Université de Montpellier II

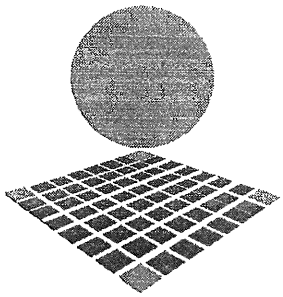
Publication de
l'Institut de Recherche sur
l'Enseignement des Mathématiques



Des statistiques à la pensée statistique

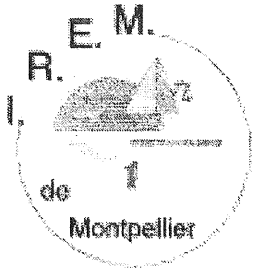
IREM de Montpellier.

2001



IREM
UNIVERSITÉ MONTPELLIER 2

cc 040 - Place Eugène Bataillon
34095 MONTPELLIER Cedex 05
Tél : 04.67.14.33.83 - 04.67.14.33.84
Fax : 04.67.14.39.09
e.mail : irem@math.univ-montp2.fr
http://www.irem.univ-montp2.fr



Des statistiques à la pensée statistique

IREM de Montpellier.

2001

PREFACE

La présence de statistiques dans les programmes de mathématiques du collège et du lycée n'est pas une nouveauté et certains seront sans doute surpris de la date à laquelle ont été écrites les lignes suivantes :

> *« Ce chapitre [Statistiques] présente un quadruple intérêt : d'abord la lecture pertinente de tableaux statistiques est maintenant nécessaire à la compréhension du fonctionnement de la société. Ensuite, c'est un excellent terrain pour des activités interdisciplinaires où les élèves peuvent faire preuve d'initiative et développer leurs méthodes de travail. En outre, apprendre à organiser, à représenter et à traiter des données fournies à l'état brut constitue une activité exemplaire de mathématisation. De même, savoir apprécier l'intérêt et les limites d'un processus de mathématisation d'une situation est une composante de toute formation scientifique. Enfin, c'est un secteur d'investissement des activités numériques, des représentations graphiques et des outils de calcul (calculatrices, ordinateurs) »*[1].

Les objectifs actuels de l'enseignement de cette partie des mathématiques dans le second degré ne contredisent en rien ces propos : le programme de seconde insiste, en effet, sur la nécessité de faire naître chez les élèves une « pensée statistique » qui vienne enrichir la culture scientifique des citoyens de demain. « *L'esprit statistique naît lorsque l'on prend conscience de l'existence de la fluctuation d'échantillonnage* »[2].

Aujourd'hui, les outils de calcul et de représentation graphique permettent de mettre en œuvre des activités qui auraient demandé, il y a quelques années, des temps de traitements beaucoup trop longs ; il est alors possible de faire vivre dans la classe les diverses phases d'une activité mathématique.

Pour mettre en œuvre ce programme, les professeurs doivent s'approprier cette « pensée statistique », réfléchir aux questions didactiques que pose la notion de hasard, s'interroger sur les questions soulevées par la modélisation, analyser le contexte mathématique des notions et outils rencontrés. Ceci demande un effort important de formation et de réflexion. Cette brochure constitue un outil précieux pour les enseignants.

---

[1] Commission Permanente de Réflexion sur l'Enseignement des Mathématiques (Direction des Lycées), *Objectifs et niveaux d'approfondissement des différents chapitres du programme de seconde*, 21 octobre 1983.

[2] Groupe Technique Disciplinaire de Mathématiques, *Document d'accompagnement du programme de seconde*, juin 2000.

L'article introductif et l'approche historique présentée permettent de mieux saisir ce que recouvre cette expression de « Pensée statistique ».

Le point de vue didactique présenté éclaire la lecture des programmes actuels et apporte des indications sur la perception qu'ont des élèves sur la notion de hasard.

La question de la modélisation est étudiée à travers l'exemple de la loi de Poisson. Les nécessaires « allers-retours » entre expérience, modèle et simulation sont mis en évidence au travers d'une activité.

Un exemple d'activité possible dans la classe est décrit et analysé au travers du jeu de « Franc carreau ». La place de l'expérience dans l'observation de l'aléatoire est rappelée et illustrée par de nombreux exemples.

La question de la génération des pseudo-aléatoires et les résultats d'arithmétique sous-jacents sont expliqués ; l'exemple du calcul d'une valeur approchée de π par une méthode de Monte-Carlo vient illustrer l'usage de ces outils.

Une bibliographie vient compléter ce travail varié.

Cet ouvrage apporte une aide aux enseignants de lycée, que ses auteurs en soient remerciés. Nous encourageons la diffusion de cette brochure et souhaitons qu'elle soit un point de départ d'un échange riche entre professeurs sur les moyens à mettre en œuvre pour que les élèves qui nous sont confiés puissent acquérir une culture mathématique adaptée aux exigences scientifiques et sociales d'aujourd'hui.

Les Inspecteurs Pédagogiques Régionaux

De l'Académie de Montpellier.

SOMMAIRE

INTRODUCTION 3

I - STATISTIQUES, STATISTIQUE

DES STATISTIQUES A LA PENSEE STATISTIQUE (J.-C. DUPERRET) 7
1. Qu'est-ce que la statistique ? 7
2. Les statistiques : un problème d'enseignement ? 7
3. Mon point de vue d'enseignant de collège 9
4. Des statistiques à la « pensée statistique » 10
5. Les premiers pas vers la statistique 11
6. Résumés et comparaison de séries statistiques 13
7. Du collège au lycée... ou des statistiques aux probabilités 16
8. Conclusion 23

Annexe 24

HISTOIRE DES STATISTIQUES ET DES PROBABILITES : UN SURVOL (M.-F. JOZEAU) 27
1. Introduction 27
2. Première tradition : les jeux 30
3. Deuxième tradition : l'arithmétique politique 32
4. Troisième tradition : la statistique descriptive 34
5. Apparition des premiers graphiques 35
6. La loi des grands nombres 36
7. Statistique au XIX^{ème} siècle 40
8. Statistique au XX^{ème} siècle 45
9. Bibliographie 48

Documents 51

II - NOUVEAUX PROGRAMMES, NOUVELLE APPROCHE

NOUVEAUX PROGRAMMES, NOUVELLE DEMARCHE (N. BASCOU) 77
1. Un enseignement important, pourquoi ? 77
2. Les principales étapes de la démarche statistique 78
3. Nouveau chapitre du programme de seconde : la statistique 78

LES ENJEUX DE L'ENSEIGNEMENT DE LA STATISTIQUE ET DES PROBABILITES AUJOURD'HUI (M. HENRY) 87
1. Leur place dans la production scientifique contemporaine 87
1. Statistique ou Probabilités ? 88
1. Enjeux et défis pour l'enseignement secondaire 90
1. Un état des lieux dans l'enseignement secondaire français 94
1. Les programmes de seconde de l'an 2000 96
1. Remarques sur les programmes de Première S et sur le projet de programme de Terminale 100
1. Vers l'enseignement de la modélisation : quelle formation des enseignants ? 103

STATISTIQUES ET PROBABILITES : UN POINT DE VUE DIDACTIQUE (M. ROCHE, M. SECO, C. VERGNE) 105
1. Un point de vue didactique 105
2. Le hasard, une question problématique et didactique 109
3. Questionnaire 112
4. Conclusion 117

Bibliographie 118

Annexe 119

III - ACTIVITES POUR LA CLASSE ET PROLONGEMENTS MATHEMATIQUES

SIMULATION ET MODELISATION : ETUDE D'UN EXEMPLE (J. FONTANA, M. NOGUES) **123**
1. De nouveaux programmes, de nouveaux problèmes 123
2. Expérience et simulations 124
3. « Ajuster » un modèle 129
4. Conclusion 131
Bibliographie 132

LE JEU DE FRANC-CARREAU. EXPERIMENTATION ET SIMULATION (M.-C. COMBES, M. LACAGE, J.-M. RAVIER, F. ROUX, J. SALLES) **133**

OBSERVER L'ALEATOIRE (A. BERNARD) **145**
1. Introduction 145
2. Les expériences aléatoires 147
3. Esquisse d'une conclusion 158

UNE APPROCHE DE LA MODELISATION AU TRAVERS DE LA LOI DE POISSON (A. GANNOUN, N. SABY) **161**
1. Introduction 161
2. La loi binomiale 162
3. Exemples 164
4. Bibliographie 164

LES NOMBRES PSEUDO-ALEATOIRES (M. JANVIER) **165**
1. Production des nombres pseudo-aléatoires 165
2. Petit théorème de Fermat 169
3. La fonction Φ d'Euler 170
4. Théorème de Lagrange 172
5. Théorème de Legendre 174
6. Racines primitives de p lorsque p est premier 175
7. Racine primitive de a, entier non premier 177
8. Calcul d'une valeur approchée de π 183
Bibliographie 187

IV - DES REFERENCES POUR ALLER PLUS LOIN

BIBLIOGRAPHIE D'APPUI POUR UN ENSEIGNEMENT DES STATISTIQUES (M.-C. COMBES, M. SAUTER, A. BERNARD) **191**
1. Articles parus dans la revue REPERES-IREM 191
2. Brochures des IREM 196
3. Revue Tangente : « Faut-il se fier aux Statistiques ? » 199

DES RESSOURCES SUR L'INTERNET POUR ENSEIGNER LA STATISTIQUE (M. JANVIER) **201**

L'IREM, CENTRE DE RESSOURCES (L. TROUCHE) **217**
1. Un centre de documentation 217
2. Un site Internet 217
3. Un centre d'édition et de diffusion de documents 218

PUBLICATIONS RECENTES **219**

Introduction

La statistique est devenue à la mode,
et c'est une position que de statistiquer

BALZAC, De la mode en littérature

La réflexion sur l'enseignement des statistiques est ancienne. Elle a d'abord été liée à une critique de l'utilisation sociale non contrôlée des statistiques[1], puis à une critique de la coupure entre l'enseignement des probabilités et celui des statistiques[2]. Mais cette réflexion était restée marginale, du fait de la marginalité même, dans le système scolaire français, de l'enseignement des statistiques.

Le renouvellement récent des programmes d'enseignement, lié à un développement important des probabilités et de la statistique, et à leur intervention accrue dans les sciences[3], se traduit par une forte augmentation de la place de ces chapitres dans le cours de mathématiques, aussi bien en collège qu'en lycée. Il se traduit aussi par un déplacement de l'objet d'étude vers la modélisation des phénomènes aléatoires. Ces changements sont donc profonds : les phénomènes qu'il s'agit de décrire, les méthodes d'étude et les concepts en jeu sont beaucoup plus complexes.

Ce renouvellement des programmes s'accompagne nécessairement d'un renouvellement de la réflexion didactique. Cette réflexion se mène depuis longtemps dans les IREM :

- la plupart d'entre eux ont constitué des équipes de recherche sur cet enseignement (Marie-France NOUGUES est responsable de l'équipe de Montpellier) ; une commission Inter-IREM très active a publié un ouvrage qui fait référence[4].
- Des « journées de formation de formateur », intitulées « *Statistique professionnelle, statistique didactique, quelle transposition ?* » ont été organisées par l'IREM de Montpellier en juin 2000. Elles donneront lieu à des Actes.
- Un GER (groupe d'Étude et de Recherches) a été constitué en 1998 par l'IUFM de Montpellier sur "l'enseignement des probabilités et des statistiques du lycée au post-bac". Il a rendu un premier rapport en octobre 2000 (une synthèse peut être consultée sur le site http://irem.univ-montp2.fr). Il devrait poursuivre ses travaux jusqu'en 2002 (Noël BASCOU en est le pilote).
- Dans le cadre du SFODEM (Suivi de FOrmation à Distance pour les Enseignants de Mathématiques), l'IREM a constitué une équipe sur « la simulation d'expériences aléatoires et le traitement statistique de données » (responsable : Michel JANVIER).

Les articles que l'on trouvera dans cette brochure seront des points d'appui pour approfondir cette réflexion. Ils viennent du réseau des IREM (que Jean-Claude DUPERRET, de l'IREM de Reims, Michel HENRY, de l'IREM de Besançon et Marie-Françoise JOZEAU, de l'IREM de Paris 7 soient ici remerciés !). Ils viennent aussi de l'ensemble des équipes de l'IREM de Montpellier : les équipes "statistique" bien sûr, mais aussi "analyse", "didactique", "informatique" et "intégration des logiciels" ont illustré, à leur façon, les nouveaux programmes d'enseignement, à partir de leur propre domaine de recherche.

---

[1] Cf. *Attention, statistiques ! Comment en déjouer les pièges*, de J. Klatzmann, 1992, Paris : La Découverte.

[2] Cf. *Le contremanuel de statistique et probabilité*, de M. Peltier, N. Rouche et M. Manderick, 1982, Bruxelles : Editions Vie Ouvrière.

[3] Cf. *La fin des certitudes,* de I. Prigogine, 1995, Paris : Editions Odile Jacob

[4] *Enseigner les probabilités au lycée, ouvertures statistiques, enjeux épistémologiques, questions didactiques et idées d'activités*, 1997, Irem de Reims. Voir aussi en IV l'ensemble des publications des IREM sur le sujet.

Cette brochure comporte quatre grandes parties :

- Deux points de vue généraux (épistémologique et didactique) sur l'enseignement des statistiques ;
- Une présentation des nouveaux programmes, suivie de deux points de vue didactiques ;
- Des idées d'activités pour la classe avec des prolongements mathématiques ;
- Des références bibliographiques et sur Internet pour aller plus loin.

En matière d'enseignement de la statistique et des probabilités, les besoins de formation continue et la nécessité de la recherche didactique sont patents. Cette brochure est une contribution à cette formation continue et à cette recherche. Une contribution qui en appelle d'autres...

Luc TROUCHE

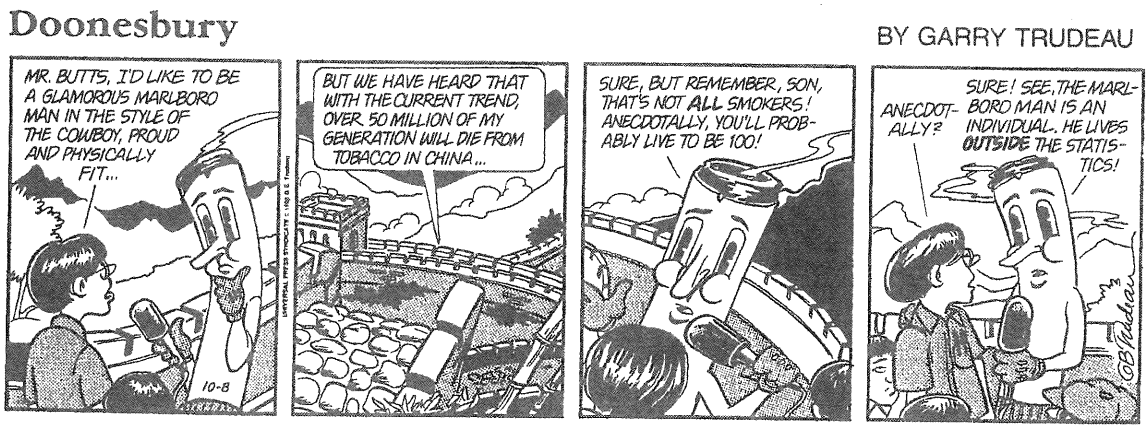
Directeur de l'IREM

Les illustrations de l'ouvrage sont extraites de *The Heart of Mathematics, An invitation to effective thinking*, de E. B. BURGER et M. STARBIRD, 2000, USA : Key College Publishing.

Cette brochure a été éditée grâce :

- aux heures attribuées pour la recherche aux animateurs de l'IREM par la DESCO (Direction des Enseignements Scolaires) et la DT (Direction des Technologies du Ministère de l'Education Nationale ;
- aux heures attribuées aux enseignants chercheurs du département de mathématiques par l'UFR Sciences de l'université Montpellier 2 ;
- aux services de secrétariat (Josiane DICK) et de reprographie (André VACHE) de l'IREM de Montpellier ;
- au soutien de la DAFP (Délégation Académique à la Formation des Personnels) et des Inspecteurs Pédagogiques Régionaux de mathématiques de l'académie de Montpellier.

I - STATISTIQUES, STATISTIQUE ?



DES STATISTIQUES A LA PENSEE STATISTIQUE

Jean-Claude DUPERRET

IREM de Reims
Membre de la Commission de Réflexion
sur l'Enseignement des Mathématiques

Résumé : *Comment définir les statistiques ? Pourquoi l'enseignement des statistiques constitue-t-il une impérieuse nécessité ? Pourquoi, pourtant, les statistiques restent-elles une "partie pauvre" de l'enseignement ? L'article propose des éléments de réponse, et des pistes pour le développement de "la pensée statistique" illustrées par des exemples d'application à l'enseignement.*

1. Qu'est-ce que la statistique ?

En 1935, W.F.WILLCOX dénombrait plus d'une centaine de définitions de la statistique. En voici quelques-unes connues :

Benjamin DISRAELI : *« Il existe trois sortes de mensonges : les mensonges, les affreux mensonges, et les statistiques ».*

Adolphe THIERS : *« ... L'art de préciser les choses que l'on ignore ».*

LAVELEYE : *« ... L'art de mentir mathématiquement ».*

MACAULEY : *« Les chiffres disent toujours ce que souhaite l'homme habile qui sait en jouer ».*

Louis ARMAND : *« Les statistiques, c'est comme le bikini, ça montre tout, mais ça cache l'essentiel ».*

Tout cela conduit à la question : **les statistiques : sciences du mensonge ?**

2. Les statistiques : un problème d'enseignement ?

Prenons pour commencer l'avis d'un spécialiste de cet enseignement, Edmond MALINVAUD, dans un extrait d'une communication faite le 27 novembre 1999 à la Commission de Réflexion sur l'Enseignement des Mathématiques :

> *1 - Pourquoi est-il si difficile d'amener des étudiants bien sélectionnés, et bien formés par ailleurs, à se familiariser avec le raisonnement sur l'aléatoire et le risque, avec le raisonnement inductif sur les données statistiques, et avec même une bonne appréciation du sens des résultats obtenus par l'économie mathématique ?*
>
> *2 - Puisqu'il y a dans les trois cas une difficulté d'accès à un champ important de la connaissance moderne, comment faciliter cet accès ? Et à quels enseignants cela revient-il ?*

Dans les trois cas, le raisonnement sur l'aléatoire, l'induction à partir de données statistiques, la recherche des résultats par l'économie mathématique, nos étudiants sont à l'aise dès lors que le professeur leur a entièrement spécifié un problème mathématique à résoudre.

Si on lui demande de préciser le champ de pertinence, c'est-à-dire le domaine d'application, du problème mathématique et de la solution qu'il y a apportée, il se sent désemparé et très hésitant. A fortiori en est-il ainsi si on ne lui a défini aucun problème mathématique et qu'on l'a laissé libre du choix du raisonnement face à une situation concrète de risque, ou face à une base de données statistiques susceptible d'éclairer un phénomène concret. Il en va de même de l'étudiant confronté à une assertion floue de théorie économique qu'il faudrait soit invalider soit confirmer.

Pourquoi cette incapacité à bien modéliser, ou même simplement à juger de la valeur d'un modèle proposé ? ....

... Les étudiants n'ont pas été assez familiarisés avec les phénomènes en cause, beaucoup moins familiarisés qu'avec les phénomènes du monde physique, considérés de plus comme certains et comme exactement observés. Il y a plus parce que, secondement, la bonne modélisation est moins évidente, donc plus délicate à trouver et moins persuasive, plus perfectible, quand on croira l'avoir trouvée.

La familiarisation exige du temps et un mûrissement progressif de la compréhension. C'est dire qu'il faut commencer assez tôt, dès l'enseignement secondaire même pour les futurs spécialistes. Chacun de ces spécialistes recevra dans sa discipline une formation beaucoup plus avancée que celle dispensée aux lycéens, mais il aura de la peine à l'assimiler alors s'il arrive à l'enseignement supérieur sans aucune préparation. Pour bien comprendre à l'université, il faut qu'il dispose déjà en quelque sorte d'une culture de base concernant le traitement des risques, l'utilisation des données statistiques, la conceptualisation des phénomènes économiques.

Quant à l'enseignement dans les lycées, nous devons prendre pour données les corps de professeurs tels qu'ils existent. J'ai des idées assez précises sur le corps des professeurs de sciences économiques et sociales. Je pourrais être long à leur sujet. Je dirai simplement que nous devons faire confiance à ces professeurs pour familiariser les lycéens qu'ils auront comme élèves avec une première étude des phénomènes économiques et sociaux. Mais nous ne pouvons pas compter sur eux pour dispenser une formation en probabilité et statistique. Cela ne peut revenir, et dans toutes les filières, qu'aux professeurs de mathématiques.

Je connais beaucoup moins ce corps de professeurs. Mais je m'en fais une idée, selon laquelle beaucoup de ces professeurs ne sont pas eux-mêmes familiarisés du tout avec le traitement des risques et avec l'induction statistique. Dès lors, la première urgence consiste à diffuser cette familiarisation de façon à ce qu'elle atteigne tous les professeurs de mathématiques des collèges et des lycées, au moins tous les jeunes professeurs.

3. Mon point de vue d'enseignant de collège

Le profond changement des programmes de mathématiques du collège de 1986 a particulièrement été marqué par l'introduction des statistiques que l'on trouve dans une rubrique nouvelle « Gestion de données »[1]. Après leur mise en place successivement en $6^{ème}$, $5^{ème}$, $4^{ème}$ et $3^{ème}$, elles font leur entrée dans le programme de seconde en 1990. Leur enseignement a très vite posé deux problèmes, intimement liés : celui de la formation des enseignants qui, pour la plupart découvraient ce domaine et celui de la part qu'il fallait leur consacrer dans le temps global alloué aux mathématiques.

