GUY BROUSSEAU

SITUATIONS FONDAMENTALES ET PROCESSUS GENETIQUES DE LA STATISTIQUE

INTRODUCTION

La publication récente du compte rendu d'une expérience ancienne est la raison de ma présence devant vous. Je ne suis pas sûr qu'en choisissant le thème « étude d'un problème curriculaire : L'enseignement des statistiques », les organisateurs avaient en tête de me voir raconter cette aventure d'une autre époque. J'avoue d'ailleurs avoir été troublé par leur demande car il me semblait que vous auriez dû préférer un cours d'un chercheur plus jeune et mieux informé des orientations du moment. J'ai tout de même accepté leur offre lorsqu'ils ont précisé qu'il serait possible de montrer à cette occasion comment fonctionne le concept de « situation » dans une recherche curriculaire. Que peut-il apporter ? Il s'agissait bien d'un défi. Il m'a conduit à revisiter les raisonnements et les résultats qui m'avaient guidé à l'époque à l'aide d'instruments développés depuis ; je me suis rapidement replongé dans un univers incommunicable dans le cadre de ce cours. Après avoir élagué le propos principal de la plupart des mises au point et des précisions que vous trouverez dans les annexes et dans les notes qui accompagnent ce cours, il nous est resté trois propos assez distincts :

- la méthode de recherche d'un *processus fondamental* pour un secteur des mathématiques, à l'aide de la théorie des situations didactiques,
- son application à la recherche des obstacles à l'enseignement
- et le compte rendu de l'expérience avec ses implications actuelles.

Le premier propos était celui qui m'intéressait le plus. Il m'a conduit à revisiter quelques idées, entre autres, la notion de situation fondamentale, et de hasarder des interprétations en termes de praxéologie. La logique aurait voulu que je traite ces questions dans cet ordre, mais les raisonnements théoriques sur des situations sont à la fois complexes et abstraits. Malgré des efforts considérables, je ne me suis pas senti capable de présenter oralement les textes que j'avais rédigés sur cette question, ni même les diaporamas qui en étaient issus. Je vais donc exposer le résultat avant d'expliquer la méthode, en me livrant ainsi à une violation caractérisée des idées didactiques que je vais m'efforcer par ailleurs de vous présenter.

1. UNE EXPERIENCE DE PREMIER ENSEIGNEMENT DES STATISTIQUES[1]

SITUATION INITIALE ET UN PROCESSUS GENETIQUE

Le but pour le professeur

Le *but* implicite du processus était de définir ce qu'est une statistique en faisant comprendre - en définissant implicitement – ce qu'est l'égalité ou plutôt *l'équivalence de deux statistiques*.

[1] Ce paragraphe résume l'article : « une expérience de premier enseignement en statistique », et l'article de Guy BROUSSEAU, Nadine BROUSSEAU, Virginia WARFIELD, "An experiment on the teaching of statistics and probabilité" *Journal of Mathematical Behavior*, 20 (2002) 363-441. Le texte en français constitue l'annexe 1 dans le CD-ROM. Voir aussi le diaporama 1

Deux statistiques sont équivalentes si et seulement si elles « définissent » le même objet. C'est-à-dire si elles l'acceptent toutes les deux comme modèle. Cette approche implique donc une stratégie explicite du *test d'hypothèse* que le processus devrait faire apparaître.
Les *données* sont des informations choisies et recherchées à la discrétion de l'actant, de façon à ce qu'il soit maître de réitérer ses demandes ou non, selon la perception qu'il a de leur utilité. Habituellement les données sont recueillies, souvent dans des conditions inconnues et il s'agit de deviner de quel objet ou système elles sont la trace. L'incertitude et donc les raisonnements portent sur l'objet. Ici l'objet de l'observation est déterminé sinon connu.
Le *milieu* capable de délivrer rapidement des informations faciles à recueillir et à interpréter est une machine de hasard, pour l'observateur, mais pas pour les actants car aucune prévision ne leur est demandée au cours de la situation d'action. Des demandes de prévisions n'apparaîtront que comme *argument rhétorique dans l'analyse des explications* fournies par les élèves pendant les situations de validation. Le seul enjeu est un savoir, pas un gain quelconque. Il est important de remarquer que ni le hasard ni la méthode du test d'hypothèse ne sont ici des objets d'étude mais que l'un est une condition de la situation et l'autre un moyen de déterminer un objet sous cette condition. La machine devait être très simple, facile à concevoir et à reproduire. L'objet X, le modèle à produire, devait exister déjà, être sûr, aussi familier que possible : la composition d'une petite collection de boules. Il n'est donc fait appel qu'à un répertoire très simple. Les diverses possibilités sont imaginables immédiatement. La situation d'action porte donc sur le choix d'une composition : « combien y a-t-il de boules blanches dans cette bouteille opaque, qui en contient 5 ? ».
La progression des *formulations* est la partie qui a été la moins clairement explicitée et la moins contrôlée. La standardisation du vocabulaire didactique aurait dû éviter les contaminations sauvages par la culture du professeur ou celle encore plus incontrôlable des élèves. Les termes ambigus se pressent en foule sur les lèvres des participants au cours des débats. Mais les faits sont plus têtus que les mots et le sens de ces formulations déficientes est correct.
La *motivation* effective n'est pas le gain, ni le goût du jeu, ni même la découverte empirique d'une loi, ce sera la mise au point d'arguments, et d'un discours convaincant, c'est-à-dire consistant, sur la vérité d'une déclaration sur un objet qui restera toujours caché, conformément au projet expliqué plus haut.
Dans ces conditions nos pourrons observer l'invention et /ou la justification immédiate de très nombreuses techniques mathématiques élémentaires dans leur rôle spécifique de la statistique. L'inventaire en est éloquent.

RESUME DU DEROULEMENT DE L'ENSEIGNEMENT

Les conditions du déroulement sont exposées dans le texte donné en annexe

La situation initiale

Le professeur :

« Votre camarade Pierre va mettre dans cette bouteille (opaque) et vide, 5 boules prises dans ce sac (opaque) qui en contient une trentaine. Venez vérifier qu'il n'y a dans le sac que des boules blanches et des boules noires ».

« Pierre, mélange les boules dans le sac, et maintenant, sans regarder, isole 5 boules que tu tiens à travers le tissu ».

« Venez constater qu'il y en a bien 5. J'introduis alors la bouteille dans le sac.

« Pierre, fais entrer les cinq boules dans la bouteille ! et maintenant bouche le goulot avec ce bouchon translucide.

« Vous êtes tous bien sûrs que dans cette bouteille il y a bien 5 boules et rien d'autre mais vous ne connaissez pas leur couleur. Nous allons essayer de savoir ce que contient cette bouteille sans jamais l'ouvrir ».

Cette question heurte évidemment les modalités des raisonnements déterministes en usage dans les classes et les élèves ne comprennent pas quel calcul ils pourraient faire pour obtenir la solution de ce problème.
Il commencent donc par essayer de regarder à travers le bouchon, puis l'un renverse la bouteille et là il peut voir la couleur de la boule qui se loge contre le bouchon.

« Il y a une boule blanche ! »

La question qui vient spontanément alors est « y en a-t-il aussi une noire ? » deux retournements de la bouteille… encore une blanche (la même ? une autre ? personne ne sait). L'idée que s'il y a une noire elle va finir par se montrer justifie des retournements jusqu'à ce que l'événement se produise (s'il se produit sinon ?). Ce ne sont évidement pas des tirages.

« Il y a une blanche et une noire… »

C'est la fin de l'épisode déterministe et rien ne saurait justifier dans ce mode de pensée de continuer à retourner cette bouteille (est-ce que la lecture d'un deuxième exemplaire d'un même journal confirme les nouvelles du premier ?).
Mais en général les enfants font des hypothèses qui vont démarrer le processus :

«- recommence cinq fois pour qu'on voie toutes les boules » dit un enfant qui a l'idée que les boules doivent se montrer successivement en bon ordre

- voilà ! il y a trois boules blanches et deux noires !

L'histoire n'a pas de raison d'aller plus loin… sauf si cette idée que les boules se montrent sagement dans le même ordre est relevée, que ce soit par le professeur ou par un élève.

« Si ce que tu dis est vrai alors en recommençant on devrait voir encore les trois blanches et les deux noires… »

Surtout ne pas s'arrêter aux hésitations des élèves qu'une prévision directe effarouche

« cette fois on voit quatre fois une blanche et une noire »

Exit sans bruit l'hypothèse de la succession régulière … tout pourrait s'arrêter là mais…

«- En tout cas il y plus de blanches que de noires… » dit un autre

- Alors on devrait continuer à voir plus de blanches que de noires si on continuait ?

- Non, si les blanches se sont montrées c'est au tour des noires (il aura compensation). »

De là une attention particulière pour les résultats des observations et pour leur enregistrement.
Il faudrait examiner toutes les évolutions possibles dans tous les cas (par exemple s'il n'y avait que des boules blanches) mais le démarrage est maintenant assuré et le professeur va pouvoir relancer des « expériences » (qui ne sont toujours pas des tirages puisqu'elles se font en situation supposée déterministe) à l'aide du moteur suivant :
Si ce qu'on voit donne des indications sur ce que contient la bouteille, alors, en reproduisant ce qu'on a fait on doit reproduire ce qu'on a vu (la question est évidemment :

« Qu’y a-t-il de commun entre ces expériences ? Que peut on voir ? ».

Ce qu’on a vu, va devenir l’image du passé de la machine de hasard, c’est-à-dire une *statistique*, ce qu’on prévoit va devenir des événements plus ou moins *probables* associés à la machine et les deux sont reliées par des hypothèses sur la structure de la machine que l’on peut facilement imaginer.
Il faut remarquer que cette situation est fondamentale à la fois pour les statistiques *et* pour les probabilités. Autrement dit, seule l’invention de ces deux concepts (et les notions afférentes) et la compréhension de leurs rapports peut résoudre ce problème de façon satisfaisante. Il serait compliqué de discuter ce point maintenant. Nous le vérifierons en suivant la genèse que ce problème peut provoquer. Cela suppose évidemment que le professeur n’apportera et n’acceptera aucun apport extérieur sur les questions cruciales que nous allons rencontrer.

Le processus

A grands traits, le processus qui s’engage dans la situation fondamentale des probabilités et des statistiques est le suivant : les enfants vont, un temps, faire des groupes de 5 observations avec l’idée de représenter ainsi le contenu de la bouteille. Ils sont très déçus de voir les résultats fluctuer mais l’idée qu’il y a plus de blanches ou plus de noires, et que ça doit se voir suffit à entretenir le processus. Bientôt ils comptent combien de fois ils ont obtenu 4 b et 1 n, 3 b et 2n ou 2 b et 3 n et en tirent argument pour renforcer leur conviction. Par exemple, à un moment donné, ils ont 5 séries (3b, 2n) contre 4 (4b et 1n) et 3 (2b, 3n).
Bientôt leur conviction est faite. Ils se déclarent sûrs de leur fait… et réclament d’ouvrir la bouteille pour vérifier ! Le professeur évidemment refuse : s’ils sont sûrs, inutile de vérifier, sinon il faut d’autre moyens de se convaincre (la probabilité n’est pas un concept expérimental).

> « Nous serons sûrs, disent les enfants, lorsque l’effectif d’une des compositions possibles dépassera les autres de 3…

Voici le résultat

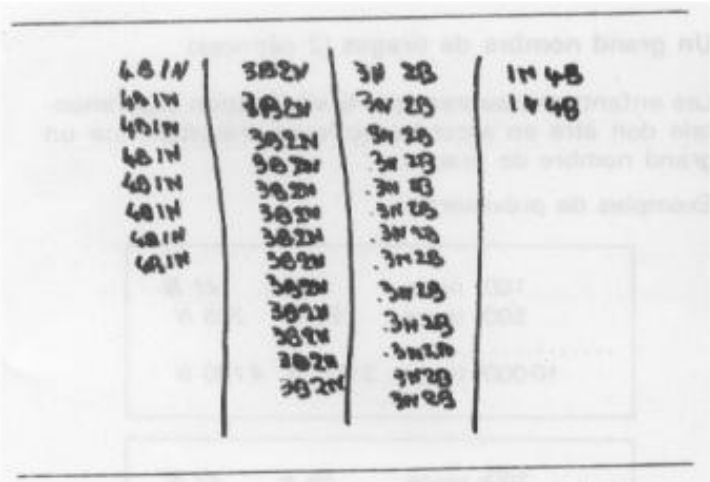


Figure 4. — Une autre manière de noter les résultats.

C’est alors que certains élèves avancent l’argument suivant :

> « Il y a autant de (3b,2n) que de (2b, 3n) mais il y a beaucoup plus de (4b, 1n) que de (1b, 4n) alors il faut conclure qu’il y a (3b, 2n).

D’autre argumentent plus radicalement :

> Il est sorti beaucoup plus de blancs (32 + 36 + 26 + 2) = 94 que de noirs (26 + 36 + 8) = 70, il y a plus de blancs que de noirs, donc c'est 3b et 2n.

