METHODES ET MODELES STATISTIQUES

PIEDNOIR Jean-Louis
I.G.

LA STATISTIQUE EN SECTION DE TECHNICIEN SUPERIEUR

Un programme de statistique est enseigné dans certaines sections de techniciens supérieurs. De plus en plus, les représentants des professions dans les organes consultatifs de l'éducation nationale demandent une initiation aux méthodes statistiques qui sont utilisées de façon croissante dans des secteurs toujours plus nombreux de la vie économique. Les progrès techniques facilitent l'enregistrement de données et ont abaissé considérablement le coût et la difficulté de leur traitement. Le développement des conduites de qualité, les besoins de connaissances précises dans le secteur tertiaire poussent également à un usage accru des techniques existantes.

La statistique met souvent le professeur du second degré mal à l'aise. Il y a peu de temps qu'elle s'est infiltrée dans les programmes, y compris ceux des collèges. Dans les cursus de mathématiques des universités, elle est peu enseignée voire absente, sauf en option. Aussi l'initiation aux méthodes et aux modèles de la statistique est en général embryonnaire. Les morceaux de statistique enseignés dans les lycées en section B et D par exemple, ou en BTS, apparaissent de deux types. Premier type : des choses élémentaires, triviales, que l'on peut apprendre tout seul, qui se prolongent par des définitions d'indicateurs telle la moyenne, la médiane, qui paraissent arbitraires. Deuxième type : des procédures faisant appel aux modèles probabilistes et donc complexes ; on est alors réduit à enseigner des recettes. De toute façon, dans les deux cas les exercices prévus pour le contrôle des connaissances sont débiles.

Les élèves, par contre, apprécient souvent un enseignement de statistique. Il s'agit pour eux d'une branche nouvelle des mathématiques et ceux qui ont eu des difficultés dans la discipline ont l'impression de pouvoir réussir sur un domaine inexploré pour eux. De plus l'aspect nécessairement pluridisciplinaire de la statistique crée un intérêt supplémentaire. On choisit, pour illustrer un cours de statistique, des exemples issus d'autres disciplines. Cette démarche montre l'aspect opérationnel des mathématiques.

En classe de technicien supérieur, il ne s'agit pas de transformer des étudiants en statisticien mais de montrer la pertinence des méthodes statistiques pour atteindre une certaine connaissance sur des populations données. A partir de quelques situations courantes, on apprend une démarche, on sait interpréter un résultat. Cela permettra au futur technicien de pouvoir dialoguer avec des spécialistes quand il faut traiter des cas plus complexes.

L'objectif poursuivi par l'exposé ci-dessous est, à partir d'exemples variés sur lesquels des méthodes statistiques ont été réellement appliquées, de faire l'analyse des situations pour lesquelles la statistique peut être employée, puis de présenter quelques modèles statistiques. Bref, on observe les moeurs de la tribu des statisticiens pour rentrer après dans un mode de pensée. Ensuite, on fera quelques remarques sur la validité des méthodes statistiques pour terminer sur quelques considérations de nature plus pédagogique.

QUELQUES SITUATION CONCRETES

Exemple 1 – On se propose d'étudier le comportement des candidats dans les différents baccalauréats généraux en 1989. Pour cela, on recueille pour chaque candidat l'ensemble des notes qu'il a obtenu aux différentes épreuves. Que peut-on tirer de cette masse de données pour tenir un discours sur les caractéristiques de chaque baccalauréat.

Exemple 2 – Un historien veut étudier le vocabulaire des hommes politiques au début du siècle et voir comment l'appartenance politique influe sur le vocabulaire employé. Pour cela, il sélectionne un certain nombre de mots et pour chaque homme politique il compte dans ses discours le nombre de fois où ces mots sont employés et note son appartenance politique. Comment représenter les rapports réciproques entre l'étiquette politique et la fréquence des mots employés.

Exemple 3 – Sur un territoire donné, un archéologue dispose d'un lot de haches néolithiques. Elles sont probablement de type et d'origine diverses. Pour chacune d'entre elles, il note des caractéristiques de formes, de composition chimique de la pierre utilisée, les principales dimensions. A partir de ces données est-il possible de conclure à une certaine homogénéité de la population ou au contraire peut-on détecter des sous-populations qui pourraient ensuite orienter les recherches à venir ?

Exemple 4 – Un commerçant reçoit une caisse de cartouches à vendre. Il veut en connaître la qualité. Pour cela il en prélève au hasard quelques unes et les expérimente. Au vu des résultats obtenus sur cet échantillon va-t-il accepter ce lot ou au contraire le renvoyer à son fabriquant ?

Exemple 5 – Une machine automatique fabrique des pièces qui sont censées être toutes identiques. De temps à autre un opérateur en prélève quelques-unes, les mesure. Va-t-il déclarer, après avoir procédé à ces essais, que la machine fonctionne bien ?

Exemple 6 – Un lots d'équipements est supposé représentatif de toute la population des équipements de la même espèce. Pour chacun d'eux on mesure la durée de vie. Que peut-on dire, du point de vue de la fiabilité, sur les équipements en question ?

Exemple 7 – Un gaz est formé de molécules. Pour chacune d'entre elles on peut déterminer à un instant donné la position, le vecteur vitesse, on peut aussi connaître le nombre de chocs entre deux instants donnés. A partir de ces mesures est-il possible de caractériser le gaz en question ?

Exemple 8 – Un radar envoie des signaux électromagnétiques, il reçoit des échos ; certains sont ceux de la cible recherchée, si elle existe, les autres sont considérés comme parasites. Quelle procédure mettre en route pour avoir la meilleures détection possible sur l'écran sans faire apparaître les signaux parasites appelés bruit ?

Exemple 9 – Au vu d'un diagnostic classique, un médecin est capable d'assigner une probabilité pour que le malade étudié ait telle ou telle maladie. Il fait ensuite procéder à des analyses. Comment modifie-t-il les probabilités données précédemment pour tenir compte des résultats des analyses.

ANALYSE DES SITUATIONS ETUDIEES

On remarque qu'aucune des situations précédentes n'est justiciable d'un schéma de causalité simple. On effectue à chaque fois plusieurs mesures et elles sont toutes différentes même si les facteurs contrôlés sont constants. La statistique se nourrit de cette variabilité. Dès qu'elle existe on peut songer à utiliser des méthodes statistiques.

Dans les exemples étudiés on peut identifier une population, appelée $\mathcal{C}$, et des individus. Sur chaque individu on effectue une mesure. A partir de ces mesures, on cherche à dire quelque chose sur la population $\mathcal{C}$; ou en d'autres termes à effectuer une mesure sur la population. Là est la démarche fondamentale de la statistique. On peut commencer à la formaliser selon le schéma ci-dessous.