Quel intérêt donner en effet à cet enseignement si on ne mesure pas son rôle de formation scientifique et social, et s'il se réduit à quelques recettes ? Il est donc indispensable de clairement définir les finalités de cet enseignement, les replacer comme partie intégrante des mathématiques, les situer dans la perspective d'une utilisation ultérieure et comme élément incontournable de la formation du citoyen.

Examinons de façon plus exhaustive un certain nombre de raisons qui militent en faveur de cet enseignement, et quelques obstacles qui empêchent son développement.

3.1. Pourquoi une telle nécessité ?

Formation du citoyen : tout individu est confronté à de nombreuses informations ; une formation statistique développera chez lui les capacités d'analyser et de synthétiser ces informations, et lui permettra d'avoir un regard critique sur celles-ci.

Invitation au travail interdisciplinaire : beaucoup d'autres disciplines utilisent les statistiques, avec deux formes d'expression privilégiées, les tableaux et les graphiques ; elles peuvent donc fournir un support pour un travail interdisciplinaire avec les mathématiques.

Liaison avec les différentes parties du programme : nombre d'activités statistiques permettent d'utiliser hors de leur contexte d'apprentissage des outils numériques, algébriques, géométriques, des modèles comme le modèle fonctionnel et celui de la proportionnalité : c'est donc l'occasion de montrer aux élèves que ces outils et modèles sont utiles et permettent l'éclairage mathématique nécessaire à une analyse et synthèse de l'information.

Préparation aux concepts probabilistes : les probabilités développées au lycée s'appuient fortement sur les statistiques, avec l'introduction de l'aléatoire et du hasard, et le passage de la fréquence statistique à la probabilité ; on peut dès le collège, sous forme d'activités, amener les élèves à une première réflexion sur des notions comme celles d'échantillon, d'aléatoire et de simulation.

Lieu privilégié pour l'utilisation de la calculatrice et de l'outil informatique : les calculatrices permettent de gérer des séries statistiques conséquentes, l'utilisation d'un tableur permet un rangement et une gestion dynamique des données, et l'éclairage immédiat sous forme de graphiques ; c'est l'endroit par excellence pour apprendre aux élèves à se servir de ces outils.

3.2. Pourquoi restent-elles encore souvent une « partie pauvre » de l'enseignement ?

Impression que ce ne sont pas de « vraies mathématiques » : si on se contente en effet de remplir des tableaux « tout prêts », de faire des graphiques sans réfléchir à la pertinence de

---

[1] Voir annexe 1 page 24.

Manque de formation des enseignants : c'est une des raisons principales, car donner du sens à un enseignement de statistiques suppose une solide formation, débordant le cadre des contenus du collège, formation qui n'a jamais été vraiment assurée.

Manque de temps : devant la difficulté à « boucler » les programmes, grande est la tentation de reléguer cette partie en fin de programme, s'il reste du temps ; ce choix est souvent argumenté par le fait que l'absence, ou la faible part accordée à ce domaine sera sans conséquence pour les élèves pour suivre en mathématiques dans les années ultérieures.

4. Des statistiques à la « pensée statistique »

Les statistiques développées au collège sont un premier niveau de la statistique : la statistique descriptive. Elles ont pour objectif essentiel « la transformation synthétique » d'informations. En ce sens, leur enseignement participe à la formation du citoyen : comprendre cette transformation, pouvoir analyser correctement, donc prudemment la synthèse effectuée. De manière plus précise, il faut faire comprendre aux élèves que le problème fondamental de la statistique descriptive est de résoudre le dilemme résultant de la transformation de données « brutes » en une synthèse qui parvienne à concilier le mieux possible deux pôles antagonistes : la « fidélité » et la « clarté ».

Un second niveau apparaît en fin de collège (sans que cela soit vraiment explicité dans les programmes) et plus clairement dans les anciens programmes de seconde : la comparaison de séries statistiques. Là prend pleinement sens la notion de « résumé ». Comment comparer en effet des séries de données brutes ? Calculer la moyenne d'une série, son écart type est souvent un exercice purement scolaire, sans véritable enjeu. Comparer deux ou plusieurs séries par leurs caractéristiques de position (mode, moyenne, médiane) et de dispersion (étendue, écart type, écart interquartile) peut conduire à une véritable activité d'analyse, d'argumentation, voire de prise de décision.

Le troisième niveau de la statistique est celui des statistiques inférentielles, où se rejoignent statistiques et probabilités. C'est le lieu de la modélisation, en vue d'étendre les informations recueillies à un domaine d'informations plus large. Se posent à ce niveau les deux problèmes fondamentaux (qui rapprochent la démarche statistique de la démarche physique) : la validité du modèle choisi, le contrôle des risques encourus. Les nouveaux programmes de seconde nous proposent une première approche de ce niveau, de façon expérimentale et non théorisée, avec l'introduction de la fluctuation d'échantillonnage.

Modéliser suppose un aller-retour entre une « réalité » et un « modèle mathématique ». Notre enseignement de mathématiques se réduit bien souvent à l'apprentissage de modèles, où les concepts perdent vite leur sens sous l'accumulation de techniques, recettes et algorithmes. La statistique est à mon sens le lieu privilégié de l'apprentissage de la modélisation, et les « réalités » que nous pouvons traiter avec ses outils ont un champ très large qui ouvre la porte à l'interdisciplinarité : physique, biologie, géographie, sciences humaines et sociales...Il nous faut cependant rester humble : modéliser des phénomènes relevant de ces différents domaines, traiter les informations qu'ils nous proposent, ne nous donnent pas forcément accès à leur compréhension.

Les trois étapes que j'ai décrites ci-dessus concourent à la formation de la « pensée statistique » : gérer un grand nombre d'informations et les synthétiser, comparer des ensembles d'informations, modéliser mathématiquement ces informations pour en tirer des conclusions « vraisemblables » et « probables » comme outils d'aide à la décision.

5. Les premiers pas vers la statistique[2]

Il s'agit de la statistique descriptive dont le premier enseignement commence en début de collège (on peut d'ailleurs se poser la question de le commencer dès la fin de l'école primaire). Aucune définition théorique de la notion de caractère (ou variable) statistique n'est donnée à ce niveau (une définition pourrait être : « caractéristique susceptible de variations observables »). Les caractères étudiés sont aussi bien qualitatifs que quantitatifs Pour les caractères qualitatifs, la distinction entre nominal et ordinal est rarement faite. Pour les caractères quantitatifs, on assiste assez rapidement à une dichotomie : un rangement « discret » lorsqu'il y a peu de modalités ; un rangement « pseudo-continu » en classes lorsqu'il y a beaucoup de modalités, ou que celles-ci sont des nombres décimaux.

5.1. Recueil et tableau de données

Les premiers exercices que rencontrent les élèves (et un nombre non négligeable des suivants !) proposent aux élèves des tableaux tout faits donnant les couples (modalité ; effectif) ou (classe ; effectif). L'objectif essentiel est en général de passer aux couples (modalité ; fréquence). Un deuxième courant d'exercices est de proposer aux élèves une série de données, et de passer des couples (individu ; modalité) aux couples (modalité ; effectif). Bien entendu, ce n'est pas en ces termes que ces exercices sont proposés. L'essentiel de l'activité est alors un travail de proportionnalité, utilisant particulièrement les pourcentages.

Limiter ce travail de recueil de données à ce seul type d'exercices est évidemment assez réducteur dans la perspective d'une véritable formation statistique. Proposer aux élèves d'aller eux-mêmes chercher les données dans d'autres discipline, ou dans la vie quotidienne donne une autre dimension à cet apprentissage. On peut même rêver d'activités plus ouvertes où la classe devra déterminer quel caractère étudié pour analyser des affirmations non forcément étayées : « il pleut plus à Lille qu'à Nice » ; « la terre se réchauffe ». Certains enseignants se lancent dans de telles activités, mais beaucoup hésitent pour les deux raisons que j'ai données en début : l'appréhension de ne pas pouvoir maîtriser l'activité par manque de formation ; le coût en temps.

5.2. Les graphiques

Nous vivons dans un monde d'images, et celles usuellement utilisées pour « montrer » un phénomène statistique sont les graphiques. C'est certainement ici que l'enseignement des statistiques a la plus grande vocation de formation du citoyen : choisir le graphique le mieux adapté à la situation, le construire en respectant des règles mathématiques ; comprendre, analyser, critiquer un graphique ; comparer différents graphiques...Le passage au graphique atténue le côté un peu réducteur de certains exercices signalés ci-dessus. L'interaction entre « tableaux » et « graphiques », l'aller-retour entre ces deux cadres, la transformation successive des informations sont autant d'éléments extrêmement formateurs.

Les principaux graphiques utilisés en collège se répartissent en deux types :

Les graphiques « de fonction » :

On représente les couples (modalité ; effectif) selon le couple de directions (horizontal ; vertical). Ce sont les diagrammes en bâtons et en barres, les histogrammes, les polygones d'effectifs. Suivant le type de caractère, l'axe horizontal présentera une structure d'ordre ou non.

---

[2] Dans le début de cette partie, je m'appuie, en le résumant beaucoup, sur l'article de Bernard PARZYSZ « *Heurs et malheurs du su et du perçu en statistique* » in Repères-IREM n°35.

Dans ce type de graphique, on veut faire apparaître que l'effectif est fonction de la modalité.

Les graphiques « de partition » :

Ce type de graphique s'adresse essentiellement à des caractères qualitatifs, ou quantitatifs considérés comme qualitatifs. Ce sont les diagrammes en bande, circulaires, semi-circulaires, elliptiques.

Dans ce type de graphiques, on veut faire apparaître que les diverses modalités déterminent une partition de la population.

Qui n'a jamais enseigné en collège ne peut mesurer les difficultés liées à cet apprentissage : choix des axes, des unités pour les graphiques de fonction, partages proportionnels pour les graphiques de partition, obtention d'un graphique juste, propre et clair... Ce travail doit conduire l'élève à comprendre les quatre qualités qu'on doit attendre d'un graphique :

- *La lisibilité* : un graphique doit être plus directement et rapidement lisible que les données chiffrées. Il ne crée pas d'informations supplémentaires, mais révèle l'essentiel masqué sous l'abondance des données.
- *La validité* : Elle est assurée par des règles mathématiques : proportionnalité aux longueurs, aux aires, aux volumes...
- *La fidélité* : Un graphique doit respecter les données et rendre fidèlement la réalité. L'impression visuelle suggérée ne doit pas conduire à déformer cette réalité comme, par exemple, la fausse perspective, ou un axe vertical qui ne démarre pas à zéro...
- *L'autosuffisance* : il doit pouvoir être compris indépendamment de la série de données qu'il représente.

5.3. Graphiques à critiquer et / ou valider

On peut dès le collège amener les élèves à « critiquer » de façon mathématique un graphique, en définissant les règles de validation de cette représentation. En voici deux exemples :

Du cercle à l'ellipse : une histoire d'affinité

Beaucoup de représentations dites « en camembert » sont souvent faites avec une ellipse. Comment contrôler la « justesse » d'un tel graphique ? Le premier réflexe est d'utiliser le rapporteur. Une telle démarche montre qu'on ne s'est pas posé la question de ce qui rend valide une telle représentation : c'est en fait, malheureusement de façon tacite, la proportionnalité des aires aux effectifs correspondants (si la représentation est de type spatiale, on s'y ramène car on a alors un cylindre en perspective). Mais sur une ellipse les aires des « parts » ne sont pas proportionnelles aux angles au centre correspondant. Pour s'en sortir, il faut construire le cercle dont l'ellipse est un affine orthogonal, utiliser alors le découpage du cercle avec le rapporteur (car là les aires sont proportionnelles aux angles) et projeter ces aires sur l'ellipse par affinité orthogonale. Pour mener une telle activité avec mes élèves, j'utilisais la méthode suivante : dessinez un « camembert circulaire » sur une page de papier ; faites alors « tourner » cette feuille de papier devant vos yeux. A partir de ce travail visuel, on construisait les règles de validation vues ci-dessus.

Partage d'un triangle équilatéral

Le théorème de Viviani dit que dans un triangle équilatéral, la somme des distances de tout point intérieur à ce triangle aux trois côtés est constante et égale à la hauteur du triangle équilatéral. Ce résultat est utilisé, particulièrement en géographie, mais évidemment sans être signalé, pour

construire des graphiques triangulaires pour des séries statistiques à trois modalités : on construit un point intérieur à un triangle équilatéral tel que les aires des trois triangles construits en joignant ce point à chacun des sommets soient proportionnelles aux effectifs des trois modalités. Ce type de représentation est souvent utilisé pour traduire une évolution : on construit ce point de façon périodique, et l'on joint de façon chronologique les point obtenus. Ce type de représentation a toujours été pour moi une mine d'activités en quatrième : tout d'abord « géométrique », avec la découverte du théorème de Viviani sous forme de problème ouvert ; ensuite « civique », avec la validation ou non de graphiques issus de manuels de géographie (il y en a un certain nombre de faux) ; enfin « créative » avec la construction de telles représentations.

6. Résumés et comparaison de séries statistiques

Dans tout ce paragraphe, il ne sera question que de séries à caractère quantitatif.

6.1. Caractéristiques de position et de dispersion

D'un point de vue mathématique, il est bien difficile de séparer ces deux notions, même si ce sera le cas dans leur enseignement. Il n'est pas certain (et c'est un euphémisme !) que beaucoup d'enseignants soient au clair là-dessus.

Une valeur centrale « a » d'une série statistique résume d'une certaine façon cette série. Elle la représentera d'autant mieux que la dispersion des valeurs autour d'elle sera petite. Tout dépend donc de la façon dont est mesuré cette dispersion, c'est-à-dire la façon dont sont calculés les écarts par rapport à cette valeur centrale.

Précisons le problème : j'ai une série statistique $(x_1, x_2, \ldots, x_n)$ que je peux considérer comme un vecteur X de $\mathbb{R}^n$. Je veux remplacer ce vecteur par un vecteur A = (a,a,…,a) qui minimise une distance donnée de $\mathbb{R}^n$ de X à la n-sectrice ensemble des vecteurs U = (u,u, ….,u). Cette définition n'est évidemment pas celle qui sera donnée aux élèves, et l'on trouvera dans le livre de Claudine ROBERT « *L'empereur et la girafe* » une façon intermédiaire, particulièrement intéressante, d'introduire ces notions en leur laissant leur sens profond.

6.2. Moyenne et écart type

Choisissons comme distance $d(X,Y) = \sqrt{\frac{1}{n}\sum (x_i-y_i)^2}$. Cette distance est une distance euclidienne, car elle dérive d'un produit scalaire, et la solution cherchée est unique : c'est la projection orthogonale de X sur la n-sectrice. Un calcul immédiat donne alors pour solution A le vecteur dont toutes les coordonnées sont égales à la moyenne de la série statistique. La distance de X à ce « vecteur moyenne » est « l'écart type ».

Ces bases théoriques étant posées, comment introduire la moyenne en collège (sachant que les élèves connaissent parfaitement son « calcul », qu'ils pratiquent depuis longtemps avec leurs notes). Une des meilleures façons est de leur demander de trouver un nombre « a » tel que si on remplace chacun des « x_i » par « a » on obtienne la même somme. Au-delà de confirmer ainsi la formule classique, les élèves retrouvent un « geste » qu'ils pratiquent souvent : quelle note dois-je avoir au prochain contrôle pour obtenir la « moyenne générale » (10, bien entendu !). Cela met aussi en évidence les propriétés « affines » de la moyenne.

Au niveau du lycée, on peut revenir sur cette définition en l'introduisant comme minimum de la dispersion des carrés des écarts autour d'un nombre x, c'est-à-dire minimum de la fonction :

$$f(x) = \sum (x_i - x)^2$$

Ce qui me paraît important, au moins au niveau des enseignants de mathématiques, c'est de comprendre en quoi ces trois approches représentent bien le même concept. Pour la première et la troisième, c'est assez immédiat. « Rapprochons donc la première et la deuxième. Cela est relativement facile à illustrer en se mettant dans $\mathbb{R}^2$: le plan est rapporté à un repère orthonormal ; on trace la première bissectrice D d'équation $y = x$.

Soit alors la série statistique (x_1, x_2) qu'on représente par le point M de coordonnées (x_1, x_2) ; on construit la droite Δ d'équation $x + y = x_1 + x_2$. Δ est orthogonale à D.

Soit A (a , a) le point d'intersection de D et Δ : il vérifie $a + a = x_1 + x_2$, mais il est aussi la projection orthogonale de M sur D. Voilà visualisées ces deux approches de la moyenne a de la série (x_1, x_2).

Allons maintenant à la quête du sens de l'écart type.

$MA^2 = (x_1 - a)^2 + (x_2 - a)^2 + \ldots$

Si S = { 1,2,3 }, a = 2 et $MA^2 = 2$

Si S = { 1,1,2,2,3,3 }, a = 2 et $MA^2 = 4$

D'où le correctif pour pouvoir comparer des séries qui n'ont pas le même nombre d'éléments : on prend pour distance d(M,A) définie par $d(M,A)^2 = \frac{MA^2}{n}$, où n est le nombre d'éléments de la série. Dans les deux cas ci-dessus, on a alors $d(M,A)^2 = \frac{2}{3}$.

D'où les définitions :

moyenne : $x = a$

écart type : $\sigma = \sqrt{\frac{MA}{n}}$

6.3. Médiane et écart absolu moyen

Choisissons comme distance $d_1(X,Y) = \frac{1}{n} \Sigma |x_i - y_i|$. Cette distance n'est pas une distance euclidienne. Rangeons les x_i dans l'ordre croissant. On obtient une nouvelle série ordonnée $(y_1, y_2, \ldots, y_n)$. Deux cas apparaissent alors :

Si n est impair (n = 2p+1), considérons le terme du « milieu, c'est-à-dire y_{p+1}. Notre problème a alors pour solution unique le vecteur dont toutes les coordonnées sont égales à y_{p+1}. Cette valeur y_{p+1} est appelée médiane de la série X.

Si n est pair (n = 2p), il n'y a pas unicité de la solution. De manière plus précise, tout vecteur de la n-sectrice ayant pour coordonnées un élément de l'intervalle « milieu » $[y_p, y_{p+1}]$ est solution de notre problème. On convient alors de prendre pour médiane le milieu de cet intervalle.

Pour introduire cette notion en collège, on peut, après avoir rangé la série dans l'ordre croissant, poser la question : existe-t-il un nombre tel qu'il y ait au maximum 50% des valeurs de la série qui lui soient supérieures, et au maximum 50% qui lui soient inférieures. Les deux cas, pairs et impairs, apparaissent alors immédiatement, et la notion de milieu ou d'intervalle milieu permet de donner une solution, et d'établir la notion de médiane.

On peut, comme pour la moyenne, visualiser ce concept de médiane. Reprenons ce qui a alors été fait pour une série (x_1, x_2) ; avec les mêmes notations, considérons les points de Δ, $A_1(x_1, x_1)$ et $A_2(x_2, x_2)$. Tout point A du segment $[A_1, A_2]$ minimise la distance « d_1 ».

De façon plus précise : $d_1(M, A) = d_1(M, A_1) = d_1(M, A_2) = |x_2 - x_1|$.

La visualisation pour une série de trois valeurs (x_1, x_2, x_3) (rangées dans l'ordre croisant) oblige à construire un parallélépipède rectangle dans lequel on peut montrer que le trajet « d_1 » le plus court est celui de M à $A_2(x_2, x_2)$.

6.4. Rangement des données par classes

Le travail repose alors sur l'hypothèse d'une répartition uniforme des données. La moyenne est alors la moyenne des centres des classes pondérés par l'effectif de la classe. La médiane est obtenue par « interpolation affine ». Cette hypothèse de répartition uniforme est assez naturellement acceptée par les élèves et, à défaut de toute autre information, elle en vaut bien une autre. Cela permet en outre de réinvestir des outils vus dans d'autres domaines des mathématiques comme, par exemple, « THALES » pour le « calcul » de la médiane.

6.5. Mode, moyenne et médiane

Il faut faire comprendre aux élèves le caractère « barycentrique » de la moyenne et le caractère « géométrique » de la médiane :

- En enlevant des valeurs « extrêmes » montrer la sensibilité de la moyenne à cette perturbation, et la stabilité de la médiane (c'est pourquoi on préfère en général la « moyenne tronquée » obtenue en enlevant aux deux extrémités de la série ordonnée un nombre d'observations correspondant au même pourcentage).
- En modifiant des valeurs « centrales », arriver au constat contraire.

Ce travail doit être complété par la notion de mode et de classe modale. A partir d'un résumé « moyenne-médiane-mode », il est intéressant d'analyser avec les élèves « l'idée » qu'on peut se faire de la série d'origine. Par exemple si, dans une entreprise, le mode des salaires est 6000 F, la médiane 8000 F, la moyenne 10000 F, on est assuré qu'il doit y avoir de « gros salaires ».

Il faut aussi donner des exemples de séries bimodales ou plurimodales pour montrer l'insuffisance des indicateurs centraux.

6.6. Travail sur la dispersion

L'étendue est un indicateur de dispersion facile à calculer, mais grossier et particulièrement inadapté lorsqu'il y a des valeurs extrêmes aberrantes. Il a cependant une certaine valeur pratique : imaginons que je veuille construire une maison au bord d'une rivière : l'information que la moyenne des crues sur ces dix dernières années est de 0,50 m ne me sera que de peu d'utilité par rapport à la connaissance de l'étendue ; si celle-ci est en effet de 6 m, je reverrais certainement ma décision. L'écart type est difficile conceptuellement, même si la plupart des calculatrices le donnent. Il n'a véritablement d'intérêt que dans le cas d'une répartition « gaussienne », car il permet de situer l'intervalle théorique où se trouve un pourcentage donné de la population centrale autour de la moyenne. Du point de vue de l'enseignement, l'écart interquartile est un indicateur plus facile d'accès puisqu'il permet de situer les 50% de la population « centrale », en utilisant un calcul analogue à celui de la médiane.