Certains, qui ne voulaient pas compter les résultats « faux » c'est-à-dire non conformes à leur intime conviction, voient là l'occasion d'augmenter ce que nous appellerions leur « corpus ». C'est ainsi qu'une distribution de distributions fait entrer la comparaison des effectifs de noirs et de blancs dans le débat. Le fait que la fréquence des événements converge plus vite que celle des distributions a été pressenti et concrètement utilisé. Les rapports de noirs et de blancs apparaîtront plus tard.
Et le professeur relance toujours le processus en disant :

> «- si vous êtes sûrs qu'il y plus de billes blanches que de noires dans la bouteille parce que vous voyez sortir plus de coups « blancs » que de coups « noirs », alors vous devez penser qu'en recommençant vous allez trouver encore, plus de coups blancs ?
>
> - Oui !
>
> - Et est ce que c'est vrai ?
>
> - Essayons… »

Ainsi les suites de tirages s'allongent, nourries de doutes et de convictions des élèves et de leur désir de vérifier et de prouver ce qu'ils pensent. Les séances sont courtes : le plus souvent dix minutes ou un quart d'heure : le temps de discuter un résultat et de décider une action.

Modélisation et tables de hasard

Certains élèves, intrigués par cette bouteille, et irrités par le refus obstiné du professeur d'ouvrir la bouteille initiale décident de prendre des bouteilles de plastique transparent pour « voir » ce qui se passe. Ils en font trois, s'accordant pour trouver sans intérêt les bouteilles qui ne contiennent que des boules de la même couleur. Le mystère ne s'éclaircit pas pour autant, alors le professeur leur demande s'ils peuvent prévoir des résultats obtenus avec ces bouteilles. Certains conservent les distributions d'autres prévoient et comptent les blancs et les noirs, mais dans les deux cas la prévision est linéaire : la distribution prochaine sera identique à la précédente, c'est en tout cas ce qu'on a avantage à prévoir.
L 'étape suivante est franchie lorsque les élèves commencent à pressentir que la meilleure prévision pour les bouteilles transparentes, utilise le contenu de la bouteille plutôt que les résultats des premiers tirages. Et aussi lorsqu'ils soupçonnent le rôle intrinsèque de la longueur des suites d'observations.

> Il y a des écarts, pensent-ils mais si on recommence on devrait voir les écarts se réduire !

Espoir fallacieux pour l'instant puisqu'on considère des effectifs, mais l'important est que cette étape fournit des raisons de faire de longues suites d'observations et de noter les résultats dans le cadre de questions pertinentes à leur sujet. Devant la longueur désespérante des tirages, de plus en plus nombreux pour satisfaire les demandes, le professeur propose des tirages déjà faits (des tables de n et b obtenues hors de la classe avec un ordinateur). Au début les élèves comptent les n et les b, par la suite, ils demandent à la machine de calculer directement les effectifs, puis beaucoup plus tard les rapports.
Dès lors, tout est en place pour calculer le rapport vers lequel on s'attend à voir la fréquence observée des blancs ou des noirs se rapprocher : 3/5 pour (3b,2n), 2/5 pour (2b, 3n) etc. Les élèves fixent eux-mêmes le nombre de tirages dont ils demandent les résultats à l'ordinateur (entré entre temps dans la classe). Les séries de fréquences observées sont reportées sur des

graphiques et les élèves « observent » que les courbes après des oscillations assez importantes viennent s'approcher des valeurs prévues.
Il leur est proposé un *jeu* qui consiste à deviner quel est le contenu d'une bouteille à la lecture d'une série de tirages qu'elle a produite. Pour cela les élèves peuvent « acheter » des séries de la longueur qu'ils veulent s'ils ont des jetons pour les payer. Cette situation simule exactement le problème du test d'hypothèse.

Le risque

Devoir payer les informations les amène à s'interroger sur les rapports entre la longueur de la série et le risque de se tromper. L'expérience est « décisive » pour eux. Autour des valeurs de convergence 0,2 (1 blanc) ; 0,4 (2b) ; 0,6 (3b) ; 0,8 (4b) ils repèrent des bandes à l'intérieur desquelles les séries finissent par entrer et à rester. En demandant 15 séries de 15 tirages 8 séries seulement sont dans la bande et 7 à l'extérieur, avec 15 séries de 20, on réussit 9 fois et on échoue 6 fois, avec des séries de 100, 15 réussites sur 15 mais avec 160 seulement 14 réussites sur 15. Les séries de 1000 tirages se resserrent beaucoup autour de la valeur théorique… Au vocabulaire près, les élèves utilisent la notion de seuil de signification et d'intervalle de décision.

Réutilisation avec une autre machine

C'est seulement lorsqu'ils ont ainsi pratiqué le genre d'expériences qui leur permet de deviner la composition d'une urne, et de prévoir ou d'interpréter des séries de fréquences qu'on change la machine de hasard et qu'on leur demande de prévoir ce qui arrivera si on utilise d'autres machines de hasard.
C'est seulement alors que l'on a pu introduire le vocabulaire des événements et des opérations sur les événements, puis le calcul sur des valeurs de « probabilité ». Tous les calculs élémentaires dits « de probabilités » ont alors pu être découverts par les élèves ou enseignés. Il faut remarquer qu'à aucun moment nous n'avons parlé formellement de probabilité, ni introduit des spéculations sur la couleur qui allait sortir au cours d'une expérience isolée. De ce fait, les grandeurs manipulées l'ont été dans le cadre d'une pensée déterministe en situation d'incertitude, tout à fait adaptée aux enfants de cet âge. L'expérience s'est étendue sur 32 séances (hors des cours de mathématiques) et dont seulement une dizaine ont duré plus d'une demi heure.

Cette expérience, malgré ses apparences probabilistes, introduit essentiellement des notions de statistiques et elle n'est pas une démarche empirique. Les situations n'ont pour effet que de stimuler des questions, jamais elles ne fournissent de réponse décisive à ces questions. Ce qui a été découvert et enseigné est une démarche, pas une preuve[2].

Le vocabulaire technique relatif à d'autres types de variables aléatoires et les statistiques classiques (moyennes, écart-type, etc.) peuvent être introduits plus tard sur cette base.

L'évaluation immédiate de l'apprentissage

Les observateurs ont posé quelques questions sur des expériences aléatoires sans machine de hasard, assez différentes mais « familières », et qui auraient pu s'interpréter facilement croyaient-ils avec les modèles étudiés avec les élèves. Mais le questionnaire était trop long,

[2] En effet la preuve de la validité de la méthode du test d'hypothèse n'est pas dans une expérience de BERNOULLI effective, mais dans la démonstration mathématique des théorèmes de convergence (central-limite ou même BIENAYME-CHEBICHEV). La validité de nos inférences sur des statistiques nécessairement finies est établie par leur stabilité à l'infini, c'est à dire par des mathématiques et non pas par des expériences.

trop compliqué. Finalement les réponses des élèves ne différaient pas significativement de celles de l'échantillon témoin. Seuls le vocabulaire et les calculs, dans les exemples étudiés, les distinguaient. La conclusion de cette évaluation n'est pas que les élèves n'avaient pas appris les concepts que nous venions de présenter. Les réponses aux exercices sur ce qu'ils avaient vu et sur quoi ils avaient réfléchi montraient le contraire. Mais elle montre les illusions des expérimentateurs qui n'avaient pas à l'époque les moyens de savoir à quel point les questions qu'ils posaient aux élèves étaient éloignées de leur expérience effective. Ce phénomène est fréquent dans l'enseignement, où l'usage abusif des analogies et les succès que leur assurent les effets Jourdain, développe chez les professeurs des illusions flatteuses sur la possibilité pour les élèves de transférer les résultats de leur enseignement vers des soi disant applications, par des voies empiriques directes (mais mystérieuses). Au contraire l'impuissance des méthodes d'évaluation à rendre compte des « connaissances » et des acquisitions de niveau taxonomique élevé conduit à des bilans trop pessimistes.

Conclusions… provisoires

L'étude de cette expérience et de ses résultats ainsi que sa comparaison avec d'autres méthodes sont renvoyées à la fin de l'article, lorsque nous connaîtrons mieux ses motifs objet du chapitre suivant.
Il faut néanmoins retenir que la situation comporte trois systèmes d'objets

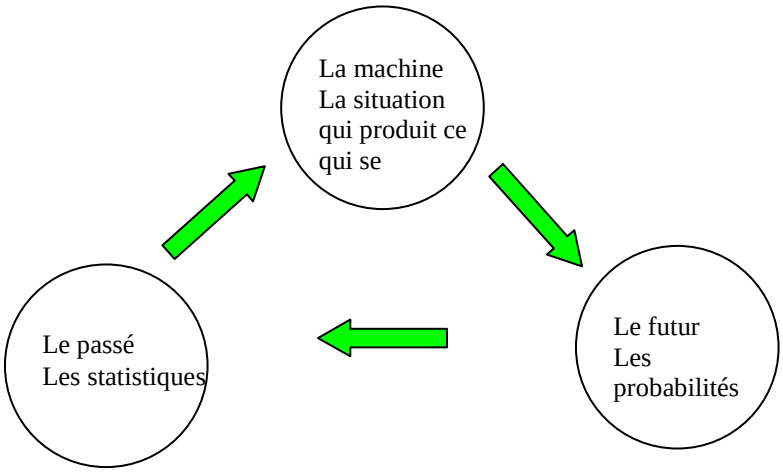
- Ce qui a été vu devient une image du passé d'une machine de hasard, c'est à dire une statistique,
- Ce que l'on cherche à reproduire sont des événements, plus ou moins probables...
- Déterminés par la structure de la machine
- Les observations répondent à des hypothèses faciles à imaginer sur les relations entre ces trois types d'objets

Elle les met en rapport :

- Les raisonnements renvoient d'un objet à l'autre
- Le contenu de la bouteille ne change pas : contenu → hypothèse (probabilité)
- Les observations reflètent le contenu de la bouteille : statistiques → contenu
- Les observations à venir doivent refléter celles du passé puisque la machine ne change pas. Hypothèse (probabilités) → statistiques

Ce qui génère un processus a trois temps dont le moteur est le suivant :

- Si ce que l'on *observe* donne des indications sur ce que *contient* la bouteille alors en reproduisant ce qu'on a fait on devrait *reproduire* ce qu'on a vu. (évidemment, la question implicite est : «Qu'est-ce que toutes ces expériences ont en commun ?»).
- Le passé → la machine → le futur



2. LA DETERMINATION DES SITUATIONS DE STATISTIQUE

DETERMINATION DE LA STATISTIQUE COMME BRANCHE DU SAVOIR

Pour commencer il convient de reprendre des définitions classiques pour identifier les différentes connaissances et activités liées aux principales notions de statistiques.
Un récent échange entre les organisateurs du thème « Statistique » à l'école d'été a bien mis en évidence les différents objets (les statistiques) et points de vue sur ce dont il s'agit (la statistique). Il y a quelques années il semblait qu'il y eut un accord tacite pour n'utiliser qu'un seul terme – singulier- pour englober l'ensemble de ces questions.

> « Statistique[3] : Science qui a pour but de faire connaître l'étendue, la population, les ressources agricoles et industrielles d'un *état* »

disait ce bon monsieur Littré emboîtant le pas à SCHNEITZEL et à ACHENWALL. Et il acceptait l'adjectif « statistiques » pour qualifier « ce qui a pour objet la statistique », (des recherches d'informations par exemple).

> « Le mot « statistique » désigne à la fois un *ensemble de données d'observation* et *l'activité* qui consiste dans leur recueil, leur traitement et leur interprétation, »

disait George MORLAT dans l'Encyclopédia Universalis. Il ne contredisait pas Murray R. SPIEGEL :

> « La statistique est l'ensemble des méthodes *scientifiques* à partir desquelles on recueille, organise , résume, présente et analyse des données et qui permettent d'en tirer des conclusions et de prendre des décisions judicieuses. »).

J'aime cette dernière définition. L'alectromancie consiste à interpréter les suites de lettres sur lesquelles un coq picore des grains de blé, afin de prendre des décisions de justice. Il y a là tout ce qui pourrait faire de l'alectromancie une branche de la statistique, sauf la scientificité.
Certains incluaient même la statistique dans la théorie des probabilités. Arthur Engel par exemple proposait[4] pour l'enseignement de ne pas les présenter séparément et les distinguait ainsi :

> « En théorie des probabilités on calcule des probabilités à partir de probabilités initialement données… en statistique c'est à partir d'observations qu'on évalue des probabilités. ».

