

СИЖАВЭЗ...

Si j'avais un laboratoire...

Jean-Paul BENZÉCRI

Il est juste, que sous ce titre, Jean-Paul Benzécri rende hommage à Jean-Pierre Fénelon et à tous ceux, dont beaucoup ont quitté ce monde, grâce à qui on a pu faire... ce qu'il me plairait qu'on fit aujourd'hui !

Il y a quarante ans, allait commencer une année scolaire mémorable, dont il est commun de déplorer les ravages¹, mais dont il faut aussi que je confesse d'abord avoir été l'heureux bénéficiaire.

Chargé par le professeur Daniel Dugué de l'enseignement du diplôme d'Études Approfondies de Statistique, je choisis pour thème du cours : l'Analyse des Données, titre dont Dugué me dit en souriant qu'il embrassait toute la Statistique ! Moi, sous cette ample bannière, j'entendais multiplier les Analyses de Correspondances : analyses par lesquelles, grâce à la patience de Brigitte Cordier² (travaillant sur un IBM 1620 : une calculette d'aujourd'hui mais que le Doyen Yves Martin nous avait offerte pour le prix d'un Château !) ... j'avais pu de Rennes, prétendre conquérir la Linguistique, l'Économie,...

Or pour analyser, il fallait des données... je résolus d'envoyer les étudiants du D.E.A. cueillir des gerbes de ces précieuses fleurs.

Dès mon premier cours, j'annonçai qu'ils devraient faire des stages pratiques. Mais cet appel, répété de semaine en semaine, ne trouvait pas d'écho. Les étudiants pensaient que, s'ils survivaient à un écrit et un oral – problème et questions de cours – personne ne pourrait leur en demander davantage ; tandis que sans potasser... ils ne gagneraient rien à l'originalité d'une excursion dans la pratique.

D'ailleurs, auraient-ils été dociles à mon projet – voire séduits par ce projet... qui les aurait accueillis en stage ?

Le joli mois de Mai 68 allait tout changer !

¹ Depuis que j'ai écrit *сижавэз*, on a fêté le 40-ème anniversaire... comme un événement dans la pensée, ou, au moins, l'idéologie ! En bref, voici mes souvenirs.

Vers avril 1968, j'avais reçu le programme des sessions d'examen pour juin ; avec tel écrit là, tel oral ici... c'était tendu à craquer... et les nouvelles de Mai ne me semblèrent, d'abord, que confirmer mes pronostics : bien plus que d'idéologie, il s'agissait de la charge d'un projet d'Instruction Publique écrasant !

J'ai été lent à reconnaître que, dans la lourde terre du projet, il avait germé des graines ...

Une charge écrasante ! on spéculait, aujourd'hui, sur la vie et sur la mort... D'abord, ce sont les techniques qui ont tout changé : les humains, de par la médecine, vivent – ou se survivent à eux-mêmes, de plus en plus longtemps ; mais incapable de suivre les techniques, ils se trouvent inaptes au service, de plus en plus tôt.

Respecter les Idées, dire loyalement ce dont on est convaincu... mais aussi, dans les débats d'idées, soupeser patiemment des faits sur lesquels tous peuvent préalablement s'accorder.

² NdR : il s'agit de Brigitte Escoffier-Cordier, professeur à l'Université de Rennes 2 décédée le 15 juillet 1994, dont la thèse soutenue en 1965 figure avec le cours de linguistique mathématique donné par Jean-Paul Benzécri à l'Université de Rennes parmi les références princeps de cette époque.

Habitant à Orléans, je n'entendais des cris de Lutèce que de lointains échos ; finalement en septembre, il fallut bien que patron et étudiants se résignent à reprendre chacun son rôle...

Avec, pour chacun, de la prudence... mais aussi quelque audace !

Depuis le mois de Mai, on avait clamé que l'Université rhumatisante, sclérosée... ne préparait pas à la Vie. D'autres que moi avaient prôné les stages ! Machinalement, sans doute, des étudiants vinrent me dire qu'ils voulaient faire des stages. Je triomphais ! « Oui, oui, », avouèrent-ils en bégayant, « vous nous l'aviez dit ».

Et comme ceux mêmes qui, en novembre 67, auraient refusé tout stagiaire, se sentirent, en septembre 68, tenus de sauver la Patrie en choyant la jeunesse... la voie s'ouvrit à l'Analyse des Correspondances ! Analyse dont la pratique a été, très tôt, associée à celle de la Classification Ascendante Hiérarchique (CAH).