LA DEMARCHE STATISTIQUE

Soit $\mathcal{C}$ une population, $I_n = \{1,2,....,n\}$ un ensemble d'individus de $\mathcal{C}$. On a $I_n \subset \mathcal{C}$ ou $I_n = \mathcal{C}$.
On appellera X l'ensemble dans lequel les individus prennent leurs mesures et x la mesure ; $x : I_n \rightarrow X$.

Sur $\mathcal{C}$ on cherche une mesure. Soit A l'ensemble dans lequel cette mesure est prise. On cherche donc $y_{\mathcal{C}} \in A$, $y_{\mathcal{C}}$ sera calculé à partir des mesures sur les individus. Chercher une procédure statistique, c'est trouver une application $g_n : X^n \rightarrow A$.

L'art de la statistique, c'est donc de trouver la ou les bonnes applications g_n. Pour cela il faudra évidemment introduire d'autres éléments.

Au niveau du langage on parlera aussi de collectif au lieu de population, d'éléments du collectif au lieu d'individus. Cela est mieux adapté à certaines situations. On peut remarquer que cette structure est très courante dans les sciences. En sociologie on a une population et des individus. En biologie un être vivant est composé de cellules. En chimie un corps pur est fait de molécules, etc...

La démarche statistique existe dans la vie courante. D'une façon intuitive, tout commerçant fait de la statistique comme Monsieur JOURDAIN faisait de la prose, sans le savoir. C'est bien à partir de l'observation des ventes dans les semaines précédentes que celui-ci procède, produit par produit, à ses commandes. Cet exemple sera repris par la suite.

On peut, sur les exemples précédents, déterminer les ensembles I_n, X et A et donc l'application x.

Exemple 1 : Prenons cinq épreuves écrites, $E = \{A,B,C,D,E\}$ l'ensemble des baccalauréats généraux, $X = E \times \mathbf{R}^5$, A sera l'ensemble des discours possibles (cf. étude sur le baccalauréat).

Exemple 2 : Soit I avec Card I = p l'ensemble des formations politiques, si k est le nombre de mots choisis, $X = I \times \mathbf{N}^k$. Une proximité entre une formation et un mot peut être exprimé comme le nombre de répétitions d'un mot dans une formation politique donnée. Alors $A = (\mathbf{R}^+)^{pk}$.

Exemple 3 : On posera J_1 l'ensemble des caractères qualitatifs de forme, J_2 l'ensemble des caractères qualitatifs géologiques et on suppose que l'on procède à mesure, donc $X = J_1 \times J_2 \times \mathbf{R}^k$.

On cherche des sous-populations, A est donc l'ensemble des partitions de I_n.

Exemple 4 : Si la cartouche fonctionne on notera 1, sinon 0. p sera la proportion de bonnes cartouches dans le lot global. $X = \{0,1\}^n$, $A = [0,1]$.

etc...

LE COLLECTIF ET LES MESURES

Le collectif n'est pas n'importe quelle collection d'individus. Il doit être relatif à une réalité cohérente ainsi que l'illustre le contre-exemple suivant. On veut étudier l'efficacité d'un médicament sur une collection d'individus composée d'hommes et de femmes. On notera H l'ensemble des hommes, F celui des femmes, T l'ensemble des individus traités, $\overline{T}$ celui des individus non traités, G l'ensemble des individus guéris, $\overline{G}$ l'ensemble des individus qui restent malades. Sur la population globale, on note les effectifs suivants :

	G	$\overline{G}$
T	120	50
$\overline{T}$	200	100

Au vu de ces résultats le médicament paraît efficace. Effectuons le même travail sur les sous-populations hommes, femmes.

Hommes

	G	$\overline{G}$
T	8	10
$\overline{T}$	60	60

Femmes

	G	$\overline{G}$
T	112	40
$\overline{T}$	140	40

L'examen de ces résultats montre à l'évidence que sur chacune des sous-populations le médicament est dangereux. On voit immédiatement que les hommes et les femmes réagissent de façon très différente à la dite maladie et que la proportion de personnes ayant subi le traitement est très différente d'une sous-population à l'autre. Cet exemple montre a contrario que la définition d'un collectif et de ses différents sous-collectifs nécessite une réflexion a priori. Le "piège" précédent est un exemple de ce qui peut arriver si les définitions ne sont pas précises et les procédures (voir plus loin) non rigoureusement observées.

Simpson fournit dans son paradoxe un exemple analogue. Il s'agit d'étudier si, dans sa sélection, une université favorise les hommes au détriment des femmes. Les résultats globaux tendraient à le montrer.

Mais il existe une "variable cachée" : le département. En effet, on s'aperçoit que dans chaque département la sélection favorise les femmes. Dans le tableau ci-dessous le numérateur donne pour chaque sexe le nombre de candidats sélectionnés, le dénominateur le nombre total de candidats.

	Hommes	Femmes
Département A	512/825=0,621	89/108=0,824
Département B	22/373=0,059	24/341=0,070
Total	534/1198=0,446	113/449=0,252

La mesure sur le collectif déduite des mesures sur les individus est aussi appelée résumé statistique. D'un point de vue naïf, on peut dire que le résumé extrait l'information interne dans les mesures faites sur les individus et qui intéresse la population. Il en résulte que le résumé doit être peu sensible à la présence ou à l'absence d'un individu particulier donné. Il doit y avoir une certaine stabilité du résumé par rapport aux variations inter-individuelles. En effet, un collectif est un entité distincte des individus qui le composent, il est d'un autre ordre. Le commerçant évoqué plus haut le sait bien intuitivement. Une seule mauvaise journée ne l'amène pas à modifier ses commandes.

Pour trouver le bon résumé statistique, il faudrait, cela paraît évident :

1/ Savoir ce que l'on cherche sur le collectif,

2/ Savoir ce que l'on veut faire de la mesure sur le collectif,

3/ Trouver, compte tenu du but fixé, les mesures pertinentes sur le collectif.

Il existe des cas où ces conditions sont remplies. On sait dans l'exemple 7, à partir de mesures sur les molécules d'un gaz dans diverses conditions, démontrer la loi de Mariotte : pression × volume = constante × température, loi physique qui peut aussi s'observer directement. Mais cet exemple est limite. Il n'est pas possible en général de vérifier la validité de la procédure statistique directement. S'il en était toujours ainsi, la statistique serait une quasi curiosité scientifique.

Souvent, le point 1 précédent ne peut être explicité complètement ; c'est le cas de l'exemple 1. Alors, on applique à ces situations des procédures statistiques connues, on examine les résultats. C'est l'intelligibilité de ces résultats par rapport à la situation de départ qui valide la procédure. La validation est alors a posteriori et non a priori par le modèle de départ.

En résumé, on peut dire que la statistique veut passer des éléments au collectif en éliminant les aspects individuels sans éliminer les individus.

L'ELABORATION DES RESUMES

Pour élaborer des résumés, il est nécessaire d'avoir sur le phénomène étudié des idées a priori, une connaissance sur la façon dont les données ont été recueillies. Sur les exemples précédents on va illustrer le propos.

