LES CAHIERS DE L'ANALYSE DES DONNÉES

J.-P. BENZÉCRI

Approximation stochastique, réseaux de neurones et analyse des données

Les cahiers de l'analyse des données, tome 22, n° 2 (1997), p. 211-220

<http://www.numdam.org/item?id=CAD_1997__22_2_211_0>

Les Cahiers de l'Analyse des Données
Vol. XXII - 1997 - n°2 - pp. 211 220.

APPROXIMATION STOCHASTIQUE, RÉSEAUX DE NEURONES ET ANALYSE DES DONNÉES

[APPR. RÉSEAUX]

J.-P. BENZÉCRI

0 Introduction

Mathématiques, algorithmes de calcul numérique, techniques du traitement de l'information, analyse statistique, structure des automates, psychologie de l'apprentissage, neurophysiologie, innovation en biologie... : autant de domaines de la science que l'on peut ou doit considérer à propos des réseaux de calcul. Beaucoup d'exemples viennent à l'esprit: la difficulté est de les ordonner sans confondre méthode avec problème, technique avec métaphysique.

1 Exemple élémentaire: le calcul de la moyenne d'un nombre indéfini de termes

Partons du cas le plus simple: le calcul d'une moyenne par approximation stochastique. Supposons que se présente à un observateur une suite de nombres: $x_1, x_2, x_3, \ldots, x_n, \ldots$ dont il faut calculer la moyenne.

Une première hypothèse est que la suite est finie, comprend exactement N nombres, N étant connu, *a priori*. En ce cas, il semble que l'observateur n'ait rien de mieux à faire que d'additionner les nombres au fur et à mesure de leur venue; se réservant de diviser le total par N, aussitôt que la séquence sera achevée.

Cette méthode n'est pourtant pas sans laisser à désirer. Supposons que les x_n soient les poids de colis que l'on devra transporter. L'observateur, ayant pour rôle de préparer ce transport, peut présumer que les colis ont tous des poids équivalents, que les derniers ne sont ni plus lourds ni plus légers que les premiers; en ce cas, il vaut la peine de tenir constamment à jour la valeur moyenne des poids déjà reçus: soit : $\mu_n = (1/n).\Sigma\{x_i \mid i=1, \ldots, n\}$.

On démontre aisément que deux valeurs successives de la moyenne peuvent être liées par la formule suivante:

$$\mu_{n+1} = \mu_n + ((x_{n+1} - \mu_n)/(n+1)) \ .$$

cette formule a le mérite de faire voir que la venue d'un (n+1)-ème nombre ne fait qu'apporter à la moyenne une correction; qui, de plus, serait nulle si ce nouveau nombre était rigoureusement égal à la moyenne des précédents.

Sans doute ce calcul par retouches successives a-t-il l'inconvénient de cumuler des petites imprécisions. Mais, d'une part, celles-ci ne s'ajoutent pas en valeur absolue, elles se compensent partiellement; et, d'autre part, le calcul séquenciel a l'avantage de signaler à l'attention de l'observateur toute erreur grossière, ou l'arrivée d'un nombre exceptionnel.

Mais surtout, le calcul séquenciel s'impose presque, si n'est pas connu le nombre N des termes dont on cherche la moyenne; particulièrement s'il s'agit d'une suite indéfinie. Et alors, le problème mathématique lui-même se modifie: il ne s'agit plus d'un calcul arithmétique élémentaire de moyenne, il s'agit d'estimer, d'après un échantillon, l'espérance mathématique d'une grandeur: x, dont on suppose qu'elle suit une loi de probabilité sur laquelle on peut faire diverses hypothèses.

2 Solution séquencielle d'un problème de calcul statistique: l'approximation stochastique en analyse factorielle

Prenons un deuxième problème, qui comme le calcul d'une moyenne, peut être décrit dans le formalisme mathématique, mais offre une réelle difficulté: l'analyse d'un tableau de correspondance $I \times J$.

Dans la pratique courante de la statistique, le tableau est donné comme un tout; on en connaît les dimensions: CardI et CardJ. Du point de vue de l'analyse numérique, la partie la plus difficile du calcul est la diagonalisation d'une matrice carrée symétrique de dimension: CardJ × CardJ ; même si, d'autre part, quand le nombre, CardI, des individus est très élevé, le calcul de la matrice, certes trivial, peut demander un grand nombre d'opérations élémentaires.