Depuis une réaction s'est amorcée et un flot croissant d'approches tend à faire éclater l'unité de la statistique. Pour éviter d'épineuses polémiques nous pourrions interpréter la définition que THURSTON donnait pour l'ensemble des mathématiques

> *la statistique est ce que font les statisticiens.*

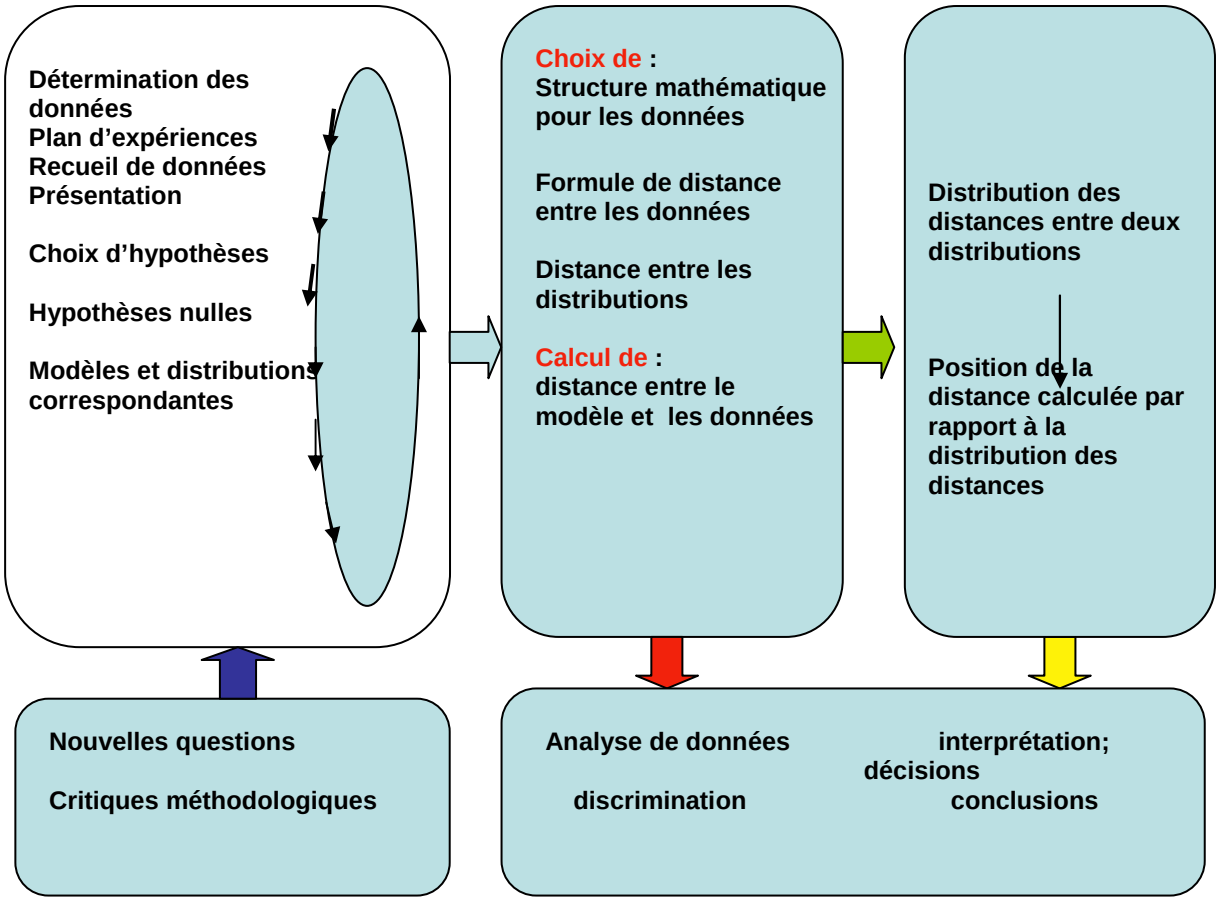
Le recensement de leurs activités ne serait d'ailleurs pas une mauvaise entrée pour la recherche de situations typiques de la statistique. Il nous conduirait à développer la praxéologie des statistiques. Dont nous donnons un aperçu avant de suivre l'autre voie, celle qui consiste à les dériver des définitions générales de l'objet de savoir « statistique ».

---

[3] De Statisticum, qui a trait à l'Etat.
[4] A. ENGEL, « l'enseignement des probabilités et des statistiques » CEDIC 1975.

PRAXEOLOGIE DE LA STATISTIQUE : LES ACTIVITES DES STATISTICIENS

Le schéma ci-dessous détaille la praxéologie de la statistique :



a) Le cycle des tâches de la statistique : détermination des données, plan d'expériences, recueil des données, organisation et présentation des données, choix d'hypothèses, hypothèses nulles, modèles et distributions correspondantes, discrimination. C'est un cycle car le processus peut commencer en un point quelconque, aussi bien par le choix d'hypothèses que par la détermination ou le recueil de données. La vérification de la pertinence, de la validité et de l'adéquation de chaque étape peut demander plusieurs cycles d'études

b) Les *techniques de la statistique*. Une technique s'appuie sur des choix successifs : choix de la structure mathématique pour les données, d'une formule de distance entre les données, d'un type de distance entre les modèles ou les distributions, calcul de la distance entre le modèle et les données. L'utilisateur choisit parmi différentes techniques celle qui lui paraît la plus appropriée

c) *Les technologies de la statistique.* Il s'agit de déterminer la distribution des distances entre deux distributions et la position de la distance observée (calculée) par rapport à cette distribution des distances. La distribution théorique des « distances » entre deux distributions se présente comme une table pour l'utilisateur, Elle est établie par des calculs mathématiques qui relèvent de divers champs théoriques, mais pour l'utilisateur, elle pourrait l'être – sous certaines conditions – par d'autres moyens (simulation, Monte-Carlo, et calculs à l'ordinateur à l'aide de fonctions pseudo aléatoires)

d) L'interprétation de l'analyse des données, et les décisions qui en découlent, les critiques méthodologiques et les nouvelles questions qui relancent éventuellement le processus, constituent un ensemble de nouvelles tâches …
e) Les théories de la statistique, en particulier les justifications épistémologiques et mathématiques. La vérification de la consistance des méthodes statistiques est l'enjeu de la théorie des probabilités et de la modélisation mathématique de la statistique.

Stratégies

Les plus utilisées des stratégies du statisticien peuvent se résumer ainsi[5]

a) Résumer ou modéliser

- *déterminer la contingence,*un ensemble d'observations *(x_i)*
- *choisir un résumé, un modèle, un représentant X (le meilleur de sa catégorie)*
- *indiquer sa valeur comme représentant* par exemple par une distance d(X, *(x_i))*
- *et si cette valeur est trop forte, si la représentation n'est pas très satisfaisante, discriminer la contingence et choisir plusieurs représentants*

b) interpréter le résumé, évaluer sa rareté par rapport à une hypothèse

- *placer le modèle, ou sa distance à la contingence par rapport a une distribution*

c) *extraire progressivement l'information contenue dans les données,*

- *expliquer une partie des variations* (la nécessité)
- *tenter d'interpréter les résidus avec des lois stochastiques* (le hasard)
- *ordonner les modèles successifs retenus en séparant le hasard de la nécessité*

Remarquons que le hasard est ainsi un résidu des explications déterministes conforme à certaines lois.

J'appellerai statisticien une personne qui effectue *une partie significative* (une sous praxéologie) *de ces activités*, c'est-à-dire qui effectue des tâches avec des techniques qu'elle contrôle par des technologies appropriées avec un minimum de connaissances théoriques : en ce sens un enquêteur qui « cueille » des données pourrait être un statisticien.

Comment extraire de cette diversité un modèle général ?

La classification des connaissances doit elle suivre celle des statisticiens ?

Depuis Hammourabi - le premier statisticien recenseur et recensé- il est apparu tellement d'objets et d'activités diverses relevant du même champ qu'il y a des possibilités de distinguer les membres de la communauté par la nature des connaissances qu'ils utilisent[6].

Il est courant de vouloir distinguer

- ceux qui s'intéressent essentiellement à l'objet X et à l'interprétation des données en vue de prendre une décision, les usagers des statistiques
- de ceux qui s'intéressent aux méthodes qui permettent ces interprétations et à leur justification : les techniciens et les théoriciens de la statistique.

Mais parmi ces derniers, d'aucuns veulent ne pas confondre ceux qui traitent ces questions à l'aide des mathématiques et les autres. La ségrégation continue entre ceux qui utilisent des mathématiques déjà faites, plus ou moins récentes, et ceux qui inventent des moyens mathématiques appropriés à des problèmes nouveaux. Et parmi ces derniers il y a évidemment ceux, qui, pour cela, doivent faire des mathématiques réellement « nouvelles » – c'est-à-dire

[5] Voir le diaporama 2 « les stratégies de l'analyse statistique » ou mieux l'ouvrage de même titre BROUSSEAU G. (1993) "Stratégies de l'analyse statistique". (cours pour les professeurs de mathématiques, 80 pages) (LADIST). Vente au DAEST Bordeaux 2.

[6] Je me souviens avoir été surpris par le mélange de ces populations lors de ma première participation à un Congrès mondial des statisticiens. Je n'ai compris qu'après l'intérêt et la signification de ce « mélange ».

qui intéressent les autres mathématiciens, et ceux qui se contentent de bricoler un modèle nouveau avec des éléments bien connus. Ces classifications sont-elles pertinentes pour notre propos ? Autrement dit forment-elles des branches indépendantes de la science statistique ? On peut en douter, mais elles interviennent de façon assez lourde dans la représentation noosphérienne de ce qu'est la statistique et dans les conceptions des enseignements dispensés[7]. Nous reviendrons plus loin sur ces points

Justification de la recherche de situations fondamentales

Mais il me semble important du point de vue éducatif de présenter aux élèves une image correcte, complète, signifiante et positive de toutes les activités liées à cette communauté et pour cela de respecter au mieux les problèmes qu'elle tend à résoudre. Cette visite bien trop rapide des activités des statisticiens aura donné une idée du vaste champ dont l'enseignement doit donner une image. On peut envisager de multiplier les situations didactiques qui illustrent chacune un aspect différent comme le suggère MOORE (voir § 9), mais cette stratégie didactique est douteuse : elle demande beaucoup de temps, elle donne une image trop riche et trop parcellaire et il se peut que jamais le sens d'ensemble ne soit transmis. Il est donc avantageux de disposer d'une « situation fondamentale » qui puisse générer ce champ. La fonction d'une situation fondamentale est de résumer ce sens global de façon à permettre ensuite de déployer le nombre et la diversité de ses occurrences suivant les nécessités et les possibilités de l'enseignement. Ce rôle est comparable à celui d'une connaissance générale par rapport aux connaissances particulières. Nous allons essayer de déterminer une telle situation fondamentale et nous allons essayer de montrer comment le raisonnement sur les situations peut y contribuer. La modélisation que nous recherchons porte sur au moins trois points : les connaissances en jeu, l'activité et les méthodes utilisées, les motivations des actants.
On peut imaginer la situation « fondamentale » comme un modèle général - une situation avec de nombreux paramètres pouvant prendre toutes sortes de valeurs et sur chaque variable des valeurs ou des variantes exclues – tel que *toute situation statistique quelle qu'elle soit entre dans le schéma*. Mais on peut l'imaginer aussi comme une situation plus particulière, où le choix des paramètres et de leurs valeurs est beaucoup plus restreint de manière à ne plus modéliser que l'essence même de la notion mathématique. Cette situation peut servir à générer les autres par d'autres processus que la simple spécification, les questions et les réponses qu'elle suggère, comme dans le processus que nous avons présenté plus haut.

LES CARTACTERISTIQUES DES SITUATIONS STATISTIQUES

Le milieu d'une situation statistique comprend au moins une suite d'informations et son recueil, son interprétation ou son résumé

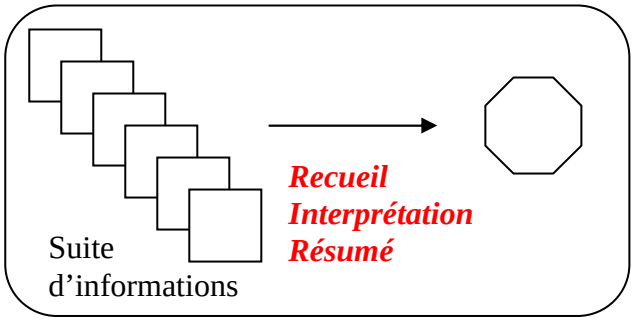


Figure 1

[7] Voir à ce sujet l'annexe 5.

L'interprétation des données

Quel est le rapport du résumé avec les données. Existe-t-il un objet commun caché ?
Les conditions de production de la suite d'informations sont-elles

- des reproductions d'une même situation ?
- ou une exploration progressive de l'objet caché ?
- ou un mélange des deux ?

Selon le cas le résumé sera un composé « logique » d'informations complémentaires ou le résultat d'une interprétation statistique

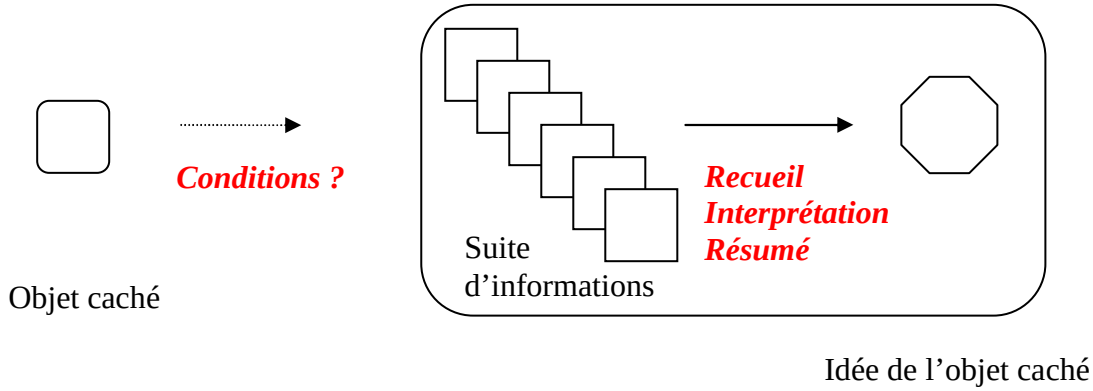


Figure 2

Le Résumé

Pour quoi faire ? La fonction du résumé est essentielle : doit-il permettre de reconstituer l'objet caché, ou simplement de permettre de communiquer de façon économique mais approximative un ensemble de données sans se préoccuper de ce qu'elles peuvent présenter de concrètement similaire ?