Dans l'exemple 1 : le collectif des candidats au baccalauréat est divisé en 5 sous-populations suivant les baccalauréats. Pour la description statistique des résultats obtenus, on fera l'hypothèse que la mesure sur chaque individu est un vecteur de l'espace euclidien à 5 dimensions.

Dans l'exemple 2 : on appellera q_{ij} la fréquence d'emploi du mot j par les individus ayant l'appartenance i ($\sum q_{ij} = 1, \forall i$) et r_{ij} la fréquence des individus d'appartenance i pour l'emploi du mot j ($\sum r_{ij} = 1, \forall j$).

Alors, les vecteurs $(q_{i1}, q_{i2},...,q_{ik})$, $i=\{1,2,...,p\}$ appartiennent au simplexe δ_k , les vecteurs $(r_{1j}, r_{2j},...,r_{pj})$, $j=\{1,2,...,k\}$ au simplexe δ_p. On suppose que la distance dite du χ^2 sur ces simplexes est adéquate pour mesurer l'écart entre deux vecteurs.

Dans l'exemple 3 : on construit sur $J_1 \times J_2 \times \mathbf{R}^k$ une distance, ou au moins un indice de dissimilarité, et on s'efforcera de mettre dans la même classe les individus proches les uns des autres.

Dans l'exemple 4 : on supposera que le procédé d'obtention des individus de l'échantillon testé est le suivant : il y a tirage au hasard des cartouches essayées et chaque cartouche a la même probabilité de figurer dans l'échantillon.

Dans l'exemple 5 : on supposera que la mesure faite sur chaque pièce est le résultat d'une variable aléatoire gaussienne de moyenne μ et d'écart-type σ. Toutes les variables aléatoires sont indépendantes.

L'exemple 6 est justifiable d'une modélisation analogue ; on supposera que la variable aléatoire durée de vie est celle d'un système qui ne vieillit pas, donc suit une loi exponentielle de paramètre λ.

Dans l'exemple 8 : on prélève un échantillon de n observations, réalisations de n variables aléatoires indépendantes admettant une densité g(.) qui est la loi des échos suspects. On a : $\forall\, x \in \mathbf{R},\ g(x) = f(x - \Delta)$.

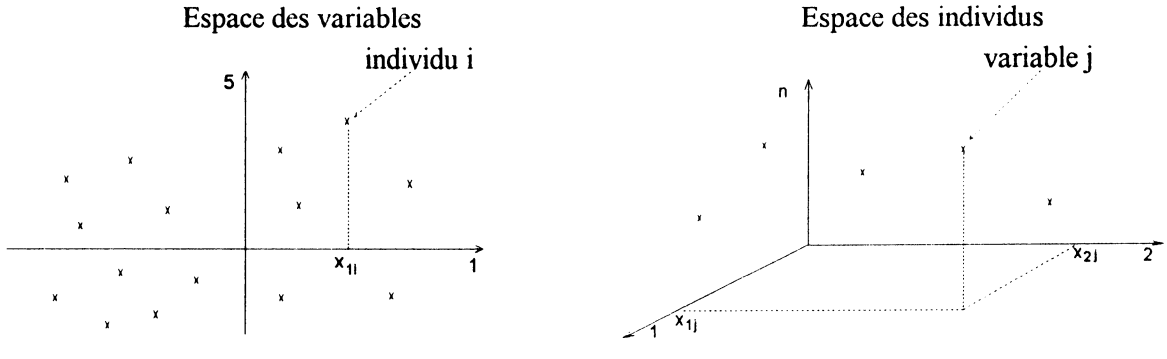
Si $\Delta = 0$ l'écho suspect n'est que du bruit, si $\Delta > 0$ il y a cible.

C'est à partir de ces hypothèses précises que l'on pourra élaborer des résumés statistiques. Leur pertinence dépend bien entendu de l'adéquation du modèle à la réalité.

On voit qu'il existe dans les exemples précédents deux grands types de modèles. Dans les exemples 1, 2, 3, la population étudiée forme tout le collectif et le modèle a priori est formalisé sous une forme géométrique, topologique ou algébrique. Les techniques mises en oeuvre forment la statistique descriptive dite aussi analyse des données. Dans les exemples 4, 5, 8, la population étudiée est un échantillon supposé extrait d'un collectif plus grand, réel ou mythique, souvent supposé infini et décrivable par une mesure de probabilité. Faire une mesure sur le collectif, c'est alors dire des choses sur la probabilité inconnue, en d'autres termes, mesurer cette probabilité. Les techniques mises en oeuvre forment la statistique inductive. C'est le jugement sur échantillon.

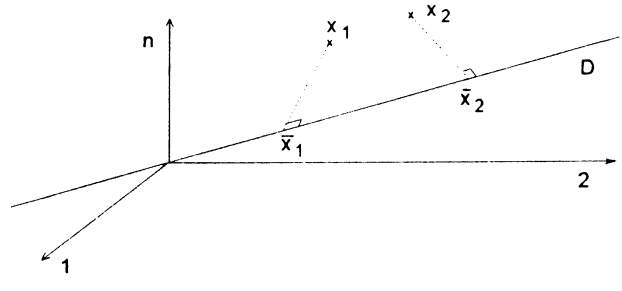
EXEMPLE DE STATISTIQUE DESCRIPTIVE

Il est hors de question de faire un panorama de l'ensemble des méthodes relevant de l'analyse des données. On se contentera des méthodes relevant du cadre euclidien. Prenons l'exemple 1. Chacun des n candidats d'un bac donné (C par exemple) est décrit par un vecteur à 5 composantes. On peut faire deux représentation duales : représenter les n individus dans l'espace E_5 des variables par un nuage de n points, ou représenter les 5 variables dans l'espace E_n des individus. Cela peut-être illustré par le schéma suivant.



A partir de ces représentations, il est possible de déterminer pour chaque variable des indicateurs de centralité, de dispersion, de chercher l'intensité de la liaison linéaire entre deux variables, d'avoir des résumés fidèles plus simples.

INDICATEUR DE CENTRALITE, DE DISPERSION, DE LIAISON LINEAIRE.



Prenons la représentation dans l'espace des individus. Il y a un cas où l'indicateur de centralité pour un ensemble de valeurs prises par une variable donnée s'impose de lui-même. C'est bien entendu le cas où toutes les observations sont identiques. On a $x_{11} = x_{12} = ... = x_{1n}$. L'ensemble des variables ayant cette propriété est la droite D n-sectrice des axes dont les coefficients directeurs sont tous égaux à 1. Si la variable X_1 n'est pas de ce type et si la description euclidienne est adéquate, on prendra comme indicateur de centralité celui de la variable qui a toutes ses coordonnées identiques et qui est la plus proche de X_1. Il suffit donc de projeter X_1 sur D. Il est facile de voir que le point $\bar{X}_1$ à toutes ses coordonnées égales à

$$\bar{x}_1 = \frac{1}{n} \sum x_{1i}$$

L'indicateur de dispersion suit aussi immédiatement du modèle euclidien : c'est la distance entre le point-variable X_1 et la droite D. On a alors

$$d^2(X_1, \bar{X}_1) = \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2$$

Pour pouvoir comparer des dispersions quand les tailles des populations diffèrent, on normalise par le facteur taille.

