

# STATISTIQUE ET ANALYSE DES DONNÉES

SOCIÉTÉ FRANÇAISE DE BIOMÉTRIE

**Résumés - Journées de Statistique, Paris, mai 1979**

*Statistique et analyse des données*, tome 4, n° 3 (1979), p. 85-93

[http://www.numdam.org/item?id=SAD\\_1979\\_\\_4\\_3\\_85\\_0](http://www.numdam.org/item?id=SAD_1979__4_3_85_0)

© Association pour la statistique et ses utilisations, 1979, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

**SOCIETE FRANÇAISE  
DE BIOMETRIE**

**(Résumés - Journées de Statistique, Paris, MAI 1979)**



ENTREES ET SORTIES DE L'ACTIVITE ECONOMIQUE SUIVANT L'AGE. APPLICATION D'UN  
PROCESSUS MARKOVIEU INHOMOGENE : FORMULATION CONTINUE ; FORMULATION DISCRETE.  
CONVERGENCE FAIBLE DES POPULATIONS.

N. BROUARD  
Institut National d'Etudes Démographiques  
27, rue du Commandeur  
75675 PARIS cédex 14

Extension des tables de mortalité, les tables de vie active supposaient une vie active continue avec des entrées en activité en deçà de l'âge model des taux d'activité, et les sorties au delà ; assez rudimentaire pour les hommes, ces tables n'étaient pas applicables aux femmes. Un modèle plus réaliste permettant des entrées et sorties multiples au cours du cycle de vie, s'imposait.

Celui que nous présentons ici est un processus de Markov, où les forces de transition d'un état d'activité à un autre, analogues à la force de mortalité, dépendent de l'âge (processus inhomogène ou certaines propriétés de convergence apparaissent néanmoins, et rappellent la théorie de la convergence faible des populations).

Une première version présente un modèle continue, une seconde un modèle discret mieux adapté au traitement des données, d'une enquête emploi de l'INSEE. Une fois le modèle connu et estimé, on étudie plusieurs phénomènes difficilement appréhendables sinon : la durée totale de l'activité ou espérance de vie active, les prises et reprises d'activité notamment pour la population féminine ou les problèmes d'activité sont actuellement cruciaux.

ANALYSE DES DONNEES EVOLUTIVES

M. CHAVANE  
INSERM, U88  
91, Bd de l'Hopital  
75634 PARIS cedex 13

Les méthodes d'analyse de données sont de plus en plus fréquemment appliquées à l'étude des données évolutives, ou plus généralement des tableaux à triples entrée. Etant donné un ensemble de  $p$  variables mesurées sur  $n$  sujets dans  $l$  situations (ou  $l$  instants). Nous avons proposé une méthode permettant de choisir les variables les plus informationnelles. Afin de mettre en évidence les variables qui expriment le mieux les modifications globales en fonction des  $l$  situations de recueil des données, une analyse en composantes est effectuée sur le tableau  $l \times q$  contenant les moyennes des  $q$  variables dans les  $l$  situations (ou aux  $l$  instants). Afin de mettre en évidence les variables qui expriment le mieux les différences inter-individuelles  $l$  analyses en composantes principales sont effectuées à partir des données recueillies dans les  $l$  situations, puis une fois déterminé le nombre  $n$  de composantes que l'on souhaite garder, les  $q$  variables les mieux corrélées avec les  $q$  premières composantes de chaque analyse sont recherchées. La méthode utilisée est comparée aux propositions d'autres auteurs. Elle est appliquée à des données électroencéphalographiques.

## REGRESSION ET SERIES CHRONOLOGIQUES

D. DACUNHA CASTELLE  
Université Paris XI, Mathématiques  
91405 ORSAY

On étudie le modèle

$$Y_t = R_t + X_t$$

Obs. Régression bruit

$R_t$  peut être une fonction périodique ou une régression polynomiale, par exemple,  $X_t$  est un bruit de mesure spectrale inconnue. On étudie les problèmes de tests et d'estimation des paramètres de  $R_t$  pour des problèmes biologiquement "vraisemblables".