Cependant, il est commun que l'ensemble, I, des individus décrits dans le tableau, ne soit qu'un échantillon d'un ensemble potentiellement infini. Ici, de même que pour le problème élémentaire de la moyenne, le mathématicien propose un schéma probabiliste: les lignes du tableau sont issues d'une loi sur l'espace de dimension CardJ. Si la loi est fixée, la matrice à diagonaliser peut, en bref, se calculer en substituant un calcul intégral aux sommations usuelles; et la diagonalisation relève des mêmes algorithmes que pour CardI fini.

Mais, dans l'esprit de l'analyse des données, ce schéma ne satisfait pas le statisticien. La loi n'est pas connue; il est illusoire de lui donner forme mathématique. On est en présence d'un processus empirique, d'une source de

données, i.e. de cas individuels, que l'expérimentation suscite ou maîtrise imparfaitement; et dont l'existence peut être bornée dans le temps si, les conditions changeant, le phénomène observé dérive sans conserver une loi constante, fût-elle aléatoire (e.g.: la pathologie change, avec l'hygiène).

À ce nouveau schéma, qu'on peut appeler pseudo-aléatoire, répond parfaitement un algorithme itératif d'approximation qui est analogue à celui qu'on a décrit pour le calcul de la moyenne (cf. [APPR. CORR.]). En bref, au fur et à mesure que surviennent les cas individuels, l'état des valeurs des facteurs sur l'ensemble J des variables, est retouché; mais, tandis que pour la moyenne, on retouche par simple addition, ici, on retouche par multiplication matricielle: à tout individu est associée une matrice carrée: CardJ × CardJ (très simple, de rang 1: qui est, en substance, le produit tensoriel d'une ligne par elle-même); et c'est par l'action de cette matrice qu'est calculé le terme de correction des facteurs.

Toutefois, une difficulté se présente, que nous n'avions même pas évoquée; à propos de la moyenne. Dans ce dernier cas, il est clair qu'au rang n, le terme correctif a pour ordre de grandeur (1/n); et, avant toute élaboration mathématique, on présume une convergence. En analyse factorielle, la correction a pour effet une sorte de rotation: et on n'évalue pas, à première vue, à quelle instabilité peut conduire la composition de rotations successives.

Mais un modèle mathématique précis, celui de l'approximation stochastique dans une algèbre normée non commutative, permet de démontrer la convergence, moyennant, d'une part, des hypothèses vraisemblables quant à la source aléatoire; et, d'autre part, des règles dans le choix d'un coefficient affectant le terme de correction (i.e., l'analogue du (1/n) apparu dans le calcul de la moyenne).

Quant à la technique du calcul, l'approximation stochastique offre le remarquable avantage d'une grande économie de mémoire. Alors que le calcul usuel commence par traiter le tableau de base: CardI × CardJ, puis la matrice à diagonaliser: CardJ × CardJ, l'approximation stochastique, quand elle poursuit l'élaboration de p facteurs, ne garde que le tableau rectangulaire: p × CardJ, des valeurs de ces facteurs sur J. Ainsi, non seulement on peut considérer un ensemble quelconque d'individus, la venue de nouveaux cas ne servant qu'à améliorer la précision d'un résultat assez vite ébauché, mais le nombre CardJ des variables ne requiert même pas l'espace CardJ × CardJ. Pour traiter 1000 variables, l'espace requis n'est pas un million, mais quelques milliers.

Un autre avantage, déjà apperçu dans le calcul de la moyenne, est que l'algorithme, procédant par retouches minimes, mais dont l'effet est potentiellement infini (pourvu que les coefficients de correction soient

convenablement choisis) a, en quelque sorte, la capacité d'absorber les erreurs introduites par un individu aberrant.

3 Pratique de l'approximation stochastique

Ces avantages étaient particulièrement appréciables, vers 1970, quand fut proposée la méthode d'approximation stochastique; et c'est pourquoi, après J.–P. FÉNELON, L. LEBART entreprit de traiter par approximation stochastique des données réelles, avec un nombre élevé, CardJ, de variables; et, des individus en nombre, CardI, fini; mais lus plusieurs fois de suite pour simuler une source indéfinie.