On a alors : $\sigma^2_{X_1} = \frac{1}{n} \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2$

S'il existe une liaison affine entre deux variables : $\forall$ i, $x_{1i} = a.x_{2i} + b$, les deux vecteurs $X_1 - \bar{X}_1$ et $X_2 - \bar{X}_2$ ont des supports parallèles. Si la liaison entre ces deux variables est proche d'une liaison affine, l'angle de ces deux variables sera proche de 0 (a>0) ou de l'angle plat (a<0). Il en résulte que l'intensité de la liaison linéaire peut être mesurer par l'angle entre les deux vecteurs ou par une de ses lignes trigonométriques.

On pose : $\rho = \cos (X_1 - \bar{X}_1, X_2 - \bar{X}_2)$. On a : $\rho = \frac{\sum (x_{1i} - x_1).(x_{2i} - x_2)}{n.\sigma_{x_1}.\sigma_{x_2}}$

ANALYSE EN COMPOSANTES PRINCIPALES

Il n'est pas facile de se représenter un espace à 5 dimensions. On va alors chercher une représentation du nuage des individus dans l'espace des variables, dans un espace à une ou deux dimensions et qui soit aussi fidèle que possible. Le cadre euclidien va permettre de répondre à la question. On va chercher le sous-espace affine à 1, 2 ou k dimensions le plus proche du nuage de points. Soit P_i le point représentatif de l'individu i, F_k un sous-espace affine à k dimensions, P'_i projection orthogonale de P_i sur F_k.

On cherche F_k tel que : $d^2(P_i, P'_i)$ soit minimum, d étant la distance euclidienne.

Un tel sous-espace passe par le point G dont les coordonnées sont les moyennes des différentes variables. Le plan minimum passe par la droite minimum. Une fois déterminé le sous-espace, on projette les points P_i sur F_k en P'_i. Le nuage des P'_i est alors le nuage le plus proche du nuage des P_i dans un espace affine à 1,2 ou k dimensions.

Le rapport : $R^2 = \frac{\sum GP'^2_i}{\sum GP^2_i}$ donne la qualité de la représentation ; il est dit rapport de l'inertie expliquée par la représentation à l'inertie totale du nuage.

Le langage utilisé rappelle la mécanique d'un système de points matériels. Si chaque point P_i est muni de la masse 1/n, on peut interpréter G comme le centre de gravité du système, F_1 comme étant l'axe principal d'inertie portant l'inertie maximum, de même (F_1, F_2) est le plan principal d'inertie portant la plus grande inertie. Chercher les composantes principales c'est déterminer les axes principaux de l'ellipsoïde d'inertie.

Une fois déterminé, ces différents axes d'inertie sont souvent interprétables en terme de caractéristiques du collectif. Ainsi dans l'exemple 1, le premier axe d'inertie est interprétable comme

représentant la performance globale des individus. Les candidats refusés ont sur cet axe une coordonnée petite, ceux qui ont la mention bien, une grande coordonnée.

Le deuxième axe oppose, au bac C, la réussite en mathématiques et celle en physique. Les candidats ayant une bonne note en maths et une mauvaise en physique ont une forte coordonnée sur ce deuxième axe. La conclusion est inversée pour ceux ayant une bonne note en physique et une mauvaise en mathématiques.

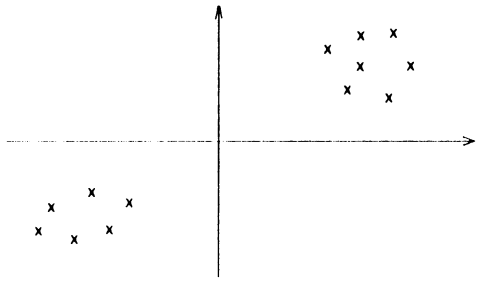
PERTINENCE DE LA REPRESENTATION

Tous les résumés présentés précédemment ne sont pertinents que si la représentation euclidienne est valide. Ainsi, par exemple, si les points représentatifs du nuage de points des individus ont, pour une variable, la présentation suivante :

———————X———————————————————XXXXXXXXXX

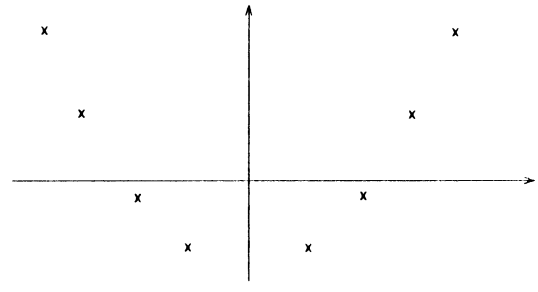
la moyenne ne peut être un bon indicateur de centralité. En effet, la présence ou l'absence de la plus petite observation influe considérablement sur la moyenne. Il n'y a plus stabilité du résumé par rapport aux variations individuelles. Le cadre euclidien n'est plus adéquat. Les fonctionnaires royaux du 18ème siècle qui voulaient résumer la capacité productive des provinces éliminaient la plus petite et la plus grande observation et faisait la moyenne tronquée. On élimine aussi des valeurs dites aberrantes.

De même, quand on étudie la liaison entre deux variables, un nuage de points tel que celui ci-dessous :



conduit à un coefficient de corrélation élevé qui ne traduit pas une liaison linéaire positive entre les deux variables. On a affaire à des données qui peuvent s'interpréter comme un mélange de deux populations distinctes.

Dans le cas suivant, on a une liaison fonctionnelle, mais un coefficient de corrélation nul. L'absence de liaison linéaire ne veut pas dire absence de liaison. L'indicateur choisi pour détecter la liaison est dans ce cas inadéquat.



STATISTIQUE INDUCTIVE

1. LE MODELE STATISTIQUE

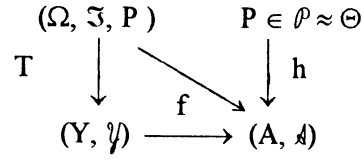
Faire de la statistique inductive, c'est définir sur le collectif étudié une mesure de probabilité, observer un échantillon de ce collectif et, à partir de cette observation, mesurer au moins partiellement la probabilité en question. On peut décrire mathématiquement cette suite d'opérations.