## UNE REPONSE A LA QUESTION "QUE FAIT UN BIOMETRICIEN AVEC UN TABLEAU INDIVIDUS x CARACTERES x EPOQUES DE MESURES ?"

P.DAGNELIE  
Faculté des Sciences agronomiques d'Etat, Statistique et  
Informatique  
Av. de la Faculté, 8  
5800 GEMBLoux, Belgique

Présentation et comparaison des différentes solutions qui peuvent être apportées au problème posé : différents modèles d'analyse de la variance à une variable, d'analyse de la covariance et d'analyse de la variance à plusieurs variables.

## MODELE DE TRANSITION ENTRE CAUSES DE DECES

P. DAMIANI, M. AUBENQUE, INSEE  
18, Bd A. Pinard  
75675 PARIS Cédex 14

On a établi, dans cette étude, un modèle de transition entre les différentes causes de décès, par sexe, pour le groupe d'âge 45-64 ans. On n'a pas tenu compte des accidents. Les données ont été corrigées pour tenir compte des décès de cause non spécifiée. On a calculé par département les taux de mortalité par cause en 1954 (moyenne 1952-1956) et en 1962 (moyenne 1960-1964). On a déterminé les régressions entre le taux de mortalité de chaque cause en 1962 et les taux de toutes les causes en 1954. A partir de ces résultats, on a établi la matrice des probabilités de transition d'une cause à l'autre au cours de cette période. On en a déduit des taux de mortalité limites par cause. On a enfin tracé des graphes représentant ces probabilités de transition.

COSINOR ET PERIODOGRAMME DE SCHUSTER

C. HAON et S. HUET  
Laboratoire de Biométrie  
I.N.R.A.-C.N.R.Z.  
78350 JOUY-EN-JOSAS

C. JACOB  
Laboratoire de Biométrie  
I.N.R.A.-C.N.R.Z.  
78350 JOUY-EN-JOSAS

Les biologistes se trouvent souvent confrontés au problème de recherche de périodicités d'une série temporelle ; dans cet exposé, on suppose que la moyenne du processus s'écrit comme une somme finie de sinusoides. Sous cette hypothèse, la recherche des périodicités peut se faire à l'aide des méthodes du COSINOR et du PERIODOGRAMME DE SCHUSTER.

En général, le COSINOR est très utilisé par les biologistes pour la recherche d'une seule périodicité ; or, son champ d'application est bien restreint :

- confirmation statistique de la périodicité d'un phénomène
- étude des influences que peuvent avoir des facteurs exogènes sur l'amplitude et la phase du phénomène périodique.

On étudie aussi les erreurs qu'entraîne une application plus large.

Par contre, le PERIODOGRAMME DE SCHUSTER, moins utilisé est bien adapté à l'étude de séries dont on recherche les périodicités.

La méthode permet de détecter les périodicités "significatives" ; c'est donc un outil de test. On calcule les amplitudes et les déphasages, c'est aussi une méthode d'ajustement. Une étude des limites et des intérêts de la méthode est faite. Des exemples sont traités.

Au vu d'une série chronologique (ou d'un échantillon) relative à une variable de mesure concernant un certain phénomène biologique périodique, il s'agit de pouvoir déterminer quelles sont les hypothèses mathématiques sous-jacentes à un tel ensemble de données et qui permettront de l'étudier statistiquement (y-a-t-il indépendance des observations d'un instant à un autre et Normalité des lois, le processus est-il stationnaire, possède-t-il une structure déterminée (que l'on retrouve d'une expérience à une autre ?)).

Selon la réponse à ces questions et selon le problème que l'on se pose :

- 1) Simple description du phénomène ;
- 2) Recherche descriptive de périodicités élémentaires (sinusoides) ;
- 3) Etude de la structure sous-jacente au processus ;
- 4) Prédiction du phénomène ;
- 5) Etude explicative.

On s'intéressera à des méthodes du type :

- 1) Analyse harmonique ;
- 2) Périodogramme ou analyse spectrale ;
- 3) Modèles à "moyenne" (régression périodique, cosinor, analyse de variance...) ;
- 4) Modèles saisonniers de Box et Jenkins ;
- 5) Modélisation : étude des solutions périodiques d'un système d'équations différentielles.

