a modified name partition $G^{(1)} = \{G_i^{(1)}\}_{i=0}^k$ by

$$G_i^{(1)} = \bigcup\{L \in \mathscr{L}(T_1)\colon \nu_1(L) = i\} \cup (E_i \cap (\Omega \setminus S_1)), \quad \text{for } 0 \leqslant i \leqslant k.$$

Note that

$$\mu(E_i \Delta G_i^{(1)}) < 2k\mu(F_1). \quad \text{for } 0 \leqslant i \leqslant k,$$

and so, by (4.1),

$$\mu(E_i \Delta G_i^{(1)}) < \eta_1, \quad \text{for } 0 \leqslant i \leqslant k. \tag{4.3}$$

We finally define the first modified stopping time $\nu^{(1)}$ to be the stopping time corresponding to the partition $\{E_i^{(1)}\}_{i=0}^k$ where

$$E_0^{(1)} = G_0^{(1)} \cup (\Omega \setminus S_1),$$
$$E_i^{(1)} = G_i^{(1)} \cap S_1, \quad \text{for } 1 \leqslant i \leqslant k.$$

In other words, when considering $\nu^{(1)}$, we reduce our attention to S_1.

We now describe the general induction step which differs from the construction of T_1.

Suppose that at stage r, where $r \geqslant 1$, T_r is a tower with m_r columns, say $T_r = (C_{rj}\colon 1 \leqslant j \leqslant m_r)$ with bases $B_{rj} = B(C_{rj})$, and $F_r = \bigcup_{j=1}^{m_r} B_{rj}$. Let $S_r = T_r^*$, let w_r be the width of the smallest level in T_r, and let H_r be the common height of all columns of T_r. Suppose also that the ν_r-name of levels in $D(T_r)$ is k.

Define P_r to be the partition of F_r given by $P_r = \{B_{rj}\}_{j=1}^{m_r}$, and let $\sigma_r = \tau_{F_r}$. Choose

$$\eta_{r+1} < w_r / 2^{r+1} H_r. \tag{4.4}$$

Since σ_r is ergodic there exists $N_r > 0$ such that for $n \geqslant N_r$, there is a collection $\xi_{r,n}$ of atoms of $\bigvee_{i=0}^{n-1} \sigma_r^{-i} P_r$ of total measure at least $\mu(F_r) - \frac{1}{2}\eta_{r+1}$ and such that, for all $B_{rj} \in P_r$ and $A \in \xi_{r,n}$,

$$|f_A(B_{rj}, n) - \mu(B_{rj})| < \tfrac{1}{2}\eta_{r+1}. \tag{4.5}$$

Recall that $f_A(B_{rj}, n)$ is a relative frequency of occurrence of the index j in the $P_r - n$-name of A with respect to σ_r. Inequalities (4.4) and (4.5) and the definition of w_r imply that, for any $n \geqslant \max(N_r, 2/\eta_{r+1})$ and any set $A \in \xi_{r,n}$, $f_A(B_{rj}, n) \geqslant 2/n$ for any j, where $1 \leqslant j \leqslant m_r$.

Now we choose $t(r+1)$ a positive integer such that

$$t(r+1) > \max(N_r, 4k/\eta_{r+1}), \tag{4.6}$$

and apply the strong Rohlin theorem on F_r to find $F_{r+1} \subset F_r$, which has $t(r+1)$ disjoint images under σ_r and is such that

$$d\left(\bigvee_{i=0}^{t(r+1)-1} \sigma_r^{-i} P_r \Big/ F_{r+1}\right) = d\left(\bigvee_{i=0}^{t(r+1)-1} \sigma_r^{-i} P_r\right) \tag{4.7}$$

and

$$\mu\left(F_r \Big\setminus \bigcup_{i=0}^{t(r+1)-1} \sigma_r^{i} F_{r+1}\right) < \frac{\eta_{r+1}}{2t(r)}. \tag{4.8}$$

Let $T_{r,1}$ be the τ_r-tower with base F_{r+1} and columns pure with respect to $\mathscr{L}(T_r)$. For $0 \leqslant i \leqslant t(r+1)-1$ and $1 \leqslant j \leqslant m_r$, (4.7) implies that $\mu(\sigma_r^{-i} B_{rj} \cap F_r) = \mu(B_{rj})\mu(F_r) > 0$. In other words, for any $0 \leqslant i \leqslant t(r+1)-1$, $\sigma_r^{i} F_{r+1}$ intersects the base of any column of T_r. Let $T_r^{(i)}$ be the tower with columns consisting of the subcolumns of C_{rj} with base $\sigma_r^{i} F_{r+1} \cap B_{rj}$, for $1 \leqslant j \leqslant m_r$. If we follow the successive images of F_{r+1} under $\tau_r = \tau_{S_r}$, we see that first F_{r+1} climbs through the subtower $T_r^{(0)}$. Further, note that if $\omega \in A(T_r^{(0)})$, then $\tau_r(\omega) \subset F_r$, and thus $\tau_r(A(T_r^{(0)})) = \tau_{F_r}(F_{r+1}) = \sigma_r(F_{r+1})$. Hence next, under τ_r, F_{r+1} climbs through the levels of $T_r^{(1)}$, and so on. Consequently

$$T_{r,1} = T_r^{(0)} * T_r^{(1)} * \ldots * T_r^{(t(r+1)-1)}.$$

So each column of $T_{r,1}$ is a product of $t(r+1)$ columns which are copies of the columns of T_r. A typical column C of $T_{r,1}$ can thus be written as

$$C = [C_{ri_0}, \ldots, C_{ri_{t(r+1)-1}}]. \tag{4.9}$$

where $1 \leqslant i_l \leqslant m_r$ and $0 \leqslant l \leqslant t(r+1)-1$, and its base is then

$$F_{r+1} \cap B_{ri_0} \cap \sigma_r^{-1} B_{ri_1} \cap \ldots \cap \sigma_r^{-(t(r+1)-1)} B_{ri_{t(r+1)-1}}.$$

We call a column of $T_{r,1}$ *regular* if there appear at least two distinct

copies of each of the C_{rj}, for $1 \leqslant j \leqslant m_r$, in (4.9). We now delete from $T_{r,1}$ all columns which are not regular and call the resulting tower $T_{r,2}$.

LEMMA 1. $\mu(T^*_{r,1}) - \mu(T^*_{r,2}) < \frac{1}{2}\eta_{r+1}$.

Proof. Let R_r be the union of all bases of the columns of $T_{r,1}$ which are not regular. Then the choice of $t(r+1)$ in (4.6) implies that if a base

$$B = F_{r+1} \cap B_{ri_0} \cap \ldots \cap (\sigma_r)^{-(t(r+1)-1)} B_{ri_{t(r+1)-1}}$$

of $T_{r,1}$ belongs to R_r then

$$B_{ri_0} \cap \ldots \cap (\sigma_r)^{-(t(r+1)-1)} B_{ri_{t(r+1)-1}}$$

belongs to $\xi^c_{r,t(r+1)}$. Thus

$$R_r \subset \xi^c_{r,t(r+1)} \cap F_{r+1},$$

and since $\xi^c_{r,t(r+1)}$ is a union of atoms of $\bigvee_{i=0}^{t(r+1)-1} (\sigma_r)^{-i} P_r$. (4.7) then implies that

$$\mu(R_r) \leqslant \mu(\xi^c_{r,t(r+1)})\mu(F_{r+1}) < \tfrac{1}{2}\eta_{r+1}\mu(F_{r+1}),$$

so. finally.

$$\mu(T^*_{r,1}) - \mu(T^*_{r,2}) < \tfrac{1}{2}\eta_{r+1}.$$

We then purify $T_{r,2}$ with respect to Q_r, the dyadic intervals of order r. and obtain a tower $T_{r,3}$. The columns of $T_{r,3}$ are subcolumns of the columns of $T_{r,2}$ and so. when considered as products of copies of columns of T_r, are regular.

We now form independent $\mathscr{u}$-blocks. the r-$\mathscr{u}$-blocks, within each column of $T_{r,3}$. This defines the tower T_{r+1}. So $T^*_{r+1} = T^*_{r,3}$ and $D(T_{r+1})^* = D(T_{r,3})^*$. We let $G^{(r+1)} = \{G_i^{(r+1)}\}_{i=0}^k$ be the partition corresponding to $G^{(r)}$ after formation of independent $\mathscr{u}$-blocks within each column of $T_{r,3}$, and we let $\mathscr{n}_{r+1}$ be the new name function associated with it. The tower T_{r+1} is thus pure with respect to $G^{(r+1)}$.

The set J_r of the levels of $T_{r,3}$ whose $\mathscr{n}_r$-name has been altered by the r-$\mathscr{u}$-blocks, is a subset of $\{D(C_{r,j})\colon 1 \leqslant j \leqslant m_r\}$ and so is included in $D(T_r)^*$. Since all levels in $D(T_r)$ have an $\mathscr{n}_r$-name equal to k. and since in forming $\mathscr{u}$-blocks we never give a name equal to 0, it follows that

$$G_0^{(r+1)} = G_0^{(r)},$$
$$G_i^{(r+1)} \supset G_i^{(r)}, \quad \text{for } 1 \leqslant i \leqslant k-1.$$
$$G_k^{(r+1)} \subset G_k^{(r)}.$$

Note that the set J_r is disjoint from $D(T_{r,3})^*$ and that since $D(T_{r,3}) \subset D(T_r)$, the α_{r+1}-name of the levels in $D(T_{r,3})$ is k. It is also clear that the formation of the r-α-blocks has not altered the names of the levels of $T_{r,3}$ included in any of the i-α-blocks, for $i \leqslant r-1$, as they are disjoint from $D(T_{i,3})^* = D(T_{i+1})^*$ and thus disjoint from $D(T_r)$. Now

$$\mu(G_i^{(r)} \Delta G_i^{(r+1)}) < 2k\mu(F_r),$$

and so, by our choice of $t(r)$,

$$\mu(G_i^{(r)} \Delta G_i^{(r+1)}) < \eta_r. \tag{4.10}$$

We finally define $S_{r+1} = T_{r+1}^*$ and let $\nu^{(r+1)}$ be the stopping time corresponding to the partition $\{E_i^{(r+1)}\}_{i=0}^k$ where

$$E_0^{(r+1)} = G^{(r+1)} \cup (\Omega \backslash S_{r+1}),$$
$$E_i^{(r+1)} = G_i^{(r+1)} \cap S_{r+1}, \quad \text{for } 1 \leqslant i \leqslant k.$$

This completes the induction step.

We are now ready to define the limiting stopping time ν'. First note that the monotonic properties of the sequence $\{G^{(r)}\}_{r \geqslant 1}$ imply that if we set $G_i = \bigcup_{r=1}^{\infty} G_i^{(r)}$, for $0 \leqslant i \leqslant k-1$, and $G_k = \bigcap_{r=1}^{\infty} G_k^{(r)}$, then $G = \{G_i\}_{i=0}^k$ is a partition of Ω. Letting $S = \bigcap_{r=1}^{\infty} S_r$, we can then define a partition $\{E_i'\}_{i=0}^k$ of Ω by

$$E_0' = G_0 \cup (\Omega \backslash S),$$
$$E_i' = G_i \cap S, \quad \text{for } 1 \leqslant i \leqslant k,$$

and we take ν' to be the stopping time corresponding to the partition $\{E_i'\}_{i=0}^k$.

Note that as $\{S_r\}_{r \geqslant 0}$ is a decreasing sequence of sets, $S_0 = \Omega$. Also

$$\mu(S) > 1 - \sum_{r=1}^{\infty} \eta_r \tag{4.11}$$

since, by (4.2), $\mu(S_0 \backslash S_1) < \eta_1$, and, by (4.8) and Lemma 1,

$$\mu(S_r \backslash S_{r+1}) \leqslant (\mu(S_r) - \mu(T_{r,1}^*)) + (\mu(T_{r,1}^*) - \mu(T_{r,2}^*))$$
$$\leqslant \tfrac{1}{2}\eta_{r+1} + \tfrac{1}{2}\eta_{r+1}.$$

LEMMA 2.

$$\mu(E_0 \Delta E_0') < 2 \sum_{r=1}^{\infty} \eta_r,$$

$$\mu(E_i \Delta E_i') < 3 \sum_{r=1}^{\infty} \eta_r, \quad \textit{for } 1 \leqslant i \leqslant k.$$

Proof.

$$\mu(E_0 \Delta E_0') \leqslant \mu(E_0 \Delta G_0^{(1)}) + \mu(\Omega \backslash S) < \eta_1 + \sum_{r=1}^{\infty} \eta_r$$

by (4.3) and (4.11). For $1 \leqslant i \leqslant k-1$.

$$E_i' = G_i \cap S \quad \text{and} \quad G_i = G_i^{(1)} \cup \bigcup_{r=1}^{\infty} (G_i^{(r+1)} \backslash G_i^{(r)});$$

thus

$$\mu(E_i \Delta E_i') \leqslant \mu(E_i \Delta G_i^{(1)}) + \sum_{r=1}^{\infty} \mu(G_i^{(r+1)} \backslash G_i^{(r)}) + \mu(\Omega \backslash S)$$

$$< \eta_1 + 2 \sum_{r=1}^{\infty} \eta_r,$$

by (4.3), (4.10), and (4.11). A similar result can be proved for $\mu(E_k \Delta E_k')$. Note that the lemma implies. in particular, that

$$\mu(E_k') > \mu(E_k) - 3 \sum_{r=1}^{\infty} \eta_r. \tag{4.12}$$

5. Approximation results

The metric chosen here will be the uniform metric d defined by

$$d(\tau, \tau') = \mu\{\omega: \tau(\omega) \neq \tau'(\omega)\}.$$

where τ and τ' are both automorphisms of $(\Omega, \mathcal{B}, \mu)$.

We shall use the following fact:

LEMMA 3 [8]. *Let ν and ν' be two stopping times for $(\Omega, \mathcal{B}, \mu, \tau)$, let $\int \nu \, d\mu = \int \nu' \, d\mu = k$. with $\{E_i\}_{i=0}^k$ and $\{E_i'\}_{i=0}^k$ their respective associated partitions. For any $\varepsilon > 0$, if $\mu(E_i \Delta E_i') < \varepsilon / k^2(k+2)$, then $d(\tau^\nu, \tau^{\nu'}) < \varepsilon$.*

We now give a preparatory result based on the following lemma which is due to Knopp.

LEMMA 4. *Let $(X, \mathcal{F}, m)$ be a probability space where $\mathcal{F}$ is the σ-algebra generated by a ring $\mathcal{R}$ of subsets of X. Suppose that there is a subset $\mathcal{I}$ of $\mathcal{R}$ such that every $R \in \mathcal{R}$ is a finite disjoint union of sets in $\mathcal{I}$. Then if $F \in \mathcal{F}$ is such that*

$$m(F \cap I) \geqslant c m(I),$$

for any $I \in \mathcal{I}$, where $c > 0$ is a constant independent of I, it follows that $m(F) = 1$.

We use the following notation: μ' represents the normalized Lebesgue measure on the set $\{\nu' \neq 0\}$, and $\mathcal{L}_r^+ = \{L \in \mathcal{L}(T_r): L \subset G_0^{(r)c}\}$.

COROLLARY 1. *If for any* $r \geqslant 1$ *and* $L_1, L_2 \in \mathscr{L}_r^+$, *we have*

$$\mu'\left(\left(\bigcup_{i=-\infty}^{\infty} (\tau^{\nu'})^i L_1\right) \cap L_2\right) \geqslant c\mu'(L_2), \tag{5.1}$$

where c *is independent of* L_2, *then, for any* $L \in \mathscr{L}_r^+$,

$$\mu'\left(\bigcup_{i=-\infty}^{\infty} (\tau^{\nu'})^i L\right) = 1.$$

Proof. Fix $r \geqslant 1$ and let $L \in \mathscr{L}_r^+$. Note that

$$\{\nu' \neq 0\} = G_0{}^c \cap S = (G_0^{(r)})^c \cap S.$$

If we let $\mathscr{R}$ be the ring of finite disjoint unions of elements of $\mathscr{L}_p^+$, where $p \geqslant r$, then it follows by construction that $\mathscr{R} \cap S$ generates the Borel σ-algebra restricted to $\{\nu' \neq 0\}$. Replacing, if necessary, L by one of its copies in T_p and using (5.1), we see that for any $L' \in \mathscr{L}_p^+$, with $p \geqslant r$,

$$\mu'\left(\left(\bigcup_{i=-\infty}^{\infty} (\tau^{\nu'})^i L\right) \cap L'\right) \geqslant c\mu'(L'). \tag{5.2}$$

The result now follows from Lemma 4.

We now prove our main result which shows that for a suitable choice of $\{\eta_r\}_{r\geqslant 1}$, the stopping-time transformation $\tau^{\nu'}$ is ergodic on $\{\nu' \neq 0\}$, has the same entropy as τ^{ν}, and is arbitrarily close to τ^{ν}.

THEOREM 3. *Let* τ *be an ergodic automorphism of* $(\Omega, \mathscr{B}, \mu)$, *and let* ν *be a stopping time for* $(\Omega, \mathscr{B}, \mu, \tau)$ *with* $\int \nu \, d\mu = k$, *where* k *is a positive integer. Then for any* $\varepsilon > 0$, *there exists a stopping time* ν' *for* $(\Omega, \mathscr{B}, \mu, \tau)$ *such that the following statements hold:*

(i) $\tau^{\nu'}$ *is ergodic in* $\{\nu' \neq 0\}$;

(ii) $h(\tau^{\nu'}) = h(\tau^{\nu})$;

(iii) $d(\tau^{\nu}, \tau^{\nu'}) < \varepsilon$.

Proof. Let $\{E_i\}_{i=0}^k$ be the partition associated with ν. By Theorem 1 (iii), E_k is a non-empty set and so there exists $\delta > 0$ such that $\mu(E_k) > \delta$. We now choose a decreasing sequence of positive numbers $\{\eta_r'\}_{r\geqslant 1}$ satisfying

$$\sum_{r=1}^{\infty} \eta_r' < \min(\delta/6, \varepsilon/3k^2(k+2)). \tag{5.3}$$

Next we construct T_1 as in §4 for $\eta_1 = \eta_1'$. We then choose $\eta_2 < \eta_2'$ satisfying (4.4) and construct T_2, and so on. That is, in general, we

choose η_{r+1} such that

$$\eta_{r+1} < \min(\eta'_{r+1}, w_r/2^{r+1}H_r), \quad \text{where } r \geqslant 1, \tag{5.4}$$

and define the stopping time ν' to be as constructed in § 4 for that choice of $\{\eta_r\}_{r\geqslant 1}$.

(i) To prove that $\tau^{\nu'}$ is ergodic in $\{\nu' \neq 0\}$, we first show that for any $L \in \mathscr{L}_r^+$, with $r \geqslant 1$,

$$\mu'\left(\bigcup_{i=-\infty}^{\infty} (\tau^{\nu'})^i L\right) = 1. \tag{5.5}$$

By Corollary 1, it will be sufficient to show that for any $L, L' \in \mathscr{L}_r^+$,

$$\mu'\left(\left(\bigcup_{i=-\infty}^{\infty} (\tau^{\nu'})^i L\right) \cap L'\right) \geqslant c\mu'(L'),$$

with a constant independent of L'.

So suppose that $L \in C_{rl_0}$ and $L' \in C_{rl_1}$, where $1 \leqslant l_0, l_1 \leqslant m_r$. Now L' is distributed as distinct copies among the columns of $T_{r,3}$; precisely, we write

$$L' \cap S_{r+1} = \bigcup_{i\in I} l'_i,$$

where l'_i is a copy of L' in a column of $T_{r,3}$ and I is a finite index set. By construction, for any $i \in I$, in the column of $T_{r,3}$ containing l'_i, there is a copy of C_{rl_0}, that we denote by l_i, above or below the copy of C_{rl_1} containing l'_i.

Recall that we have formed independent α-blocks within each column of $T_{r,3}$ and that $\nu^{(r+1)}$ is the associated stopping time. Therefore we can apply Proposition 2 to each l_i and l'_i. For the sake of simplification, we shall include more powers of $\tau^{\nu(r+1)}$ than those given by Proposition 2 and hence we have the following inequality:

$$\mu\left(\left(\bigcup_{j=-H_{r+1}}^{H_{r+1}} (\tau^{\nu(r+1)})^j l_i\right) \cap l'_i\right) \geqslant \mu(l'_i)/k.$$

Now regrouping over $i \in I$, we obtain

$$\mu\left(\left(\bigcup_{j=-H_{r+1}}^{H_{r+1}} (\tau^{\nu(r+1)})^j L \cap S_{r+1}\right) \cap (L' \cap S_{r+1})\right) \geqslant \mu(L' \cap S_{r+1})/k. \tag{5.6}$$

Now for $r \geqslant 1$ and by the choice of η_{r+1} in (5.4),

$$\mu(S_{r+1} \setminus S) < \sum_{i=r+1}^{\infty} \eta_{i+1} < \sum_{i=r+1}^{\infty} \frac{w_i}{H_i} \frac{1}{2^{i+1}}.$$

and so

$$\mu(S_{r+1}\backslash S) < w_{r+1}/4H_{r+1}. \tag{5.7}$$

Hence writing $L \cap S_{r+1}$ as $L \cap S_{r+1} = (L \cap S) \cup (L \cap (S_{r+1}\backslash S))$ and similarly for $L' \cap S_{r+1}$, we deduce, from (5.6) and (5.7), that

$$\begin{aligned}\mu\left(\left(\bigcup_{j=-H_{r+1}}^{H_{r+1}} (\tau^{\nu(r+1)})^j (L \cap S)\right) \cap (L' \cap S)\right) \\ \geqslant (\mu(L' \cap S_{r+1})/k) - (2H_{r+1}+3)\mu(S_{r+1}\backslash S) \\ \geqslant (\mu(L' \cap S_{r+1})/k) - w_{r+1}. \end{aligned} \tag{5.8}$$

We now notice that, by (4.6),

$$w_{r+1} < \frac{1}{t(r+1)} < \frac{\eta_{r+1}}{4k} < \frac{w_r}{4k}. \tag{5.9}$$

and that as L' is a level in T_r,

$$\begin{aligned}\mu(L' \cap S_{r+1}) &> \mu(L') - \eta_{r+1} \\ &> \mu(L') - \mu(L')/2^{r+1}. \end{aligned} \tag{5.10}$$

Combining (5.9), (5.10), and (5.8), we then have

$$\mu\left(\left(\bigcup_{j=-H_{r-1}}^{H_{r+1}} (\tau^{\nu(r+1)})^j (L \cap S)\right) \cap (L' \cap S)\right) \geqslant \frac{\mu(L')}{k} - \frac{\mu(L')}{2^{r+1}k} - \frac{\mu(L')}{4k}$$

and therefore in any case

$$\begin{aligned}\mu\left(\left(\bigcup_{j=-H_{r-1}}^{H_{r-1}} (\tau^{\nu(r+1)})^j (L \cap S)\right) \cap (L' \cap S)\right) &\geqslant \mu(L')/2k \\ &\geqslant \mu(L' \cap S)/2k. \end{aligned} \tag{5.11}$$

Note that the iterates of $L \cap S$ under $\tau^{\nu(r+1)}$ which make (5.11) hold are all *within* the columns of T_{r+1} and so the formation of u-blocks of order $i \geqslant r+1$ will not affect them. So using (5.11), we finally obtain

$$\begin{aligned}\mu\left(\left(\bigcup_{j=-\infty}^{\infty} (\tau^{\nu_i})^j (L \cap S)\right) \cap (L' \cap S)\right) \\ \geqslant \mu\left(\left(\bigcup_{j=-H_{r+1}}^{H_{r+1}} (\tau^{\nu_i})^j (L \cap S)\right) \cap (L' \cap S)\right) \\ \geqslant \mu(L' \cap S)/2k, \end{aligned}$$

or equivalently, as $(\tau^{\nu_i})S = S$ and normalizing, we have

$$\mu'\left(\left(\bigcup_{j=-\infty}^{\infty} (\tau^{\nu_i})^j L\right) \cap L'\right) \geqslant \mu'(L')/2k.$$

Thus (5.5) holds. Now, since the sets in $\{(\bigcup_{r=1}^{\infty} \mathscr{L}_r^+) \cap S\}$ generate the Borel σ-algebra restricted to $\{\nu' \neq 0\}$, the ergodicity of $\tau^{\nu'}$ on $\{\nu' \neq 0\}$ follows by a standard approximation argument (see [**3**, § 6], for instance).

(ii) Recall that $\mu(E_k) > \delta$ and, by (5.3), $\sum_{r=1}^{\infty} \eta_r' < \frac{1}{6}\delta$. Using (4.12), we see that it then follows that the set E_k' is non-empty and hence that $\int \nu' \, d\mu = k$. As a consequence of (2.2), we have

$$h(\tau^{\nu'}) = kh(\tau) = h(\tau^{\nu}).$$

(iii) If we use Lemma 2 and equation (5.3), we see that

$$\mu(E_i \Delta E_i') < 3 \sum_{r=1}^{\infty} \eta_r < \frac{\varepsilon}{k^2(k+2)}, \quad \text{for } 0 \leqslant i \leqslant k.$$

Lemma 3 then implies that $d(\tau^{\nu}, \tau^{\nu'}) < \varepsilon$.

COROLLARY 2. *The conclusions* (i) *and* (iii) *of Theorem* 3 *still hold when* $\int \nu \, d\mu = +\infty$.

Proof. We remarked earlier that we can approximate a stopping-time transformation τ^{ν} with $\int \nu \, d\mu = +\infty$ by a stopping-time transformation τ^{ν_k} with $\int \nu_k \, d\mu = k < \infty$.

We are now in a position to study indirectly other ergodic properties of stopping-time transformations by using Theorem 3 and known density results for induced transformations.

THEOREM 4. *Let* τ *be an ergodic automorphism of* $(\Omega, \mathscr{B}, \mu)$, *and let* ν *be a stopping time for* $(\Omega, \mathscr{B}, \mu, \tau)$.

(i) *For any* $\varepsilon > 0$, *there exists a stopping-time transformation* τ^{ν_1} *which is strong mixing on* $\{\nu_1 \neq 0\}$ *and such that* $d(\tau^{\nu}, \tau^{\nu_1}) < \varepsilon$.

(ii) *Suppose that* τ *has positive entropy. For any* $\varepsilon > 0$, *there exists a stopping-time transformation* τ^{ν_2} *which is a Kolmogorov automorphism on* $\{\nu_2 \neq 0\}$ *and such that* $d(\tau^{\nu}, \tau^{\nu_2}) < \varepsilon$.

Proof. (i) In [**4**] Friedman and Ornstein proved that given any ergodic automorphism τ, the sets A such that τ_A is strong mixing are dense.

So given $\varepsilon > 0$, we first use Corollary 2 to obtain a stopping time ν' such that $\tau^{\nu'}$ is ergodic on $\{\nu' \neq 0\}$ and $d(\tau^{\nu}, \tau^{\nu'}) < \frac{1}{2}\varepsilon$. Let $\tau^{\nu'}|_{\{\nu' \neq 0\}}$ denote the restriction of $\tau^{\nu'}$ to $\{\nu' \neq 0\}$. Then there exists a set $A \subset \{\nu' \neq 0\}$ such

that

(a) $\mu(A)/\mu(\{\nu' \neq 0\}) > 1-\frac{1}{6}\varepsilon$. and

(b) $(\tau^{\nu'}|_{\{\nu'\neq 0\}})_A$ is strong mixing.

Let τ^{ν_1} be the transformation defined by

$$\tau^{\nu_1}(\omega) = \begin{cases} (\tau^{\nu'})_A(\omega), & \text{if } \omega \in A, \\ \omega, & \text{if } \omega \in A^c. \end{cases} \tag{5.12}$$

Then clearly τ^{ν_1} is a stopping-time transformation. Moreover, since $\tau^{\nu'}|_{\{\nu'\neq 0\}} = (\tau^{\nu'})_{\{\nu'\neq 0\}}$, we have

$$\begin{aligned} \tau^{\nu_1}|_{\{\nu_1\neq 0\}} &= (\tau^{\nu'})_A \\ &= (\tau^{\nu'})_{\{\nu'\neq 0\}\cap A} \\ &= (\tau^{\nu'}|_{\{\nu'\neq 0\}})_A, \end{aligned}$$

which is strong mixing by (b). Also

$$d(\tau^{\nu'}, \tau^{\nu_1}) < 2\mu(\{\nu' \neq 0\}\backslash A) + \mu(A^c \backslash \{\nu' = 0\}) < \tfrac{1}{2}\varepsilon.$$

which in turn implies that $d(\tau^{\nu}, \tau^{\nu_1}) < \varepsilon$.

(ii) In [7], Ornstein and Smorodinsky proved that given any ergodic automorphism τ of positive entropy. the sets A such that τ_A is a K-automorphism are dense.

As in (i) we use Corollary 2 to obtain a stopping time ν' with finite expectation such that $\tau^{\nu'}|_{\{\nu'\neq 0\}}$ is ergodic and $d(\tau^{\nu}, \tau^{\nu'}) < \frac{1}{2}\varepsilon$. We now check that the entropy of $\tau^{\nu'}|_{\{\nu'\neq 0\}}$ is strictly positive. We have

$$\begin{aligned} h(\tau^{\nu'}|_{\{\nu'\neq 0\}}) &= h((\tau^{\nu'})_{\{\nu'\neq 0\}}) \\ &= kh(\tau) \cdot \mu(\{\nu' \neq 0\}), \end{aligned}$$

where $\int \nu' \, d\mu = k < \infty$. $\mu(\{\nu' \neq 0\}) \geqslant \mu(E_k') > 0$. and we have used (2.2) to deduce the last equality. As $h(\tau) > 0$ by hypothesis. it then follows that $h(\tau^{\nu'}|_{\{\nu'\neq 0\}}) > 0$. and we can then choose a set A, with $\mu(A)/\mu(\{\nu' \neq 0\}) > 1-\frac{1}{6}\varepsilon$. such that $(\tau^{\nu'}|_{\{\nu'\neq 0\}})_A$ is a K-automorphism. We then define τ^{ν_2} similarly to τ^{ν_1} in (5.12).

REFERENCES

1. R. M. Belinskaya. 'Partitions of Lebesgue space in trajectories defined by ergodic automorphisms'. *Funkcional. Anal. i Priložen* 2 (1968) 4–16; *Functional Anal. Appl.* 2 (1968) 190–99.
2. —— Generalised degrees of automorphism and entropy', *Sibirsk. Mat. Ž.* 11 (1970) 739–49; *Siberian Math. J.* 11 (1970) 559–66.
3. N. A. Friedman. *Introduction to ergodic theory* (van Nostrand. New York, 1970).
4. —— and D. S. Ornstein. 'Ergodic transformations induce mixing transformations', *Adv. in Math.* 10 (1973) 147–63.

5. S. KAKUTANI. Induced measure preserving transformations', *Proc. Imp. Acad. Sci. Tokyo* 19 (1943) 635–41.
6. J. NEVEU. 'Temps d'arrêt d'un système dynamique', *Z. Wahrsch. Verw. Gebiete* 13 (1969) 81–94.
7. D. S. ORNSTEIN and M. SMORODINSKY. 'Ergodic flows of positive entropy can be time changed to become K-flows', *Israel J. Math.* 26 (1977) 75–83.
8. S. T. RICHARDSON and K. M. WILKINSON. 'Stopping time transformations and towers', *Z. Wahrsch. Verw. Gebiete* 48 (1979) 259–84.
9. V. A. ROHLIN. Lectures on the entropy theory of measure preserving transformations', *Uspehi Mat. Nauk* 22 (1967) 3–56; *Russian Math. Surveys* 22 (1967) 1–52.

University of Warwick

Present address:
*U.*170 *INSERM*
16 *bis avenue Paul-Vaillant-Couturier*
94800 *Villejuif, France*

DEUXIEME PARTIE

PROCESSUS SPATIALEMENT DEPENDANTS : CONVERGENCE VERS LA NORMALITE, TESTS D'ASSOCIATION ET APPLICATIONS

ON THE VARIANCE OF THE SAMPLE CORRELATION BETWEEN TWO INDEPENDENT LATTICE PROCESSES

J. Appl. Prob. **18**, 943–948 (1981)
Printed in Israel
0021–9002/81/040943–06 $00.85

ON THE VARIANCE OF THE SAMPLE CORRELATION BETWEEN TWO INDEPENDENT LATTICE PROCESSES

SYLVIA RICHARDSON* AND
DENIS HEMON,* *Institut National de la Santé et de la Recherche Médicale*

Abstract

Consider two stochastically independent, stationary Gaussian lattice processes with zero means, $\{X(u), u \in \mathbb{Z}^2\}$ and $\{Y(u), u \in \mathbb{Z}^2\}$. An asymptotic expression for the variance of the sample correlation between $\{X(u)\}$ and $\{Y(u)\}$ over a finite square is derived. This expression also holds for a wide class of domains in $\mathbb{Z}^2$. As an illustration, the asymptotic variance of the correlation between two first-order autonormal schemes is evaluated.

AUTONORMAL SCHEMES; LATTICE PROCESSES; SAMPLE CROSS-CORRELATION

1. Introduction

We shall consider throughout an infinite square lattice, $\mathbb{Z} \times \mathbb{Z}$, and random variables $X(u) = X_{(u_1,u_2)}$, $Y(v) = Y_{(v_1,v_2)}$ associated with each point (u_1, u_2) or (v_1, v_2) of the lattice. The processes $\{X(u)\}$ and $\{Y(v)\}$ are assumed to be mutually independent and both stationary Gaussian with zero means and finite variance σ_X^2 and σ_Y^2 respectively. Let C_n, $n \geqq 1$, be any square region in $\mathbb{Z}^2$ containing n^2 points, $C_n = \{(i, j);\ s \leqq i \leqq s + n - 1;\ t \leqq j \leqq t + n - 1\}$ for some integers s, t. We consider in this paper the variance of the sample correlation coefficient r_{XY} between the n^2 observations from $\{X(u)\}$, $\{Y(u)\}$ over C_n:

$$r_{XY} = \frac{\sum_{u \in C_n} X(u)Y(u)}{\left[\sum_{u \in C_n} X(u)^2\right]^{\frac{1}{2}} \left[\sum_{u \in C_n} Y(u)^2\right]^{\frac{1}{2}}}.$$

Interest in lattice processes has arisen recently in many contexts, see Besag (1974) and Tjøstheim (1978) for further references. For time series, the asympto-

Received 7 July 1980; revision received 2 October 1980.

* Postal address: U170 Statistiques, 16 bis avenue Paul Vaillant Couturier, 94800 Villejuif, France.

tic variance and covariances of the cross correlation were derived by Bartlett (1935), (1966), and their asymptotic distribution by Hannan (1970). These results are commonly used in the analysis of relationships between two time series (Haugh and Box (1977), Pierce (1979)). Similarly, spatial cross-correlation coefficients have been proposed for studying relationships between spatial series (Bennett (1980)).

2. The asymptotic variance of r_{XY}

Variances and covariances will be denoted as follows:

$$C_{XX}(u, v) = E[X(u)X(v)], \qquad C_{YY}(u, v) = E[Y(u)Y(v)]$$
$$C_{XY}(u, v) = E[X(u)Y(v)],$$

in particular $\mathrm{Var}[X(u)] = C_{XX}(0,0) = \sigma_X^2$, where in this case 0 is the point in $\mathbb{Z}^2$ with both coordinates equal to 0. It will be supposed that the autocovariances of $\{X(u)\}$ and $\{Y(u)\}$ are dominated by exponentially decreasing functions of their coordinates, that is there exists θ_1, θ_2, $0 < \theta_1, \theta_2 < 1$, such that

$$(1) \qquad \begin{aligned} |C_{XX}(u, v)| &\leqq c_1 \theta_1^{|u_1 - v_1| + |u_2 - v_2|} \\ |C_{YY}(u, v)| &\leqq c_2 \theta_2^{|u_1 - v_1| + |u_2 - v_2|} \end{aligned}$$

where c_1, c_2 are positive constants. Note that Condition (1) follows from Cauchy's inequalities whenever the autocovariance generating functions of $\{X(u)\}$ and $\{Y(u)\}$ have a Laurent expansion in the neighbourhood of the unit torus, as is the case for the autonormal schemes of Besag (1974) or the simultaneous autoregressive schemes of Whittle (1954). It will be shown that

$$(2) \qquad \lim_{n \to \infty} n^2 \mathrm{Var}(r_{XY}) = \frac{1}{\sigma_X^2 \sigma_Y^2} \sum_{v \in \mathbb{Z}^2} C_{XX}(0, v) C_{YY}(0, v),$$

where, due to (1), the series on the right-hand side is absolutely convergent.

To obtain this result, we first derive $\mathrm{Var}[\Sigma_{u \in C_n} X(u)Y(u)]$. Since $\{X(u)\}$ and $\{Y(u)\}$ are independent and both have mean 0,

$$\mathrm{Var}\left[\sum_{u \in C_n} X(u)Y(u)\right] = \sum_{u \in C_n} \sum_{w \in C_n} C_{XX}(u, w) C_{YY}(u, w).$$

For $0 < |v_1|, |v_2| \leqq n - 1$, we let $D(v) = D_{(v_1, v_2)} = \{(u, w) \in C_n \times C_n ;\ u_1 - w_1 = v_1,\ u_2 - w_2 = v_2\}$. With this notation,

$$\sum_{u \in C_n} \sum_{w \in C_n} C_{XX}(u, w) C_{YY}(u, w) = \sum_{\substack{|v_1| \leqq n-1 \\ |v_2| \leqq n-1}} \sum_{(u,w) \in D(v)} C_{XX}(u, w) C_{YY}(u, w).$$

Note that the number of couples (u, w) in $D(v)$ is equal to the number of points in the intersection of C_n and $T_{-v}(C_n)$, the translation of C_n by $-v$, that is

$(n-|v_1|)\cdot(n-|v_2|)$. Because $\{X(u)\}$ and $\{Y(u)\}$ are stationary $C_{XX}(u, w) = C_{XX}(0, v)$ for $(u, w) \in D(v)$, so that

$$\operatorname{Var}\left[\sum_{u\in C_n} X(u)Y(u)\right] = \sum_{\substack{|v_1|\le n-1\\ |v_2|\le n-1}} (n-|v_1|)\cdot(n-|v_2|)C_{XX}(0, v)C_{YY}(0, v).$$

Hence, using Condition (1),

$$\frac{1}{n^2}\operatorname{Var}\left[\sum_{u\in C_n} X(u)Y(u)\right] = \sum_{\substack{|v_1|\le n-1\\ |v_2|\le n-1}} C_{XX}(0, v)C_{YY}(0, v) + O\left(\frac{1}{n}\right). \tag{3}$$

Next, we expand $\operatorname{Var}(r_{XY})$ in powers of n^{-2} using standard techniques. Let

$$A_n = \left[\frac{1}{n^2}\sum_{u\in C_n} X(u)Y(u)\right]^2, \qquad B_n = \left[\frac{1}{n^2}\sum_{u\in C_n} X(u)^2\right]\cdot\left[\frac{1}{n^2}\sum_{u\in C_n} Y(u)^2\right],$$

so that $E(r_{XY}^2) = E(A_n/B_n)$. The first term in the expansion of $E(r_{XY}^2)$ in powers of n^{-2} is given by

$$\frac{E(A_n)}{E(B_n)} = \frac{1}{n^4\sigma_X^2\sigma_Y^2} E\left[\left(\sum_{u\in C_n} X(u)Y(u)\right)^2\right].$$

To show that $E(r_{XY}^2)$ equals $E(A_n)/E(B_n)$ plus terms of order n^{-4} we must consider the variances and covariances of A_n and B_n. Note that relation (3) implies that $E(A_n)$ is of order n^{-2}. Now

$$E(A_n^2) = \frac{1}{n^8}\sum_{u,v,w,x\in C_n} E[X(u)X(v)X(w)X(x)]E[Y(u)Y(v)Y(w)Y(x)],$$

but since $\{X(u)\}$ is Gaussian,

$$\begin{aligned} &E[X(u)X(v)X(w)X(x)] \\ &\quad = C_{XX}(u, v)C_{XX}(w, x) + C_{XX}(u, w)C_{XX}(v, x) + C_{XX}(u, x)C_{XX}(v, w), \end{aligned}$$

and consequently $E(A_n^2)$ can be expressed as the sum of nine terms, three symmetric in X, Y and six cross-product terms. Each symmetric term is equal to $E(A_n)^2$ and hence of order n^{-4}. All cross-product terms are of similar form and so we consider only one. If $\theta = \max(\theta_1, \theta_2)$, and $\|u - v\| = |u_1 - v_1| + |u_2 - v_2|$, we have:

$$\begin{aligned} &\frac{1}{n^8}\sum_{u,v,w,x\in C_n} C_{XX}(u, v)C_{XX}(w, x)C_{YY}(u, w)C_{YY}(v, x) \\ &\quad \le \frac{c_1c_2^3}{n^8}\sum_{u,v,w,x\in C_n} \theta^{\|u-v\|+\|w-x\|+\|v-x\|+\|u-w\|} \\ &\quad \le \frac{c_1c_2}{n^4}\left(\sum_{v\in Z^2}\theta^{\|v\|}\right)^2. \end{aligned}$$

Consequently $\mathrm{Var}(A_n)$, and similarly $\mathrm{Var}(B_n)$, $\mathrm{Cov}(A_n, B_n)$ are of order n^{-1}. Similarly $E(r_{XY}) = 0$ up to order n^{-2} so that the first term in the expansion of $\mathrm{Var}(r_{XY})$ is given by

$$\mathrm{Var}(r_{XY}) = \frac{1}{n^4 \sigma_X^2 \sigma_Y^2} \mathrm{Var}\left[\sum_{u \in C_n} X(u)Y(u)\right] + O\left(\frac{1}{n^4}\right). \tag{4}$$

Using (3) and (4), (2) follows immediately.