On appelle Ω l'ensemble de tous les échantillons a priori possibles ; Ω est l'espace fondamental. C'est un espace probabilisé, il est donc muni d'une tribu d'événements $\mathfrak{I} \subset \wp(\Omega)$ et d'une probabilité P. P étant inconnu, on a $P \in \mathcal{P}$ ensemble des lois de probabilité a priori possibles. $\mathcal{P}$ peut en général être mis en bijection avec un ensemble Θ appelé ensemble des paramètres. On veut connaître des choses sur Θ. Cela peut se formaliser en introduisant l'ensemble A des actions à mener et une application $h : \Theta \rightarrow A$. Pour la suite on supposera A probabilisable, donc muni d'une tribu d'événements $\mathcal{A}$.

Faire de la statistique c'est associer à un échantillon observé une action à mener. On cherche donc une application S, mesurable: $S : (\Omega, \mathfrak{I}) \rightarrow (A, \mathcal{A})$. S est aussi appelé une stratégie. S en général ne peut

déterminé directement mais à partir d'un résumé intermédiaire. Un résumé des données est une application mesurable $T : (\Omega, \mathfrak{I}) \to (Y, \mathcal{Y})$. T est appelée une statistique ($S = f_o T$).

On peut maintenant faire le schéma global de la statistique inductive.



Dans le cas fréquent où l'échantillon est interprétable comme la réalisation de n variables aléatoires indépendantes, on a alors :

$\Omega = X^n, \quad \mathfrak{I} = \mathcal{X}^{\otimes n}, \quad P = \Pi^{\otimes n}$

où $(X, \mathcal{X}, \Pi)$ est l'espace probabilisé formalisant une expérience. $(\Omega, \mathfrak{I}, P)$ est alors un espace produit.

Illustration du modèle

On reprend les exemples de l'introduction pour illustrer la formalisation proposée.

Exemple 4 - Pour le lot de cartouches supposé illimité, on pose : $X = \{0,1\}$, $\mathcal{X} = \wp(X)$. Soit p la proportion de bonnes cartouches. On a donc $\Omega = \{0, 1\}^n$ et si $k(\omega)$ est le nombre de bonnes cartouches dans l'échantillon, on écrit $P(\{\omega\}) = p^{k(\omega)}.(1-p)^{n-k(\omega)}$. On a $\Theta = [0, 1]$.

Si on veut connaître cette proportion, on pose : $A = [0, 1]$, et h est l'application identique.

Si on veut seulement savoir si $p \leq p_0$ ou $p > p_0$, on écrit $A = \{0, 1\}$ et $h(p) = 0$ si $p \leq p_0$, $h(p) = 1$ si $p > p_0$.

Exemple 5 : On s'intéresse au diamètre des pièces produites, on suppose que celui-ci suit une loi de Gauss de moyenne μ et d'écart type σ. La machine est considérée comme bien réglée si $\mu = \mu_0$. On a alors :

$X = \mathbf{R}, \quad \Omega = \mathbf{R}^n, \quad \Theta = \mathbf{R} \times \mathbf{R}^+, \quad A = \{0,1\}$

$$\frac{dP}{d\lambda_n}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{n/2}\,\sigma^n}\, e^{-\frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

, où λ_n est la mesure de Lebesgue de $\mathbf{R}^n$.

$h(\mu, \sigma) = 0$ si $\mu = \mu_0$, $h(\mu, \sigma) = 1$ sinon.

Exemple 6 : la durée de vie d'un équipement est un nombre réel positif donc : $X = \mathbf{R}^+$, $\Omega = (\mathbf{R}^+)^n$.

Si cet équipement ne vieillit pas , on montre simplement que la loi est exponentielle, c'est-à-dire que l'on peut écrire :

$$P\left(\prod_{i=1}^{n} [x_i, +\infty[\right) = \prod_{i=1}^{n} \exp(-\lambda . x_i), \quad \lambda \in \mathbf{R}^+$$

Donc $\Theta = \mathbf{R}^+$, $1/\lambda$ est la durée moyenne de vie que l'on désire connaître. Il en résulte que $A = \mathbf{R}^+$ et h est l'application identique.

Exemple 8 : Dans les problèmes de détection - radar, échos de bruit et échos de cible peuvent être considérés comme des réalisations de variables aléatoires indépendantes de même loi pour chacun des deux types. La loi du bruit change suivant les conditions atmosphériques, celle de la cible se déduit de celle du bruit par une simple translation. On est amené alors à procéder comme suit. On prélève un échantillon de m observations de bruit et un échantillon de n observations de ce qui peut être une cible. Soit f la densité de probabilité du bruit et Δ l'intensité de l'écho cible. On a $\Omega = \mathbf{R}^{n+m}$, $\Theta = \mathbf{R}^+ \times \mathcal{J}$, où $\mathcal{J}$ est l'ensemble des densités de probabilité continues.

$$\frac{dP_\theta}{d\lambda}(x_1,\ldots,x_m,x_{m+1},\ldots,x_{m+n}) = \prod_{i=1}^{m} f(x_i) \cdot \prod_{i=m+1}^{m+n} f(x_i - \Delta) \quad A = \mathbf{R}^+, \quad h(\Delta, f) = \Delta.$$

CHOIX D'UNE STRATEGIE STATISTIQUE

Pour choisir une stratégie statistique il est nécessaire de se donner des critères de choix. On peut se restreindre à certaines classes d'applications, se donner un préordre sur l'ensemble des stratégies. Pour cela, il faut enrichir le modèle précédent :

Stratégie convergente : on connaît en probabilité la loi des grands nombres. Soit un événement E de probabilité $p > 0$, n répétitions indépendantes, f_n la fréquence empirique d'apparition de E. On peut démontrer : $\forall \varepsilon > 0$, $P_p(\{|f_n - p| > \varepsilon\}) \to 0$ quand $n \to \infty$, P_p étant la probabilité quand p est "l'état de la nature".

Ce théorème permet de dire que quand on réalise un grand nombre d'expériences indépendantes, on approche de très près la vraie loi de probabilité. A noter que pour exprimer la proximité entre probabilité et fréquence, on est obligé d'employer le concept de probabilité. D'où la difficulté de définir ce dernier comme limite de la fréquence empirique. Quoi qu'il en soit de cette discussion épistémologique, il est naturel de concevoir la probabilité comme limite de la fréquence empirique. Ce point de vue est celui des statisticiens fréquentistes.

Dans ces conditions, il est naturel d'exiger qu'une stratégie statistique soit d'abord convergente. Pour formaliser cela il faut plonger la situation particulière dans une suite de situations. Un problème statistique δ c'est la donnée de : $\forall n$, $\delta_n = (X^n, \Theta, A, h ; P_{n,\theta})$.

Supposons que A soit un espace métrique muni d'une distance d.

La stratégie $S_n : X^n \to A$ sera dite convergente si :

$\forall \varepsilon > 0, \forall \theta \in \Theta$, $P_{n,\theta}(\{\omega / d[S_n(\omega), h(\theta)] > \varepsilon\}) \to 0$, quand $n \to \infty$,

En terme intuitif cela se dit : la stratégie choisie approche la valeur cherchée quand n est grand.