Le départ était pris... avec Pierre Leroy, puis Madame Laraise... les étudiants du DEA, bien conduits, se répandaient partout, pour la cueillette !

Analyse des Scrutins

Je ne pense pas que le Suffrage Universel réponde à la fonction qu'on lui attribue...

Généralement, il s'agit, au fond, de problèmes trop complexes pour que les électeurs eux-mêmes en puissent décider ; et la compétence des hommes à qui l'on confie de décider ne peut être bien appréciée.

Mais les hommes politiques, qui prônent une décision, ou prétendent décider, eux-mêmes, ultérieurement..., parlent aux électeurs : et ceux-ci votent selon l'idée qu'ils se font du rapport qu'il y a entre, d'une part, ce qu'ils entendent dire, et d'autre part, ce qu'ils voudraient qu'on fit pour eux.

Il y a là une Correspondance qu'il faut analyser.

Par deux fois au labo du quai Saint Bernard, on a confié des données en souriant, comme si on vidait une corbeille à papier dans la mémoire d'un ordinateur... et l'Analyse a montré des structures devant lesquelles le sourire s'est changé en grimace !

En 2007, on a voté, en France : pour choisir, d'abord, un Président ; et, ensuite, des Députés.

Au 1^{er} tour, il y avait, disons, dix candidats pour la Présidence; dont la plupart ne pouvaient être élus, mais offraient aux électeurs l'occasion de manifester, pour une fois !, leur humeur, mauvaise ou bonne [de même, quand j'étais étudiant, je profitais des soirées d'Élection, pour suivre les dépouillements, dans les lieux publics, que, d'ordinaire, on ne peut visiter en flânant...].

Soit **I**, un ensemble de bureaux de vote (ou de circonscriptions dont chacune peut comprendre plusieurs bureaux) ; **J_s**, l'ensemble des attitudes possibles à un scrutin « *s* », abstention comprise. On a un tableau de correspondance ks , avec à l'intersection de la ligne *i* et de la colonne *j*, le nombre $ks(i,j)$ des électeurs qui, inscrits dans le bureau *i*, ont adopté l'attitude *j*.

Que produira cette analyse ? l'effet Guttman ? Un 1^{er} axe opposant la « droite » à la « gauche »... un 2^{ème} axe opposant le centre aux extrêmes ; peut-être... mais les attitudes de faible poids ne sont pas toutes extrémistes, et certaines ont une vive couleur... et où sera l'abstention ?

L'ensemble **J_s** des attitudes une fois choisi, la représentation obtenue pour cet ensemble dépendra de l'ensemble **I** des circonscriptions considérées.

Et on peut juxtaposer des tableaux croisant un même ensemble **I**, avec plusieurs ensembles **J_s**. Par exemple, si **I** est l'ensemble des bureaux de vote décidant du choix d'un unique député, on verra le lien entre choix du Président et choix du Député.

Avec, ici comme partout, les grimaces ! Il faudra que le prix des logements dans *i*, dépende de la nuance révélée par la correspondance avec **J**...

[Est publié dans les Cahiers de l'Analyse des Données, Vol XX, 1984, un article de H.M. Badran, qui donne des exemples d'analyse de scrutins, avec une bibliographie des analyses analogues déjà parues avant cet article.]

Et les programmes d'action associés à une attitude *j* (abstention exceptée...) ? En compilant les déclarations des candidats, on relèvera une liste **L**, de mots ou de locutions ; d'où un tableau de correspondance entre **L** et **J** (**J** sans l'abstention).

Ce tableau **L** x **J** peut être analysé tel quel. Un candidat ne tient pas le même discours, au 1^{er} tour, *j1*, et au 2^{ème}, *j2* : car à *j2*, assuré de posséder en propre les électeurs qui l'ont suivi à *j1*, il voudra aussi rallier la clientèle de candidats éliminés au 1^{er} tour.

De plus, en passant par **J**, on pourra construire une correspondance entre *I* (ensemble des bureaux) et **L** (liste des mots...).

Etc. etc.

Physique

La physique de la matière progresse au cours des générations par les efforts conjugués de l'observation, de l'expérimentation, de la théorie, du calcul ... selon un programme que propose la philosophie.