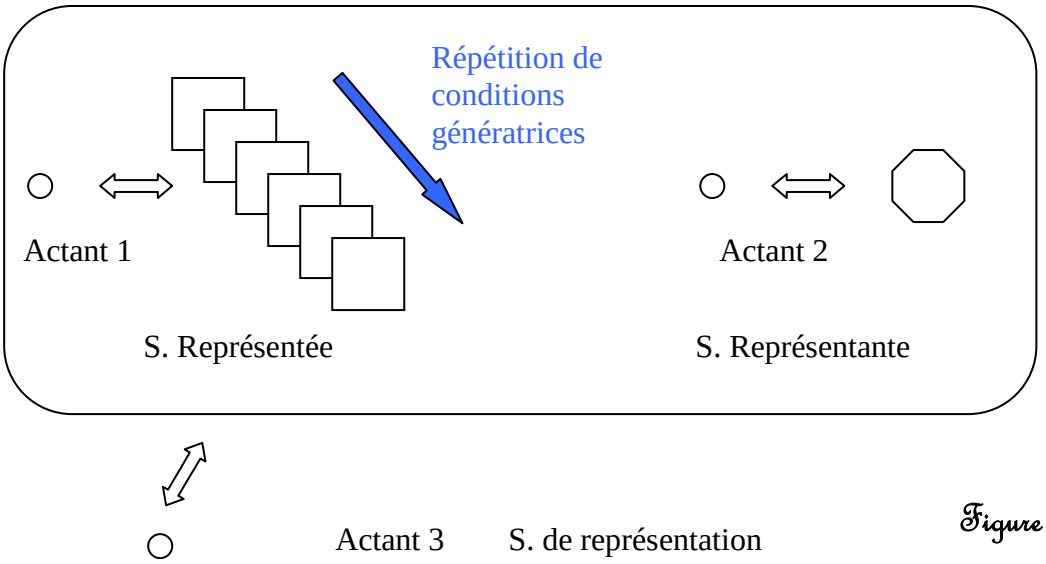


Figure 3

Il faut identifier

- Une situation de production des données,

- Une situation du constructeur ou de l'utilisateur du résumé
- Et une situation impliquant les deux rapports : à la suite de données et à son interprétation : par exemple une situation de représentation[8].

Généralité de cette situation

Cette situation recouvre donc les activités très générales suivantes :

- Reconnaître, connaître les objets et les propriétés,
- créer un répertoire d'identification
- Comparer des objets
- Comparer des classes d'objets, des propriétés
-

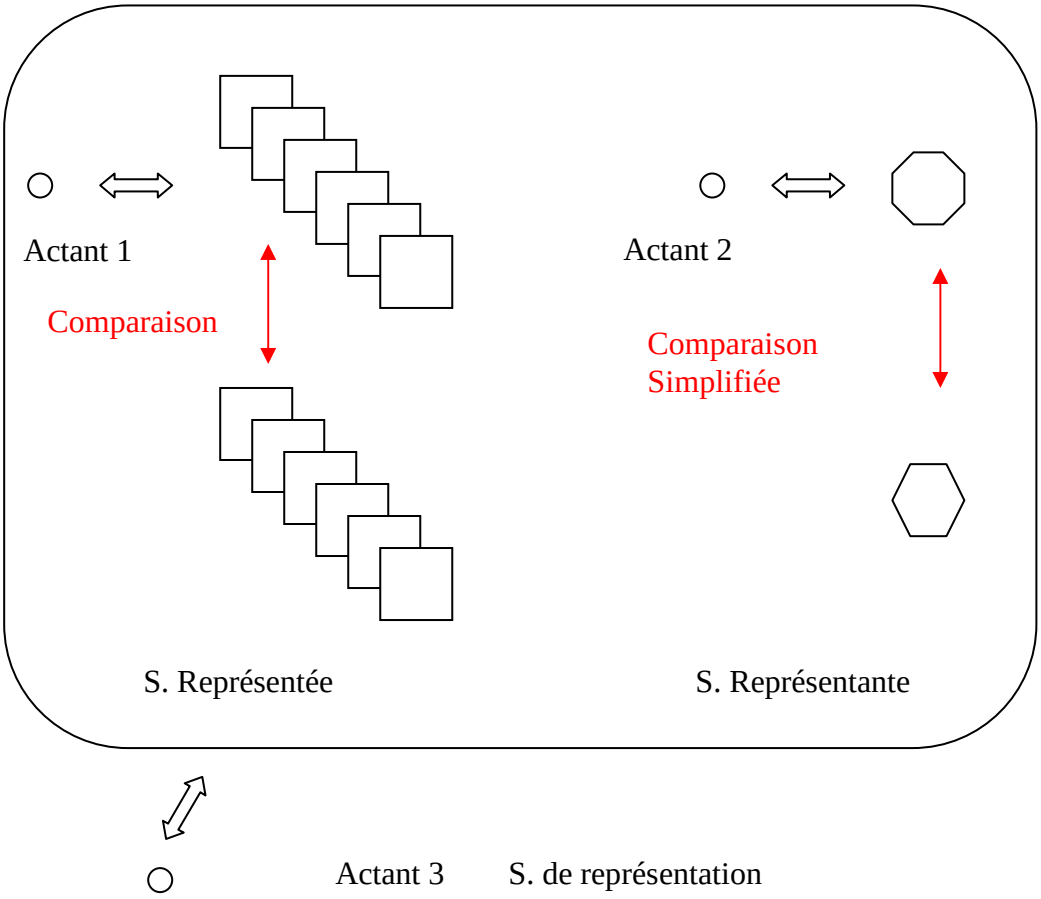


Figure 4

Par conséquent cette situation sert à identifier les occurrences d'un même objet
Et simplifier le répertoire d'identification des objets eux-mêmes.

Exemple 1 : passer d'une liste d'objets désignés par des noms ou des listes de propriétés propres à une désignation commune avec un répertoire simplifié.
Désignation et caractérisation d'objets et d'ensembles
Exemple 2 : représenter une propriété de l'ensemble importante pour un actant

La statistique (comme modèle implicite d'action) apparaît ici comme un moyen spontané d'élaboration de toute connaissance.

[8] Guy BROUSSEAU, « représentants et représentations », à paraître

LES STRATEGIES DES STATISTICIENS

Position

Le statisticien étudie la situation de l'actant 3, la modélise et justifie cette modélisation.

Suivant ce schéma le *dénombrement* d'une collection et la mesure d'une grandeur relèvent de la statistique. Les conventions culturelles ont produit des conditions d'emploi qui permettent d'utiliser les nombres mesures comme des énoncés binaires sans recourir à des distributions de probabilités sur des intervalles.

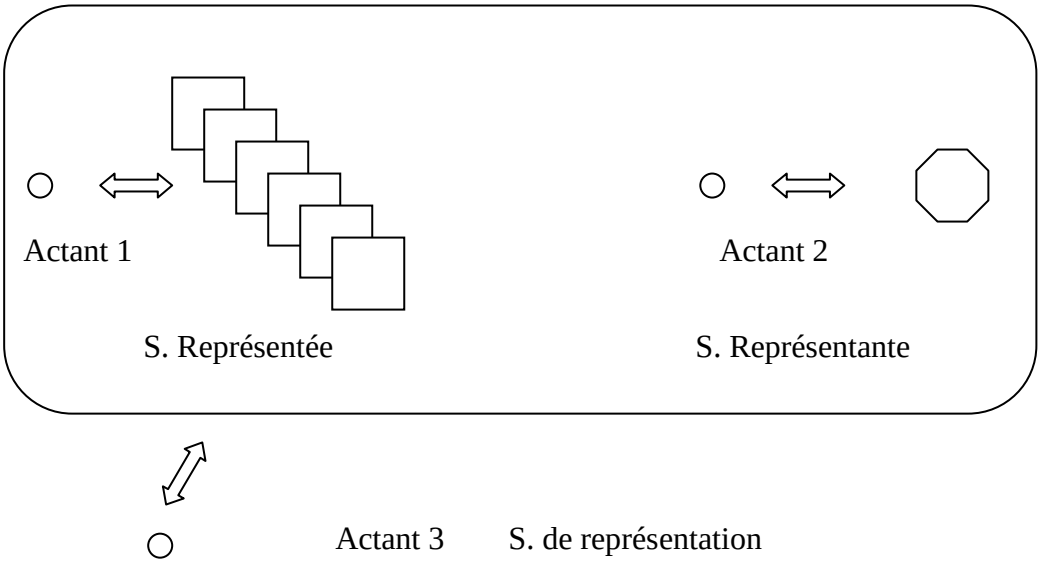


Figure 5

En revenant au schéma de la représentation, pour confronter les informations et l'objet représenté, il est plus courant de les placer dans une même structure mathématique : « nominale » (P(E)), ordonnée, ou numérique.

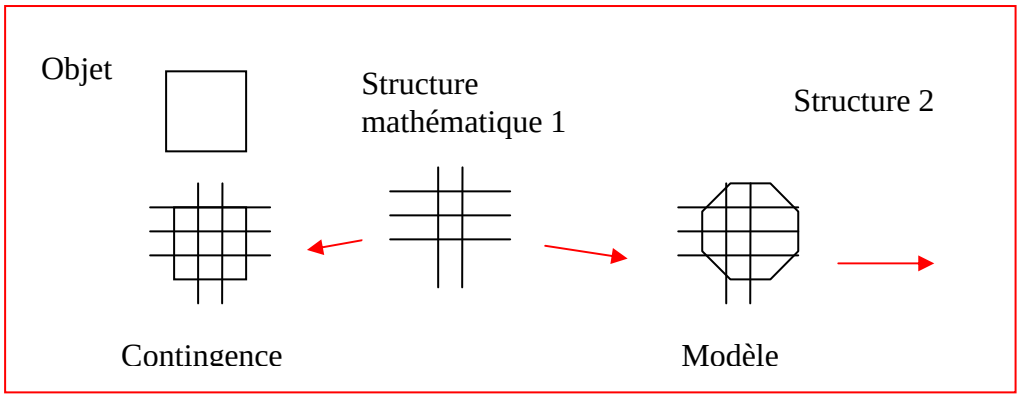


Figure 6

L'ensemble des suites de données, supposées issues d'observations répétées, reproduites et exprimées dans une structure mathématique aussi dénuée d'interprétation que possible est « *la contingence* », ou statistiques primaires, les résumés, représentants, paramètres etc. calculés sont des statistiques dérivées.

L'incertitude

Suivant les situations, l'incertitude est considérée comme un élément du milieu de la situation ou comme un élément de la connaissance du sujet. La différence est très importante
Par exemple une mesure est entachée d'erreurs et la valeur représentante est approximative, mais ni les conditions de la mesure, ni les moyens de l'établir (le répertoire de détermination) ne sont habituellement incertains pour l'actant. Son incertitude est une caractéristique objective de la situation.
Dans d'autres circonstances comme la reconnaissance des structures en l'analyse des données, ou le test d'hypothèse, c'est la connaissance de l'actant qui en jeu ; le répertoire qui lui sert pour identifier les conditions ou l'objet produit est incertain, pour lui en tout cas. Il se trouve alors dans une incertitude subjective. C'est notamment le cas lorsque l'actant est en situation d'élaborer ses processus statistiques.

Incertitude
sur l'existence d'un « objet caché »

Incertitudes objectives	Incertitudes subjectives
sur la reproduction de conditions « identiques »	sur le répertoire qui permet d'identifier ces conditions
sur les propriétés de l'objet caché	sur le répertoire de détermination de l'objet

Première stratégie, la plus générale : représentants et ressemblances

a) représenter des ressemblances et les dissemblances entre la contingence et un *représentant* par des fonctions de distance ou de proximité
- entre deux objets,
- entre un objet et un ensemble d'objets

b) choisir une de ces fonctions : distances (Hamming, valeur absolue, euclidiennes absolues ou relatives, métriques, angulaires, ultramétriques etc.) ou indices selon l'usage dans la situation représentante. Pour des classes d'objets, interpréter l'agrégation des distances comme de l'inertie

Rapporter une distance ou un effectif à un maximum (fréquence, distance relative) prolonge le dénombrement et préfigure toutes les comparaisons futures

c) choisir le meilleur représentant de cette catégorie, celui qui est à la plus petite distance de la contingence

d) indiquer la valeur de cette représentation par cette distance (qui représente la « variation » de la distance choisie entre les objets de la contingence et le représentant.)
La grande variété des structures mathématiques (ensembles, nombres, n-uplets vecteurs, variables aléatoires, jeux, etc.) des objets de ces structures et des types de distances ou de dépendances fonctionnelles génère le champ très vaste de la statistique descriptive et de l'analyse de données.
Cette variété est un des obstacles à son exploration pragmatique et aléatoire. L'implication de mathématiques très variées et parfois très complexes aussi.
Les justifications du choix de l'un ou de l'autre de ces paramètres sont très limitées

Deuxième Stratégie : une mesure unifiée et spécifique : la rareté

Comment comparer la taille d'une puce et celle d'un éléphant

a) L'idée fondamentale : représenter la distance d'un objet à une collection d'objets par la « proportion » des objets de la collection plus éloignés que lui (plus rares).