Dans l'exemple des cartouches, si on prend comme stratégie $S_n(\omega) = k(\omega)/n$ (fréquence empirique des bonnes cartouches), on a bien, si p est la proportion de bonnes cartouches :

$\forall \varepsilon > 0$, $P_p(\{|S_n - p| > \varepsilon\}) \to 0$ quand $n \to \infty$.

C'est la loi des grands nombres. La distance sur [0,1] choisie est la distance classique sur **R**.

Fonction de coût : Pour introduire un préordre sur l'espace des stratégies on introduit une mesure de l'erreur faite en remplaçant ce que l'on cherche h (0) par la décision statistique s (ω).

Pour cela on introduit une fonction de coût : $L : \Theta \times A \to \mathbf{R}^+$

Le coût de la décision statistique est donc : $L(\theta, S_n(\omega))$ si θ est l'état de la nature, et ω le constat expérimental, S_n la stratégie choisie. Ce nombre dépend de ω donc du hasard, c'est le résultat d'une loterie. Depuis Pascal, pour comparer entre elles les loteries on compare leurs espérances mathématiques. On définit ainsi la fonction de risque de la stratégie S_n par :

$$R(\theta, S_n) = E_\theta(S_n) = \int_\Omega L(\theta, S_n(\omega)) dP_\theta(\omega)$$

Une stratégie S_n sera préférée à la stratégies S_n^* ($S_n \succ S_n^*$) si et seulement si

$\forall \theta \in \Theta$, $R(\theta, S_n) \leq R(\theta, S_n^*)$.

On a un préordre partiel sur l'espace des stratégies.

Prenons l'exemple 4 du lot de cartouches pour illustrer le calcul de la fonction de risque. On a $\Theta = A = [0, 1]$. Il est d'usage de prendre la fonction de coût quadratique $L(\theta, a) = (\theta - a)^2$.

La stratégie S_n qui consiste à estimer la proportion p par la fréquence empirique f_n a comme fonction de risque : $R(p, f_n) = E_p((f_n - p)^2) = Var(f_n) = p.(1-p) / n$.

La stratégie triviale qui consiste à décider p=1/2 sans regarder l'échantillon a pour fonction de risque $(p-1/2)^2$. On voit immédiatement que ces deux stratégies ne sont pas comparables.

Le préordre précédent ne permet donc pas de choisir entre ces deux stratégies. C'est pour cela que l'on est amener à restreindre l'ensemble des stratégies. Il est bien évident que la stratégie triviale n'est pas convergente, par exemple. On va introduire le cas de l'estimation qui est un autre concept.

CAS DE L'ESTIMATION

On dit que l'on a un problème d'estimation quand l'ensemble A est un espace vectoriel normé. Dans les cas classiques : $A = \mathbf{R}^k$. On dit alors que S_n estime $h(\theta)$, S_n est l'estimateur, $S_n(\omega)$ l'estimation déduite de l'échantillon.

Un critère possible de sélection d'une stratégie est le suivant. On veut qu'en moyenne les estimations différentes qui seraient possibles, si on pouvait répéter l'épreuve statistique, coïncident avec la chose à estimer. Cela peut se traduire par : $E_\theta(S_n) = h(\theta)$.

L'estimateur S_n est alors dit sans biais. Dans l'exemple des cartouches, on a bien : $E_p(F_n) = p$.

Dans le cas où $A = \mathbf{R}$, introduire la fonction de coût quadratique et se restreindre à des estimateurs sans biais revient à comparer les variances des estimateurs. Cette variance est la fonction de risque de l'estimateur sans biais.

Estimer un paramètre, c'est attribuer à ce paramètre une valeur que l'on espère proche de la valeur réelle. On est alors tenté d'indiquer la précision de cette approximation et donc de donner une borne supérieure à l'erreur commise. Mais comme on est dans un modèle probabiliste, celui-ci sera formalisé en termes probabilistes.

ESTIMATION PAR INTERVALLE

On se contentera d'étudier le cas où $A = \mathbf{R}$. Pour définir la précision on veut trouver un intervalle qui, avec une forte probabilité, recouvre la valeur cherchée. Pour cela, on se fixe un risque α et on cherchera un intervalle I tel que $P_\theta(\{\omega / I(\omega) \ni h(\theta)\}) = 1 - \alpha$.

Si $I = [a,b]$ cela peut s'écrire : $P_\theta(\{\omega / a(\omega) \leq h(\theta) \leq b(\omega)\}) = 1 - \alpha$.

Attention, c'est bien a et b qui sont aléatoires, $h(\theta)$ est inconnu mais certain. Il ne faut pas confondre aléatoire et inconnu. Cette situation d'estimation peut-être illustrée par l'exemple 6 sur les durées de vie. On veut estimer $1/\lambda$, durée moyenne de vie.

On appelle $X_1,...,X_n$ les variables aléatoires durées de vie des équipements à observer. On propose comme estimation la moyenne des durées de vie observées. On posera :

$\widehat{1/\lambda} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$. On a : $E(\widehat{1/\lambda}) = 1/\lambda$ et $Var(\widehat{1/\lambda}) = 1/\lambda\sqrt{n}$.

On montre que $n\lambda\ \widehat{1/\lambda}$ suit la loi Γ_n, ce qui permet de construire l'intervalle de confiance.
Fixons $\alpha = 0{,}10$, on cherche a_n et b_n tels que $\Gamma_n(a_n) = 0{,}05$ et $1 - \Gamma_n(b_n) = 0{,}05$.

où : $\Gamma_{n+1}(x) = \int_0^x e^{-t}\frac{t^n}{n!}dt$

On a $P_\lambda(a_n < n\lambda.\widehat{1/\lambda} < b_n) = 0{,}90$ et $P_\lambda(n/b_n.\widehat{1/\lambda} < 1/\lambda < n/a_n.\widehat{1/\lambda}) = 0{,}90$.

On dit qu'avec une confiance de 90% la durée moyenne de vie est comprise entre $n/a_n\ \widehat{1/\lambda}(\omega)$ et $n/b_n\ \widehat{1/\lambda}(\omega)$.

La confiance n'est pas la probabilité. Ce terme rappelle que la procédure ayant abouti au constat fait devait réussir avec une probabilité 0,9.

PROBLEMATIQUE DES TESTS

Dans l'exemple 5 la décision à prendre est dichotomique. Doit-on oui ou non faire régler la machine. Dans un premier temps, il en est de même dans l'exemple 8 : y a-t-il oui ou non écho, si oui quelle est son intensité ? Quand l'ensemble des décisions à prendre ne comporte que deux éléments on peut écrire :

$A=\{0,1\}$.