L'article hep-ph/0610012, article que j'ai pris sur internet, (arXiv :hep-ph/0610012 ; en France, site : <http://fr.arxiv.org>) considère deux expériences gigantesques : Tevatron et LHC.

Il s'agit, ici comme là, d'ensembles de résultats de mesures effectuées après des chocs de deux particules élémentaires se mouvant avec une très grande énergie.

D'un seul choc, résulte un faisceau très complexe de particules. Pour observer ce faisceau, il faut un appareillage monumental... dans ce qu'on observe, ce qu'on présume mériter d'être élaboré ne peut être choisi que par une informatique inimaginable il y a vingt ans... et le choix étant fait, reste à faire l'élaboration !

Du Tevatron, déjà en service, au LHC, que l'on prépare, le progrès nous paraît être en bref que :

- ✓ dans le Tevatron, le choc initial est entre des particules primaires bien connues, des protons ; le prodige étant que ces particules sont mues comme par une différence de potentiel de 10^{12} volts...
- ✓ dans le LHC, on suscitera des chocs entre des particules, des hadrons, résultant de chocs préalables.

Ici et là, l'interprétation de ces chocs est fondée sur une théorie selon laquelle ce qu'on appelle communément particule élémentaire, disons un proton, est en soi, à très petite échelle, un système d'infra-éléments (appelés : quarks) ; lesquels ne peuvent pas être directement observés ; parce qu'ils n'existent que confinés au sein des particules que l'on appelle élémentaires mais qui sont en fait complexes.

Ainsi, dans le choc de deux protons, on veut saisir, comme phénomène élémentaire, le choc d'un infra-élément de l'un avec un infra-élément de l'autre. Dans le choc, les protons sont détruits ; mais les infra-éléments sont réintégrés en particules. Et les mesures portent finalement sur des gerbes de particules usuelles. De ces gerbes, on veut remonter aux infra-éléments. Dans le choc des deux infra-éléments, on présume que se produisent d'abord d'autres infra-éléments ; à la connaissance desquels on veut remonter ; en amont des gerbes, seules observées. Au plus simple, on reconnaîtrait dans les gerbes des sous-gerbes, issues chacune d'un infra-élément. Cela est impossible...

Mais, partant d'une théorie, on peut simuler, par Monte-Carlo, des gerbes finales.

Et, par une sorte de CAH, proposer une décomposition, de la gerbe finale simulée, en sous-gerbes.

De ce parcours, tout en calcul, mais fondé sur un schéma théorique, résulte un ensemble potentiel de gerbes subdivisées ; ensemble qui a certains caractères statistiques globaux.

Le processus de décomposition peut aussi être appliqué à un ensemble réel de gerbes... ensemble réel en ce qu'il est issu des observations sur Tevatron ou LHC.

On sera satisfait de la théorie si les caractères statistiques globaux des pertes réelles sont proches de ceux des gerbes simulées.

On a dit : une sorte de CAH... c'est ici que les vétérans de l'Analyse des données prétendent éclairer la naïveté des physiciens !

Le §2 de hep-ph/0610012, pp.5-31,

« jet algorithms »,

considère des algorithmes d'agrégation : divers quant au choix de centres initiaux, quant au critère, distance angulaire ou inertie... ;

- ✓ sans considérer : qu'au même ensemble (nous avons dit : à la même gerbe) on pourrait appliquer une CAH en plusieurs étapes successives (par exemple : **I**, de la gerbe donnée à un ensemble intermédiaire ; et **II**, de cet ensemble au schéma retenu), chaque étape ayant son propre critère ;
- ✓ sans considérer : que l'essentiel n'est pas que l'algorithme produise une décomposition où l'on reconnaisse, tel quel, le choc initial entre infra-éléments ; mais que la théorie physique soit confirmée parce que les caractères statistiques globaux des gerbes réelles sont proches de ceux des gerbes simulées...

Dans « jet algorithms », on redoute de multiplier les traitements divers : mais il se peut que l'expérience acquise en CAH (agrégation autour de centres variables, etc.) permette d'aller plus vite qu'on ne l'a fait jusqu'ici ; et, donc, permette de multiplier les essais.

Cosmologie

Dès les débuts de la théorie de la Relativité Générale (≈ 1917), on a voulu décrire, l'Univers, dans son ensemble : le Cosmos ; description dans l'espace-temps ; donc description d'une histoire.