UN EXEMPLE D'ETUDE DE TABLEAU INDIVIDUS x CARACTERES x EPOQUES DE MESURE :

LES COURBES DE CROISSANCE DES 4 PREMIERS ORGANES DU BLE

L'ANALYSE DE DONNEES CHRONOLOGIQUES : QUELQUES ESSAIS

Y. ESCOUFIER  
C.R.I.G.  
Av. d'Occitanie  
34075 MONTPELLIER

M. GRANGER, J.P. MASSON, J.P. PAGES  
E.N.S.A. Rennes  
65, rue de St Brieuc  
35042 RENNES

Trois points sont abordés dans cette étude.

Les questions biologiques. Nous avons cerné les principales questions biologiques que l'on peut se poser à partir d'un tel tableau. Puis, en proposant un cadre de référence, nous les avons formalisées.

Réflexion sur la mesure de la croissance. Pour mesurer la croissance entre les instants  $i$  et  $i+1$ , on peut utiliser des accroissements de longueur bruts. Ce faisant on ne prend en compte qu'un aspect du phénomène : d'autres mesures peuvent être envisagées.

Utilisation de méthodes factorielles. A propos d'un tableau individus x caractères x époques de mesure de nombreuses méthodes factorielles peuvent être envisagées. Nous en avons appliquées plusieurs en indiquant quels aspects des données elles étaient les plus aptes à mettre en évidence.

On se propose d'étudier quelques exemples d'analyse de données chronologiques pour essayer de faire le point sur des techniques d'accès facile aujourd'hui. On s'interroge sur la pertinence des réponses que fournissent ces techniques et sur les fondements mathématiques des procédures mise en oeuvre. Les exemples traitent de coupes successives de luzerne, de courbes de croissance, de comparaison de matrices de covariances, de comparaison de tableaux de contingence.

DISCRIMINATION SUR DONNEES CHRONOLOGIQUES

J.P. NAKACHE, A. MALLET  
INSERM U 88  
16 bis, av. P.V. Couturier  
94800 VILLEJUIF

L'analyse discriminante basée sur la règle de décision baryessienne fournit, à partir de mesures relevées sur un ensemble d'individus répartis en  $k$  groupes bien définis a priori, des probabilités ponctuelles d'appartenance aux  $k$  groupes.

Dans de nombreux cas, notamment en médecine, où les paramètres mesurés à différents temps  $t_0, t_1, \dots, t_l$ , il s'agit de définir des probabilités à un temps  $t$  donné en tenant compte de l'information depuis l'instant initial  $t_0$  jusqu'à  $t$ .

Pour déterminer ces probabilités qu'on peut considérer comme cumulées par opposition aux probabilités ponctuelles, un modèle déterministe ou aléatoire est utilisé dans une première étape en supposant la tendance moyenne par groupe, linéaire ou polynomiale en  $t$ , avec aléas autocorrelés pour tenir compte du caractère chronologique des données.

Les paramètres du modèle, estimés par la méthode du maximum de vraisemblance sont ensuite utilisés pour définir la règle  $\zeta$  l'un des groupes.

Cette méthode est illustrée par un exemple médical concernant le pronostic et la thérapeutique au cours du temps pour des malades admis à l'hôpital dans le coma et en état de choc à la suite d'une overdose de barbiturique.

ANALYSE SPECTRALE DE MODELES A COMPARTIMENTS

E. POMMIES  
I.N.A.P.G., Chaire de Mathématiques appliquées  
16, rue Claude Bernard  
75231 PARIS Cédex 05

Dans de nombreux modèles, en particulier les modèles à compartiments, se pose le problème d'ajustement d'une fonction  $f(t)$  à une combinaison d'exponentielles du type

$$\sum_{i=1}^m a_i e^{-\lambda_i t}, \quad a_i \in \mathbb{R}^+, \quad \forall i, i = 1, \dots, m$$

La méthode proposée permet d'estimer simultanément le nombre et la valeur des coefficients  $\{a_i, \lambda_i\}$  :

Un programme informatique en langage Fortran a été exécuté.



## COMPORTEMENT DES LAPINS ET PROCESSUS PONCTUELS

E. JOLIVET  
Laboratoire de Biométrie  
I.N.R.A.-C.N.R.Z.  
78350 JOUY-EN-JOSAS

La statistique des processus ponctuels est un outil qui semble délaissé dans les applications, tout au moins en France. Il faut dire qu'il est rare de trouver de "bonnes données" relevant de ces méthodes. Sur des "séries d'événements" correspondants aux repas d'un lapin, on utilisera les outils de la statistique des processus ponctuels, et on comparera les résultats à ceux obtenus par des méthodes plus connues, mais moins bien adaptées.