Un résumé particulièrement intéressant d'une série statistique est « la boite à moustache » (box-plot). Introduites pour la première fois en 1977 par John TUKEY dans son ouvrage « Exploratory Data Analysis », ces boîtes de dispersion sont un moyen de représenter une série statistique par un graphique qui illustre la forme de sa distribution, en particulier sa symétrie, sa valeur centrale ainsi que sa dispersion.

6.7. Comparaison de séries statistiques

Comparer des phénomènes est un des enjeux fondamentaux des statistiques. Cela peut se faire de façon purement qualitative en comparant par exemple des graphiques. Cela peut se faire en comparant des résumés. Cela peut enfin se faire de façon quantitative, par exemple en étudiant la « corrélation » de deux séries. Un des enjeux de l'enseignement est alors d'apprendre à ne pas confondre « corrélation » et « causalité ». Si par comparaison de graphiques, ou par étude de corrélation, j'arrive à la conclusion que l'évolution du nombre de morts par cancer dus au tabac sur ces dix dernières années aux Etats Unis se modélise de la même façon que l'évolution du nombre de cigognes en Alsace, y voir une relation de causalité est un choix plus que hasardeux !

7. Du collège au lycée... ou des statistiques aux probabilités

7.1. Evolution de l'enseignement de la statistique

Avant d'aborder cette partie de mon exposé, il m'apparaît nécessaire de faire un rapide historique de l'évolution des programmes.

Avant 1990 :

- On enseigne le modèle (probabilité) en 1ère et terminale.
- On enseigne de manière indépendante de la statistique descriptive dans certaines classes de lycée.
- C'est dans les classes post-bac que se fait le lien avec la statistique inférentielle.

A partir de 1991 :

La statistique descriptive a été enseignée en collège et seconde. On enseigne conjointement modèle et modélisation en première avec l'approche fréquentiste des probabilités, qui repose sur la stabilisation de la fréquence. Cela présente un inconvénient mathématique :

- en statistique, les suites sont finies, et rien ne prouve l'existence et l'unicité de la « limite stabilisée » ;
- en analyse, les suites sont infinies, la limite est unique et ne dépend pas des premiers termes de la suite.

A partir de 2000 :

On introduit en seconde une sensibilisation à la modélisation, mais sans mettre en place le modèle. Pour cela on s'appuie sur la simulation (qui devient *ipso facto* le modèle).

Examinons maintenant les conséquences de ces différents choix en illustrant ce passage si délicat des statistiques aux probabilités.

7.2. Les paquets de café

Avez-vous déjà acheté un paquet de café de 500 g ? Evidemment oui, me répondrez-vous ! Je suis au regret de vous dire que cette probabilité est très faible, voire nulle (suivant l'angle sous lequel on se place).

A un objet réel, la réalisation effective d'une mesure ne fait correspondre en fait un nombre déterminé que si les conventions sociales le déclarent. Dans des situations où cette convention n'existe pas, l'image d'un objet par une mesure est en fait un intervalle (erreur, tolérance, intervalle de confiance), ou plus précisément une distribution (de fréquences, de probabilité).

Revenons dans le domaine des statistiques. Si on va dans des grandes surfaces de distribution, et que l'on pèse tous les paquets de café au gramme près (c'est une convention !) d'une même marque, on va constater des fluctuations : dans la première pour 100 paquets, on aura un minimum à 478 g et un maximum à 525 g, avec une moyenne de 501 g ; dans une autre ces chiffres deviendront respectivement 475, 522, et 499 ; si on passe à un « échantillon » de 1000 paquets, ils deviendront 468, 545, et 500. On constate qu'en augmentant la taille de l'échantillon, la moyenne se stabilise à 500 g (ce qui prouve l'honnêteté des fabriquants de paquets de café !), mais que l'étendue augmente.

Ce phénomène est celui introduit dans le nouveau programme de seconde sous l'appellation « fluctuation d'échantillonnage ». Le monde dans lequel nous vivons est un « monde probable ». Mais si nous comparons tous les histogrammes obtenus à partir de ces échantillons, nous constaterons une certaine permanence de forme, qui sera d'autant plus stable que l'échantillon sera plus grand : pour les spécialistes, ils reconnaîtront immédiatement le modèle « gaussien ». Derrière cette fluctuation « statistique », il semble donc qu'il y ait un invariant mathématique : c'est la notion de modèle.

Il n'en reste pas moins que celui qui partira avec un paquet de café de 475 g se sentira « floué », alors que celui qui en aura eu un de 520 g aura l'impression d'être un « veinard ». C'est la notion de « chance », que nous allons traduire de façon un peu plus mathématique avec celle de « hasard ».

7.3. Une première idée du hasard

Les statistiques sont certainement la discipline scientifique expérimentale la plus ancienne que pratique l'homme. Tout le commerce repose sur elles : quel stock dois-je commander en fonction des résultats des années passées ? Une telle décision essaie d'éliminer le « hasard » Et pourtant ce hasard est omniprésent dans nos croyances, et souvent dans nos décisions. Examinons deux extrêmes :

- Une première conception du hasard répond à un sentiment de « justice », et pour cela il doit être « proportionnel ». C'est le cas de tous les jeux de « hasard » comme le loto. Nul n'admettrait qu'il ne soit pas régi par une loi d'équiprobabilité. Mais cela se traduit par une grande confusion entre statistiques et probabilités comme en témoignent les journaux spécialisés : le 7 est en bonne forme ; le 14 devrait rattraper son retard ; le 18 est en période noire ; jouez l'outsider, le 49 ! Cela va plus loin : qui oserait jouer 1,2,3,4,5,6 ? Cette grille ne paraît pas « normale » au niveau du « hasard ». Il faut cependant noter que si toutes les grilles ont la même « chance », le gain est lui lié au choix de la grille. On touche ici à la notion de variable aléatoire : le vrai problème du loto n'est pas la chance de gagner, mais de gagner beaucoup. Les renseignements statistiques que se refuse évidemment de donner la « Française des jeux » pourraient aider à cette finalité.

- Une seconde conception du hasard répond à un sentiment de « fatalité » : c'est la reproduction d'évènements dont la probabilité est très faible, généralement appelée « loi des séries ». On retrouve cette conception dans la croyance populaire sous la forme de dictons : « jamais deux sans trois » ; « un malheur n'arrive jamais seul »...

Ces deux extrêmes rejoignent deux visions du monde antagoniste :

- L'une « déterministe » à l'excès, tel D'ALEMBERT qui pensait que si, dans un jeu de pile ou face, « pile » était sorti trois fois, alors la probabilité de tirer « face » était supérieure à $\frac{1}{2}$.
- L'autre « aléatoire » à l'excès : « Si je traverse une route, ou je me fais écraser par une voiture, ou non » qui attribue alors à l'événement « se faire écraser par une voiture en traversant une route » une probabilité de $\frac{1}{2}$.

7.4. Les premiers pas dans l'aléatoire

Une première approche de « l'aléatoire » dès les classes de collège et en seconde me paraît une bonne chose, à condition d'éviter toute théorisation. C'est le lieu d'activités motivantes et attrayantes pour les élèves, et leur première rencontre avec ce type de phénomènes et la réflexion que cela engendre. Cela donne d'autre part une autre dimension aux statistiques qu'ils pratiquent usuellement.

Il faut cependant séparer à ce niveau deux types d'activités que la théorie confond[3] : le tirage d'un échantillon d'une population ; la répétition d'un événement aléatoire. Si la première amène la question : « Ai je choisi un bon échantillon ? » ou de façon plus mathématique « La fréquence du caractère étudié observée dans l'échantillon est elle proche de celle de la population totale ? », la seconde pose la question « Ai je bien attendu la stabilisation de la fréquence ? » ou de façon plus mathématique « La fréquence f_p choisie par arrêt au bout de p répétitions est elle proche de la fréquence lim f_n, si elle existe, que je choisis comme probabilité ? ». Si les deux relèvent de la même problématique, à savoir la difficulté ou l'impossibilité d'avoir accès à la population totale, soit qu'elle soit trop grande, soit a fortiori qu'elle soit infinie, elles se différencient au niveau du recueil de l'information, la première ne dépendant pas de l'ordre, la seconde résultant d'une chronologie.

7.5. Echantillon

Les médias nous « baignent » quotidiennement dans des sondages, des simulations, des enquêtes statistiques...Tout cela semble passer par le mot magique « échantillon ». Pourquoi travailler sur des échantillons ? Comment les constituer ? Quelle crédibilité accorder aux résultats qu'on en tire ? Voilà, à mon sens, les questions qu'il faut aborder avec les élèves, dès le collège, pour mener une première réflexion sur cette notion difficile mais omniprésente dans notre quotidien. Voici deux exemples qui peuvent permettre d'amener les élèves à comprendre la nécessité de travailler sur des échantillons :

- Je suis un fabricant de « pétards ». Voyant arriver le 14 juillet, je m'inquiète de savoir si mes pétards sont en bon état. Pour être sûr de l'état de mon stock, je peux évidemment tous les essayer. Je pourrais alors affirmer : j'en avais 97% de bons. Mais je n'aurai plus de stock !
- Je suis un homme politique, et j'aimerais « prévoir » les résultats des prochaines élections nationales. Je peux évidemment interroger tous les français votants. Mais comment les contacter tous ? Où ? Combien de temps me faudra-t-il ?

---

[3] Voir mon article « *L'apprenti fréquentiste* » in Repères-IREM n° 21.

La réponse est de travailler sur une partie de la « population ». Dans le premier cas, on essaiera par exemple 1 pétard sur 150. Dans le second, on se contentera d'interroger 1000 personnes. Une telle partie est un échantillon. Une première question fondamentale est : « Qu'est ce qu'un bon échantillon ? ». C'est une partie de la population qui, pour la question que j'étudie, me donne sensiblement les mêmes réponses que la population toute entière. Dans le premier exemple, il faudra que l'échantillon me donne approximativement 97%. Dans le second, il devra me donner la photographie « miniaturisée » du résultat des élections.

Une question corollaire encore plus fondamentale est « Comment faire un bon échantillon ? La réponse immédiate est : « par un tirage au hasard ». Mais comment faire un tirage au hasard ? Par exemple, pour le cas des élections, je peux décider de me mettre sur une place dans le centre d'une ville et d'interroger les 1000 premiers passants. On sent bien tout de suite que les résultats ne seront pas les mêmes si je suis dans une ville de province, dans une grande ville, ou dans une ville de banlieue.

En fait, pour éviter les « catastrophes » que pourrait amener le hasard sur l'enquête menée, on décide en général de constituer un échantillon à partir de critères objectivement définis. Une méthode particulièrement employée est la « méthode des quotas », parfaitement accessibles aux élèves dès les premières classes du collège[4].

7.6. Réalité et mathématique : statistiques et probabilités

Pour aller plus loin, et illustrer ce que nous proposent les nouveaux programmes de seconde, je vais m'appuyer sur un tableau :

Réalité	*Mathématique*
Statistique descriptive	Probabilité
Données observées	Données calculées
Résultats empiriques	Résultats théoriques
Distribution de fréquences	Loi de probabilité
Moyenne empirique	Espérance théorique

- Modéliser une expérience aléatoire, c'est lui associer une loi de probabilité.
- Les distributions de fréquence varient, le modèle est un invariant.
- La réalité est liée à la notion de variabilité, la modélisation a pour objectif de rendre le monde intelligible et prévisible.

Le schéma ci-après traduit comment va fonctionner ce passage des statistiques aux probabilités :
- On cherche dans un premier temps à modéliser les « données expérimentales ».
- Un modèle mathématique ayant été trouvé, on va recréer des données calculées par simulation (en général pour des raisons d'économie).

[4] Voir par exemple l'article de l'IREM de Montpellier (Marie-Claire COMBES et al.) in *« Des chiffres et des lettres »* (brochure de la commission 1er cycle).

Données expérimentales	Modélisation Simulation	Théorie des probabilités

Une des confusions du nouveau programme de seconde pourrait être de faire croire aux élèves que la simulation est un instrument de modélisation, alors que, comme je viens de le dire plus haut : sans modélisation, pas de simulation possible.

7. 7. Simulation

En statistique, choisir au hasard sous-entend une loi équirépartie (tout le problème est sur quel ensemble !) Le principe de la simulation est donc de produire des séries équiréparties (en général de chiffres ou de nombres). Mais cela pose d'entrée un problème philosophique. Pour l'illustrer, suivons BOREL avec son singe typographe (repris par Evar EKELAND dans son livre « Au hasard ») : ce singe, immortel, tape « au hasard » des 0 et des 1. Dire qu'il les tape au hasard veut dire qu'il y a une hypothèse d'équirépartition, mais qu'aucune règle ne régit la sortie des 0 et des 1 (en particulier, il ne peut s'agir de la sortie alternative de 0 et de 1). Ce singe doit donc pouvoir taper à un moment « Proust » converti en binaire. Il doit pouvoir taper à un moment donné son œuvre complète...Mais alors, on est en train de se redonner une règle !

Maintenant que nous sommes persuadés que nous ne pouvons pas espérer une simulation parfaite, examinons quelques procédés plus ou moins bons :

- Tables de hasard.
- Annuaire téléphonique.
- Tirage au sort (de boules, de dés, d'étiquettes,...)
- Ouvrir un gros livre et noter la page.
- Numéros de plaque minéralogique.
- Touche « Random » de la calculatrice.
- Utilisation de l'ordinateur.

L'homme serait-il meilleur que la « machine » ? Demandons lui d'écrire au hasard 200 chiffres 0 ou 1. Utilisons un générateur aléatoire de « machine ». Peut-on reconnaître ces deux listes ? La réponse est oui, avec une « bonne probabilité ». Il y a un obstacle psychologique chez l'humain à écrire plus de 5 fois le même terme de façon consécutive.

Avec une machine, la probabilité d'avoir au moins 6 données consécutives égales est :

- pour 100 chiffres : 0,8
- pour 200 chiffres : 0,96

Avec une machine, la probabilité d'avoir 5 données consécutives égales est :

- pour 50 chiffres : 0,82
- pour 100 chiffres : 0,97
- pour 200 chiffres : 0,999

Il ne faudrait pas croire qu'on ne peut simuler que des lois équiréparties. Mais, pour des lois non équiréparties, il faut procéder à un « recodage » de la simulation.

7.8. Les spaghettis

Les spaghettis sont à l'origine d'une activité que j'ai proposée à mes élèves de quatrième en 1988, dans le feu de l'expérimentation des nouveaux programmes : elle consistait à couper « au hasard » des spaghettis en trois morceaux pour voir si on pouvait faire un triangle ayant comme côtés chacun des trois morceaux ; l'objectif était d'établir « l'inégalité triangulaire ».

Je passe sur les péripéties que vous pourrez lire plus en détail dans mon article *« L'apprenti fréquentiste »* déjà cité, et j'arrive à ce que je considérais comme une « modélisation » de ce problème et que j'ai proposé d'expérimenter à ces mêmes élèves en troisième : tirer trois nombres au hasard avec la touche « Random » de la calculatrice, et regarder si ils pouvaient être les trois côtés d'un triangle. Le passage aux statistiques s'est fait en leur demandant de faire chacun 20 « tirages ». La fréquence obtenue en prenant tous les résultats était voisine de 0,25. Il m'a fallu quelque temps pour trouver un « modèle probabiliste » donnant du sens à ce résultat. Me plaçant dans l'espace rapporté à un repère orthonormal, je considérais le cube formé à partir des trois vecteurs de base : ce cube, en tant que solide, me donnait tous les points de coordonnées (x,y,z) résultats possibles de tirages décrits ci-dessus. Pour que (x,y,z) puisse être un triplet donnant un triangle, il me fallait « enlever » à ce cube trois tétraèdres correspondant à l'intersection de ce cube avec les trois demi-espaces d'inéquations respectives : $x > y + z$; $y > x + z$; $z > x + y$. On vérifie alors aisément qu'il reste « exactement » la « moitié » du cube.

Examinons tous les non dits de cette activité :

- Je suppose que la probabilité d'obtenir un triangle dans les conditions ci-dessus existe.
- Je modélise en supposant que tout point d'un spaghetti a la même probabilité d'être un point de fracture.
- Je simule, à partir de cette hypothèse d'équirépartition, en imaginant que mon spaghetti est un segment de longueur 1.
- Je cherche un modèle géométrique me donnant la probabilité cherchée par une autre voie que la seule approche fréquentiste (stabilisation de résultats statistiques) : celui-ci est un cube où chaque point a la même probabilité d'être tiré. J'utilise alors une hypothèse de probabilité uniforme pour déclarer que la probabilité cherchée est un rapport de volumes.

Critiquons maintenant cette « modélisation » :

- Je passe du continu au discret sur mon spaghetti (Random ne me donne que trois chiffres après la virgule), et du discret au continu lorsque je passe à la modélisation par rapport de volumes.
- Le problème que j'ai modélisé n'est pas celui de la « fracture » d'un spaghetti en trois, mais la fracture de trois spaghettis en deux : les collègues de Montpellier[9] ont attiré mon attention sur ce point, et m'ont proposé deux façons de couper « au hasard » un cube : soit il se coupe d'un coup en trois, la modélisation géométrique conduisant alors à une probabilité de 0,25 ; soit il se coupe en deux temps, et l'on arrive alors à une probabilité de $\ln 2 - 1/2$, voisine de 0,19. Ces résultats peuvent être approchés de façon fréquentiste par simulation.

Critiquons maintenant à un autre niveau :

- Si on fait l'expérience « physique », on arrive à plus de 70% en moyenne de triangles : l'hypothèse de l'équirépartition ne rend nullement compte de la réalité physique (c'est assez facile à comprendre).
- Notre travail de modélisation porte donc sur un « spaghetti abstrait », mais il nous faut cependant bien le « voir » pour comprendre.

---

[9] Voir dans cette même brochure l'article de Joëlle FONTANA et Maryse NOGUES.

- Cette abstraction étant faite, la modélisation conduit à deux résultats différents : c'est l'imprécision du « au hasard » qui amène à ce paradoxe, à rapprocher du paradoxe de BERTRAND. Définir le « au hasard », c'est définir les conditions d'expérimentation.

7.9. Loi de BENFORD

Je vais terminer cet exposé par un problème que m'a exposé Claudine ROBERT et que je trouve particulièrement signifiant.

Le tableau ci-dessous donne la fréquence d'apparition du premier chiffre de nombres pris respectivement :

Colonne 2 : 1000 nombres du Monde daté du vendredi 23 avril 1999 ;
Colonne 3 : 914 nombres d'un historique de compte de la société Gilibert ;
Colonne 4 : nombres d'habitants de 1229 communes obtenus lors du recensement de 1992.

Premier chiffre	Le monde	Gilibert	Commune	Loi de Benford
1	0,322	0,317	0,321	0,301
2	0,151	0,161	0,168	0,176
3	0,108	0,142	0,133	0,125
4	0,099	0,088	0,081	0,097
5	0,073	0,070	0,087	0,080
6	0,081	0,061	0,067	0,067
7	0,055	0,070	0,055	0,058
8	0,065	0,040	0,045	0,051
9	0,046	0,050	0,044	0,046

Deux constats s'imposent :
- On est bien loin de l'équiprobabilité (qui est pourtant notre premier réflexe).
- Les trois expériences donnent des résultats vraiment proches.

Claudine ROBERT a alors multiplié les expériences, constaté que les résultats étaient invariants par changement d'échelle ; elle a alors, de façon mathématique, trouvé une loi modélisant ce phénomène, la loi de BENFORD :

la probabilité que le premier chiffre à gauche dans l'écriture en base 10 soit i = 1,..., 9 est :

$$\log (1 + 1/i) \text{ (logarithme décimal)}$$

La dernière colonne du tableau ci-dessus donne les fréquences théoriques obtenues par calcul avec cette loi. La modélisation par cette loi apparaît comme très bonne d'un point de vue qualitatif.

Mais, avoir modélisé mathématiquement nous donne-t-il le sens profond du phénomène ? Evidemment, non !

Pour donner du sens à ce phénomène, j'ai essayé d'imaginer une simulation (qui ne peut reposer sur le tirage « au hasard » de nombres) : un chronomètre se déclenche, et j'écris régulièrement les entiers naturels dans l'ordre croissant ; le chronomètre s'arrête « au hasard », et je fais mes comptes ! Il y a donc bien du hasard là-dedans, mais pas là où on le croit.

Tout cela me direz-vous n'est que jeu de mathématicien : ceux qui se sont fait « épingler » par le fisc qui utilisait cette loi pour vérifier leur comptabilité n'en sont pas complètement convaincus !