Après un demi-siècle d'efforts conjugués de la spéculation théorique et des observations, s'est finalement imposé le schéma du « Big Bang » ; terme, d'abord introduit par dérision, afin d'évoquer, par le bruit d'une explosion, la croissance, au cours du temps, des dimensions spatiales de l'Univers.

Maintenant, on affirme même qu'il s'agit d'une croissance accélérée ; qui pourrait aboutir à une catastrophe finale : « Big Rip », la déchirure...

L'histoire du Cosmos repose sur l'ensemble de ce que présentement on en peut voir : des faits d'autant plus éloignés dans le Passé qu'ils sont plus éloignés dans l'Espace.

La nouvelle Astronomie fondée non seulement sur l'observation du rayonnement visible mais sur tout rayonnement... considère particulièrement les explosions dans le cosmos, de systèmes stellaires ; de ce qu'on appelle « supernova »...

Il s'agit d'élaborer des observations, non d'analyser des expériences...

Ici, il faut avoir une typologie des « supernova » et de leurs explosions, dans leur diversité : car c'est par des événements de même nature et de même type que l'on construit le système global du Passé et de l'Espace.

Voici deux articles dont on voudrait analyser les données : L'article astro-ph/0311622, article pris sur internet, (arXiv:astro-ph/0311622 ; en France, site : <http://fr.arxiv.org>) a pour titre : « Cosmological parameters from supernova observations : A critical comparison of three data sets »

Alors que, dans hep-ph/0610012, (Tevatron et LHC) un chapitre entier traite, à sa manière, d'Analyse des données (de CAH...), telle que nous l'entendons, le titre même de : astro-ph/0311622 annonce des calculs d'ajustement de modèles... croisements de paramètres avec ellipses d'erreur. Et, compte tenu de l'arrivée de données nouvelles, on peut dire que l'article est déjà ancien.

Tout en reconnaissant les faiblesses de leur méthode, les auteurs manifestent, dès le début, leur défiance vis-à-vis des modèles complexes.

Cependant, malgré l'hétérogénéité des données, on souhaiterait considérer la « supernova » comme un objet multidimensionnel. Cela est même indispensable : même si c'est sur quelques dizaines de cas seulement, on doit s'assurer que les observations que l'on conjugue sont prises d'un même point de vue.

Il faudrait chercher, sur internet, des données récentes et rallier les auteurs aux méthodes que nous cultivons.

L'article astro-ph/0701520, pris également sur internet : (arXiv:astro-ph/0311622 ; en France, site : <http://fr.arxiv.org>): « Gamma-ray Bursts in the Swift era » concerne, lui-aussi, la Cosmologie, mais c'est un article tout récent, fondé sur de formidables observations par satellites, par fusées... abondance d'informations multidimensionnelles !

Le titre du §2 : « Classes of gamma Ray Bursts » nous arrête par le mot de : « Classes »...

Jusqu'ici les « Classes » ont été définies d'après une seule variable ; la classification change selon la variable adoptée. L'auteur n'est pas satisfait des typologies ainsi obtenues...

L'abondance des données, qui sont non seulement des variables isolées mais des courbes..., invite à entreprendre de constituer des ensembles homogènes, pour l'Analyse de Correspondance, la CAH.

La physique offre aussi matière à une

sociologie !

Qu'il s'agisse de Cosmologie ou de schémas théoriques nouveaux, les auteurs qui traitent un problème donné restent divisés en plusieurs groupes ; qu'on devine d'après la bibliographie... Analyse de Correspondance à faire :

T (ensemble des textes d'article) x **B** (ensemble des auteurs cités) ;

Analyse sociologique dont la physique elle-même pourrait bénéficier ; car il n'y a pas seulement variété de personne, mais d'outil et de point de vue.

Une phrase entendue :

Il sera désormais interdit de brancher la

climatisation

Dans un local où la température est inférieure à 26°...

Cette phrase n'a pas de sens !

Où mettra-t-on le thermomètre ? Si la lumière du soleil entre par des fenêtres y exposera-t-on un thermomètre ? ... l'état de la température d'un local est une grandeur multidimensionnelle.

Dans un local donné, *loc*, on placera un ensemble **J** de thermomètres *j* : sans doute vaut-il mieux que chacun soit enveloppé : par exemple, plongé dans un vase transparent contenant de l'eau (ou tout autre liquide) ; car ainsi, on appréciera mieux l'effet de la chaleur (ou du froid...) sur qui est dans le local.