Expression (2) can be extended to sequences of domains in $\mathbb{Z}^2$ which are not necessarily square-shaped. Consider an infinite sequence of domains $\{D_n\}_{n \geq 1}$ with $\lim_{n \to \infty} (b_n / d_n) = 0$, where d_n is the number of lattice points in D_n and b_n the number of lattice points on the boundary of D_n. Comparing the number of points in D_n and in $D_n \cap T_{-v}(D_n)$, for v fixed in $\mathbb{Z}^2$, it can easily be shown that (2) still holds in this case. On the other hand, if we now consider an infinite sequence of rectangles of fixed width m and increasing length n, an expression other than (2) can be obtained for the asymptotic variance of r_{XY}, namely

$$\lim_{n \to \infty} nm \, \mathrm{Var}(r_{XY}) = \frac{1}{\sigma_X^2 \sigma_Y^2} \sum_{\substack{v_1 \in \mathbb{Z} \\ 0 \leq |v_2| \leq m-1}} C_{XX}(v, 0) C_{YY}(v, 0)$$

$$- \frac{1}{m} \sum_{\substack{v_1 \in \mathbb{Z} \\ 0 \leq |v_2| \leq m-1}} |v_2| \, C_{XX}(v, 0) C_{YY}(v, 0),$$

an expression which is Bartlett's result for time series when $m = 1$. Result (2) is therefore specifically related to sequences of domains such that boundary effects can be asymptotically neglected.

3. Application to symmetric first-order autonormal schemes

As an illustration of (2), suppose that $\{X(u)\}$ and $\{Y(u)\}$ are both stationary symmetric first-order autonormal schemes with autocovariance generating functions given by

$$F_X(z_1, z_2) = \frac{1}{1 - a(z_1 + z_1^{-1} + z_2 + z_2^{-1})},$$

$$F_Y(z_1, z_2) = \frac{1}{1 - b(z_1 + z_1^{-1} + z_2 + z_2^{-1})},$$

$$1 - \delta < |z_1|, |z_2| < 1 + \delta, \quad |a| < \tfrac{1}{4}, \quad |b| < \tfrac{1}{4}.$$

The existence of such processes is discussed in Rozanov (1967) and Moran (1973b).

As noted by Besag (1972), the autocovariances of many spatial processes are not readily expressed in an analytically tractable form. However, using a remark of Quenouille (1947) for time series, $\Sigma_{v \in \mathbb{Z}^2} C_{XX}(0, v) C_{YY}(0, v)$ can be calculated

as the constant term in the Laurent expansion of $F_X(z_1, z_2)F_Y(z_1, z_2)$. Denoting this constant term by $P_1(a, b)$, we have

$$P_1(a, b) = \sum_{\substack{n=0 \\ m+n \text{ even}}}^{\infty} \sum_{m=0}^{\infty} a^n b^m \binom{m+n}{\frac{1}{2}(m+n)}^2,$$

so that

$$P_1(a, b) = \begin{cases} \dfrac{1}{a-b}\left[a\sum_{t=0}^{\infty}\binom{2t}{t}^2 a^{2t} - b\sum_{t=0}^{\infty}\binom{2t}{t}^2 b^{2t}\right], & a \neq b \\ \sum_{t=0}^{\infty}(2t+1)\binom{2t}{t}^2 a^{2t}, & a = b. \end{cases} \tag{5}$$

Moran (1973a) noted that

$$\sum_{t=0}^{\infty}\binom{2t}{t}^2 a^{2t} = \frac{2}{\pi}K(16a^2)$$

where $K(16a^2) = \int_0^{\pi/2}(1 - 16a^2\sin^2\theta)^{-\frac{1}{2}}d\theta$, is the complete elliptic integral of the first kind. Similarly, using standard results (Abramowitz and Stegun (1965), 15.1.1., 17.3.10) we have

$$\sum_{t=0}^{\infty}(2t+1)\binom{2t}{t}a^{2t} = \frac{2}{\pi}\frac{1}{1-16a^2}E(16a^2),$$

where $E(16a^2) = \int_0^{\pi/2}(1 - 16a^2\sin^2\theta)^{\frac{1}{2}}d\theta$, is the complete elliptic integral of the second kind.

Consequently

$$P_1(a, b) = \begin{cases} \dfrac{2}{\pi}\dfrac{1}{a-b}[aK(16a^2) - bK(16b^2)], & a \neq b \\ \dfrac{2}{\pi}\dfrac{1}{1-16a^2}E(16a^2), & a = b. \end{cases} \tag{6}$$

Moran (1973a) also showed that $\sigma_X^2 = (2/\pi)K(16a^2)$, $\sigma_Y^2 = (2/\pi)K(16b^2)$. We conclude from (2) and (6) that for $a \neq b$

$$\lim_{n\to\infty} n^2 \operatorname{Var}(r_{XY}) = \frac{\pi}{2}\frac{1}{a-b}\left[\frac{a}{K(16b^2)} - \frac{b}{K(16a^2)}\right]$$

and for $a = b$

$$\lim_{n\to\infty} n^2 \operatorname{Var}(r_{XY}) = \frac{\pi}{2}\frac{1}{1-16a^2}\frac{E(16a^2)}{[K(16a^2)]^2}.$$

Numerical evaluation of these last two expressions shows that $n^2\operatorname{Var}(r_{XY})$ increases more than exponentially with a and b.

Acknowledgements

We should like to thank Drs J. K. Ord and C. M. Triggs for helpful discussions and Mrs M. Guerrois for typing the manuscript.

References

ABRAMOWITZ, M. AND STEGUN, I. A. (1965) *Handbook of Mathematical Functions*. Dover, New York.

BARTLETT, M. S. (1935) Some aspects of the time-correlation problem in regard to tests of significance. *J. R. Statist. Soc.* **98**, 536–543.

BARTLETT, M. S. (1966) *An Introduction to Stochastic Processes*, 2nd edn. Cambridge University Press.

BENNETT, R. J. (1980) *Spatial Time Series*. Pion, London.

BESAG, J. E. (1972) On the correlation structure of some two-dimensional stationary processes. *Biometrika* **59**, 43–48.

BESAG, J. E. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B **36**, 192–236.

HANNAN, E. J. (1970) *Multiple Time Series*. Wiley, New York.

HAUGH, L. D. AND BOX, G. E. P. (1977) Identification of dynamic regression (distributed lag) models connecting two time series. *J. Amer. Statist. Assoc.* **72**, 121–130.

MORAN, P. A. P. (1973a) A Gaussian Markovian process on a square lattice. *J. Appl. Prob.* **10**, 54–62.

MORAN, P. A. P. (1973b) Necessary conditions for Markovian processes on a lattice. *J. Appl. Prob.* **10**, 605–612.

PIERCE, D. A. (1979) R^2 measures for time series *J. Amer. Statist. Assoc.* **74**, 901–910.

QUENOUILLE, M. H. (1947) Notes on the calculation of autocorrelations of linear autoregressive schemes. *Biometrika* **34**, 365–367.

ROZANOV, YU. A. (1967) On the Gaussian homogeneous fields with given conditional distributions. *Theory Prob. Appl.* **12**, 381–391.

TJØSTHEIM, D. (1978) Statistical spatial series modelling. *Adv. Appl. Prob.* **10**, 130–154.

WHITTLE, P. (1954) On stationary processes in the plane. *Biometrika* **41**, 434–449.

AUTOCORRELATION SPATIALE : SES CONSEQUENCES SUR LA CORRELATION EMPIRIQUE DE DEUX PROCESSUS SPATIAUX

AUTOCORRELATION SPATIALE : SES CONSEQUENCES SUR LA CORRELATION EMPIRIQUE DE DEUX PROCESSUS SPATIAUX

Sylvia RICHARDSON, Denis HEMON (*)

RESUME

L'existence d'autocorrélations pose des problèmes quant aux tests d'indépendance stochastique de deux processus spatiaux.

Pour apprécier de façon quantitative ces problèmes, on étudie la variance du coefficient de corrélation empirique des deux processus. Cette variance est évaluée pour différentes valeurs des autocorrélations et pour plusieurs types de processus. Il ressort clairement de ces évaluations que l'influence des autocorrélations ne peut être négligée. Ceci conduit à discuter les approches proposées pour tester l'indépendance de deux processus spatiaux.

1. INTRODUCTION

Les méthodes classiques d'analyse statistique visant à tester l'indépendance de deux variables aléatoires X et Y supposent que l'on dispose d'un échantillon de réalisations indépendantes du couple (X, Y).

Cependant lorsque ces variables sont liées à un espace géographique, elles présentent le plus souvent un certain degré d'autocorrélation spatiale. Ceci pose des problèmes du point de vue de l'analyse statistique, comme le soulignent LEBART [15], CLIFF et ORD [8] et UNWIN et HEPPLE [26]. Ces auteurs remarquent en particulier que si l'on néglige l'autocorrélation, on est souvent conduit à sous-estimer le risque de première espèce des tests employés. Pour apprécier de façon quantitative ces problèmes, nous étudions ici la variance du coefficient de corrélation empirique.

Comme dans le cas de séries temporelles, l'existence d'une corrélation entre deux processus spatiaux peut résulter de l'existence de deux tendances.. L'estimation et l'interprétation d'une tendance spatiale a fait l'objet de nombreux travaux, dont on trouvera une bibliographie dans UNWIN et HEPPLE [26]. Aussi ne consi-

(*) INSERM U.170, 16 Bis av. P.V. Couturier, 94800 VILLEJUIF.

Mots-Clés : **Processus spatiaux, autocorrélation, corrélation croisée.**

dérons-nous que des processus stationnaires définis en chaque point d'un quadrillage régulier, processus étudiés en particulier par BESAG [6] et TJØSHEIM [25].

Soient $\{X(u)\}$ et $\{Y(u)\}$ deux processus stationnaires, gaussiens, centrés et de variances finies σ_X^2 et σ_Y^2 respectivement, définis en tout point $u = (u_1, u_2)$ de $Z \times Z$. Les autocovariances de $\{X(u)\}$ seront notées :

$$C_{XX}(u, v) = E[X(u)\, X(v)], \quad C_{YY}(u, v) = E[Y(u)\, Y(v)],$$

en particulier $\mathrm{Var}[X(u)] = C_{XX}(0, 0) = \sigma_X^2$, 0 représentant le point de Z^2 dont les deux cordonnées sont égales à zéro. Pour tout sous-ensemble D_n de Z^2 contenant d_n points, le coefficient de corrélation empirique r_{XY} entre les observations de $\{X(u)\}$, $\{Y(u)\}$ sur D_n est défini par :

$$r_{XY} = \frac{\sum_{u \in D_n} X(u)\, Y(u)}{\left[\sum_{u \in D_n} X(u)^2\right]^{1/2} \left[\sum_{u \in D_n} Y(u)^2\right]^{1/2}}$$

Sous l'hypothèse d'indépendance mutuelle entre $\{X(u)\}$ et $\{Y(u)\}$ nous avons démontré précédemment ([23]) que la variance asymptotique de r_{XY} est donnée par :

$$\lim_{n \to \infty} d_n \,\mathrm{Var}(r_{XY}) = \frac{1}{\sigma_X^2 \sigma_Y^2} \sum_{v \in Z^2} C_{XX}(0, v)\, C_{YY}(0, v), \qquad (1)$$

pour toute suite de domaines D_n de Z^2 où les effets de bords s'estompent à l'infini, (c'est-à-dire tels que $\lim_{n \to \infty} \frac{b_n}{d_n} = 0$, b_n étant le nombre de points de Z^2 sur la frontière de D_n). Notons que la série intervenant dans (1) converge absolument pour une large classe de processus, en particulier ceux dont les autocovariances $C_{XX}(0, v)$ et $C_{YY}(0, v)$ sont dominées par une fonction exponentielle décroissante du type $c\theta^{|v|}$, avec $|\theta| < 1$ et $c > 0$, [23].

L'expression (1) est analogue à celle de BARTLETT dans le cas de séries temporelles, [2], [3]. Remarquons qu'il suffit que l'un des processus $\{X(u)\}$, $\{Y(u)\}$ soit non autocorrelé pour que la variance asymptotique de r_{XY} ait la même valeur que dans le cas d'un échantillon de d_n couples (X, Y) indépendants c'est-à-dire $1/d_n$. Ceci est par exemple le cas en expérimentation agronomique où des parcelles disposées selon un lattice reçoivent un traitement tiré au sort. Par contre si les deux processus $\{X(u)\}$ et $\{Y(u)\}$ présentent tous deux une autocorrélation positive la relation (1) montre que la variance asymptotique de r_{XY} est supérieure à $1/d_n$.

Au § 2 nous évaluerons numériquement cette variance pour différents types de processus spatiaux. Au § 3 nous discuterons les différentes approches qui ont été proposées pour tester l'indépendance stochastique de deux processus spatiaux.

2. EVALUATION DE LA VARIANCE ASYMPTOTIQUE DE r_{XY} POUR PLUSIEURS TYPES DE PROCESSUS

2.1. PROCESSUS DONT L'AUTOCOVARIANCE DECROIT GEOMETRIQUEMENT

Considérons tout d'abord le cas où $\{X(t), t \in \mathbf{Z}\}$ et $\{Y(t), t \in \mathbf{Z}\}$ sont deux processus autoregressifs d'ordre 1 sur $\mathbf{Z}$. Dans ce cas $\{X(t), t \in \mathbf{Z}\}$ et $\{Y(t), t \in \mathbf{Z}\}$ sont également des processus de Markov, leur autocovariance est donnée par :

$$C_{XX}(0, k) = E[X(0)X(k)] = \sigma_X^2 \lambda^{|k|}, \; k \in \mathbf{Z}, \; |\lambda_1| < 1$$

$$C_{YY}(0, k) = E[Y(0)Y(k)] = \sigma_Y^2 \lambda^{|k|}, k \in \mathbf{Z}, \; |\lambda_1| < 1. \qquad (2)$$

La variance asymptotique de r_{XY}, défini sur un intervalle de n points, est égale à :

$$\lim_{n \to \infty} n \, \mathrm{Var}\,(r_{XY}) = \frac{1 + \lambda_1 \lambda_2}{1 - \lambda_1 \lambda_2}, \qquad (3)$$

formule obtenue par BARTLETT [2].

Cherchant à étendre la forme d'autocovariance donnée par [2] à un processus dans le plan, MARTIN [16] considère les processus "doublement géométriques" dont la structure stochastique est celle du produit de deux processus unidimensionnels satisfaisants (2) ; pour ces processus :

$$C_{XX}(0, v) = \sigma_X^2 \lambda_1^{|v_1| + |v_2|}, \; |\lambda_1| < 1, \; v \in \mathbf{Z}^2.$$

Contrairement au cas temporel $\{X(u), u \in \mathbf{Z}^2\}$ n'est plus un processus markovien et l'équation "autorégressive" qui le définit n'a pas d'interprétation naturelle. Si l'on suppose que $\{X(u)\}$ et $\{Y(u)\}$ sont tous deux des processus doublement géométriques, de paramètre λ_1 et λ_2 respectivement, l'expression (1) devient alors l'extension directe à deux dimensions de la formule (3) :

$$\lim_{n \to \infty} d_n \, \mathrm{Var}\,[r_{XY}] = \left(\frac{1 + \lambda_1 \lambda_2}{1 - \lambda_1 \lambda_2}\right)^2 \qquad (4)$$

2.2. PROCESSUS MARKOVIENS ET "AUTOREGRESSIFS" DES PLUS PROCHES VOISINS

L'autocovariance de beaucoup de processus spatiaux n'est pas exprimable sous une forme analytique simple, BESAG [5]. Cependant l'expression $\sum_{v \in \mathbf{Z}^2} C_{XX}(0, v) \, C_{YY}(0, v)$ peut être calculée comme terme constant du produit des développements de Laurent des fonctions génératrices d'autocovariance des processus $\{X(u)\}$ et $\{Y(u)\}$.

a) Processus markoviens

Nous supposons que $\{X(u)\}$ et $\{Y(u)\}$ sont tous deux des processus stationnaires gaussiens, markoviens des plus proches voisins, définis par les fonctions

génératrices des autocovariances suivantes :

$$F_X(z_1, z_2) = [1 - a(z_1 + z_1^{-1} + z_2 + z_2^{-1})]^{-1}$$
$$F_Y(z_1, z_2) = [1 - b(z_1 + z_1^{-1} + z_2 + z_2^{-1})]^{-1} \quad (5)$$

$1 - \delta < |z_1|, |z_2| < 1 + \delta$, $|a| < 1/4$, $|b| < 1/4$. L'existence de tels processus a été discutée par ROZANOV [24] et MORAN [19].

Ces processus satisfont aux relations suivantes :

$$X_{u_1,u_2} = a(X_{u_1-1,u_2} + X_{u_1+1,u_2} + X_{u_1,u_2-1} + X_{u_1,u_2+1}) + \epsilon_{u_1,u_2}, (u_1, u_2) \mathbf{Z} \times \mathbf{Z}. \quad (6)$$

Si $\partial X(u) = \{X_{u_1-1,u_2}, X_{u_1+1,u_2}, X_{u_1,u_2-1}, X_{u_1,u_2+1}\}$ est l'ensemble des plus proches voisins de X(u), chaque variable aléatoire $\epsilon(u)$ intervenant dans (6) est gaussienne ; conditionellement à $\partial X(u)$ elle est indépendante de $\{X(u)\}$ avec $|v_1 - u_1| + |v_2 - u_2| > 1$ et centrée.

L'expression de la variance asymptotique de r_{XY} que l'on calcule à partir des fonctions génératirces des autocovariances (5) est donnée dans l'appendice. Cette expression, qui fait intervenir les intégrales elliptiques des paramètres a et b, a été tabulée par intégration numérique avec trois décimales de précision (Table 1), les valeurs des paramètres a et b sont choisies de telle sorte que les autocorrélations d'ordre 1*, ρ_X et ρ_Y, varient de 0,0 à 0,7 par pas de 0,1.

TABLE 1

Valeur asymptotique de $d_n \mathrm{Var}(r_{XY})$ pour deux processus Markoviens

ρ_X \ ρ_X	0,00	0,100	0,200	0,300	0,400	0,500	0,600	0,700	b
0,000	1,00								0,000
0,100	1,00	1,042							0,095
0,200	1,00	1,088	1,195						0,168
0,300	1,00	1,138	1,321	1,558					0,2123
0,400	1,00	1,192	1,469	1,866	2,444				0,2355
0,500	1,00	1,248	1,637	2,251	3,270	4,989			0,24565
0,600	1,00	1,307	1,824	2,720	4,419	7,917	15,903		0,24918
0,700	1,00	1,368	2,025	3,256	5,883	12,432	33,523	140,603	0,24994
a	0,00	0,095	0,168	0,2123	0,2355	0,24565	0,24918	0,249949	

ρ_X et ρ_Y sont les autocorrélations d'ordre 1 de $\{X(u)\}$ et $\{Y(u)\}$ respectivement. Les valeurs correspondantes de a et b peuvent être lues respectivement en bas et à droite. Pour $a > b$, les valeurs sont obtenues par symétrie.

b) Processus autorégressifs

Supposons maintenant que $\{X(u)\}$ et $\{Y(u)\}$ soient des processus stationnaires gaussiens associés aux fonctions génératrices d'autocovariance suivantes :

$$F_X(z_1, z_2) = [1 - a(z_1 + z_1^{-1} + z_2 + z_2^{-1})]^{-2}$$

$$F_Y(z_1, z_2) = [1 - b(z_1 + z_1^{-1} + z_2 + z_2^{-1})]^{-2}$$

$1 - \delta < |z_1|, |z_2| < 1 + \delta$, $|a| < 1/4$, $|b| < 1/4$. Ces processus, qui furent considérés pour la première fois par WHITTLE [27], correspondent aux équations autorégressives :

$$X_{u_1,u_2} = a(X_{u_1-1,u_2} + X_{u_1+1,u_2} + X_{u_1,u_2-1} + X_{u_1,u_2+1}) + \epsilon_{u_1,u_2} \qquad (7)$$

dans lesquelles la suite d'innovations $\{E(u)\}, u = (u_1, u_2)\ Z \times Z\}$, est une suite de variables aléatoires gaussiennes indépendantes, d'espérance nulle. Notons que pour le processus markovien (6), $\epsilon(u)$ et $\epsilon(v)$ sont corrélées quand $|u_1 - v_1| + |u_2 - v_2| \leqslant 2$. L'expression de la variance asymptotique de r_{XY} en fonction des paramètres des processus est donnée dans l'appendice et a été également tabulée (Table 2).

*L'autocorrélation d'ordre 1, ρ_X, est définie comme le coefficient de corrélation entre X(u) et l'un quelconque de ses plus proches voisins.

TABLE 2

Valeur asymptotique de $d_n \mathrm{Var}(r_{XY})$ pour deux processus autoregressifs

ρ_Y \ ρ_X	0,00	0,100	0,200	0,300	0,400	0,500	0,600	0,700	0,800	0,900	b
0,000	1,00										0,0000
0,100	1,00	1,041									0,0493
0,200	1,00	1,085	1,178								0,0945
0,300	1,00	1,130	1,282	1,458							0,1335
0,400	1,00	1,178	1,394	1,655	1,960						0,165
0,500	1,00	1,227	1,515	1,878	2,320	2,864					0,1902
0,600	1,00	1,279	1,647	2,130	2,743	3,530	4,540				0,2099
0,700	1,00	1,332	1,788	2,411	3,237	4,347	5,844	7,894			0,225
0,800	1,00	1,387	1,940	2,729	3,823	5,369	7,581	10,826	15,893		0,2364
0,900	1,00	1,444	2,094	3,096	4,541	6,710	10,045	15,403	24,838	44,846	0,2447
a	0,000	0,0493	0,0945	0,1335	0,165	0,1902	0,2099	0,225	0,2364	0,2447	

ρ_X et ρ_Y sont les autocorrélations d'ordre 1 de $\{X(u)\}$ et $\{Y(u)\}$ respectivement. les valeurs correspondantes de a et b peuvent être lues respectivement en bas et à droite. Pour a > b, les valeurs sont obtenues par symétrie.

En comparant les résultats obtenus pour les différents processus envisagés (formules (3), (4), Tables 1,2), on remarque tout d'abord que l'effet de l'autocorrélation intrinsèque des processus {X(u)} et {Y(u)} sur la variance du coefficient de corrélation est numériquement bien plus important dans le cas de processus spatiaux que dans celui des séries temporelles. Par exemple pour $\rho_X = \rho_Y = 0{,}3$, la "taille corrigée" d'un N-échantillon est de N/1,20 pour une série temporelle satisfaisant (2), tandis qu'elle est de N/1,56 pour le processus markovien dans $\mathbf{Z}^2$ défini par (6).

Par ailleurs les valeurs données pour les processus markoviens (Table 1) sont plus élevées que celles correspondant aux processus autorégressifs (Table 2) et aux processus doublement géométrique. Néanmoins quand ρ_X est plus petit que 0,3, les valeurs données pour les trois processus sont voisines.

3. DISCUSSION

L'analyse de la corrélation de deux séries temporelles est classiquement conduite en étudiant les corrélations croisées ou le spectre croisé. La plupart des concepts utilisés dans ce type d'analyse peuvent être étendus au cas des séries spatiales. Avant d'envisager de façon détaillée les différentes méthodes proposées, il est bon de rappeler les difficultés intrinsèques à l'étude des processus spatiaux.

La première de ces difficultés tient à l'absence d'un ordre naturel sur $\mathbf{Z}^2$. Ceci exclue en particulier la possibilité d'utiliser la distinction entre le passé et le futur pour interpréter les corrélations, PIERCE et HAUGH [22]. De ce point de vue, il est clair que l'analyse de séries à la fois spatiales et temporelles est particulièrement informative, GRANGER [11]. Une seconde difficulté tient à l'arbitraire du choix de la taille des unités géographiques considérées : on sait en effet que les résultats des analyses statistiques dépendent du degré d'agglomération de ces unités. Enfin, les observations sont souvent disposées de manière irrégulière dans l'espace ce qui interdit une analyse spectrale classique. Dans le cas d'un quadrillage irrégulier, on peut toutefois donner un sens à une décomposition hiérarchique en variations locales, départementales et régionales, comme le proposent par exemple CLIFF et ORD [9] et CURRY [10].

Les coefficients de corrélation croisés sont couramment employés dans l'analyse de la liaison de deux séries temporelles, HAUGH et BOX [14], PIERCE [21]. HANNAN [13] a établi la normalité asymptotique de ces coefficients sous certaines hypothèses. BENNETT [4] a suggéré d'entendre cette approche à l'étude de la liaison entre deux processus spatiaux. Si l'on se place dans le cas de processus stationnaires sur $\mathbf{Z}^2$ et d'un domaine d'échantillonnage régulier assez grand, on peut estimer les autocorrélations $\frac{C_{XX}(0,v)}{\sigma_X^2}$, $\frac{C_{YY}(0,v)}{\sigma_Y^2}$ d'un grand nombre de processus de manière consistante, GUYON et PRUM [12]. On peut alors estimer de façon consistante une approximation de la variance de r_{XY} en utilisant une version tronquée de la formule (1).

Plusieurs auteurs ont étudié des modifications du modèle de régression linéaire. LEBART [15] fixe la structure de la matrice de variance-covariance des résidus en la décomposant suivant des "niveaux de contiguités" décroissants. Cela

lui permet d'estimer les paramètres du modèle par la méthode des moindres carrés généralisés et d'apprécier la perte d'information due à la dépendance spatiale. Cette décomposition de la matrice de variance-covariance peut s'appliquer au cas où les variables sont mesurées aux sommets d'un graphe quelconque ; elle suppose que la covariance de deux variables dépende uniquement de leur niveau de contiguité sur le graphe et qu'elle soit nulle au delà d'une certaine distance. Généralisant les modèles de régression des séries chronologiques, ORD [20] étudie différents modèles mixtes régressifs-autorégressifs dans lesquels interviennent une matrice de voisinage supposée connue à un facteur de proportionalité près, ρ. Il discute la qualité des estimateurs de régressions et de ρ et compare l'estimateur du maximum de vraisemblance à plusieurs alternatives. Comme dans le cas de Lebart, ces modèles ne supposent pas une structure régulière de l'espace et la matrice de voisinage est adaptée au cas particulier considéré. Une approche empirique visant à prendre en compte l'effet des autocorrélations dans le modèle linéaire a été proposée par MARTIN [17]. Elle consiste à transformer $\{X(u)\}$ et $\{Y(u)\}$ en leur différence d'ordre 1, c'est-à-dire à supposer essentiellement $\rho = 1$ avant d'effectuer une régression linéaire. Martin étudie l'effet de cette procédure par des méthodes de Monte Carlo et montre qu'en supposant $\rho = 1$ les estimateurs des paramètres de régression sont en général meilleurs qu'en négligeant les autocorrélations ($\rho = 0$).

Plusieurs auteurs se sont préoccupés de la définition d'une mesure de corrélation sur des unités que l'on pouvait en principe subdiviser ou regrouper. Les méthodes proposées sont toutes fondées sur l'idée d'une décomposition hiérarchique des variations de chaque processus. CURRY [10] estime les paramètres d'un modèle dans lequel la valeur prise par le processus Y en un point u est une combinaison linéaire des différentes composantes hiérarchiques du processus X. CLIFF et ORD [9] utilisent une analyse de variance emboîtée en postulant que la corrélation entre $\{X(u)\}$ et $\{Y(u)\}$ est dûe à un facteur commun à chaque niveau hiérarchique. BESAG [7] propose un test non paramétrique qui fait intervenir toutes les permutations d'unités à l'intérieur d'un même bloc hiérarchique.

Les méthodes décrites ci-dessus apportent des solutions partielles au problème général de l'analyse de la corrélation entre deux processus spatiaux. Certaines présentent un caractère empirique impliquant la connaissance de la structure de la matrice de variance-covariance. D'autres adoptent des méthodes développées pour les séries chronologiques, dans ce cas la validité de certains résultats asymptotiques utilisés reste à établir. Les problèmes d'inférence statistique concernant l'indépendance de deux processus spatiaux restent donc encore ouverts, l'importance de l'effet des autocorrélations justifie que des travaux complémentaires leur soient consacrés.

REMERCIEMENTS

Nous remercions vivement Micheline IMBERT pour la préparation de ce manuscrit.

APPENDICE

a) LA VARIANCE ASYMPTOTIQUE DE r_{XY} DANS LE CAS DE DEUX PROCESSUS DE MARKOV

Le terme constant $P_1(a, b)$ du produit $F_X(z_1, z_2) \cdot F_Y(z_1, z_2)$ est égal à

$$P_1(a, b) = \begin{cases} \dfrac{1}{a-b}\left[a \sum_{t=0}^{\infty} \binom{2t}{t}^2 a^{2t} - b \sum_{t=0}^{\infty} \binom{2t}{t}^2 b^{2t} \right] & , a \neq b \\ \sum_{t=0}^{\infty} (2t+1) \binom{2t}{t}^2 a^{2t} & , a = b \end{cases} \tag{8}$$

Les séries constituant (i) peuvent être exprimées au moyen des séries hypergéométriques de Gauss, ABRAWOWITZ et STEGUN [1] 15.1.1 :

$$F(k, \ell, m, z) = \sum_{t=0}^{\infty} \frac{(k)_t (\ell)_t}{(m)_t} \frac{z^t}{t!}$$

$(k)_t = k(k+1)\ldots .(k+t-1)$. On obtient

$$\sum_{t=0}^{\infty} \binom{2t}{t}^2 a^{2t} = F(1/2, 1/2, 1, 16a^2)$$

$$\sum_{t=0}^{\infty} (2t+1) \binom{2t}{t}^2 a^{2t} = F(3/2, 1/2, 1, 16a^2)$$

On peut utiliser les relations qui lient les séries hypergéométriques aux intégrales elliptiques de la 1ere et 2e espèce quand on veut les évaluer numériquement :

$$F(1/2 \,.\, 1/2 \,.\, 1 \,, z) = \frac{2}{\Pi} K(z)$$

$$F(3/2 \,, 1/2 \,.\, 1 \,, z) = \frac{2}{\Pi} \frac{1}{1-z} E(z)$$

avec

$$K(z) = \int_0^{\Pi/2} (1 - z \sin^2 \theta)^{-\frac{1}{2}} \, d\theta \quad \text{et} \quad E(z) = \int_0^{\Pi/2} (1 - z \sin^2 \theta)^{\frac{1}{2}} \, d\theta \,,$$

ABRAMOVITZ et STEGUN [1] 17.3.9 et 17.3.10.

En conséquence

$$P_1(a, b) = \begin{cases} \dfrac{2}{\Pi} \dfrac{1}{a-b} [aK(16a^2) - bK(16b^2)] & , a \neq b \\ \dfrac{2}{\Pi} \dfrac{1}{1-16a^2} E(16a^2) & , a = b \,. \end{cases}$$

MORAN [18] a remarqué que :

$$\sigma_X^2 = \frac{2}{\Pi} K(16a^2) , \sigma_Y^2 = \frac{2}{\Pi} K(16b^2) .$$

On obtient donc finalement pour a ≠ b :

$$\lim_{n \to \infty} d_n \text{ Var } [r_{XY}] = \frac{\Pi}{2} \frac{1}{a - b} \left[\frac{a}{K(16b^2)} - \frac{b}{K(16a^2)} \right],$$

tandis que pour a = b :

$$\lim_{n \to \infty} d_n \text{ Var } [r_{XY}] = \frac{\Pi}{2} \frac{1}{1 - 16a^2} \frac{E(16a^2)}{[K(16a^2)]^2}$$

b) LA VARIANCE ASYMPTOTIQUE DE r_{XY} DANS LE CAS DE DEUX PROCESSUS AUTOREGRESSIFS

Le terme constant $P_2(a , b)$ du produit $F_X(z_1 , z_2) \cdot F_Y(z_1 , z_2)$ est égal à

$$P_2(a , b) = \begin{cases} \dfrac{2}{\Pi} \dfrac{1}{(a - b)^2} \left[\dfrac{b^2}{1 - 16b^2} E(16b^2) + \dfrac{a^2}{1 - 16a^2} E(16a^2) \right] \\ + \dfrac{2}{\Pi} \dfrac{2ab}{(a - b)^3} [bK(16b^2) - aK(16a^2)] \qquad , a \neq b \\ F(5/2 , 1/2 , 1 , 16a^2) + 20a^2 F(7/2 , 3/2 , 2 , 16a^2) , a = b . \end{cases}$$

On peut exprimer $F(5/2 , 1/2 , 1 , z)$ et $F(7/2 , 3/2 , 2 , z)$ en fonction des intégrales elliptiques $K(z)$ et $E(z)$ à l'aide des relations de récurrence suivantes :

$$5 F\left(\frac{7}{2}, \frac{3}{2}, 2 , z\right) = 2 F\left(\frac{5}{2}, \frac{3}{2}, 1 , z\right) + 3 F\left(\frac{5}{2}, \frac{3}{2}, 2 , z\right)$$

$$3 F\left(\frac{5}{2}, \frac{3}{2}, 2 , z\right) = 2 \frac{(1 - z)}{z} \left[F\left(\frac{5}{2}, \frac{3}{2}, 1 , z\right) - F\left(\frac{5}{2}, \frac{1}{2}, 1 , z\right) \right]$$

$$2 F\left(\frac{5}{2}, \frac{3}{2}, 1 , z\right) = \frac{2}{1 - z} \left[4 F\left(\frac{5}{2}, \frac{1}{2}, 1 , z\right) - 3 F\left(\frac{3}{2}, \frac{1}{2}, 1 , z\right) \right]$$

$$F\left(\frac{5}{2}, \frac{1}{2}, 1 , z\right) = \frac{1}{3(1 - z)} \left[2(2 - z) F\left(\frac{3}{2}, \frac{1}{2}, 1 , z\right) - F\left(\frac{1}{2}, \frac{1}{2}, 1 , z\right) \right]$$

Finalement remarquant que dans ce cas $\sigma_X^2 = P_1(a , a)$ et $\sigma_Y^2 = P_1(b , b)$ est défini par (8), on obtient :

$$\lim_{n \to \infty} d_n \text{ Var } [r_{XY}] = \frac{P_2(a , b)}{P_1(a , a) P_1(b , b)}$$

REFERENCES

[1] M. ABRAMOWITZ and I.A. STEGUN. — *Handbook of Mathematical Functions.* Dover, New-York, 1965.

[2] M.S. BARTLETT. — Some aspects of the time-correlation problem in regard to tests of significance. *J.R. Statist. Soc.*, 98, 536-543, 1935.

[3] M.S. BARTLETT. — *An Introduction to Stochastic Processes*, 2nd edn. Cambridge University Press, Cambridge, 1966.

[4] R.J. BENNETT. — *Spatial Time Series.* Pion, London, 1980.

[5] J.E. BESAG. — On the correlation structure of some two-dimensional stationary processes. *Biometrika*, **59**, 43-48, 1972.

[6] J.E. BESAG. — Spatial interaction and the statistical analysis of lattice system (with discussion) *J.R. Statist, Soc.*, **B 36**, 192-236, 1974.

[7] J.E. BESAG and P.J. DIGGLE. — Simple Monte Carlo tests for spatial pattern. *J.R. Statist. Soc.*, **C 26**, 327-333, 1977.

[8] A.D. CLIFF and J.K. ORD. — Model building and the analysis of spatial pattern in human geography (with discussion). *J.R. Statist. Soc.*, **B 37**, 297-348, 1975.

[9] A.D. CLIFF and J.K. ORD. — *The effects of spatial autocorrelation on geographical modelling.* University of Warwick Technical Report, 1978.

[10] L. CURRY. — A bivariate spatial regression operator. *Canadian Geographer*, **16**, 1-14, 1972.

[11] C.W.J. GRANGER. — Spatial data and time series analysis. In *Studies in Regional Sciences* (A.J. Scott, ed), 1-24. London : Pion, 1969.

[12] X. GUYON et B. PRUM. — Estimations et tests relatifs aux processus spatiaux réguliers du second ordre, Publication n° 201, Université d'Orsay, 1976.

[13] E.J. HANNAN. — *Multiple Time Series.* John Wiley and Sons, New-York, 1970.

[14] L.D. HAUGH and G.E.P. BOX. — Identification of dynamic regression (distributed lag) models connecting two time series. *J. Amer. Statist. Assoc.*, 72, 121-130, 1977.

[15] L. LEBART. — Analyse statistique de la contiguité. *Pub. Inst. Stat. Univ. Paris*, **18**, 81-112, 1969.

[16] R.J. MARTIN. — A subclass of lattice processes applied to a problem in planar sampling. *Biometrika*, **66**, 209-17, 1979.

[17] R.L. MARTIN. — On spatial dependence, bias, and the use of first spatial differences in regression analysis. *Area*, **6**, 185-94, 1974.

[18] P.A.P. MORAN. — A Gaussian Markovian process on a square lattice. *J. Appl. Prob.*, **10**, 54-62, 1973.

[19] P.A.P. MORAN. — Necessary conditions for Markovian processes on a lattice. *J. Appl. Prob.*, **10**, 605-612, 1973.

[20] J.K. ORD. — Estimation methods for models of spatial interaction. *J. Am. Statist. Assoc.*, **70**, 120-6, 1975.

[21] D.A. PIERCE. – R^2 Measures for time series. *J. Amer. Statist. Assoc.*, **74**, 901-910, 1979.

[22] D.A. PIERCE and L.D. HAUGH. – Causality in temoral systems, characterizations and a survey. *J. Econometrics*, **5**, 265-293, 1977.

[23] S.T. RICHARSON and D. HEMON. – On the variance of the sample correlation between two independant lattice processes. *J. Appl. Prob.*, **18**, 943-8, 1981.

[24] Yu. A. ROZANOV. – On the Gaussian homogeneous fields with given conditional distributions. *Theor. Prob. Appl.*, **12**, 381-391, 1967.

[25] D. TJOSTHEIM. – Statistical spatial series modelling. *Adv. Appl. Probl.*, **10**, 130-154, 1978.

[26] D.J. UNWIN and L.W. HEPPLE. – The statistical analysis of spatial series. *The Statistician*, **23**, 211-27, 1974.

[27] P. WHITTLE. – On stationary processes in the plane. *Biometrika*, **41**, 434-449, 1954.

VITESSE DE CONVERGENCE DU THEOREME DE LA LIMITE CENTRALE POUR DES CHAMPS FAIBLEMENT DEPENDANTS

Z. Wahrscheinlichkeitstheorie verw. Gebiete
66, 297–314 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants

Xavier Guyon[1] et Sylvia Richardson[2]

[1] Laboratoire de Statistiques, Bâtiment 425, Université de Paris XI, F-91405 Orsay, France
[2] Laboratoire de Statistiques, Université de Paris V, 45, rue des Saints Pères, F-75006 Paris, France

Résumé. Nous étudions la vitesse de convergence du théorème de la limite centrale pour des champs de $\mathbb{Z}^d$, faiblement dépendants: m-dépendant ou α-fortement mélangeant. Dès que le champ est dans $L^{2+\delta}, \delta>0$, la vitesse de convergence obtenue est $\sigma_n^{-(\delta\wedge 1)}$ avec un facteur $(\log\sigma_n)^a$ qui intervient quand α est à décroissance exponentielle et dans le cas m-dépendant quand $\delta\geqq 1$. Le cas où α est à décroissance puissance est aussi étudié. Ces résultats ne font intervenir ni la stationarité, ni la géométrie des domaines sur lesquels le T.L.C. est étudié[1].

§ 1. Introduction

Nous nous intéressons dans ce travail à la vitesse de convergence dans le théorème de la limite centrale (T.L.C.) pour $X=\{X_j, j\in\mathbb{Z}^d\}$ un champ defini sur $\mathbb{Z}^d$, centré, faiblement dépendant, vérifiant:

$$\sup_j \|X_j\|_{2+\delta}=\|X\|_{2+\delta}<\infty, \qquad \delta>0. \tag{1-1}$$

Deux types de faible dépendance sont étudiés classiquement: la m-dépendance et les concepts de mélange. Parmi ceux-ci nous nous limiterons ici à une notion de α-mélange fort.

La litterature dans le cas $d=1$ est abondante, traitant également le cas de processus accroissements de martingale (voir par exemple Ibragimov [10], Hall et Heyde [11] et Bolthausen [2]).

Pour le cas m-dépendant sans hypothèse de stationnarité, Petrov [17],

[1] Au moment où est soumis cet article, les auteurs on pris connaissance de l'article de H. Takahata [23] qui étudie exactement le même problème. La condition de mélange (Dobrushin) est moins forte que la nôtre et les résultats sont obtenus en utilisant la technique de Stein. Pour des champs m-dépendant et sous une condition L^8, la vitesse obtenue est optimale tandis que pour des champs α-fortement mélangeant à décroissance exponentielle et sous une condition $L^{8+\delta}$, $\delta>0$, cette vitesse optimale est ralentie par un facteur $(\log\sigma_n)^d$.

Egorov [8], Sergin [20] et Maejina [13] utilisent la technique de Bernstein (formation de paquets) et améliorent successivement les vitesses obtenues. Avec l'article de Sergin [21], la vitesse optimale est obtenue. Sur la base d'une autre technique d'évaluation d'espérence conditionnelle et sous une condition L^8, Stein [22] obtient également cette vitesse optimale dans le cas stationnaire et une vitesse presque optimale pour un processus α-fortement mélangeant, α étant à décroissance exponentielle. Bolthausen [4] obtient aussi une vitesse optimale dans le cas de chaines de Markov récurrentes et fortement mélangeantes.

S'inspirant de l'article de Stein, Tikhomirov [24], sur la base d'une évaluation directe de la dérivée de la fonction caractéristique $f_n(t)$ de la somme étudiée, obtient des résultats analogues pour un processus stationnaire sous la seule condition $L^{2+\delta}$, $\delta>0$. L'hypothèse de stationnarité est levée par Schneider [19]. Comme on va le voir dans le cas spatial, ni la stationnarité, ni la géométrie des domaines sur lesquels on somme X, ne jouent un rôle particulier dans le problème de vitesse de convergence. Dans ce travail, nous suivrons, en la généralisant à $\mathbb{Z}^d$, la technique de Tikhomirov.