De façon précise, L. LEBART a conjugué l'approximation stochastique avec d'autres méthodes d'analyse numérique. Voici comment: L. L. lit d'abord, deux fois de suite, l'ensemble I des n individus disponibles; en appliquant l'approximation stochastique comme s'il s'agissait de 2n individus issus d'une source ouverte ; (l'expérience a de plus, montré qu'il était avantageux de lire la suite des cas en alternant le sens de lecture: i.e. de 1 à n, puis de n à 1). Puis L. L. fait encore plusieurs (e.g. 3) lectures de I afin de multiplier par une puissance de la matrice d'inertie, les vecteurs axiaux fournis par l'approximation stochastique. Ensuite, L. L. projette le nuage N(I) sur les axes ainsi calculés en deux étapes; et effectue, pour le nuage projeté, une analyse usuelle exacte; qui améliore l'orientation des axes.

Depuis 1970, les moyens de calculs ont progressé au-delà de toute attente. L'espace de mémoire se mesure en méga-octets; l'algorithme SYMQR de diagonalisation est beaucoup plus rapide que celui d'approximation stochastique. Selon nous, la limite supérieure de CardJ n'est plus imposée par le coût du calcul, mais par la difficulté de concevoir des ensembles très complexes de données dont la cohérence se prête à analyse.

Certains statisticiens en jugent autrement. Des notices, publiées par M. MORFIN, nous ont fait connaître des travaux visant à réduire le rang de très grandes matrices; notamment celles croisant un ensemble I de textes avec un ensemble J de mots. En fait, ces matrices sont de très faible densité (i.e. comportent un taux élevé de cases nulles); on ne peut en extraire une structure stable et interprétable; mais leur représentation approchée par une matrice de faible rang permet, en quelque sorte, de se propager rapidement dans le réseau des documents.

Quelque avantage qu'on ait pu touver à un tel parcours, nous pensons que l'analyse documentaire doit plutôt se fonder sur une vue précise et confirmée de la structure du corpus; telle que celle que nous cherchons dans l'étude des classiques grecs. En considérant d'abord les mots les plus fréquents, on obtient une première typologie; où, par subdivision, on distingue des domaines dont

on peut voir la structure sans que jamais le thème d'une recherche particulière ne soit isolé. Les mots caractéristiques appellent d'autres mots; et des textes, liés à ceux-ci par un tableau de correspondance qui n'est pas lacunaire.

En somme, en explorant la voie de l'approximation stochastique, l'analyse factorielle a suggéré des problèmes mathématiques d'une certaine originalité; mais elle n'a pas trouvé un algorithme numérique efficace. En revanche, elle a rencontré certains problèmes majeurs de la technique contemporaine du traitement de l'information et aussi de la philosophie. C'est ici qu'apparaîtra un autre terme du titre du présent exposé: les réseaux de neurones.

4 Analyse discriminante et mécanisme du perceptron

Dans la première moitié du XX-ème siècle, alors que les moyens de calculs étaient, relativement aux nôtres, dérisoires, des statisticiens, tels le grand FISHER, posaient déjà le problème de l'analyse discriminante; problème familier au médecin qui cherche un diagnostic, comme au naturaliste qui doit déterminer l'espèce d'une plante ou d'un animal.

En terme mathématique, l'individu, cas clinique ou être vivant, est assimilé à un ensemble de variables de format déterminé; donc à un point d'un espace multidimensionnel; disons, d'un espace vectoriel. Dans cet espace, les entités à discriminer, maladies ou espèces, sont des ensembles de cas, i.e., de points; tout cas nouveau devant être affecté à l'un ou l'autre de ces ensembles.

Ici, la notion d'ensemble doit recevoir une forme précise. Certes, on imagine d'abord, sans peine, que chaque espèce occupe une véritable partie d'un plan: par exemple, un domaine polygonal. [Et nous ne cacherons pas que trop de publications, proposant une méthode de discrimination, se bornent à en faire l'essai, nullement convaincant, sur un tel schéma.]

Mais quand le naturaliste interoge le statisticien, ce n'est pas pour distinguer un rat d'un écureuil, mais, tel L. BELLIER dans l'étude du genre *Praomys*, pour distinguer entre deux espèces ou sous-espèces voisines dont l'individualité même n'est pas encore admise par tous les spécialistes (cf. Traité: IC n°5). Dès lors, la multiplicité des dimensions est essentielle; surtout si les données qu'on peut mesurer avec précision dans tous les cas, ne suffisent pas à distinguer, sans erreur, les deux espèces. Avec une telle description, non seulement il n'y a pas de cloison de forme simple; mais les ensembles à discriminer empiètent.