Pour *loc*, on fera des mesures à diverses heures, en diverses saisons... ; dans un ensemble **I** de cas. Pour chaque cas, *i*, on aura une suite $k(i, \mathbf{J})$ de températures ; suite indicée par *j*. D'où, pour *loc*, un tableau $k(\mathbf{I}, \mathbf{J})$, ayant autant de colonnes qu'il y a de thermomètres ; et autant de lignes que l'on aura considéré de cas.

Du tableau $k(\mathbf{I}, \mathbf{J})$, on essaiera divers codages... ; par l'analyse factorielle et la CAH, on décrira ce que peut être, globalement, la température du local *loc*.

En considérant, un ensemble de **L** de locaux, *loc1*, *loc2*, ... (locaux d'un même type auquel on s'intéresse), on acquerra une vue globale de ce qu'est la température.

Restera à décider de la clim...

Considérer ici :

Météo...

Le temps, même en une seule ville, n'est pas une température, ou un intervalle dans lequel varie la température ; avec pluie ou sans pluie ; etc. c'est une image globale : il faut apprendre à classer ces images !

Autre phrase : 17% des français sont

Catholiques...

Voire : se disent Catholiques.

Être Catholique : croire ce qu'enseigne l'Église catholique et s'appliquer à être devant Dieu et devant les hommes, ce qu'elle enseigne qu'on doit être !

Aller à l'église chaque Dimanche ; ou y aller parfois, en observant les fêtes de Noël, de Pâques...

Confier à l'Église de célébrer les mariages, les funérailles...

Ou encore : n'être ni Calviniste, ni Musulman, ni Luthérien, ni athée, ni Israélite...

Pour savoir ce qu'on peut vouloir dire du lien qu'on a avec la religion en général, il faut d'abord, avec diverses personnes qui ont pensé à cela, et qui s'expriment avec aisance, avoir des entretiens libres.

D'après ce qu'on aura enregistré de ces entretiens, on entrera dans une suite d'étapes

Entretiens **libres**

Questionnaire **ouvert**

Questionnaire **fermé**

Typologie des ensembles de réponses

Grandes Surfaces...

Comment on traite les clients !

Comment les clients s'en ressentent-ils ?

Connaître la clientèle, c'est connaître le lien entre l'ensemble **C** des clients et l'ensemble **V** des produits vendus...

Un tableau de correspondance **C x V** ?

En général, l'ensemble des achats faits par un même individu client, « c » n'est pas connu : la caisse n'enregistre que des « paniers », « p », ce qu'un client a pris en une seule fois : on analyserait donc une correspondance **P x V**.

Mais certains magasins ont des programmes de fidélité : ce sera l'occasion de recueillir des données sur les clients tels qu'ils se conduisent sur une période. Tous les paniers successivement remplis par un même client « c » peuvent être cumulés sur une seule ligne « c » du tableau de correspondance : ainsi, on aurait **C x V**.

Mais que mettre en colonnes ? On ne peut prendre pour **J** l'ensemble **V** des articles vendus par le magasin : l'ensemble des produits vendus est trop grand... il change d'un jour à l'autre... l'attrait d'un même produit varie avec son prix ; mais on peut considérer diverses partitions de **V** : l'ensemble **R** des rayons du magasin : alimentation, vêtements, quincaillerie ; ou, plutôt, les subdivisions de l'alimentation : fruits et légumes frais, viandes, conserves, laitages... en subdivisant encore : conserves de luxe, conserves ordinaires...

L'analyse de correspondance permet de vérifier, d'améliorer une première partition, choisie a priori. Voici comment !

Lignes : ensemble **C** des clients (ou ensemble **P** des paniers)

Colonnes : une partition **G** de l'ensemble **V** des articles vendus, partition grossière prise a priori ;

F : une partition de **V** plus fine que **G** : **F** a priori, elle-aussi.

On analyse : **C x G**, avec en supplémentaire **C x F**, chaque élément « g » de **G** est éclaté suivant une partie « gF » de **F**... d'après les résultats de l'analyse on fait une classification de

F : d'après cette classification, proprement critiquée, on peut remanier la partition **G**, substituer à **G** une nouvelle partition **G'** de **F** : certaines classes de **G** étant subdivisées, d'autres fondues.