Dans l'étude des théorèmes limites pour les champs ($d \geqq 2$) deux notions de mélange apparaissent. L'une où seule la distance entre les deux sous-ensembles sur lesquels on examine les tribus intervient (Deo [5], Gorodetskii [9]). L'autre, introduite par Dobrushin [7] dans le contexte des champs de Gibbs, fait intervenir également les sous-ensembles eux-mêmes, parfois uniquement par l'intermédiaire de leur cardinal (Nahapetian [14], Neaderhauser [15, 16], Bolthausen [3]). C'est la première notion, plus forte, que nous considérons.

Dans le cas m-dépendant, la vitesse de convergence du T.L.C. a été étudiée par Leonenko [12] qui obtient le même résultat que Petrov dans le cas $d \geqq 1$ et par Prakasa-Rao [18]. C'est la conjecture annoncée par ce dernier qui a motivé notre travail: sa conjecture est inexacte et la forme du domaine n'intervient pas. Le cas de champs mélangeants est étudié par Neaderhauser [16] et Takahata [23] (voir note (1)). Berkes et Morrow [1] donnent une évaluation de l'approximation des sommes partielles multidimensionnelles par un drap brownien.

2. Notations et resultats

Soit $X=(X_j)_{j\in\mathbb{Z}^d}$ un champ centré vérifiant (1 — 1); soit (A_n) une suite strictement croissante de domaines de $\mathbb{Z}^d$ à laquelle sont associées les suites (S_n), (σ_n):

$$S_n=\sum_{A_n} X_j, \qquad \sigma_n^2=\operatorname{Var} S_n. \tag{2-1}$$

La seule condition que devra vérifier la suite (A_n), relative au processus X, est:

$$\liminf \sigma_n^2 \cdot |A_n|^{-1}=\alpha>0 \tag{2-2}$$

($|A|$ est le cardinal de A).

Soit ϕ la fonction de répartition de la normale réduite. La distance entre S_n renormalisée et la normale réduite est évaluée par:

$$\Delta_n = \sup_x |P(S_n/\sigma_n \leqq x) - \phi(x)|. \tag{2-3}$$

On considérera par la suite la notion de mélange fort suivante: soit A, B deux parties de $\mathbb{Z}^d$, $d(A, B)$ leur distance, où la distance entre deux points x, y de $\mathbb{Z}^d$ est celle du sup: $d(x, y) = \sup_{1 \leqq i \leqq d} |x_i - y_i|$, et $\mathcal{F}_A$, $\mathcal{F}_B$ les σ-algèbres engendrées par X sur A, B.

On dira que le champ X est α-mélangeant si, pour tout couple de parties A, B, on a:

$$\sup_{E \in \mathcal{F}_A,\, F \in \mathcal{F}_B} |P(E \cap F) - P(E)\,P(F)| \leqq \alpha(d(A, B)) \tag{2-4}$$

où $\alpha(d) \to 0$ quand $d \to +\infty$.

On dira que la champ est m-dependant si $\alpha(d) = 0$ pour $d \geqq m$. On a les résultats suivants:

Théorème 1. *Soit X un champ centré et m-dépendant de $\mathbb{Z}^d$, (A_n) une suite strictement croissante de domaines de $\mathbb{Z}^d$, vérifiants* (1-1), (2-2). Alors:

a) *Si* $0 < \delta < 1$, $\Delta_n = O(\sigma_n^{-\delta})$.

b) *Si* $\delta \geqq 1$, $\Delta_n = O((\operatorname{Log} \sigma_n)^{\frac{d-1}{2}} \cdot \sigma_n^{-1})$

Notant $a \wedge b$ l'inf de a et de b.

Théorème 2. *Supposons que X soit α-mélangeant, α à décroissance exponentielle; alors:*

$$\Delta_n = O[(\operatorname{Log} \sigma_n)^{d[(1-\delta) \wedge 2]} \cdot \sigma_n^{-(\delta \wedge 1)}].$$

Si de plus X est dans $L^{1+\delta}$, $\delta > 0$, alors:

$$\Delta_n = O[(\log \sigma_n)^d \cdot \sigma_n^{-1}].$$

Théorème 3. *Supposons que X soit α-mélangeant, où α est à décroissance puissance: $\alpha(m) = O(m^{-a})$. Posons:*

$$a = \beta \cdot \frac{2d(2+\delta)\,[(1+\delta) \wedge 2]}{\delta \cdot (\delta \wedge 1)}.$$

Alors, dès que $\beta > 1$, on a le contrôle suivant de Δ_n:

$$\Delta_n = O(\sigma_n^{-E(\beta, \delta)}), \quad \text{où}$$

$$E(\beta, \delta) = (\delta \wedge 1) \cdot \frac{2(\beta - 1)}{2\beta + (\delta \wedge 1)}.$$

3. La décomposition de Tikhomirov

Les résultats énoncés au §2 sont obtenus en évaluant la distance entre $f_n(t)$, la fonction caractéristique de S_n, et $\exp\left(-\frac{t^2}{2}\right)$ puis en appliquant l'inégalité de

Esseen (voir [10]). A cette fin, Tikhomirov utilise un développement de $f_n'(t)$.

Nous employerons les notations suivantes:

$$S_j^{(l)} = \frac{1}{\sigma_n} \sum_{A_n \cap \overline{B(j,lm)}} X_j, \quad l \geqq 1, \qquad S_j^{(0)} = \frac{1}{\sigma_n} \sum_{j \in A_n} X_j$$

$$\xi_j^{(l)} = e^{it(S_j^{(l-1)} - S_j^{(l)})} - 1, \quad l \geqq 1,$$

où $B(j, lm)$ représente la boule de centre j et de rayon lm, et $\overline{B(j, lm)}$ son complément.

Pour $k > 2$, on a par induction ([24]):

$$f_n'(t) = \frac{i}{\sigma_n} \Big[\sum_{j \in A_n} E(X_j e^{itS_j^{(1)}}) + \sum_{r=2}^{k} \sum_{j \in A_n} E\left(X_j \prod_{l=1}^{r-1} \xi_j^{(l)} e^{itS_j^{(r)}} \right) + \sum_{i \in A_n} E\left(X_j \prod_{l=1}^{k} \xi_j^{(l)} e^{itS_j^{(k)}} \right) \Big].$$

En écrivant:

$$E\left(X_j \prod_{l=1}^{r-1} \xi_j^{(l)} e^{itS_j^{(r)}} \right) = E\left(X_j \prod_{l=1}^{r-1} \xi_j^{(l)} \{ e^{itS_j^{(r)}} - E(e^{itS_j^{(r)}}) \} \right) + E\left(X_j \prod_{l=1}^{r-1} \xi_j^{(l)} \right) E(e^{itS_j^{(r)}})$$

et en ajoutant et soustrayant $f_n(t)$ de $E(e^{itS_j^{(r)}})$ dans le 2ème membre, on obtient:

$$f_n'(t) = T_0 + \sum_{r=2}^{k} (T_1(r) + T_2(r) + T_3(r)) + R_k \tag{3-1}$$

avec

$$T_0 = \frac{i}{\sigma_n} \sum_{j \in A_n} E(X_j e^{itS_j^{(1)}})$$

et pour $2 \leqq r \leqq k$, en notant $a_j(r) = E\left(X_j \prod_{l=1}^{r-1} \xi_j^{(l)} \right)$

$$T_1(r) = \frac{i}{\sigma_n} \sum_{j \in A_n} a_j(r) f_n(t), \qquad T_2(r) = \frac{i}{\sigma_n} \sum_{j \in A_n} a_j(r) \{ E(e^{itS_j^{(r)}}) - f_n(t) \}$$

$$T_3(r) = \frac{i}{\sigma_n} \sum_{j \in A_n} E\left(X_j \prod_{l=1}^{r-1} \xi_j^{(l)} \{ e^{itS_j^{(r)}} - E(e^{itS_j^{(r)}}) \} \right)$$

$$R_k = \frac{i}{\sigma_n} \sum_{j \in A_n} E\left(X_j \prod_{l=1}^{k} \xi_j^{(l)} e^{itS_j^{(k)}} \right).$$

Dans le cas d'un processus m-dépendant, T_0 et $T_3(r)$ sont nuls et la décomposition (3-1) se réduit à:

$$f_n'(t) = \sum_{r=2}^{k} (T_1(r) + T_2(r)) + R_k. \tag{3-2}$$

Dans la décomposition (3-1) ou (3-2) de $f_n'(t)$, $T_1(2)$ contribuera un terme en $-tf_n(t)$ à l'équation différentielle. La vitesse de convergence sera obtenue en évaluant l'ordre des termes restants. En écrivant l'équation différentielle sous la forme:

$$\begin{aligned} &f_n'(t)-tf_n(t)+a_n(t)f_n(t)+h_n(t) \text{ avec} \\ &a_n(t)=a_{n,0}\,\theta_{n,0}(t)+a_{n,1}\,\theta_{n,1}(t)\,|t|+a_{n,2}\,\theta_{n,2}(t)\,|t|^{(1+\delta)\wedge 2},\ \delta>0 \\ &h_n(t)=b_{n,0}\,\hat{\theta}_{n,0}(t)+b_{n,1}\,\hat{\theta}_{n,1}(t)\,|t|+\sum_{\Delta} b_{n,s}\,\hat{\theta}_{n,s}(t)\,|t|^s \end{aligned} \tag{3-3}$$

où Δ est une partie finie de $]1, +\infty[$; $a_{n,j}, b_{n,j}$ réels positifs, on utilisera:

Lemme 3.1. *Supposons que $f_n(t)$ vérifie* (3-3), *les fonctions θ, $\hat{\theta}$ étant continues, uniformément bornées en n, j, t pour $|t| \leqq T_0(n)$, $a_n = \max\{a_{n0}, a_{n1}, a_{n2}\}$ tendant vers* 0.

Soit

$$\theta_{n,2}=\sup_{|t|\leqq T_0(n)}|\theta_{n,2}(t)|, \qquad T_1(n)=\inf\left(T_0(n), \frac{1}{4a_{n,2}\,\theta_{n,2}}\right),$$

alors pour n grand:

$$\Delta_n \leqq \inf_{T\leqq T_1(n)}\left(\max\left\{\varepsilon_n(T), \frac{1}{T}\right\}\right)$$

où

$$\varepsilon_n(T)=\int_{-T}^{T}\left|\frac{f_n(t)-e^{-t^2/2}}{t}\right|dt \leqq Ba_n + C\left[b_{n,0}+b_{n,1}\log T+\sum_{\Delta} b_{n,s}\,T^{s-1}\right]$$

B, C étant des constantes finies.

4. Vitesse de convergence pour un processus *m*-dependant

4.1. Lemmes fondamentaux

Dans ce qui suit, nous utiliserons à plusieurs reprises les résultats suivants:

$$\begin{aligned} &|e^{ix}-1| \leqq x \\ &e^{ix}-1-ix=\theta_j(x)\,|x|^{(1+\delta)\wedge 2} \quad \text{où} \quad \theta_j(x) \quad \text{est bornée sur } \mathbb{R} \text{ par } \theta_j. \end{aligned} \tag{4-1}$$

Lemme 4.1. $|a_j(r)| \leqq c_d \|x\|_2 \left(\dfrac{|t|\,b_m(r-1)^{\frac{d-1}{2}}}{\sigma_n}\right)^{r-1} \times (r-1)^{\frac{d-1}{4}} = a(r),\ r \geqq 2,$

où

$$c_d=(2\sqrt{2\pi})^d \quad et \quad b_m=(2m)^d\sqrt{d}\,\|X\|_2\,e^{\frac{-(d-1)}{2}}.$$

Démonstration. $|a_j(r)| \leqq E\left|X_j \prod_{\substack{l=1 \\ l \text{ pair}}}^{r-1} \xi_j^{(l)} \prod_{\substack{l=1 \\ l \text{ impair}}}^{r-1} \xi_j^{(l)}\right|.$

Comme les $\{\xi_j^{(l)}, l \text{ pair (resp. impair)}\}$ sont indépendants entre eux, en appliquant l'inégalité de Hölder, on obtient:

$$|a_j(r)| \leqq \|X\|_2 \prod_{l=1}^{r-1} \|\xi_j^{(l)}\|_2 .$$

Soit $C_{n,l}$ l'intersection de A_n et de la couronne de centre j et de rayons $[(l-1)m, lm]$.

En utilisant (4-1):

$$\|\xi_j\|_2^2 < \frac{t^2}{\sigma_n^2} E\Big(\Big(\sum_{C_{n,l}} X_j\Big)^2\Big),$$

$$E\Big(\sum_{C_{n,l}} X_j\Big)^2 \leqq (2m)^d \times 2dm(2lm)^{d-1} \|X\|_2^2$$

à cause de la m-dépendance. On obtient donc:

$$|a_j(r)| \leqq \|X\|_2 \left(\frac{|t|}{\sigma_n}(2m)^d \sqrt{d}\, \|X\|_2\right)^{r-1} ((r-1)!)^{\frac{d-1}{2}}$$

et la démonstration est achevée en utilisant l'inégalité de Sterling.

Lemme 4.2.

$$|T_2(r)| \leqq a(r) \frac{|t|}{\sigma_n^2} \{|A_n|^{\frac{1}{2}} r^d B_1(m, \|X\|_2) + |A_n| r^{\frac{d}{2}} B_2(m, \|X\|_2) |f_n(t)|\}$$

avec

$$B_1(m, \|X\|_2) = 3^{d/2}(2m)^{3d/2} \|X\|_2$$

et

$$B_2(m, \|X\|_2) = (2m)^d \|X\|_2 .$$

Démonstration. Posons:

$$\eta_j(r) = e^{it(S_j^{(r)} - S_j^{(0)})} - 1, \qquad r \geqq 2$$

$$|T_2(r)| \leqq \frac{1}{\sigma_n}\left\{E\Big|\sum_{j \in A_n} a_j(r)[\eta_j(r) - E(\eta_j(r))]\Big| + |f_n(t)|\ \Big|\sum_{j \in A_n} a_j(r) E(\eta_j(r))\Big|\right\}$$

$$E\Big|\sum_{j \in A_n} a_j(r)[\eta_j(r) - E(\eta_j(r))]\Big| \leqq a(r) \cdot \sum_{A_n \times A_n} \operatorname{cov}(\eta_j(r), \eta_p(r))^{1/2}$$

$\operatorname{cov}(\eta_j^{(r)}, \eta_p^{(r)}) = 0$ dès que $d(j, p) > 3rm$. Dans le cas contraire, en utilisant (4-1) et la m-dépendance, on obtient:

$$\operatorname{cov}(\eta_j^{(r)}, \eta_p^{(r)}) \leqq \frac{t^2}{\sigma_n^2}(2m)^d (2rm)^d \|X\|_2^2,$$

et de manière analogue:

$$|E(\eta_j(r))| \leqq \frac{|t|}{\sigma_n}(2m)^{d/2}(2rm)^{d/2} \|X\|_2$$

ce qui permet de conclure la démonstration.

Lemme 4.3.

$$T_1(2)=f_n(t)\Big(-t+|t|^{(1+\delta)\wedge 2}\frac{|A_n|}{\sigma_n^{(2+\delta)\wedge 3}}\theta_\delta(t)(2m)^{d[(1+\delta)\wedge 2]}\|X\|_{2+\delta}^{(2+\delta)\wedge 3}\Big)$$

où $\theta_\delta(t)$ *est une fonction bornée sur* $\mathbb{R}$.

Démonstration. Utilisant (4-1):

$$E(X_j\xi_j^{(1)})=E[X_j\{it(S_j^{(0)}-S_j^{(1)})+\theta_\delta(t(S_j^{(0)}-S_j^{(1)}))\,|t|^{(1+\delta)\wedge 2}\,|S_j^{(0)}-S_j^{(1)}|^{(1+\delta)\wedge 2}\}]$$

Notons:

$$\sum_{j\in A_n}E(X_jS_j^{(0)})=\sigma_n,$$

$$\sum_{j\in A_n}E(X_jS_j^{(1)})=0 \qquad \text{à cause de la } m\text{-dépendance.}$$

La démonstration est conclue en évaluant:

$$\begin{aligned}|E(X_j|S_j^{(0)}-S_j^{(1)}|^{(1+\delta)\wedge 2})| &\leqq \|X\|_{2+\delta}E[|S_j^{(0)}-S_j^{(1)}|^{[(1+\delta)\wedge 2](\frac{2+\delta}{1+\delta})}]^{\frac{1+\delta}{2+\delta}}\\ &\leqq \|X\|_{2+\delta}\sigma_n^{-[(1+\delta)\wedge 2]}(2m)^{d[(1+\delta)\wedge 2]}\|X\|_{2-\delta}^{(1-\delta)\wedge 2}.\end{aligned}$$

4.2. La décomposition de $f_n(t)$

On utilisera la décomposition (3-2) et l'on se place sous l'hypothèse (2-2).

Examen de $T_1(2)$. D'après le lemme 4-3. $T_1(2)$ contribuera une expression de la forme:

$$-tf_n(t)+\frac{(2m)^{d[(1+\delta)\wedge 2]}}{\sigma_n^{\delta\wedge 1}}\theta_{n,2}^*|t|^{(1+\delta)\wedge 2}f_n(t)$$

à l'équation différentielle. La vitesse de convergence sera donc limitée (Lemme 3.1) par $a_{n,2}^*=\sigma_n^{-(\delta\wedge 1)}$ avec une constante en m égale à $(2m)^{d[(1-\delta)\wedge 2]}$.

Examen de R_k. Soit $\varepsilon>0$. k_n le premier entier supérieur à $\varepsilon\log\sigma_n$ et $T_0(n)=\frac{\sigma_n}{eb_m(\varepsilon\log\sigma_n)^{(d-1)/2}}$. Alors pour $k=k_n-1$, $|t|\leqq T_0(n)$, utilisant le Lemme 4.1.:

$$|R_k|\leqq\frac{|A_n|}{\sigma_n}c_d\|X\|_2\left(\frac{1}{e}\right)^k k^{\frac{d-1}{4}}.$$

Choisissons maintenant ε assez grand pour que:

$$\left(\frac{1}{e}\right)^{\varepsilon\log\sigma_n-1}\times(\varepsilon\log\sigma_n)^{\frac{d-1}{4}}<\sigma_n^{-2}$$

alors $R_k=\frac{1}{\sigma_n}\hat{\theta}_{n,0}(t)$ avec $|\hat{\theta}_{n,0}(t)|<c_d\|X\|_2$.

Examen des termes $\sum_3^k T_1(r)$, $k=k_n-1$. En appliquant à nouveau le Lemme 4.1 et en faisant sortir du crochet une puissance $(1+\delta)\wedge 2$, on obtient pour $|t|\leqq T_0(n)$:

$$\left|\sum_{r=3}^{k}\frac{i}{\sigma_n}\sum_{j\in A_n}E\left(X_j\prod_{l=1}^{r-1}\xi_j^{(l)}\right)\right|\leqq\frac{|A_n|}{\sigma_n}\left(\frac{|t|\,b_m}{\sigma_n}\right)^{(1+\delta)\wedge 2}\cdot\left(\sum_{r=3}^{\infty}\left(\frac{1}{e}\right)^{r-[(2+\delta)\wedge 3]}\cdot(r-1)^{\left(\frac{d-1}{4}\right)[(3+2\delta)\wedge 5]}\right)$$

$$\Rightarrow\sum_3^r T_1(r)=\frac{|t|^{(1+\delta)\wedge 2}}{\sigma_n^{\delta\wedge 1}}\theta_{n,2}(t)f_n(t),$$

avec $|\theta_{n,2}(t)|\leqq b_m^{(1+\delta)\wedge 2}\cdot e^{-3}\cdot\Gamma\left(\frac{5d-1}{4}\right)$.

Examen des termes $\sum_2^k T_2(r)$, $k=k_n-1$. D'après le Lemme 4.2, ces termes font intervenir deux expressions:

(a) $L_1=\frac{|t|}{\sigma_n^2}|A_n|^{1/2}B_1(m,\|X\|_2)\sum_{r=2}^{k}a(r)r^d$

(b) $L_2=\frac{|t|}{\sigma_n^2}|A_n|\,B_2(m,\|X\|_2)|f_n(t)|\sum_{r=2}^{k}a(r)r^d$.

L'évaluation de L_1 où L_2 se fait de manière semblable, en appliquant le Lemme 4.1 et en faisant sortit du crochet une puissance $(\delta\wedge 1)$.

On obtient pour $|t|\leqq T_0(n)$:

$$L_1\leqq\frac{|t|^{(1-\delta)\wedge 2}}{\sigma_n^{(1-\delta)\wedge 2}}\theta_2(t)$$

avec $|\theta_2(t)|\leqq B_1(m,\|X\|_2)c_d\|X\|_2 b_m^{\delta\wedge 1}e^{-3}\Gamma\left(\frac{7d+1}{4}\right)$

$$L_2\leqq\frac{|t|^{(1+\delta)\wedge 2}}{\sigma_n^{\delta\wedge 1}}$$

avec $|\theta_3(t)|\leqq B_2(m,\|X\|_2)\times c_d\|X\|_2 b_m^{\delta\wedge 1}e^{-3}\Gamma\left(\frac{5d+1}{4}\right)$.

Par conséquence, pour $|t|\leqq T_0(n)$:

$$\sum_{r=2}^{k}T_2(r)=\frac{|t|^{(1+\delta)\wedge 2}}{\sigma_n^{\delta\wedge 1}}\theta'_{n,2}(t)f_n(t)+\frac{|t|^{(1+\delta)\wedge 2}}{\sigma_n^{(1+\delta)\wedge 2}}\hat{\theta}_{n,2}(t)$$

avec $|\theta'_{n,2}(t)|\leqq|\theta_3(t)|$ et $|\hat{\theta}_{n,2}(t)|\leqq\theta_2(t)$.

4.3. La vitesse de convergence

On déduit du §4.1 que pour $|t| \leqq T_0(n)$, $f_n''(t)$ satisfait à l'équation suivante:

$$f_n''(t) = -tf_n(t) + a_{n,2}|t|^{(1+\delta)\wedge 2}\theta_{n,2}(t) f_n(t) + b_{n,0}\hat{\theta}_{n,0}(t) + b_{n,2}(t)|t|^{(1+\delta)\wedge 2}$$

où $a_{n,2} = \sigma_n^{-(\delta\wedge 1)}$, $b_{n,0} \leqq \sigma_n^{-1}$ et $b_{n,2} = \sigma_n^{-[(1+\delta)\wedge 2]}$.

Nous devons maintenant distinguer deux cas:

(a) $\delta < 1$: $T_1(n) = \inf(T_0(n), a_{n,2}^{-\frac{1}{2}}) = a_{n,2}^{-\frac{1}{2}}$

$$\varepsilon_n(T) \leqq \frac{B}{\sigma_n^{\delta}} + C\left[\frac{1}{\sigma_n} + \frac{T^{\delta}}{\sigma_n^{(1+\delta)\wedge 2}}\right], \qquad 0 \leqq T \leqq T_1(n)$$

et la vitesse de convergence est donc en $\sigma_n^{-\delta}$.

(b) $\delta \geqq 1$: $T_1(n) = \inf(T_0(n), a_{n,2}^{-\frac{1}{2}}) = T_0(n) = \dfrac{\sigma_n}{eb_m(\varepsilon \log \sigma_n)^{(d-1)/2}}$

$$\varepsilon_n(T) \leqq \frac{B}{\sigma_n} + C\left[\frac{1}{\sigma_n} + \frac{T}{\sigma_n^2}\right], \qquad 0 \leqq T \leqq T_1(n).$$

La vitesse de convergence est donc limitée par $\dfrac{1}{T_1(n)}$ ce qui conclut la démonstration du Théorème 1 (b).

Remarque. Pour $\delta < 1$, la vitesse optimale est atteinte. Le seul terme restrictif est le coefficient de $|t|^{1+\delta}$ dans $T_1(2)$. Dans ce cas il n'est pas nécessaire de pousser le développement de $f_n''(t)$ jusqu'à un ordre $k = k_n - 1$ non borné. Il suffirait de choisir $k = \left[\dfrac{1}{1-\delta}\right] + 1$ pour obtenir cette vitesse.

5. Cas d'un champ α-mélangeant: démonstration des théorèmes 2 et 3

Si X est α-mélangeant, nous avons les inégalités suivantes (cf. Hall et Heyde [11]).

$$|Y_1| \leqq C_1, |Y_2| \leqq C_2 : |E(Y_1\, Y_2) - E(Y_1)\, E(Y_2)| \leqq 4C_1\, C_2\, \alpha(Y_1, Y_2) \qquad (\alpha)$$

($\alpha(Y_1, Y_2)$ est à prendre en un sens naturel).

Cett inégalité se généralise sans difficultés à:

$$|E(Y_1\, Y_2 \ldots Y_r)| \leqq \prod_1^r |E(Y_k)| + 4(r-1)\, C_1 \ldots C_r\, \alpha(m)$$

dès que $|Y_k| \leqq C_k$, $k = 1, r$, les Y étant deux à deux "distantes" de m.

$$|X| \leqq C,\ Y \in L^p,\ p > 1:$$

$$|E(XY) - E(X)\, E(Y)| \leqq 6C\, \|Y\|_p\, \alpha(X, Y)^{1-\frac{1}{p}} \qquad (\beta)$$

$X \in L^p$, $Y \in L^q$, $\frac{1}{p}+\frac{1}{q}<1$:

$$|E(XY)-E(X)E(Y)| \leqq 8 \|X\|_p \|Y\|_q \alpha(X,Y)^{1-\frac{1}{p}-\frac{1}{q}} \qquad (\gamma)$$

Soit la condition:

$$C^2 = \sum_0^\infty v^{d-1} \alpha(v)^{\frac{\delta}{2+\delta}} < \infty \qquad \text{(C0)}$$

On a le résultat préliminaire:

Lemme 5.0. *Sous* (C0). $E(\sum_A X_j)^2 \leqq 16 \cdot C^2 \cdot |A| \, \|X\|^2_{2+\delta}$.

Démonstration. Utilisant (γ) avec $p=q=2+\delta$, on obtient:

$$E(\sum_A X_j)^2 \leqq 8 \|X\|^2_{2+\delta} |A| \cdot \sum_A \alpha(j)^{\frac{\delta}{2+\delta}}.$$

Le résultat s'obtient en constatant que le nombre de j à une distance v de l'origine est majoré par $2d\,v^{d-1}$.

5.1. Les trois lemmes fondamentaux

Soit m un entier positif.

Lemme 5.1. *Supposons que t vérifie:* $S = 4d \frac{|t|}{\sigma_n} (2m)^d \, C \|X\|_{2+\delta} \leqq 1$. *Alors:*

$$|a_j(r)| \leqq \|X\|_{2+\delta} \Big\{ 6r \cdot 2^r \alpha(m)^{\frac{\delta}{2+\delta}} + C_1 (r-1)^{\frac{5(d-1)}{4}} \cdot \Big[C_2 \frac{|t|}{\sigma_n} m^{\frac{d}{2}} (r-1)^{\frac{d-1}{2}} \Big] e(r) \Big\} = a(r)$$

avec: $e(r)=1$ *si* $r=2$: $e(r)=2p$ *si* $r=2p+1$ *ou* $2p+2$, $p \geqq 1$

$$C_1 = (2\sqrt{2\pi})^{\frac{d-1}{2}}, \qquad C_2 = 4d\, 2^{\frac{d}{2}} \cdot C \cdot e^{\frac{1-d}{2}} \|X\|_{2+\delta}.$$

Démonstration. Cas, $r-1=1$: Il suffit d'utiliser l'inégalité de Hölder habituelle pour le produit $X_j \cdot \xi_j^{(1)}$ et le Lemme 5.1.1.

Cas, $r-1 \geqq 2$: Des $\xi_j^{(l)}$ apparaissent pour l pair et l impair. Notons Π' le produit relatif aux l impair, Π'' relatif aux l pair. Par Hölder;

$$|a_j(r)| \leqq \|\Pi' \xi_2^{(l)}\|_2 \, \|X_j \Pi'' \xi_j^{(l)}\|_2. \qquad (5\text{-}1)$$

Utilisant (α) pour Π', (β) avec $p=\frac{2+\delta}{2}$ pour $X_j \Pi''$, on a:

$$E|\Pi' \xi_j^{(l)}|^2 \leqq r \cdot 2^{r+1} \alpha(m) + \Pi' E(\xi_j^{(l)})^2$$

$$E|X_j \Pi'' \xi_j^{(l)}|^2 \leqq 6 \|X\|_{2+\delta}^2 \cdot 2^{r-1} \alpha(m)^{\frac{\delta}{2+\delta}} + \|X\|_{2+\delta} E(\Pi'' \xi_j^{(l)})^2.$$

L'inégalité (α) peut à nouveau être utilisée pour évaluer $E(\Pi'')^2$. Réunissant ces majorations dans (5-1), on obtient:

$$|a_j(r)| \leqq \|X\|_{2+\delta}(u+v)$$
$$u = 6r\,2^r\,\alpha(m)^{\delta/2+\delta}, \qquad v' = \Pi' E(\xi_j^{(l)})^2, \qquad v'' = \Pi'' E(\xi_j^{(l)})^2 \tag{5-2}$$
$$\text{et } v = \max\{v', v''\}.$$

Le leme 5.0. permet d'évaluer v', v''

$$v' \leqq (\Pi' l)^{\frac{d-1}{2}} S^2[\tfrac{r}{2}], \qquad v'' \leqq (\Pi'' l)^{\frac{d-1}{2}} S^2[\tfrac{r-1}{2}]$$

($[a]$ est la partie entière de a); utilisant la majoration de Sterling d'un factoriel et $|S| \leqq 1$, on obtient:

$$v \leqq (r-1)^{\frac{3(d-1)}{4}} \cdot C_1 \cdot \left(\frac{r-1}{e}\right)^{\frac{(r-1)(d-1)}{2}} S^2[\tfrac{r-1}{2}].$$

L'examen $r = 2p+1$, $r = 2p+2$ pour $p \geqq 1$ conduit directement au résultat anoncé.

Lemme 5.2. *Majoration du terme $T_2(r)$.*
Dès que $m \geqq 2^{d-2} \cdot C^2$, on a:

$$|T_2(r)| \leqq a(r) \frac{|t|}{\sigma_n^2} \{|f_n(t)|\, C_3 |A_n| (rm)^{\frac{d}{2}} + C_4 |A_n|^{\frac{1}{2}} (rm)^{\frac{3d}{2}}\}$$

avec: $C_3 = 4.2^{d^2}\sqrt{d}\,\|X\|_{2+\delta} \cdot C$; $C_4 = 2^{2(d-1)}\,\|X\|_{2+\delta}$.

Démonstration. La démonstration suit la même démarche que celle du Lemme 4.2. dont nous reprenons les notations.

Dans un premier temps, il faut majorer $\sum_{j,p} \operatorname{cov}(\eta_j(r), \eta_p(r))$. Utilisant l'inégalité (γ) avec $p = q = 2+\delta$, on obtient:

$$\Big|\sum_{A_n \times A_n} \operatorname{cov}(\eta_j(r), \eta_p(r))\Big| \leqq 8|A_n|\,\|\eta(r)\|_{2+\delta}^2 \sum_{A_n} \alpha(d(B(0, rm), B(j, rm)))^{\frac{\delta}{2+\delta}}$$

où $\|\eta(r)\|_{2+\delta}$ majore les $\|\eta_j(r)\|_{2+\delta}$.

— Si $d(0, j) \leqq 2rm$, la distance entre les deux boules est nulle. Il y a $(4rm)^d$ tels j.

— Si $d(0, j) = 2rm + v$, $v > 0$, la distance entre les deux boules est v. Il y a $(4rm + 2v)^{d-1}$ tels j. Une suite de majorations élémentaires permet d'obtenir, sous (CO):

$$\sum_{j \in A_n} \alpha(d(B(0, rm), B(j, rm))) \leqq 2^{2d-1}(rm)^d$$

dès que $m \geqq 2^{d-2} \cdot C^2$. D'autre part:

$$\|\eta(r)\|_{2+\delta} \leqq \frac{|t|}{\sigma_n} (2rm)^d \|X\|_{2+\delta}.$$

Ceci conduit à:

$$E\left|\sum_{A_n} a_j(r)(\eta_j(r) - E\eta_j(r))\right| \leqq 2^{2(d+1)} |A_n|^{\frac{1}{2}} \frac{|t|}{\sigma_n} a(r)(rm)^{\frac{3d}{2}} \|X\|_{2+\delta}$$

D'autre part, on a:

$$\left|\sum_{A_n} a_j(r) E(\eta_j(r))\right| \leqq a(r)|A_n| \|\eta(r)\|_2.$$

Le Lemme 5.0. permet alors de conclure.

Lemme 5.3. *Evaluation de* $T_1(2)$

$$T_1(2) = \left(-t + \theta_1(t)|A_n| \alpha(m)^{\frac{\delta}{2-\delta}} \cdot t + \theta_2(t) \frac{m^{d[(1+\delta)\wedge 2]}}{\sigma_n^{\delta \wedge 1}} |t|^{(1+\delta)\wedge 2}\right) f_n(t)$$

où

$$|\theta_1(t)| \leqq 8\alpha \|X\|_{2+\delta}$$

$$|\theta_2(t)| \leqq 2^{d[(1+\delta)\wedge 2]} \alpha \theta_\delta \|X\|_{(2-\delta)\wedge 3}^{(2-\delta)\wedge 3} \qquad (\alpha, \theta_\delta \textit{ définis en } (2.2) \textit{ et } (4.1)).$$

Démonstration. La seule modification par rapport au Lemme 4.3 est l'évaluation de $E(X_j S_j^{(1)})$. Utilisant (γ) avec $p = q = 2+\delta$ et l'inégalité de Minkowski pour $\|S_j^{(1)}\|_{2+\delta}$, on obtient le lemme.

Par la suite les constantes ou les majorants de certaines fonctions de t ne seront pas explicités: de telles constantes notées $\theta \ldots$ seront universelles une fois fixées δ, d, X et (A_n).

5.2. La décomposition de $f_n'(t)$

On part de la décomposition (3.1) de $f_n'(t)$.

Majoration de $T_0 + \sum_2^k T_3(r)$: par application de (β)

$$|T_0| \leqq 6\alpha\sigma_n \|X\|_{2-\delta} \alpha(m)^{\frac{\delta+1}{\delta-2}}$$

$$|T_3(r)| \leqq 6\alpha \cdot \sigma_n \cdot 2^r \|X\|_{2+\delta} \alpha(m)^{\frac{\delta+1}{\delta+2}}$$

et donc:

$$\left|T_0 + \sum_2^k T_3(r)\right| \leqq \theta(t) \alpha(m)^{\frac{\delta+1}{\delta-2}} \cdot 2^k \tag{5-3}$$

Examen de $T_1(2)$: (Lemme 5.3)

$$T_1(2) = f_n(t)\{-t + \theta_1(t)\tilde{a}_{n1} t + \theta_2(t) a_{n2} |t|^{(1+\delta)\wedge 2}\}$$

$$\text{où: } \tilde{a}_{n,1} = |A_n| \alpha(m)^{\frac{\delta}{\delta-2}}, \qquad a_{n2} = m^{d[(1+\delta)\wedge 2]} \sigma_n^{\delta \wedge 1} \tag{5-4}$$

Examen de R_k: Soient $\varepsilon>0$, $K>0$. Choisissons:

$$\cdot\, k=k_n=[\varepsilon \log \sigma_n]+1$$

$$\cdot\cdot\, |t| \leqq T_0(n)=\sigma_n/(e^K C_2 m^{\frac{d}{2}} (\varepsilon \log \sigma_n)^{\frac{d-1}{2}})$$

Si:

$$0 \leqq k \leqq k_n, \left[C_2 \frac{|t|}{\sigma_n} m^{\frac{d}{2}} (k-1)^{\frac{d-1}{2}} \right] \leqq e^{-K}$$

En

$$k=k_n: k_n^{\frac{5(d-1)}{4}} \left[C_2 \frac{|t|}{\sigma_n} m^{\frac{d}{2}} (k_n-1)^{\frac{d-1}{2}} \right]^{k_n-1} \leqq (\varepsilon \log \sigma_n)^{\frac{5(d-1)}{4}} \sigma_n^{-K\varepsilon}$$

A $\varepsilon>0$ fixé, choisir $K(\varepsilon)$ tel que ce second membre soit inférieur à σ_n^{-2}. On obtient par application du Lemme 5.1:

$$R_{k_n}=\theta_3(t)\, k_n\, 2^{k_n}\, \sigma_n\, \alpha(m)^{\frac{\delta}{\delta+2}} + \theta_4(t)\, \sigma_n^{-1} \tag{5-5}$$

Examen de $\sum_3^{k_n} T_1(r)$: $\left|\sum_3^{k_n} T_1(r)\right| \leqq \theta \cdot \sigma_n f_n(t) \sum_3^{k_n} a(r)$.

Dans la majoration de $a(r)$ deux termes apparaissent: le terme en $\alpha(m)$ et le terme complémentaire: de ce second terme, sortons du crochet une puissance $(1+\delta)\wedge 2$ (possible puisque $r \geqq 3$). En constatant que la série:

$$\sum_r r^a\, e^{-K(\varepsilon) e(r)} < \infty$$

on en déduit:

$$\sum_3^{k_n} T_1(r) = f_n(t) \left\{ \theta_5(t)\, \sigma_n\, k_n\, 2^{k_n}\, \alpha(m)^{\frac{\delta}{2+\delta}} + \theta_6(t) \frac{m^{\frac{d}{2}[(1+\delta)\wedge 2]}}{\sigma_n^{\delta\wedge 1}} |t|^{(1+\delta)\wedge 2} \right\} \tag{5-6}$$

Examen de $\sum_2^{k_n} T_2(r)$: Utilisant le Lemme 5.2, sortant cette fois-ci du crochet une puissance $\delta\wedge 1$ (possible puisque $r \geqq 2$), on obtient de façon analogue, une première quantité facteur de $f_n(t)$, une autre non facteur de $f_n(t)$:

$$f_n(t) \left\{ \theta_7(t)\, m^{\frac{d}{2}}\, \alpha(m)^{\frac{\delta}{\delta+2}}\, k_n^{\frac{d}{2}+1}\, 2^{k_n} |t| + \theta_8(t) \frac{m^{\frac{d}{2}[(1+\delta)\wedge 2]}}{\sigma_n^{\delta\wedge 1}} |t|^{(1+\delta)\wedge 2} \right\} \tag{5-7}$$

$$\frac{|t|}{\sigma_n} m^{\frac{3d}{2}} \{\theta_9(t) + \theta_{10}(t)\, k_n^{1+\frac{3d}{2}}\, 2^{k_n}\, \alpha(m)^{\frac{\delta}{\delta+2}}\} \tag{5-8}$$

Reportant les évaluations (5-3) à (5-8) dans le Lemme 3.1 d'intégration, on obtient:

$$
\begin{aligned}
&a_{n0} = k_n 2^{k_n} \alpha(m)^{\frac{\delta}{2+\delta}},\\
&a_{n1} = \max\{\tilde{a}_{n1}, m^{\frac{d}{2}} k_n^{\frac{d}{2}+1} 2^{k_n} \alpha(m)^{\frac{\delta}{2+\delta}}\}, \qquad \tilde{a}_{n1} = |A_n| \alpha(m)^{\frac{\delta}{2-\delta}},\\
&a_{n2} = m^{d[(1+\delta)\wedge 2]}/\sigma_n^{\delta\wedge 1},\\
&b_{n0} = \max\{\sigma_n^{-1}, \sigma_n k_n 2^{k_n} \alpha(m)^{\frac{\delta}{2+\delta}}\},\\
&b_{n1} = \frac{m^{\frac{3d}{2}}}{\sigma_n} \max\{1, k_n^{\frac{3d}{2}+1} 2^{k_n} \alpha(m)^{\frac{\delta}{\delta+2}}\}.
\end{aligned}
\tag{5-9}
$$

On va choisir ε et $m = m(n)$ tels que:

$\tilde{a}_{n1}$ est de l'ordre de a_{n2},
a_{n1} est de l'ordre de $\tilde{a}_{n1}$,
a_{n0} et b_{n0} sont d'un ordre inférieur à a_{n2},
b_{n1} est de l'ordre de $m^{\frac{3d}{2}} \cdot \sigma_n^{-1}$.

Alors:

$$
\Delta_n = O\left(\max\left\{a_{n2}, T_1(n)^{-1}, \frac{\log T_1(n)}{\sigma_n} m^{\frac{3d}{2}}\right\}\right)
$$

où

$$
T_1(n) = \inf\{T_0(n), a_{n2}^{-1}\}.
$$

Dès que a_{n2} tend vers 0 les cinq conditions ci-dessus se réduisent aux deux premières, soit, en notant $\tilde{\varepsilon} = \varepsilon \log 2$:

$$
\sigma_n^2 \alpha(m)^{\frac{\delta}{2-\delta}} \leqq \frac{m^{d[(1+\delta)\wedge 2]}}{\sigma_n^{\delta\wedge 1}} \tag{C1}
$$

$$
m^{\frac{d}{2}} (\log \sigma_n)^{1+\frac{d}{2}} \sigma_n^{\tilde{\varepsilon}} \alpha(m)^{\frac{\delta}{\delta+2}} \leqq a_{n2} \tag{C2}
$$

5.3. *Cas d'un mélange exponentiel*

Le choix $m(n) = a \log \sigma_n$, avec a assez grand et $\tilde{\varepsilon} < 1$ assure (CO) et (C1)–(C2) pour tout $m \geqq m(n)$. Pour ce choix:

$$
T_0(n) = C \cdot \frac{\sigma_n}{(\log \sigma_n)^{d-\frac{1}{2}}}, \qquad T_1(n) = a_{n2}^{-1}
$$

La vitesse est en $(\log \sigma_n)^{d[(1+\delta)\wedge 2]}/\sigma_n^{\delta\wedge 1}$.