Une puce qui n'a qu'un millième des puces plus grandes qu'elle est plus grande (dans sa catégorie) qu'un éléphant donné, si 30% des éléphants sont plus grands que lui.
Pour effectuer la comparaison, la collection est distribuée sur la grandeur de comparaison. Les méthodes consistent à

b) Générer des distributions parentes,
   - soit de façon empirique
   - soit de façon théorique

c) Comparer de ce point de vue (de la rareté ou de la vraisemblance) toute statistique observée ou calculée (objet, collection, nombre, mais aussi distance entre objets de la contingence, ou entre contingence et représentants, ou entre représentants …) à la distribution parente (empirique ou théorique) dont elle est supposée être extraite (comme échantillon).

Ainsi cette idée substitue à la comparaison d'objets de même nature mathématique, une comparaison entre un objet et une distribution de cet objet obtenue par la répétition d'une grand nombre d'expériences similaires. L'explication dans une situation donnée est renvoyée à des connaissances qui ne peuvent être acquises que dans une situation formée par la répétition de cette expérience… La borne de ce processus est évidemment le coût de ces répétitions.

d) Envisager suivant ce critère l'acceptation ou le rejet de l'hypothèse :

« tel objet est un échantillon de tel ensemble parent »

Pour cela envisager la reproduction de la situation du statisticien et évaluer les deux espèces de risques qu'il peut prendre.
L'importante conséquence de cette stratégie est qu'il faut rejeter la symétrie classique entre une affirmation et sa négation. Ne pas pouvoir rejeter une hypothèse n'implique pas qu'on doive l'accepter.
La méthode des juges permet de dire si une propriété ou un objet sont réellement « connus » par une population, si le terme signifie quelque chose. Cet exemple montre la profonde parenté de la situation statistique théorique avec la genèse de la connaissance.

La théorie des probabilités

La théorie des probabilités et les mathématiques qu'elles rassemblent sont des moyens

- d'assurer la consistance de ces raisonnements statistiques finitistes en montrant le comportement à l'infini des paramètres attachés aux répétitions supposées mais nécessaires à ces raisonnements.
- de calculer des distributions correspondant à des hypothèses diverses
- de donner un modèle du hasard dans la stratégie suivante

3ième Stratégie. Séparer le hasard et la nécessité

L'idée consiste à diminuer la variabilité de la contingence par des lois, des fonctions des dépendances, des liens de causalité et d'expliquer le résidu par des modèles aléatoires.
Concrètement cette stratégie consiste à

- décomposer la variance d'une variable ou d'une statistique en
  - variance expliquée par un modèle « déterministe »
  - et en variance résiduelle attribuée provisoirement au hasard
- puis à analyser le rapport de la variance expliquée à la variance totale (ou à la variance résiduelle). Si ce rapport est trop faible l'explication est rejetée
- à comparer ce rapport - pour déterminer s'il est grand ou faible - à une distribution obtenue dans des conditions assez générales selon la deuxième stratégie (dist. de Fisher),

- à essayer d'effectuer des partitions (ou discriminations) du corpus si aucune explication ne peut rendre compte de l'ensemble des données.
- à trouver un modèle stochastique qui rende compte du résidu, ce qui met fin au processus d'analyse

Un raffinement de cette stratégie consiste à organiser l'extraction progressive de relations déterministes en fonction des questions du statisticien et des risques qu'il accepte (exemple la régression polynomiale.
Toutes les activités des statisticiens, toutes les méthodes qu'ils emploient peuvent être placées selon ce plan, mais il est possible d'en envisager d'autres, il ne s'agit que d'une description.

LES FORMES DE SITUATIONS ET DE PROCESSUS FONDAMENTAUX

L'élaboration du processus présenté plus haut repose sur une conception de la genèse des connaissances et de leur organisation que nous sommes obligés de préciser maintenant (faute de l'avoir rédigé ailleurs)

1. Première forme de situation fondamentale : situation typique ou paradigmatique

Nous venons de caractériser les éléments fondamentaux des *situations* qui caractérisent la statistique telle que les propositions de départ la définissaient.
Pour déterminer une situation statistique particulière, il suffit alors de préciser comme « milieu » les éléments d'une situation envisagée ci-dessus (figure 3 ou autre) : contingence, structure, modèle (effectif ou supposé) représentation types de structure mathématique de comparaison etc. d'identifier une stratégie de statisticien… et de l'intégrer dans une justification par un usage ou par la réponse à une question…

> Ainsi devoir calculer la moyenne et l'étendue d'une série de données numérique peut représenter la statistique, si nous pouvons justifier cette activité par une situation qui la rend nécessaire.

Nous obtenons ainsi une première forme de « situation fondamentale » de la statistique. Nous pourrions l'appeler situation typique, ou représentative ou générique. Par le jeu de ses diverses variables et de leurs bornes, elle détermine un champ où figure n'importe laquelle des tâches simples ou complexes mais spécifiques d'un statisticien, et elle ne génère que celles là pour peu que les modes de justification soient « scientifiques » comme le désirait Spiegel. Cette proposition reste à vérifier.
La diversité des combinaisons et des intentions ainsi engendrées est très grande. Une collection aussi nombreuse et diversifiée de situations particulières et un schéma aussi général paraissent peu utiles à la statistique. Toutefois, cette forme de définition permet d'organiser systématiquement le champ des connaissances de statistique de différentes façons : suivant les connaissances mathématiques utilisées, suivant l'ordre axiomatique de la théorie de la mesure, suivant la complexité des modèles, ou suivant les types d'applications. Cette souplesse est intéressante pour l'enseignement – déjà pour les concepteurs de curriculums - car elle favorise la diversité des choix didactiques

> Par exemple elle permet de multiplier les exemples isomorphes à un schéma donné pour en favoriser l'acquisition.

2. Deuxième forme de situation fondamentale : situation significative

A l'opposé de cette présentation assez amorphe du champ des situations statistiques il peut être utile de le générer par un autre type de situation et par d'autres types de dérivation.

Il s'agit de trouver une situation qui permette de justifier d'un coup *l'ensemble* des stratégies et des concepts caractéristiques des statistiques, de les articuler par leur *fonctions* réciproques, par leur sens, même si chaque composante n'était obtenue que par des techniques frustres et une technologie pauvre. Cette situation doit donc intégrer toutes les justifications connues de ces composants. Le remplacement de ces composants premiers par d'autres, plus sophistiqués, peut alors se faire au fur et à mesure de la complexification des questions et des acquisitions sans que le sens de l'ensemble soit perdu.
L'étude des stratégies a fait ressortir l'importance clé, de la deuxième : Elle complète et tend à justifier la première et sert de base à la troisième. Une situation fondamentale des statistiques de la deuxième forme devrait par conséquent avoir

- pour solution le test d'une hypothèse, portée sur une situation répétable
- et pour objet - pour les actants – non pas seulement de dire si cette hypothèse est vraie, mais de s'accorder sur le caractère probant ou non des raisonnements statistiques obtenus.

3. Troisième forme de situation fondamentale : situation initiale d'un processus génétique

La développement d'une science s'effectue au cours d'une suite de circonstances où se sont posées des questions, qui ont appelé des solutions, lesquelles ont permis d'investir de nouvelles situations où se sont posées de nouvelles questions … Ces chaînes ont croisé d'autres chaînes, se sont prolongées ou ses sont interrompues et ont formé un réseau par lequel les connaissances se sont développées, transformées modifiées sous l'effet conjugué de nécessités de toutes sortes et de hasards heureux ou malheureux.
Il serait donc absurde de prétendre condenser tous les détails d'un pareil phénomène en une simple « situation ». ? D'ailleurs une image statique et anhistorique serait obligatoirement une image fausse de la connaissance puisqu'elle trahirait sa dimension évolutive et historique
Plutôt qu'une situation unique, une suite de situations pourrait mieux « représenter » en même temps les questions, les réponses, les corrections, les formulations et les justifications de la connaissance visée. Le débat porterait alors sur la qualité de la représentation.
L'épistémologie tend à établir une représentation véridique de ce processus historique. Cette histoire de la genèse d'une connaissance est beaucoup trop complexe pour fournir un modèle utilisable directement en didactique. Chercher une représentation même simplifiée et significative de la *genèse historique* de la statistique, comme nous avons cherché une représentation simplifiée des conditions caractéristiques de sa mise en œuvre et de son étude, ne serait donc pas adaptée pour l'enseignement.
Il s'agit de ne retenir que ce qui est significatif et explicatif de la statistique telle que nous la concevons aujourd'hui, mais dont le sens est construit dans un processus qui simule le processus historique, qui lui emprunte ses lois, ses faits significatifs, qui prend en charge les obstacles épistémologiques légitimes, constitutifs de la connaissance et évite les autres, et qui élimine les hasards et les errements de l'histoire réelle.
En conclusion une situation fondamentale de cette troisième forme se présente comme la dernière d'une suite de situations telle que chacune résulte de la précédente et prépare la suivante avec le minimum d'apports extérieurs. Les situations, les questions et les connaissances qui s'y manifestent doivent se justifier par des causes et des raisons internes. La situation initiale joue dont un rôle essentiel dans le processus qu'elle va générer, elle doit contenir la question à laquelle le processus va répondre.

En conclusion :
Une situation est fondamentale du premier type si elle vise à fournir un modèle qui, par le jeu de ses variables et de leurs limitations, peut convenir à n'importe quelle situation où cette notion intervient.

Exemple « Une situation statistique est une situation où un sujet essaie de reconnaître un objet dans une collection de situations fluctuantes et/ou avec un répertoire incertain ».

Une situation est fondamentale du deuxième type si elle vise à servir de référence, à représenter symboliquement au besoin, ce qui est essentiel dans les objets et dans leurs relations, de façon à pouvoir y rattacher des situations effectives par « le sens », par des « représentations ou par des transformations diverses.

Exemple : celle qui est finalement présente dans l'esprit des élèves à la fin du processus « comment se convaincre de la vérité d'un fait qui reste incertain en répétant des prises d'informations ».

Une situation est fondamentale du troisième type si elle peut engendrer un *processus* qui aboutit à la connaissance de la notion par le jeu des questions qu'elle conduit à se poser, et des réponses qu'elle appelle.

La situation « combien de boules blanches dans cette bouteille » appartient à ce type. Elle aboutit à la connaissance d'une situation du deuxième type

LES PROPRIETES DIDACTIQUES DES PROCESSUS GENETIQUES

La présentation d'une branche des mathématiques à l'aide d'une situation initiale significative et d'un processus génétique fondamental permet-elle d'améliorer l'enseignement de la notion visée ? Il peut sembler en effet que l'on risque ainsi perdre le précieux résultat de l'activité scientifique, de multiplier les temps de recherches stériles, de noyer les connaissances visées dans des circonstances particulières et de les bloquer dans des formes provisoires.

Dans les processus didactiques les acquisitions nouvelles sont obtenues par des mises en relation de connaissances connues selon des processus connus. Par exemple un exposé axiomatique d'une théorie dérive des théorèmes nouveaux par le jeu du modus ponens sur des axiomes et sur des théorèmes connus, mais il existe bien d'autres types de relations explicites ou implicites entre des connaissances. Ce processus de construction est en général interne au savoir. Il ne répond pas à des questions ni a des situations précises, ou alors seulement à des questions liées à l'organisation des connaissances. C'est seulement une fois acquis que le savoir peut être « appliqué » c'est-à-dire utilisé dans des situations particulières. Cette phase d'« application » est souvent le point faible des enseignements basés sur des organisations « arbitraires » du champ. Les exemples de statistique se présentent alors comme des techniques déterritorialisées au sens de CHEVALLARD et les élèves ne voient le rapport que ces exemples épars peuvent avoir avec les « applications » qu'ils rencontreront.

Dans un processus génétique, les connaissances nouvelles engagées dans une situation sont bien construites en relation avec des connaissances acquises (en prolongement ou en opposition avec elles) mais elles répondent à des questions issues de cette situation. Dans la mesure où cette situation est significative de la connaissance (au sens 2) elle a d'étroites relations avec toutes les situations du champ. Les relations entre les étapes de l'apprentissage maintiennent cette relation entre la construction logique, l'adéquation sémantique et l'adaptation économique du savoir construit. Avec cette subordination réciproque des situations, des techniques et des théories, ses principaux avantages sont la préservation du sens et la répétition des circonstances.

Evidemment il est nécessaire que les situations du processus aient les propriétés nécessaires à la création des connaissances locales : qu'elles justifient l'action, la formulation, la justification et l'institutionnalisation aux moments voulus, comme dans tout processus didactique.

3. DIFFICULTES ET OBSTACLES A LA PENSEE STATISTIQUE

Nous allons essayer de montrer comment les obstacles peuvent être rattachés à l'étude de la situation fondamentale et parfois détectés avant de faire l'objet de recherches empiriques effectives.