La partition **F** pouvant elle aussi être modifiée, en supprimant la distinction entre un « f1 » et un « f2 » que l'analyse a montré être très proches... ou encore, en faisant au sein d'un « f » une distinction qu'on avait hésité à faire, de peur d'être perdu avec un **F** de cardinal trop grand : car même si la puissance de calcul est grande aujourd'hui, on ne peut analyser que ce qu'on a... or qu'y a-t-il dans un tableau lacunaire ?

L'attrait d'un même produit peut varier au cours du temps : ce qui est cher en novembre peut être bon marché en juillet : de ce point de vue la partition **F**, ou même la partition initiale **G**, pourra tenir compte du temps : ne pas mettre dans la même classe « tomate en Juillet » et « tomate en Novembre »... ; L'origine du produit compte aussi : « tomate de France » ou « tomate importée »...

Les

étiquettes de prix

sont un autre poème !

A quoi bon écrire 3,99 comme si le client avait peur de 4 ?

Certains, dont je suis, lisent directement 4, même si on a écrit 3,99, ou même 3,87.

Mais quel est, en général, l'effet des étiquettes ?

On songe à une typologie psychologique des clients fondée sur le jeu d'un questionnaire.

On propose à un ensemble **C** de clients un ensemble **E** de produits étiquetés... une dizaine par exemple. Et puis on demande aux clients de dire, de mémoire, le prix que porte l'étiquette « e ». Est-ce que, pour 3,87 on retient 3,80 ? ou 3,50 ? ou 3-et-quelque ?

En général, les produits à prix avantageux ont des emballages ternes ; afin que seuls les clients austères les achètent ; les autres clients allant gaiement à la couleur.

La ruse n'a pas de limite : parfois on veut faire passer un prix pour plus avantageux qu'il n'est... banal !

Mais parfois, on veut, au contraire que le prix soit vu plus haut qu'il n'est !

Exemple : un produit à 1,01 non loin d'un produit semblable à 1,17 ; celui-ci paré d'un emballage charmeur. Un client austère prendra à 1,01 sans sourire à l'emballage : mais un autre, sans même raisonner, ira à l'emballage, n'ayant rien perçu de la différence camouflée des prix... (car, pour certains regards, $0,99 \neq 1,01$; mais $1,01 \approx 1,17$).

Quel est le véritable effet de toutes ces ruses, sur les uns et sur les autres ?

Ou, pour revenir à l'analyse des paniers d'achats : à côté de « 1,02 », « 1,17 » a-t-il moins de succès que « 0,99 » ? Et auprès de qui ...

Le magasin ne peut pas laisser échapper les clients austères (d'autant que l'austérité d'un client n'est pas constante : il peut craquer... et un client austère peut ne l'être que pour lui-même, non pour les siens...); mais le magasin ne veut pas qu'en cultivant les austères, il perde quoi que ce soit de ce qu'il attend des autres !

Mais aussi, un client n'est pas toujours un individu, ce peut être une famille : l'une économe, l'autre fantaisiste ! et un même individu, qui a sagement fait son marché, peut craquer devant les reflets d'un emballage.

Qu'en résultera-t-il pour la typologie des produits ?

Y-a-t-il des produits de luxe qui séduisent plus particulièrement certaines catégories de clients austères ? La typologie mettra ces produits de luxe près des produits austères, au moins sur tel ou tel axe issu de l'analyse de correspondance ?

Sur

l'analyse de la parole

règne, parmi les linguistes, la phonologie : le discours est une suite de phonèmes ; comme le texte d'une langue alphabétique est une suite de lettres. Mais en observant des sons enregistrés, on voit d'autres éléments avec, de par le jeu même de la voix émise, des anticipations d'un son par un autre. Puisque les langues changent par degrés, au cours des siècles, une opposition dichotomique doit pouvoir apparaître ou disparaître. Prenons l'exemple de la « ceta » de l'espagnol. Par l'histoire et l'orthographe, ce son, généralement reconnu pour non voisé, se rattache au « zed » voisé du français. La diachronie, ici comme ailleurs oblige à reconnaître une zone de transition. De même avec l'étude synchronique des dialectes. Un linguiste (dont nous regrettons de ne pas connaître le nom), travaillant dans l'aire du franco-provençal, nous a dit que les variations de prononciation d'un village à l'autre étaient telles que les locuteurs pouvaient se vanter de faire, dans la prononciation de deux sons, une distinction absente du parler d'une localité voisine ; cependant que le linguiste ne l'entendait ni ici, ni là !