Pour Sir Ronald FISHER, compte tenu des moyens de calcul dont il dispose et des images que, dans ce contexte, il s'est ainsi accoutumé à traiter avec élégance et rigueur, un ensemble est assimilé à une distribution normale multidimensionnelle; avec, e.g., pour une espèce, un point moyen et, pour régir la dispersion autour de ce centre, une matrice de variance et covariance,

définissant une distribution en cloche dont la densité décroît par ellipsoïdes concentriques; la masse totale de la distribution pouvant être fixée, non à 1, mais à une valeur proportionnelle à la probabilité relative de l'espèce.

Selon ce modèle, le domaine de chacune des espèces s'étend à tout l'espace [ce qui, à la vérité, sort des limites du réel]; mais un cas individuel, représenté par un point M, est affecté à celle des distributions qui, en M, l'emporte sur les autres par la densité. Ainsi peuvent être construites des hypersurfaces de séparation délimitant le domaine afférent à chaque espèce.

En fait, même regardée comme une approximation, l'hypothèse de normalité des distributions n'est vérifiée que rarement. Le mérite du modèle est qu'il a permis de considérer le problème fondamental de la discrimination alors que, répétons-le, les moyens de calcul étaient dérisoires.

Avec l'avènement des ordinateurs, l'idée cybernétique, issue de Norbert WIENER, d'expliquer et d'imiter la vie selon des équations différentielles régissant des processus aléatoires, apparaît praticable. Le projet de la reconnaissance automatique des formes vient à l'ordre du jour. Vers 1955, on espérait que la machine à percevoir, (à lire l'écriture manuscrite, à reconnaître les paroles et les visages...,) fontionnerait bientôt.

Ainsi fut conçu le "Perceptron" de ROSENBLATT. En bref, la machine recevant sur une couche de cellules sensibles, ou rétine d'entrée, des images simples, ou stimuli, relevant de plusieurs classes distinctes, élabore une réponse, au travers d'un réseau de cellules aboutissant à une série de lampes. Initialement, les liens entre cellules ont des valeurs arbitraires. Au cours d'un processus d'apprentissage, sont présentés des stimuli divers, l'utilisateur donnant la réponse désirée, i.e. celle des lampes qui aurait dû s'allumer. Le Perceptron, traitant ces données selon un programme universel, modifie la forces des liens entre cellules; et est censé, au terme de l'apprentisage, fournir, par lui-même, des réponses généralement correctes.

Ce système ingénieux connut en son temps quelques succès, parfois spectaculaires; mais qui n'étaient pas à la mesure des besoins pratiques. On l'oublia, peut-être, pendant vingt ans. Mais les principes en sont revenus aujourd'hui, dans les "réseaux neuronaux".

Ce n'est pas le lieu de décrire les capteurs, de sons ou d'images, dont sont munies les machines en service ou en projet. Du point de vue mathématique, la machine, avec ses connexions, est comme un programme dont le système de coefficients est décrit par un point: C, dans un espace de dimension convenable. De même, les stimuli sont comme des points: S, d'un autre espace.

Dans l'apprentissage, les stimuli, sont issus d'une source aléatoire. En des termes familiers aux psychologues, on dira que les renforcements des liens, élaborés en tenant compte de la réponse exacte, doivent, selon un processus d'approximation stochastique, converger vers une structure apte à fournir, par elle-même, à tout stimulus, la réponse exacte.

Ainsi que nous l'annoncions en introduction, spéculation mathématique, jeu algorithmique, problèmes techniques majeurs, biologie, philosophie... se sont ici donné rendez-vous. Sans précipitation, exposons d'abord ce que nous attendons de l'efficacité de l'approximation stochastique.

5 Efficacité relative de l'approximation stochastique comparée à d'autres méthodes

Au §2, l'approximation stochastique a fourni un algorithme, non dépourvu d'élégance, pour résoudre un problème numérique classique, posé par l'analyse factorielle. De façon précise, Ludovic LEBART, a même pu proposer, pour les divers problèmes évoqués succintement aux §§2 et 3, des solutions qui rentrent exactement dans le cadre formel des Perceptrons et des réseaux considérés par les spécialistes d'aujourd'hui.