Obstacles ontogénétiques

Nous avons remarqué que le schéma de situation fondamentale typique que nos avons commencé à discuter plus haut pouvait être retenu comme modèle de construction de toute connaissance. Un sujet acquiert ses connaissances dans la rencontre d'un petit nombre de situations reconnues ou signalées comme similaires, au cours desquelles il les corrige et diminue leur caractère incertain. Par une sorte d'inférence statistique, il construit ainsi un répertoire qui lui permet d'identifier des objets des propriétés et des circonstances. Que ce répertoire lui vienne du plus profond de son capital génétique construit au cours de millions d'années de vie animale puis humaine, qu'il lui vienne de la transmission, intentionnelle ou non d'une culture ou qu'il lui vienne de ses acquisitions spontanées, des ajustements sont nécessaires. L'exercice d'un répertoire pour reconnaître des objets exige des ajustements réciproques. L'origine de l'incertitude de la reconnaissance d'un objet est ainsi double :

- d'une part le répertoire, quoique bien fixé peut ne déterminer qu'une catégorie mais pas un objet unique (exemple : « la planche mesure 1, 50 mètre » peut indiquer « entre 1,495 et 1,505 »)
- d'autre part le répertoire peut être composé de catégories floues assemblées par des connecteurs multimodaux

Les connaissances intuitives et implicites et leur développement obéissent à des lois qui sont loin d'être connues. Elles ont une part d'universalité qui leur permet d'être « comprises » ou enseignées par frayage, mais la prise de décisions dans une société complexe où il faut rendre des comptes exige qu'une partie au moins des répertoires soient composés de variables binaires et d'énoncés non contradictoires. L'importance de ne pas être mis en contradiction est une base universelle de règlement des rapports entre humains. Nous en avons tiré les conséquences en adoptant le modèle de LORENZEN comme archétype des situations mathématiques. La mise au point des répertoires logiques et déterministes accompagne dans l'éducation des enfants la dénomination des objets et des actes les plus simples. Les résistances des enfants à l'entrée dans ce système sont l'indice de sa source culturelle et la trace de l'obstacle qu'à dû constituer la pensée stochastique primitive à l'établissement de la rationalité.

La situation caractéristique (fig. 3) peut se décliner pour produire l'émergence des catégories logiques. Nous l'avons montré dans nos études sur la désignation des objets, des listes d'objets, des propriétés et des caractères distinctifs (le jeu des trésors) et sur leurs dérivés. Ces études prolongeaient celles de WERMUZ et de PIAGET qui explicitaient les conditions d'émergence des foncteurs et des prédicats par la décantation des composantes contextuelles à partir de prédicats amalgamés et de préfoncteurs.

Ainsi la pensée logique n'est pas un prolongement « naturel » de la pensée stochastique et les modèles qui tendaient à le croire ont échoué. (Il faudra s'en souvenir au moment de la défense trop naïve de certaines méthodes d'analyses de données). La première est en compétition avec la seconde et lui fait obstacle dès la petite enfance. Elle recevra par la suite l'appui d'une culture abondante et ancienne. C'est pourquoi il n'est pas possible d'envisager de dériver les connaissance de statistique du processus didactique bien connu : modèles implicites → formulation → validation.

Lors de l'émergence historique ou didactique d'une connaissance scientifique de la statistique nous nous trouverons en présence d'un double obstacle : d'une part, celui de la pensée rationnelle et déterministe, acquise par tant d'efforts par la culture et par les enfants au cours de leur développement, et d'autre part celui de la pensée stochastique primitive renforcée par toute sorte d'obstacles épistémologiques que nous allons survoler maintenant.

Les difficultés liées à la dualité des langages

Nous trouvons aussi une difficulté liée à la dualité des moyens linguistiques de la détermination des objets et des classes : par des noms propres ou par des propriétés, par des listes de noms ou par des propriétés d'ensembles, etc. Cette dualité de langage établit une sorte d'équivalence linguistique et « logique » entre un élément, un élément générique et la classe qu'il parcourt. Les mêmes mots désignent les objets et les classes, les fonctions et leurs images etc. L'abus, dénoncé mais fermement maintenu en mathématiques, est partout assez anodin sauf quand le passage de l'ensemble à son représentant ou l'inverse sont problématiques. L'usage sans précaution de la formulation logique classique dans les situations statistiques sera une source permanente d'illusions et d'erreurs pour les débutants, et peut être un obstacle.

Obstacles épistémologiques

Les connaissances « empiriques » et ces modèles spontanés se composent d'une foule de raisonnements et d'objections dont le recensement sort du cadre de notre étude et qui forment des *obstacles épistémologiques spécifiques* (au sens de Bachelard[9]).

a) Représentant et dispersion

L'intérêt de l'actant pour un certain représentant d'une collection d'informations (il faudrait utiliser un adjectif comme « statisticiable » pour indiquer qu'il existe une situation répétée qui les produit) est interprété par une situation propre. Elle détermine donc le genre de représentant cherché.

Par exemple, s'agissant des températures de cet été, un journaliste voudra mettre en évidence leur caractère exceptionnel. Il citera et comparera des *maximums*. Yves CHEVALLARD parlait à ce sujet de situations de compétitions (banales dans les sports) qu'il opposait à des situations de coopération orientées vers la recherche d'un *représentant moyen*. Ce n'est que lorsque le journaliste voudra s'occuper de la durée de la canicule qu'il devra commencer à devoir chercher à intégrer les valeurs ponctuelles et donc à s'intéresser aux tendances centrales. *Ainsi notre approche nous conduit à opposer, non pas la *moyenne* à la *dispersion*, comme on le fait fréquemment[10] mais un *représentant individuel* (de tendance centrale ou extrémal) à un *indice collectif de dispersion*.

L'existence de deux classes de situations, les unes favorables à l'étude de la représentation, les autres à celle de la dispersion, a été observée depuis longtemps. Il n'y a pas de preuve formelle que l'une fasse obstacle à l'autre et de nombreuses situations les conjuguent sans encombre. Pourtant nous avons des traces historiques des débats sur l'interprétation de l'une et de l'autre (les querelles sur l'homme moyen[11] par exemple).

A ce sujet il est intéressant de noter que le rapport entre la capacité pour un élément d'être un représentant de sa classe est souvent lié à la possibilité de l'actant de lui attribuer une

[9] Gaston BACHELARD. *La formation de l'esprit scientifique* Vrin 1938 réed.1983

[10] Elle a été évoquée récemment par un membre de l'Académie pour expliquer le retard des français dans la connaissance et l'utilisation des statistiques. (Cette hypothèse qui rappelle assez malencontreusement celles fondées jadis sur « l'âme des peuples[10] » m'a paru un peu badine).

[11] L'œuvre de QUETELET et les critiques de E. DURKHEIM et de M. HALBWACHS

propriété spécifique : le billet 1234567 paraît un moins bon représentant de l'ensemble des billets d'une loterie que le billet 1805532.

Mais l'intérêt pour les tendances centrales correspond bien à une centration sur les lois et les nécessités et que l'intérêt porté aux dispersions ou aux fluctuations correspondent bien à une mise en doute des premières et à une centration sur le hasard. L'opposition des deux est claire, constitue-t-elle un couple d'obstacle ? Cela peut y ressembler car, du moins, on peut voir une sorte de moteur de la recherche et de l'évolution de la statistique, illustrées de nombre de débats.

b) l'observateur de données et l'actant engagé

Les deux types de situations que nous venons d'évoquer ont en commun de placer le sujet (l'actant qui examine les données et leur représentant) en position réflexive, extérieure, intemporelle, par rapport aux situations productrices des données. Cette situation est conforme à notre schéma. Le « statisticien » est hors de l'action. C'est un sage qui évoque une suite d'événements passés et y cherche pour des raisons quelconques, quelque régularité, quelque singularité ou quelque leçon.

Mais la série des situations productrices de la série des données ne peut pas être entièrement ignorée, et avec elle l'évocation du point de vue de l'actant (n°1), celui qui est engagé dans chacune de ces situations. Pour lui chaque donnée est le résultat d'une action en cours, elle a une temporalité, un enjeu, et le résultat des actions passées est engagé les décisions de l'action suivante, les statistiques se font naturellement prévisions.

L'introduction de la temporalité, de l'ordre, de l'évolution des données préfigure la sortie des situations de description statique et l'entrée dans les jeux de la prévision auxquels sont attachés une autre famille d'obstacles, ceux attachés à la pensée probabiliste mais qui vont largement déborder sur la statistique. Nous devons l'examiner pour des raisons de didactique car l'introduction des statistiques est presque toujours associée à des considérations probabilistes ou au moins fréquentistes.

Avec la centration sur la situation répétitive, productrice des données, le doute est alors renvoyé

- au sujet : l'obstacle « personnaliste » (chance, vertus, adresse de l'actant)
- au milieu : l'obstacle « animiste » (les conditions ou les objets favorables, bienveillants ou malveillants)
- à la règle du jeu :
  - à une justice immanente : la nature exerce une loi de compensation ou au contraire d'amplification (l'imbécile « loi des séries »)
  - à un être supérieur qui décourage l'analyse :
    - volonté et desseins impénétrables → fatalisme
    - obstacles moraux : chercher à comprendre est une offense à Dieu (arbre de la science.. du bien et du mal)
  - absence de règles, dieu indifférent
  - A l'enjeu : les probabilités des événements varieraient en fonction de l'enjeu. L'obstacle que constitue l'investissement de l'actant n°1 de notre schéma, dans les enjeux de chacune de ses actions répétées est tout à fait typique du rôle d'un obstacle à la fois de point d'appui (le calcul des probabilité est né du jeu et de la distinction entre l'espérance mathématique et la probabilité) et de difficultés récurrentes associées à ce point de vue. Nous y reviendrons dans l'analyse des obstacles didactiques.
- ou à une combinaison de ces éléments

L'entrée en scène du temps dans l'analyse des ensembles de données introduit le concept de *cause* avec une nouvelle source d'incertitude tout dépend-il de tout ? C'est seulement notre incapacité à suivre les conjonctions de multiples causes qui nous empêche d'utiliser notre vision déterministe et nos concepts logicistes (GAUSS 1777-1855).
On voit que l'histoire de la statistique est profondément liée à la conception du monde et de la culture, ce qui ne devrait pas resté ignoré dans la conception de sa didactique.

c) Le renoncement aux situations logiques élémentaires et son prix

En fondant l'attitude scientifique moderne sur l'expérience du doute DESCARTES marque sans qu'il y paraisse trop une étape importante dans le développement de l'esprit statistique

d) L'objet caché : mythe, réalité objective ou instrument subjectif ?

Le fait de savoir si une suite de données est l'expression répétée d'un objet caché ou non ne s'est pas posé avec une grande acuité avant que l'on interprète de façon temporelle la production de ces données et qu'on veuille les rattacher à une situation reproductible donc qui introduit une part de prévisible. Mais dès lors que l'on accède au concept de modèle aléatoire, les questions précédentes se reposent quel est le composant qui supporte le hasard :

- les événements sont-ils objectivement aléatoires ? Ce point de vue « objectiviste » est celui des » *CARDAN 1501-1578, PASCAL 1623–1662 et JACQUES BERNOULLI 1654-1705)*
- ou le langage des probabilités n'est-il qu'une forme de relations logiques, un langage sans rapport direct avec les événements ? *C'est le point de vue « logiciste » de KEYNES J. M. 1883- 1946, d'Harold JEFFREYS F.R.S. 1891-1989, et de Charles Sanders PEIRCE 1839-1914)*
- la probabilité est seulement la limite d'une fréquence. Ce point de vue empiriste des « probabilités statistiques » est celui de *POISSON 1781-1840.*
- Le point de vue BAYES-fiduciaires raffine celui de POISSON : la statistique formalise seulement l'induction : ce sont l'information de l'expérimentateur et sa confiance qui évoluent au fil des répétitions. Le calcul des probabilités n'est que l'instrument de cette dialectique, en statistique ce sont les questions du statisticien et ses doutes qui dirigent le choix des méthodes. BAYES (1702-1761) [12].

Ces différents points de vue s'opposent et chacun semble déterminer un obstacle spécifique aux autres. L'entremêlement des dates pourrait faire croire le contraire à qui penserait que les obstacles tombent les uns après les autres dans une chronologie simple et inéluctable.
Le rôle du temps et celui de l'interprétation causale des faits est particulièrement visible dans les résistances au point de vue Bayes-fiduciaire.

