

TEMOIGNAGE :
PREMIER CONTACT AVEC L'A.M.D.

J.J. BENET

=====

RESUME : A partir de son expérience limitée de profane découvrant l'Analyse multidimensionnelle, l'auteur en présente les principes, et l'intérêt de cette méthode maintenant accessible à tout possesseur d'un micro-ordinateur.

*
* *

Profane, utilisateur novice de l'Analyse Multi Dimensionnelle (A.M.D.), je souhaite faire partager mes réflexions concernant la découverte de cette discipline.

"Les charmes de Paris au mois d'août"

Un des avantages d'être "juilletiste" est de permettre de consacrer dans le calme d'un laboratoire en léthargie estivale (théorique), l'essentiel de son activité à un gros dossier.

L'A.M.D. fait partie de ceux-là.

Disposant depuis peu d'un logiciel de traitement statistique des données (STAT-ITCF, créé par Microsoft, sous le nom de microstat, puis traduit, corrigé, complété par l'I.T.C.F. ; programmes d'A.M.D. développés d'après l'ouvrage de Foucard) adapté à notre ordinateur du Laboratoire associé I.N.R.A. (un vieux coucou, largement dépassé par certains ordinateurs familiaux actuels), il était possible de consacrer "quelques heures" à cette découverte.

En comptant aussi les heures de lecture d'ouvrages divers, cela représente quand même une centaine d'heures (c'est pas mal) pour ce qui n'est qu'une toute première approche (c'est peu).

• "L'A.M.D., qu'est-ce que c'est ?"

Question piège, autant que "qu'est-ce que l'épidémiologie ?"

Envisageons selon plusieurs points de vue (... : autant de dimensions).

- ◊ C'est tout d'abord une affaire de chapelle : les statisticiens ont viscéralement l'esprit de clocher, et ignorent, méprisent délibérément les "autres". Les règlements de compte parsèment les entre-lignes des publications.

Il est plus facile de parler d'Ecole française, menée par BENZECRI, le phare de l'analyse factorielle des correspondances.

* Ecole nationale vétérinaire - Service des Maladies contagieuses -
94704 Maisons-Alfort Cedex.

On peut l'opposer (dans tous les sens du terme), soit aux autres (en bloc !), soit d'une façon plus précise à une Ecole d'inspiration anglo-saxonne.

Laissons-les à leurs explications, pour ne retenir que l'essentiel : l'abondance de documents des premiers, rendant accessible à tous les niveaux de compréhension cette chose complexe, l'A.M.D. De nombreux ouvrages couvrent aussi bien le domaine théorique qu'ils fourmillent de cas expliqués en détail. Le talent pédagogique est uniforme : on sent la patte d'un maître qui a imposé une rigueur dans l'exposé, répondant au souci d'être bien perçu. Les autres font partie de l'élite, et vous le font bien sentir : sans un bagage minimum de bac maths + 4 ans de maths, point de salut (... j'ai craqué).

L'A.M.D. ? ... Mais c'est très simple

C'est vrai, "il n'y a qu'à" rentrer les données, choisir le programme approprié, et laisser mijoter quelques heures. Ensuite, "il n'y a qu'à" lire les listages (on ne dit plus "listing") :

- . pour savoir quelles sont les conditions de validité des résultats,
- . pour tenter d'interpréter les premières sorties.

Généralement, ce n'est pas satisfaisant ; on recommence donc, en bricolant par ci par là (un codage modifié, des sujets ou des variables mis de côté, selon une stratégie judicieuse).

Finalement, on parvient tout de même à une projection intéressante, qui ouvre des perspectives sur le sujet étudié.

Ce n'est pourtant pas fini ! Il faut encore vérifier si ces résultats sont stables et détecter de possibles artefacts.

C'est laborieux, certes, comme toutes les tâches rigoureuses, de laboratoire. Mais simple dans sa démarche.

L'A.M.D., ce n'est que de la cuisine

Les opposants de l'A.M.D. traitent les étapes précédentes de cuisine : c'est vrai, ça mijote ; c'est vrai, on retire, on ajoute des morceaux, tout cela pour parvenir à ses fins... Cela ne fait pas noble : une bonne formule mathématique, ça au moins c'est sérieux.