L'ensemble Θ qui représente tous les "états de la nature" a priori possibles peut alors être dichotomisé $\Theta = \Theta_0 \cup \Theta_1$ avec $\Theta_0 \cap \Theta_1 = \varnothing$ et $\Theta_0 = h^{-1}(\{0\})$.

Décider 0 c'est dire que le paramètre θ qui régit le phénomène est un élément de Θ_0. Ainsi dans l'exemple 5, cela revient à affirmer que la moyenne du diamètre des pièces est μ_0. Quand décide-t-on 0 ? Quand la règle est l'application S, on pose : $C = S^{-1}(\{1\})$ qui est appelée région critique du test.

Si on examine les erreurs on se trouve devant le cas de figure suivant :

Décision du statisticien — Etat de nature

	Θ_0	Θ_1
0	Bien	Erreur 2
1	Erreur 1	Bien

Pour pouvoir introduire un préordre partiel sur les règles de décisions, il faut mesurer les erreurs.

L'erreur 1 peut être représentée par la fonction : $P_\theta(C)$, $\theta \in \Theta_0$. $P_\theta(C)$ est la probabilité de décider 1 quand l'état de la nature est θ, ici $\theta \in \Theta_0$.

L'erreur 2 peut être représentée par la fonction : $1 - P_\theta(C)$, $\theta \in \Theta_1$. $1 - P_\theta(C)$ est la probabilité de décider 0 quand l'état de la nature est θ, $\theta \in \Theta_1$.

Jusqu'à présent, les deux cas possibles jouent des rôles parfaitement symétriques. Mais dans les problèmes de décision dichotomique, il n'en est pas en général ainsi. L'une des hypothèses, disons $\theta \in \Theta_0$ est privilégiée. Elle correspond par exemple à une théorie scientifique et on ne l'abandonnera que si on a de très sérieuses raisons pour cela. Ou bien, comme dans l'exemple 5, rejeter l'hypothèse que la machine est bien réglée conduit à tout un processus qui coûte cher. On ne le mettra en route que si l'hypothèse choisie a priori est peu plausible au vu des résultats.

On traduit mathématiquement ce choix de l'hypothèse nulle de la façon suivante. On va fixer a priori une borne à l'erreur 1. On choisira une règle de décision S, donc une région C telle que

$$\sup_{\theta \in \Theta} P_\theta(C) \leq \alpha$$

Une fois l'erreur 1 choisie, un test sera meilleur qu'un autre si l'erreur 2 est plus petite pour le premier que pour le second et ce quel que soit $\theta \in \Theta_1$. On appelle puissance du test C la fonction : $\beta_C(\theta) = P_\theta(C)$.

Le test C_1 sera préféré à C_2 (C_1 â C_2) si, et seulement si $\forall\, \theta \in \Theta_1$, $\beta_{C_1}(\theta) \geq \beta_{C_2}(\theta)$.

Le choix de α dépend, lui, de la confiance que l'on a, a priori, dans l'hypothèse Θ_0, α sera d'autant plus petit que l'on a une confiance plus grande dans l'hypothèse Θ_0, ou bien que l'abandon de l'hypothèse Θ_0 conduit à des complications diverses et coûteuses. Ainsi dans les problèmes tels que ceux de l'exemple 5, on prend souvent α=0,01. Les psychologues, les biologistes choisissent traditionnellement α=0,05 ou 0,10.

Dans l'exemple 8 de détection de signaux radar, le nombre α a une interprétation concrète. Pendant un temps de surveillance donné, le nombre de fois où il apparaît un point sur l'écran alors qu'il n'y a pas de cible est proportionnel à α, ce que le radaristes appellent taux de fausse-alarme. De même la puissance du test est la probabilité de faire apparaître un point sur l'écran quand il y a cible. Accepter trop de faux échos, c'est prendre le risque de mobiliser pour rien des moyens importants et trop souvent. On limite donc ce taux et on prend souvent $\alpha = 10^{-5}$ ou 10^{-6}. Le réglage de ce seuil est d'ailleurs possible, sur les appareils existants, par l'utilisateur.

Un test sera dit sans biais s'il est toujours meilleur que la stratégie triviale qui consiste à décider sans regarder les observations. On a alors $\beta_C(\theta) \geq \alpha$, $\forall\, \theta \in \Theta_1$ (prendre comme stratégie triviale : je décide $\theta \in \Theta_1$ avec la probabilité α quel que soit l'échantillon).

CHOIX D'UNE PROCEDURE

Qu'il s'agisse de test, d'estimation ou d'autres problèmes statistiques, il faut ensuite mettre au point des procédures présentant un certain nombre de qualités prescrites a priori et ayant, au vu des critères développés ci-dessus, de bonnes performances.

C'est l'objet d'une partie de la statistique mathématique de faire ce travail. Il a été commencé au début du siècle par K. Pearson ; il s'est développé entre les deux guerres pour les modèles gaussiens avec Sir Ronald Fisher et son équipe, en particulier. Il se continue depuis la guerre en cherchant toujours plus de généralité ou en travaillant sur des modèles dont jusqu'à présent la complexité avait interdit la mise au point ou l'étude de procédures adéquates.

LA STATISTIQUE BAYESIENNE

Dans ce qui précède, on a présenté la probabilité comme une caractéristique objective du phénomène étudié dont on pouvait s'approcher en observant un grand nombre de répétitions. Dans l'optique Bayésienne une opinion sur les choses est une interprétation possible de la probabilité. On peut démontrer d'ailleurs qu'un système cohérent d'opinions peut se traduire par une distribution de probabilité.

Reprenons le schéma statistique. On a vu que l'ensemble $\mathcal{P}$ des lois de probabilité a priori possibles pouvaient être mis en correspondance biunivoque avec un ensemble Θ appelé ensemble des paramètres. On supposera que Θ est un espace probabilisable. On va munir cet espace d'une loi de probabilité Q qui exprime l'opinion a priori du statisticien (ou d'un spécialiste) sur les états de la nature. Il existe des cas où cette opinion a priori est le résultat d'une construction, reflète l'expérience du praticien. L'exemple 9 du diagnostic médical illustre cette situation. L'examen clinique permet au médecin de donner une probabilité a priori à chaque maladie possible.

On peut maintenant utiliser cette probabilité a priori. On supposera que toutes les probabilités admettent des densités par rapport aux mesures σ-finies μ sur $(\Omega,\mathfrak{I})$ et ν sur $(\Theta,\mathfrak{R})$; $q(\theta)$ est la densité de probabilité de Q par rapport à ν, $p(\omega,\theta)$ est la densité de probabilité de P_θ par rapport à μ.

Appliquons le théorème de Bayes. Dans le cas fini celui-ci peut s'énoncer ainsi : soit $(A_1,..,A_n)$ une partition de Ω, B un événement. En posant P(A/B) la probabilité de A sachant B, on peut facilement démontrer :

$$P(A_i/B) = P(B/A_i).P(A_i) \,/\, \Sigma\, P(B/A_i)\, P(A_i)$$

$P(A_i/B)$ s'interprète comme la probabilité de la "cause" A_i quand l'événement B se produit et $P(B/A_i)$ la probabilité de B quand "A_i agit" ; $P(A_i)$ la probabilité a priori de la cause A_i.