Mais, d'une part, il ne semble pas que, dans l'esprit de l'approximation stochastique et des réseaux, on aurait pu concevoir une méthode précise d'analyse factorielle; et, d'autre part, cette méthode une fois conçue sous d'autres inspirations, il apparaît que les calculs se font bien plus vite par l'analyse numérique usuelle que par approximation stochastique. Quant à l'efficacité, cette première comparaison est donc défavorable aux réseaux.

Quant à la discrimination, problème propre pour lequel les réseaux ont été conçus, une comparaison nous est offerte par quatre articles publiés dans les cahiers n°1 et n°2, de 1996. Dans ces articles, ([DONNÉES RÉSEAUX], [CODAGE DISCRI.], [EXON INTRON], [STRUCTURE INTRON],) des jeux de données assez complexes, utilisés d'ordinaire comme banc d'essai pour les réseaux, sont traités en soumettant les données à un codage préalable, au vu d'histogrammes afférents aux classes à discriminer. En bref, l'analyse de correspondance résout le problème de discrimination au moins aussi bien que ne le font les réseaux; et le coût en calcul est si bas, qu'il suffit d'un micro-ordinateur travaillant quelques minutes; au lieu de longues heures d'apprentissage sur une machine plus puissante.

[On a récemment publié le compte-rendu d'une étude de discrimination où, avant d'entrer dans les réseaux neuronaux, les données sont traduites en terme de facteurs issus d'une analyse dont le schéma n'est pas précisé: il nous paraît que c'est là transmettre au réseau un problème déjà virtuellement résolu; pour se donner ensuite l'avantage de réussir là où, d'ordinaire, on échoue].

On objectera ici que, pour chacun des problèmes traités dans les articles cités de *CAD*, le statisticien a consacré plusieurs heures à l'examen approfondi des données; et que l'on ne songe pas, présentement, à tenir la gageure de laisser seule au prise avec les données, une machine munie seulement d'un programme universel d'analyse.

Il est vrai que ce projet d'une machine autonome séduit. Mais nul ne l'a encore réalisé. Quand, en vue d'une application déterminée, (par exemple pour reconnaître les sons d'une langue,) on a recours aux réseaux de neurones, c'est au prix de centaines d'heures de programmation qu'on aboutit, dans le champs choisi, à une discrimination imparfaite qui ne s'étend pas directement ailleurs (e.g. à une autre langue).

Une autre expérience nous a fait apprécier la portée et les limites de l'approximation stochastique. Dans le modèle d'instauration d'un code, (cf. [INST. CODE (1, 2, 3)]), est simulé le comportement de sujets qui communiquent entre eux pour désigner des objets, points d'un espace A, par des signaux, points d'un espace B. Partant d'un code liant aléatoirement A et B, on aboutit, par renforcement des liens ayant produit des communications réussies, à un code stable, de rendement appréciable, et dont la structure offre, par morceaux, une sorte de correspondance cartographique entre A et B.

Certes, un programme d'analyse factorielle, détermine directement la structure globale d'un ensemble A; mieux que ne le fait l'algorithme d'instauration de code, qui met A en relation avec B. Mais l'intérêt de cet algorithme vient de la pauvreté des informations qu'il traite: des relations de contiguïté entre éléments d'un ensemble. On peut dire que, par le code, il range les objets suivant une forme qui ne lui a pas été indiquée globalement; qu'il s'élève du local au global.

6 La genèse de l'ordre: naissance ou reconnaissance d'une forme

Le terme même de réseau de neurones évoque l'ordre du vivant, à propos d'un algorithme. Ainsi, celui-ci est investi de l'autorité de celui-là; et, en retour, le psychologue et le biologiste, acquièrent, dans leurs propres spéculations, l'assurance que donne la rigueur des mathématiques.

Pour l'heure, on admet volontiers un codage universel des comportements par des circuits ou réseaux neuronaux. Mais est-il vraisemblable que la technologie du traitement de l'information, mise au service de l'intelligence des distingués vertébrés que nous sommes, soit inférieure à celle qui, dans le génome, régit la forme d'une limace (cf. [FLÈCHE], §2). Il pourrait exister des médiateurs chimiques très complexes du comportement. À tout le moins, la labilité des contacts synaptiques (entre neurones du cerveau) dépasse celle qu'autorise la variation des coefficients du réseau associé à un algorithme.

Quant à la structure, l'algorithme n'a pas reçu toute l'inspiration du vivant...