C'est clairement affiché par les tenants : l'A.M.D., c'est de la cuisine ! Mais attention, c'est un Art ! C'est-à-dire qu'il faut du goût (celui des choses bien faites), de l'esprit (celui des choses nouvelles) pour mener à ce qui, tout compte fait, est une sorte de création.

Mais enfin qu'est-ce que c'est ?

Un moyen d'appréhender une réalité complexe dans sa totalité, un outil d'OBSERVATION.

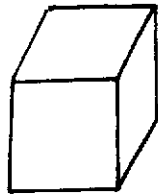
Prenons un problème simple, un phénomène étudié, qui évolue selon les caprices d'une variable. A chaque valeur de l'une correspond une valeur de l'autre. L'être humain n'a encore rien fait de mieux pour percevoir immédiatement cette réalité que d'en faire une représentation graphique : x en abscisses, et y en ordonnées. L'observation de la courbe permet ensuite l'interprétation.

C'est un moyen d'accéder à la connaissance d'un phénomène, très avantageux, car il permet de condenser sur une feuille de papier une information qui peut être abondante. Toutefois, il peut être dangereux et conduire à des erreurs d'interprétation si les données sont mauvaises, ou si la représentation n'est pas bonne (erreur d'échelle par exemple).

Mais cela, tout le monde le sait.

Le problème se complique pour trois variables simultanées : il faut trois axes orthonormés, donc un volume. A moins d'utiliser une astuce de représentation, "en perspective", c'est-à-dire en calculant les coordonnées de façon à rendre compte de la réalité tridimensionnelle, tout en la projetant sur une feuille de papier.

La figure suivante est :



- . un cube
 - . un polyèdre suggérant l'idée d'un cube.
- (Cherchez la bonne réponse)

Avec quatre variables (quatre dimensions) s'arrêtent nos talents d'analyse et de synthèses simultanées. Alors pour " n " variables...

L'A.M.D., c'est ÇA, ni plus ni moins : une projection sur 1 plan (parmi d'autres) d'une représentation multidimensionnelle.

Les génériques de la télévision se complaisent dans une débauche d'effets électroniques plus étonnants les uns que les autres. Sous nos yeux, se transforment des paysages (donc des volumes), qui évoluent en particulier selon des changements d'angle sous lequel on les regarde. Pourtant, notre écran est réellement plat (vous vérifiez tout à l'heure). Donc, tout est fait pour nous suggérer l'idée du volume, en mouvement, à partir d'une représentation sur un plan.

Pourquoi insister sur ce point, qui paraît bien banal : les dessinateurs, les peintres sont rompus à ces principes ! C'est vrai, et l'on peut ainsi apprécier leur talent... Mais pour un ordinateur (responsable de ces merveilles télévisuelles), point de talent, hormis celui des concepteurs du logiciel : tout est mathématique, ou informatique. Chaque image est construite par dérivé d'une description initiale, par exemple celle d'un volume (un cube) ; sa représentation sur l'écran est une transformation de la réalité, qui seule parle à notre perception.

Ensuite, le cube se met en mouvement : c'est-à-dire que les modifications de la description initiale résultant de transformations mathématiques, qui donneront finalement une nouvelle représentation. Cela image, par image. Nous pourrions dire, autant que d'angles de point de vue.

Pourquoi "Multi"-dimensionnel ?

Certes les problèmes des représentations en 2 ou 3 dimensions d'un objet sont complexes, mais ils se comprennent. Mais d'où viennent les autres dimensions ?

... C'est toute la difficulté du concept "multi-dimensionnel". En fait, le mot "dimension", utilisé en A.M.D. n'a pas du tout le sens trivial que nous lui donnons communément (longueur, largeur, hauteur). C'est une commodité de langage, et même de pensée, pour rendre compte de réalités appartenant au monde mathématique. La dimension mathématique, c'est une variable. Autant de variables, autant de dimensions.

Si l'on peut admettre ce point de vue mathématique, la compréhension du profane est heurtée lorsqu'on prétend projeter sous ses yeux sur une feuille de papier une image d'un "nuage" en n dimensions, qui n'existe que dans l'imagination des mathématiciens.

Si le mathématicien sait reconnaître sur la feuille une projection de son nuage, le profane lui, voit une réalité tangible : des points, repérés par des axes orthonormés.

C'est bien là le piège :

"CE QU'ON VOIT N'EST PAS CE QUI EST"

Ce qu'on voit n'est qu'une image, une transformation de la réalité (complexe) de manière à nous la rendre perceptible.