On appelle $q(\theta,\omega)$ la densité de la loi de probabilité a posteriori Q_ω sur Θ par rapport à la mesure ν. Le théorème de Bayes s'écrit :

$$q(\theta,\omega) = \frac{p(\omega,\theta).q(\theta)}{\int_\Theta p(\omega,\theta).q(\theta).\nu(d\theta)}$$

C'est à partir de cette densité a posteriori que s'élaborent les stratégies statistiques dites alors bayésiennes.

Quand on peut définir une fonction de coût $L : \Theta\times A \to \mathbf{R}$, $L(\theta,a)$ est le coût de l'action a quand l'état du système est θ. On introduit alors des risques a priori et a posteriori de l'action a.

Risque a priori $\tau_Q(a) = \int_\Theta L(\theta,a).q(\theta).\nu(d\theta)$ Risque a posteriori $\tau_Q(a,\omega) = \int_\Theta L(\theta,a).q(\theta,\omega).\nu(d\theta)$

La stratégie optimale quand l'opinion a priori est Q est celle qui minimise le risque a posteriori.

On peut écrire sous réserve d'existence $\tau_Q(S(\omega),\omega) = \min_{a\in A} \tau_Q(a,\omega)$.

Le point de vue Bayésien permet donc de trouver des stratégies optimales au sens bayésien. On montre que si $\forall\, \theta \in \Theta$, $q(\theta) > 0$, les stratégies bayésiennes sont convergentes au sens précédent.

QUELQUES REMARQUES POUR FINIR

L'exposé ci-dessus a voulu illustrer la démarche statistique par la modélisation de quelques situations particulières suivie des méthodes de traitement les plus usitées. Mais la statistique intervient également en amont. Etant donné un problème, comment mener les expériences pour avoir les renseignements souhaités, pour optimiser le nombre d'expériences suivant leur coût. En aval, il existe aussi des techniques destinées à

des conclusions dépend de l'adéquation du modèle à la réalité décrite et du mode de recueil des informations.

Du point de vue de l'enseignement, la statistique est une discipline très riche. A un certain niveau on y fait des mathématiques relativement sophistiquées. A un niveau plus modeste, on peut montrer un champ d'application des mathématiques très vaste. En outre, elle oblige à un travail interdisciplinaire au plein sens du terme.

Les données traitées sont toujours issues d'une autre discipline. La mise au point du modèle représentatif, l'interprétation des résultats supposent que le statisticien entre dans la problématique du spécialiste demandeur et que ce dernier soit capable de comprendre ce qu'est un phénomène aléatoire, une probabilité, une confiance. Sans cette compréhension au moins intuitive, il risque de faire des erreurs graves au niveau des conclusions qu'il tire pour sa discipline.

Si on veut intéresser les élèves à la statistique, il faut avant tout faire ressortir la démarche, son intérêt scientifique et pour cela utiliser de nombreux exemples. Le traitement numérique concret des données peut maintenant être considérablement allégé grâce à l'utilisation des logiciels ad'hoc, voire de calculatrices. On a donc en plus une application intéressante de l'informatique. Il est toujours intéressant d'aller jusqu'au bout des calculs afin de pouvoir faire l'interprétation des résultats...

La pratique montre que pédagogiquement l'étude de la statistique permet d'intéresser à la mathématique des élèves, des étudiants, qui jusque là, étaient rétifs à cette discipline. Son caractère fortement interdisciplinaire permet ce déblocage.

La statistique voit actuellement son champ d'application s'étendre considérablement. La demande d'enseignement ne peut donc que croître au niveau des formations à finalité professionnelle. De plus, pouvoir interpréter des données existantes est un acte que tout citoyen voulant se construire une opinion doit faire. Il n'y a qu'à voir le nombre de bêtises qu'écrivent les journalistes à propos des sondages ou de la correction des variations saisonnières, pour se convaincre de l'importance de la statistique aussi dans les mathématiques pour tous.

BREVE BIBLIOGRAPHIE

Cette bibliographie ne prétend pas à l'exhaustivité, il existe des ouvrages remarquables et adaptés aux besoins de ceux qui s'intéressent à l'enseignement de la statistique jusqu'au niveau Bac+2 et qui n'y figurent pas. Elle veut seulement guider le premier choix de ceux qui découvrent cette discipline.

– W. PETER Countring for something, statistical principal and personalities. Ed. Sprinper – Verlag
Ouvrage simple, donne une présentation historique simple de la statistique.

– A.P.M.E.P. Analyse de données (2 tomes)
Une présentation simple, pédagogique, et intéressante des méthodes de description de la statistique.

– ROZENGARD Probabilité et statistique en recherche sociale. Ed. Dunod.
Initiation aux probabilités et aux statistiques de niveau moyen. On utilise les mathématiques, mais de nombreux résultats sont admis.

– FOUCART, BENSABERT, GARNIER. Méthodes pratiques et de la statistique. Ed. Dunod.
Exposé très pratique des méthodes statistiques, mais les idées sous jacentes ne sont pas toujours très bien mises en valeur.

– J.J. DROESBEKE Eléments de statistique, collection SMA, Ed. Ellipses.
Bon compromis entre la statistique appliquée et le modèle mathématique, intéressant, complet.

– A. MONFORT Cours de probabilité. Ed. Economica.
Pour ceux qui veulent plonger dans le calcul des probabilités et lire un exposé complet de bon niveau écrit en utilisant le langage de la théorie de la mesure (le seul adapté).

– Ph. TASSI Méthodes statistiques. Ed. Economica.
Exposé classique de la statistique inférentielle. Nécessité de connaître un peu le calcul des probabilités.

– ANDERSON, SCLOVE. Statistical analysis data, Ed. The Scientific press.
Exposé des principales méthodes de statistiques descriptive et inférentielle avec de nombreux exemples, lecture agréable.

– Th WONACOTT et R. WONACOTT. Statistique (économie gestion, sciences humaines). Ed. Economica.
Présentation claire des méthodes statistiques (surtout pour la statistique inférentielle) ; livre à lire, de très nombreux exemples, une mine d'exercices possibles, mais les applications sont surtout orientées vers les sciences humaines et économiques.

– SARDADI, VINCZE. Mathematical methods of statistic quality control. Ed. Académic Press.
Initiation aux probabilités, à la statistique inférentielle avec des applications au contrôle de qualité essentiellement et un peu à la fiabilité, beaucoup d'exemples.

–INTER IREM TECHNIQUE. Fiabilité.
Toutes les procédures statistiques au programme des BTS maintenance et de quelques autres illustrées par des exemples issus des examens.