Obstacles didactiques

Nous distinguerons les obstacles microdidactiques et les obstacles macrodidactiques. Les premiers sont attachés aux choix didactiques supposés totalement libres d'un ordre et d'un mode d'introduction des différents concepts de statistique ainsi que celui des situations où ils se manifestent. Les obstacles microdidactiques sont les obstacles « nouveaux » provoqués par ces choix. Le fait pour une méthode didactique d'utiliser, d'augmenter, de diminuer ou d'effacer l'effet d'obstacles épistémologiques est tout à fait important, mais même sous des formes différentes, les obstacles épistémologiques restent identifiables.
Les obstacles macrodidactiques proviennent des limitations dans les choix didactiques imposées par les pratiques et les connaissances ayant cours dans la société, dans les institutions, du fait même du fonctionnement de ces institutions. De façon un peu simplifiée

[12] D'autres conceptions ont existé comme celle de « probabilités psychologiques » (COHEN 1960) pour analyser les notions de hasard, de chance, d'adresse, l'appréciation des causes, de l'incertitude et de la confiance, comme attitudes ou comportements psychologiques mais elles ne définissent pas un objet consistant.

les obstacles macrodidactiques sont ceux qui résultent de l'incapacité des grands systèmes socio-culturels à assumer leurs responsabilités didactiques.

a) Obstacles microdidactiques

Ce sont autant des générateurs d'obstacles que des obstacles
- Les approches philosophiques et épistémologiques, comme obstacles didactiques auxquels répond le point de vue minimaliste du simple « calcul des probabilités »
- L'absorption de la statistique par les probabilités depuis l'accession du calcul au rang de théorie grâce à la théorie de la mesure a fait de la pensée probabiliste un obstacle à la pensée statistique. En particulier l'interprétation prématurée des probabilités par des fréquences – pour des raisons d'illustration – constituent un effet « Jourdain » qui enlève toute justification aux théorèmes de convergence et donc à l'étude mathématique des statistiques qui peuvent passer au rang de commentaires inutiles pour des auteurs comme MOORE
- La subordination de la problématique proprement statistique à l'ergonomie didactique et/ou mathématique qui modifie la nature, la signification et la fréquence des questions à traiter. La recherche de la « rentabilité mathématique et didactique d'un enseignement » crée des obstacles par la confusion qu'elle introduit sur les relations entre les situations et les méthodes (peut être la déterritorialisation à laquelle Y. CHEVALLARD fait allusion [CHEVALLARD, 1978]). Cette confusion a une cause : le souci des professeurs de « rentabiliser » leurs enseignement les conduit naturellement à chercher à appliquer le plus possible les notions déjà enseignées dans les leçons nouvelles, et à choisir dans les secteurs nouvellement abordés les concepts qui auront le plus de chance de s'appliquer dans les questions mathématiques étudiées par la suite. Cette tendance justifiée par des raisons purement didactiques favorise évidemment les mathématiques dites « pures » au détriment des mathématiques appliquées. Les questions fondamentales de probabilités ou de statistiques sont ainsi *cannibalisées* par des questions mathématiques étrangères au sujet.

Cette tendance est particulièrement marquée en France où elle est liée à l'histoire même des recherches en statistiques et où le rapport des professeurs de mathématiques aux statistiques reproduit et accentue ce phénomène[13].

Les textes d'accompagnement des nouveaux programme 2002 (les programmes sont toujours nouveaux comme le Beaujolais) indiquent :

> « Ainsi les expériences de référence classiques (le plus souvent simulées en terminale) sont indispensables pour comprendre la théorie des probabilités, elles ne sont cependant pas de nature à convaincre les élèves de l'importance de cette théorie en mathématiques comme dans les autres sciences. Aussi, le programme de probabilité de la série scientifique a *partie liée* avec un autre chapitre important du programme de mathématique de terminale concernant l'intégration (loi uniforme sur un intervalle borné et loi exponentielle) et une convergence thématique forte apparaît avec le chapitre « Radioactivité » du programme de physique : en physique on étudie la radioactivité.

Ce qui montre que l'accent est mis davantage sur l'importance des applications de cette théorie aux sciences et aux mathématiques elles-mêmes que sur sa validité mathématique et surtout sur sa signification épistémologique. Les problèmes relatifs aux croyances erronées et aux modèles spontanés sont ignorés. Et il est vraisemblable qu'il s'agit de convaincre les professeurs au moins autant que les élèves de l'importance de ces applications.

[13] Ce point a fait l'objet d'une étude comparée du rôle social des statistiques et de la recherche en statistique en France et aux Etats-Unis (1850-1970) et de son influence sur l'enseignement, communiquée mais non publiée (annexe 5)

- L'apparition des ordinateurs et des générateurs pseudo aléatoire a profondément bouleversé et la statistique, qui peut utiliser des instruments mathématiques complexes, et même des méthodes non maîtrisées comme l'analyse neuronale et les probabilités dont le calcul des systèmes et des automates est puissamment facilité. L'usage de ces instruments dans l'enseignement permet déjà des effets JOURDAIN extrêmement violents (par exemple faire apparaître les modèles pseudo-aléatoires et la convergence des fréquences vers eux comme des lois empiriques) qui ne peuvent manquer d'avoir des conséquences[14] didactiques et scientifiques.

b) Obstacles macrodidactiques (ou socio-culturels)

L'intégration du milieu, y compris le milieu socio-culturel et professionnel des enseignants et des élèves dans les objets de la théorie des situations permet aujourd'hui d'identifier d'autres types d'obstacles dits « macro-didactiques ». Ces obstacles à la diffusion d'une connaissance proviennent des rapports des diverses institutions d'une société à ce savoir. En statistique notamment de nombreux auteurs (par exemple Daniel SCHWARTZ) ont noté que la Grande Bretagne et les Etats-Unis bénéficiaient de conditions socioculturelles beaucoup plus favorables à cet enseignement que la France. Les difficultés proviendraient des représentations (au sens de la sociologie) que ses font ces institutions de la façon de pratiquer, d'apprendre ou d'enseigner ce savoir en le reliant aux mathématiques (esprit cartésien, individualisme, méconnaissance de la variance etc.), qui entraîneraient des difficultés récurrentes. Malgré l'existence de solutions microdidactiques et la possibilité évidente de les mettre en œuvre, le développement ne se ferait pas pour des raisons de ce genre. Le phénomène s'observe dans différents secteurs des mathématiques mais chaque fois de manière spécifique à un certain savoir. Ce type d'obstacles est particulièrement évident en statistique et demanderait une étude qui sort du cadre de ce cours.
L'analyse du processus de 1972 montrerait comment il a évité toute une série d'obstacles en changeant la position des élèves en retournant la situation en quelque sorte. Au lieu d'être des joueurs soucieux des résultats prochains ou des variations de la machine de hasard, ils ont été d'emblée des analystes dans un « sondage » sans population parente apparente, armés de leurs seuls raisonnements « déterministes ». Le test d'hypothèse apparaît comme ce qu'il est : un moyen d'éliminer les hypothèses les plus improbables en contrôlant le seuil de risque. La question de la reproduction à l'infini de l'opération et de la stabilisation progressive des paramètres reste ouverte aux mathématiques. Les élèves traitent les probabilités comme des limites de fréquences, mais les ont apprises implicitement dans des circonstances bayésiennes comme des résumés d'informations progressives sur un système, ce qui peut permettre par la suite une élucidation de ce type.

RETOUR SUR LES EXPERIENCES DE 1971-73 ET CONCLUSIONS

Les résultats

Je renvoie les auditeurs de ce cours au texte de l'article de référence pour le détail de ce que nous a enseigné cette expérience dans les domaines de la psychologie, de la didactique de l'épistémologie et de la sociologie de l'éducation.
Du point de vue de l'enseignement de la statistique et des mathématiques, il nous semble que les élèves ont compris l'usage de toutes les techniques mathématiques utilisées (rapports,

[14] Effet Jourdain : le professeur obtient de l'élève une réponse spontanée à un problème banal puis identifie cette connaissance avec une connaissance complexe voire scientifique de telle façon que l'élève est abusé sur la valeur de son acquisition.

fréquences, comparaison de nombres et de suites, graphiques, etc.). Mais le premier résultat est que les élèves ont pu connaître véritablement les principaux objets de la statistique et des probabilités dans leurs relations essentielles. Ils ont pu en comprendre le rôle et en utiliser les techniques. C'était le point important de la démonstration. Il restait bien sûr à continuer le processus dans deux directions :

- le « reconnaître » dans d'autres exemples de façon à le détacher d'un contexte nécessaire mais beaucoup trop particulier pour lui donner sa valeur de paradigme de la statistique, le déterritorialiser, sans lui faire perdre son sens.
- l'étudier, l'institutionnaliser c'est-à-dire transformer en objets de savoirs les connaissances issues de cette première rencontre, qui, très riche ne peut pas porter à leur terme tous les apprentissages de toutes les notions utilisées, utiliser les techniques, définir les notions etc.

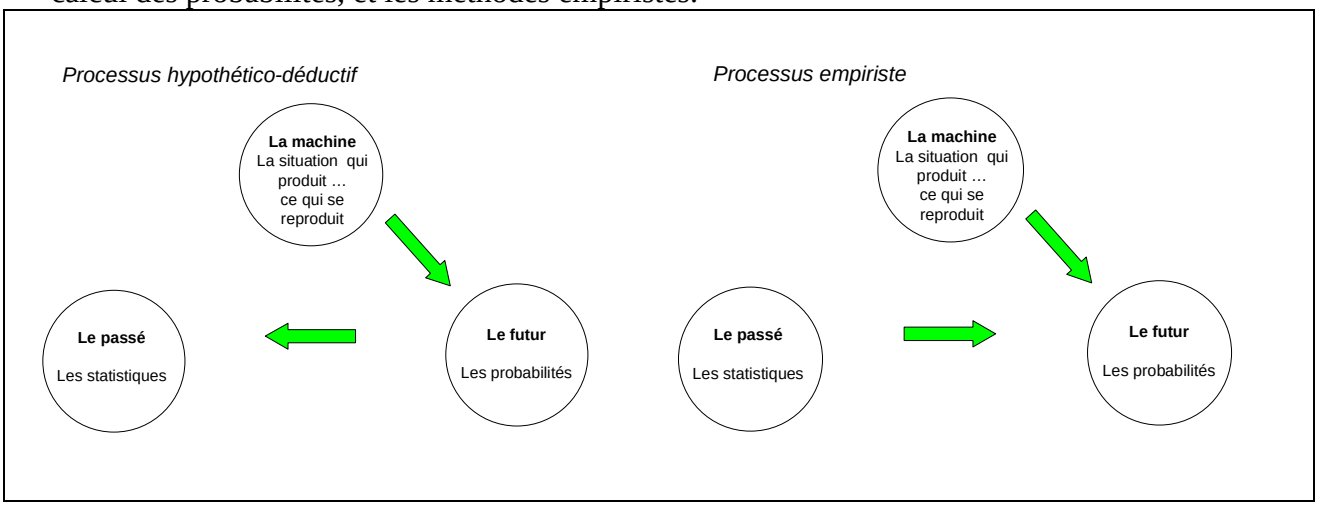
Accessoirement nous avons été satisfaits de prouver la possibilité de faire se dérouler cette longue suite d'activités, de constater que les calculs que nous avions faits a priori des variables de la situation étaient corrects, et de voir que les élèves se sont prêtés à l'expérience avec, dans l'ensemble, une très bonne volonté et quelquefois avec enthousiasme.

Accessoirement encore, nous avons observé les grandes difficultés des professeurs à assumer que leur cours dépende d'un processus aléatoire. C'est peut être pourquoi, leurs collègues en l'absence d'appui préfèreront souvent à la responsabilité d'une construction du savoir, le confort de la position empiriste et dogmatique. Leur justification est alors : le monde est comme ceci ou mon savoir est comme cela. Nous en avons conclu que nous ne pouvions pas recommander de faire « reproduire » ces activités hors du contrôle par des statisticiens didacticiens.

Pourtant dans notre projet de l'époque, cette situation et le processus qui la suivait devant en faire une situation de référence destinée à éclairer et replacer toutes les rencontres que l'on pouvait organiser au collège puis au lycée avec la variété des exemples. Comme nous l'avions fait auparavant avec les naturels et par la suite avec les rationnels ou la géométrie, nous visions à produire une situation et un processus de référence, comme moyen optimal d'organiser la construction du savoir. Mais notre traité ne pouvait être reçu que comme un exercice parmi d'autres ce qui lui enlevait toute signification et toute efficacité.

Comparaison avec d'autres méthodes

Il est intéressant de comparer la situation expérimentale proposée ci-dessus aux situations classiques proposées par les enseignants. Dans celles-ci ont peut distinguer les processus hypothético-déductif où les statistiques sont les « preuves expérimentales » de la validité du calcul des probabilités, et les méthodes empiristes.



Ces méthodes utilisent les relations entre les trois systèmes fondamentaux, mais elles ne sont pas dialectiques.

Dans ces méthodes les élèves répètent des expériences qu'ils ont analysées eux-mêmes préalablement avec le calcul probabilités. Ils tirent des conclusions au simple vu de la proximité entre la série empirique observée et les probabilités calculées. Ils écartent ainsi toute référence à la sécurité de leurs conclusions.
Avec les ordinateurs et les générateurs de suites pseudo aléatoires il est facile de « donner à voir » la spectaculaire convergence des fréquences, le resserrement des fluctuations de faisceaux de séries autour des valeurs de convergence comme nous l'avons fait avec l'animation, et de donner ces phénomènes comme des lois naturelles… Ici on procède a une double falsification, pire encore, en appuyant la conviction des élèves sur une argumentation empirique, comme si on pouvait produire une preuve expérimentale directe de la validité du calcul des probabilités et de la convergence.

Cette position enlève toute motivation à la recherche de la consistance, donc à l'analyse mathématique. Elle coupe court à toute justification de la statistique inférentielle et de son approche probabiliste, et on peut satisfaire le projet de MOORE : débarrasser la statistique de l'encombrante mathématique.