Reprenons : à quoi sert l'A.M.D. ?

L'A.M.D. sert à appréhender une réalité multiple, dans son ensemble.

Un tableau de données (de 1 lignes et c colonnes) est un être (fictif) multidimensionnel. On peut en avoir décidé ainsi au départ, en jugeant intéressant de rapprocher et comparer les différentes variables (colonnes). Ou bien, on peut ne pas avoir d'a priori, et simplement chercher si un tableau de données collectées possède une structure particulière (à moins que les données ne soient réparties au hasard dans le tableau).

L'A.M.D. sert alors à VOIR cette réalité complexe (qui n'est pourtant qu'une abstraction !).

• Mais, comment ça marche ?

"C'est étudié pour !"

Le principe est simple : le bon vieux théorème de Pythagore ($h^2 = a^2 + b^2$), un peu "travaillé", simplement. C'est en fait sur ce point que les Ecoles se querellent.

Le premier contact avec une A.M.D. est celui de la stupeur : en fonction d'axes sans dénomination, sont représentés pêle mèle, des individus et des variables. Incompréhensible au premier abord. Choquant au deuxième.

Pourtant, c'est très simple (ou tout au moins tel qu'expliqué par Fenelon, auquel la démonstration est empruntée).

Prenons l'exemple d'un tableau de données, dans lequel pour chaque individu ont été mesurés la taille (T) et le poids (P).

Supposons que le choix des unités de mesure a été résolu (mètre ? centimètre ? kilogramme ? gramme ?), de même que celui du système de représentation (combien de centimètres pour représenter une unité de poids ou de taille ; choix des axes : P horizontalement ou verticalement ...?). Ce problème est général et concerne toute représentation graphique. Il est pourtant très important puisqu'il détermine les caractéristiques de l'image obtenue.

L'image obtenue nous est familière. En fonction de l'une des variables (taille) est représenté le point correspondant aux valeurs de l'autre variable (poids). Elle nous montre un "nuage de points".



Pourtant, insistons bien : ce nuage n'est qu'une image parmi toutes (les images possibles), en fonction de divers choix de modalités de représentation. Quant au phénomène représenté, c'est une abstraction (la relation entre T et P), qui bien que ne possédant que 2 dimensions, n'a aucune existence physique dans l'espace qui nous est familier. Si la taille est bien l'une de nos 3 dimensions triviales le poids est à l'évidence une composante d'autres dimensions.

Cette représentation est commode, car toutes les valeurs du tableau sont représentées par un point ; elle me permet donc d'appréhender globalement un tableau (ici à 2 dimensions).

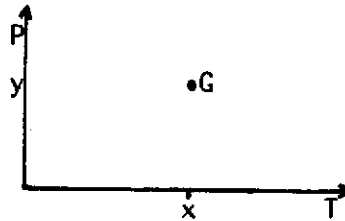
Elle demeure encore trop complexe pour permettre des comparaisons. Il faudrait la REDUIRE à quelques valeurs qui résument toute l'information du nuage.

N.B. : Si les concepts mathématiques (même élémentaires) effarouchent certains, ils peuvent passer directement au paragraphe "L'A.M.D., une révolution dans le processus des connaissances".

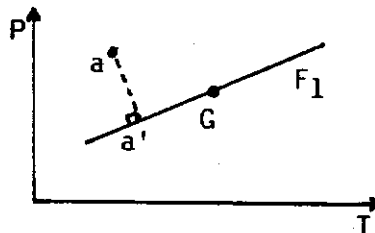
"REDUCTION" : la réduction des données dans la statistique classique conduit à déterminer les paramètres de position (moyenne), et de dispersion (variance).

Cela moyennant un a priori impératif : celui de la distribution des valeurs selon une loi commune. Profitons en pour souligner que l'A.M.D. n'a pas besoin de cet a priori (rarement vérifié en biologie).

Nous pouvons réduire le nuage à son centre de gravité :



Pour retrouver la notion de tendance, il faut un axe ; c'est l'axe passant par le centre de gravité, et qui rend le mieux compte de la morphologie du nuage.



La construction de l'axe résulte d'une contrainte imposée, de réduire au minimum les écarts entre cette droite et l'ensemble des points du nuage (les mathématiciens disent : "distance").