## BIAIS LIES AUX ERREURS DE DATATION

P. LAZAR  
Institut Gustave Roussy  
16 bis, av. P.V. Couturier  
94800 VILLEJUIF

Une erreur de datation est susceptible d'introduire un biais important dans l'évaluation des corrélations entre variables dépendant du temps. On montrera comment ce biais dépend de l'espérance et de la variance de l'erreur de datation.

## UTILISATION DE LA DIVERGENCE DE JEFFREYS POUR L'ANALYSE DE DONNEES CHRONOLOGIQUES MULTIDIMENSIONNELLES

J.P. NAKACHE, A. CHEVALIER  
INSERM U88  
91, Bd de l'Hôpital  
75634 PARIS Cedex 13

On entend par données chronologiques multidimensionnelles, des données qui se présentent sous la forme d'une suite de  $l$  tableaux correspondant à différents temps. Chacun des tableaux contient, en lignes, les  $n$  individus de l'échantillon, et en colonnes, les  $p$  paramètres relevés sur ces individus.

Comme la plupart des méthodes d'analyse de données évolutives non basées sur un modèle, l'analyse proposée procède en deux étapes :

- i) Analyse inter-tableaux,
- ii) Recherche d'un référentiel commun à tous les temps.

La méthode proposée sera comparée aux techniques qui lui sont proches, à savoir : la double analyse en composantes principales de J.M. Bourroche, et la méthode utilisant les opérateurs d'Escoufier.

APPLICATIONS DE LA MODELISATION EN PHARMACOCINETIQUE

J.L. STEIMER, A. MALLET  
INSERM U88, Dept. Biomathématiques  
91, Bd de l'Hôpital  
75634 PARIS Cédex 13

LES PRINCIPAUX CONCEPTS SCIENTIFIQUES CORRELATIFS DU TEMPS EN PHYSIQUE  
ET EN BIOLOGIE

A.A. SANCHES  
Université de Nice  
Faculté des Lettres et Sciences Humaines  
98, Bd Edouard-Herriot  
06036 NICE Cedex

Dans les interprétations théoriques des disciplines scientifiques bien formalisées, le temps apparaît toujours, à l'analyse, comme un cofacteur privilégié d'un petit nombre de variables-type qui lui sont reliées d'une manière stable, claire et rigoureusement déterminée. L'appréhension intuitive (et l'exploration systématique) des relations -de type nécessaire, ou axiomatique) qui raccordent le temps à ces catégories fondamentales de variables permet, dans les différentes branches des sciences physiques, (i) le développement d'une structure cohérente de concepts simples et opérationnels, (ii) une formalisation rigoureuse des interprétations, et (iii) le perfectionnement théorique et pratique de la technique de mesure qui y atteint aujourd'hui, comme chacun sait, un très haut degré d'exactitude et d'efficacité.

La pharmacocinétique concerne l'analyse, la représentation, la prédiction... des phénomènes cinétiques impliqués dans le devenir d'un médicament après son introduction dans l'organisme (répartition, métabolisme, élimination...). En général, les données expérimentales se présentent sous la forme d'un certain nombre de mesures (concentrations sanguines, quantités dans les urines) recueillies à divers instants après administration du produit chez un certain nombre de sujets. Alliées aux développements récents de l'Analyse des Systèmes, les méthodes classiques de la Statistique permettent le traitement de telles phénomènes (souvent, modèle de compartiments), la seconde l'évaluation non linéaire. Diverses applications de cette approche générale sont présentées. D'une part elle permet de traiter de façon originale ces problèmes courants tels que le calcul des paramètres pharmacocinétiques pour un sujet et la comparaison de formes pharmaceutiques sur la base de protocoles standardisés. D'autre part, elle permet d'apporter des solutions à ces problèmes tels que l'évaluation des paramètres pharmaceutiques d'une population de sujets à partir de données hétérogènes, la définition de posologies optimales lors de traitements au long cours, et d'envisager la recherche de formes galéniques adaptées à un objectif thérapeutique particulier.