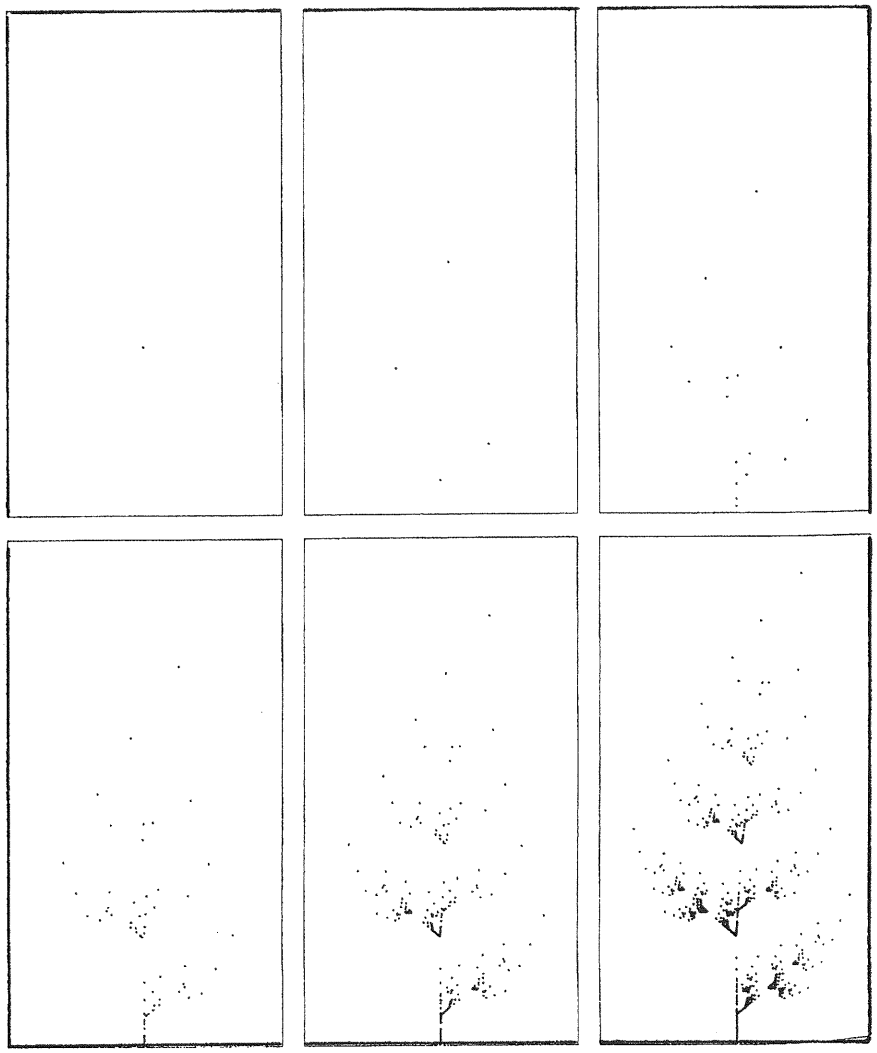
8. Conclusion

Je la laisse à LAPLACE qui, en 1812, affirmait :

> ***« Et si l'on observe ensuite que dans les choses qui peuvent ou non être soumises au calcul, la théorie des probabilités apprend à se garantir des illusions, il n'est pas de science qu'il soit plus utile de faire entrer dans le système de l'instruction publique. »***

Mais je me permettrai de la compléter par une réflexion plus personnelle :

> ***« La formation de la pensée statistique, çà n'est pas l'école du mensonge, c'est celle de l'humilité ».***



ORGANISATION ET GESTION DE DONNEES AU COLLEGE

Objectifs de la 6ème Initiation à la lecture, à l'interprétation et à l'utilisation de diagrammes, tableaux et graphiques, et à l'analyse critique. La réalisation de tels objectifs contribue à l'éducation civique.		
Contenus	**Compétences exigibles**	**Commentaires**
Exemples issus d'activités : **- à base numérique** Relevés statistiques	Appliquer un taux de pourcentage.	On se servira de ces exemples pour : lire et établir des relevés statistiques sous forme de tableaux ou de représentations graphiques, éventuellement en utilisant un ordinateur.
Objectifs du cycle central Acquisition de quelques outils statistiques utiles dans d'autres disciplines et dans la vie de tout citoyen. Initiation à la lecture, à l'interprétation, à la réalisation et à l'utilisation de diagrammes, tableaux et graphiques et à l'analyse critique de ceux-ci.		
5ème		
Contenus	**Compétences exigibles**	**Commentaires**
Lecture, interprétation, représentations graphiques de séries statistiques. Classes, effectifs.	Lire et interpréter un tableau, un diagramme à barres, un diagramme circulaire ou semi-circulaire. Regrouper des données statistiques en classes, calculer des effectifs. Présenter une série statistique sous la forme d'un tableau, la représenter sous la forme d'un diagramme ou d'un graphique.	Il importe d'entraîner les élèves à lire et à représenter des données statistiques en utilisant un vocabulaire adéquat. Le calcul d'effectifs cumulés n'est pas une compétence exigible mais il pourra être entrepris, en liaison avec les autres disciplines dans des situations où les résultats auront une interprétation. Le choix de la représentation est lié à la nature de la situation étudiée. La notion de fréquence est notamment utilisée pour comparer des populations d'effectifs différents, et faire le lien avec la proportionnalité. Les écritures 4/10, 2/5, 0,4 (ou en notation anglo-saxonne 0.4 ou .4), 40% qui peuvent être utilisées pour désigner une fréquence, permettent d'insister sur les diverses représentations d'un même nombre.
Fréquences.	Calculer des fréquences.	

4ème		
Effectifs cumulés, fréquences cumulées. Moyennes pondérées. Initiation à l'utilisation de tableurs-grapheurs.	Calculer des effectifs cumulés, des fréquences cumulées. Calculer la moyenne d'une série statistique. Calculer une valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles.	L'élève sera confronté à des situations courantes où la méthode de calcul est à remettre en cause : par exemple, les différences constatées entre la moyenne annuelle des notes d'un élève calculée à partir de l'ensemble des notes de l'année ou à partir de la moyenne des moyennes trimestrielles. Les tableurs-grapheurs, utilisés dès la 5ème en technologie, introduisent une nouvelle manière de désigner une variable : par l'emplacement de la cellule où elle se trouve dans un tableau. Cette nouveauté est un enrichissement pour des utilisations dont on pourra donner des exemples. Pour les graphiques des choix successifs sont proposés, ils conduisent naturellement à examiner leur pertinence pour l'utilisation d'une situation donnée.
Objectifs de la 3ème Poursuivre l'étude des paramètres de position et aborder l'étude de la dispersion. Prendre l'habitude de s'interroger sur la signification des nombres utilisés, sur l'information apportée par un résumé statistique et donc sur la perte d'information, sur les possibilités de généralisation, sur les risques d'erreurs d'interprétation et sur leurs conséquences possibles.		
Contenus	**Compétences exigibles**	**Commentaires**
Caractéristiques de position d'une série statistique Approche de caractéristiques de dispersion d'une série statistique. Initiation à l'utilisation de tableurs-grapheurs en statistique.	Une série statistique étant donnée (sous forme de liste ou de tableau, ou par une représentation graphique), proposer une valeur médiane de cette série et en donner la signification. Une série statistique étant donnée, déterminer son étendue ou celle d'une partie donnée de cette série.	Il s'agit essentiellement d'une part, de faire acquérir aux élèves les premiers outils de comparaison de séries statistiques, d'autre part de les habituer à avoir une attitude de lecteurs responsables face aux informations de nature statistique. On repère, en utilisant effectifs ou fréquences cumulés, à partir de quelle valeur du caractère on peut être assuré que la moitié de l'effectif est englobée. Les exemples ne devront soulever aucune difficulté au sujet de la détermination de la valeur de la médiane. L'étude de séries statistiques ayant même moyenne permettra l'approche de la notion de dispersion avant toute introduction d'indice de dispersion. On introduira l'étendue de la série ou de la partie de la série obtenue après élimination de valeurs extrêmes. On pourra ainsi aborder la comparaison de deux séries en calculant quelques caractéristiques de position et de dispersion, ou en interprétant des représentations graphiques données. Les tableurs que l'on peut utiliser sur tous les types d'ordinateurs permettent, notamment en liaison avec l'enseignement de la technologie, d'appliquer de manière rapide à des données statistiques les traitements étudiés.

HISTOIRE DES STATISTIQUES ET DES PROBABILITES : UN SURVOL

Marie-Françoise JOZEAU
IUFM de Versailles
Equipe M. :A.T.H. de l'IREM de Paris 7

Résumé : *Jusqu'alors, l'enseignement de la statistique en France, dans les classes du secondaire, voire du primaire, ne permettait pas aux élèves d'aborder les questions auxquelles « tout citoyen à l'aube du 21ème siècle » devrait être sensibilisé. Depuis 2000, cependant, les programmes amorcent une évolution qui devrait permettre aux lycéens de s'initier à la pensée statistique.*
Un parcours historique peut nous aider à comprendre les choix effectués pour l'enseignement de cette branche des mathématiques et à saisir les rapports existants entre statistique et probabilités[1].

1. Introduction

ETYMOLOGIE

- **Statisticus 1672 « relatif à l'état »**

  (latin moderne formé à partir de l'italien)

Statitica 1633 dérivé de Statista

« l'homme d'état »

Introduit en France vers 1660

Introduit en Allemagne vers 1750 :

Statistik

Introduit en Angleterre vers 1798 :

Statistics

L'histoire de la statistique comporte deux volets : un volet externaliste et un volet internaliste.

a) Les différentes notions, les différents outils de la statistique se sont élaborés, forgés, mis en place dans des domaines les plus variés (voir document 1).

Leur première apparition ne se fait ni dans des traités de mathématique, ni dans des comptes-rendus de travaux de recherches mathématiques. Il faut consulter, par exemple, des ouvrages d'actuaires, d'artillerie, de biométrie, etc... L'histoire des institutions statistiques et des congrès statistiques permet de suivre les échanges à ce sujet entre différents pays et aussi de comprendre les politiques spécifiques à chaque pays.

[1] Cet article est le compte-rendu d'une conférence sur l'histoire des statistiques faite lors d'un stage pour des enseignants du second degré.

Ars conjectandi (1713)
Ouvrage de Jacques BERNOULLI (1654-1705)
Cet ouvrage paraît à Bâle, 8 ans après la mort de Jacques. Il aurait pu le publier dès 1689. Mais, il cherche, en vain, pendant 15 ans des applications de ce calcul.

b) La statistique mathématique utilise le calcul des probabilités ; il est important alors de connaître l'histoire de ce calcul, calcul où le hasard, le probable sont présents et de **la loi des grands nombres [J.BERNOULLI 1713]** loi qui établit « un pont » entre le certain et le probable.

La statistique mathématique s'est constituée au 20ème siècle. Un essai de généalogie (voir document 2) permet de mettre en évidence différentes traditions, correspondant à des zones géographiques différentes, qui, depuis le 18ème siècle, ont permis son émergence : statistique descriptive allemande, arithmétique politique anglaise, jeux et probabilités françaises. Puis, une théorie des erreurs avec l'invention de la méthode des moindres carrés voit le jour simultanément en France avec LEGENDRE en 1805 et en Allemagne avec GAUSS. Une première synthèse sera faite par LAPLACE en 1810. Ensuite, le belge QUETELET (1846) sera surtout sensible à la notion de moyennes et aux régularités, tandis que les anglais GALTON (1889), K. PEARSON (1896) et YULE (1897) le seront à la notion de variabilité et d'écarts à la moyenne : ils donneront naissance à la biométrie. Parallèlement, le calcul des probabilités progresse avec, par exemple, les français POISSON (1837), COURNOT (1843)... puis les anglais EDGEWORTH (1885), KEYNES (1921), et enfin le russe KOLMOGOROV qui axiomatise les probabilités en 1933. Pour des raisons institutionnelles et de tradition culturelle, la France « bourbakiste » recevra très vite cette axiomatisation dans l'université et sera moins réceptive, dans les milieux universitaires, à la statistique inférentielle élaborée en Angleterre par R. FISCHER (1925) pour les problèmes d'estimation, par E. PEARSON (1928) pour les test d'hypothèse, par J. NEYMANN et E. PEARSON pour la théorie moderne de l'échantillonnage.

Ce bref parcours peut aider à comprendre l'immobilisme des programmes du secondaire, ancrés dans la tradition française des jeux (avec les problèmes de dénombrement) et l'axiomatisation des probabilités par la théorie des ensembles. Il est peut-être temps de prendre en compte les autres approches !

1.1. Les statistiques ou la statistique ?

Ce vocable possède deux acceptations distinctes :
- un ensemble de données ;
- un ensemble de méthodes pour traiter les données.

Cette distinction date du XXème siècle :
- quand sont diffusées les techniques de la régression et de la corrélation, à partir du centre de biométrie de K. PEARSON ;
- puis la statistique inférentielle développée au laboratoire expérimental d'agriculture de Ronald FISCHER.

On peut s'étonner, à juste titre, que près de vingt siècles séparent la collecte des données (pensons aux premiers recensements) de la statistique inférentielle. Le premier est du domaine du **Certain** tandis que le deuxième est du domaine du **Probable**.
Pour passer de l'un à l'autre, il est nécessaire de faire intervenir **LE HASARD**.
Cet obstacle fut difficile et long à franchir !

Le hasard est une notion étrangère à la pensée primitive ; on se réfère alors au divin. Plus tard, l'idée de parier sur la vie et sur la mort est elle aussi étrangère à la morale chrétienne. Les superstitions sur le hasard ne permettront pas à la composante stochastique (aléatoire) de se développer ; elle sera maîtrisée bien plus tard que la composante déterministe.

Géométrie du hasard

PASCAL

1654

Vers le XVII^{ème} siècle une approche quantitative du hasard est amorcée. En effet, il faut avoir la possibilité de **penser le hasard**, ce qui va être permis par les diverses évolutions de la société : jeux de hasard, capitaux à risque nécessaires au commerce maritime (usure), contrats d'assurances, rentes viagères ...

Les théologiens discutent : un tel enrichissement « sans cause légitime » est-il licite ?
Un problème d'équité se pose au moment des partages[2]. Ainsi, la notion d'**espérance** apparaît avant celle de probabilité. Les espérances calculées doivent être telles que les diverses alternatives soient indifférentes aux joueurs.

Contrats aléatoires

Un avoir certain dans le présent ⟷ échangé avec ⟷ Enrichissement aléatoire dans l'avenir

1.2. La préhistoire

Au commencement était l'idée de dénombrement, d'inventaire, de recensement. Ainsi, par exemple, des tablettes d'argile datant de la civilisation de Sumer (5ème-2ème s av. J.C) contiennent des listes de biens et d'hommes. Les premiers recensements sont réalisés par les Hébreux, les Chinois puis plus tard par l'Inde, le Japon, la Grèce et Rome.
MAIS la statistique ici se contente de rassembler, de compter et parfois de classer les données. Il ne s'agit pas encore de les traiter.
Il faudra attendre plus de 20 siècles pour que LA STATISTIQUE se constitue. Comme nous l'avons dit ci-dessus, plusieurs traditions participent à sa naissance.

[2] Voir le problème du Chevalier DE MERE, le problème des partis....

2. Première tradition : les Jeux

D'origine essentiellement française, cette tradition remonte au XVI[ème] siècle ; on en trouve trace chez CARDAN ainsi que chez GALILEE[3].

Blaise PASCAL présente ainsi cette **Géométrie du hasard** :

> *« Et puis un traité tout à fait nouveau, d'une matière absolument inexplorée jusqu'ici, savoir : la répartition du hasard dans les jeux qui lui sont soumis, ce qu'on appelle en français faire les partis des jeux.[...] Et grâce à la géométrie, nous l'avons réduite avec tant de sûreté à un art exact, qu'elle participe de sa certitude et déjà progresse audacieusement. Ainsi, joignant la rigueur des démonstrations de la science à l'incertitude du hasard, et conciliant ces choses en apparence contraires, elle peut, tirant son nom des deux s'arroger à bon droit ce titre stupéfiant : **la géométrie du hasard** »*[4].

Quelques repères chronologiques

CARDAN (1501-1576) : quelques recherches sur les problèmes de jeux.

FERMAT (1601-1665) et PASCAL (1623-1662) : correspondance en 1654 au sujet de la règle des partis, c'est-à-dire de partage de la mise lorsqu'un joueur désire se retirer avant la fin de la partie.

HUYGENS (1629-1695) : *De ratiocinus in ludo alea*, 1657. Ce livre constitue le premier ouvrage sur le sujet.

Les premiers traités de probabilités sont dus à :

MONTMORT (1678-1719) *Essai d'analyse sur les jeux de hasard*, 1708.

MOIVRE (1667-1754) *De mensura sortis*, 1711.

Doctrine of chances, 1718.

Quelques problèmes célèbres

Luca PACIOLI (1445-1514)

> *« Deux camps jouent à la balle ; chaque manche est de 10 points et il faut 60 points pour gagner ; la mise totale est de 10 ducats. Il arrive que, le jeu ne puisse s'achever. On demande ce que touche de la mise totale chacun des 2 camps, lorsque l'un a 50 points et l'autre 20 points » (Summa de arithmetica 1509)*[5].

Voici la solution que propose PACIOLI :

> *« Les deux camps peuvent faire en tout 110 points (à savoir quand il manque une manche à chacun et que l'un gagne la dernière). Puis considérons quelle partie*

---

[3] Gerolamo CARDANO, *Liber de ludo alea,* in *Opera Omnia,* vol.1, Lyon, 1663 ; écrit sans doute vers le milieu du XVI[ème] siècle.

Galilei GALILEO, *Considerazione sopra il giuco dei dadi* in *Opere,* t.14, Firenze, 1755, 1[ère] édition en 1718 ; la date de rédaction est vers 1620.

[4] PASCAL, *Œuvres complètes,* Seuil, 1988.

[5] La première édition date de 1494.

de ces 110 points ont obtenu chacun de ces deux camps. Le premier en a obtenu 5/11 et le second 2/11. Par conséquent, ils doivent retirer de la mise totale des parties proportionnelles à ces nombres dont la somme est 7/11. On les calculera ainsi : 7/11 gagnent 10 ducats ; combien gagnent 5/11 et 2/11 ? On verra que le premier camp aura 7 ducats 1/7 et le second 2 ducats 6/7 ».

Nicolas TARTAGLIA donne son avis sur cette solution en 1556 dans *La prima parte del general di numere e misure :*

« Sa règle ne me paraît ni bonne, ni belle, parce que s'il arrive qu'un parti ait 10, et l'autre rien, et que l'on procédât selon sa règle, le premier devrait tirer le tout et le second rien ; et ce serait tout à fait déraisonnable que, pour 10 points, il doive tirer le tout ».

- Des problèmes de dés

"En combien de coups peut-on espérer faire sonnez[6] avec deux dés ?"
Chevalier de MERE (1654).

« Si on entreprend de faire un six avec un dé, il y a avantage de l'entreprendre en 4, comme de 671 à 625.
Si on entreprend de faire sonnez avec deux dés, il y a désavantage de l'entreprendre en 24.
Et néanmoins 24 est à 36 (qui est le nombre des faces des deux dés) comme 4 à 6 (qui est le nombre des faces d'un dé) ».

PASCAL commentera cette démonstration dans une lettre à FERMAT daté du 29 juillet 1654 :

« Voilà quel était son scandale qui lui faisait dire hautement que les propositions n'étaient pas constantes et que l'arithmétique se démentait : mais vous en verrez bien aisément la raison par les principes où vous êtes ».

Remarquons que, comme chez nos élèves, l'idée de proportionnalité est toujours première !

- Pour l'histoire du problème des partis, combien célèbre, nous renvoyons à quelques textes contemporains l'étudiant de façon approfondie[7] et analysant entre autre la correspondance

---

[6] Sortir un double six

[7] Voir par exemple : P.J. About et M. Boy, *La correspondance de Blaise Pascal et de Pierre de Fermat, la Géométrie du Hasard ou le début du Calcul des Probabilités,* Les cahiers de Fontenay n°32, 1983 ; Cahiers pédagogiques de philosophie et d'histoire des mathématiques, fasc. 4 : *Pascal et les probabilités,* CRDP-IREM de Rouen, 1993 ; IREM de PARIS VII, collection M/.A.T.H *Mathématiques : Approche par les textes historiques,* n°61, 1986 et Mnémosyne n°6, 1993 ; IREM de POITIERS, *Thèmes pour l'enseignement de la statistique et des probabilités,* 1994 ; Commission inter-IREM Statistiques et probabilités, *Enseigner les probabilités au lycée,* réseau des IREM avec le soutien de la DLC, 1997.

PASCAL-FERMAT datant de 1654. POISSON présente ainsi ce problème dans *Recherches sur la probabilité de jugement* en 1837 :

> *« Un problème relatif aux jeux de hasard, proposé à un austère janséniste par un homme du monde, a été l'origine du calcul des probabilités. Il avait pour objet de déterminer la proportion suivant laquelle l'enjeu doit être partagé entre les joueurs, lorsqu'ils conviennent de ne pas achever la partie, et qu'il leur reste à prendre, pour la gagner, des nombres de points inégaux. Pascal en donna le premier la solution, mais pour le cas de deux joueurs seulement ; il fut ensuite résolu par Fermat, dans le cas général d'un nombre quelconque de joueurs ».*

Quelques remarques sur le vocabulaire
On ne dit pas, par exemple, que la probabilité de gagner est de 6/8 mais qu'il y a 6 chances contre 2 chances de gagner. La notion « d'espérance » chez Huygens : Prop 1 : si j'espère a ou b, et que l'un ou l'autre puisse m'advenir aussi facilement, on doit dire que mon espérance vaut (a+b)/2 (note bas de page [8]).

3. Deuxième tradition : l'arithmétique politique

L'expression est inventée par Sir William PETTY (1623-1687), membre de la Royal Society de Londres. Charles DAVENANT (1656-1714) définit ainsi l'arithmétique politique : *« L'art de raisonner avec des chiffres sur des matières touchant au gouvernement »*. Il s'agit, effectivement, d'une approche scientifiquement établie et destinée à éclairer les décisions politiques. Dans *L'Esprit des Lois, Critique de l'Economie Politique,* MONTESQUIEU et Karl MARX s'en sont inspirés. Mais la paternité légitime de l'arithmétique politique reviendrait plutôt à John GRAUNT (1620-1674)[9]. Comme il apparaît, cette tradition se situe en Angleterre, dans la deuxième moitié du XVII[ème] siècle[10].

En effet, à cette époque, le besoin d'élaborer des statistiques pour des objectifs précis se fait sentir : lever les impôts, payer des frais d'expéditions militaires, contrôler le comportement religieux de la population (baptêmes, décès...). Jusqu'alors, l'arithmétique appliquée aux faits politiques, sociaux ou culturels intervient dans trois domaines : les calculs eschatologiques[11],

---

[8] Christian Huygens, *Ratiociniis in aleae ludo,* 1657 ; voir aussi une traduction française de ses Oeuvres, tome 14, 1888-1950, La Haye.

[9] *Natural and Political Observations upon the Bills of Mortality ... of the City of London,* 1661. Cette oeuvre révolutionnaire connaît cinq éditions en 14 ans. Il existe une traduction française due à E. VILQUIN, Paris, INED, 1977.

[10] En France, les statistiques n'étaient accessibles qu'aux administrateurs, alors qu'en Angleterre, elles étaient à la disposition de tout le monde.