(Est publié dans les Cahiers de l'Analyse des Données, Vol. XIV, 1989, un article sur les Fondements de la Reconnaissance de la Parole, avec une bibliographie des recherches analogues déjà parues avant cet article.)

Quant à la

Natalité,

chacun, selon ses convictions rappelle des devoirs ou revendique des droits : il n'est pas au programme de notre Colloque de disserter en cette matière ; nous nous bornerons à dire que le ferme énoncé des principes suffit ; la répétition n'ayant pas, par elle-même, valeur de preuve : en rassemblant des données en vue d'analyses, on offre à chacun l'occasion de voir, en conscience, la prise qu'ont, sur le réel, ses propres assertions.

Voici d'abord un exemple rare, et, à ce que nous sachons, jamais évoqué, d'une donnée numérique – disons : unidimensionnelle – mesurée avec certitude.

Selon leurs convictions, les uns accusent la légalisation de l'avortement d'avoir été à l'origine de la multiplication d'une pratique ; les autres, mettent au crédit de la loi cette multiplication.

Mais considérons une pyramide des âges pour la France. Entre la largeur de la génération née au cours de l'année précédant la promulgation de la loi, et celle de la génération née au cours de l'année suivant la loi, il y a un rétrécissement qui mesure l'effet de la loi : la différence entre le nombre des avortements « après » et « avant ».

Les deux nombres sont inconnus... mais la différence se lit : car, parce que toutes les choses (mortalité naturelle, désirs des couples ou des mères isolées...) sont égales d'ailleurs (car elles n'éclatent pas en une année), joue une cause unique : la variation de la loi. Or cette cause a un effet qui n'est qu'une fraction du total (inconnu, certes, répétons-le ; mais dont il suffit de savoir l'ordre de grandeur).

Amis comme ennemis sont ainsi déboutés devant le Tribunal de la Statistique unidimensionnelle.

La cause majeure, certes multidimensionnelle, est le changement des mœurs.

A priori, ce changement échappe à toute analyse quantitative... Et bien non !

Partons de la Chanson du Bois de Chaville :

à la cueillette du Muguet, on est parti à deux... on en revient trois.

Pas de panique ! Il faudra se marier : inutile d'essayer de dissimuler pourquoi : quelques mois après la conception, ce sera le mariage... quelques mois après : plutôt quatre que neuf, assurément... la naissance. Viendront encore, au rythme annuel, deux autres naissances ... puis l'Organisation des Pays Exportateurs de Latex y mettra un terme.

Mais si l'OPEL défaille, pas de panique non plus... ce sera une famille de quatre enfants.

Affirmation gratuite ! direz-vous.

Non : projet de recueil de données dans les registres de l'État-Civil. L'individu de référence pouvant être la Naissance : les variables étant la date (pour une première naissance, conception aux congés-payés, sans doute, moins printanier mais plus fréquent qu'au muguet...); l'âge de la mère, l'état matrimonial (à la naissance ; et, d'abord, à la conception) ; plus précisément la distance entre la date éventuelle du mariage et celle de la naissance (après, ou avant...).

Maintenant, variables complémentaires ; l'année de la naissance... le lieu et la catégorie socioprofessionnelle pour la mère.

Il y a dans un tel corpus tout ce qu'il faut pour voir changer les mœurs : de la fraîcheur – innocente non, mais humaine – de la quatrième Rép. à une solitude sans tendresse : le hasard sans l'amour.

et la **Bourse** ?

Même si on le suit d'ordinaire sur un unique indice, le niveau d'une place financière, l'état d'un marché est un profil multi-dimensionnel... les rapports entre places ne se résument pas adéquatement sur les rapports mutuels de leurs seuls indices... les frémissements conjoints des marchés permettent de fixer à temps l'attention sur une place ou une valeur.

On a dans C.A.D. des analyses la plupart dues à A.G. Hathout : voici deux extraits de ces analyses que leur auteur poursuit aujourd'hui.

Tokyo

Variation des cours de 19 titres cotés à la Bourse de Tokyo, du 30 décembre 1986 au 30 mars 1990 (CAD, Vol XV – 1990 – n°2, pp. 231-238).

Analyse d'un tableau rectangulaire, (150 x 19), donnant de semaine en semaine, les cours de 19 titres cotés à la bourse de Tokyo.