Remarques. i) Dans le cas $d = 1$, X stationnaire, $A_n = [1, n]$ et $\delta > 1$, le résultat peut être amélioré. Le Théorème 1 de [25] est applicable (α est à décroissance exponentielle)

$$
\left\| \sum_A X_j \right\|_r \leqq K|A|^{1/2} \qquad \text{si } 2 \leqq r < 2+\delta. \tag{5-10}
$$

L'évaluation de a_{n2} dans le Lemme 5.3 est ainsi modifiée:

$$E(|X_j| |\sum_A X_k|^{(1+\delta)\wedge 2}) \leqq \|X\|_{2-\delta} \|\sum_A X_k\|_r)^{(1+\delta)\wedge 2}$$

avec $r=[(1+\delta)\wedge 2]\dfrac{2+\delta}{1+\delta}$. Si $\delta>1$, alors $2<r<2+\delta$. On en déduit que a_{n2} $=m\cdot\sigma_n^{-1}$. Donc si X est dans $L^{3+\gamma}$, $\gamma>0$, la vitesse est en $(\log\sigma_n)/\sigma_n$.

ii) Dans le cas $d=1$, et sans stationarité, mais sous une condition $L^{4+\gamma}\gamma>0$, on obtient la même vitesse en utilisant l'inégalité (5-10) démontrée en [6] pour les entiers pairs $r\geqq 4$. Quand $d\geqq 2$ et si X est dans $L^{4+\gamma}, \gamma>0$, la vitesse de convergence est en $(\log\sigma_n)^d/\sigma_n$. En effet, l'inégalité (5-10) est une conséquence du lemme suivant, qui présente un intérêt propre:

Lemme 5.4. *Soit q un entier supérieur ou égal à 2, X dans $L^{q+\delta}$, $\delta>0$, α-mélangeant, satisfaisant:*

$$C=\sum_{k\geqq 0} k^{d(q-[\frac{q}{2}])-1}\,\alpha(k)^{\frac{\delta}{q+\delta}}<\infty.$$

Alors, pour toute partie A de $\mathbb{Z}^d$, on a:

$$E(\sum_A X_j)^q \leqq K\cdot|A|^{[\frac{q}{2}]}$$

où K est une constante finie ne dépendant que de q, d, X.

Démonstration. $E(\sum_A X_j)^q=\sum_{A^q} E(X_{j1}, X_{j2}, \ldots, X_{jq})$.

Notons $J=\{j_1, j_2, \ldots, j_q\}$, $D=D(J)=\max_{i=1,q}\{d(\{j_i\}, J\setminus\{j_i\})\}$. X étant centré, on a, utilisant (γ):

$$|E(X_{j_1} X_{j_2}\ldots X_{j_q})| \leqq 8\,\alpha(k)^{\frac{\delta}{q+\delta}}\,\|X\|_{q+\delta}^q$$

Reste à dénombrer l'ensemble des J de A^q tels que $D(J)=k$.
Ce dénombrement résultera de:

Lemme 5.5. *Il existe un recouvrement de J constitué d'au plus $\left[\dfrac{q}{2}\right]$ boules, centrées en des points de J, de rayon $3^q D(J)$, l'un au moins des points de J étant exactement à distance $D(J)$ d'un des centres.*

Admettons provisoirement ce lemme. Le nombre de J tels que $D(J)=k$ est donc majoré par: $q!\,|A|^{[q/2]}(2^d d\cdot k^{d-1})\,(2\cdot 3^q k)^{d(q-[q/2]-1)}$ et donc:

$$|E(\sum_A X_j)^q| \leqq 8q!\,\|X\|_{q+\delta}^q\cdot 2^d\cdot d\cdot(2\cdot 3^q)^{d(q-[\frac{q}{2}]-1)}\sum_{k\geqq 0} k^{d(q-[\frac{q}{2}])-1}\,\alpha(k)^{\frac{\delta}{q+\delta}}.$$

Démontrons le lemme auxiliaire. Il est facile de le vérifier pour $q=2, 3$. Supposons que j_1 réalise $D(J)$, j_2 étant le point le plus proche de j_1. Deux cas se présentent:

1er cas: $B(j_2, D(J)) \cap \{j_3, \ldots, j_q\} = \emptyset$. Alors $D(j_3, j_4, \ldots, j_q)$ est inférieur ou égal à $D(J)$, et donc avec $\left[\frac{q-2}{2}\right]$ boules, de rayon $3^{q-2} D(J)$, centrées en certains points de $j_3, \ldots, j_q$, on recouvre $j_3, \ldots, j_q$. Adjoindre alors la boule $B(j_2, D(J))$ et augmenter les rayons donne le résultat annoncé à l'ordre q.

2ème cas: Supposons que j_3 appartienne à $B(j_2, D(J))$. Montrons alors que $D(j_2, j_3, \ldots, j_q) \leqq 2D(J)$. Si cela n'est pas vrai, alors $D' = D(j_2, \ldots, j_q) > 2D(J)$. Soit j' le j qui réalise D'. Alors $j' \neq j_2$ puisque $d(j_2, j_3) \leqq D(J)$. Donc:

$$d(j', j_1) \geqq D' - D(J) > D(J)$$

et donc:

$$d(\{j'\}, J \setminus \{j'\}) > D(J)$$

ce qui contredit l'hypothèse que $D(J)$ est réalisé en j_1.

On peut donc recouvrir $j_2, \ldots, j_q$ avec $\left[\frac{q-1}{2}\right]$ boules centrées en certains de ces j, de rayon $3^{q-1} \times 2D(J)$. Il suffit d'augmenter de $D(J)$ le rayon de la boule contenant j_2 pour englober j_1. Donc $\left[\frac{q-1}{2}\right]$ boules de rayon $3^{q-1} \times 2D(J) + D(J)$ recouvre J. D'où le résultat.

Pour obtenir la vitesse annoncée, il suffit alors, dans l'inégalité (5-10) de majorer la norme r par la norme 4. On en déduit, comme dans la remarque:

$$a_{n2} = (\log \sigma_n)^{d/2} \sigma_n.$$

5.4. Cas d'un mélange à décroissance puissance

$$\alpha(m) = O(m^{-a}), \qquad a > 0$$

(C0) est équivalente à: $a\delta > d(2+\delta)$,

(C1) est vérifiée si $m = m(n) = \sigma_n^h$, $h = \dfrac{(2+\delta)[(2+\delta) \wedge 3]}{a\delta + d(2+\delta)[(1+\delta) \wedge 2]}$.

Choisissons alors $m = m(n)$:

Le terme a_{n2} tend vers 0 dès que:

$$a = \beta \frac{2d(2+\delta)[(1+\delta) \wedge 2]}{\delta \cdot (\delta \wedge 1)} \qquad \text{avec } \beta > 1 \tag{A}$$

La condition (A) implique (C0). D'autre part, on vérifie facilement que si $\tilde{\varepsilon} < 1$, et si (A) est vérifiée, (C2) est vraie. Pour ce choix $m(n)$:

$$T_0(n) = \frac{\sigma_n}{(\log \sigma_n)^{\frac{d-1}{2}} \sigma_n^{\frac{d}{2} \cdot h}} \geqq a_{n2}^{-1}$$

et donc:

$$\Delta_n = 0\left(\max\left\{a_{n2}, \frac{\log T_0(n)}{\sigma_n} m(n)^{\frac{3d}{2}}\right\}\right)$$

Sous (A), a_{n2} réalise ce max:

$$\Delta_n^{\cdot} = O(a_{n2}), \quad \text{avec:}$$

$$a_{n2} = \sigma_n^{-E(\beta,\delta,d)}$$

$$E(\beta,\delta,d) = \delta \wedge 1 - d[(1+\delta)\wedge 2]\frac{(2+\delta)[(2+\delta)\wedge 3]}{a\delta + d(2+\delta)[(1+\delta)\wedge 2]}$$

$$= (\delta \wedge 1)\cdot\frac{2(\beta-1)}{2\beta+(\delta\wedge 1)} = E(\beta,\delta).$$

La condition (A) porte explicitement sur δ, d, β. L'exposant dans la vitesse ne dépend de d que par l'intermédiaire de (A).

References

1. Berkes, I., Morrow, G.J.: Strong invariance principales for mixing random fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete **57**, 13–37 (1981)
2. Bolthausen, E.: Exact convergence rates in some martingale central limit theorem. Ann. Probability **10**, 672–688 (1982)
3. Bolthausen, E.: On the central limit theorem for stationary random fields. Ann. Probability **10**, 1047–1050 (1982)
4. Bolthausen, E.: The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete **60**, 283–289 (1982)
5. Deo, C.M.: A functional central limit theorem for stationary random fields. Ann. Probability **3**, 708–715 (1975)
6. Doukhan, P., Portal, F.: Principes d'invariance faibles, avec vitesse dans un cadre mélangeant. Prépublication Orsay n° 83T28 (1983)
7. Dobrushin, R.L.: The description of a random field by its conditional distribution and its regularity condition. Theory Probability Appl. 13, 197–227 (1968)
8. Egorov, V.A.: Certain limit theorems for m-dependent random variables. Liet. Mat. Rink **10**, 51–59 (1970) (en Russe)
9. Gorodetskii, V.V.: The invariance principle for stationary random fields satisfying the strong mixing condition. Theory Probability Appl. **27**, 380–385 (1982)
10. Ibragimov, I.A., Linnik, Yu.V.: Independent and stationary sequences of random variables. Groningen: Wolters-Nordhoff 1971
11. Hall, P., Heyde, C.C.: Martingale limit theory and its application. New York: Academic Press 1980
12. Leonenko, N.N.: Estimate of the rate of convergence in the central limit theorem for m-dependent random fields. Mathematical Notes **17**, 76–78 (1975)
13. Maejima: A non uniform estimate in the central limit theorem for m-dependent random variables. Keio Engineering Reports **31**, 15–20 (1978)
14. Nahapetian, B.S.: The central limit theorem for random fields. In Multicomponent Random Systems, pp. 531–542, ed. by R.L. Dobrushin and Ya.G. Sinai. New York: Marcel Dekker 1980
15. Neaderhauser, C.C.: Limit theorems for multiply indexed mixing random variables, with application to Gibbs random fields. Ann. Probability **6**, 207–215 (1978)
16. Neaderhauser, C.C.: Some limit theorems for random fields. Comm. Math. Phys. **61**, 293–305 (1978)
17. Petrov, V.V.: On the central limit theorem for m-dependent variables. Selected Transl. Math. Statist. Probability **9**, 83–88 (1970)
18. Prakasa Rao, B.L.: A non uniform estimate of the rate of convergence in the central limit theorem for m-dependent random fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete **58**, 247–256 (1981)

19. Schneider, E.: On the speed of convergence in the random central limit theorem of ϕ-mixing processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete **58**, 125–138 (1981)
20. Shergin, V.V.: An estimate of the remainder term in the central limit theorem for m-dependent random variables. Lith. Math. J. vol **16**, 637–641 (1977) (trad. de Litov Math. Sbornik, vol. **15**, 245–250 (1976)
21. Shergin, V.V.: On the convergence rate in the central limit theorem for m-dependent random variables. Theory Probability Appl. **24**, 782–796 (1979)
22. Stein, C.: A Bound for the error in the normal approximation of a sum of dependent random variables. Proc. Berkeley Sympos. Math. Statist Probability 2, 583–603 Univ. Calif. (1972)
23. Takahata, H.: On the rates in the central limit theorem for weakly dependent random fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete **64**, 445–456 (1983)
24. Tikhominov, A.N.: On the converge rate in the central limit theorem for weakly dependent random variables. Theory Probability Appl. **25**, 790–809 (1980)
25. Yokoyama, R.: Moment bounds for stationary mixing sequences. Z. Wahrscheinlichkeitstheorie verw. Gebiete **52**, 45–57 (1980)

Reçu le 22 Avril 1982; en forme revisée le 5 Décembre 1983

TESTING THE ASSOCIATION BETWEEN TWO SPATIAL PROCESSES : A REVIEW OF DIFFERENT APPROACHES

TESTING THE ASSOCIATION BETWEEN TWO SPATIAL PROCESSES
A REVIEW OF DIFFERENT APPROACHES

by

Sylvia RICHARDSON

Laboratoire de Statistiques Médicales, Université de Paris V

et

Institut National de la Santé et de la Recherche Médicale, Villejuif

France

Abstract

This paper presents a short review of the statistical problems involved in testing association between two autocorrelated variables.

Résumé

Cet article contient une brève synthèse des problèmes statistiques rencontrés lors d'un test d'association entre deux variables auto-corrélées.

1. INTRODUCTION

A question which is frequently asked in statistics is whether the apparent association between two variables is due to something other than chance. Of particular interest are cases where the variables are observed at a variety of spatial locations.

Typical examples are found in the fields of geography and regional science, for instance when relating consumption of agricultural output and road accessibility (Cliff and Ord [12]). Examples in sociology and political science are described by Doreian [17], with reference for instance to voting behaviour and socio-economic or political factors. In epidemiology, etiological clues to environmental risk-factors are sometimes sought through their joint analysis with disease incidence or mortality maps (Doll [16]).

The data in these examples consist of a set A of N locations, and a pair of variables indexed by their locations. These variables will typically exhibit some degree of spatial autocorrelation. The first step in studying the relationship between these variables is often the testing of statistical independence between pairs of variables, $\{(X_\alpha, Y_\alpha)\ ,\ \alpha \in A\}$. This problem cannot be solved by the usual methods based on correlation and linear regression which require that both $\{X_\alpha\ ,\ \alpha \in A\}$ and $\{Y_\alpha\ ,\ \alpha \in A\}$ represent independent samples.

The consequences of neglecting the autocorrelation in the X and Y processes were first pointed out in the case of stationary times series by Bartlett [3]. Bartlett calculated the asymptotic variance of the empirical correlation coefficient r_{XY} . When X and Y are mutually independent and normally distributed, this variance depends only upon the autocorrelations of both processes. This formula was extended to stationary spatial process in $\mathbb{Z}^2$ by Richardson and Hémon [32]. Numerical calculations given in [33] show how, when X and Y are both positively autocorrelated, first-order nearest neighbour isotropic autoregressive or Markovian processes, the asymptotic variance of r_{XY} increases steeply with the first-order autocorrelation of each process. Moreover, for two dependent processes, Bivand [7], in a Monte Carlo study, made similar observations about the increase in the standard deviation of Fisher's Z transformation of r_{XY} .

In regression analysis of Y on X the effect of autocorrelation in the dependent variables and in the error term has been analysed by several authors. In the case of positive autocorrelation, Johnston [26] for time series, Cliff and Ord [12] and Bennett [4] for spatial series, show that the variance of the slope is inflated and that the least-square estimator of the variance of the errors is biased downwards, leading to over significant t and F statistics. An illustration of the bias between the true variance of the slope and the OLS variance estimate is given by Martin [29] in a Monte Carlo study. Moran's I statistic for testing autocorrelation has been extended by Cliff and Ord [12] to the testing of autocorrelation among

regression residuals. Further work in this area was done by Brandsma and Ketellapper [10] who compared the power of different tests for autocorrelation among regression residuals and concluded that overall Moran's test with ordinary least squares residuals seems to perform best.

The problem of dealing with spatial autocorrelation in testing for association may be tackled in various ways. Time series methods such as the prewhitening of each series before analysis of the cross correlations could be extended. Alternatively, standard measures of association such as the correlation coefficient can be adapted to take account of autocorrelation. This approach will be developed in a subsequent paper and is outlined in Clifford and Richardson [13]. Moreover classical regression analysis can be extended to take specifically into account the spatial structure of the data. Finally new indices of association can be proposed.

2. REVIEW

In time series, Haugh [23] has proposed that univariate time series model should be fitted to each series. The cross correlation coefficients between the innovation processes are then calculated. This set of cross correlation coefficients is found to be asymptotically independent and normally distributed and can be used to perform a χ^2 test of independence of the two series. He also uses prewhitening as a first step in identifying (dynamic) regression models relating the two series. Using this approach, Pierce [31] found only weak evidence of a relationship between pairs of economic time series which were traditionally considered as related. In a subsequent paper, Geweke [18] compares several tests of independence between stationary time series, in particular Haugh's test and an F test on the regression parameters of a mixed regressive-autoregressive model between the two series. He concludes that in many cases the rate of type II error of Haugh's test is larger than that F test of the regression coefficients. For spatial series, no comparable study has been done to date although the prewhitening of series has been discussed by Griffith [20]. It is likely that a similar conclusion to Geweke would hold and that tests of independence or estimates of regression coefficients based on the original series will be more efficient than those based on their residuals after prewhitening.

As for time series, spatial structure has been traditionally accounted for in regression models in two ways, either by considering a mixed simultaneous regressive-autoregressive model :

$$Y = L(Y) + X\beta + \varepsilon \,, \qquad (1)$$

where ε is independent of Y and $L(Y)_i$ represents some linear function of the value of Y in locations j, $j \neq i$ or by considering a regression model with

autocorrelated error term :

$$Y = X\beta + U \ , \qquad (2)$$

where the spatial autocorrelation is reflected in the structure of Σ , the variance covariance matrix of the error U (see, for example, Cliff and Ord [12], Upton and Fingleton [37]).

The interpretation of β is essentially different for the regressions (1) and (2), hence these equations should not be treated as alternatives but the choice between them should be motivated by the problem considered (Cox [15]). In (1) the spatial autocorrelation is concentrated on Y and β represents the effect of X when the influence of neighbouring values of Y has already been taken into account, whereas for (2) the autocorrelation is generated solely by U .

There are cases when equation (1) does not provide a plausible explanatory model, for instance when relating the incidence of a non contagious disease to environmental factors, since the neighbouring disease rates have no direct causal influence. On the other hand, for some examples, a propagation effect of Y is conceivable like in models of regional income diffusion considered by Haining [21], or those on the Huk rebellion considered by Doreian [17].

The estimation of models (1) and (2) was first discussed by Ord [30]. The autoregressive term in (1) is usually modelled as : $L(Y) = \rho WY$, where W is a matrix of weights representing contiguity introduced by Cliff and Ord [12], and ρ is a parameter summarising the overall level of autocorrelation which has tc be estimated. If the error term is multivariate normal with uncorrelated components, the likelihood of Y can be written down and the parameters can be estimated, usually by direct search. The least-square estimates are not consistent for the simultaneous representation in (1) but Cliff and Ord [12] and Besag [5] have pointed out that this is not the case in the equivalent Markovian scheme.

Mardia and Marshall [28] have studied the asymptotic properties of the maximum likelihood estimators for model (2). Supposing that Y is a Gaussian process and that its matrix of variance-covariance Σ can be parametrised in term of a finite number of parameters θ , they give conditions which ensure the consistency and the asymptotic normality of $(\hat{\beta},\hat{\theta})$, the maximum likelihood estimates of (β,θ) .

Models for Σ have been basically of two kinds, either a specific autoregressive model for U is supposed to hold :

$$U = \rho WU + \varepsilon \ , \quad \varepsilon \sim N(0 \ , \sigma^2 I) \ , \qquad (3)$$

which implies that $\Sigma = \sigma^2 (I - \rho W)^{-1} (I - \rho W')^{-1}$, or the elements of Σ are assumed

to depend on distance with a functional form suggested by the variogram of the ordinary least-square regression residuals.

Equation (3) has been fitted by Bodson and Peeters [9], Cliff and Ord [12] and Bivand [8]. Equations (2) and (3) can be combined to give :

$$Y = \rho WY + X\beta - \rho WX\beta + \varepsilon , \qquad (4)$$

and so giving a particular case of a larger class of regression models in which Y is also regressed on WX . Tests of the adequacy of (4) within this larger class are discussed by Burridge [11]. A Bayesian analysis of model (3) has been given by Hepple [24]. Using a uniform diffuse prior for $\beta = (\beta_1, \beta_2)$, ρ and $\log \sigma$, Hepple shows how the conditional posterior distribution of β_2 is sensitive to value of ρ and he derives the bivariate posterior distribution for β_2 and ρ , with mode corresponding to the maximum likelihood estimates of β_2 and ρ .

The arbitrary nature of the definition of W is a common criticism of model (3). Arora and Brown's [2] alternative approach either needs the availability of spatial data at different time intervals or assumes that the spatial autocorrelation is entirely created through an unknown common error term. Their approach does not apply to data on which spatial autocorrelation decreases with distance.

The choice of parametrisation of W by plotting the variogram of the least square residuals has been discussed by Ripley [34], Cook and Pocock [14] and Agterberg [1]. Care has to be taken when interpretating the variogram since it is sensitive to the number of pairs of data points used to estimate the empirical covariance at a particular distance.

The number of pairs will vary with the distance, typically increasing at first. Riplay advises the use of cross-validation by successive deleting of data points to assess the fit of the model chosen for the covariance function. Cook and Pocock fitted an exponential function decreasing with distance, whilst Agterberg uses a quadratic one.

Once a particular parametrisation is chosen, the parameters may be estimated by maximum likelihood, either by grid search or by iteration. Mardia and Marshall [28] recommend updating the parameters by scoring, Cliff and Ord [12] suggest finding a conditional maximum for ρ , β being fixed, then estimating β by generalised least squares and iterating until convergence; a convergence however which is not necessarily guaranteed (see Haining [22]).

Other ways to model W , for instance by a spectral decomposition (Streitberg [35]) or with a finite number of contiguity matrices of increasing order (Lebart [27] have been proposed but are infrequently used.

Clearly, the regression methods described above are computationally involved and entail a certain degree of arbitrariness. Bivand [8] discusses model selection and introduces the use of Akaike's criterion. In a trend surface analysis, Haining [22] compares three different models for Σ in (2).

A non parametric index of association based on the locations of the ranks of two spatially defined variables $\{(X_\alpha, Y_\alpha)\ ,\ \alpha \in A\}$ has been introduced by Tjøstheim [36] and generalised by Hubert and Golledge [25]. This index is defined as the sum of the distances between the locations of similar ranks, with a suitably chosen distance function.

In a simulation study, Glick [19] illustrates the performance of this test in comparison to Spearman's rank correlation. When the X process has no autocorrelation and the Y process corresponds to a simple spatial shift of the X process, Tjøstheim's index Λ detects an association whilst Sperman's rank correlation is non significant. The situation is reversed when the dependence between the X and the Y process has no spatial component. The randomisation model used by Tjøstheim and, Hubert and Golledge to generate a reference distribution against which observed values of Λ are compared under the null hypothesis is that of a spatial redistribution of one of the variables, the other staying fixed. This full set of permutation clearly does not preserve the autocorrelation of the variable and hence the significance of Λ is not correctly assessed in this case. One would need to consider a smaller set of permutations which would respect in some way the spatial autocorrelation of the variable. This would also be the case for the Monte Carlo method proposed by Besag and Diggle [6] for the detection of spatial association between two sets of continuous quadrat counts over a simple region.

3. CONCLUSION

In this review we have tried to outline different approaches to the problem of testing association between spatial variables. Further work on the modification of existing measures, in particular with the aim of taking into account the spatial scale of the processes would be of interest and of practical use. Studies comparing the power of these methods would give useful insight into this complex problem.

REFERENCES

[1] Agterberg, F.P. (1984), "Trend surface analysis", in *Spatial Statistics and Models*, edited by E.L. Gaile and C.J. Willmott. Dordrecht, Reidel.

[2] Arora, S.S. and M. Brown (1977), "Alternative approaches to spatial autocorrelation : an improvement over current practice". *International Regional Science Review*, 2(1).

[3] Bartlett, M.S. (1935), "Some aspects of time-correlation problem in regard to test of significance". *Journal of the Royal Statistical Society*, 98, 536-543.

[4] Bennett, R.J. (1979), *Spatial Time Series*. London, Pion.

[5] Besag, J.E. (1975), "Statistical analysis of non-lattice data". *The Statistician*, 24(3), 179-195.

[6] Besag, J.E. and P.J. Diggle (1977), "Simple Monte Carlo tests for spatial pattern". *Applied Statistics*, 26(3), 327-333.

[7] Bivand, R.S. (1980), "A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations". *Quaestiones Geographicae*, 6, 5-10.

[8] Bivand, R.S. (1984), "Regression modeling with spatial dependence : an application of some class selection and estimation methods". *Geographical Analysis*, 16(1), 25-37.

[9] Bodson, P. and D. Peeters (1975), "Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. An application to a Belgian labour-demand function". *Environment and Planning A*, 7, 456-472.

[10] Brandsma, A.S. and R.H. Ketellapper (1979), "Further evidence on alternative procedures for testing of spatial autocorrelation among regression disturbances", in *Exploratory and Explanatory Statistical Analysis of Spatial Data*, edited by Bartels and Ketellapper.

[11] Burridge, P. (1981), "Testing for a common factor in a spatial autogresssion model". *Environment and Planning A*, 13, 795-800.

[12] Cliff, A.D. and J.K. Ord (1981), *Spatial Processes. Models and Applications*. London, Pion.

[13] Clifford, P. and S. Richardson (1985), "Testing the association between two spatial processes". *Statistics and Decisions*, Suppl. issue, 2, 155-160.

[14] Cook, D.G. and S.J. Pocock (1983), "Multiple regression in geographic mortality studies with allowance for spartially correlated errors". *Biometrics*, 39, 361-371.

[15] Cox, D.R. (1981), "Statistical analysis of time series : some recent developments". *Scandinavian Journal of Statistics*, 8, 93-115.

[16] Doll, R. (1980), "The epidemiology of cancer". *Cancer*, 45, 2475-2485.

[17] Doreian, P. (1981), "Estimating linear models with spatially distributed data", in *Sociological Methodology*, edited by Leinhardt and Samuel. San Francisco, Jossey-Bass Publishers, 359-388.

[18] Geweke, J. (1981), "A comparison of tests of the independence of two covariance-stationary time series". *Journal of the American Statistical Association*, 76(374), 363-373.

[19] Glick, B.J. (1982), "A spatial rank order correlation measure". *Geographical Analysis*, 14(2), 177-181.

[20] Griffith, D.A. (1980), "Towards a theory of spatial statistics". *Geographical Analysis*, 12(4), 325-339.

[21] Haining, R. (1985), "Income diffusion and regional economic structure". Presented to the NATO advanced studies institute, Hanstholm, Denmark.

[22] Haining, R. (1985), "Trend surface models with global and local scales of variation". Presented at a conference on Applied Spatial Statistics, Cambridge College of Arts and Technology, April 18, 1985.

[23] Haugh, L.D. (1976), "Checking the independence of two covariance-stationary time series : a univariate residual crosscorrelation approach". *Journal of the American Statistical Association*, 71, 378-385.

[24] Hepple, L.W. (1984), "Bayesian analysis of the linear model with spatial dependence", in *Spatial Statistics and Models*, edited by Gaile and Reidel. Dordrecht, Reidel.

[25] Hubert, L.J. and R.G. Golledge (1982), "Measuring association between spatially defined variables : Tjøstheim's index and some extensions". *Geographical Analysis*, 14(3), 273-278.

[26] Johnston, J. (1972), *Econometric Methods*, 2nd edition. New York, McGraw Hill.

[27] Lebart, L. (1969), "Analyse statistique de la contiguité". *Publ. Inst. Stat. Univ. Paris*, 18, 81-112.

[28] Mardia, K.V. and R.J. Marshall (1984), "Maximum likelihood estimation of models for residual covariance in spatial regression". *Biometrika*, 71(1), 135-146.

[29] Martin, R.L. (1974), "On spatial dependence, bias, and the use of first spatial differences in regression analysis". *Area*, 6, 185-194.

[30] Ord, K. (1975), "Estimation methods for models of spatial interaction". *Journal of the American Statistical Association*, 70, 120-125.

[31] Pierce, D.A. (1977), "Relationships and the lack thereof between economic time series, with special reference to money and interest rates". *Journal of the American Statistical Association*, 72(357), 11-26.

[32] Richardson, S. and D. Hémon (1981), "On the variance of the same correlation between two independent lattice processes". *Journal of Applied Probability*, 18, 943-948.

[33] Richardson, S. and D. Hémon (1982), "Autocorrélation spatiale : ses conséquences sur la corrélation empirique de deux processus spatiaux". *Revue de Statistique Appliquée*, 30(1), 41-51.

[34] Ripley, B.D. (1981), *Spatial Statistics*. New York, Wiley.

[35] Streitberg, B. (1979), "Multivariate models of dependent spatial data", in *Exploratory and Explanatory Statistical Analysis of Spatial Data*, edited by Bartels and Ketellapper.

[36] Tjøstheim, D. (1978), "A measure of association for spatial variables". *Biometrika*, 65(1), 109-114.

[37] Upton, G. and B. Fingleton (1983), *Spatial Data Analysis by Example*. New York, Wiley.

ASSESSING THE SIGNIFICANCE OF THE CORRELATION BETWEEN TWO SPATIAL PROCESSES

BIOMETRICS 45, 123–134
March 1989

Assessing the Significance of the Correlation Between Two Spatial Processes

Peter Clifford,[1] Sylvia Richardson,[2] and Denis Hémon[3]

[1] Mathematical Institute, University of Oxford, 24–29 St Giles, Oxford OX1 3LB, England

[2] INSERM U170, 16 Avenue Paul Vaillant Couturier, 94807 Villejuif Cedex, France, and Laboratoire de Statistiques Médicales, Université de Paris V, 45, rue des Saints Pères, 75006 Paris, France

[3] INSERM U170, 16 Avenue Paul Vaillant Couturier, 94807 Villejuif Cedex, France

SUMMARY

Modified tests of association based on the correlation coefficient or the covariance between two spatially autocorrelated processes are presented. These tests can be used both for lattice and nonlattice data. They are based on the evaluation of an effective sample size that takes into account the spatial structure.

For positively autocorrelated processes, the effective sample size is reduced. A method for evaluating this reduction via an approximation of the variance of the correlation coefficient is developed. The performance of the tests is assessed by Monte Carlo simulations. The method is illustrated by examples from geographical epidemiology.

1. Introduction

The calculation of correlation coefficients for variables that are observed at a variety of different spatial locations has suggested intriguing hypotheses about the relationship between environmental factors and disease pathologies (Armstrong and Doll, 1975; Hoover and Fraumeni, 1975). It is well known that the null distribution of r, the product moment correlation coefficient, is influenced by spatial and temporal autocorrelation (Student, 1914; Bartlett, 1935; Richardson and Hémon, 1981).

In a preliminary report, Clifford and Richardson (1985) have suggested a method of approximating the critical values of r. Their method depends on an estimate of the variance of r. In this paper we investigate a general method of obtaining such an estimate and show that the test which results is related to a test based on the standardised covariance between the processes. Theoretical properties of the distribution of r are reviewed in Section 2. In Section 3 we report on an empirical simulation study of the performance of these procedures and in Section 4 we apply our methods to test the association between cigarette consumption, industrial risk factors, and deaths from lung cancer for French départements.

Some authors have devised measures of association that involve transformation of the data before standard techniques are applied. Student (1914) advocated a form of what would now be called prewhitening, and more recent work, such as that of Haugh (1976) with time series and Davies and Jowett (1958) can be seen as developments of this local approach, in that filters are applied to the data in order to reduce temporal or spatial

---

Requests for reprints should be addressed to the second author.

Key words: Correlation coefficient; Geographical epidemiology; Monte Carlo simulations; Significance tests; Spatial processes.

autocorrelation. However, the application of this approach to irregularly spaced data presents a number of problems that have not been thoroughly explored and information is lost by filtering techniques.

The problem of testing the association between autocorrelated variables can also be tackled by regression techniques (Ord, 1975) when a relationship between a dependent variable and a set of independent variables is postulated. Cook and Pocock (1983), in their study of the association between water hardness and heart disease, considered a regression model with stationary autocorrelated errors, the autocorrelation decreasing exponentially with distance. Mardia and Marshall (1984) have demonstrated the asymptotic properties of this approach. In general, these techniques involve a substantial amount of computing time and depend on an explicit parametric model for the autocovariance.

A nonparametric index of association has been proposed by Tjøstheim (1978) and generalised by Hubert and Golledge (1982). However, they assess the significance of the index with reference to a randomisation distribution that involves the spatial redistribution of one of the variables, the other staying fixed. This full set of permutations clearly does not preserve the autocorrelation of the permuted variable and hence the significance of the index is not correctly assessed.

2. Modified Tests of Association

2.1 *Basic Properties of r*

We are interested in data sets which consist of a set A of N locations numbered from 1 to N and a set of pairs of observations (X_α, Y_α), $\alpha \in A$, where each pair is indexed by its location. The correlation coefficient is then given by

$$r = \frac{s_{XY}}{s_X s_Y}, \tag{2.1}$$

where $s_{XY} = N^{-1} \sum_A (X_\alpha - \bar{X})(Y_\alpha - \bar{Y})$ is the sample covariance, $s_X^2 = N^{-1} \sum_A (X_\alpha - \bar{X})^2$, $s_Y^2 = N^{-1} \sum_A (Y_\alpha - \bar{Y})^2$, and where $\bar{X} = N^{-1} \sum_A X_\alpha$ and $\bar{Y} = N^{-1} \sum_A Y_\alpha$. If, conditional on X, the elements of Y are normal i.i.d. variables or conditional on Y, the elements of X are normal i.i.d., then r has the standard null distribution with p.d.f.

$$f_N(r) = \frac{(1 - r^2)^{(N-4)/2}}{B[\frac{1}{2}, \frac{1}{2}(N - 2)]}, \quad r \leqslant 1, \tag{2.2}$$

where B is the beta function.

Critical values of r are usually obtained from t-tables since $(N - 2)^{1/2} r/(1 - r^2)^{1/2}$ has a t-distribution with $N - 2$ degrees of freedom under these assumptions. This is also the t-statistic that is calculated in testing the significance of the linear regression either of Y on X or of X on Y.

2.2 *The Standardised Covariance*

For independent samples the correlation coefficient can be thought of as a standardised covariance. If the elements of Y are i.i.d., then conditional on $X = x$ the sample covariance s_{XY} has mean zero and variance $N^{-2} \sum_A (x_\alpha - \bar{x})^2 \sigma_Y^2$ where σ_Y^2 is the variance of the elements of Y. To standardise s_{XY}, the unknown quantity σ_Y^2 is replaced by an unbiased estimate $\sum_A (Y_\alpha - \bar{Y})^2/(N - 1)$ and s_{XY} is divided by the resulting estimate of its standard deviation. This leads to the expression $(N - 1)^{1/2} r$, whose significance would be approximately assessed with reference to tables of the normal distribution, relying on a central limit theorem to justify the approximation. Thus, a test based on the standardised covariance is equivalent to a test based on r.

In the general case, we suppose now that **X** and **Y** are independent but that both **X** and **Y** are multivariate normal vectors with constant means and variance–covariance matrices $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Y$, respectively. The conditional variance s_{XY} is given by

$$N^{-2} \sum_{\alpha,\beta} (X_\alpha - \bar{X})(X_\beta - \bar{X})\text{cov}(Y_\alpha, Y_\beta), \tag{2.3}$$

which is equal to $N^{-2} \sum_{\alpha,\beta} (X_\alpha - \bar{X})(X_\beta - \bar{X})\ \text{cov}(Y_\alpha - \bar{Y}, Y_\beta - \bar{Y})$ since $\sum_\alpha (X_\alpha - \bar{X}) = 0$. Replacing $\text{cov}(Y_\alpha - \bar{Y}, Y_\beta - \bar{Y})$ by the unbiased estimate $(Y_\alpha - \bar{Y})(Y_\beta - \bar{Y})$ gives the expression $s_X^2 s_Y^2$, as a trivial estimate of the conditional variance of s_{XY}. It is clear that no progress can be made until some plausible restrictive structure is imposed on $\mathbf{\Sigma}_Y$.

2.3 *Structure for* $\mathbf{\Sigma}_X$ *and* $\mathbf{\Sigma}_Y$

We will assume that the set of locations A is a subset of some larger set Ω. For stationary processes Ω can, in principle, be arbitrarily large. For real spatial data, Ω is finite, perhaps equal to A itself. We assume that the set of all ordered pairs of elements of Ω can be divided into strata $S_0, S_1, S_2, \ldots$, so that the covariances within strata are constant, i.e., $\text{cov}(X_\alpha, X_\beta) = C_X(k)$ and $\text{cov}(Y_\alpha, Y_\beta) = C_Y(k)$, $(\alpha, \beta) \in S_k$, $k = 0, 1, \ldots$. Of course, if $(\alpha, \beta) \in S_k$, then $(\beta, \alpha) \in S_k$ for consistency. For stationary processes the stratification is indexed by directional lags. For isotropic processes the number of strata is reduced in the lattice case and when the data are irregularly spaced the strata can be indexed by a discrete distance function. The general formulation is flexible enough to permit spatially inhomogeneous variances and other aspects of nonstationarity. With this structure (2.3) becomes

$$N^{-2} \sum_k N_k \left[\frac{\sum_{A_k} (X_\alpha - \bar{X})(X_\beta - \bar{X})}{N_k} \right] C_Y(k), \tag{2.4}$$

where N_k is the cardinality of A_k, $A_k = (A \times A) \cap S_k$, and the summation over A_k is for all ordered pairs $(\alpha, \beta) \in A_k$.

For stationary processes, Clifford and Richardson (1985) have suggested using

$$\hat{C}_Y(k) = \sum_{A_k} (Y_\alpha - \bar{Y})(Y_\beta - \bar{Y})/N_k \tag{2.5}$$

as an estimate of $C_Y(k)$ for values of k corresponding to small spatial lags and shrinking $\hat{C}_Y(k)$ to zero otherwise. This was partially for computational convenience. Here we propose to consider the inclusion of all lags. The resulting estimate of the conditional variance of s_{XY} is therefore

$$N^{-2} \sum_k N_k \hat{C}_X(k) \hat{C}_Y(k). \tag{2.6}$$

It does not rely on an arbitrary notion of what constitutes a small lag, it is invariant to shifts in the mean, and it is unbiased for periodic processes. It has the additional advantage of being symmetric in X and Y, so that it is also the estimate of the conditional variance of s_{XY} given Y. Note that when $S_0 = A \times A$, (2.6) reduces to $N^{-1} s_X^2 s_Y^2$.

2.4 *The Standardised Covariance and the Modified t-Test*

Using the estimate (2.6), the standardised covariance, W, becomes

$$W = N s_{XY} \left[\sum_k N_k \hat{C}_X(k) \hat{C}_Y(k) \right]^{-1/2} \tag{2.7}$$

If we consider the correlation coefficient, in Appendix 1 it is shown that to the first order,

$$\sigma_r^2 \approx \frac{\text{var}(s_{XY})}{\text{E}(s_X^2)\text{E}(s_Y^2)}. \tag{2.8}$$

We therefore take as our estimate of σ_r^2

$$\hat{\sigma}_r^2 = \frac{\sum N_k \hat{C}_X(k)\hat{C}_Y(k)}{N^2 s_X^2 s_Y^2}. \tag{2.9}$$

We consider a modified t-test for the correlation coefficient by approximating the critical values of r by percentage points of the p.d.f. $f_{\hat{M}}(r)$, where $\hat{M} = 1 + \hat{\sigma}_r^{-2}$ and f is defined by (2.2). The quantity $(\hat{M})$ M can be thought of as an (estimated) "effective sample size" that takes into account the spatial autocorrelation in the variables **X** and **Y**. Note that when $S_0 = A \times A$, $\hat{M} = N + 1$. For positively autocorrelated processes, the estimated effective sample size is typically less than N. If one of the processes has negative autocorrelation it is, in principle, possible that the effective sample size will be larger than N.

Comparing (2.7) and (2.9), we see that

$$W = (\hat{M} - 1)^{1/2} r, \tag{2.10}$$

where $\hat{M} = 1 + \hat{\sigma}_r^{-2}$.

In Section 3 we consider the performance of the test obtained by assuming that W has a N(0, 1) distribution under the null hypothesis and compare this test with that obtained by assuming that r has p.d.f. $f_{\hat{M}}(r)$, that is, using a t-statistic with $\hat{M} - 2$ in place of $N - 2$.

3. Monte Carlo Simulations

The performance of the modified t-test or the test based on the standardised covariance has been assessed by Monte Carlo simulation in both a lattice and a nonlattice case.

3.1 *The Lattice Case*

Method We simulated stationary first-order isotropic simultaneous autoregressive processes defined by

$$X_{s,t} = a(X_{s-1,t} + X_{s+1,t} + X_{s,t-1} + X_{s,t+1}) + \varepsilon_{s,t}, \tag{3.1}$$

where $\{\varepsilon_{s,t}\}$ is a sequence of i.i.d. N(0, 1), $0 < |a| < \frac{1}{4}$. This class of processes has been widely discussed in the modelling of spatial patterns (Whittle, 1954; Besag, 1974; Cliff and Ord, 1975) and is relatively easy to simulate (see Appendix 2).

The simulation was performed on a DPS7 C.I.I. Honeywell Bull. A spectral decomposition similar to that given by Besag in the discussion of Bartlett's (1978) paper for autonormal processes was used to generate processes on a 26 × 26 lattice with zero on the boundary and then restricted to middle 12 × 12, 16 × 16, and 20 × 20 squares. This restriction was sufficient to render negligible the influence of the boundary condition.

For each simulation, a 26 × 26 field of i.i.d. N(0, 1) variables was first generated using a polar algorithm. Values of a equal to 0, .0945, .165, .2099, .2364 were chosen to give values of the nearest neighbour autocorrelation, $\rho_X(1)$, equal to 0, .2, .4, .6, and .8, respectively.

For each value of a, 500 pairs (X_α, Y_α), $\alpha \in A$, were generated. Taking advantage of the lattice structure, each process **X** can also be rotated by 90° or reflected with respect to the diagonals, leading to eight copies (four rotations and/or two reflections) of the original process. These copies have the same set of estimated autocovariances, $\{\hat{C}_X(k)\}$, which therefore is calculated only once. Thus, 4,000 trials, dependent in groups of eight, were obtained for each value of $\rho_X(1)$ and $\rho_Y(1)$.