Considérons le point a. Il est à la distance a-G du point G, qui le situe dans le nuage par rapport à G.

La droite F_1 permet de projeter le segment a G par le segment a'G : ce dernier "représente" sur l'axe F_1 la distance a G. Le point a' est une image de a sur F_1 .

On peut refaire la démonstration pour tous les points du nuage : l'axe F_1 est porteur de l'information concernant le nuage initial. C'est un axe factoriel.

Seulement, il ne porte pas toute l'information. Il reste une variabilité résiduelle, correspondant à la distance aa'. On peut donc construire un deuxième axe, perpendiculaire au premier, qui sera le deuxième axe factoriel, celui de la variabilité résiduelle.

Ce deuxième axe F_2 est lui aussi porteur d'information.

On peut donc comparer la quantité d'information de chacun des axes : c'est une fonction du poids de l'individu (c'est-à-dire des valeurs des variables pour l'individu), et de sa distance au centre de gravité, c'est donc une inertie.

Chacun des deux axes est donc porteur d'une certaine "INERTIE", expression empruntée à la physique (en raison du principe de la méthode), pour rendre compte de l'information.

Ici, la totalité de l'inertie portée par les deux axes est de 100 p. cent, car il n'y a que deux variables. Avec d'autres variables, le total de deux axes serait très inférieur.

Ces axes sont donc totalement différents des axes de coordonnées initiaux (x, y). Nous devons les appréhender comme une forme de représentation du nuage initial ; on leur attribue une notion de qualité de représentation (entre autres) selon le pourcentage d'information (inertie) dont ils sont porteurs.

Etendons ce processus à plusieurs dimension (plus de 2).

Pour 3 variables, nous imaginons facilement un nuage de points, comme un volume, dans un espace abstrait. Nous percevons fort bien que la projection du nuage sur un plan ne donnera qu'une image de la réalité.

Supposons un nuage ayant la "forme" d'une banane. Une projection sur un plan placé selon son plus grand axe donnera une information importante, dont l'image est assez difforme : on retrouvera la forme allongée, en croissant.

Une autre projection, selon un plan perpendiculaire au premier apportera une information supplémentaire, concernant la forme circulaire. Nous avons perdu, en revanche, les premiers éléments d'information.

Pour un nombre de variables supérieur à trois, nous pouvons continuer d'imaginer un nuage de points, dans un espace abstrait comportant autant de dimensions que de variables.

Mais chaque projection ne comporte que deux axes factoriels. Donc l'information est réduite par rapport à la réalité. La représentation ne sera satisfaisante qu'en étudiant de nombreuses représentations possibles, pour vérifier que l'image perçue n'est pas trop éloignée de la réalité.

Arrêtons là cette présentation élémentaire, abusive dans sa simplification, pour renvoyer aux ouvrages (excellents) signalés par ailleurs, qui sont plus justes, plus pédagogiques.

■ L'A.M.D., une révolution dans les processus de connaissance

L'A.M.D. est un concept bien connu des spécialistes de cette discipline mais qui peut désormais devenir accessible au plus grand nombre grâce aux performances des micro-ordinateurs. Pour ces nouveaux utilisateurs c'est un concept nouveau.

◇ L'A.M.D. est un CONCEPT

C'est une façon de concevoir une réalité dans son ensemble, quelle qu'en soit la complexité. Elle s'oppose à la conception analytique, ou linéaire, qui ne peut appréhender que les réalités pouvant être réduites à quelques variables seulement, plus aisément perceptibles.

◊ L'A.M.D. c'est aussi une méthode et des techniques

Cette méthode, ces techniques permettent d'utiliser le concept, de l'appliquer en pratique.

- ◊ Cette nouveauté nécessite une tournure d'esprit particulière, pour "admettre" d'évoluer dans une dimension du raisonnement qui ne soit plus linéaire. De ce fait, elle a une valeur formative, en laissant l'esprit libre à l'observation, en rejetant tous les a priori. Nous entraînant dans un monde étrange, qui ne nous est pas familier par ses rapports (ses dimensions), mais qui contient des objets qui eux le sont, l'A.M.D. nous place intensément en situation d'observation.
- ◊ Elle a elle-même valeur pédagogique : forçant à "admettre" ses postulats de départ (concernant les particularités des représentations graphiques) non seulement elle forme par sa pratique à l'appréhension de l'A.M.D., mais elle conduit à prendre conscience de la précarité des représentations que nous avons de la réalité.