[11] Il s'agit du calcul de l'âge du monde ; supputation sur la date du Jugement Dernier à partir des écritures et prophéties. NAPIER et même NEWTON s'y sont intéressés.

les recensements fiscaux et les registres paroissiaux d'état civil. Le premier domaine a contribué à rendre habituel l'usage de l'arithmétique à des fins prévisionnelles. Les deux autres domaines, plus anciens, se caractérisent par leur imprécision et une absence totale de méthodes. L'apport anglais consiste en l'introduction d'une méthode rigoureuse et clairement énoncée s'appuyant sur des sources relativement crédibles et soumises à une vraie critique : les registres paroissiaux. En effet, l'église tient un état civil à cause de la complexité du Droit Canon en matière matrimoniale. A Londres, se pratique, depuis 1519, l'établissement de bulletins de mortalité pour la surveillance des épidémies. Ils servent de matières premières à J. GRAUNT ; ce dernier utilise les renseignements fournis par les bulletins pour établir et vérifier certaines assertions controversées : causes réelles des décès, étendue exacte de la population de Londres, rapport entre le nombre de femmes et d'hommes... Un tabou est brisé ; il a osé tenter de déchiffrer les secrets de Dieu. Il a désacralisé la mort qui devient un sujet d'analyse et de calcul. Sa méthode est très simple : usage des 4 opérations, éclairée par des tableaux chronologiques où il transfère les données numériques des bulletins.

> ***« C'est sans doute dans ce passage de la notion de liste à celle de tableau que l'on peut situer la naissance de la statistique ».*** E. VILQUIN

Les tableaux autorisent sommes, calculs de fréquences relatives, moyennes, amplitudes, accroissement relatif...
GRAUNT ajoute à cette méthode un esprit critique vis-à-vis des chiffres, des sources... ; il se questionne sur les erreurs possibles. Il innove en extrapolant les résultats de Londres à l'Angleterre[12] mais, malgré tout, il reste un homme de son temps, étant persuadé que les lois naturelles qui régissent la vie et la mort ont leur source dans un décret divin.
PETTY affirme aussi la primauté de la quantification et de l'examen expérimental. Dans sa préface, il écrit : « *J'ai adopté la méthode qui consiste à m'exprimer en termes de nombres, poids et mesures* »[13]. Mais son œuvre manque de rigueur ; il se contente de l'approximatif, ne donne aucune explication ni sur sa manière de procéder, ni sur l'usage éventuel du document qu'il présente.

Bien que la notion de table de mortalité soit déjà présente chez J. GRAUNT, c'est HALLEY Edmund (1656-1742) qui, en 1693[14], est le premier à construire une table de mortalité en s'appuyant sur des observations (voir documents 3 et 4) alors que J. GRAUNT avait utilisé une fonction arbitrairement définie ; il l'utilise pour calculer la rentabilité du système anglais des tontines[15].

---

[12] Plus tard, LAPLACE utilisera la méthode du coefficient multiplicateur pour évaluer la population française à partir d'un échantillon (cf. *Théorie Analytique des probabilités,* T.2, 1812). Cette méthode sera reprise aussi par QUETELET pour évaluer la population des Pays-Bas (cf. *Correspondance Mathématique et Physique,* T.5, 1829).
[13] W. PETTY, *Political Arithmétic,* écrit vers 1673, publié en 1690.
[14] *An estimate of the degree of the Mortality of Mankind...* In Philosophicals Transactions p.596-610 (1693).
[15] Pour approfondir l'histoire des tables de mortalité, voir J. DUPAQUIER, *L'invention de la table de mortalité,* PUF, 1996.

La diffusion de l'arithmétique politique s'effectue dans toute l'Europe. Cependant, en Allemagne, elle a du mal à s'installer à cause de CONRING qui perpétue la tradition aristotélicienne. Ce qui nous conduit à la dernière tradition évoquée.

4. Troisième tradition : la statistique descriptive

« *Une façon de classer des savoirs hétéroclites plus facile à retenir, à utiliser par les politiques...* ». Ainsi, Hermann CONRING (1606-1681) définit-il la statistique descriptive allemande. Il invente le mot « stati » pour désigner cette école.

La statistique allemande ne met pas l'accent sur des méthodes quantitatives comme l'arithmétique politique. Pourtant le pôle que constitue cette statistique est significatif en tant que façon de représenter le monde social. Il y apparaît un effort pour penser et décrire la diversité territoriale d'un Etat national. Cette statistique se forme dans un espace géographique où règnent trois cents Princes sur des petits pays en ruine avec un pouvoir absolu. Ces princes devaient maintenir leur indépendance à la fois contre l'Empereur et contre la France. Ils furent donc amenés à rechercher les compétences en matière de droit et d'administration. Pour décrire les caractéristiques de l'état national, il est donc nécessaire de rassembler tous les renseignements issus de chaque pays. En 1655, l'ouvrage Teutscher Fürsten Staat de Ludwig VON SECKENDORF (1626-1692) paraît alors qu'il est conseiller à la cour de Gotha. Ce livre définit, à l'usage des administrateurs, le fonctionnement politique et économique d'un Etat allemand type. Il propose des modèles pour une description approfondie de chaque région, avec une analyse de ses richesses naturelles. Pour les recensements, ses modèles permettent d'énumérer les feux, les chefs de famille et quelques catégories professionnelles. Les renseignements sont consignés dans des tableaux croisés, ce qui est une innovation intéressante (voir documents 5 et 6) ; contrairement aux récits de voyageurs, où les descriptions prennent place dans un discours linéaire, l'auteur, en utilisant les deux dimensions de la feuille (lectures horizontale et verticale), permet un accès beaucoup plus facile à l'information, permettant ainsi des comparaisons plus rapides et plus efficaces entre deux régions, par exemple.

CONRING, quant à lui, applique à l'étude des phénomènes démographiques la vieille grille de recherches des causes matérielles (le territoire et le nombre d'hommes), finales (les buts que s'assigne la communauté), formelles (le droit, la coutume, la nature du pouvoir du Prince) et efficientes (l'administration, l'appareil bureaucratique et judiciaire).

Avec ACHENWALL, les comparaisons entre Etats s'effectuent désormais selon 6 critères : la description du pays, sa population et ses forces économiques, la structure de l'Etat, son administration, sa santé et ses intérêts extérieurs.
Cette tradition ne doit pas être sous-estimée ; elle a permis de mettre l'accent sur la récolte des données et sur leur organisation afin de procéder à des comparaisons. Certes, l'appel à

l'arithmétique est absent, mais cette tradition montre bien le travail nécessaire avant de traiter les données.

> **Critique des « faiseurs de tableaux » :**
> « *Ces pauvres imbéciles répandent l'idée insensée que l'on peut comprendre la puissance d'un Etat en ne connaissant que sa superficie, sa population, son revenu national et le nombre des animaux broutant alentour* ».
> « *Les machinations auxquelles se livrent ces criminels de statisticiens-politiques dans leurs efforts pour exprimer tout par des chiffres[...] sont méprisables et ridicules au-delà de toute expression* ». Extraits de V. JOHN.

Ces trois traditions se fondent sous la révolution et l'Empire en France. CONDORCET, LAPLACE y participent. Le rôle de QUESNAY est important au XVIII^{ème} siècle. Il a l'idée d'essayer de créer un modèle qui simule la société par une nomenclature (tradition allemande) et par des mesures (tradition anglaise). L'idée de modèle a des connotations diverses : descriptive, causale, normative et probabiliste (tradition française). Ainsi, chez QUESNAY on retrouve déjà l'aspect descriptif, l'aspect explicatif et l'aspect prescriptif.

L'idée de modèle empirique est née, mais les outils pour en éprouver la solidité ne seront construit qu'un siècle plus tard avec QUETELET.

5. Apparition des premiers graphiques

Les premières représentations graphiques apparaissent en même temps que les tableaux, empruntant leur forme aux traités de mathématiques : figures géométriques, tracés de courbes, coordonnées rectangulaires.
Un grand évènement est le graphique inventé par Christian HUYGENS en 1669 pour représenter la mortalité ; la courbe tracée représente le nombre de survivants en fonction de l'âge d'après la table de mortalité de GRAUNT (voir document 7).

L'anglais William PLAYFAIR introduit les figures géométriques en 1786, dans un atlas commercial et politique. Chaque pays est représenté par un cercle de surface proportionnelle à sa superficie, avec des graphiques de couleur pouvant rendre compte de la population, du revenu et de l'évolution du commerce (voir document 8). Il invente à cette occasion le graphique en bâtons. Il cherche à systématiser ce genre de représentation sous le nom d'arithmétique linéaire.

Plus tard, A. DE HUMBOLT, en 1811, reprendra ce type de graphiques mais il remplacera les cercles par des carrés.

6. La loi des grands nombres

Les frères BERNOULLI (voir documents 9 et 10)
Jacques (Jacob) BERNOULLI (1654-1705). Jean (Johann) BERNOULLI (1667-1748). Daniel (fils de Jean) BERNOULLI (1700-1782). Nicolas BERNOULLI (1687-1759) (frère de Jacques et Jean).

Jacques BERNOULLI
*Diplômé de théologie. *1679 ; 1682 : 2 voyages d'étude à Paris, auxPays-Bas et en Angleterre. *Université de Bâle : assistant en physique expérimentale. *1682 *Théorie des comètes* *1683 *Théorie de la gravitation* *1687 Professeur à l'université de Bâle *1686 à 1689 dans son journal scientifique, il développe la démonstration du Théorème qui fait l'objet du 5ème chapitre de la 4ème partie de l'***Ars Conjectandi,*** publié en 1713.

6.1. Le projet de BERNOULLI

Il s'agit, pour BERNOULLI, d'établir le nombre très grand d'expériences qu'il faut faire d'un événement pour que l'on puisse estimer la probabilité de cet événement, en l'encadrant par 2 valeurs limites, avec une chance quantifiable très petite de se tromper. C'est l'objet de la quatrième partie de *l'Ars Conjectandi.* Cette œuvre paraît à Bâle en 1713 alors que Jacques est mort depuis 8 ans. Le 25 décembre 1713, MONTMORT écrit à Nicolas, neveu de Jacques : « *Il est vrai, ce que vous dites, qu'il y a peu de choses à apprendre pour nous, qui avons beaucoup pensé à ces matières. Le livre ne laisse pas d'être excellent* ». En effet, entre temps, MONTMORT, DE MOIVRE ont écrit au sujet des probabilités. **En 1713, le livre arrive trop tard et manque son but :** l'application de cet art à « la vie civile, économique et sociale ».

Comment expliquer cette parution si tardive ? Dès 1685, Jacques commence à travailler sur des problèmes de jeux de hasard en méditant l'ouvrage DE HUYGENS *Du calcul dans les jeux de hasard.* Il aurait pu publier l'*Ars Conjectandi* à partir de 1689, mais il cherche en vain pendant 15 ans des applications à son théorème, comme il s'en ouvre à LEIBNIZ dans une lettre du 3/10/1703 :

> « *J'ai déjà terminé la plus grande partie de mon Livre, mais il me manque encore la partie essentielle dans laquelle je montre comment les bases de l'art de conjecturer peuvent s'appliquer à la vie civile, sociale et économique* ».

Jacques demande à LEIBNIZ des données juridiques (en particulier le traité de DE WITT) mais vainement... LEIBNIZ semble hostile au projet de Jacques et pourtant, en 1714, après la mort de celui-ci, il semble s'approprier son idée :

> *« La possibilité de déterminer le nombre de cas qui permettent de calculer les degrés de probabilités d'un événement « a postiori » d'une façon aussi certaine que si nous la connaissions « a priori » ».*

Ce qu'il appelle **L'ART DE CONJECTURER**

Les péripéties de la publication

- La veuve demande à HERMANN, ancien élève de Jacques, de rassembler les papiers qui pourraient être publiés.
- En 1706, SAURIN dans le Journal des Savants décrit correctement le manuscrit.
- En 1706, HERMANN publie dans *Acta Eruditorum* un éloge dont LEIBNIZ retire les passages relatifs aux probabilités.
- SAURIN suggère au frère Jean de faire la publication. Mais, les deux frères étant fâchés, la veuve interdit à Jean l'accès aux papiers. Jean écrit à SAURIN de s'adresser au neveu pour ne pas décevoir le public.
- En 1709, LEIBNIZ et MONTMORT s'intéressent à la publication.
- Nicolas, neveu de Jacques et Jean, utilise des passages, sans doute transmis par HERMANN, pour sa thèse en 1709. Il s'efforce de trouver une solution pour publier le texte entier et le 26/02/1711, il écrit à MONTMORT. Le fils de Jacques donne son accord, mais alors, la veuve s'y oppose. Au début de l'année 1713, Nicolas séjourne chez MONTMORT, mais le fils de Jacques a fait faire la correction par un certain BRINGOLF, et vend les droits aux éditeurs THURNEYSEN.

Jean écrit alors à Nicolas le 15/05/1712 :

> *« Il va en résulter quelque chose de beau si ni l'éditeur, ni le correcteur, ni l'imprimeur ne comprennent rien à la matière... »*

VARIGNON lui aussi s'en émeut auprès de Jean et Nicolas, mais l'édition sort quand même ! Nicolas préfacera le texte s'adressant au lecteur (voir document 11).

Cet ouvrage comporte 4 parties : le traité de HUYGENS avec la résolution des problèmes proposés, un traité de combinatoire, l'application de ces outils à la résolution de problèmes relevant des jeux de hasard et enfin la si célèbre proposition ! (voir document 12).

Aide à la lecture de la proposition :

On se place dans le cas où il y a deux issues favorables ou non : A et $\overline{A}$. **La probabilité a priori est connue ;** les cas observés a postiori fournissent une estimation de la proportion des cas a priori. Ainsi, la fréquence observée fournit une estimation de la probabilité. BERNOULLI, par son théorème répond en quelque sorte à la question : Quelle confiance peut-on accorder à cette estimation ?

Remarquons le vocabulaire : il parle de cas féconds et de cas stériles.

Il pose Card A = r Card $\overline{A}$ = s r + s = t

Pour nous, $p(A) = \frac{r}{t}$

BERNOULLI prend $\frac{r}{s} = \frac{p(A)}{p(\overline{A})}$

Le vocabulaire est celui lié au jeu de hasard à l'époque (remarquons que $\frac{r}{s} = \frac{10kr}{10ks}$

On a p(A) qui vérifie : $\frac{r-1}{t} < p(A) < \frac{r+1}{t}$

Posons, de façon moderne,

X une variable aléatoire égale au nombre de succès observés (événements fertiles) ou égale au nombre de fois où A survient.

N nombre d'observations.

$Y = \frac{X}{N}$, fréquence relative de A

Soit c > 0 un grand nombre

Soit E : « $\frac{r-1}{t} \leq Y \leq \frac{r+1}{t}$ »

Il cherche N tel que : $\frac{p(E)}{p(\overline{E})} > c$

Traduisons : $p(E) > cp(\overline{E})$

$P(E) > c(1-p(E))$

$(1+c)p(E) > c$

$p(E) > \frac{c}{1+c}$

$p(E) > 1 - \frac{1}{1+c}$

On retrouve un théorème limite : lim p(E) = 1, c'est-à-dire que :

$$\forall c>0, \exists N, 1-\frac{1}{1+c} < P_N\left(\frac{r-1}{t} \leq X \leq \frac{r+1}{t}\right) \leq 1$$

On peut choisir t pour resserrer les limites.

Soyons plus rigoureux : soit p la proportion inconnue, alors en posant $\varepsilon = \frac{1}{t}$ et N=nt on a :

$$\forall \varepsilon > 0, \forall c > 0, \exists N \text{ tel que } p\left(\left|\frac{X}{N} - p\right| \leq \varepsilon\right) > cp\left(\left|\frac{X}{N} - p\right| \geq \varepsilon\right) \text{ c'est-à-dire } p\left(\left|\frac{X}{N} - p\right| > \varepsilon\right) < \frac{1}{1+c}$$

Remarques :

1) r et s sont des nombres entiers ; donc ici, $p = \frac{r}{r+s}$ ne décrit pas tout l'intervalle [0 ;1] ; il ne prend que quelques valeurs. Le but de BERNOULLI est simplement de montrer que le rapport exact $\frac{r}{r+s}$ peut être reconnu comme une « certitude morale » pour un nombre N suffisamment grand. Ainsi, a moins que N soit un multiple de r+s, il n'est pas possible

d'obtenir $\frac{X}{N}=\frac{r}{r+s}$ exactement. Et même si N est multiple de r+s, la probabilité que cette égalité soit observé diminue quand N augmente. BERNOULLI laisse entendre que l'on obtient seulement une approximation mais ces exemples montrent qu'il cherche une valeur exacte ! Chez BERNOULLI, ε n'est jamais quelconque, il prend $\frac{1}{r+s}$.

2) Sa preuve est très rigoureuse.

3) Le fait que BERNOULLI ne cherche pas à se rapprocher d'une proportion p, mais de la valeur exacte va freiner le développement des statistiques. Il n'a pas de recul conceptuel pour juger son entreprise, voire la condamner.

4) BERNOULLI, comme LAPLACE, pense que dans tous les domaines (les jeux de hasard n'étant qu'un cas particulier), l'incertitude n'est qu'une apparence, un défaut de notre entendement ; Elle n'est pas dans les choses mais dans la méconnaissance que nous en avons. *« Tout ce qui bénéficie sous le soleil de l'être ou du devenir, passé, présent ou futur, possède toujours en soi et objectivement une certitude totale »*. Il définit ainsi la probabilité : *« La probabilité [...] est un degré de la certitude et en diffère comme la partie diffère du tout »*.

6.2. La mathématique sociale de CONDORCET (1743-1794)

Le projet de CONDORCET est d'introduire la science mathématique dans des domaines où elle était absente jusqu'à maintenant : sciences politiques et morales, mortalité, élection, loterie, assurances maritimes, rentes viagères, assurances sur la vie, justice criminelle... La mathématique sociale a pour objet les hommes, les choses ou les deux à la fois. Il écrit en 1785 dans *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des lois* :

> *« On verra combien, si cette science était plus répandue, plus cultivée, elle contribuerait et au bonheur et au perfectionnement de l'espèce humaine »*.

Les statistiques criminelles sont apparues dans la deuxième moitié du XVIII^{ème} siècle, marquée par des erreurs judiciaires (comme par exemple l'affaire CALLAS), et par des tentatives de réformes judiciaires. CONDORCET découvre des anomalies dans l'organisation judiciaire et politiques des sciences modernes, avec le paradoxe des votes qui met en évidence la formation d'une réponse collective contradictoire à partir de réponses individuelles cohérentes.

Ainsi, CONDORCET apparaît comme un précurseur du mouvement visant à une approche scientifique des questions de comptabilité et de prévision des actes humains dans lequel il faut situer les travaux de QUETELET.

7. Statistique au XIX^{ème} siècle

7.1. Méthode des moindres carrés et loi normale

De nouvelles questions voient le jour au début du XIX^{ème} siècle. Les astronomes, les géodésiens, les artilleurs dans l'exercice de leur métier ont besoin de répondre à ce type de questions : *quel milieu choisir pour décrire au mieux une série statistique ? Comment résumer une diversité ? Comment faire de l'un à partir du multiple ?*

LEGENDRE, en 1805, essaye de répondre à celle-ci : *comment combiner les observations effectuées dans des conditions différentes, afin d'obtenir les meilleures estimations possibles de plusieurs grandeurs astronomiques ou terrestres, liées entre elles par une relation linéaire ?* Par exemple, un angle est mesuré un grand nombre de fois avec des instruments différents et dans des circonstances différentes ; comment alors utiliser au mieux cette abondance de mesures, compte-tenu qu'elles ne vérifient jamais parfaitement la relation théorique géométrique (par exemple, dans un triangle la somme des angles est égale à 180°) mais laisse subsister un léger écart (ou erreur ou résidu) ? Il faut combiner de façon optimale ces équations de conditions pour parvenir à une estimation des grandeurs cherchées. LEGENDRE propose de **minimiser la somme des carrés des écarts.** Cette méthode connue sous le nom de **méthode des moindres carrés** est introduite à la même époque par GAUSS. Une querelle de priorité a lieu, la publication de GAUSS étant postérieure à celle de LEGENDRE alors qu'il a utilisé cette méthode avant LEGENDRE ! (voir annexe 1). Cette méthode va se répandre rapidement (voir document 13).

La synthèse de LAPLACE-GAUSS résulte de deux traditions : la tradition décrite ci-dessus et celle des mathématiciens philosophes probabilistes ; ces derniers interrogent *le mieux possible* qui devient *comment estimer le degré de confiance que mérite une estimation ?*
Pour formuler la loi de probabilité d'une estimation statistique, il faut d'abord convenir de telles lois pour les erreurs élémentaires de chaque observation, puis les combiner mathématiquement pour en déduire une loi pour la statistique calculée (voir document 14).

Citons quelques exemples :	
Simpson 1757	$-a\lvert x\rvert + b$
Laplace 1774	$\frac{m}{2} e^{-m\lvert x\rvert}$
Laplace 1777	$\frac{1}{2a} \log \frac{a}{\lvert x\rvert}$
Gauss 1810	e^{-x^2}

La forme gaussienne e^{-x^2} s'impose au moins pour deux raisons : bon accord avec les observations et bonnes propriétés mathématiques.

Toute la statistique du XIX^{ème} siècle va vivre sur cette loi, loi dite normale, à partir de QUETELET et de son homme moyen.

7.2. Adolphe QUETELET (1796-1874)

La moyenne permet d'estimer un objet existant (ex : mesure d'un angle en géodésie) ou comme chez QUETELET, de créer des entités nouvelles.

> Adolphe BERTILLON (1876) distingue 3 moyennes :
>
> 1) objective : un objet unique soumis à plusieurs mesures ;
>
> 2) subjective : tendance centrale si la distribution peut s'ajuster à une loi binomiale ou normale ;
>
> 3) arithmétique : pure fiction si loi non normale.