The proposed statistics are denoted by W' and $t_{\hat{M}-2}$, where $\hat{M}$ is the estimated effective sample size. $\hat{M} = 1 + \hat{\sigma}_r^{-2}$. If expression (2.6) for the estimated variance of s_{XY} gave an inadmissible negative estimate, it was replaced by the product $N^{-1}s_X^2 s_Y^2$, which is the estimate in the case of no autocorrelation. The standard t-statistic for r based on N observations is denoted by t_{N-2}. For the three statistics, W', $t_{\hat{M}-2}$, and t_{N-2}, a 5% nominal rejection level was chosen. For testing $t_{\hat{M}-2}$, the integer part of $\hat{M}$ was taken. The empirical variance of the rejection levels was estimated by first averaging the rejection indicator function over the eight dependent pairs (X_α, Y_α) (where X_α is obtained by rotations/reflections) and then calculating the empirical variance of this average over the 500 independent simulations.

Results Figure 1 illustrates the poor performance of the standard t-test in the presence of positive autocorrelation. The observed Type I error rates are represented for the three lattice sizes together with their 95% confidence interval. Note that the observed Type I error rates are significantly larger than 5%, even for low values of $\rho_X(1)$ and $\rho_Y(1)$.

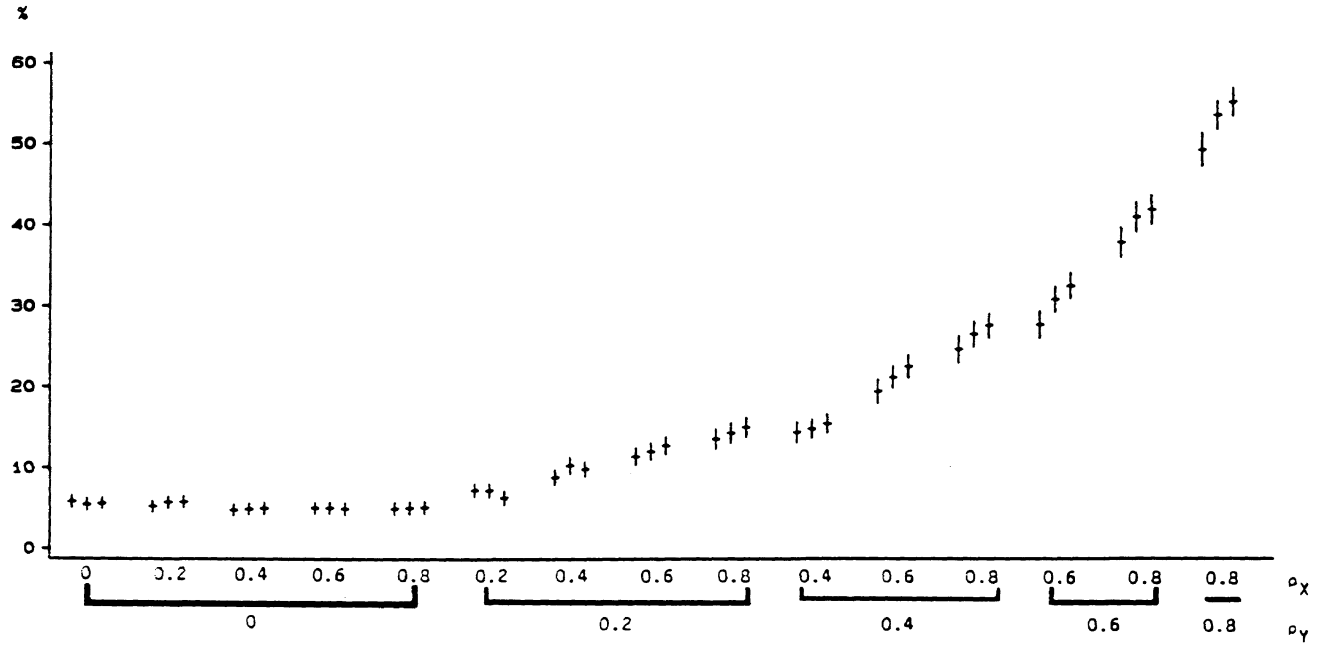


Figure 1. Standard test of the correlation coefficient: 95% confidence intervals for the proportion of Type I errors for a 5% nominal test in the case of two mutually independent simultaneous autoregressive processes on a lattice. For each value of the nearest neighbour autocorrelation for **X** or **Y**, $\rho_X(1)$ or $\rho_Y(1)$, the confidence intervals are plotted for three lattice sizes: 12 × 12, 16 × 16, 20 × 20.

When both **X** and **Y** are highly autocorrelated, observed Type I error rates vary between 25% and 55%, thus clearly showing that the testing procedure needs to be modified.

For the two proposed tests, the observed Type I error rates ranged from 4.2% to 5.85% for the W' test (Figure 2) and from 4.1% to 5.9% for the $t_{\hat{M}-2}$ test, and are thus close to the nominal 5% level. Note that when one of the processes has no autocorrelation, W' and $t_{\hat{M}-2}$ perform as well as the standard t-test. The confidence interval did not include the 5% nominal level in only two cases [20 × 20, $\rho_X(1) = .2$, $\rho_Y(1) = .2$ and 16 × 16, $\rho_X(1) = .6$, $\rho_Y(1) = .8$]. Inadmissible estimates were rare: they did not occur more than twice per lattice size in all the trials except in the case $\rho_X(1) = 0$, $\rho_Y(1) = .8$, where there were three negative estimates for the 12 × 12 lattice.

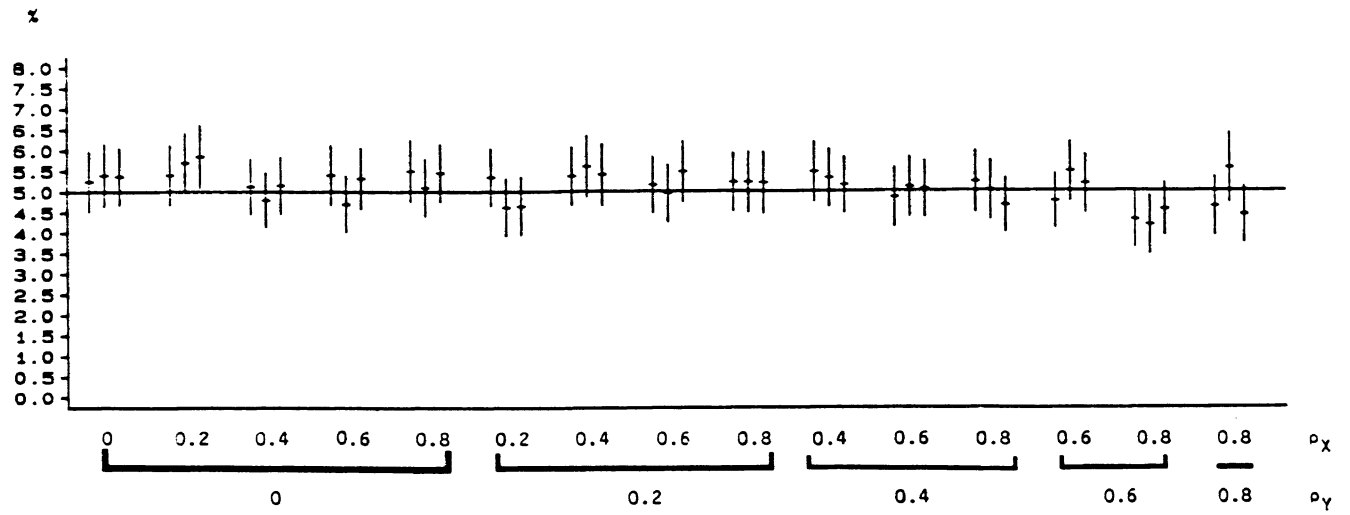


Figure 2. Test based on the standardised covariance W: 95% confidence intervals for the proportion of Type I errors for a 5% nominal test in the case of two mutually independent simultaneous autoregressive processes on a lattice.

A comparison of the empirical variance of r (averaged over the 4,000 trials) and of the average v_r of the estimated variance $\hat{\sigma}_r^2$ given by (2.9), was also made. For small to moderate autocorrelations, there is practically no difference between the empirical variance of r and v_r. As the autocorrelation increases, v_r is consistently too low. That (2.6) is negatively biased as an estimate of the variance of s_{XY} can be easily seen in the case $\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y$.

Quantile plots of the distribution of the W statistic show that it has short tails compared with the normal distribution in the extreme case of $\rho_X(1) = \rho_Y(1) = .8$. The departure from normality is confirmed by a Kolmogorov–Smirnov test, which is significant at the 5% level but not at the 1% level. The tendency for short tails also occurs in other cases of $\rho_X(1)$ and $\rho_Y(1)$, but is more marked for higher autocorrelation. Nevertheless, the departure from normality of W and the order of the negative bias in v_r do not seem to be quantitatively altering the levels of the W and t_{N-2} tests even in the strongly autocorrelated cases.

3.2 *Nonlattice Case*

Choice of the model The structure of the network and the type of spatial dependence were both chosen in order to be similar to that of examples in geographical epidemiology which will be discussed in the next section.

The coordinates of the points of the network were identified with the geographical locations of the administrative centers ("préfectures") of French départements. The variables considered are the mortality rate for lung cancer (LC) for men, the cigarette sales (CS) per inhabitant, and the percentage of metal workers (MW) and the percentage of textile workers (TW) with respect to the male working population in each département.

The spatial structure of these variables was investigated by means of a variogram. In this analysis, $N = 82$ locations were retained after grouping the départements around Paris into one area. The distances between the centres of départements were partitioned into 15 classes of 50-km intervals each. This gives 15 strata $S_1, \ldots, S_{15}$: the stratum $S_0 = \{(\alpha, \alpha)\ \alpha \in A\}$. The number of strata chosen should take into account a balance between the sampling fluctuations of the estimated autocovariances when the number of strata is large and the introduction of bias when the number of strata is small. In making this judgement it is helpful to compute the observed semivariogram for several cases. For

the present data set we have found that the results were not sensitive to the choice of the number of strata.

The observed variogram of LC, i.e., the plot of

$$N_k^{-1} \sum_{(\alpha,\beta)\in S_k} (X_\alpha - X_\beta)^2, \qquad k = 1, \ldots, 15,$$

against the average distance, d_k, for départements in S_k, is shown in Figure 3. The variograms of CC and MW were similar and exhibited also an upward trend of fairly linear shape with increasing distance. Hence, a disc model for the covariance matrix (Ripley, 1981, p. 55) seemed appropriate and we chose it to simulate a spatially dependent process on this irregular network.

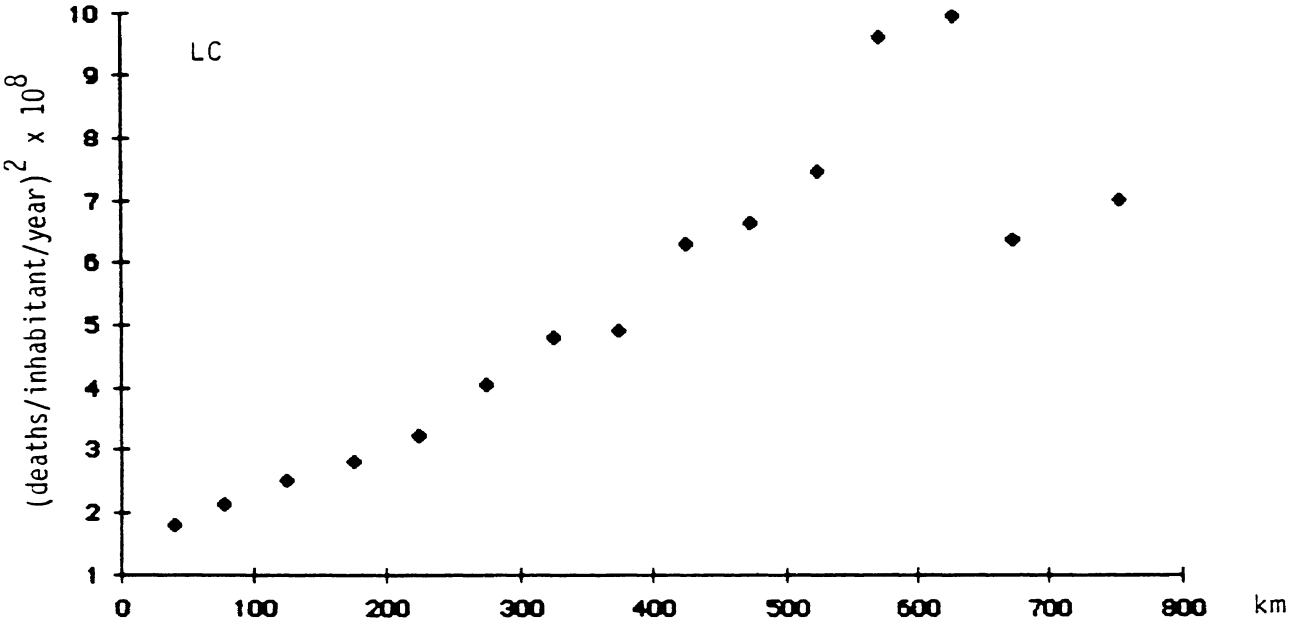


Figure 3. Variogram of the lung cancer mortality (LC). Fifteen classes of distance are considered: the number of ordered pairs in each class is (82; 400; 582; 674; 764; 822; 812; 726; 630; 476; 304; 172; 94; 58; 40). The abscissa corresponds to the average distance in kilometres within each class.

Method The average distance between départements in the first stratum S_1 is approximately 40 km. The parameters of the disc models for **X** and **Y** were chosen to be such that the autocorrelation at distance 40 km is equal to .2, . . . , .9. We also denote these by $\rho_X(1)$ or $\rho_Y(1)$.

For each chosen value of $\rho_X(1)$, the matrix $\boldsymbol{\Sigma}_X$ was triangularised ($\boldsymbol{\Sigma}_X = \mathbf{L}\mathbf{L}^T$) and then a realisation of **X** was obtained by first generating a vector of N i.i.d. N(0, 1) and then multiplying this vector by **L**. In each case 500 pairs of mutually independent processes (**X**, **Y**) were simulated and the statistics W and t_{W-2} calculated for each pair (**X**, **Y**) as in Sections 2.2 and 2.3, where the autovariances $\hat{C}_X(k)$ are defined as

$$\hat{C}_X(k) = N_k^{-1} \sum_{(\alpha,\beta)\in S_k} (X_\alpha - \bar{X})(X_\beta - \bar{X}).$$

The same procedure as in the lattice case was adopted for an inadmissible negative estimate for the variance of s_{YY}. A 5% nominal level was chosen to assess the performance of W, t_{W-2}, and t_{X-2}.

Results Figure 4 demonstrates the increased proportion of Type I errors of the standard t-test in the case of positive autocorrelation of both processes. Figure 5 shows the performance of the W' test, which is indistinguishable from the performance of the $t_{\hat{N}-2}$ test. All observed Type I error rates, even in the most highly autocorrelated case, were close to 5%. No inadmissible negative estimate, $\hat{\sigma}_\varepsilon^2$, arose. Kolmogorov–Smirnov tests performed on the distribution of W' were significant in two cases (.2 × .6 and .9 × .9).

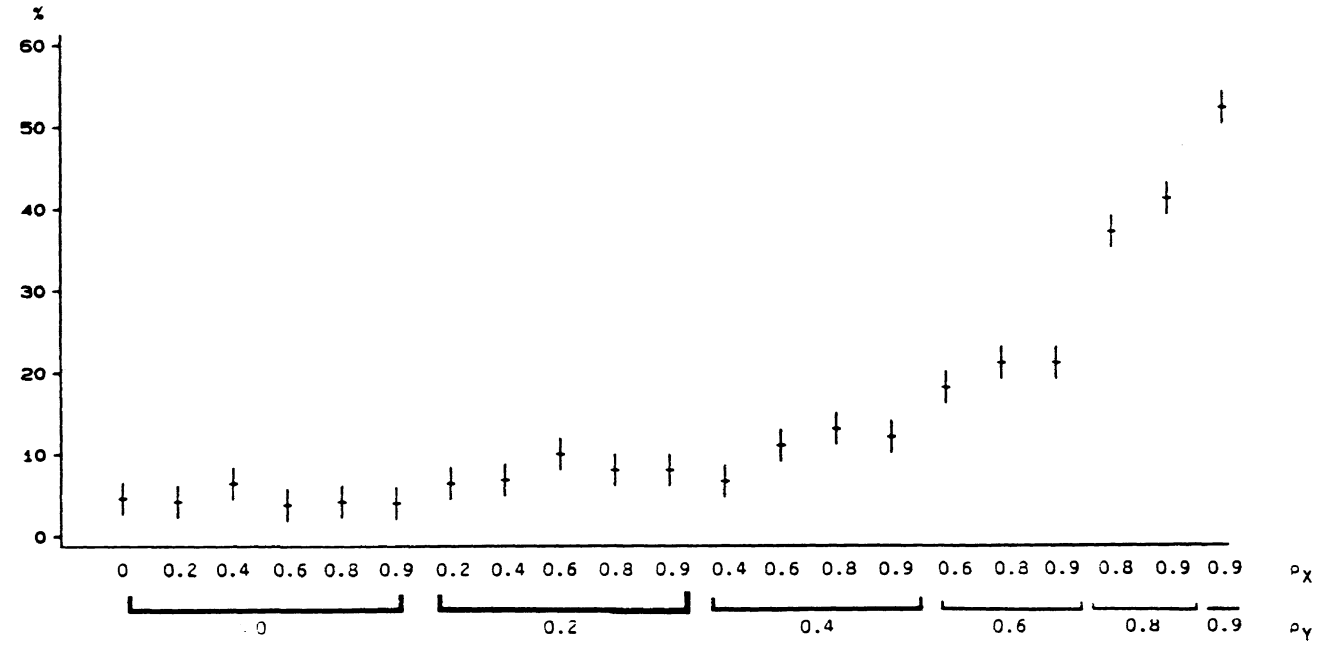


Figure 4. Standard test of the correlation coefficient: 95% confidence intervals for the proportion of Type I errors for a 5% nominal test in the case of two mutually independent processes generated by a disc model on a network of 82 points. (The parameters $\rho_X(1)$ and $\rho_Y(1)$ of the disc models for **X** and **Y** respectively are equal to the autocorrelation at 40 km.)

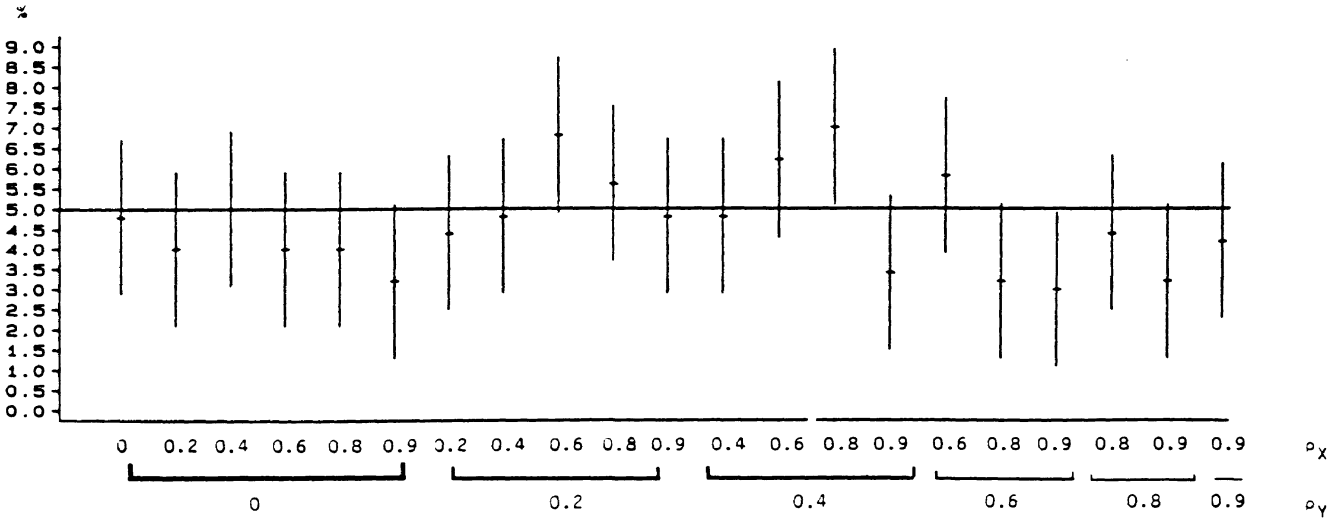


Figure 5. Test based on the standardised covariance W': 95% confidence intervals for the proportion of Type I errors for a 5% nominal test in the case of two mutually independent processes generated by a disc model on a network of 82 points. (The parameters $\rho_X(1)$ and $\rho_Y(1)$ of the disc models for **X** and **Y** respectively are equal to the autocorrelation at 40 km.)

4. Examples

Our examples concern the relationship between lung cancer, smoking, and industrial factors. We calculate W and $t_{\hat{N}-2}$.

4.1 *The Data*

For 82 départements we considered male lung cancer mortality rate (LC) standardised over the age 35–74 and over a 2-year period, 1968–1969; cigarette sales per inhabitant (CS) in 1953 (a 15-year time lag was chosen to account for the delay between exposure and the onset of the pathology); and demographic data on the percentage of employed males in the metal industry (MW) and the textile industry (TW) recorded by census in 1962.

4.2 *Results*

Using the standard test based on r, there is a highly significant positive association both between LC and CS, and between LC and MW. The association between LC and TW, on the other hand, is less strong but still significant at the 1% level (Table 1). The W and $t_{\hat{N}-2}$ statistics for these three examples are shown in Table 1. One can see a substantial reduction of the degrees of freedom when the autocorrelation is taken into account.

Table 1
Comparison of the significance levels for tests of the association between lung cancer mortality rates and several risk factors given by standard test, W, and $t_{\hat{N}-2}$ tests

	r	t_{N-2}[a]	W	$\hat{M}$	$t_{\hat{N}-2}$
Cigarette sales per inhabitant (1953) (CC)	.76	10.48 $P \approx 10^{-21}$	2.94 $P = .0032$	15	4.22 $P = .001$
% male workers in metal industry (1962) (MW)	.63	7.16 $P \approx 10^{-11}$	2.48 $P = .0136$	16	3.00 $P = .01$
% male workers in textile industry (1962) (TW)	.28	2.57 $P = .01$	1.51 $P = .13$	30	1.52 $P = .15$

[a] Standard test (t-transformation with 80 d.f.).

For CS and MW the CS effective sample size is about 20% of the original sample size. Consequently, the significance levels are reduced but even after this "adjustment" these two factors are statistically significantly associated with lung cancer. For TW the significance disappears after adjustment.

These results can be considered in good agreement with current knowledge concerning life style and occupational risk factors for lung cancer (Schottenfeld and Fraumeni, 1982).

5. Discussion

Our study of the correlation coefficient has been motivated by its widespread use in epidemiology, where relatively small data sets of up to 100 irregularly spaced points are encountered. In this context, positively autocorrelated **X** and **Y** are most commonly observed.

We have confirmed both theoretically and empirically that the uncritical use of the correlation coefficient for testing association between positively autocorrelated processes leads to an inflated proportion of Type I errors, and have shown theoretically that the magnitude of this inflation is consistent with a reduction in the effective sample size. We have then investigated how the correlation coefficient behaves when it is adjusted to take account of this effect. We have done this in two related ways, the W and $t_{\hat{N}-2}$ tests, and

have shown that when adjustment is made the Type I error rate is much closer to the nominal level. These methods do not require the identification of a particular parametric model for the type of spatial autocorrelation and they cope equally well with regularly and irregularly spaced points. Isotropy need not be assumed. The strata can be defined by orientation as well as distance in the spirit of Granger (1969) because equations (2.7) and (2.9) can be easily adapted to any partition of the set of pairs of locations. Only simple calculations of autocorrelations, which can be done on small computers, are involved.

The detection of association between spatial data sets is not a simple problem. In this paper we have investigated one particular approach. As well as having reservations about the assumption of stationarity, in a detailed statistical analysis it should be recognised that association can exist simultaneously at a number of different geographical scales. The correlation coefficient is a single omnibus statistic that averages the scale-dependent association. Thus, for example, it is possible that negative association at small scales is swamped by positive association at large scales. We have not attempted to explore the ways in which these scale-dependent associations can be separated out. This is an important area of research.

Acknowledgements

The authors wish to thank Annie Mollié and Nicole Le Moual for computing assistance. This work was supported by Euratom Contract BI6-0126-F.

Résumé

Des tests d'association modifiés, portant sur le coefficient de corrélation ou la covariance entre deux processus spatiaux autocorrelés, sont proposés. Ces tests peuvent être utilisés à la fois pour des données observées sur un lattice ou sur un domaine irrégulier. Il sont fondés sur l'évaluation d'un nombre d'observation corrigé qui prend en compte la structure spatiale de chaque processus.

Le nombre d'observations corrigé est inférieur à la taille de l'échantillon quand l'autocorrélation de chaque processus est positive. Une méthode pour évaluer cette réduction par l'intermédiaire d'une approximation de la variance du coefficient de corrélation est développée. La performance des tests est évaluée par des simulations de Monte-Carlo. La méthode est illustrée par des exemples d'épidémiologie géographique.

References

Armstrong, B. and Doll, R. (1975). Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *International Journal of Cancer* **15**, 617–631.

Bartlett, M. S. (1935). Some aspects of the time-correlation problem in regard to tests of significance. *Journal of the Royal Statistical Society* **98**, 536–543.

Bartlett, M. S. (1978). Nearest neighbour models in the analysis of field experiments. *Journal of the Royal Statistical Society, Series B* **40**, 147–174.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

Cliff, A. D. and Ord, J. K. (1975). Model building and the analysis of spatial pattern in human geography. *Journal of the Royal Statistical Society, Series B* **37**, 297–348.

Clifford, P. and Richardson, S. (1985). Testing the association between two spatial processes. *Statistics and Decisions*, Supp. issue **2**, 155–160.

Cook, D. G. and Pocock, S. J. (1983). Multiple regression in geographic mortality studies with allowance for spatially correlated errors. *Biometrics* **39**, 361–371.

Davies, H. M. and Jowett, G. H. (1958). The fitting of Markoff serial variation curves. *Journal of the Royal Statistical Society, Series B* **20**, 120–142.

Granger, C. W. J. (1969). Spatial data and time series analysis. In *London Papers in Regional Science* 1. *Studies in Regional Science*, A. J. Scott (ed.), 1–24. London: Pion.

Haugh, L. D. (1976). Checking the independence of two covariance stationary time series: A univariate residual cross-correlation approach. *Journal of the American Statistical Association* **71**, 378–385.

Hoover, R. and Fraumeni, J. F. (1975). Cancer mortality in U.S. counties with chemical industries. *Environmental Research* **9**, 196–207.
Hubert, L. J. and Golledge, R. G. (1982). Measuring association between spatially defined variables: Tjostheim's index and some extensions. *Geographical Analysis* **14**, 273–278.
Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–146.
Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70**, 120–126.
Richardson, S. and Hémon, D. (1981). On the variance of the sample correlation between two independent lattice processes. *Journal of Applied Probability* **18**, 943–948.
Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.
Schottenfeld, D. and Fraumeni, J. F. (1982). *Cancer Epidemiology and Prevention*. Philadelphia: W. B. Saunders.
Student (W. S. Gosset) (1914). The elimination of spurious correlation due to position in time or space. *Biometrika* **10**, 179–181.
Tjøstheim, D. (1978). A measure of association for spatial variables. *Biometrika* **65**, 109–114.
Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434–449.

Received June 1987; *revised June* 1988.

Appendix 1

The Variance of r

Let $\boldsymbol{\Sigma}_\xi$ and $\boldsymbol{\Sigma}_\eta$ denote the variance–covariance matrices of the vectors with elements $\xi_{ij} = X_{ij} - \bar{X}$ and $\eta_{ij} = Y_{ij} - \bar{Y}$.

If we write

$$r = \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\eta}/(\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}\,\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{1/2}$$

then

$$\mathrm{var}(r) = \mathrm{E}\ \mathrm{tr}\left\{\frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}}{\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}}\frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}}{\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}}\right\} = \mathrm{tr}\left(\mathrm{E}\,\frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}}{\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}}\,\mathrm{E}\,\frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}}{\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}}\right) \tag{A.1}$$

by the independence of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$.

Expanding $\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}$ about its expectation, $k_1 = \mathrm{tr}(\boldsymbol{\Sigma}_\eta)$, we have

$$\frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}}{\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}} = \frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}}{k_1}\left[1 - \frac{\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta} - k_1}{k_1} + \frac{(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta} - k_1)^2}{k_1^2} - \cdots\right]. \tag{A.2}$$

Note that in the case $\boldsymbol{\Sigma}_\eta = \mathbf{I}$ we have $k_1 = N - 1$ so that we would be expanding in inverse powers of $N - 1$. To evaluate the expectation of (A.2) we calculate

$$\mathrm{E}(\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}})^2 = 2\boldsymbol{\Sigma}_\eta^2 + k_1\boldsymbol{\Sigma}_\eta$$

and

$$\mathrm{E}(\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}})^3 = 8\boldsymbol{\Sigma}_\eta^3 + 2k_2\boldsymbol{\Sigma}_\eta + 4k_1\boldsymbol{\Sigma}_\eta^2 + k_1^2\boldsymbol{\Sigma}_\eta,$$

where $k_2 = \mathrm{tr}(\boldsymbol{\Sigma}_\eta^2)$.

If we make the plausible assumption that the maximum eigenvalue of $\boldsymbol{\Sigma}_\eta$ is bounded above by λ_B as $N \to \infty$, then using the inequality $\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2) \leqslant \lambda_B\ \mathrm{tr}(\boldsymbol{\Sigma}_\eta)$, we have the second-order asymptotic expression

$$\mathrm{E}\left(\frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}}{\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}}\right) = \frac{\boldsymbol{\Sigma}_\eta}{k_1} - 2\,\frac{\boldsymbol{\Sigma}_\eta^2}{k_1^2} + 2\,\frac{\boldsymbol{\Sigma}_\eta k_2}{k_1^3}. \tag{A.3}$$

Finally, substituting (A.3) and a similar expression for $\mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}/\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi})$ into (A.1), we have to the first order

$$\sigma_r^2 = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}_\xi)}{k_1 j_1}. \tag{A.4}$$

and to the second order

$$\sigma_r^2 = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}_\xi)}{k_1 j_1} - \frac{2\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2 \boldsymbol{\Sigma}_\xi)}{k_1^2 j_1} + \frac{2\mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}_\xi)}{k_1 j_1}\left(\frac{k_2}{k_1^2} + \frac{j_2}{j_1^2}\right) - \frac{2\mathrm{tr}(\boldsymbol{\Sigma}_\eta \boldsymbol{\Sigma}_\xi^2)}{j_1^2 k_1}. \tag{A.5}$$

where $j_n = \mathrm{tr}(\boldsymbol{\Sigma}_\xi^n)$.

In the case $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_\xi$ the second-order term in (A.5) becomes

$$-4\left\{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_\eta^3)}{k_1^3} - \frac{\mathrm{tr}(\boldsymbol{\Sigma}_\eta^2)\mathrm{tr}(\boldsymbol{\Sigma}_\xi^2)}{k_1^4}\right\} = -4\left\{\sum_{i=1}^{N-1} p_i^3 - \left(\sum_{i=1}^{N-1} p_i^2\right)^2\right\}$$

$$= -4\left\{\sum_{i=1}^{N-1}\left(p_i - \sum_{i=1}^{N-1} p_i^2\right)^2 p_i\right\} \leq 0,$$

where $p_i = \mu_i / \sum_{i=1}^{N-1} \mu_i$, $\{\mu_i\}_{i=1}^{N-1}$ being the eigenvalues of $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\xi$ (matrices of rank $\leq N-1$), i.e., the first-order approximation is no larger than the second-order approximation.

Note that in the case where (i) $\boldsymbol{\Sigma}_\xi$ and $\boldsymbol{\Sigma}_\eta$ commute, (ii) $\boldsymbol{\Sigma}_\eta$ is proportional to an idempotent matrix, and (iii) the eigenvalues of $\boldsymbol{\Sigma}_\xi$ are zero whenever the associated eigenvalues of $\boldsymbol{\Sigma}_\eta$ are zero, then the first-order term in (A.4) is exactly equal to σ_r^2. This follows from the remark that in this case r has p.d.f. $f_M(r)$ with effective sample size $M = 1 + \mathrm{rank}(\boldsymbol{\Sigma}_\eta)$, where $f_M(r)$ is the standard distribution of r defined in (2.2).

APPENDIX 2

Simulation of Simultaneous Autoregressive Process

For a process given by (3.1) with zero boundary, i.e., $X_{s,0} = X_{0,t} = X_{n+1,t} = X_{s,n+1} = 0$ ($s, t = 0, 1, \ldots, n+1$), the spectral decomposition of $\boldsymbol{\Sigma}_X$ can be deduced from the expression given by Besag [discussion of a paper by Bartlett (1978)]. The process can be simulated by computing $\mathbf{P}^T \mathrm{diag}(\delta_i^{1/2})\mathbf{Z}$, where $\mathbf{Z}$ is a vector of independent N(0, 1) variables and where $\mathbf{P}^T \mathrm{diag}(\delta_i)\mathbf{P}$ is the spectral decomposition of $\boldsymbol{\Sigma}_X$. If we denote the matrix $\{X_{s,t}\}_{s,t=1}^n$ by $\mathbf{X}$ and rearrange the vector of $\mathbf{Z}$ into an $n \times n$ matrix whose elements are Z_{ij} ($i, j = 1, 2, \ldots, n$), then, exploiting the properties of $\mathbf{P}$, we find that $\mathbf{X}$ is simulated by

$$\mathbf{X} = \mathbf{Q}\{Z_{ij}\lambda_{ij}^{-1/2}\}\mathbf{Q},$$

where

$$\lambda_{ij} = 1 - 2a\{\cos[\pi i/(n+1)] + \cos[\pi j/(n+1)]\}$$

and

$$Q_{ij} = [2/(n+1)]^{1/2}\sin[\pi ij/(n+1)], \quad n \text{ even}.$$

The computational cost of a single realisation is therefore of the order of n^3 operations or $N^{3/2}$ since $N = n^2$. In contrast, a single triangulation $\mathbf{L}\mathbf{L}^T$ of $\boldsymbol{\Sigma}_X$ requires an order of N^3 operations and each realisation involves an order of N^2 operations.

TESTING ASSOCIATION BETWEEN SPATIAL PROCESSES

TESTING ASSOCIATION BETWEEN SPATIAL PROCESSES

Sylvia RICHARDSON*, Peter CLIFFORD+

1. Introduction

This paper is concerned with a topic in spatial statistics, that of testing for association between two spatial processes. This question arises frequently in various fields and examples abound in geography and regional sciences (Cliff and Ord, 1981), geology (Malin and Hyde, 1982), sociology (Doreian, 1981)... In the epidemiology of chronic diseases, etiological clues are sometimes sought by studying joint geographic variations of environmental risk factors and disease rates (Doll 1980). The examples discussed in the later part of this paper will be taken from this field.

Throughout the data will consist of a set A of N locations and pairs of variables (X_α, Y_α), $\alpha \in A$, indexed by their location. These variables will be spatially autocorrelated.

In the first part a method for approximating the critical values of the product moment correlation coefficient r_{XY} will be summarised. This method has been described in detail in Clifford, Richardson, Hémon (1989).

In the following section we give some complementary results on the performance of the tests for small domains and on their power. We also discuss the influence of the choice of the partition of the covariance structure. Finally we give some examples, comparing our results with those of Monte Carlo tests and giving pivotal confidence intervals for the regression coefficient.

2. Modified tests of associations

We have devised modified tests of association based either on s_{XY}: the empirical covariance between pairs of observations $(X_\alpha, Y_\alpha))$, $\alpha \in A$, or based on r_{XY} : the corresponding empirical correlation coefficient. We shall use the notation :

$$\bar{X} = N^{-1}(\Sigma X_\alpha), \quad s_{XY} = N^{-1} \Sigma (X_\alpha - \bar{X})(Y_\alpha - \bar{Y}), \; s_X^2 = N^{-1} \Sigma (X_\alpha - \bar{X})^2$$

and similarly for $\bar{Y}$ and s_Y^2.

* INSERM U.170, 16 Avenue Paul Vaillant-Couturier
94807 VILLEJUIF Cedex, France
+ Mathematical Institute, University of Oxford,
24-29 St Giles, Oxford OXI 3LB

Method

Suppose that X and Y are independent but that both X and Y are multivariate normal vectors with constant means and variance-covariance matrices Σ_X and Σ_Y respectively. A stratified structure for Σ_X and Σ_Y is imposed. Pairs in A x A are divided into strata $S_0, S_1, S_2, \ldots$ such that the covariances within strata remain constant,

ie $\quad \text{cov}(X_\alpha, X_\beta) = C_X(k) \quad$ if $(\alpha,\beta) \in S_k$.

An estimate of the conditional variance of s_{XY} is then derived :

$$N^{-2} \sum_k N_k \hat{C}_X(k) \hat{C}_Y(k) \quad , \qquad (1)$$

N_k is the number of pairs in strata S_k and $\hat{C}_X(k)$ (respectively $\hat{C}_Y(k)$) is the estimated autocovariance :

$$\hat{C}_X(k) = \sum_{S_k} (X_\alpha - \bar{X})(X_\beta - \bar{X}) / N_k$$

Thus the estimate takes into account the autocorrelation of both X and Y.

Further it can be shown that, to the first order, the variance of r_{XY}, σ^2_r is :

$$\sigma^2_r = \frac{\text{Var}(s_{XY})}{E(s_X^2)\, E(s_Y^2)} \quad , \qquad (2)$$

which leads to the following estimate :

$$\hat{\sigma}^2_r = \frac{\Sigma N_k \hat{C}_X(k) \hat{C}_Y(k)}{N^2 s_X^2 s_Y^2} \qquad (3)$$

Note that an equivalent expression for $\text{Var}(s_{XY})$ is $N^{-2}\ \text{tr}(\Sigma_\xi \Sigma_\eta)$ where Σ_ξ and Σ_η are the variance-covariance matrices of the centered vectors $X - \bar{X}$ and $Y - \bar{Y}$.

In the classical non autocorrelated case, when either Σ_X or $\Sigma_Y = I$, it can be shown that the approximation given by (2) is exact and that r_{XY} follows a t-distribution with N-2 d.f (t_{N-2}). Further : $N = 1 + (\sigma^2_r)^{-1}$

In general, an estimated effective sample size, $\hat{M}$, is defined by the relationship

$$\hat{M} = 1 + (\hat{\sigma}^2_r)^{-1} \quad ,$$

where $\hat{\sigma}^2_r$ is given by (3). A modified **t**-test : $t_{\hat{M}-2}$ is proposed which rejects the null hypothesis of no association when :

$$| (\hat{M}-2)^{1/2} \; r(1 - r^2)^{-1/2} | > t^{\alpha}_{\hat{M}-2}$$

where $t^{\alpha}_{\hat{M}-2}$ is critical value of the **t**-statistic with $\hat{M}-2$ d.f.

Equivalently a standardised covariance can be used :

$$W = N \, s_{XY} \, (\Sigma \, N_k \, \hat{C}_X(k) \, \hat{C}_Y(k))^{-1/2}$$

and tested as a standard normal relying upon central limit theorems for spatially dependent variables.

Results on the performance of W and $t_{\hat{M}-2}$

The performance of these tests under the null hypothesis of stochastic independence between X and Y was first assessed by Monte Carlo simulations for two models :

a) X and Y were generated on 3 lattice sizes (12x12, 16x16, 20x20) as nearest neighbour isotropic autoregressive gaussian processes, with 1st order autocorrelation $\rho_X(1)$ and $\rho_Y(1)$ ranging from 0.2 to 0.8 ;

b) X and Y were generated on the grid of the administrative centres of the French départements as gaussian variables with a disc model for their autocovariance (Ripley 1981) and arbitrarily defined 1st order autocorrelation ranging from 0.2 to 0.9.

In both cases and for several levels of autocorrelation in X or Y, 500 trials were executed with a nominal rejection level of 5%.

For the two statistics, $t_{\hat{M}-2}$ and W, the type I errors found did not vary in any systematic way with the level of autocorrelation and fluctuated around the nominal 5% level (the confidence interval excluded 5% in only 2 cases out of 45 simulated in (a)). This contrasted with the performance of the standard **t**-test procedure based on N-2 d.o.f. where the type I error increased systematically with the autocorrelation to reach values around 50 % in the highly autocorrelated cases (Clifford, Richardson, Hémon, 1989).

3. Comparison of W and $t_{\hat{M}-2}$ on small lattices

Based on our first results the performance of the statistics $t_{\hat{M}-2}$ and W seemed indistinguishable. The number of sample points was reasonably large and a difference between their respective performances could be better highlighted by studying smaller samples.

Results from further simulations on smaller lattices of sizes 6x6, 8x8, 10x10 are shown in Figure 1. The type I errors of the W statistic exhibit a systematic downward trend with increasing autocorrelation which is not apparent for the modified $t_{\hat{M}-2}$ statistic. Consequently, for small domains, the use of the modified $t_{\hat{M}-2}$ statistic is preferable to that of W.

We also note that the convergence of W to normality is slower as the autocorrelation increases. Figure 2 shows a Q.Q. plot of W, i.e. quantiles of a standard normal distribution against the sample quantiles of W for 500 independent trials, in the case $\rho_X(1) = \rho_Y(1) = 0.8$ and a 12x12 lattice. We note that the distribution of the W statistic has short tails compared to the normal distribution. The departure from normality is confirmed by a Kolmogorov-Smirnov test which is significant at the 5% level but not at the 1% level.