La réalité à laquelle nous sommes confrontés est complexe.

Les processus usuels de raisonnement sont de type analytique, linéaire : nous ne pouvons appréhender qu'une toute petite partie des mécanismes à la fois, et jamais de façon synthétique. Nous avons réussi à expliquer ce qui était à notre portée par ces modalités intellectuelles, mais la limite est inéluctable : la complexité ne peut qu'échapper à ces processus conventionnels de travail.

Le pire est bien que nous n'en avons pas conscience : aussi longtemps que nous travaillons, nous croyons pouvoir un jour aboutir. Les processus de représentation de la réalité nous donnent une image très fautive. Nous ne savons pas combien cette image est mauvaise. Nous ne savons pas combien ce processus est mauvais.

L'A.M.D. permet non seulement d'accéder à une connaissance jusque là interdite, mais elle ouvre les yeux sur l'extraordinaire relativité de notre représentation du monde : elle fournit DES images de la réalité, dont on sait très bien qu'elles ne sont que des représentations (puisqu'elles sont des projections sur une feuille, et les conditions particulières de lecture nous forcent à nous en souvenir).

L'A.M.D. est critiquée justement de ce point de vue : les images produites peuvent être fausses, par les biais de représentation, mais aussi les biais de conception. En effet, le choix des unités de mesure, ou des modalités de codage équivaut à des hypothèses concernant la conception de la réalité étudiée. Les conséquences peuvent effectivement être très lourdes. Les critiques sont donc justifiées...

Mais l'A.M.D. a justement l'avantage de clairement montrer à l'observateur tous les pièges de la perception :

- . non seulement ceux de la représentation,
- . mais aussi ceux de la conception.

Les autres modalités immédiates sont d'autant plus dangereuses que l'on n'a pas conscience de ces biais, qui sont pourtant rigoureusement les mêmes.

- ◊ Prenons un seul exemple (parmi tous ceux amplement développés dans les ouvrages d'A.M.D.), celui de l'effet de la Trinitrine selon l'état initial du malade (en insuffisance cardiaque) : selon les auteurs, ce médicament pouvait avoir un effet favorable, ou désastreux. Le recensement de différents signes cliniques initiaux n'avait pas permis malgré toutes les ressources de la statistique classique de déceler quelle pouvait être la typologie des malades pour qui la trinitrine était contre-indiquée. On en restait au constat de résultats contradictoires selon les auteurs et inexpliqués.

L'A.M.D. a permis de connaître cette typologie, et donc de modifier le pronostic de cette thérapeutique, ce que n'avait pu obtenir une démarche conventionnelle.

- ◊ Cet exemple permet aussi de situer le bon vieux "sens clinique" des praticiens (médecins, vétérinaires) comme une véritable approche multidimensionnelle. Elle s'est constituée au fil de l'expérience, et elle a fait ses preuves. Elle a l'inconvénient d'être incommunicable.

L'A.M.D. elle, est accessible à tous. Elle est communicable : on peut l'apprendre. C'est aussi un "media" : elle permet la communication de réalités complexes.

Si son principe est simple, elle demande toutefois beaucoup d'expérience pour savoir l'utiliser et beaucoup de travail pour obtenir des résultats satisfaisants.

CONCLUSION :

Cette expérience de l'A.M.D. permet de vérifier un phénomène bien connu des spécialistes : l'avènement informatique conduit à un bouleversement des rapports entre les choses et les gens. L'A.M.D., de conception pourtant ancienne, est désormais accessible à tout possesseur d'un micro-ordinateur. Le remarquable outil d'observation va redonner leurs lettres de noblesse à toutes les disciplines faisant appel à l'observation, en leur ouvrant les champs, jusque là interdits, de la complexité.

Ainsi se trouvera posée avec une acuité inégalée, la position du savant par rapport à la réalité qu'il a charge d'étudier. Les années à venir risquent d'être passionnantes.

*
* *

Nous remercions Messieurs Josse (J.), Monicat (F.) tous deux du Centre d'Ecopathologie multi-espèce Rhône-Alpes (Villeurbanne), et Moutou (F.), Tuffery (G.) pour leur aimable lecture critique, et leurs nombreux conseils et suggestions.