Adolphe QUETELET (1796-1874) est astronome de formation. Il est donc un habitué des méthodes permettant d'estimer au mieux la mesure d'un objet. Il utilise les statistiques en démographie et le calcul des probabilités dans les sciences humaines. Il étudie la diversité de caractères physiques facilement mesurables (voir documents 15 et 16 : exemples de la mesure de la taille des conscrits français et celle du tour de poitrine des militaires écossais) et cherche une unité sous cette diversité.

Comment concilier la diversité des traits physiques et les régularités observées ? Deux idées fortes guident QUETELET : la régularité et les formes de distribution. Il construit graphiquement la distribution des fréquences des tailles observées (ici, on mesure une fois plusieurs objets) ; elle ressemble à la distribution des résultats d'une série de mesures, entachées d'erreurs, pour une grandeur donnée soit une « courbe en cloche » (voir document 16 : là, on mesure plusieurs fois le même objet)[16]. La combinaison entre la régularité de la moyenne et la forme normale de la distribution lui permet d'inférer une cause constante. On assiste ici au premier ajustement à un modèle au sens contemporain : ressemblance entre la distribution de la taille des conscrits et celle des mesures imparfaites d'une grandeur unique. Ainsi, dans les années 1830, QUETELET rapproche les régularités observées par les bureaux de

[16] La distribution des tailles des soldats du Doubs n'est pas d'allure gaussienne et unimodale. Elle présente étrangement deux bosses (distribution bimodale). Adolphe BERTILLON (1876) en déduit que cette forme bizarre résulte de la superposition de deux distributions normales, et que la population du Doubs est un mélange de deux ethnies : les Burgades et les Celtes. Alain DESROSIERES (1993) suggère plutôt que la tranche correspondant au deuxième maximum est celle qui permet d'accéder à la cavalerie et au génie, et d'échapper ainsi à l'infanterie.

statistiques administratives[17] et les régularités que présentent les mesures astronomiques (n'oublions pas que QUETELET est directeur de l'Observatoire de Bruxelles).

QUETELET infère que ces écarts par rapport à la valeur centrale sont de même nature !

Il en déduit des lois statistiques autonomes par rapport aux individus. Il suppose l'existence, au-dessus des cas particuliers contingents, d'**un homme moyen idéal**, dont ces cas sont des copies imparfaites. Il utilise l'histoire métaphorique du roi de Prusse qui, très admiratif d'une statue parfaite d'Apollon, en commande 1000 copies à 1000 sculpteurs du royaume. Les copies ne sont pas parfaites et se distribuent avec des écarts par rapport au modèle, de la même manière que les individus concrets se distribuent autour de l'homme moyen. Le Créateur est alors l'équivalent du roi de Prusse, et les individus sont des réalisations imparfaites du modèle parfait : « *Les choses se passent comme si la cause créatrice de l'homme, ayant formé le modèle du type humain, eût ensuite, en artiste jaloux, brisé son modèle, laissant à des artistes inférieurs le soin des reproductions* ».

QUETELET étend cette notion aux comportements moraux (crimes, suicides...). L'homme moyen de QUETELET réunissant toutes les caractéristiques moyennes des mesures effectuées (bras et jambes, crânes, poids, taille, tour de poitrine...) a, bien sûr, fait l'objet de nombreuses critiques, par exemple :

* COURNOT en 1843 dans *Exposition de la théorie des chances et des probabilités* indique qu'un homme dont les attributs corporels seraient les moyennes de ces mêmes attributs mesurés pour beaucoup *d'individus « bien loin d'être en quelque sorte le type de l'espèce, serait tout simplement un homme impossible ou du moins rien n'autorise jusqu'ici à le concevoir comme possible »*. Il souligne de plus que les relations existant entre les variables ne sont pas prises en compte ; elles ne sont pas indépendantes.
* BERTRAND en 1889 dans *Calcul des probabilités* prend l'exemple d'une sphère moyenne obtenue à partir du rayon moyen, de la surface moyenne et du volume moyen : « *Aucune concession n'est possible, nulle sphère n'est difforme. Un homme malheureusement peut l'être et Mr QUETELET en profite* ».

Quelle naïveté ! L'homme moyen est paré de toutes les vertus. Mais sa perfection réside dans la modération ; il évite tous les excès. L'accent est mis sur la recherche d'une moyenne, idéal centriste. Un débat de philosophie politique se met en place mettant en lumière une tension existant entre un fatalisme collectif et la liberté individuelle[18], débat que l'on retrouve en médecine.

---

[17] En Belgique, QUETELET organise des recensements ; il organise en 1853, le premier congrès international de statistiques, dans le but de produire un grand nombre de données.

[18] Cf. J. BOUVERESSE *L'homme probable Robert Musil, le hasard, la moyenne et l'escargot de l'histoire.* L'éclat, 1993.

7.3. Statistique médicale

Il est difficile, en médecine, de penser en même temps les situations singulières et les régularités statistiques. Depuis longtemps, la maladie et son traitement par un médecin constituent un évènement unique et singulier. Dans les milieux médicaux, deux positions différentes s'affrontent : le docteur LOUIS est d'accord pour des méthodes numériques (la statistique) tandis que AMADOR considère la médecine comme un art fondé sur l'intuition et l'instinct du praticien (la tradition). Mais, face aux épidémies et sa prévention, le médecin va peu à peu céder car alors le problème collectif appelle des solutions globales. La maladie devient alors un TOUT dont on recherche les causes constantes c'est-à-dire les facteurs favorisant sa diffusion ; il intervient alors des calculs de taux moyen, de mortalité de population classée suivant divers critère. Le corps médical accepte cette apparition des nombres dans l'exercice de leur métier car cela leur permet d'intervenir dans des débats publics et dans l'organisation de la société, par contre le corps médical refuse les techniques quantitatives en matière de clinique médicale, ne voulant pas soumettre à la statistique la comparaison des diverses méthodes pour soigner une maladie.

Crée en 1829, La *Revue d'hygiène publique et de médecine légale* servira de support pour des enquêtes portant sur les groupes sociaux les plus exposés à la misère, l'alcoolisme... En 1832, VILLERME (1782-1863) hygiéniste, ami de QUETELET, écrit un article sur l'épidémie de choléra. Les médecins sont partagés entre deux explications : celle des contagionistes affirmant que la maladie est transmise par le contact avec les malades et celle des infectionnistes pensant que la maladie n'est pas contagieuse mais que sa diffusion est favorisée par l'insalubrité des logements. Ces derniers s'appuient sur des statistiques de taux de mortalité obtenues suivant les rues de Paris et le niveau économique des habitants. Ils veulent intervenir sur le plan politique afin d'obtenir une amélioration des conditions d'hygiène. La découverte en 1883 du vibrion du choléra donnera raison aux deux groupes de médecins puisque sa diffusion est favorisée par l'absence d'égouts.

Claude BERNARD, fondateur de la médecine expérimentale, fait progresser la méthode scientifique et pourtant, il est hostile à la méthode numérique ; le médecin ne peut pas, dit-il, soigner « en moyenne ». Il faut d'après lui, analyser complètement, sans incertitude ; ce qui ne peut pas être réalisé par des méthodes statistiques synonymes pour lui de manque de rigueur et d'à peu-près.

Avec QUETELET, l'universalisme triomphe ; avec BERNARD, l'individualisme est à l'honneur. Deux outils nouveaux vont permettre de penser les deux : la régression et la corrélation. Ils sont l'invention, à partir du milieu du XIX^{ème} siècle, d'une florissante école anglaise. Cinq anglais sont fondateurs de la statistique mathématique : GALTON (médecin), EDGEWORTH (économiste), WELDON (biologiste), PEARSON (mathématicien, philosophe, politique), YULE (ingénieur, physicien). Ils fabriquent un nouveau langage, façonné collectivement mais

articulé avec d'autres langages car ils ont des formations, des engagements intellectuels et politiques différents.

7.4. Francis GALTON (1822-1911) et Karl PEARSON (1857-1936)

Ils vont réunir (voir document 2) diverses approches : les questions d'hérédité étudiées par DARWIN, les distributions normales des attributs de l'espèce humaine mises en place par QUETELET et les techniques d'ajustement utilisées dans la théorie des erreurs de mesure.

Leur projet diffère cependant de celui de QUETELET ; ce dernier s'intéresse à l'homme moyen et non à la distribution relative des hommes non moyens alors que GALTON et PEARSON portent leur attention sur les différences entre les individus, sur **la variabilité de leurs attributs** (aptitudes naturelles). Cependant, GALTON conserve l'idée de la distribution normale des attributs humains autour d'une valeur moyenne. Sous l'influence de son cousin germain DARWIN, il abandonne la médecine qui l'intéresse peu et préfère se consacrer à l'anthropologie. Il se tourne vers la statistique, après avoir eu connaissance des travaux de QUETELET, et se pose des questions sur l'hérédité biologique. Il veut transposer et appliquer la théorie de l'évolution à l'espèce humaine, dans une perspective d'amélioration de celle-ci : l'eugénisme est en germe.

GALTON utilise la loi normale comme une loi de déviation permettant de classer les individus et non comme une loi des erreurs. Les faits pertinents sont **les écarts par rapport à la moyenne** qui ne sont plus des parasites à éliminer comme chez les astronomes ou les géodésiens ; d'où la prédominance de notions tels que la médiane et les quartiles pour ordonner les données. L'erreur probable de QUETELET devient la déviation standard puis l'écart type ; le mot variance apparaît.

Il met en place un outil nouveau : **la régression**. En 1877, il étudie la relation existant entre la taille des enfants, prise en écart par rapport à la moyenne de leur génération, et celle des parents et constate que l'écart à la moyenne diminue, c'est-à-dire régresse. GALTON veut comprendre ce phénomène, mais il n'est pas mathématicien. Par contre, il possède une grande imagination expérimentale. Il fait donc construire une machine la « quinconce » en 1873, connue sous le nom de planche de Galton, et réalise une expérience sur la récolte des petits pois avec des semences étalonnées (voir document 17 et 18). Il met ainsi en évidence comment deux dispersions successives peuvent se combiner. Il introduit alors une mesure numérique de cette régression : notre coefficient de régression. Il constate que les coefficients de régression de x en y ou de y en x sont les mêmes si les variables sont normalisés en termes de variabilité ; il appelle alors ce coefficient, coefficient de co-relation ou de corrélation et le note r comme l'initiale de régression. GALTON utilise dans son calcul non l'écart type mais l'écart interquartile. Son rôle dans le développement des divers quantiles est important.

Profitant de l'exposition internationale de la santé à Londres en 1885, il mesure les tailles des parents et des enfants des diverses familles se présentant au stand qu'il a installé, afin de poursuivre ses recherches. Il constate là aussi un effet de régression vers la moyenne des tailles enfants par rapport à celles des parents ; pour l'interprétation, il fait appel au mathématicien Hamilton DICKSON qui reconnaît aussitôt une loi normale à deux dimensions.[19] Le français BRAVAIS avait déjà, en 1846, identifié cette loi probabiliste dans le cadre de la théorie des erreurs de mesures mais n'en avait pas tiré de conclusion pertinente quant aux variations conjointes de deux variables.

Ces avancées conceptuelles sont essentielles, mais très peu formalisées ; ainsi, la pente de la droite de régression est déterminée graphiquement, sans aucune optimisation du type méthode des moindres carrés.

A la même époque, Adolphe BERTILLON (1853-1914), expert à la préfecture de police de Paris veut identifier d'une façon unique les personnes afin de découvrir les délinquants et les criminels. Il mesure diverses parties d'un même corps humain et essaye d'établir des relations entre les données obtenues sans pour autant s'interroger sur d'éventuelles dépendances entre ces grandeurs. Les calculs de corrélation permettent d'apporter des réponses à cette question[20].

8. Statistique au XX^{ème} siècle

8.1. K. PEARSON (1857-1936)

Sa contribution en statistique est importante et très variée. Il met en place un laboratoire de biométrie très actif où sont façonnés les outils mathématiques de la statistique moderne et participe à la fixation de certains termes et définitions.

Dans la première édition, en 1892, de *La grammaire de la Science,* aucune allusion à la statistique n'est faite. Par contre, dans la troisième édition de 1911, PEARSON utilise la corrélation et la décrit en ces termes : « une opération qui fait époque dans l'histoire des idées ». Entre les deux éditions, il a rencontré GALTON et ses questions sur l'hérédité biologique[21]. Ayant reçu une formation mathématique au King's Collège de Cambridge, son regard sur les travaux de GALTON s'en nourrit. Il met en forme mathématique la régression et la corrélation pour étayer la construction eugéniste de GALTON et il étudie des généralisations du coefficient de corrélation r. Il applique les statistiques aux problèmes biologiques de l'hérédité et de l'évolution.

---

[19] Des lectures différentes seront faites de l'œuvre de GALTON, par exemple par CHEYSSON et par DURKHEIM.

[20] GALTON propose une solution pour identifier les personnes : utiliser les empreintes digitales.

[21] Une traduction française de la troisième édition de cette œuvre sera faite en 1912 par Lucien MARCH, directeur de la Statistique générale de la France, ancêtre de l'INSEE.

En 1892, il s'attaque aux mesures de crabes et crevettes effectuées par WELDON. Il en examine les distributions et leurs normalités. Observant des dissymétries, il essaye alors de les décomposer en deux ou plusieurs lois normales. Il cherche une représentation mathématique s'ajustant le mieux possible aux observations en indiquant des moments d'ordre successifs pour estimer des paramètres « méthode des moments ». Il semble donc comme QUETELET chercher une ou plutôt plusieurs causes constantes. Une discussion autour du réalisme de cette modélisation a lieu. Comment tenir compte de ce que l'on sait déjà par ailleurs ? Quels critères retenir pour éliminer d'éventuelles observations aberrantes ?

De 1893 à 1912, il écrit 18 articles sous le titre *Mathematical Contribution to the Theory of Evolution.* Certains de ses travaux sont consacrés à la dispersion ; il introduit le terme de standard déviation en 1893, c'est-à-dire l'écart type. Il mesure la variance des estimateurs pour en déterminer la précision et met en place le test du chi-deux en 1900. Une longue dispute l'opposera à FISHER au sujet de la taille des échantillons. Il travaille à partir de grands échantillons alors que ce dernier en utilise des petits et se permet malgré tout d'en déduire des corrélations pour trouver des causes.

Il est co-fondateur avec WELDON et GALTON puis éditeur de la revue *Biometrika* datant de 1901 et construit des tables de statistiques pour donner un outil de travail aux utilisateurs de cette discipline, favorisant ainsi sa diffusion et son utilisation dans divers domaines :

a) Maths et biologie donnent la biométrie.
b) Maths et économie donnent économétrie.
c) Maths et psychologie donnent la psychométrie.

8.2. Yule UDNY (1871-1951)

Après des études d'ingénieur à Londres et un début de carrière de physicien en Allemagne, il devient élève de K. PEARSON puis son assistant à partir de 1893. Il s'attaque en 1895 au problème de la pauvreté et de l'assistance. Pour aider ceux qui réforment la loi sur les pauvres (datant de 1834), il essaye de répondre aux questions suivantes : « Quelle part de la baisse du paupérisme est due aux changements administratifs de la gestion de l'assistance et quelle part à d'autres causes ? », « L'ampleur de l'assistance à domicile ne contribue-t-elle pas à entretenir, sinon à accroître la pauvreté ? » Un nouvel objet est crée : « le paupérisme ». La première étude « économétrique » est réalisée, mais de nombreuses questions ne sont que pressenties ; non explicitées, elles restent sans réponse à l'époque : comment définir ce nouvel objet ? Quelle est sa signification ? Est-il mesurable ? Ne dépend-il pas d'une codification sociale ? Dans l'article *On the Theory of Correlation* de 1897, il utilise l'ajustement par la méthode des moindres carrés, dans un contexte autre que la théorie des erreurs, pour définir la droite de régression comme celle qui minimise la somme des carrés des écarts des observations à la droite ajustée.

Sa contribution dans l'étude des séries chronologiques est importante de 1920 à 1930 : il introduit le corrélogramme et met en place le coefficient d'autocorrélation permettant de mesurer la corrélation entre des observations distantes d'un intervalle de temps fixé a priori.

Par son enseignement, par les discussions qu'il a avec ses collègues et par ses articles, il contribue grandement au développement de la statistique. Son livre *Introduction to the Theory of Statistics* est publié pour la première fois en 1911. Il connaîtra 14 éditions jusqu'en 1950 ; la dernière est écrite en collaboration avec Maurice KENDALL.

8.3. Gosset William SEALY (1876 - 1937)

Après avoir obtenu des diplômes de mathématiques et de sciences naturelles à Oxford, il est engagé en 1899 dans la firme de bière Guiness, à Dublin. En 1930, il met en place dans cette brasserie, un contrôle de fabrication à partir d'un petit nombre de prélèvements ; les observations sont donc effectuées sur des petits échantillons et non vérifiables sur la population totale. Il est amené de ce fait à estimer les variances et les lois de distributions de paramètres calculées sur des observations ne satisfaisant évidemment pas la « loi des grands nombres ».
Il est à l'origine, en 1908, de « la loi de Student », Student étant le nom sous lequel il publie ses articles. Deux objets apparaissent : l'estimation des paramètres et les tests d'hypothèses. Le premier est un point de vue subjectif car il suppose le modèle vrai ; les informations issues du modèle et des données sont combinées mais rien ne dit s'il y a accord. Le deuxième s'appuie sur l'idée fréquentiste ; on éprouve l'accord entre les données et les informations issues du modèle quitte à changer de modèle.

8.4. Fisher Ronald AYLMER (1890-1962)

Il se dirige vers les mathématiques, s'intéressant aussi à la biométrie et la génétique. Il connaît les travaux de MENDEL (1865), ceux de GALTON et de PEARSON. Il étudie la théorie des erreurs dans un manuel de AIRY et l'intérêt qu'il y porte lors d'observations astronomiques, le conduit à des problèmes statistiques.

Il est un des pionniers de la théorie de l'estimation. Cette loi est apparue dans un contexte de statistique expérimentale. En effet, FISHER travaille de 1919 à 1933 dans un manoir anglais Rothamsted Agricultural Experiment Station, laboratoire expérimentale d'agriculture, où il obtient des résultats importants ; il y développe entre autre l'analyse de la variance. Il introduit, en 1921, la notion de vraisemblance d'un paramètre et inaugure des tests d'hypothèses à partir de simulation. Comme GOSSET, il travaille sur des petits échantillons, ne pouvant procéder dans son laboratoire qu'à un nombre restreint d'essais contrôlés. Il invente alors une technique de « randomisation » introduisant ainsi l'aléa probabiliste dans la démarche expérimentale. Avec GOSSET, il est obligé d'introduire des notations distinctes pour

désigner, d'une part, un paramètre théorique, celui d'une distribution probabiliste, et d'autre part, l'estimation de ce paramètre, calculée à partir d'observations en nombre limité. Il n'est pas possible en effet de négliger l'écart entre la valeur théorique et la valeur estimée. Cette démarche marque un tournant décisif, rendant possible une statistique inférentielle appuyée sur des modèles probabilistes.

Avant de conclure, signalons que les enquêtes par sondages probabilistes n'apparaissent que vers la fin du XIX^{ème} siècle avec le norvégien KIAER. Les premiers calculs et les intervalles de confiance datent de 1906 avec l'anglais BAWLEY et les méthodes de stratification de 1934 avec NEYMANN. Les idées de FISHER sont diffusées pour la première fois en France en Mai-Juin 1929 lors d'une conférence à l'Institut Henri Poincaré intitulée « Méthode statistique dans les sciences expérimentales ». L'année suivante, les chaînes de MARKOV sont introduites à Paris. Mais l'hiver 1937-38, la présentation des travaux de KOLMOGOROV à Paris, reçoit un très bon accueil des milieux universitaires français, le bourbakisme étant un bon terrain de réception à cette nouvelle théorie.

Jusqu'en 1958, une thèse de mathématiques n'était pas reconnue si elle ne portait que sur des probabilités et a fortiori sur des statistiques.

On comprend mieux ainsi les difficultés que l'on rencontre pour l'introduction d'un tel enseignement dans les classes du secondaire en France.

9. Bibliographie[1]

Sources primaires

BERNARD C., 1865, *Introduction à l'étude de la médecine expérimentale,* Paris, Baillère.

BERTILLON L.A., 1863, *De la méthode dans l'anthropologie à propos de l'influence des milieux sur la coloration des téguments,* in Bulletin de la société d'anthropologie de Paris, 4, pp.223-242.

BERTILLON L.A., 1876, *La théorie des moyennes statistiques,* in Journal de la société statistique de Paris, pp.265-308.

BERTRAND J., *Calcul des probabilités,* 2nd édition, 1907.

BRAVAIS A., 1846, *Analyse mathématique sur les probabilités des erreurs de situation d'un point,* in Mémoires présentés par divers savants à L'Académie Royale des Sciences de l'Institut de France, 9, pp.255-332.

[1] Nous n'indiquons dans cette bibliographie que des ouvrages relatifs à la statistique. De ce fait, ne figurent pas d'ouvrages de probabilités.

COURNOT A., *Exposition de la théorie des chances et des probabilités,* Hachatte, Paris, 1843 ; réed. en 1984, in Œuvres Complètes de Cournot, Vol.1, B. Bru, Vrin, Paris.

DICKSON, HAMILTON J.D., 1886, *Appendix,* (to Galton, 1886) in Proceedings of the Royal Society of London, 40, pp.63-66.

EDGEWORTH F., 1881, *Mathematical Psychics. An essay on the application of mathematics to the Moral Sciences,* C.Kegan Paul, Londres.

EDGEWORTH F., 1885, *Calculus of Probabilities Applied to Psychical Research, I,* in Proceedings of the Society for Psychical Research 3, pp.190-208.

EDGEWORTH F., 1892, *The law of Error and Correlated Averages,* in Philosophical Magazine, 5° séries, 34, pp.429-438 et pp.518-526.

FISHER R., 1922, *On the mathematical foundations of theorical statistics,* in Philosophical Transactions of the Royal Society of London(A), 222:309-368.