L'analyse sur une période de plusieurs années appelle l'attention sur les événements fondamentaux ; et elle permet d'interpréter des écarts de faible amplitude sur les graphiques issus d'une analyse de correspondance.

A posteriori, la période apparaît marquée par deux mouvements fondamentaux :

Reprise, inscrite sur l'axe 1, des électroniques et de l'industrie lourde (acier) au détriment des bancaires et des financières ;

Reprise, inscrite sur l'axe 2, des technologiques au détriment des financières et de l'acier.

Chacune des deux reprises survient après une phase de refuge (banque Sumitomo...)

Habitué à guetter les frémissements du marché sur moins d'un an, nous avons vu, ici, se distinguer tendance structurelle et mouvements conjoncturels.

Se signale la naissance d'un nouveau grand, SONY, qui tend à devenir le n°1 mondial, non seulement dans l'électronique de loisir mais aussi dans l'audiovisuel et la communication.

Paris

Variation des cours de 38 titres cotés à la bourse de Paris, de décembre 1986 à juin 1990 (CAD, vol XV – 1990 – n°3, pp. 349-358).

Contrairement aux événements boursiers qui frappent le KABUTOCHO, il ne se passe rien que de rond au Palais BRONGNIART, comme l'atteste l'effet Guttman liant les quatre premiers facteurs issus de l'analyse de correspondance.

Rien que de rond ? La période considérée comprend pourtant un « krach » et quelques velléités de « mini-krach » ! Mais ces accidents de parcours, n'ont affecté que l'indice global de la place de Paris sans provoquer, dans le mouvement du profil des actions, de brusques changements d'orientation, tels que ceux remarqués à Tokyo.

Or et monnaie

J.-P. Benzécri vous a-t-il déjà annoncé l'apoptose du \$?

le \$ baisse... il baissera encore ! méga-inflation... le \$ ne vaudra plus rien... les USA auront largué leurs naïfs créanciers...

alors ils instaureront une nouvelle unité, un chryso\$, gagé sur l'or (matière dont personne n'est aussi riche que les dépôts des USA)

et pourquoi apoptose ? le mot vient de la biologie, et tout dictionnaire digne du titre d'encyclopédique vous en dira plus que je n'en saurai dire !

Dans la montée générale des cours de l'or, les frémissements relatifs des diverses monnaies, face au \$, peuvent annoncer l'apoptose à qui en attend la venue...

En CAD, Vol XVIII – 1993 – n°1, pp. 27-46, A.G. Hathout et Sagombaye Nodjiram considèrent : « Séries de cours et séries de quotient dans l'analyse comparée des monnaies ». À cet ensemble de séries, on adjoindrait celles des cours de l'or sur les principales places financières...

; Und so weiter, und so fern!

Bibliographie :

Traité : L'Analyse des Données:

Tome 1 : La Taxinomie ;

Tome 2 : L'Analyse des Correspondances.

Collection d'exposés élémentaires :

HIST0 : Histoire et Préhistoire de l'Analyse des Données.

ENS1 : Enseignement destiné à tous les utilisateurs.

ENS2 : Abrégé théorique, destiné à ceux qui ont, en Mathématiques, des connaissances telles que celles d'un étudiant de 1-er Cycle.

PRA3 : Analyses en Linguistique et Lexicologie : exemples et méthode.

PRA4 : Analyses en Médecine, Pharmacologie et Physiologie clinique.

PRA5 : Analyses en Sciences Économiques.

tous ces livres ont été édités par DUNOD,

à l'exception de PRA4, publié par STATMATIC³.

Un exposé général en langue anglaise a été rédigé avec T.K.Gopalan :

Correspondence Analysis Handbook, ed. Marcel Dekker Inc.

De 1975 à 1997 ont paru XXII volumes de la revue :

C.A.D. Les Cahiers de l'Analyse des Données, DUNOD, Paris.

Les articles de **physique** cités viennent du site :

(arxiv :hep-ph/0610012 ; en France, site : <http://fr.arxiv.org>)

Recueil de programmes conçus pour tous les ordinateurs actuels :

Correspondence Analysis and Data Coding with Java and R, Fionn Murtagh⁴,
ed. Chapman & all / CRC

³ NdR : 4, rue de Fécamp, 75012 Paris, statmatic@wanadoo.fr.

⁴ NdR : <http://astro.u-strasbg.fr/~fmurtagh/mda-sw/correspondances/>