La pensée évolutionniste se complait dans le projet d'un mécanisme universel qui, à l'épreuve d'un apprentissage, élabore une structure, apte à résoudre un problème. Il y a des évolutions, mais les voies n'en sont pas toutes connues.

Des exposés généraux postulent, implicitemement, que les processus d'évolution sont stables; et convergent vers une trajectoire; laquelle, dans un espace abstrait généralement non défini, poursuit, avec l'efficacité maxima, l'optimum d'un critère, lui-même suggéré sans précision. L'expérience des systèmes mécaniques asservis, conçus dès avant le milieu du XX-ème siècle, montre, au contraire, que la convergence vers l'optimum est difficile à assurer; même dans les processus les plus simples (cf. [CONVERGENCE MODÈLES]).

Il n'y a donc pas lieu d'attendre, de l'approximation stochastique sur un réseau, la solution implicite des problèmes les plus complexes. Certes, même imparfaite, toute convergence d'un processus vers une forme nous intéresse. Mais l'optimum ne s'offre qu'à une stratégie globale.

En particulier, nous croyons que la reconnaissance de la parole requiert une phonologie nouvelle, fondée non sur les traits pertinents de l'émission des sons articulés, mais sur d'autres traits pertinents; qu'on doit trouver dans le seul signal sonore. À ce projet théorique, nous espérons que servira l'analyse des données multidimensionnelles.

Références bibliographiques

L. BELLIER , d'après M.-O. LEBEAUX: "L'étude de la forme du squelette de la tête chez le *Praomys* (petit rongeur africain)"; in *L'Analyse des Données*, I: La Taxinomie; ICn°5; Dunod, (1973);

J.-P. BENZÉCRI : "Approximation stochastique dans une algèbre normée non commutative"; in *Bull. Soc. Math. France.*; T.97, pp.225-241; (1969);

J.-P. FÉNELON : *Deux contributions à une programmathèque d'analyse des données: analyse factorielle des préférences, approximation stochastique de l'analyse des correspondances*; Thèse; Université de Paris VI, (1973);

[APPR. CORR.] : J.-P. BENZÉCRI : "L'approximation stochastique en analyse des correspondances"; in *CAD*, Vol.VII, n°4, pp.387-394; (1982);

[BERGSON] : J.-P. BENZÉCRI : "Qualité et quantité: la grandeur et l'espace selon BERGSON et en analyse des données"; in *CAD*, Vol.VII, n°4, pp.395-412; (1982);

[MEM. STOCH.] : J.-P. BENZÉCRI : "Intégration de l'information dans la mémoire et approximation stochastique"; in *CAD*, Vol.X, n°4, pp.495-498; (1985);

[FLÈCHE] : J.-P. BENZÉCRI : "Système de la narture et flèche du temps"; in *CAD*, Vol.XVIII, n°1, pp.97-118; (1993);

[INST. CODE (1, 2, 3)] : J.-P. BENZÉCRI : "Sur l'instauration d'un code"; in *CAD*, Vol.XX, n°3, pp.301-372; (1995);

[CONVERGENCE MODÈLES] : J.-P. BENZÉCRI : "Convergence des processus et modèles d'économie libérale et de philogénèse"; in *CAD*, Vol.XX, n°4, pp.473-482; (1995);

F. MURTAGH : "Application de l'analyse factorielle et de l'analyse discriminante à des données colligées pour être soumises à des réseaux de cellules"; [DONNÉES RÉSEAUX], in *CAD*, Vol.XXI, n°1, pp.53-74; (1996).

HASSAN HAMOUD Anwar : "Diversité des codages permis en analyse discriminante: exemple de données cytologiques"; [CODAGE DISCRI.], in *CAD*, Vol.XXI, n°1, pp.75-82; (1996);

F. MURTAGH , J.-P. BENZÉCRI : "Discrimination des jonctions entre exon et intron dans les séquences d'acide désoxyribonucléique"; [EXON INTRON], in *CAD*, Vol.XXI, n°2, pp.133-148; (1996);

A. M. ALKAYAR : "Structure des introns au voisinage d'un point limite entre exon et intron sur une séquence d'ADN"; [STRUCTURE INTRON], in *CAD*, Vol.XXI, n°2, pp.149-164; (1996);

L. LEBART : "*Réseaux de neurones et analyse des correspondances*"; (1997);

M. MORFIN : "La SVD en Statistique Multidimensionnelle"; in *Les Cahiers de l'IMA*; n° 97-1; (1997).