FISHER R., 1925, *Statistical Methods for research workers,* Oliver and Boyd, Edinburgh.

FISHER R., 1959, *Statistical Methods and Scientific Inference,* 2nd ed. Edinburg: Oliver and Boyd.

GALTON F., 1869, *Hereditary Genius An Inquiry into its Laws and Consequences,* Macmillan, Londres.

GALTON F., 1885, *Regression Towards Mediocrity in Hereditary Stature,* in Journal of the Anthropological Institute, 15, pp.246-263.

GALTON F., 1889, *Natural Inheritance,* Macmillan, Londres.

GALTON F., 1909, *Essays in Eugenics,* Eugenics Education Society, Londres.

GAUSS C.F., 1855, *Méthodes des moindres carrés.* Mémoire sur la combinaison des observations, traduction de J. Bertrand, Paris.

NEYMANN J., PEARSON E., 1928, *On the Use and Interpretation of certain Test Criteria for Purposes of statistical Inference,* in Biometrika, 20A, pp.175-240 et pp.263-294.

PEARSON K., 1896, *Mathematical Contributions to the Theory of Evolution, III: Regression, heredity and panmixia,* in Philosophical Transactions of the Royal Society of London (A), 187, pp.253-318.

PEARSON K., 1912, *La grammaire de la Science,* traduit par Lucien Mach, Alcan, Paris. 1ère édition 1892.

PEARSON K., 1920, *Notes of the History of Correlation,* in Biometrika, 12, pp.25-42.

PEARSON K., HERON D., 1913, *On Theories of Association,* in Biometrika, 9, pp.159-315.

QUETELET, A., 1835, *Sur l'homme et le développement de ses facultés ou Essai de physique sociale,* 2 vol., Paris, Bachelier.

QUETELET A., 1832, *Sur la possibilité de mesurer l'influence des causes qui modifient les éléments sociaux,* in Correspondance mathématique et publique, 7, pp.321-346.

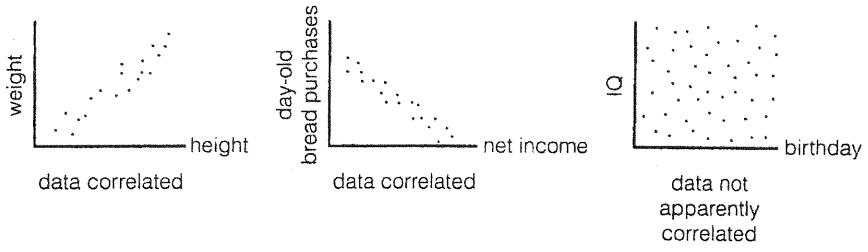
TUKEY, P.A., 1977, *Exploratory Data Analysis,* Reading, Addison-Wesley.

TUKEY, P.A., 1983, *Graphical Methods for Data Analys,* Pacific Grove, Wadsworth and Brooks Cole Publi. Company.

VILLERME L.R., 1845, *Sur l'institution par le gouvernement belge d'une commission centrale de statistique, et observations sur les statistiques officielles publiées en France par les divers ministères*, in Journal des économistes, tome XV, pp.130-148.
VILLERME L.R., 1829, *Mémoire sur la taille de l'homme en France,* in Annales d'hygiène publique et de médecine légale 1, pp.351-399.
YULE U., 1895, *On the correlation of total pauperism with proportion of out-relief, I, all ages,* in Economic Journal 5, pp.603-611.
YULE U., 1897,*On the Theory of Correlation,* in Journal of the Royal Statistical Society, 60, pp.812-854.

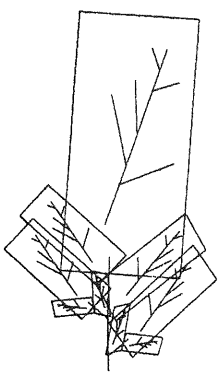
Sources secondaires

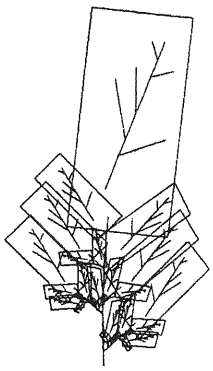
BENZECRI J.-P., 1982, *Histoire et préhistoire de l'analyse des données,* Dunod, Paris.
DESROSIERES A., 1993, *La politique des grands nombres,* Paris, Editions la découverte.
DROESBEKE J.-J., TASSI P., 1990, *Histoire de la statistique,* PUF, « Que sais-je ? » n°2527, Paris.
DUPAQUIER J., 1985, *Histoire de la démographie,* Perrin, Paris.
HALBWACHS H., 1913, *La Théorie de l'homme moyen. Essai sur Quetelet et la statistique morale,* Alcan, Paris.
INSEE, 1977, *Pour une histoire de la statistique, tome 1et 2,* Economica, Paris, réédité en 1987.
KENDALL M., PLACKETT R.L., 1977, *Studies in the History of Statistics and Probability,* Griffin, Londres.
LAZARSFELD P., 1970, *Notes sur l'histoire de la quantification en sociologie,* in Philosophie des sciences sociales, Gallimard, Paris, pp.75-162.
MAIRESSE J., (éd), 1988, *Estimations et sondages. Cinq contributions à l'histoire de la statistique,* Economica, Paris.
PORTER T.M., 1986, *The Rise of Statistical Thinking,* Princeton University Press, Princeton.
STIGLER S.M., 1986, *The history of statistics,* Belknapp Press of Harvard University Press, Cambridge.



DOCUMENTS

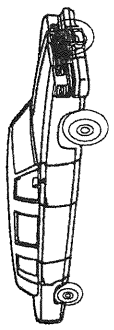
(En relation avec l'article de M.-F. JOZEAU)

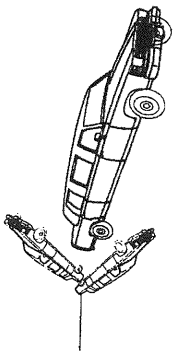


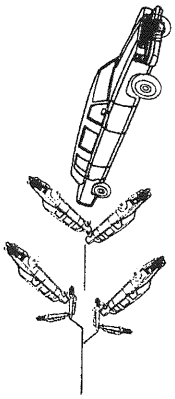




Barnsley's Fern

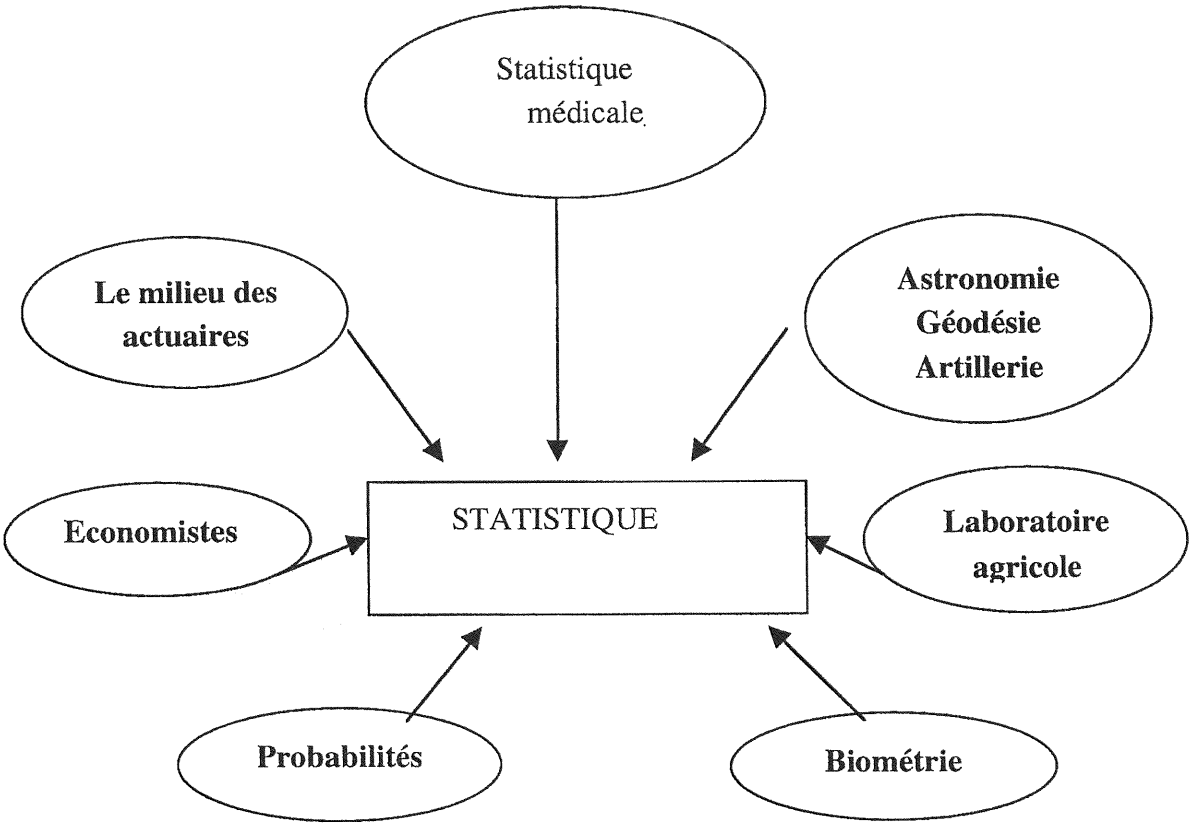






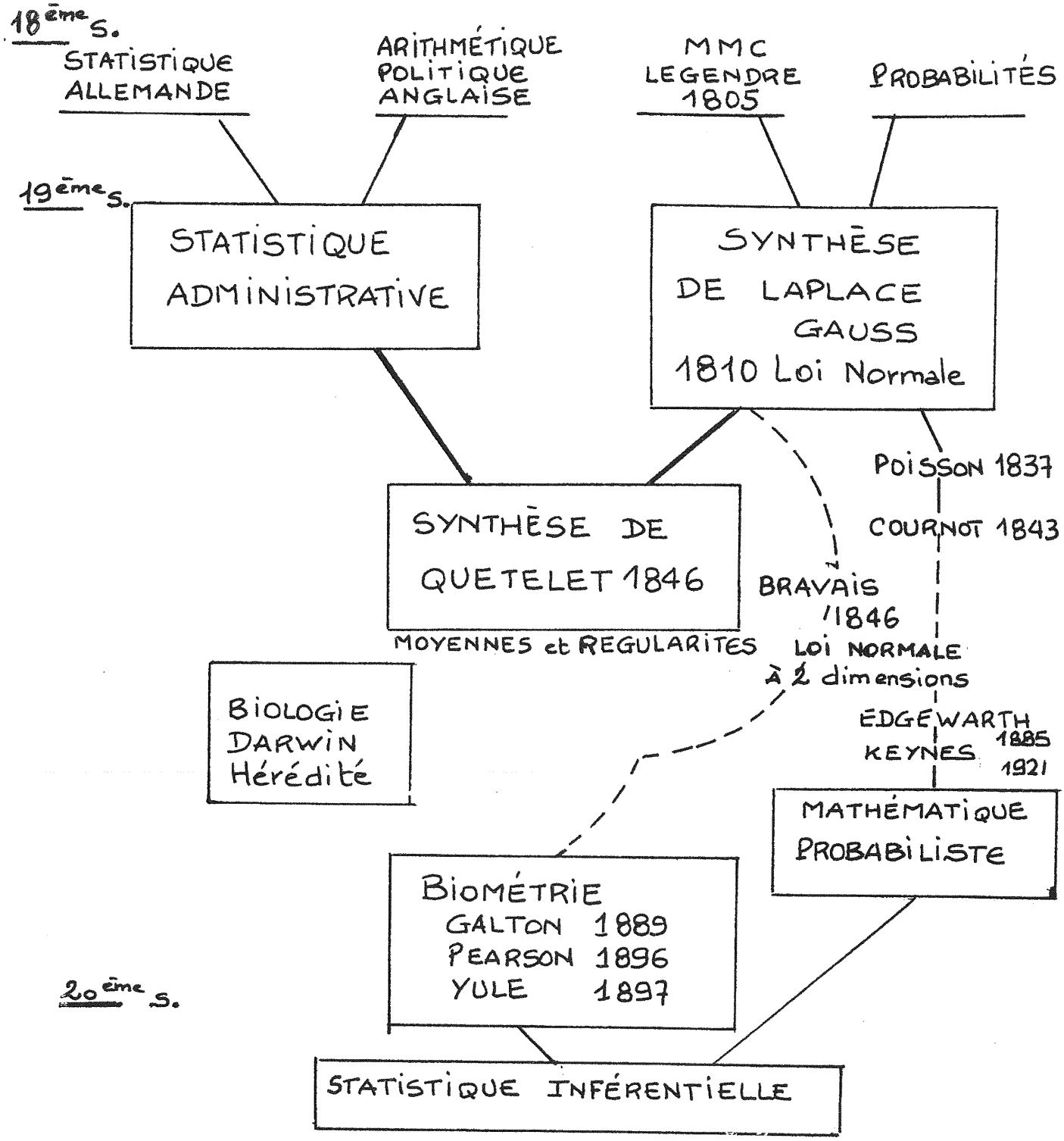


Document 1


Statistique médicale
Le milieu des actuaires
Astronomie
Géodésie
Artillerie
Economistes
STATISTIQUE
Laboratoire agricole
Probabilités
Biométrie

Document 2

Généalogie des Statistiques Inférentielles



1925 R. FISHER : PROBLÈME D'ESTIMATION
1928 E. PEARSON : LES TESTS D'HYPOTHÈSE
1928 J. NEYMANN } THÉORIE MODERNE
E. PEARSON } DE L'ÉCHANTILLONAGE

Pour terminer ce parallele entre les Problêmes ſur les Jeux, & les queſtions qu'on peut propoſer ſur les choſes œconomiques, politiques & morales, il faut obſerver que dans ces dernieres comme en celles des Jeux, il y a une eſpece de Problêmes qu'on pourra reſoudre en obſervant ces deux regles; 1°. borner la queſtion que l'on ſe propoſe à un petit nombre de ſuppoſitions, établies ſur des faits certains; 2°. faire abſtraction de toutes les circonſtances auſquelles la liberté de l'homme, cet écueil perpetuel de nos connoiſſances, pourroit avoir quelque part. Il eſt à croire que M. Bernoulli avoit égard à ces regles dans la quatriéme partie de ſon Ouvrage, & il eſt certain qu'avec ces deux reſtrictions on pourroit traiter pluſieurs ſujets ou de politique ou de morale avec toute l'exactitude des verités geometriques.

C'eſt ce qu'a fait admirablement M. Halley dans un Memoire qui ſe trouve dans les Tranſactions philoſophiques d'Angleterre, num. 196, où ce ſçavant Anglois entreprend de déterminer le degré de la mortalité du genre humain. Ce morceau eſt plein de choſes curieuſes, dont le Lecteur verroit ici l'extrait avec plaiſir : Mais cette Préface étant peut-être déja trop longue, je n'en rapporterai qu'une qui eſt traitée par l'Auteur avec beaucoup de fineſſe. C'eſt une methode pour déterminer ſur quel pied ſe doivent regler les rentes à fond perdu. Il donne une Table toute calculée pour les differens âges de cinq en cinq années depuis un an juſqu'à ſoixante & dix. Cette Table fait voir combien étoit avantageux aux Anglois le parti que leur faiſoit alors le Roy Guillaume, en donnant 14 pour cent par année de rente viagere, ce qui eſt à peu près la ſeptiéme partie du fond. On voit par cette Table qu'une perſonne âgée de dix ans n'en devroit avoir que la treiziéme partie, & un homme âgé de trente-ſix ans la onziéme, & enfin que l'interêt de dix pour cent n'eſt dû qu'aux perſonnes âgées de quarante-trois à quarante-quatre ans. Il pouſſe encore cette idée plus loin, & il examine ſur quel pied devroit ſe regler une rente viagere qui ſeroit ſur la tête de deux ou de pluſieurs perſonnes de differens âges.

An Eſtimate of the Degrees of the Mortality *of Mankind, drawn from curious* Tables *of the* Births *and* Funerals *at the City of* Breſlaw; *with an Attempt to aſcertain the Price of* Annuities *upon* Lives. *By* Mr. E. Halley, *R.S.S.*

THE Contemplation of the *Mortality* of *Mankind*, has beſides the *Moral*, its *Phyſical* and *Political* Uſes, both which have been ſome years ſince moſt judiciouſly conſidered by the curious Sir *William Petty*, in his *Natural* and *Political* Obſervations on the Bills of *Mortality* of *London*, owned by Captain *John Graunt*. And ſince in a like Treatiſe on the Bills of *Mortality* of *Dublin*.

[---]

From theſe Conſiderations I have formed the *adjoyned Table*, whoſe Uſes are manifold, and give a more juſt *Idea* of the *State* and *Condition* of *Mankind*, than any thing yet extant that I know of. It exhibits the *Number* of *People* in the City of *Breſlaw* of all Ages, from the *Birth* to extream *Old Age*, and thereby ſhews the Chances of *Mortality* at all *Ages*, and likewiſe how to make a certain Eſtimate of the value of *Annuities* for *Lives*, which hitherto has been only done by an imaginary *Valuation*: Alſo the *Chances* that there are that a *Perſon* of any *Age* propoſed does live to any other *Age* given; with many more, as I ſhall hereafter ſhew. This *Table* does ſhew the *number* of *Perſons* that are living in the *Age* current annexed thereto, as follows:

Age. Curr.	Perſons.	Age. Curr.	Perſons	Age. Curr.	Perſons	Age. Curr.	Perſons	Age. Curr.	Perſons	Age. Curr.	Perſons
1	1000	8	680	15	628	22	586	29	539	36	481
2	855	9	670	16	622	23	579	30	531	37	472
3	798	10	661	17	616	24	573	31	523	38	463
4	760	11	653	18	610	25	567	32	515	39	454
5	732	12	646	19	604	26	560	33	507	40	445
6	710	13	640	20	598	27	553	34	499	41	436
7	692	14	634	21	592	28	546	35	490	42	427
Age Curr	Perſons.	Age. Curr.	Perſons	Age. Curr.	Perſons	Age Curr	Perſons	Age. Curr.	Perſons	Age. Curr.	Perſons
43	417	50	346	57	272	64	202	71	131	78	58
44	407	51	335	58	262	65	192	72	120	79	49
45	397	52	324	59	252	66	182	73	109	80	41
46	387	53	313	60	242	67	172	74	98	81	34
47	377	54	302	61	232	68	162	75	88	82	28
48	367	55	292	62	222	69	152	76	78	83	23
49	357	56	282	63	212	70	142	77	68	84	20

Age.	Perſons.
7	5547
14	4584
21	4270
28	3964
35	3604
42	3178
49	2709
56	2194
63	1694
70	1204
77	692
84	253
100	107
	34000
	Sum Total.

Thus it appears, that the whole People of *Breſlaw* does conſiſt of 34000 *Souls*, being the Sum *Total* of the Perſons of all Ages in the *Table*.