4. Power of the modified tests

The power of the modified tests was assessed under a simple alternative hypothesis of a linear regression between Y and X : $H_1 : Y = aX + W$, $X \sim N(\mu_X, \Sigma_X)$, $W \sim N(\mu_W, \Sigma_W)$ and X and W independent. It is difficult to calculate theoretically the power of the modified tests because their distribution under H_1 is not precisely known. Their power can be assessed by simulations.

Two independent spatially autocorrelated processes X and W were generated on the grid of the administrative centers of French départements as gaussian variables with a disc model for their autocovariance. Without loss of generality $\sigma^2_X = \sigma^2_W$ was chosen and hence the correlation ρ_{XY} between X and Y was only dependent on the parameter a. Five hundred trials were carried out for several levels of autocorrelation in X and W and for the values $\rho_{XY} = 0.2$ and 0.4. The grid contained N = 82 points. Results for higher values of ρ_{XY} are not reported because the power of the tests was very close to 1. In these simulations, the power of W and $t_{\hat{M}-2}$, was evaluated with a 5 % nominal level.

On the other hand it might be interesting to calculate the power $\pi_T(s_{XY})$ of a test on the covariance s_{XY}, similar to W but where the estimate : $N^{-2} \sum N_k \hat{C}_X(k) \hat{C}_Y(k)$, of the variance of s_{XY} is replaced by its theoretical value under H1 :

$$N^{-2} \text{ tr } (\Sigma_\xi \Sigma_\eta) = N^{-2} [a^2 \text{ tr}(\Sigma_\xi^2) + \text{tr}(\Sigma_\xi \Sigma_\theta)].$$

where Σ_θ denotes the variance-covariance matrix of the centered vector $W - \bar{W}$ and Σ_ξ and Σ_η are defined as before.

In order to carry out the calculation of $\pi_T(s_{XY})$, it is necessary to suppose that the distribution of $\sqrt{N}\, s_{XY}$ under H1 is approximately normal. This approximation can be justified by central limit theorems if appropriate hypotheses are placed on the rate of decrease of the autocovariances of X and W as a function of the lag.

The expectation and the variance of s_{XY} under H1 are given by :

$$E_{H1}(N s_{XY}) = a \text{ tr } (\Sigma_\xi)$$

$$V_{H1}(N s_{XY}) = 2a^2 \text{ tr } (\Sigma_\xi^2) + \text{tr } (\Sigma_\xi \Sigma_\theta).$$

The traces of the matrices $\Sigma_\xi\Sigma_\theta$ or their product can be expressed in terms of Σ_X and Σ_W and thus evaluated for specific models for Σ_X and Σ_W.

The power $\pi_T(s_{XY})$ of the test of the covariance using the statistic $Ns_{XY}\, tr(\Sigma_\xi\, \Sigma_\eta)^{-1/2}$ for a bilateral test of nominal level α is thus equal to $1 - [\Phi(M_1) - \Phi(M_2)]$

with :

$$M_1 = \frac{C_\alpha [a^2 tr(\Sigma_\xi^2) + tr(\Sigma_\xi\Sigma_\theta)]^{1/2} - a\, tr(\Sigma_\xi)}{[2a^2 tr(\Sigma_\xi^2) + tr(\Sigma_\xi\Sigma_\theta)]^{1/2}}$$

$$M_2 = \frac{-C_\alpha [a^2 tr(\Sigma_\xi^2) + tr(\Sigma_\xi\Sigma_\theta)]^{1/2} - a\, tr(\Sigma_\xi)}{[2a^2 tr(\Sigma_\xi^2) + tr(\Sigma_\xi\Sigma_\theta)]^{1/2}}$$

Φ being the N(0,1) distribution function and C_α being such that $P\{|N(0,1)| > C_\alpha\} = \alpha$.

It is also interesting to be able to compare the power observed by simulations to a reference value. Along these lines, we thus also calculated the power of the classical test of r_{XY} ; $\pi_{N^*}(r)$, in a case which would be compatible with the observed empirical variance of r_{XY}, v_e, estimated by the Monte Carlo simulations. Recall that in the case of non autocorrelated variables X and Y and large samples, the variance of r_{XY} is approximately equal to $(1 - \rho^2_{XY})^2 / N-1$ for a sample of N observations. For autocorrelated X and Y, we thus computed an approximately equivalent sample size, N* :

$$N^* = 1 + (1 - \rho^2_{XY})^2 / v_e$$

This number N* was used to compute the reference value, $\pi_{N^*}(r)$, power of the classical test of r_{XY} based on N* observations.

A summary of the results is given in Table1a and 1b. In those tables, the observed power of W is only given since it is almost identical to that of $t_{\hat{M}\text{-}2}$. Overall all the powers, whether observed or calculated, are close. The only differences are seen a few cases of high autocorrelation for either X and W.

In summary we can say that the "theoretical power" $\pi_T(s_{XY})$ gives a good approximation of the observed power in most cases and does not require Monte-Carlo simulations. Furthermore, the modified W and $t_{\hat{M}\text{-}2}$ tests have comparable power to that of a classical test based on an "equivalent" number of observations N*.

5. Choice of the partition for the covariance structure of X and Y.

Computations needed in order to apply the modified tests to a data set are straight forward but rely upon a choice of strata $\{S_0, S_1, S_2, \ldots\}$ of AxA, on each of which the covariance of the two processes is assumed to be constant.

In Table 2 the performance of the $t_{\hat{M}-2}$ statistic is investigated for different choices of strata in the case of the irregular grid network and gaussian disc models. Six partitions were defined, ranging from 5 to 21 strata and corresponding to different discretisations of the distance between pairs of locations (assuming isotropy).

Overall the type I error was close to 5 % for most of the partitions. As the autocorrelation increased, the type I error for the 5-strata partition was inflated whereas there was a stability in the observed error rate when the number of strata increased. The results for the W statistic were similar.

This confirmed that a balance has to be reached between choosing too few classes which can bias expression (1) or too many resulting in less precise estimates of the autocovariances. Clearly in the case simulated where the autocorrelation decreases smoothly with distance, the performance of the $t_{\hat{M}-2}$ statistic is robust to various choices of partitions.

6. Confidence interval for regression coefficient

Once we have a test for independence it is possible to construct a confidence interval for the regression coefficient b, in the model

$$Y = a1 + bX + Z$$

where Z is a process independent of X and 1 is a vector with unit elements. The confidence interval for b is the set of values of b which we would not actually reject i.e the set of b such that Y - bX has no significant correlation with X. Defining :

$f_\alpha = X_\alpha - \bar{X}$ and $g_\alpha = Y_\alpha - \bar{Y}$ the standardised covariance between Y - bX and X is :

$$W_b = \frac{(g^T f - b f^T f)}{\sqrt{\Sigma_k N_k \hat{C}_X(k) \hat{C}_{Y-bX}(k)}}$$

where $\hat{C}_{Y-bX}(k) = \hat{C}_Y(k) + b^2 \hat{C}_X(k) - 2b N_k^{-1} \Sigma_{S_k} f_\alpha g_\beta$.

This standardised covariance can be used as a pivot to obtain a confidence interval for b. We do not reject the null hypothesis when

$$|W_b| < C_\alpha \quad \text{where : } P\{|N(0,1)| > C_\alpha\} = \alpha.$$

Therefore the confidence interval is :

$$\{b : (g^T f - b f^T f)^2 \leq C_\alpha^2 \sum N_k \hat{C}_X(k)\ \hat{C}_{Y-bX}(k)\} \qquad (4)$$

$$\text{or } b - T_2 \pm (T_2^2 + T_1 - T_1 T_3 - 2bT_2 + b^2 T_3)^{1/2} (1 - T_3)^{-1}$$

where $b = s_{XY} / s_X^2$; $T_1 = d_X \sum_k N_k \hat{C}_Y(k) \hat{C}_X(k)$; $T_2 = d_X \sum_k N_k \hat{C}_X(k) \hat{C}_{XY}(k)$
$T_3 = d_X \sum_k N_k \hat{C}_X^2(k)$ and where $\hat{C}_{XY}(k) = \sum f_\alpha\ g_\alpha / N_k$ and $d_X = C_\alpha^2 N^{-2}\ s_X^{-4}$.

7. Examples

Our examples concern the relationship between lung cancer, smoking and industrial factors. We calculate W, $t_{\hat{M}-2}$ and a confidence interval for the regression coefficient for each example. To provide an additional check on the performance of our tests, we also carried out a Monte Carlo test based on the disc model for cases in which such a model is plausible.

The data

For 82 départements we considered male lung cancer mortality rate (LC) over a 2 year period, 1968-1969, standardised over the age 35-74, cigarette sales per inhabitant (CS) in 1953 (a fifteen year time lag was chosen to account for the delay between exposure and the onset of the pathology) and demographic data on the percentage of employed males in the metal industry (MW) and the textile industry (TW) recorded by census in 1962.

The coordinates of the points of the network were identified with the geographical locations of the administrative centers ("préfectures") of French départements. The spatial structure of these variables was investigated by means of a variogram. In this analysis, N = 82 locations were retained after grouping the départements around Paris into one area. The distances between the centres of départements were partitionned into 15 classes of 50 kilometres intervals each. This gives 15 strata S_1, ... S_{15}.

The observed variograms, i.e the plot of $N_k^{-1} \sum_{(\alpha,\beta)\ \varepsilon\ S_k} (X_\alpha - X_\beta)^2$, $k = 1,\ldots 15$, against the average distance, d_k, for départements in S_k, for the four variables considered are shown in Figure 3. Note that the last two classes contain few pairs and hence have a large variability. Three of the variograms (LC, CS, MW) exhibit clearly an upward trend with increasing distance. Up until the 10th class, the shape of this trend is fairly linear with increasing distance, thus compatible with the disc model for the covariance matrix discussed by Ripley (1981). The variogram for TW gives little indication of any spatial autocorrelation.

Results

Using the standard test base on r_{XY} there is a highly significant positive association both between LC and CS and between LC and MW. The association between LC and TW on the other hand is less strong but still significant at the 1 % level (Table 3). The W and $t_{\hat{M}-2}$ statistics for these 3 examples are shown in Table 3. One can see a substantial reduction of the degrees of freedom when the autocorrelation is taken into account. We note that this occurs also for the case LC-TW which is surprising if one recalls the shape of the variogram of TW. A possible explanation for this is that the geographically small départements, which are over-represented in the first few strata, are atypical for this particular variable (TW). For CS and MW the effective sample size is about 20 % of the original sample size. Consequently the significance levels are reduced but even after this "adjustement" these two factors are statistically significantly associated with lung cancer. For TW the significance disappears after adjustement.

The first two examples were also investigated by a Monte Carlo test. A disc model for the covariance was fitted by maximum likelihood. The parameters were found by direct search and corresponded to autocorrelation ρ (1) of 0.91, 0.85 and 0.91 respectively for LC, CS and MW.

For each example, 1000 pairs of mutually independent variables, with covariance given by the estimated disc model, were generated and the observed correlation coefficient was ranked among the 1000 generated coefficients. The significance levels obtained are given in Table 3. The agreement between them and the significance levels of the W or $t_{\hat{M}-2}$ tests is better for CS than for MW ; this is possibly due to a better fit of the disc model for the CS variable. Confidence intervals, given by (4), for the regression coefficients were also calculated. We note that they are not symmetric.

8. Discussion

In this paper we have studied the properties of modified tests of the empirical correlation coefficient between two spatial processes. We have shown that by a simple adjustment, correct level of significance can be reached and that the power under a simple linear alternative is compatible with that of a standed test in an equivalent situation. These tests can be applied both to regularly and irregularly spaced points and can be considered as a first step in an analysis of association when detailed spatial modelling is not suitable.

The performance did not vary much when different strata of equal covariance were chosen. On small lattices, the modified $t_{\hat{M}-2}$ statistic is better than the standardised covariance.

Application of these tests to data may also give pivotal confidence interval for the regression coefficient. Furthermore, if one is prepared to model the observed covariance structure, Monte Carlo tests of association can be performed. In the examples investigated, the results from the two types of testing procedures were close.

It would be interesting to develop distribution-free tests of association based on permutations and to compare their performance to that of the proposed modified tests in the case of non gaussian spatial distributions.

REFERENCES

- Cliff, A.D. and J.K. Ord (1981), Spatial Processes. Models and Applications. London, Pion.

- Clifford, P. Richardson, S. and Hémon, D. Assessing the significance of the correlation between two spatial processes, (1989) to appear in Biometrics.

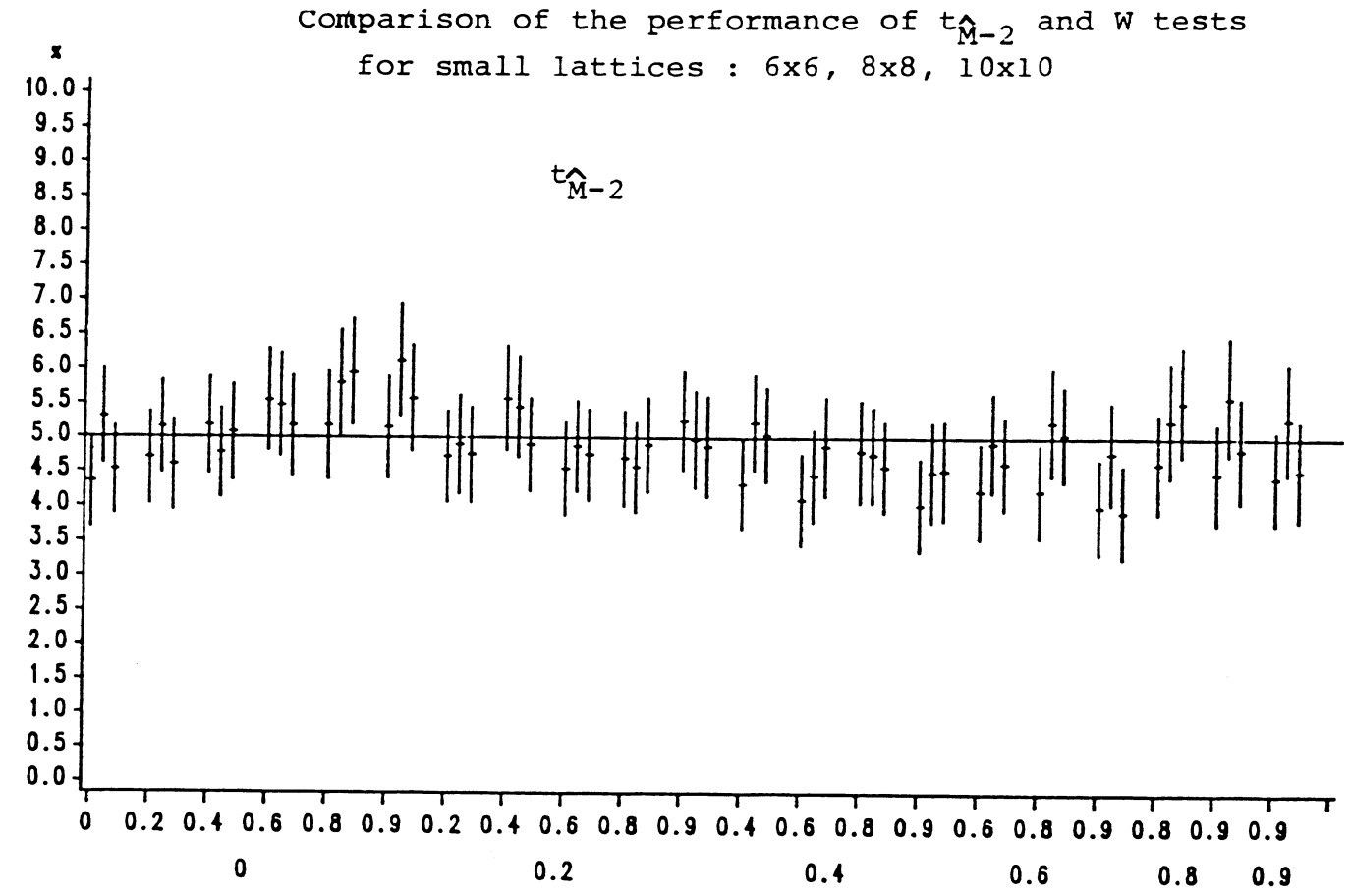
- Doll, R. (1980). The epidemiology of cancer. Cancer, 45, 2475-2485.

- Doreian, P. (1981), Estimating linear models with spatially distributed data, in Sociological Methodology, edited by Leinhardt and Samuel. San Francisco, Jossey-Bass Publishers, 359-388.

- Malin, S.R.C. and Hide, R. (1982). Bumps on the core-mantle boundary : geomagnetic and gravitational evidence revisted. Phil. Trans. R. Soc. Lon A, 306, 281-289.

- Ripley, B.D. (1981), Spatial Statistics. New York, Wiley.

Figure 1



95 % C.I. for the % of type I errors for a 5 % nominal test

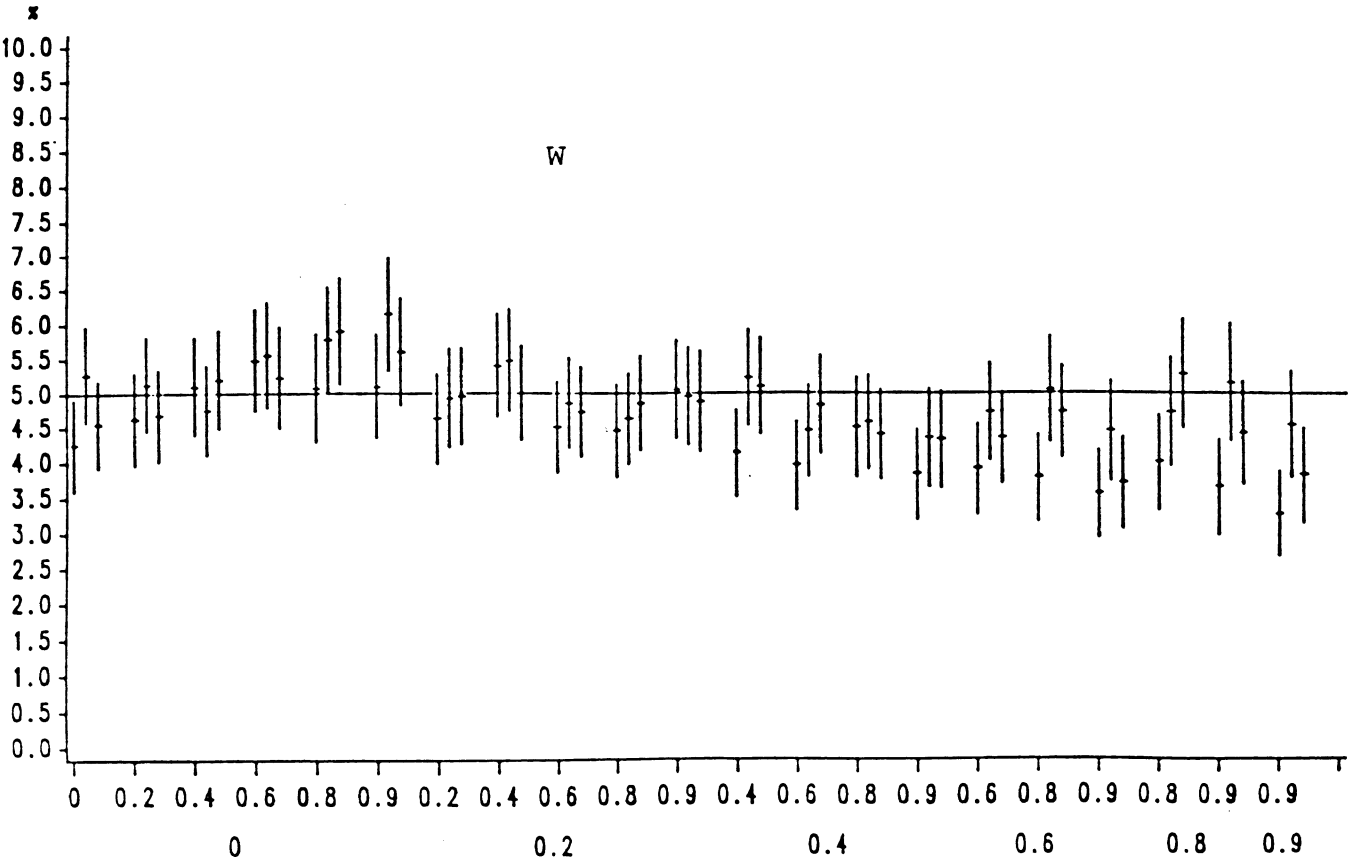
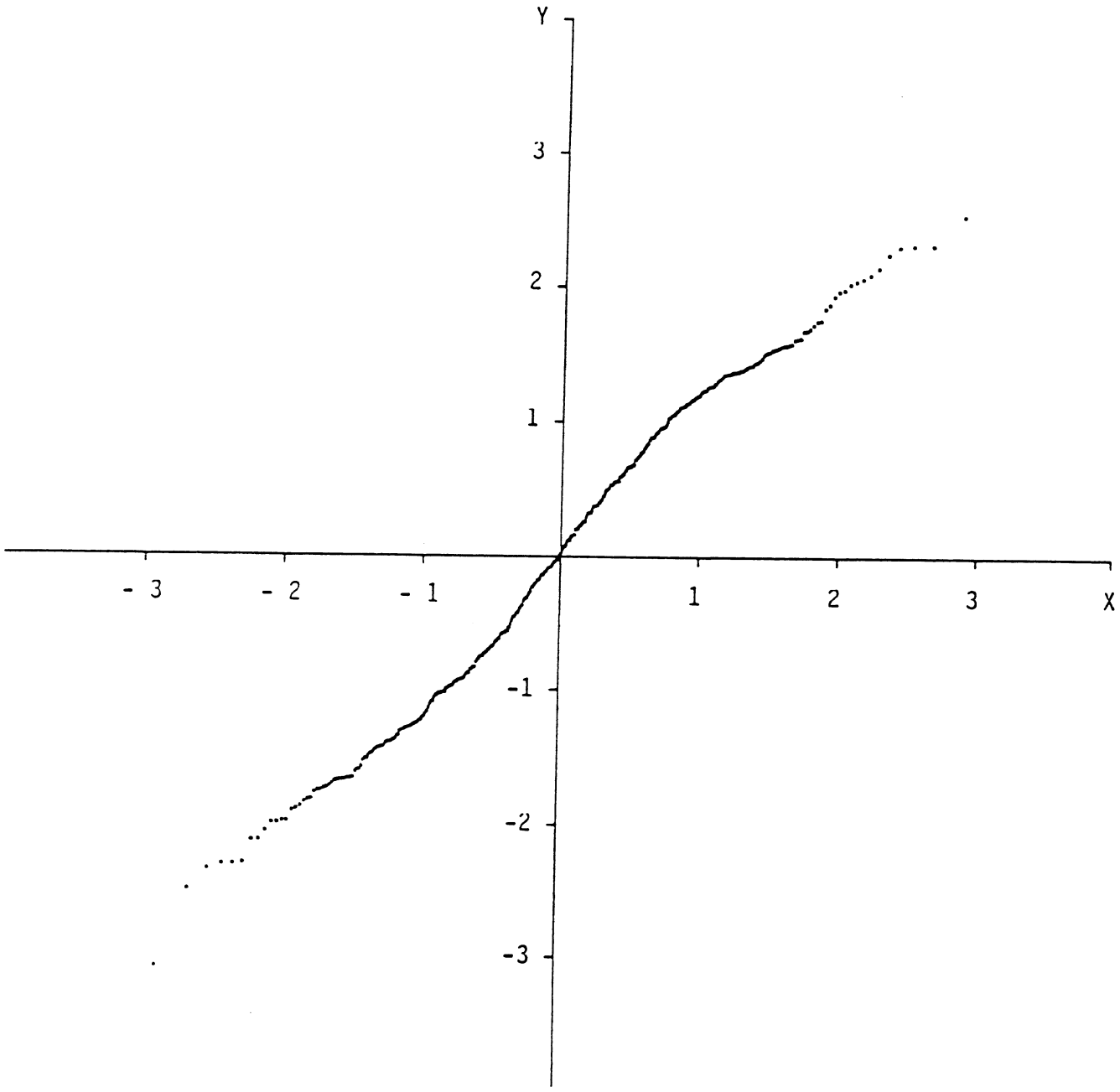


Figure 2

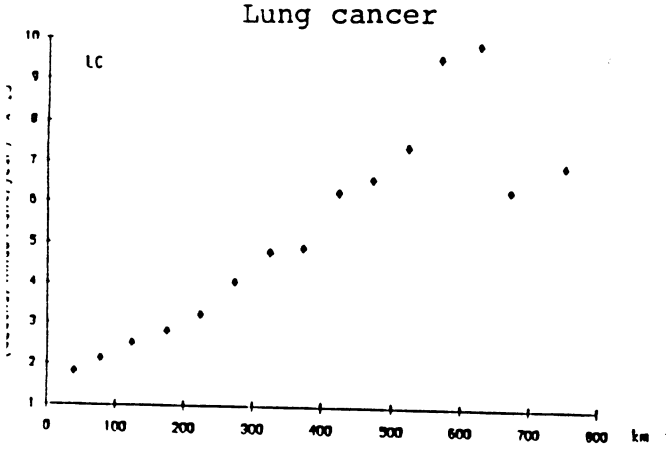


QQ plot (sample quantiles (Y) against quantiles of a standard normal distributions (X)) of 500 trials for the W statistic with two mutually independent simultaneous autogressive processes (12x12 lattice, $\rho_X(1) = \rho_Y(1) = 0.8$)

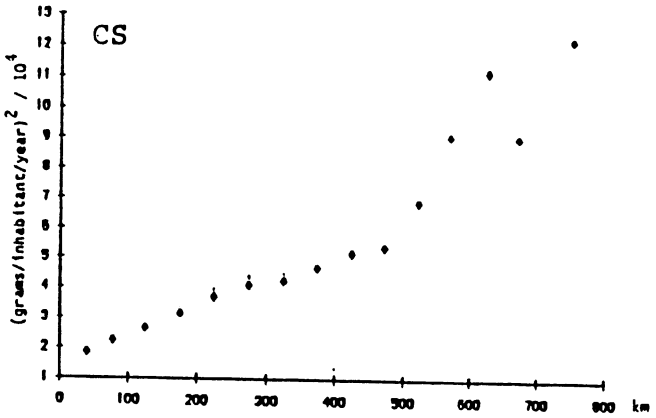
Figure 3

VARIOGRAMS OF THE VARIABLES CONSIDERED IN &7

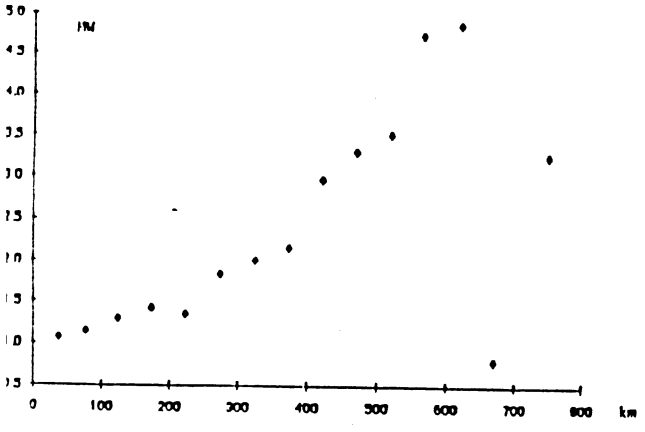
Lung cancer



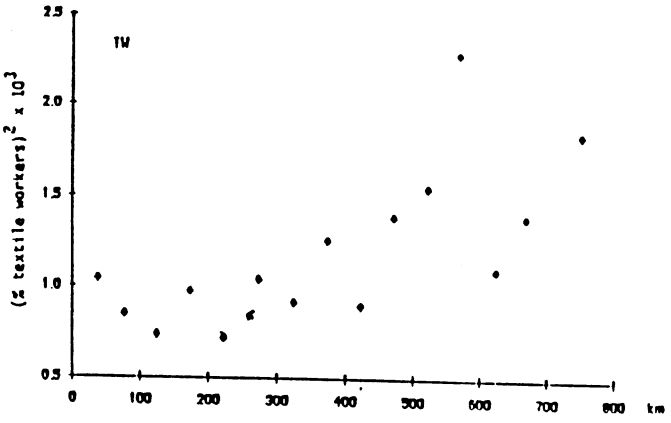
Cigarette sales/inh



% workers in metal industry



% workers in textile industry



Fifteen classes of distance are considered. The number of pairs in each class is : 82, 400, 582, 674, 764, 822, 812, 726, 630, 476, 304, 172, 94, 58, 40.

Table 1a

Power of the modified tests: results concerning the testing of the correlation between X and $Y = aX + W$ where both X and Y are spatially autocorrelated, X and W independent and of equal variance and a is chosen so that the correlation ρ_{XY} between X and Y takes the value 0.2.

ρ_W	ρ_X	$\rho_{xy} = 0.2$				
		0.0	0.2	0.4	0.6	0.8
	N*	79	82	78	79	84
	$\pi_{N^*}(r)$	0.42	0.44	0.42	0.42	0.44
0.0	power of W	0.42	0.44	0.43	0.39	0.35
	$\pi_T(s_{XY})$	0.44	0.44	0.43	0.42	0.37
		76	74	67	71	63
0.2		0.41	0.40	0.36	0.38	0.34
		0.42	0.42	0.39	0.34	0.30
		0.44	0.41	0.39	0.36	0.32
		80	78	66	57	54
0.4		0.42	0.42	0.36	0.32	0.30
		0.44	0.36	0.37	0.34	0.21
		0.44	0.39	0.36	0.32	0.27
		78	75	61	46	35
0.6		0.41	0.40	0.33	0.25	0.20
		0.48	0.36	0.36	0.28	0.20
		0.45	0.39	0.34	0.27	0.20
		78	66	54	39	19
0.8		0.41	0.36	0.30	0.22	0.12
		0.52	0.48	0.42	0.28	0.12
		0.48	0.40	0.33	0.23	0.13

500 simulations were carried out. The observed power of W is compared with $\pi_{N^*}(r)$ and $\pi_T(s_{XY})$ (cf § 4). The observed power of t_{M-2} is almost identical to that of W. Standard deviations and confidence intervals for the observed proportions can be calculated according to binomial sampling.

Table 1b

Power of the modified tests: results concerning the testing of the correlation between X and Y= aX + W where both X and Y are spatially autocorrelated, X and W independent and of equal variance and a is chosen so that the correlation ρ_{XY} between X and Y takes the value 0.4.

ρ_W	ρ_X	$\rho_{xy} = 0.4$ 0.0	0.2	0.4	0.6	0.8
	N*	83	87	81	70	63
0.0	$\pi_{N^*}(r)$	0.97	0.97	0.96	0.93	0.90
	power of W	0.97	0.96	0.96	0.92	0.87
	$\pi_T(s_{XY})$	0.94	0.93	0.92	0.89	0.74
0.2		82	73	76	68	55
		0.96	0.94	0.95	0.90	0.86
		0.96	0.96	0.94	0.92	0.83
		0.94	0.91	0.90	0.85	0.70
0.4		84	67	61	63	49
		0.97	0.92	0.89	0.90	0.81
		0.97	0.92	0.91	0.88	0.79
		0.94	0.91	0.87	0.81	0.65
0.6		83	69	66	44	30
		0.97	0.93	0.92	0.77	0.58
		0.96	0.95	0.93	0.81	0.56
		0.94	0.90	0.85	0.74	0.54
0.8		68	52	51	33	20
		0.92	0.83	0.82	0.63	0.39
		0.97	0.93	0.89	0.76	0.42
		0.95	0.91	0.84	0.68	0.38

500 simulations were carried out. The observed power of W is compared with $\pi_{N^*}(r)$ and $\pi_T(s_{XY})$ (cf § 4). The observed power of $t_{\hat{M}-2}$ is almost identical to that of W. Standard deviations and confidence intervals for the observed proportions can be calculated according to binomial sampling.

Table 2

PERCENTAGE OF TYPE I ERRORS OF THE $t_{M-2}^{\wedge}$ STATISTIC FOR DIFFERENT PARTITIONS OF THE COVARIANCE STRUCTURE

Number of strata in each partition	5	9	13	15	17	21
$\rho_X^{(1)} = \rho_Y^{(1)} = 0$	0.056	0.052	0.05	0.056	0.054	0.054
$\rho_X^{(1)} = \rho_Y^{(1)} = 0.2$	0.04	0.04	0.036	0.032	0.034	0.03
$\rho_X^{(1)} = \rho_Y^{(1)} = 0.4$	0.062	0.046	0.05	0.05	0.052	0.05
$\rho_X^{(1)} = \rho_Y^{(1)} = 0.6$	0.066	0.058	0.054	0.056	0.052	0.052
$\rho_X^{(1)} = \rho_Y^{(1)} = 0.8$	0.068	0.058	0.048	0.04	0.046	0.046

	r	t_{N-2}*	W	$\hat{M}$	$t_{\hat{M}-2}$	Monte Carlo+	95%CI for $\hat{\gamma}$
Cigarette sales per inhabitant (1953) (CS)	0.76	10.48 $p \simeq 10^{-21}$	2.94 p=0.0032	15	4.22 p=0.001	2/1000	0.78 [0.54;0.88]
% male workers in metal industry (1962) (MW)	0.63	7.16 $p \simeq 10^{-11}$	2.48 p=0.0136	16	3.00 p=0.01	45/1000	0.29 [0.11;0.36]
% male workers in textile industry (1962) (TW)	0.28	2.57 $p \simeq 0.01$	1.51 p=0.13	30	1.52 p=0.15	-	0.18 [-0.07;0.37]

Table 3 : Comparison of the significance levels for tests of the association between lung cancer mortality rates and several risk factors given by standard test, W and $t_{\hat{M}-2}$ tests and Monte Carlo simulations, $\hat{\gamma}$ is the estimated regression coefficient.

* standard test (t transformation with 80 d.f.)

\+ Monte Carlo significance levels over 1000 simulations.

A METHOD FOR TESTING THE SIGNIFICANCE OF GEOGRAPHICAL CORRELATIONS WITH APPLICATION TO INDUSTRIAL LUNG CANCER IN FRANCE

A method for testing the significance of geographical correlations with application to industrial lung cancer in France

Sylvia RICHARDSON

SUMMARY

This paper discusses some of the problems encountered when using tests of significance in geographical epidemiology studies where the variables analysed will typically exhibit some spatial autocorrelation.

A test of partial correlations between spatially autocorrelated variables is presented. This test is based on an evaluation of an effective sample size which takes into account the spatial structure. Its performance is assessed via Monte Carlo simulations. The method proposed is applied to studying the relationship between male lung cancer rate and specific industries.

Key words :
significance tests, partial correlation, spatial processes, geographical epidemiology, lung cancer.

1. *Introduction*

Ecological correlation studies aim to analyse the association between a set of variables defined on groups. Among these, geographical correlation studies are particularly concerned with variables measured as averages over geographical units and the study of their joint variations.

In epidemiology, studies of the geographical distribution of incidence or mortality rates for particular diseases have been widely used for obtaining some clues about the etiology of those diseases[1,2,3]. In particular, the correlation between the spatial variation of exposure indicators, like dietary practice or industrial employment, and health indicators have been investigated[4,5]. Ecological analysis has also been used to check epidemiological hypotheses suggested by other approaches such as animal experiments, case-control or cohort studies[6,7].

Examples of geographical correlation studies are also found in other domains like regional science, for instance when relating agricultural output and road accessibility[8], or sociology with reference for instance to voting behaviour and socio-economic factors[9].

When considering results from such studies, it is necessary to be aware of the methodological problems which arise from their design. These problems fall broadly into two categories : problems of interpretation due to the nature of the data considered, and statistical problems connected with the spatial structure of the data. Indeed, the data units can rarely be considered as independent replicates and often present some spatial autocorrelation.

In epidemiology, the limitations which stem from the nature of the data have been reviewed by many authors[10-13]. They are related to numerous factors : the potential non uniformity in diagnosis and registration of causes of death, the difficulty in adequately defining the population at risk, the difficulty in finding data on health and exposure on the same geographical units, the use of average

exposure when exposure is heterogeneous within the population and the difficulty in considering appropriate time lags between exposure and disease.
A discussion of the biases encountered when trying to estimate quantitatively relative risks from ecological studies is given in Richardson, Stücker and Hémon[14] and Greenland and Morgenstern[15].

Furthermore, geographical data often exhibit a regular component in these spatial variations analagous to a trend or a gradient. Such gradients, if they are present in both the exposure and the health indicator rates, make the interpretation of the joint variations difficult since many potential confounder variables may also show the same regular spatial gradient. This is clearly illustrated in the study of cardiovascular mortality and water hardness of Pocock et al[16] where a northwest to southeast gradient is observed for both variables which is also related to climatic and industrial variables in Great Britain. At an individual level, this situation would be equivalent to having a confounding factor whose distribution nearly parallels that of the risk factor investigated. This has suggested a further step in the analysis : correlating the residual variations in the exposure and the health variables after removing a trend component or gradient which is obtained by regression of each of the variables on the coordinates of the geographical units. This is equivalent to computing the partial correlation between the exposure and the health variables after conditioning on the coordinates of the geographical units. A significant correlation between both the non-adjusted and adjusted observations would strengthen a causal interpretation.

Similarly one might want to adjust the relationship investigated on some geographically distributed confounding factors. Nevertheless, the residual variations after the removing of confounding factors and/or a spatial gradient might still be spatially autocorrelated. In a multifactorial chronic disease for instance, this residual autocorrelation might arise from the spatial structure of other unidentified or unmeasurable risk factors. In the case of autocorrelation

among the residuals the classical methods for testing the significance of the association are not applicable.

In this paper, we will describe a method for assessing the significance of a partial correlation coefficent in the presence of spatial autocorrelation. This method is an extension of a modified test of correlation developped in Clifford, Richardson and Hémon[17] to the case of several variables. The proposed method is outlined in section 3, and some results on its performance are given which were obtained by Monte Carlo simulations. This section is preceded by a review of some available methods in geographical correlation studies where, in particular, the modified test of correlation between two spatially autocorrelated variables is briefly described as well as the fitting of multivariate regression models with a variance-covariance spatially parametrised error matrix. In a final section, the test proposed is applied to an example concerning the relationship between lung cancer mortality for men in France and particular industries. This relationship is investigated at the level of the French *départements* and adjustment is made on cigarette sales. The results clearly illustrate the interest of performing an adjustment for spatial autocorrelation. The effect of adjusting a linear spatial gradient is also discussed on the examples and through some simulations.

2. *Review*

The data usually considered in geographical correlation studies is a set A of N locations and variables indexed by their locations. These variables will typically exhibit some degree of spatial autocorrelation. A first step in the analysis is often the calculation and testing of correlation coefficents between pairs of variables (X_α, Y_α), $\alpha \varepsilon A$.

The consequences of neglecting the autocorrelation in the X and Y variables were first pointed out by Bartlett[18] in the case of stationary time series. When the autocorrelations are positive, the significance level of classical tests

are overestimated. This phenomenon is more accentuated for spatial series. For instance, when X_α and Y_α are both spatially autocorrelated with nearest neighbour autocorrelation around 0.8, the observed significance level of a 5 % nominal test of the correlation coefficent can be increased up to 50 % [17].

The situation is similar in regression analysis. The effect of neglecting the autocorrelation in the dependent variables or in the error term has been pointed out by several authors, Johnston[19] for time series, Cliff and Ord[8] for spatial series. Inflated values of the t and F statistics are again found when there is positive autocorrelation even though the estimation of the regression coefficient is not biased.

A modified test of the correlation coefficient r_{XY} has been proposed by Clifford, Richardson and Hémon[17] for testing the association between a pair of variables (X_α, Y_α). It consists in modifying the degrees of freedom of the classical test on r_{XY}. An effective sample size is calculated based on an estimation of the variance of r_{XY} which takes the internal autocorrelations of X and Y into account.

Suppose that X and Y are independent but that both X and Y are multivariate normal vectors with constant means and NxN variance-covariance matrices Σ_X and Σ_Y respectively. It can be shown that, to the first order, the variance of r_{XY}, σ^2_r is :

$$\sigma^2_r = \frac{\mathrm{Var}(s_{XY})}{E(s_X^2)\, E(s_Y^2)} \quad , \qquad (1)$$

where s_{XY} denotes the empirical covariance between pairs of observations (X_α, Y_α), $\alpha \varepsilon A$ and s_X^2 (respectively s_Y^2) the empirical variances of X (respectively Y).

To be able to estimate var (s_{XY}), one needs to impose a stratified structure for Σ_X and Σ_Y. More precisely pairs in A x A are divided into strata $S_0, S_1, S_2, \ldots$ such that the covariances within strata remain constant, i.e. $\mathrm{cov}\,(X_\alpha, X_\beta) = C_X(k)$, if $(\alpha,\beta) \varepsilon S_k$.

An estimate of the conditional variance of s_{XY} is then derived :

$$N^{-2} \Sigma N_k \hat{C}_X(k)\ \hat{C}_Y(k) \quad , \qquad (2)$$

where N_k is the number of pairs in strata S_k and $\hat{C}_X(k)$ (respectively $\hat{C}_Y(k)$) is the estimated autocovariance :

$$\hat{C}_X(k) = \sum_{S_k} (X_\alpha - \bar{X})(X_\beta - \bar{X}) / N_k \quad .$$

Thus the estimate takes into account the autocorrelation of both X and Y.