Une comparaison plus précise fait ressortir les différences entre les situations, entre les demandes du professeur et entre les productions des élèves.

a) Comparaison des situations

S. empiriques :	S. expérimentale
•La machine est connue •Souvent, elle est analysée a priori avec les probabilités •Deux machines physiquement identiques sont tenues pour statistiquement équivalentes •Consigne: –répéter des « tirages » –Ajouter les résultats, etc. •La fréquence est définie et indiquée comme objet d'observation	•La machine n'est pas entièrement connue, mais les possibilités sont simples •Aucune analyse a priori •L'utilisation de la machine n'est pas indiqué, ni l'objet des observations •Suggestion : Si les faits passés (statistiques) sont déterminés par la composition de la machine, alors les « nouveaux » faits devraient être similaires aux passés

b) Demandes du professeur

S. empiriques	S. expérimentale
•Réitérer les tirages (a. r) un nombre arbitraire de fois •Observer le comportement de la fréquence et ses fluctuations •Distinguer la valeur limite malgré les fluctuations	•Deviner la composition de la machine •Se convaincre de la vérité de sa conclusion y pouvoir l'argumenter

•(Par exemple écarter l'hypothèse d'une distribution uniforme),
•Mettre en rapport les résultats avec une distribution uniforme à l'aide d'un calcul de combinatoire d'après la composition de la machine

c) Connaissances» que les élèves doivent produire

S. empirique
•L'expérience s'approche de la théorie
•Les élèves croient que l'expérience doit nécessairement «confirmer la théorie»
•Preuve : si la série s'écartait obstinément de la prévision (expérience truquée à l'insu du professeur) la situation didactique serait sans solution et la machine serait rejetée

•S. expérimentale
•Réitérer les tirages et noter les observations
•Dégager les objets statistiques de l'observation : fréquences et distributions
•Les hypothèses
•Les raisons de la réitération ou de la décision de terminer l'observation. Las raisons évoluent suivant les hypothèses successives
•Le principe du test d'hypothèses
•Comment dire si deux machines sont statistiquement équivalentes ou non.

Confrontation avec les recommandations de l'Académie des sciences aujourd'hui

En ce qui concerne la conception du dispositif, il est intéressant de le confronter aux recommandations de deux rapports récents de l'académie des sciences et aux projets de la nouvelle réforme. Ce sera un moyen d'évaluer leur actualité, leur modernité devrais-je dire. Les principales suggestions sont issues d'un article de MOORE rapporté par ESCOUFIER
MOORE propose de

> « - 1. Favoriser les éléments de pensée statistique : a) le besoin de données b) l'importance de la production de données c) l'omniprésence de la variabilité d) la mesure et la modélisation de la variabilité
>
> - 2. Incorporer plus de données et de concepts, moins de principes et de démonstrations, partout où c'est possible, calculs automatiques et graphiques. Un cours d'introduction devrait a) relier étroitement les données au réel (mais pas purement réalistes) b) développer des concepts de statistique : causalité contre association, expérimental contre observationnel, et études longitudinales contre croisements de sections, c) recourir aux ordinateurs plutôt qu'aux formules de calcul,d) traiter les démonstrations comme d'importance secondaire
>
> - 3. Accueillir et encourager les apprentissages actifs à l'aide des alternatives suivantes au cours magistral : a) groupes de résolution de problèmes et de débats, b) exercices de laboratoires, c) démonstrations basées sur des générateurs de données, d) présentations écrites et orales, e) projets de groupe plutôt que projets individuels »

Chaque suggestion des deux premiers points parait avoir été écrite pour caractériser notre expérience (sauf le 2d qui nous paraît non pertinent dans notre expérience et fortement négatif). Le troisième point convient aussi, bien qu'il ne soit pas clair ici si ces options sont

liées à la construction du savoir comme dans notre expérience où s'il s'agit d'options pédagogiques générales établies dans des études indépendantes : par exemple débats n'ont d'intérêt que s'ils portent sur des points importants et signifiants pour les élèves et pour la connaissance en jeu.
De son côté, le rapport de la CREM[15] poursuit, presque dans le même sens :

> « L'objectif d'une initiation aux probabilités et à la statistique aux niveaux collège et lycée est d'enrichir le langage, de repérer les questions de nature statistique, de définir des concepts qui fonderont un mode de pensée pertinent, rassurant, remarquablement efficace. Les modes usuels de représentation graphique (histogrammes, diagrammes en bâtons notamment), c'est-à-dire les éléments de base du langage graphique de la statistique sont aujourd'hui enseignés en collège et une introduction à l'aléatoire, appuyée sur le calcul des probabilités et la simulation, est proposée dans les nouveaux programmes de lycée. »

Toutes les situations évoquées sont comprises dans la nôtre.

L'épineuse question des différents niveaux scolaires

Le processus que nous venons d'étudier peut-il être utilisé à tous les niveaux scolaires ? Sa construction générale n'a pas dépendu de conditions particulières de la part des élèves, il peut donc être « montré » comme situation fondamentale signifiante. Par contre il est évident qu'il ne peut pas être reproduit à l'identique par des élèves, comme processus quasi-constructiviste. A priori, les élèves plus âgés ont reçu d'autres instruments conceptuels qu'ils tenteront de mettre en œuvre (adroitement ou non, là n'est pas la question). Par exemple il est improbable qu'ils groupent les tirages par 5 pour représenter le contenu de la bouteille et étudier la distribution de ces groupes de tirages.
Nous renvoyons l'étude expérimentale et théorique de ce problème à l'atelier de Joël Briand.

Conclusions

Cet article avait trois buts

- Présenter les idées et les résultats d'une recherche vieille de trente ans et estimer son actualité.
- montrer comment la recherche de situations fondamentales d'une connaissance mathématique peut conduire à des études de didactique théorique et d'ingénierie
- accessoirement essayer de présenter de façon organisée les situations utilisables en didactique de la statistique, nous en avons évoqué quelques unes.

Il m'était difficile de les atteindre avec une seule flèche. Aussi il m'a semblé indispensable de mettre à la disposition des auditeurs ou des lecteurs de ce cours un certain nombre de textes qui le précèdent, le précisent ou le prolongent
La première partie du cours est précisée dans l'article de référence
La deuxième partie a esquissé une méthode de recherche d'une situation de ce type. Il n'était pas possible de poursuivre cette étude dans le cadre de ce cours, mais on trouvera quelques compléments dans l'annexe 4.
Elle nous a conduit à un inventaire beaucoup trop rapide des obstacles épistémologiques et didactiques. J'aurais aimé m'attarder en particulier sur le septième, celui des obstacles macro didactique parce qu'il est particulièrement important en statistique (cf Annexe 5).
La question la plus importante est celle des différents niveaux scolaires où ces processus sont utilisables. Cette question sera étudiée et débattue dans l'atelier de Joël Briand.

[15] Dirigée par J. P. KAHANE

BIBLIOGAPHIE

L'expérience d'enseignement qui est évoquée dans ce texte a été répliquée deux années à l'Ecole Jules Michelet de Talence en 1972 et 1973. Elle faisait partie d'un ensemble d'autres études expérimentales rapportées dans l'« enseignement élémentaire des mathématiques » N°11 de l'IREM de Bordeaux, intitulé « vers un enseignement des probabilités à l'école élémentaire. Sa description a été publiée en 1974 et en 2002.

BROUSSEAU G. (1974). "L'enseignement des probabilités à l'école élémentaire" in *Compte-rendus de la XXVIe rencontre de la C.I.E.A.E.M.* IREM de Bordeaux.
Guy BROUSSEAU, Nadine BROUSSEAU, Virginia WARFIELD, (2002) "An experiment on the teaching of statistics and probability" *Journal of Mathematical Behavior*, 20 (2002) 363-441
BROUSSEAU G. et BROUSSEAU Nadine et WARFIELD Ginger,
"Une expérience d'enseignement des statistiques et des probabilités" non publié en français

Environnement BIBLIOGRAPHIQUE de l'article initial (1972)

Mathématiques

Marc BARBUT Mathématiques des sciences humaines Vol 1 et 2 P. U. F.

Statistique

FOURGEAUD G. et FUCHS A., *Statistique*, Dunod (1967)
GUERBER Louis, HENNEQUIN Paul Louis, *pour apprendre à conjecturer… initiation à la statistique* Bibliothèque d'enseignement mathématique APMEP (1967)
MURRAY R. SPIEGEL *Théorie et applications de la statistique*. Série Schaum 1972
PIATIER André *Statistique descriptive et initiation à l'analyse* (THEMIS) PUF 1966
VAN DER WARDEN B. L. *Statistique mathématique* trad. Guinchat et Degenne, Dunod (1967)

Logique, épistémologie :

[1] Benjamin MATALON Epistémologie des probabilités, in *Logique et connaissance scientifique*, Jean Piaget Editeur Encyclopédie de la pléiade (1967) NRF Paris*[16]

Psychologie

[2] J. COHEN Hasard adresse et chance (1963) PUF *
[3] M.F. EHRLICH Les théories de l'apprentissage in *le comportement verbal* (1970) DUNOD Paris *
[4] ROUANET Les modèles stochastiques d'apprentissage (1965) GAUTHIER-VILLARS Paris

Epistémologie Génétique

[5] J. PIAGET et B. INHELDER La genèse de l'idée de hasard chez l'enfant (1951) PUF Paris

[16] Les ouvrages signalés par * contiennent une bibliographie substantielle

Didactique

[6] D. GILIS et B. HERAUD Introduction des probabilités à l'élémentaire, in *la mathématique à l'école élémentaire* (1972) APMEP Paris
[7] G. BROUSSEAU Processus de mathématisation, in *la mathématique à l'école élémentaire* (1972) APMEP Paris
[8] Maurice GLAYMANN et Tamas VARGA, Les Probabilités à l'école (1973) CEDIC , Paris

et aussi par la suite…

CHEVALLARD Y. *Notes pour la didactique de la statistique,* IREM d'AIX-Marseille 1978

ENGEL Arthur, *L'enseignement des probabilités et de la statistique*. (2 vol.) CEDIC 1975

ROUANET Henry, LE ROUX Brigitte, BERT Marie-Claude. *Statistique en Sciences humaines ; procédures naturelles* Dunod 1987

ORUS-BAGUENA Pilar *Place de la logique spontanée des élèves dans le contrat didactique*

rôle possible d'un enseignement en analyse typologique dans la scolarité. Thèse de Didactique des Mathématiques. Université Bordeaux I 1992

BROUSSEAU G. (1993) "Stratégies de l'analyse statistique". (cours pour les professeurs de mathématiques, 80 pages) (LADIST).

ACADEMIE DES SCIENCES. *La Statistique.* Rapport sur la science et la technologie n°8 Juillet 2000 TEC & DOC

RAPPORT AU MINISTRE DE L'EDUCATION NATIONALE sous la direction de Jan Pierre KAHANE *« L'enseignement des sciences mathématiques* » Odile Jacob 2002

A propos de formation en statistique

Jean Claude REGNIER, A propos de la formation en statistique. Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique » *Questions éducatives Didactique des mathématiques* déc. (2002) N°22-23 Université de Saint Etienne

ANNEXES

Annexe 1. BROUSSEAU G. et BROUSSEAU Nadine et WARFIELD Virginia, "Une expérience d'enseignement des statistiques et des probabilités"
Annexe 2. Résumé des principales recherches de l'IREM de Bordeaux sur l'enseignement des statistiques et des probabilités (1971-75)
Annexe 3 Difficultés et Obstacles (Glossaire)
Annexe 4 Le séquençage des observations de classes
Annexe 5. Souvenirs et réflexions sur la S.F. en statistique
Annexe 6 Réflexions sur l'influence des sociétés savantes et de la noosphère sur l'enseignement des statistiques en France.

DIAPORAMAS

Diaporama 1. L'expérience d'enseignement
Diaporama 2. Genèse des situations didactiques
Diaporama 3 .Difficultés et obstacles.

TABLE DES MATIERES

INTRODUCTION 1
1. Une expérience de premier Enseignement des statistiques 1
situation initiale et un processus genétique 1
Le but pour le professeur 1
Résumé du déroulement de l'Enseignement 2
La situation initiale 2
Le processus 4
Modélisation et tables de hasard 5
Le risque 6
Réutilisation avec une autre machine 6
L'évaluation immédiate de l'apprentissage 6
Conclusions… provisoires 7
2. La détermination des situations de statistique 8
Détermination de la statistique comme branche du savoir 8
Praxéologie de la statistique : les activités des statisticiens 9
Stratégies 9
La classification des connaissances doit elle suivre celle des statisticiens ? 10
Justification de la recherche de situations fondamentales 11
Les Cartactéristiques des situations statistiques 11
L'interprétation des données 12
Le Résumé 12
Les stratégies des statisticiens 14
Position 14
L'incertitude 15
Première stratégie, la plus générale : représentants et ressemblances 15
Deuxième Stratégie : une mesure unifiée et spécifique : la rareté 15
La théorie des probabilités 16
3ième Stratégie. Séparer le hasard et la nécessité 16
Les formes de situations et de processus fondamentaux 17
1. Première forme de situation fondamentale : situation typique ou paradigmatique 17
2. Deuxième forme de situation fondamentale : situation significative 17
3. Troisième forme de situation fondamentale : situation initiale d'un processus génétique 18
Les propriétés didactiques des processus génétiques 19
3. Difficultés et obstacles à la pensée statistique 20
Obstacles ontogénétiques 20
Les difficultés liées à la dualité des langages 21
Obstacles épistémologiques 21
Obstacles didactiques 23
Retour sur les expériences de 1971-73 25
Les résultats 25
Comparaison avec d'autres méthodes 26
Confrontation avec les recommandations de l'Académie des sciences aujourd'hui 28
L'épineuse question des différents niveaux scolaires 29
Conclusions 29
Bibliogaphie 30
Annexes (Dans le CD-ROM) 31
Diaporamas 31