Equations (1) and (2) lead to the following estimate of the variance of r :

$$\hat{\sigma}^2_r = \frac{\Sigma N_k \hat{C}_X(k) \hat{C}_Y(k)}{N^2 s_X^2 s_Y^2} \quad . \qquad (3)$$

In the classical non autocorrelated case, when either Σ_X or $\Sigma_Y = I$, it can be shown that the approximation given by (1) is exact and that $(N-2)^{1/2} r_{XY} / (1-r^2_{XY})^{1/2}$ follows a t-distribution with N-2 degrees of freedom (d.f.) (t_{N-2}). Further : $N=1+(\sigma^2_r)^{-1}$

In general an estimated effective sample size, $\hat{M}$, is defined by the relationship

$$\hat{M} = 1 + (\hat{\sigma}^2_r)^{-1} \quad ,$$

where $\hat{\sigma}^2_r$ is given by (3). A modified t-test : $t_{\hat{M}-2}$ is proposed which rejects the null hypothesis of no association when :

$$| (\hat{M}-2)^{1/2} \; r(1 - r^2)^{-1/2} | > t^\alpha_{\hat{M}-2} \qquad (4)$$

where $t^\alpha_{\hat{M}-2}$ is critical value of the t-statistic with $\hat{M}$-2 degrees of freedom.

Monte Carlo simulations carried out both for lattice and non lattice models show that the $t_{\hat{M}-2}$ statistic has indeed a correct significance level[17].

In the case of several variables, classical regression analysis can be extended. The spatial structure is taken into account either by considering a mixed simultaneous regressive-autoregressive model :

$$Y = L(Y) + X\beta + \varepsilon \qquad (5)$$

where $\{\varepsilon_\alpha\}$ are i.i.d. $N(0, \sigma^2)$, and $L(Y)_\alpha$ represents some linear function of the values of Y in neighbouring locations $\gamma, \gamma \neq \alpha$; or by considering a regression model with autocorrelated error term :

$$Y = X\beta + U \qquad (6)$$

where the spatial autocorrelation is reflected in the structure of Σ_U, the variance-covariance matrix of the error U, which is parametrised in terms of a finite number of parameters θ.

The interpretation of β is essentially different for the regressions (5) and (6). In model (5), β represents the effect of X_α on Y_α when the influence of neighbouring values of Y_α has already been substracted. When relating the incidence of a non contagious disease to environmental factors, model (5) is not plausible since disease rates in neighbouring areas do not have a direct causal influence and model (6) should be chosen. In model (6), it is assumed that the autocorrelation of the variable Y comes from an underlying structural variable U.

The estimation of models (5) and (6) with an autocorrelation structure modelled via a contiguity matrix of weights W, was first discussed by Ord[20]. Mardia and Marshall[21] have studied the asymptotic properties of the maximum likelihood estimators for the model (6). Assuming that Y is a gaussian process, they give conditions which ensure the consistency and the asymptotic normality of $(\hat{\beta},\hat{\theta})$ the maximum likelihood estimators of (β, θ). Conditionally upon $\hat{\theta}$, the variance-covariance matrix for $\hat{\beta}$ can be calculated and hence tests of the regression coefficients can be performed. Because of the conditionality, the standard errors of $\hat{\beta}$ might be underestimated, which leads some authors to prefer a Bayesian approach[22].

In using model (6), one is left with choosing how to parametrise Σ_U and to perform a numerical maximisation of the likelihood function which can lead to

involved computing and is fraught with difficulties. A study of Warnes and Ripley[23] showed that the Fisher scoring technique usually ensures convergence only to the nearest local maximum. Moreover Ripley[24] reports some simulation results where the global maximum found is well away from the true value. Models used for Σ_U have been basically of two kinds, either involving directly the contiguity matrix[20], or supposing that the elements of Σ_U depend upon the distance between pairs of locations with a functional form suggested by the variogram of the ordinary least square regression residuals. This is the approach chosen by Cook and Pocock[25] in their study of cardiovascular mortality and environmental factors (in particular water hardness). They fitted an exponentially decreasing function of distance for Σ_U and still found a statistically significant link with water hardness.

In summary, generalised regression where the variance-covariance matrix of the errors follows a specific spatial model involves some degree of arbitrariness and is often computationally difficult in practice. We have thus found it interesting to develop a simple test of partial correlation which does not depend on parametric modelling.

3. *A modified test for partial correlation*

Our aim is to show how the modified t_{M-2} test described in §2 can be simply extended to test partial correlations. For the sake of clarity the method is going to be described and assessed for testing the association between two variables (Y_α, Z_α) adjusted on a third one, X_α, $\alpha \in A$. Its generalisation to any number of adjustment variables is straightforward.

We suppose that the 3N vector $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ follows a multivariate normal distribution with mean $\begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{pmatrix}$ and variance-covariance matrix given by :

$$\begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} & \Sigma_{XZ} \\ \Sigma_{YX} & \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZX} & \Sigma_{ZY} & \Sigma_{ZZ} \end{pmatrix}.$$

Then the joint distribution of (Y,Z) conditional on X is also a multivariate normal with marginals given by $N(\mu_Y - \Sigma_{XY}\,\Sigma_{XX}^{-1}\mu_X,\ \Sigma_{YY} - \Sigma_{XY}\,\Sigma_{XX}^{-1}\,\Sigma_{YX})$ and $N(\mu_Z - \Sigma_{XZ}\,\Sigma_{XX}^{-1}\mu_X,\ \Sigma_{ZZ} - \Sigma_{XZ}\,\Sigma_{XX}^{-1}\,\Sigma_{ZX})$ respectively. We can therefore test the hypothesis : $\rho_{YZ.X} = 0$, where $\rho_{YZ.X}$ is the partial correlation between Y and Z conditional on X, by testing that the correlation between the residuals of the regression of Y on X and of Z in X is zero. Hence, the method outlined in §2 equations (1) to (4) can be extended to test the partial correlation by considering the relevant variance-covariances in the conditional distributions of Y and Z given above and carrying out the same adjustment of the degrees of freedom (d.f.). This adjustment will now take into account spatial autocorrelation in the conditional distributions. The same approach can be used if conditionally on X, (Y,Z) is multivariate normal even if X is not a random variable as it is the case when X is a vector of spatial coordinates.

In practice, this implies using the modified $t_{\hat{M}-2}$ statistic on the residuals of the linear regression of Y on X and of Z on X respectively. These residuals need therefore to be estimated. We are proposing to do this by ordinary least squares (O.L.S.), thus ignoring the autocorrelation, since the O.L.S regression estimates are unbiased.

In summary the following steps are followed :

1 - regress Y on X by O.L.S giving estimated residuals $\hat{U}$

2 - regress Z on X by O.L.S giving estimated residuals $\hat{V}$

3 - test the correlation coefficient between $\hat{U}$ and $\hat{V}$ using the modified test statistic $t_{\hat{M}-2}$ given in formula (4) with

$$\hat{M} = (\text{var}\ \ r_{\hat{U},\hat{V}})^{-1} + 1 \qquad (7).$$

If the residuals U and V are non autocorrelated, the variance of the partial correlation coefficient is $(N-2)^{-1}$, hence $\hat{M} = N-1$ as usual.

In order to check whether the estimation of the residuals by O.L.S could influence the performance of the test, a Monte-Carlo simulation was carried out to check the type I errors and to give some indication on the power of the proposed method.

Simulation model

In order to stay close to the epidemiological examples which will be analysed in the last section, the simulation was carried out on an irregular grid of points defined by the administrative centers of the French *département* (N = 82). Spatial dependence was introduced directly on the variance-covariance matrix of the multivariate normal distributions by considering a disc model. In this model, the covariance between the values of the variable in 2 areas (represented by their grid points) is defined as proportional to the intersection area of 2 discs of common radius R, centered on the grid points (see Ripley[26] p. 55). By varying the length of R, models with different degrees of autocorrelation can be generated. The shape of the covariance function of this model shows a fairly linear decrease with distance. This shape is similar to the observed variograms ; i.e. the plot of $N_k^{-1} \sum_{(\alpha,\gamma)\,\varepsilon\, S_k} (X_\alpha - X_\gamma)^2$ against the average distance between locations in S_k, of a number of variables such as lung cancer mortality rate or proportion of workers in specific industries (Figure 1). These will be considered in the next section. The parameter R of the disc model is chosen so that the autocorrelation for a distance of 40 km between points is equal to 0.2, 0.5 or 0.8. We denote this autocorrelation by $\rho(1)$ indexed by the name of the variable.

To simulate a process X on A, for each chosen value of $\rho_X(1)$, a NxN matrix Σ_X is generated following the disc model. Σ_X was then triangularised, $\Sigma_X = LL^T$,

and a realisation of X distributed as $N(0,\Sigma_X)$ was obtained by first generating a vector of N i.i.d. N (0,1) and then premultiplying this vector by L.

The distances between the centers of *départements* were partitioned into 15 classes of 50 km intervals each. This gives 15 strata S_1, ... S_{15} ; the stratum $S_O = \{(\alpha,\alpha), \alpha \varepsilon A\}$.

-Figure 1-

We now describe the model for partial correlation which was chosen. First three mutually independent autocorrelated variables, X, V and W are generated. Then we define : $Z = aX + V$, $Y = cX + dZ + W$. (8)
For simplification, the autocorrelation structure of X,V and W are chosen to be identical. Therefore $\rho_{YZ.X}$ depends only on the value of d : $\rho_{YZ.X} = d/(d^2+1)^{1/2}$.
The test statistic $t_{\hat{M}-2}$ is computed as described in §2 steps 1 to 3 and formula (7) using $\{\hat{C}_{\hat{U}}(k)\}$ and $\{\hat{C}_{\hat{V}}(k)\}$, k = 1,... 15, the estimated autocovariances in each of the 15 strata of the O.L.S residuals : $\hat{U}$ and $\hat{V}$, from the regression of Y on X and Z on X respectively.

Five thousand simulations were carried out for $\rho_{YZ.X}$ = 0.0, 0.2 and 0.4 and $\rho_X(1) = \rho_V(1) = \rho_W(1)$ equal to 0.0, 0.2, 0.5 and 0.8. The values of a and c were fixed. Several sets of a and c were tried and were found not to influence the results.

RESULTS

Type I errors of the modified statistic were first checked by setting d=0 in (8) and choosing 5 % and 1 % nominal levels. The results are summarised in Table 1. At the 5 % nominal level the performance of the modified $t_{\hat{M}-2}$ statistic is clearly satisfactory as the value 5 % belongs to all the confidence intervals and there is no systematic variation with increasing autocorrelation. At the 1 % level the modified $t_{\hat{M}-2}$ statistic becomes slightly conservative for the highest autocorrelation but is otherwise satisfactory. As an illustration, the empirical

variance of $r_{YZ.X}$ is also shown and one can see that it increases significantly with $\rho(1)$. As expected, the type I errors of the non adjusted standard test of partial autocorrelation are greatly inflated for high values of $\rho(1)$. The results are similar to those obtained in an earlier work for the test of the single correlation coefficient[17].

In choosing in practice the number of strata for the partition of AxA, a compromise has to be reached between choosing to few strata resulting in a bias in formula (2) or too many leading to imprecise estimates of the autocovariances in each strata. Further simulations were carried out to check the robustness of the performance of the $t_{\hat{M}-2}$ statistic with respect to the choice of partitions. Using the same random number generator seed as in Table 1 and 5000 trials, we evaluated the type I error for isotropic partitions ranging from 5 classes of around 150 km each to 21 classes of around 35 km each, in the most highly autocorrelated case ($\rho_X(1) = \rho_Y(1) = 0.8$). The observed type I errors stayed overall close to their nominal level and decreased progressively from 6.4 % [5.7 %-7.1 %] for 5 strata to 4.8 % [4.2%-5.4%] for 21 strata for a 5 % nominal level and from 1.1 % [0.86 % - 1.45 %] to 0.56 % [0.35 % - 0.76 %] for a 1 % nominal level. This indicates that at the 5 % level the test is slightly over significant when too few strata are used and that a finer partition with number of strata between 11 and 17 is preferable. In a case (like the one simulated) where the autocorrelation decreases smoothly with distance, one can thus see that the performance of the $t_{\hat{M}-2}$ statistic is robust to various choices of partitions.

Table 1

The power of the modified test was then investigated for value of $\rho_{YZ.X}$ equal to 0.2 and 0.4 and a 5 % nominal significance level. Higher values of $\rho_{YZ.X}$ led to a power close to 100 % and were less informative.

In order to have a reference value for the power the modified $t_{\hat{M}-2}$ statistic, we also calculated the power of the standard test of $\rho_{YZ.X}$ in a case which would be compatible with the observed empirical variance of $r_{YZ.X}$, v_e, estimated by

Monte-Carlo simulations. Recall that in the case of non autocorrelated variables X, Y and Z, the variance of $r_{YZ.X}$ is approximatively equal to $(1 - \rho^2_{YZ.X})^2 / N-2$ for a sample of N observations when N is large. For autocorrelated X, Y and Z we thus computed an equivalent sample size, N* :

$$N^* = 2 + (1 - \rho^2_{YZ.X})^2 / v_e \quad . \quad (9)$$

This number N* was used to compute the reference value π, power of the standard test of $\rho_{YZ.X}$ based on N* observations.

A summary of the results is given in Table 2. On the first line is indicated v_e, the empirical variance of $r_{YZ.X}$. One can see that v_e is smaller for $\rho_{YZ.X} = 0.4$ than for $\rho_{YZ.X} = 0.2$ for the same autocorrelations. This is similar to the case of non correlated variables. Further v_e increases noticeably with increasing $\rho(1)$. N* and π are given respectively on the 2nd and 3rd lines. Note that as expected N* is fairly constant for the same autocorrelation. The power of the test $t_{\hat{M}-2}$ is given on the 4th line. Clearly the test $t_{\hat{M}-2}$ perform satisfactorily as its power is comparable to π.

-Table 2-

§4 Examples

We consider in this section examples concerning the relationship between male lung cancer and some industrial factors after adjustment on smoking. Among the different cancer sites, male lung cancer has been frequently associated with industrial exposure[27,28] and the estimation of the proportion of lung cancer due to occupational risk factors has been the subject of recent debates[29]. To illustrate and discuss further the method proposed we present some results on 4 industrial branches : metal, general engineering, mining and textiles.

The data

Male lung cancer mortality rate has been standardised over the age 35-74 and over a 2 year period, 1968-1969. The data were provided by the French National Institute for Health and Medical Research (INSERM) at the scale of the French *départements*. Cigarettes sales data were compiled by the French Nationalised Tobacco Company (SEITA). To take into account the time lag between smoking and the onset of a lung pathology, cigarettes sales per inhabitant were recorded in 1953. Demographic data on the percentage of employed males in the metal industry, in general engineering, in the textile industry and in mining were taken from the 1962 census (INSEE). After the grouping of the *départements* around Paris into one area and the exclusion of 4 others owing to the poor quality of the data, N = 82 locations were retained.

Results for the modified $t_{\hat{M}-2}$ test

Separate results for the 4 industrial branches are presented in Table 3. Simple correlations, partial correlations after adjustment on smoking or after both adjustment on smoking and a linear gradient were tested by the standard methods and using the modified $t_{\hat{M}-2}$ test.

Generally one can see that by including more explanatory variables the autocorrelation in the residuals decreases and so the effective sample size $\hat{M}$ increases. As expected, the difference between the standard and the $t_{\hat{M}-2}$ tests, which can be substantial, is more marked where there is more autocorrelation.

The results obtained are quite different for the 4 industries. For metal and general engineering industry, there is a statistically significant association with male lung mortality rate after adjustment on cigarette sales both with and without the removal of a linear gradient. For mining, the significant association after adjustment on smoking is not observed after the removal of a linear gradient. For the textile industry, no significant association is found before adjustment with the modified $t_{\hat{M}-2}$ statistic whilst this would not be the case if only standard test were

carried out. After adjustment on both cigarette sales and a linear gradient, the association is also clearly non significant. Note that in all the examples except mining the adjustment solely on cigarette sales leads to slightly larger $t_{\hat{M}-2}$ statistic.

-Table 3-

In Table 3, the $t_{\hat{M}-2}$ statistic is calculated using the 15 isotropic strata defined in §3. To further study the influence of the number of strata we have calculated as an example values of the $t_{\hat{M}-2}$ statistic for metal industry adjusted on cigarette sales (2nd line in Table 3) for 9 isotropic partitions ranging from 5 classes to 21 classes. The values of $t_{\hat{M}-2}$ varied little, from 3.61 for 5 classes to 3.34 for 21 classes ; similarly the effective sample size $\hat{M}$ varied only from 37 (5 classes) to 32 (21 classes). When these calculations were carried out for the other examples there is again little change in the results. This is in agreement with the simulations described earlier.

In a further analysis, it was thought interesting to also include other industrial factors in the adjustments. We chose the metal industry and general engineering for this purpose since they are the only two factors clearly significantly related to lung cancer. The results are presented inTable 4. Taking metal industry, with an adjustment on general engineering, or taking general engineering, with an adjustment on metal industry, did not overall modify the results found in Table 3 and the relationships are confirmed. It is interesting to note that for mining, adjustment on both industries had a similar effect to that of adjusting for a gradient with the association with lung cancer mortality becoming non significant and the residual autocorrelation disappearing. For the textile industry, the adjustment for both industries gives a similar result to that for cigarette sales with no clear evidence of any link with lung cancer mortality.

-Table 4-

The inclusion of a linear gradient, whilst strengthening the interpretation of the association of lung cancer mortality with metal or general enginnering industry has on the contrary, weakened the observed association with mining or textile. We now discuss in some detail the problems connected with the existence of gradients.

Gradient influence

The adjustment of gradients always involves some degree of arbitrariness in the definition of the appropriate trend. Furthermore, the observed variations in the residuals may be of a smaller order of magnitude than the original variations thus leading to less precision in the estimates. Hence, one needs to be cautious when interpretating a non significant result after de-trending.

We have investigated the effect of omitting to adjust a gradient on the performance of the standard and modified t_{M-2} test of the correlation coefficient. Simple models were chosen for two variables X_α and Y_α, $\alpha \varepsilon A$. The centred and standardised coordinates of a point α in A will be denoted by i_α and j_α. Define :

$$X_\alpha = U_\alpha + a i_\alpha + b j_\alpha \quad \text{and} \quad Y_\alpha = V_\alpha + a' i_\alpha + b' j_\alpha \qquad (10)$$

with U_α and V_α mutually independent variables, $U_\alpha \sim N(0, \Sigma_U)$, $V_\alpha \sim N(0, \Sigma_V)$, Σ_U and Σ_V generated by a disc model with equal parameter $\rho_{(1)}$. X_α and Y_α defined by (10) will not be stationary and the autocorrelation between 2 points will depend on their coordinates.

The effect of the linear gradient on the expectation of the covariance between X and Y leads to a term proportional to aa' + bb' involving Σi_α^2 and Σj_α^2, corrected by a term involving $\Sigma i_\alpha j_\alpha$. The correction term is small since our centered and standardised coodinates are nearly orthogonal. Hence the effect is mainly related to aa' + bb' which is maximum when a' = a , b' = b and minimum when a' = -b ; b' = a. The effect of the gradient on the correlation coefficient is more complicated since a,a',b,b' are also involved in the denominator and a first order approximation for the expectation of r_{XY} can be calculated. For the sake of

simplicity in the following discussion we relate the results to aa' + bb'. Values of the first order approximation are indicated in brackets in Table 5. Five hundred simulations were carried out for different choices of aa' + bb' (keeping a + b = a' + b'), 3 levels of autocorrelation and a nominal 5 % significance level. The results are presented in Table 5.

-Table 5-

For the standard test and $\rho_{(1)} = 0$, type I errors are clearly increasing with aa' + bb'. The situation is similar when some autocorrelation is present ($\rho_{(1)} = 0.4$) with overall higher type I errors. When the autocorrelation is high ($\rho_{(1)} = 0.8$) the gradient influence on the overall type I error decreases in regard to the autocorrelation effect.

The modified $t_{\hat{M}-2}$ test is based on the estimated autocovariances $\{\hat{C}_X(k)\}$ and $\{\hat{C}_Y(k)\}$ even in the case $\rho_{(1)} = 0$. Due to the linear trend, these autocovariance functions are quite regular and reach moderately large negative values for high k. From formula (2), it is easy to see that the adjustment in the degrees of freedom for $t_{\hat{M}-2}$ is strongest when $\{\hat{C}_X(k)\}$ and $\{\hat{C}_Y(k)\}$ are quite similar, whether positive or negative. This happens clearly in the extreme case a' = a , b'= b of parallel gradients ; but also to a slightly lesser extent in the case aa' + bb' = 0 since values of the variable $\{Y_\alpha\}$ would be equal to those of $\{X_\alpha\}$ if the set of coordinates were rotationaly invariant by $\pi/2$ which is nearly the case in France. This explains that for aa' + bb' = 0, the d.f. are overadjusted and the statistic $t_{\hat{M}-2}$ is conservative. For the other values of aa' + bb', the correction of the d.f. reduces the type I error in comparison to that of the standard test but nevertheless it does not always stay close to 5 %. When $\rho_{(1)} = 0.4$ or 0.8, the intrinsic autocorrelation prevails more and more on the gradient so that irrespective of the gradient, all the type I errors are around 5 % for $\rho_{(1)} = 0.8$.

In conclusion it is worth calculating $t_{\hat{M}-2}$ both before and after de-trending. Even in the presence of a gradient, $t_{\hat{M}-2}$ may realise a satisfactory adjustment if the intrinsic autocorrelation in the variables is strong.

Concluding remarks

In this paper, we have outlined a simple test for partial correlations between spatially autocorrelated variables which is particularly suitable for ecological correlation studies. The performance of this test was shown to be satisfactory by Monte Carlo simulations. The application of this method to data sets is straightforward and only requires simple computing. The examples analysed concern the relationship between male lung cancer mortality rate and some industrial factors. They were chosen in order to illustrate that geographical epidemiology studies, when analysed with concern for the spatial structure, can identify risk factors which are related to those found in individual epidemiological studies and that this can be done quite easily with the modified test proposed. It would be interesting to compare our results with those of a generalised regression with spatially modelled error structure.

At the level of individual epidemiological studies, working in the metal industry, ship building and motor vehicule construction has been recognised to present a carcinogenic risk for lung cancer due to possible exposure to arsenic, chromium, benzo(a) pyrene, nickel or asbestos. Concerning the mining industry, the association is specifically reported for arsenic, iron-ore, asbestos or uranium mining but not for coal mining. Finally, there is no reported relationship between lung cancer and the textile industry[29]. Hence, the associations observed in our results using the modified $t_{\hat{M}-2}$ statistic with the metal industry and general engineering are particularly interesting as well as the non-association with the textile industry. Using standard methods would lead somewhat to misleading results concerning the textile industry. For the mining industry, the association is bordeline as it can be accounted for by a simple spatial gradient structure or by other industrial factors. As the epidemiological evidence does not involve all types of mining, a more detailed indicator of exposure would need to be analysed.

It is also interesting to note that part of the autocorrelation in the dependent variable is explained by the inclusion of confounding factors in the model. In this light, the gradient can be thought of as a proxy confounding variable. Finally one has to recall that as discussed in Greenland[15] confounding at the geographical level by a covariate can occur under broader conditions than for individual epidemiological studies and that geographic adjustment might be insufficient to control for it.

Acknowledgements

The author wishes to thank Chantal Guihenneuc and Virginie Lasserre of the Laboratoire de Statistiques Médicales, Université de Paris V for computing assistance and Evelyne Przybilski for secretarial assistance. This work was supported by a EURATOM contract n° BI6-126F.

REFERENCES

(1) Doll, R. The epidemiology of cancer. Cancer, 45,2475-85 (1980).

(2) Doll, R. ed. The geography of disease. British Medical Bulletin. Published for the British Council by Churchill Livingstone (1984).

(3) Hutt, M.S. and Burkitt, D.P. The geography of non-infectious disease. Oxford University Press Oxford (1986)

(4) Armstrong, B., Doll, R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer, 15, 617-631 (1975).

(5) Blot, W.J. and Fraumeni, J.F. Geographic patterns of oral cancer in the United States : etiological implications. J Chron Dis, 30, 745-757 (1977)

(6) Graham, S. Methodological problems in ecologic studies of the asbestos cancer relationship. Envir Res, 25, 35-49 (1981)

(7) Cantor, K.P., Hoover, R., Mason, T.J, Mc Cabe, L.J. Associations of cancer mortality with halomethanes in drinking water. J Natl Cancer Inst, 61, 979-985 (1978).

(8) Cliff, A.D. and Ord, J.K. Spatial Processes. Models and Applications. London, Pion (1981).

(9) Doreian, P. "Estimating linear models with spatially distributed data", in Sociological Methodology, edited by Leinhardt and Samuel. San Francisco, Jossey-Bass Publishers, 359-388 (1981).

(10) Breslow, N.E. and Enstrom, J.E. Geographic correlations between cancer cancer mortality rates and alcohol-tobacco consumption in the United States. J Natl Canc Inst, 53, 631-639 (1974).

(11) Stravraky, K.M. The role of ecologic anaysis in studies of the etiology of disease : a discussion with reference to large bowel cancer. J Chron Dis, 29, 435-444 (1976).

(12) Davies, J.M. and Chilvers, C. The study of mortality variations in small administrative areas of England and Wales, with special reference to cancer. J Epidem C, 34, 87-92 (1980).

(13) Morgenstern, H. Uses of ecological analysis in epidemiologic research. Am J Publ Health, 72, 1336-44 (1982).

(14) Richardson, S., Stücker, I., Hémon, D. Comparison of relative risks obtained in ecological and individual studies : some methodological considerations. Int J Epidemiol, 16, 111-120 (1987).

(15) Greenland, S., Morgenstern, H. Ecological bias, confounding, and effect modification. Int J Epidemiol, 18, 269-274 (1989).

(16) Pocock, S.J., Cook, D.G. and Shaaper, A.G. Analysing geographic variation in cardiovascular mortality : methods and results. J.R.Statist. Soc A, 145, 313-341 (1982).

(17) Clifford, P., Richardson, S. and Hémon, D. "Assessing the significance of the correlation between two spatial processes" Biometrics, 45(1), 123-134 (1989).

(18) Bartlett, M.S. Some aspects of the time-correlation problem in regard to tests of significance. J. R. Statist. Soc., 98, 536-543 (1935).

(19) Johnston, J. Econometric Methods, 2nd edition. New York, McGrawHill (1972).

(20) Ord, K. Estimation methods for models of spatial interaction. J. Am. Statist. Ass, 70, 120-126 (1975).

(21) Mardia, K.V. and Marshall, R.J. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, 71, 135-146 (1984)

(22) Hepple, L.W. Bayesian analysis of the linear model with spatial dependence, in Exploratory and Explanatory Statistical Ananlysis of Spatial Data pp 179-199 (Bartels, C.P.A and Ketellapper, R.H., Eds). Martinus Mijhoff : Boston (1979).

(23) Warnes, J.J. and Ripley, B.D. Problems with likelihood estimation of covariance functions of spatial Gaussian processes. Biometrika, 74(3), 640-642 (1987).

(24) Ripley, B.D. Statistical inference for spatial processes. Cambridge University Press, Cambridge (1988).

(25) Cook, D.G. and Pocock, S.J. Multiple regression in geographic mortality studies with allowance for spatially correlated errors. Biometrics, 39, 361-371 (1983).

(26) Ripley, B.D. Spatial statistics. New York, Wiley (1981).

(27) Pastorino, U., Berrino, F., Gervasio, A., Pesenti, V., Riboli, E., Crosignomi, P. Proportion of lung cancers due to occupational exposure. Int J Cancer, 33, 231-237 (1984).

(28) Benhamou, S., Benhamou, E., Flamant, R. Occupational risk factors of lung cancer in a French case-control study. Br J Ind Med 45, 231-233, (1988).

(29) Simonato, L., Vineis, P. and Fletcher, A. Estimates of the proportion of lung cancer attributable to occupational exposure. Carcinogenesis, 9, 1159-1165 (1988).

Table 1

Type I errors (per cent) of the $t_{\hat{M}-2}$ statistic for testing at 5 % and 1 % nominal rejection levels the partial correlation coefficient $r_{YZ.X}$ between $Z = aX + U$ and $Y = cX + W$ after adjusting on X.

5000 simulations are carried out for several levels of autocorrelation in X,U and W ($\rho_X(1)=\rho_U(1)=\rho_W(1)$), with $a = 2c = 0.204$. As reference, the empirical variance of $r_{YZ.X}$ and the type I errors (per cent) for the classical test of $r_{YZ.X}$ based on N = 82 observations are also indicated.

autocorrelation $\rho(1)$ for variables X,U and W	0.0	0.2	0.5	0.8
empirical variance of $r_{YZ.X}$	0.0128	0.0137	0.0176	0.0522
% type I errors for $t_{\hat{M}-2}$ [95 % CI] *5 % nominal level*	5.5 % [4.87% - 6.13%]	5.44 % [4.81% - 6.07%]	5.54 % [4.91% - 6.17%]	5.16 % [4.55% - 5.77%]
% type I errors for the t test * of $r_{YZ.X}$ based on N observations [95 % CI] *5 % nominal level*	5.56 % [4.93% - 6.19%]	6.5 % [5.81% - 7.18%]	10.48 % [9.63% -11.33%]	36.42 % [35.09% - 37.75%]
% type I errors for $t_{\hat{M}-2}$ [95 % CI] *1 % nominal level*	1.14 % [0.85% - 1.43%]	1.2 % [0.9% - 1.5%]	0.92 % [0.65% - 1.18%]	0.72 % [0.48%- 0.95%]
% type I error for the t test *of $r_{YZ.X}$ based on N observations [95 % CI] *1 % nominal level*	1.22 % [0.92% -1.52%]	1.76 % [1.40% - 2.10%]	3.06 % [2.58% - 3.53%]	22.96 % [21.79% - 24.13%]

* test of $\sqrt{N-3}\ r_{YZ.X} / (1 - r^2_{YZ.X})^{1/2}$ as a t_{N-3} distribution

Table 2

Power of the modified $t_{\hat{M}-2}$ statistic for testing at a 5 % nominal rejection level the partial correlation coefficient $r_{YZ.X}$ between $Z = aX + U$ and $Y = cX + dZ + W$ after adjusting on X.

500 simulations are carried out for several levels of autocorrelation in X,U and W ($\rho_X(1) = \rho_U(1) = \rho_W(1)$), with $a = 2c = 0.204$. N^* defined in (9) is an equivalent sample size base on v_e, the observed variance of $r_{YZ.X}$ and π is the power of a standard test of $r_{YZ.X}$ based on N^* observations.

$\rho_{YZ.X}$	$\rho(1)$	0.0	0.2	0.5	0.8
0.2	v_e	0.0111	0.0121	0.0156	0.0465
	N^*	85	78	61	22
	π	44. 8 %	40.4 %	33.3 %	10.8 %
	power of $t_{\hat{M}-2}$	43.4 %	41.7 %	34.6 %	13.6 %
0.4	v_e	0.0087	0.0094	0.0120	0.0372
	N^*	83	77	61	21
	π	96.4 %	95 %	89.1 %	41.7 %
	power of $t_{\hat{M}-2}$	96 %	95.4 %	91 %	46.8 %

Table 3

Comparison of the significance levels for tests of simple and partial correlations between male lung cancer mortality rates and industrial risk factors (adjusted on cigarette sales or on cigarette sales and a linear gradient) given by standard tests and $t_{\hat{M}-2}$ tests.

	r	standard test (N = 82) (values of the t_{N-2} statistic and significance level)	Modified $t_{\hat{M}-2}$ test (values of the $t_{\hat{M}-2}$ statistic and significance level)	$\hat{M}$
Metal industry workers				
no adjustment	0.63	7.16 $p < 10^{-9}$	3.00 p = 0.01	16
adjustment on cigarette sales	0.52	5.46 $p < 10^{-6}$	3.46 p = 0.002	34
adjustment on cigarettes sales and a linear gradient	0.35	3.34 p = 0.002	2.72 p = 0.01	55
General engineering workers				
no adjustment	0.43	4.23 $p < 10^{-4}$	2.72 p = 0.01	35
adjustment on cigarette sales	0.37	3.58 $p < 10^{-3}$	2.86 p = 0.007	53
adjustment on cigarettes sales and a linear gradient	0.26	2.41 p = 0.02	2.32 p = 0.02	76
Mine workers				
no adjustment	0.33	3.16 p = 0.003	2.37 p = 0.02	47
adjustment on cigarette sales	0.24	2.26 p = 0.03	2.42 p = 0.02	94
adjustment on cigarettes sales and a linear gradient	0.14	1.27 NS	1.26 NS	81
Textile industry workers				
no adjustment	0.28	2.57 p = 0.01	1.52 p = 0.14	30
adjustment on cigarette sales	0.26	2.40 p = 0.02	1.91 p = 0.07	53
adjustment on cigarettes sales and a linear gradient	0.11	1.02 NS	0.96 NS	73

Table 4

Significance levels of partial correlations between male lung cancer mortality rates and industrial factors after respective adjustments on metal industry and/or general engineering given by the standard tests and $t_{\hat{M}-2}$ tests.

	r	standard test (N = 82) (values of the t_{N-2} statistic and significance level)	Modified $t_{\hat{M}-2}$ test (values of the $t_{\hat{M}-2}$ statistic and significance level)	$\hat{M}$
Metal industry workers				
adjustment on general engineering	0.58	6.33 $p < 10^{-7}$	3.32 $p = 0.004$	24
adjustment on cigarette sales, general engineering and a linear gradient	0.34	3.28 $p = 0.002$	2.62 $p = 0.02$	53
General engineering workers				
adjustment on metal industry	0.33	3.09 $p = 0.003$	2.81 $p = 0.007$	68
adjustment on cigarette sales, metal industry and a linear gradient	0.25	2.33 $p = 0.03$	2.20 $p = 0.04$	73
Mine workers				
adjustment on metal industry and general engineering	0.17	1.5 NS	1.54 NS	86
adjustment on cigarette sales, metal industry, general engineering and a linear gradient	0.11	0.96 NS	0.98 NS	86
Textile industry				
adjustment on metal industry and general engineering	0.23	2.07 $p = 0.05$	1.88 $p = 0.07$	68
adjustment on cigarette sales, metal industry, general engineering and a linear gradient	0.18	1.63 NS	1.51 NS	70

Table 5

Proportion of type I errors for the standard test and the $t_{\hat{M}-2}$ test of the correlation coefficient between the variables $X_\alpha = U_\alpha + ai_\alpha + bj_\alpha$ and $Y_\alpha = V_\alpha + a'i_\alpha + b'j_\alpha$ where U_α and V_α are mutually independent spatially autocorrelated disc processes with equal parameter $\rho(1)$.

500 simulations are carried out in each case. The nominal rejection levels of the standard and $t_{\hat{M}-2}$ test chosen is 5 %.

Linear gradient slopes for Y

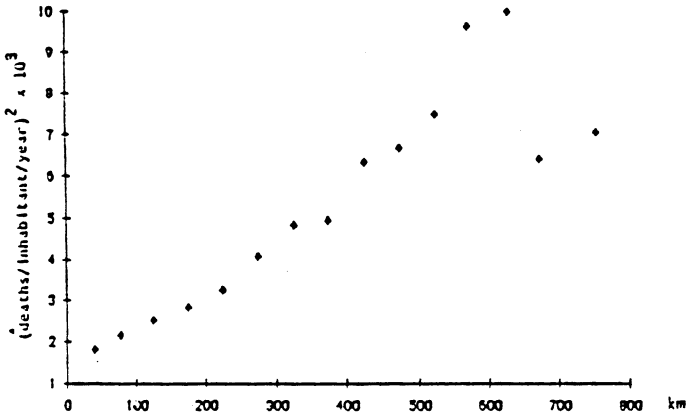
[the linear gradient slopes for X are fixed with a=0.1 and b= 0.4]

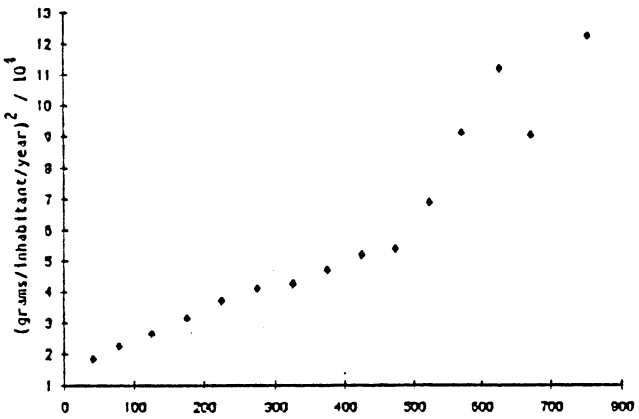
% type I errors		a' = 0.4 b' = -0.1 aa'+bb' = 0 (0.00)*	a' = 0.35 b' = 0.05 aa'+bb' = 0.075 (0.043)*	a' = 0.3 b' = 0.2 aa'+bb' = 0.11 (0.091)*	a' = 0.2 b' = 0.3 aa'+bb' =0.14 (0.119)*	a' = 0.1 b' = 0.4 aa'+bb'=0.17 (0.143)*
$\rho(1) = 0$	standard t	4.6 %	6.8 %	13.2 %	16 %	26.4 %
	$t_{\hat{M}-2}$	1.8 %	2.8 %	6.2 %	7.8 %	14 %
$\rho(1) = 0.4$	standard t	10.4 %	12.2 %	16.8 %	21.8 %	28.6 %
	$t_{\hat{M}-2}$	3.8 %	4.8 %	5.2 %	8.6 %	11 %
$\rho(1) = 0.8$	standard t	35.4 %	46.2 %	43.6 %	50 %	46%
	$t_{\hat{M}-2}$	2.6 %	4 %	5.2 %	5.2. %	5 %

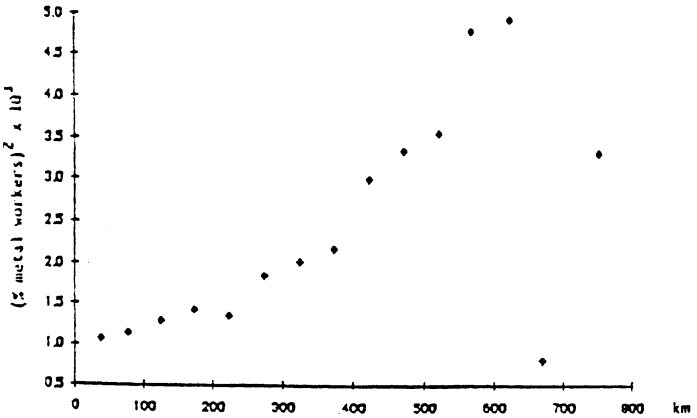
* first order approximation for the expected value of r_{XY}

Figure 1

Variograms of lung cancer mortality rate, cigarette sales per inhabitant and proportion of workers in the metal industry







Fifteen classes of distance for the plot of $N_k^{-1} \Sigma(X_\alpha - X_\gamma)^2$ against the average distance between locations in S_k are considered.

The number N_k of pairs in each class is : 82, 400, 582, 674, 764, 822, 812, 726, 630, 476, 304. 178, 94, 58, 40.

COMPARISON OF RELATIVE RISKS OBTAINED IN ECOLOGICAL AND INDIVIDUAL STUDIES : SOME METHODOLOGICAL CONSIDERATIONS

International Journal of Epidemiology
© International Epidemiological Association 1987

Vol 16, No. 1
Printed in Great Britain

Comparison of Relative Risks Obtained in Ecological and Individual Studies: Some Methodological Considerations

SYLVIA RICHARDSON*†, ISABELLE STÜCKER* AND DENIS HÉMON*

Richardson S (INSERM, U170, 16 av. Paul Vaillant-Couturier, 94807 Villejuif Cedex, France), Stücker I and Hémon D. Comparison of relative risks obtained in ecological and individual studies: some methodological considerations. *International Journal of Epidemiology* 1987, **16**: 111–120.

This paper is concerned with the problem of estimating relative risks from ecological correlation studies. In the first part, some of the biases encountered when analysing aggregated data are discussed and in particular attention is focused on the shape of the dose-response relationship obtained from aggregated and non-aggregated data, on the need for extrapolation and on the scale of aggregation. In the second part some empirical observations are made on these points by means of four examples concerning the relative risk between smoking and different pathologies. The estimates of relative risks derived from French geographical data and from case control or cohort studies are compared. The performance of ecological studies is discussed with respect to the strength of the risk factor considered, the geographical distribution of counfounding factors and the adjustment of different models.

Studies of the geographical distribution of incidence or mortality rates for particular diseases have been widely used by epidemiologists in formulating hypotheses about the aetiology of those diseases.[1–5] These studies, where the variables concerned are averages measured on groups of people, are often referred to as 'ecological correlation studies'. How to interpret and to assess the statistical significance of an observed, ecological correlation has been discussed by many authors.[6–13] Our aim is to compare the quantification of the dose-response relationship obtained by geographical studies to that assessed at the individual level, an approach first used by Béral.[14]

ESTIMATION OF RELATIVE RISK FROM ECOLOGICAL DATA: THEORETICAL CONSIDERATIONS

Comparison of the Shape of the Dose-Response Relationship between Aggregated and Non-Aggregated Data

Consider the relation:

$$f(x) = P(D \mid X = x) \qquad (1)$$

which models the probability of contracting the disease D if an individual is exposed to the risk factor X taking the value x. Suppose that inside a group G the distribution of X has mean μ_G and that it is only possible to observe the relationship between μ_G and the mean level of the disease in the group Y_G:

$$Y_G = g(\mu_G) = E_X [P(D \mid X)]. \qquad (2)$$

It is assumed here that $E_X [P(D \mid X)]$ can be indexed solely by μ_G. Let us compare in some examples estimations of the relative risk (RR) derived from (1) and (2).

(a) Firstly let us consider the case where f is linear,

$$f(x) = \alpha + \beta x \qquad (3)$$

then clearly g is also linear with the same coefficients

$$Y_G = \alpha + \beta \mu_G. \qquad (4)$$

As suggested by Béral,[14] (4) can be used to estimate the relative risk $RR(x_0)$ for an average exposure $\mu_G = x_0$:

$$RR(x_0) = (\beta/\alpha)\, x_0 + 1.$$

$RR(x_0)$ can therefore be equally estimated from equation (3) or (4), that is from individual or ecological studies. So in the linear case with one risk factor there is no mathematical bias resulting from using aggregated data.

(b) Secondly if f is convex, then using Jensen's inequality, it follows that

$$Y_G = g(\mu_G) = E_X(f(X)) \geq f(\mu_G).$$

In particular in the classical exponential case,

$$f(x) = c\, e^{\gamma x}, \gamma > 0 \qquad (6).$$

Then we have

$$Y_G = g(\mu_G) = E_X(ce^{\gamma x}) = c \exp(\gamma \mu_G) \times b(\gamma, X) \qquad (7)$$

* INSERM U170 Unité de Recherches Epidémiologiques et Statistiques sur l'Environnement et la Santé, 16 av. Paul Vaillant-Couturier, 94807 Villejuif Cedex, France.
† Laboratoire de Statistiques Médicales, Université de Paris V 45 rue des Saints Pères, 75006 Paris, France.

where the bias coefficient $b(\gamma, X)$ can be expressed in terms of the cumulants of the distribution of X and will in general depend on γ, μ_G and the parameters of f.[15] Consequently (7) will lead to a biased estimate of $RR(x_0)$ unless $b(\gamma, X)$ cancels from the numerator and the denominator of $RR(x_0)$. This is the case for instance when X follows a normal distribution (μ_G, σ_μ^2) under the condition that σ_μ^2 does not vary over the range of μ, a condition which would rarely hold.

In general the bias in the estimate of the relative risk will be more pronounced the larger the values of μ, the wider its range and the more pronounced the convexity of f.

Average response at zero exposure

In most cases, the range of the distribution of the ecological risk factor will not include zero. An extrapolation towards a non-exposed group will be necessary in order to compute an ecological relative risk. This extrapolation is subject to errors and even if two models are indistinguishable in terms of fit on the range of the ecological risk factor, they might lead to quite different risk estimates at the non-exposed level. The problem of extrapolation can also occur in individual studies where only a few cases have low exposure. In both individual and ecological studies information about relative risks can still be derived by considering non-zero reference exposure within the range of the risk factor.

Scale of aggregation

The geographical units on which the ecological analysis is based are divisible. Hence ecological studies can often be performed at several levels which may not give identical results. Checking the stability of the results in relation to the geographical scale of the analysis is important for their interpretation.[16–22]

Assuming that the variables can be considered at two levels, either ungrouped or averaged over groups (symbolized by G), the total variation of a variable (var (X)) can be decomposed into the average variation A = E[var(X | G)] within each group and the variation B = var [E(X | G)] of the mean between groups:

var (X) = A+B = E [var (X | G)] + var [E (X | G)].

Similarly for the covariance:

cov (X,Y) = A′+B′ = E [cov (X, Y | G)] + cov [E (X | G), E (Y | G)].

When the slope of the individual association of X and Y does not vary from one area to another, several situations can occur:

(a) $A'/A = B'/B$ (Figure 1a). In this case the slope of the between group variation of X and Y is equal to the within groups one. Hence $(A'+B')/(A+B) = A'/A$ and so the evaluation of the slope is the same using total or intra-group variation.

(b) $A'/A \neq B'/B$ (Figures 1b and c). In this case there is an 'ecological cross level bias' caused by the presence of a confounding factor which alters the intra-group variation of X and Y either by amplifying it (Figure 1b) or by reducing it (Figure 1c). This confounding factor can influence the analysis either at the ungrouped or at the grouped level.

The same reasoning applies to situations where small areas can be grouped into larger regions. When the slope of the relationship changes from one level to another, one may consider this as the result of a confounding factor which varies with the level of the analysis.

ESTIMATING RELATIVE RISKS FROM ECOLOGICAL DATA: SOME EXAMPLES

In the following examples French ecological relative risks between cigarette consumption and lung,

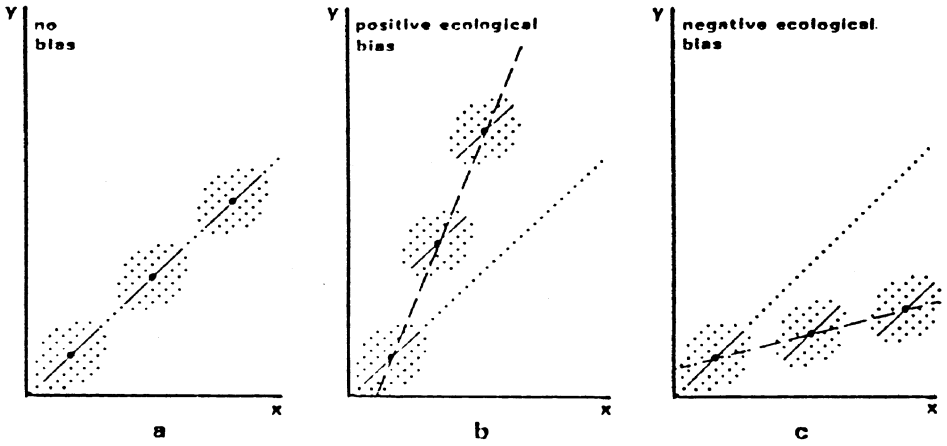


FIGURE 1 *Aggregation and ecological bias.*

bladder, oesophagus cancer and coronary heart disease (CHD) mortality are compared with those estimated in individual studies. Confidence intervals for the relative risk are calculated by Fieller's method.[23]

Mortality Rates

Throughout we consider the mortality in France for two time periods 1968–1969 and 1974–1976 corresponding to national census. The mortality rates have been standardized by the direct method using ten-year age groups from 35–44 to 65–74 and the 1968 population. From the 86 administrative areas (départements) for which both cigarette consumption and mortality rates were available for those dates, four were excluded from the analyses because the proportion of deaths due to unspecified causes was too large.[24]

Cigarette Consumption

Damiani and Massé have estimated tobacco consumption per age and per sex for the 22 French regions[25,26] under the assumption that individual and ecological relative risks are equal. Their estimates cannot therfore be used in the present study which aims at discussing that assumption.

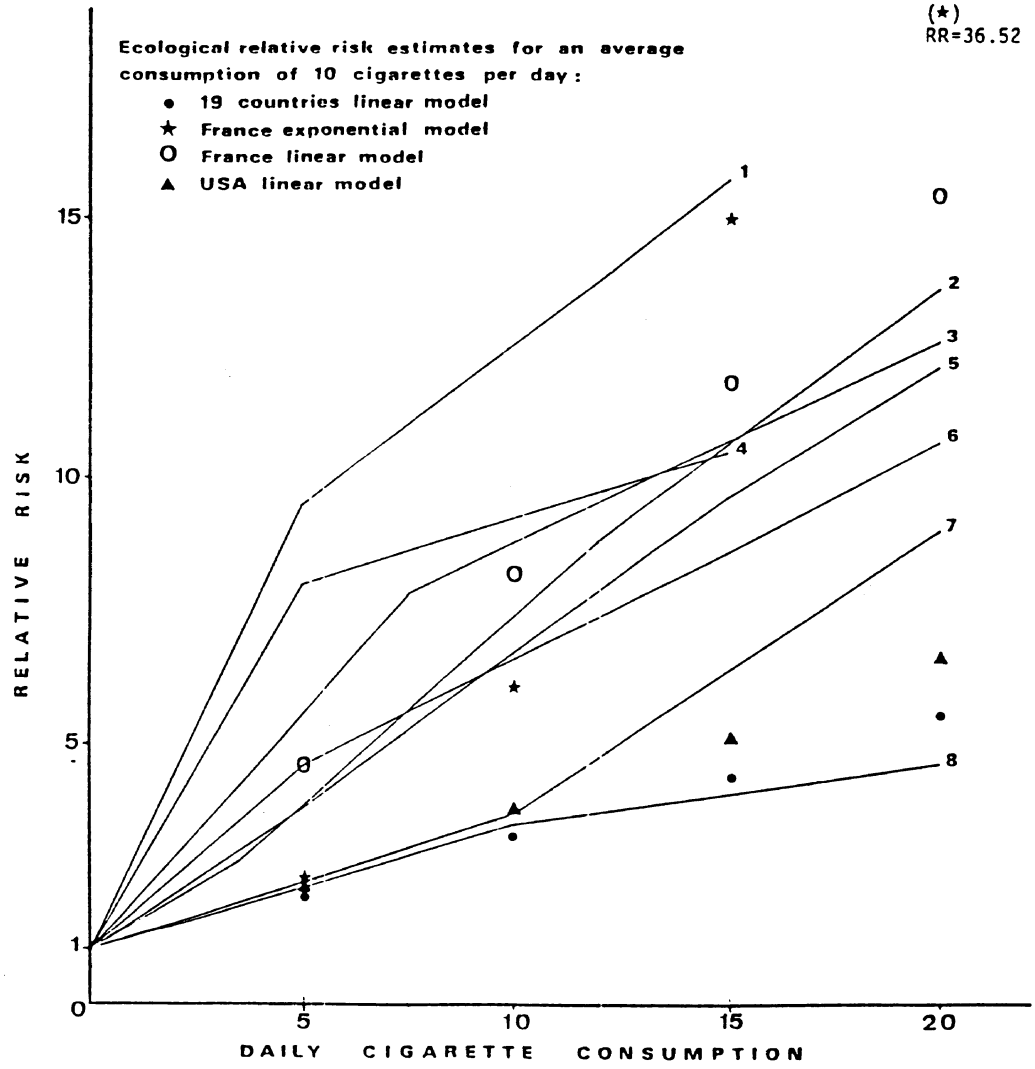


FIGURE 2 *Relative risk for lung cancer as a function of current number of cigarettes smoked per day estimated in eight cohort studies and in ecological studies.*
Cohort studies as quoted in:[29]

1 Best et al., *1961*
2 Cederlof et al., *1975*
3 Doll et al., *1972*
4 Hammond and Horn, 1958
5 Dorn et al., *1974*
6 Hammond et al., *1976*
7 Weir et al., *1970*
8 Hirayama, 1977

It is necessary to consider an average time lag of 15 years between cigarette sales and mortality rates. The number of cigarettes sold per inhabitant for each French 'département' was recorded in 1953 and 1957 by the French nationalized tobacco company.[27] To correlate mortality rates in the age range 35–74 to cigarette sales which occurred 15 years before, we assessed retrospectively the proportion of the total cigarette sales (in 1953 or 1957) which is attributable solely to men aged 20–59. We were able to estimate this proportion using the results of a large survey carried out in 1960 on smoking habits in France.[28] These estimates will be used throughout the rest of the study and referred to as cigarette consumption for men.

Lung Cancer and Cigarette Consumption

In Figure 2, are represented relative risks for lung cancer mortality (ICD 162) among men as a function of current consumption of cigarettes per day estimated in eight major cohort studies.[29] The relative risks corresponding to 5, 10, 15 or 20 cigarettes per day estimated geographically in France by a linear model between the standardized mortality rate for lung cancer in men in 1968 and cigarette consumption for men in 1953 are also shown in this figure.

The values corresponding to daily consumptions of 15 or 20 cigarettes are far outside the range of the average consumption in French 'départements', which is 2 to 10 cigarettes per day (Figure 3). The comparison between cohort estimates and ecological estimates should only be judged on the 2–10 cigarettes per day range where there is a good agreement with a linear model fitted to the French data.

In this figure are also represented the same relative risks estimated by a linear model from international data and from US data. Correlation between national

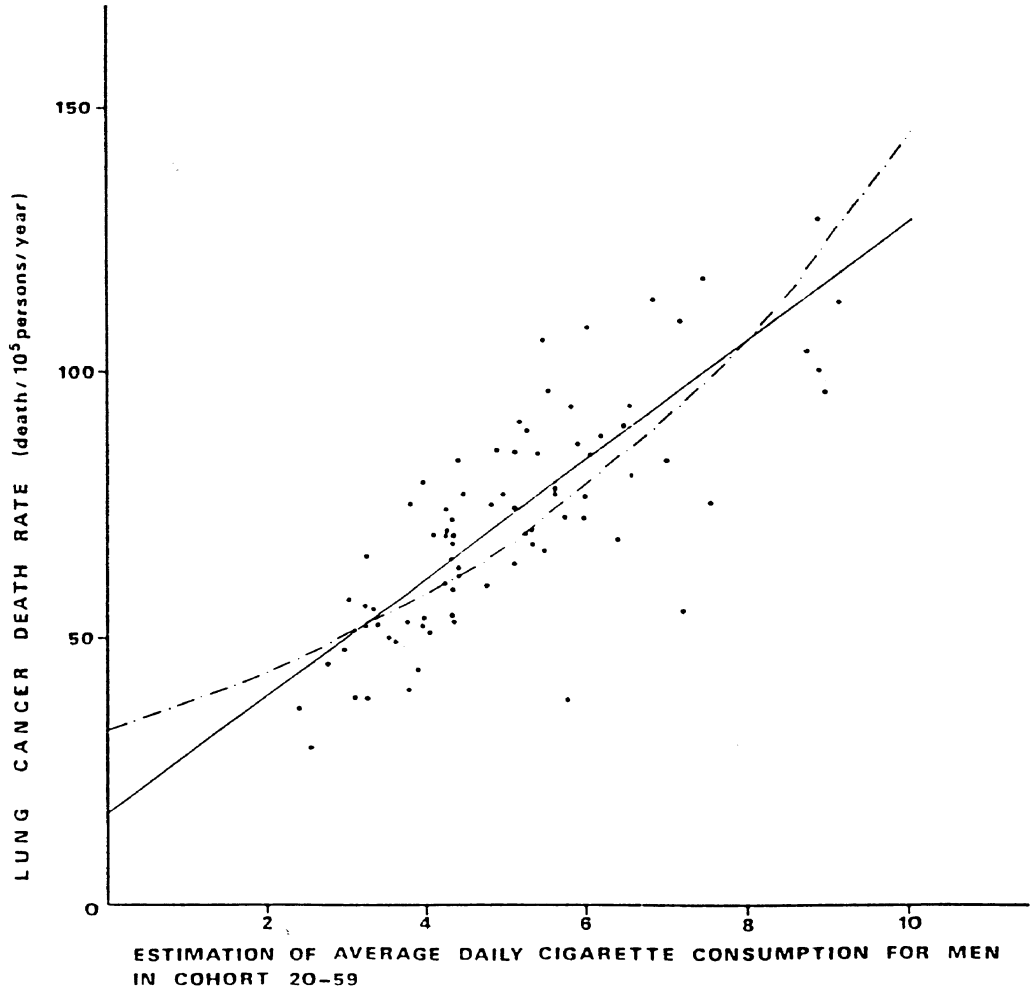


FIGURE 3 *Lung cancer mortality and average cigarette consumption per day in 82 French départements: fitting of linear (——) or exponential (–·–) model.*

lung cancer mortality and cigarette sales has been discussed by many authors.[1,24,30] Data on lung cancer mortality in 1975 or 1976 for 19 countries (age-standardized over the range 35–74 using the European population as reference) were taken from WHO statistics[31] and the corresponding national cigarette sales per adult (in 1957) from Lee.[32] The US data represented come from Fraumeni.[33] The relative risks estimated for the international and US data sets are much smaller than those given by the cohort studies or the French ecological data. This could be the consequence of the different ways of collecting data. Furthermore, US and international data on cigarettes sales and mortality rates were not restricted in this study to the most relevant age and sex groups thus resulting possibly in a dilution effect. In Table 1 estimates for different time lags and different models are given.

For the purpose of illustration, a consumption of 10 cigarettes per day was chosen, but the results are similar within the range 2–10 cigarettes per day. All correlation coefficients corresponding to this table are highly significant, and vary between 0.75 and 0.79. All ecological estimates given in Table 1 are within the range of the individual relative risks cited in the US Surgeon General's Report.[29] Further, estimates corresponding to 1953 and 1957 cigarette consumption are very close together. On the other hand, estimates based on 1968 death rates lie closer to the median value of relative risk obtained from individual studies than those for 1975 (Figure 2).

Similarly the estimates given by the linear model correspond better to cohort studies in the 2 to 10 cigarettes range than those given by the exponential model (Figure 2 and Table 1). This is understandable when looking at the fairly linear shape of the individual dose-response relationship in this range. Nevertheless there is a very good fit of both models on the French data (Figure 3). The exponential estimates of the relative risk are lower than the linear ones because the baseline mortality rates for non-smokers given by the two models are different.

Health and exposure variables often exhibit a regular component in their spatial variations analogous to a time trend which makes the interpretation of the joint variations difficult since many potential confounding variables might also show the same regular spatial gradient.[34] This suggests a further step in the analysis, the multiple regression of the health variable on the exposure variable and on trend components expressed in terms of the coordinates of the geographical units. A causal interpretation would be strengthened if a statistically significant correlation between exposure and health indicators is observed both with and without adjustment for a spatial trend. Table 2 shows that for both the linear and the exponential model, the regression coefficients between lung cancer mortality and cigarette consumption after the adjustment of a linear gradient are hardly modified.

Bladder Cancer and Cigarette Consumption

In Figure 4, the relative risk for daily cigarette consumption and bladder cancer mortality (ICD 188) estimated from the case-control studies of Wynder,[35] Cartwright,[36] Howe,[37] and Vineis[38] are represented. In contrast to Figure 2, only two studies clearly demonstrate a dose-response relationship. As for lung cancer the geographical estimates of the relative risk for the French data analysed and for the US data from Lee[32] are shown in Figure 4, but in this case the relative risk estimates based on French ecological data are higher than those of the individual studies.

In Table 3 we note that the relative risks do not vary between 1968 and 1975 and that the linear model gives only slightly larger estimates than the exponential one. Correlation coefficients corresponding to this table were all statistically significant and vary between 0.52 and 0.65.

In the case of bladder cancer we therefore observe some discrepancy between ecological estimates and

TABLE 1 *Lung cancer mortality and cigarette consumption: estimation of the relative risk corresponding to 10 cigarettes per day, from French ecological data*

	Linear model		Exponential model	
	1953[a]	1957[b]	1953[a]	1957[b]
1968 Death rate	8.2 (5.8 : 11.3)(*)	8.2 (5.7 : 11.6)(*)	6.0 (4.2 : 8.6)(*)	5.1 (3.6 : 7.2)(*)
1975 Death rate	5.3 (4.1 : 6.7)(*)	5.2 (4.0 : 6.6)(*)	4.7 (3.5 : 6.2)(*)	4.1 (3.1 : 5.4)(*)

(*) 95% confidence intervals.
Evaluation of cigarette consumption (a) in 1953, (b) in 1957.

TABLE 2 *Simple and partial correlations between lung cancer, bladder cancer and CHD mortalities and cigarette consumption after adjustment for a linear trend.*

Fitted model	Estimated slope	% variance explained	Significant(*) explanatory variables
Lung cancer			
$y = \alpha + \beta X + \varepsilon$	7.76×10^{-7}	57.3%	Cigarette consumption
$y = \alpha + \beta X + a_1U + a_2V + \varepsilon$	7.84×10^{-7}	67.9%	Cigarette consumption, latitude
$\log y = \alpha + \beta X + \varepsilon$	1.09×10^{-3}	55.3%	Cigarette consumption
$\log y = \alpha + \beta X + a_1U + a_2V + \varepsilon$	1.10×10^{-3}	63.7%	Cigarette consumption, latitude
Bladder cancer			
$y = \alpha + \beta X + \varepsilon$	1.07×10^{-7}	30.5%	Cigarette consumption
$y = \alpha + \beta X + a_1U + a_2V + \varepsilon$	0.71×10^{-7}	35.4%	Cigarette consumption, longitude
$\log y = \alpha + \beta X + \varepsilon$	0.97×10^{-3}	27.2%	Cigarette consumption
$\log y = \alpha + \beta X + a_1U + a_2V + \varepsilon$	0.61×10^{-3}	33.1%	Cigarette consumption, longitude
CHD			
$y = \alpha + \beta X + \varepsilon$	5.37×10^{-7}	22.1%	Cigarette consumption
$y = \alpha + \beta X + a_1U + a_2V + \varepsilon$	5.46×10^{-7}	32.9%	Cigarette consumption, latitude
$\log y = \alpha + \beta X + \varepsilon$	0.34×10^{-3}	21.1%	Cigarette consumption
$\log y = \alpha + \beta X + a_1U + a_2V + \varepsilon$	0.35×10^{-3}	31.2%	Cigarette consumption, latitude

y = standardized mortality (men) 1968. U: longitude.
x = cigarette consumption per inhabitant 1953. V: latitude.
(*) $p < 0.05$.

individual studies. A possible explanation may be found through failure to take into account ecological confounding factors which would be positively correlated with the distribution of cigarette consumption. Thus the slope of the regression between bladder cancer mortality and cigarette consumption would also include part of the effect of these confounding factors and lead to an overestimation of the relative risk. A

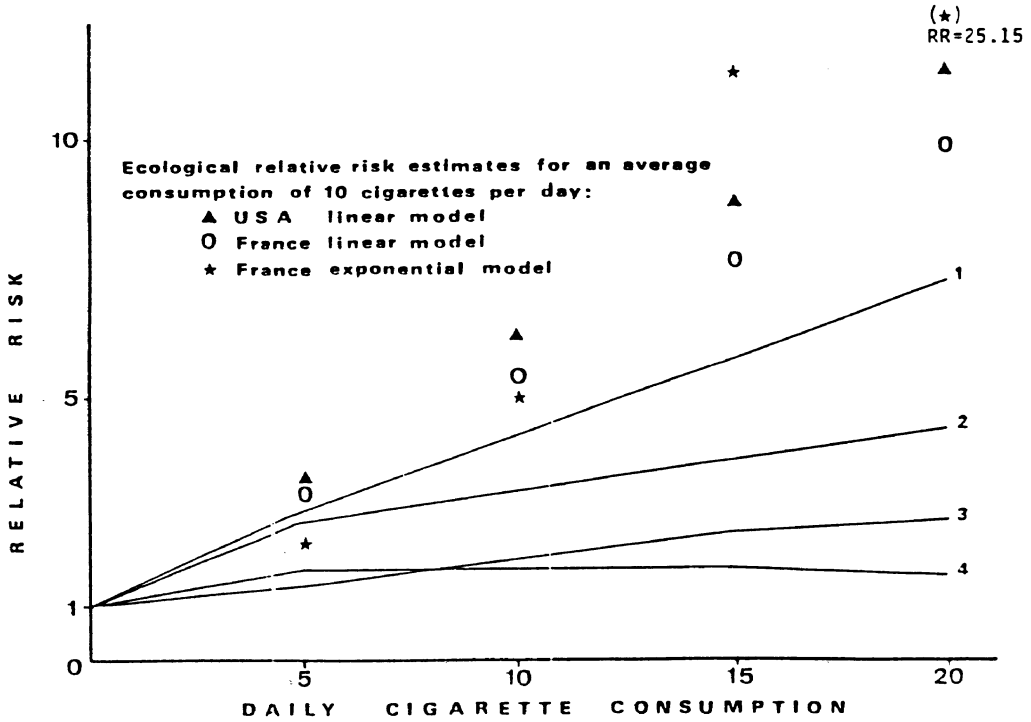


FIGURE 4 *Relative risk for bladder cancer as a function of current number of cigarettes smoked per day estimated in four case-control studies and in ecological studies.*
1 Vineis[38] — 2 Howe[37] — 3 Wynder[35] — 4 Cartwright[36]

good candidate for such a confounding factor is the exposure to some industrial risk factors.[39] Indeed, we know from consumer surveys[28] that cigarette consumption is higher among industrial workers. When a linear gradient is adjusted on ecological data (Table 2) the slope of the dose-response relationship varies more than for lung cancer, thus pointing to the likely influence of confounding factors having an heterogeneous spatial distribution.

Coronary Heart Disease and Cigarette Consumption

Figure 5 shows the relative risks for coronary heart disease (ICD 410–414) with respect to the current number of cigarettes smoked per day estimated in five cohort studies,[40] together with the ecological estimates for the French data analysed or for the US data taken from Friedman.[41] Estimates are given both for the standardized mortality and for the age group 45–54.

In Figure 5 one sees a good agreement between the individual and ecological studies and in Table 3 it is shown that the estimates for the 45–54 age group are higher than those for the 35–74 one. This tallies with the Framingham study[42] where the effect of cigarette consumption is more pronounced among middle-aged men than for those over 65. Just as for lung cancer the estimates for 1968 seem to give a better fit than those for 1975. All the correlation coefficients corresponding to Table 3 are statistically significant and similar for the two periods but are of a lower order of magnitude (from 0.35 to 0.47) than for lung or bladder cancer. The linear and exponential models give very close estimates due to a less pronounced convexity of the exponential curve. Adjustment of a linear trend does not modify the slopes of the dose-response relationships (Table 2).

TABLE 3 *Ecological relative risks (French data) estimates for an average consumption of 10 cigarettes per day*

Cause and year of death	Linear model	Exponential model
Bladder cancer 1968 (a)	5.4 (3.3 : 8.3)(*)	5.0 (2.8 : 8.9)(*)
Bladder cancer 1975 (b)	5.3 (3.6 : 7.6)(*)	4.6 (3.0 : 6.9)(*)
CHD 1968 (a)	1.8 (1.4 : 2.2)(*)	1.8 (1.4 : 2.2)(*)
CHD 1968: age (45–54) only (a)	2.4 (1.6 : 3.2)(*)	2.5 (1.7 : 3.7)(*)
CHD 1975 (b)	1.5 (1.2 : 1.8)(*)	1.5 (1.2 : 1.9)(*)
CHD 1975: age (45–54) only (b)	2.0 (1.5 : 2.6)(*)	2.0 (1.4 : 2.7)(*)

(*) 95% confidence interval.
Evaluation of cigarette consumption (a) in 1953, (b) in 1957.

Oesophagus Cancer and Cigarette Consumption

The consumption of alcoholic beverages and of tobacco products has long been suspected of increasing the risk of oesophagus cancer (ICD 150).[43–46] The ecol-

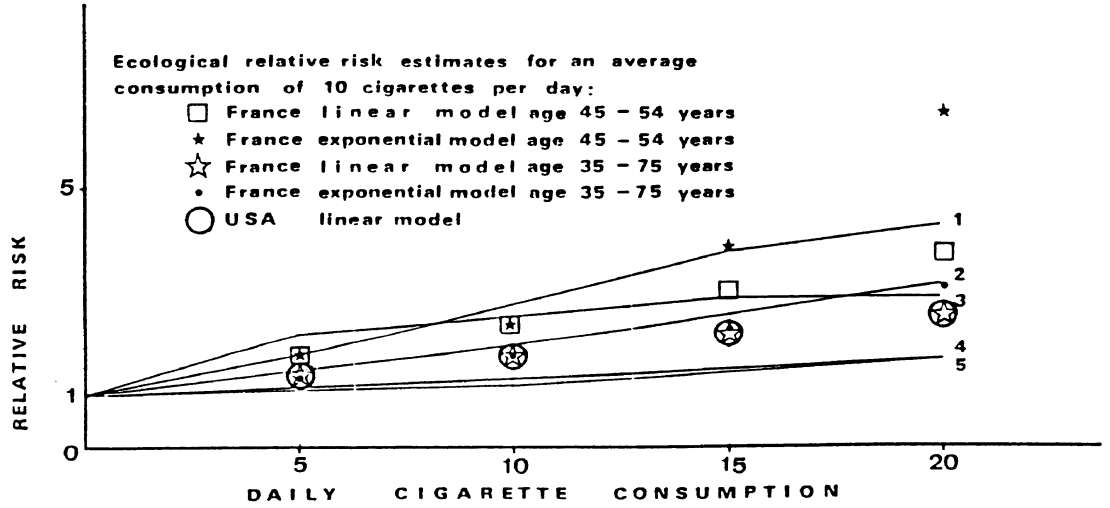


FIGURE 5 *Relative risk for coronary heart disease as a function of current number of cigarettes smoked per day estimated in five cohort studies and in ecological studies.*
Cohort studies as quoted in:[40]

1 Veterans Administration study
2 British Physicians study
3 American Cancer Society study
4 Framingham study
5 Pooling Project

ogical analysis presented concerns 1968 death rates and cigarette consumption for the year 1953. A grouping of the deaths from alcoholism (ICD 291, 303) and cirrhosis of the liver (ICD 571) was used as an indicator of alcoholic beverage consumption and these causes will be referred to as alcoholic mortality.

The simple correlation between standardized oesophagus cancer mortality for men and cigarette consumption is negative and statistically significant ($r = -0.28$) and the same result holds for the partial correlation coefficient when alcoholic mortality is taken into account, whether an additive model ($r = -0.27$) or a multiplicative one ($r = -0.28$) is fitted. Looking at Figure 6, a group of eight contiguous 'départements' situated in Normandy or Brittany clearly stands out by their very high standardized oesophagus mortality and their low cigarette consumption. When these eight 'départements' are excluded the correlation coefficients between oesophagus cancer mortality and cigarette consumption becomes non-significant for the additive and the multiplicative model.

Thus cigarette consumption does not appear to play a part in the variations by 'départements' of the male mortality for oesophagus cancer in France. This is not due to a parallelism in the geographical trends of alcoholic mortality and cigarette consumption since in France they display nearly independent variations at the 'département' level although they are correlated among individuals.

In a recent analysis of the geographical pathology of cancer of the oesophagus, Day[47] points out that there

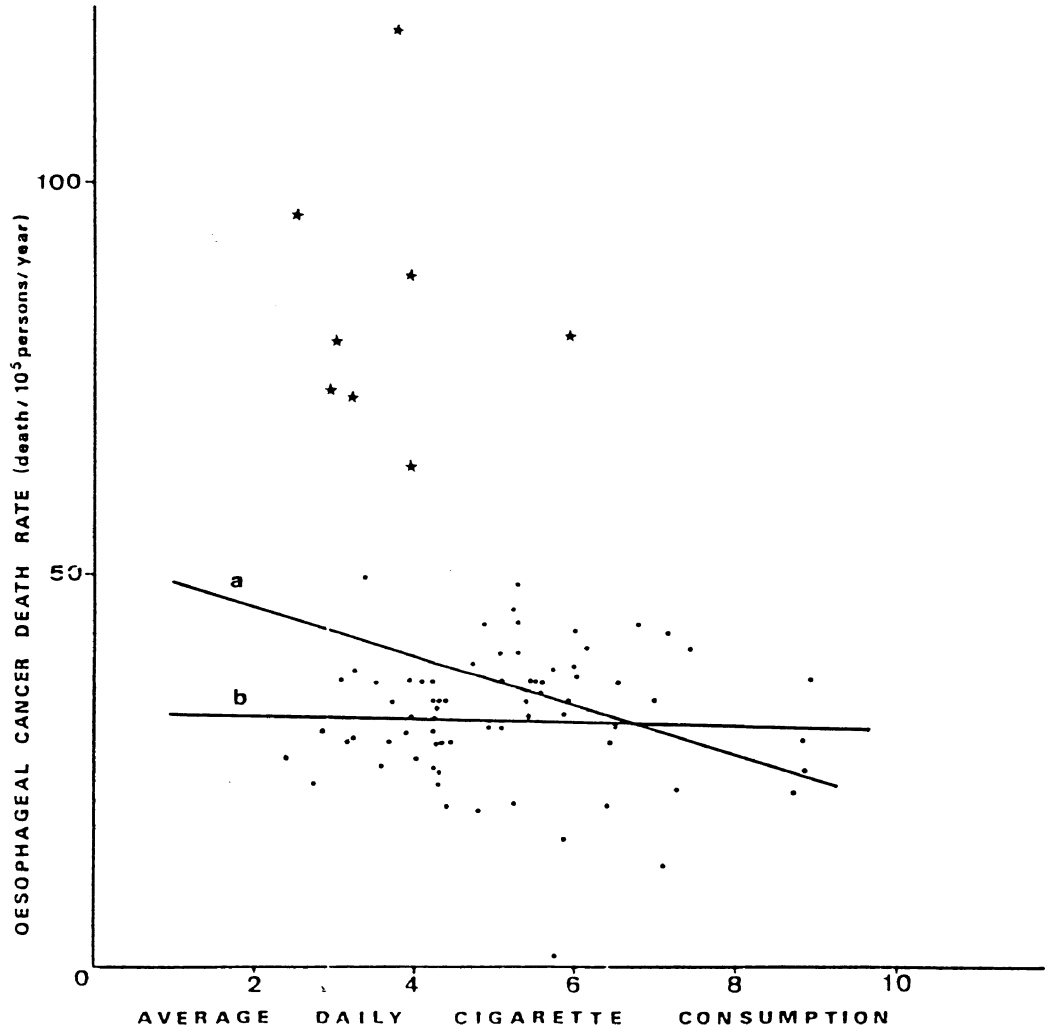


FIGURE 6 *Oesophagus cancer mortality and average daily cigarette consumption in 82 French départements: regression line (a) including all départements, (b) excluding eight contiguous départements * situated in the west of France.*

are some geographical contrasts, for instance within Normandy and Brittany, that cannot be explained solely in terms of variations of alcohol or cigarette consumption. It seems that even in industrial countries nutritional deficiencies which could interact with alcohol consumption might play an important part in the aetiology of the disease. Hence finding the precise role played by smoking is a difficult task, the more so as the dose-response relationship is weaker with respect to smoking than with respect to alcohol consumption.[48] In this context ecological analysis in France at the 'département' level does not contribute to our understanding of the aetiology of oesophagus cancer.

CONCLUSIONS

In this paper the biases that one encounters when trying to estimate a relative risk from an ecological study are discussed.

For smoking and lung cancer, a good match between individual and ecological relative risk estimates is obtained. In this case smoking is an overwhelming risk factor with a fairly linear dose-response relationship in the low and medium dose range. Furthermore, time lags were included and a careful evaluation of the relevant exposure was possible. Using a time lag is necessary but its magnitude is difficult to interpret since both mortality and consumption have been averaged over different age groups. This example also illustrates that the choice of the model has consequences on the estimates because of the frequent need for extrapolation to zero.

In the case of bladder cancer and cigarette consumption the ecological relative risk found is higher than the one given by individual studies. This example brings out the importance of considering the geographical distribution of confounding factors. This is also underlined by the modification of the regression coefficients when spatial trends are adjusted.

As in the case of bladder cancer smoking is not necessarily the dominant risk factor involved in CHD, but here the geographical estimates show a better agreement with the individual studies. The difference between the two examples probably stems from the different geographical distribution of possible confounding factors. Whilst cigarette consumption and industrial exposure are very likely to be positively correlated, this is presumably not the case for the other important risk factors for CHD such as diet. In this example of a multifactorial disease the closeness of the geographical and individual estimates of the relative risks might thus be partially coincidental.

The last example on smoking and oesophagus cancer illustrates that, in the case of a complex multifactorial disease, ecological analysis techniques are likely to fail to bring out aetiological factors when performed with routinely collected data at a regional level.

ACKNOWLEDGMENTS

This work was supported in part by the Radiation Protection Programme of the Commission of the European Communities, contract BIO-F-515-82-F.

The authors would like to thank Dr F Hatton, who kindly provided the mortality data used in the present study, Miss N Lincot for her technical assistance and Mrs M Guerrois for typing the manuscript.

REFERENCES

[1] Doll R. The epidemiology of cancer. *Cancer* 1980; **45:** 2475–85.

[2] Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int J Cancer* 1975; **15:** 617–31.

[3] Blot WJ, Mason TJ, Hoover R, Fraumeni JF. Cancer by county: etiologic implications. *Origins of human cancer* 1977.

[4] Graham S. Methodological problems in ecologic studies of the asbestos cancer relationship. *Envir Res* 1981; **25:** 35–49.

[5] Cantor KP, Hoover R, Mason TJ, McCabe LJ. Associations of cancer mortality with halomethanes in drinking water. *J Natl Canc Inst* 1978; **61:** 979–85.

[6] MacMahon B, Pugh TF. *Epidemiology: principles and methods.* Boston, Little Brown, 1970.

[7] Breslow NE, Enstrom JE. Geographic correlations between cancer mortality rates and alcohol-tobacco consumption in the United States. *J Natl Canc Inst* 1974; **53:** 631–9.

[8] Stravraky KM. The role of ecologic analysis in studies of the etiology of diseases: a discussion with reference to large bowel cancer. *J Chron Dis* 1976; **29:** 435–44.

[9] Davies JM, Chilvers C. The study of mortality variations in small administrative areas of England and Wales, with special reference to cancer. *J Epidemiol Community Health* 1980; **34:** 87–92.

[10] Richardson ST, Hémon D. On the variance of the sample correlation between two independent lattice processes. *J Appl Prob* 1981; **18:** 943–8.

[11] Cook DG, Pocock SJ. Multiple regression in geographic mortality studies with allowance for spatially correlated errors. *Biometrics* 1983; **39:** 361–71.

[12] Cliff AD, Ord JK. *Spatial processes, models and applications.* London, Pion, 1981.

[13] Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Publ Health* 1982; **72:** 1336–44.

[14] Béral V, Chilvers C, Fraser P. On the estimation of relative risk from vital statistical data. *J Epidemiol Community Health* 1979; **33:** 159–62.

[15] Kendall M, Stuart A. *The advanced theory of statistics*, vol 1, London, Charles Griffin, 1977.

[16] Robinson WS. Ecological correlations and the behaviour of individuals. *Am Sociol Rev* 1950; **15:** 351–57.

[17] Goodman LA. Some alternatives to ecological correlation. *Am J Sociol* 1959; **64:** 610–25.

[18] Thomas EN, Anderson DL. Additional comments on weighting values in correlation analysis of areal data. *Annals of the Association of American Geographers* 1965; **55:** 492–505.

[19] Alker HR. A typology of ecological fallacies. In: M. Dogan and S.

Rokkan (Eds) *Quantitative ecological analysis in the social sciences*, Cambridge, Ma, MIT Press, 1969.

[20] Clark WAV, Avery KL. The effects of data aggregation in statistical analysis. *Geographical Analysis* 1976; **8:** 428–33.

[21] Firebaugh G. A rule for inferring individual level relationship from aggregate data. *Am Sociol Rev* 1978; **43:** 557–72.

[22] Openshaw S. Ecological fallacies and the analysis of areal census data. *Envir Pl-A* 1984; **16:** 17–31.

[23] Finney DJ. *Statistical methods in biological assay*, London, Charles Griffin, 1952.

[24] Doll R, Peto R. Avoidable risks of cancer in the U.S. *J Natl Canc Inst* 1981; **66:** 1191–1308.

[25] Damiani P, Massé H. Mortalité par cause et tabac: application d'un modèle de liaison et évaluation de la consommation de tabac par sexe et par âge. *Journal de la Société de Statistique de Paris* 1980; **121:** 81–89.

[26] Damiani P, Massé H. Liaison de la mortalité par cause avec l'ensemble des consommations de tabac et d'alcool. *Journal de la Société de Statistique de Paris* 1981; **122:** 174–81.

[27] Atlas of tobacco consumption in France, SEITA 1953–1957.

[28] Etude du marché français du tabac, Enquête générale 1960, SEMA.

[29] U.S. Department of Health and Human Services. The health consequences of smoking. *Cancer: A report of the Surgeon General.* Department of Health and Human Services, Public Health Service, Office on Smoking and Health, 1982.

[30] Stocks P. Cancer mortality in relation to national consumption of cigarettes, solid fuel, tea and coffee. *Br J Canc* 1970; **24:** 215–25.

[31] World Health Statistics. Geneva, World Health Organization. 1979.

[32] Lee PN. *Tobacco consumption in various countries.* London, Tobacco Research Council, 1975.

[33] Fraumeni J. Cigarette smoking and cancers of the urinary tract: geographic variation in the United States. *J Natl Canc Inst* 1968; **41:** 1205–11.

[34] Lazar P. Geographical correlations between disease and environmental exposures. *Proceedings of the European Symposium on Medical Statistics*, Rome, 1980.

[35] Wynder EL, Goldsmith R. The epidemiology of bladder cancer. *Cancer* 1977; **40:** 1246–68.

[36] Cartwright RA, Adib R, Appleyard I *et al.* Cigarette smoking and bladder cancer: an epidemiological inquiry in West Yorkshire. *J Epidemiol Community Health* 1983; **37:** 256–63.

[37] Howe GR, Burch JD, Miller AB *et al.* Tobacco use, occupation, coffee, various nutrients, and bladder cancer. *J Natl Canc Inst* 1980; **64:** 701–13.

[38] Vineis P, Frea B, Uberti E, Ghisetti V, Terracini B. Bladder cancer and cigarette smoking in males: a case-control study. *Tumori* 1983; **69:** 17–22.

[39] Vineis P, Segnan N, Costa G, Terracini B. Evidence of a multiplicative effect between cigarette smoking and occupational exposures in the aetiology of bladder cancer. *Cancer Let* 1981; **14:** 285–90.

[40] Kannel WB, Framingham MD. Update on the role of cigarette smoking in coronary artery disease. *Am Heart J* 1981; **101:** 319–28.

[41] Friedman GD. Cigarette smoking and geographic variation in coronary heart disease mortality in the United States. *J Chron Dis* 1967; **20:** 769–79.

[42] Dawber TR. *The Framingham study.* Cambridge, Mass, Harvard University Press, 1980.

[43] Wynder EL, Bross IJ. A study of etiological factors in cancers of the oesophagus. *Cancer* 1961; **14:** 389–413.

[44] Schwartz D, Flamant R, Lellouch J *et al.* Alcool et cancer. Résultats d'une étude rétrospective. *Rev Fr Etudes Clin Biol* 1962; **7:** 590–604.

[45] Tuyns AJ, Péquignot G, Jensen OM. Le cancer de l'oesophage en Ille et Vilaine en fonction des niveaux de consommation d'alcool et de tabac. *Bulletin du cancer* 1977; **64:** 1: 45–60.

[46] Tuyns A. Oesopheal cancer in non-smoking drinkers and in non-drinking smokers. *Int J Canc* 1983; **32:** 443–4.

[47] Day NE. The geographic pathology of cancer of the oesophagus. *British Medical Bulletin* 1984; **40:** 329–34.

[48] Breslow NE, Day NE. *The analysis of case-control studies.* Lyon: IARC Scientific Publications no 32, 1980.

(Revised version received April 1986)