Document 4bis

Edmud HALLEY (1656 - 1742)

Classes d'âge	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*
1	1 000	8	680	15	628	22	586	29	539	36	481
2	855	9	670	16	622	23	579	30	531	37	472
3	798	10	661	17	616	24	573	31	523	38	463
4	760	11	653	18	610	25	567	32	515	39	454
5	732	12	646	19	604	26	560	33	507	40	445
6	710	13	640	20	598	27	553	34	499	41	436
7	692	14	634	21	592	28	546	35	490	42	427

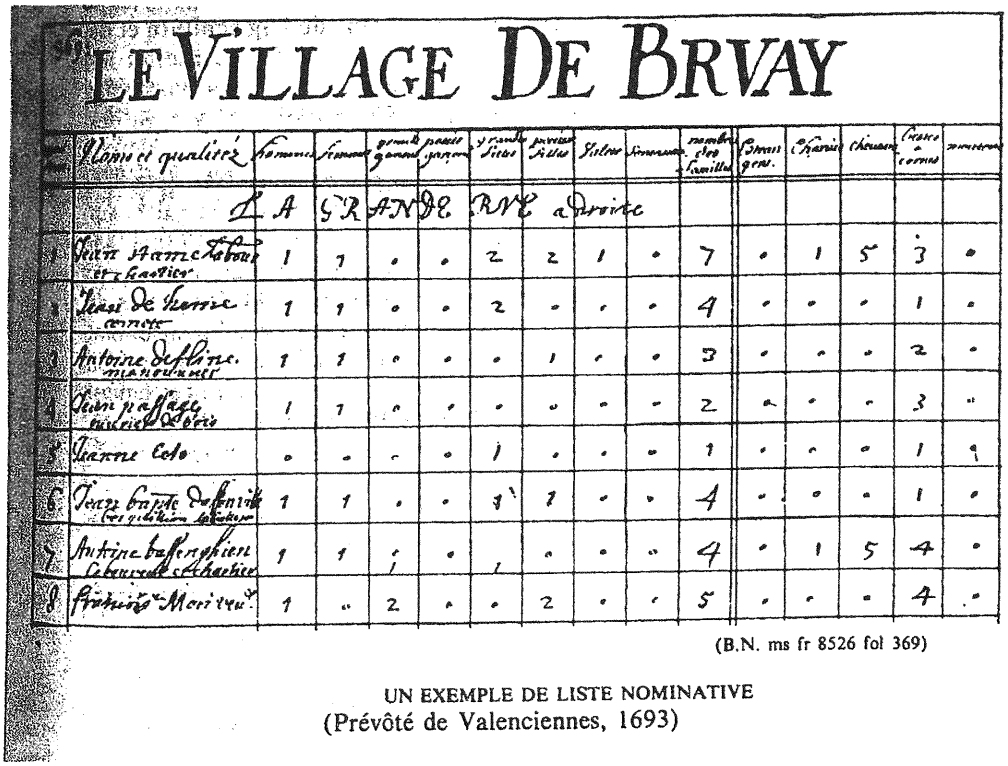
Classes d'âge	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*	*Classes d'âge*	*Pers.*
43	417	50	346	57	272	64	202	71	131	78	58
44	407	51	335	58	262	65	192	72	120	79	49
45	397	52	324	59	252	66	182	73	109	80	41
46	387	53	313	60	242	67	172	74	98	81	34
47	377	54	302	61	232	68	162	75	88	82	28
48	367	55	292	62	222	69	152	76	78	83	23
49	357	56	282	63	212	70	142	77	68	84	20

Sommes des colonnes	
Age	*Personnes*
7	5 547
14	4 584
21	4 270
28	3 964
35	3 604
42	3 178
49	2 709
56	2 194
63	1 694
70	1 204
77	692
84	253
100	107

Chastellenie de... Paroisse de S. N. Ruë Fr. ≈1686

MAISONS	NOMS ET QUALITEZ	HOMMES	FEMMES	GRANDS GARÇONS	GRANDES FILLES	PETITS GARÇONS	PETITES FILLES	SERVANTES	VALETS	NOMB. DES FAMILLES
1	Mr le Conte de... Seigneur du lieu, y résidant actuellement	1	1	2	0	0	0	6	2	12
1	Mre Philbert de Lorme, Curé dudit lieu	1	0	0	0	0	0	1	1	3
1	Mre Guillaume Vial, Vicaire	1	0	0	0	0	0	1	1	3
1	Mᵉ Jean Prevost, Scindic	1	1	2	1	3	1	1	0	10
1	Mᵉ Pierre la Plume, Procureur Fiscal	1	0	1	0	0	2	2	1	7
1	Mᵉ Matthias, Mᵉ d'Escolle	1	1	2	1	3	1	0	0	9
1	Claude de Champs, Laboureur	1	0	1	2	0	0	1	2	7
1	Pierre du Bois, Charpentier	1	0	1	0	1	0	0	0	3
1	Nicolas Marteau, Mareschal	1	1	0	0	0	2	1	1	6
1	0 0	0	0	0	0	0	0	0	0	0
1	Nicolle Lamion, veuve de Pierre de la Herse, Laboureur	0	1	1	1	1	0	0	1	5
1	Les Enfans de Thomas l'Aleine, Cordonnier	0	0	2	0	1	0	0	0	3
1	Mᵉ George Malaisé, Bourgeois rentier	1	1	0	2	1	2	1	2	10
1	Jacques Brot, Tavernier	1	1	0	2	0	2	1	1	8
1	Jean du Rivet, Savetier	1	0	2	1	1	0	0	0	5
1	Toussaint Bobine, Tisserand	1	1	1	0	0	2	0	0	5
1	George la Lime, Serrurier	1	1	1	0	0	2	1	1	7
1	Edme Fagot, Beucheron	1	0	2	1	0	0	0	0	4
1	Noël Serment, Vigneron	1	1	0	2	1	0	0	0	5
Total 18	18 Familles	16	10	18	13	12	14	16	13	112

CADRE DE RECENSEMENT PROPOSÉ PAR VAUBAN



LE VILLAGE DE BRVAY

	Noms et qualitez	Hommes	Femmes	grands garçons	petits garçons	grandes filles	petites filles	Filles	Servantes	nombre des familles	Hommes gens.	Chevaux	Chevaux à cornes	Moutons
	LA GRANDE RVE à droite													
1	Jean Hamel Laboureur et Charretier	1	1	0	0	2	2	1	0	7	0	1	5	3	0
2	Jean de Herrie Berger	1	1	0	0	2	0	0	0	4	0	0	0	1	0
3	Antoine de Flaine Manouvrier	1	1	0	0	0	1	0	0	3	0	0	0	2	0
4	Jean passage Cabaretier Boucher	1	1	0	0	0	0	0	0	2	0	0	0	3	0
5	Jeanne Coto	0	0	0	0	1	0	0	0	1	0	0	0	1	0
6	Jean Baptiste Defrenne Cordonnier Cabaretier	1	1	0	0	1	1	0	0	4	0	0	0	1	0
7	Antoine Bassenghien Laboureur et Charretier	1	1	0	0	1	0	0	0	4	0	1	5	4	0
8	Antoine Morieux	1	0	2	0	0	2	0	0	5	0	0	0	4	0

(B.N. ms fr 8526 fol 369)

UN EXEMPLE DE LISTE NOMINATIVE
(Prévôté de Valenciennes, 1693)

Tabell

über die abtey N.

welche beſitzet

Das cloſter N. worzu gehören.	Die herrlichkeit N. deren zubehör.	Das gut N.
Marckflecken N.	—	—
Dorff N.	Dorff N. diſſeits des grabens.	Untergericht zu N.
10 unterthanen zu N.	8. unterthanen zu N.	lehns-voigtey über 8. güther zu N.
Zehend zu N.		½ zehend zu N.
—	Der ſee N.	hohe u. niedere jagd zu N.

zu der in dieſem §. entworffenen tabell können noch ferner die thore, vornehmſte gaſſen, ja auch zumahl an oertern, ſo nicht waſſerreich, die brunnen gebracht werden, wiewohl, wo der raum zu enge ſalten ſollte, ſolches in jedes orts ſpecial-tabelle, ſamt andern etwan noch merckwürdigen ſachen geworffen wird, wie ich ſolches verſchiedentlich obſerviret habe.

Tabell

Tabell über die zugehörungen der ämter, der herrſchafften, ſtädte oder dörffer des amts oder herrſchafft N. N.

Darinn befinden ſich	Stadt N.	Flecken N.	Dorff N.	Weller N.
Feuerſtätten	N.	N.	N.	N.
Herrſchafftge-bäude.	N.	N.	N.	N.
Pfarr- oder andere kirchen.	N.	N.	N.	N.
Clöſter oder collegia.	N.	N.	N.	N.
Schulen.	N.	N.	N.	N.
Hoſpitalien.	N.	N.	N.	N.
Siechen-häuſer und dergleichen.	N.	N.	N.	N.
Rath-häuſer.	N.	N.	N.	N.
Gemeind-gilden- oder kauff-häuſer.	N.	N.	N.	N.
Adeliche ſitze.	N.	N.	N.	N.
Frey-höfe.	N.	N.	N.	N.
Brau-höfe.	N.	N.	N.	N.
Mahl-mühlen.	N.	N.	N.	N.
Andere mühlen.	N.	N.	N.	N.

Wo-

MODÈLES DE TABLEAUX DONNÉS PAR SECKENDORF
(édition de 1737)

Document 7

Conservée à Leyde, dans la collection HUYGENS, cette courbe donne le nombre de survivants en fonction de l'âge, d'après la Table de GRAUNT.

« Sur la ligne droite d'embas sont marquez les aages des personnes et sur les 6 il y a une perpendiculaire de 64 parties parce que de 100 personnes selon la table angloise il en reste 64 a l'âge de 6 ans. Sur le 16 il y a une perpendiculaire de 40 parties parce qu'a l'aage de 16 ans il reste 40 personnes des 100 qui estoient conçues, et ainsi du reste. Et par tous les points ou bouts de ces perpendiculaires j'ay mené la ligne courbe 64, 40, 25 &c. Si je veux scavoir maintenant combien il reste de personnes après les 20 années de 100 enfans conçus, Je prens sur la ligne d'embas l'aage de 20 ans au point A d'ou ayant erigé une perpendiculaire qui rencontre la courbe en B, je dis que AB, qui pris sur l'eschelle d'embas fait presque 33 parties, est le nombre des personnes qui de 100 conçus atteignent l'aage de 20 ans. que si je veux scavoir en suite combien il reste raisonnablement a vivre a une personne de 20 ans par exemple, je prens la moitié de BA et l'ajuste en DC entre la courbe et la droite en sorte qu'elle soit perpendiculaire à la dernière. Et j'ay AC pour les annees qui restent a vivre a la dite personne, qui font pres de 16 ans, comme il paroit par les divisions dont chacune est une année. La raison est, que la perpendiculaire DC estant la moitié de BA que marquoit le nombre d'hommes qui restent des 100, 20 ans après la conception, a scavoir 33, cette DC tombant sur 36 de la droite marquera qu'il reste la moitié de 33 c'est à dire 16 $\frac{1}{2}$ hommes après la 36 année. Donc puis que des 33 personnes de 20 ans la moitié meurt d'ordinaire dans les prochains 16 ans, on peut gager avec egal avantage qu'une personne de 20 ans vivra encore 16 ans. On trouvera de mesme que la vie d'un enfant conceu doit estre taxée à 11 ans au lieu que mon frere contoit 18 et 2 mois. »

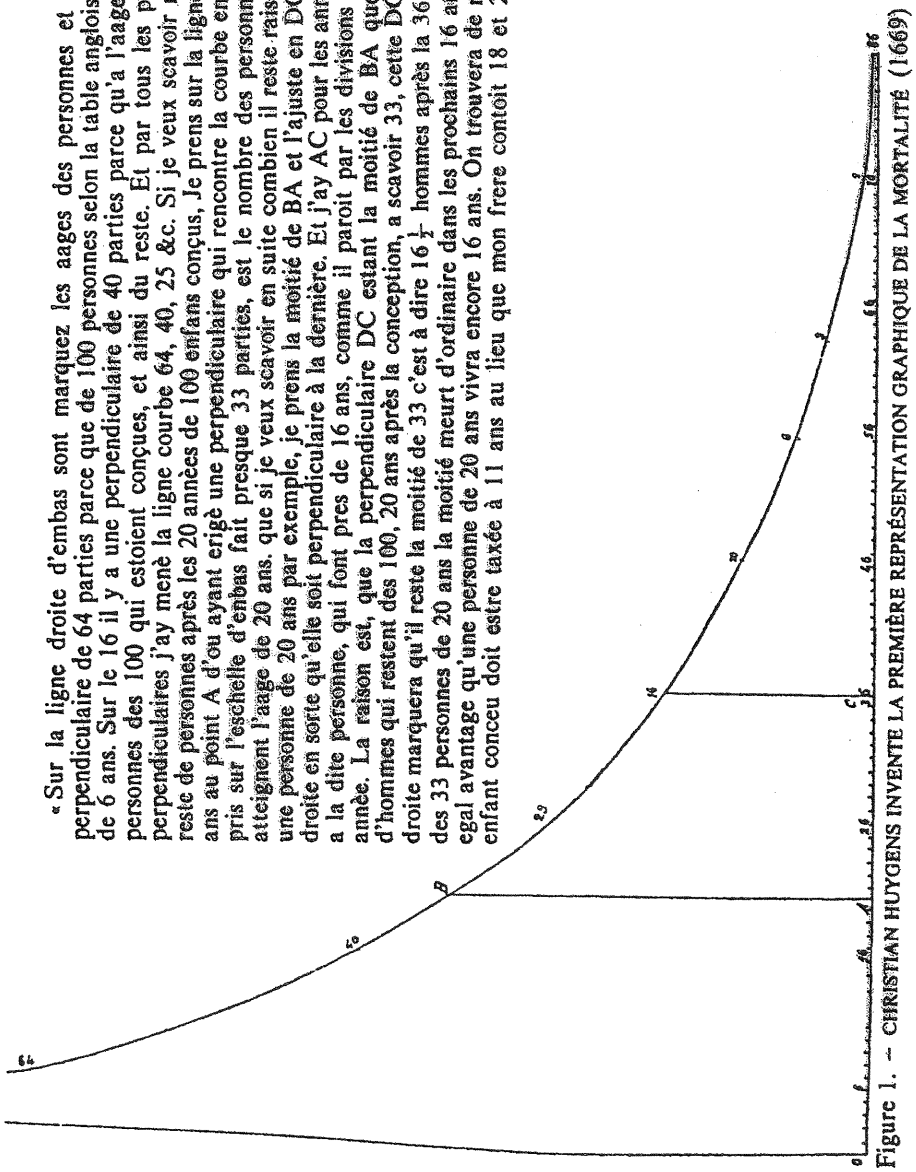
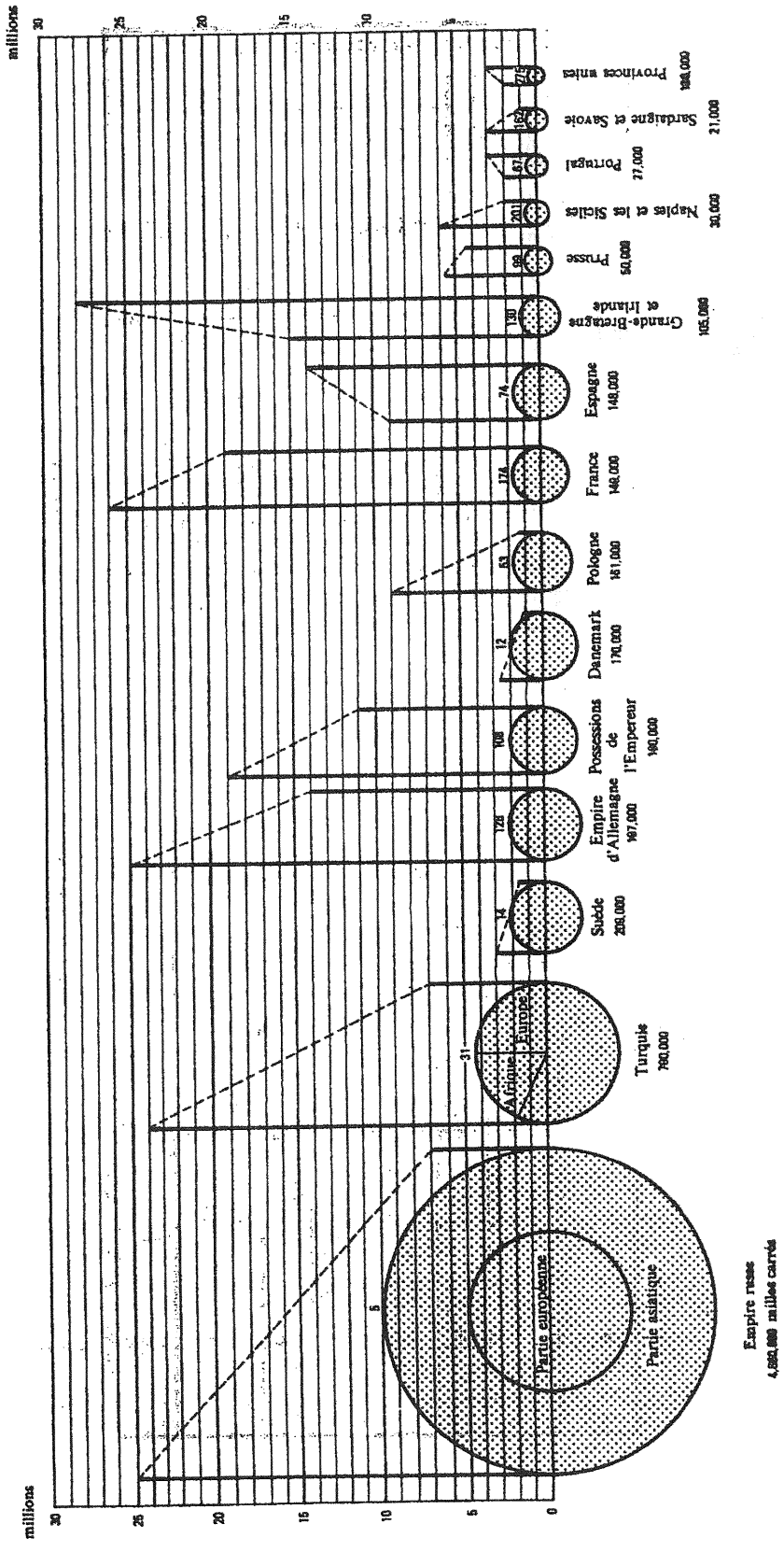


Figure 1. — CHRISTIAN HUYGENS INVENTE LA PREMIÈRE REPRÉSENTATION GRAPHIQUE DE LA MORTALITÉ (1669)

Document 8

William PLAYFAIR (1759 - 1823)
Figures géométriques in "Commercial and Political Atlas" 1786.



Document 9



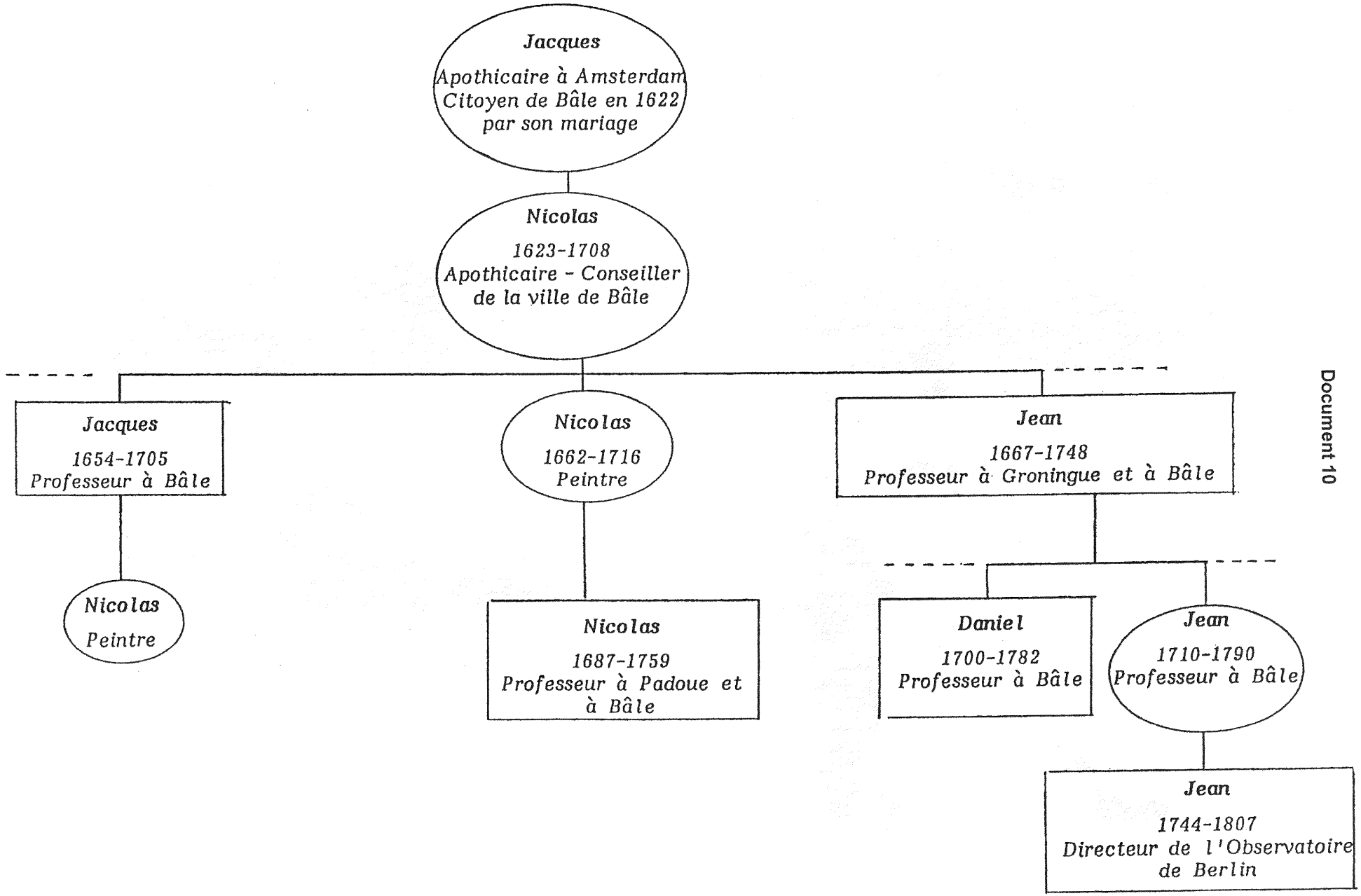
Jacob Bernouilli (1654-1705)



Johann Bernouilli (1667-1748)



Daniel Bernouilli (1700-1782)


Jacques
Apothicaire à Amsterdam
Citoyen de Bâle en 1622
par son mariage
Nicolas
1623-1708
Apothicaire - Conseiller
de la ville de Bâle
Jacques
1654-1705
Professeur à Bâle
Nicolas
Peintre
Nicolas
1662-1716
Peintre
Nicolas
1687-1759
Professeur à Padoue et
à Bâle
Jean
1667-1748
Professeur à Groningue et à Bâle
Daniel
1700-1782
Professeur à Bâle
Jean
1710-1790
Professeur à Bâle
Jean
1744-1807
Directeur de l'Observatoire
de Berlin

Document 11

Nicolas BERNOULLI au lecteur

Aujourd'hui, enfin, après une longue attente, paraît le Traité posthume de mon Oncle sur l'Art de conjecturer, et cela grâce aux Frères Thurneysen. Ceux-ci, désireux de plaire au public, ont pris soin de faire imprimer à leurs frais le Manuscrit de l'Auteur qu'ils ont l'extraordinaire usage qu'a dans la vie civile cette partie de la connaissance dont peu de personne ont jusqu'ici traité et qui concerne la mesure des probabilités. Comment et jusqu'à quel point l'auteur a-t-il exécuter son projet, on en trouvera l'examen dans les Mémoires de l'Académie Française Royale des Sciences de l'année 1705 et dans le Journal des Savants de Paris de l'année 1706. L'Auteur a divisé cette Œuvre en quatre parties. La I^{ère} contient le Traité de l'Illustre Huygens de Ratiociniis in Aleae Ludo, avec des notes, Traité que mon oncle a jugé devoir être mis en tête de son ouvrage, car on y trouve les premiers éléments de l'Art de Conjecturer. La II^{ème} partie embrasse la Théorie des Permutations et des Combinaisons, théorie absolument nécessaire pour mesurer les probabilités ; dans la IIIème partie il en a développé les applications aux divers Tirages au sort et Jeux de Hasard. Quant à la IV^{ème} partie par laquelle il aurait voulu montrer l'application et le rapport des parties précédentes aux affaires civiles, morales et économiques, affligé depuis longtemps d'une mauvaise santé et enfin prévenu par la mort elle-même, il l'a laissée inachevée. Les Editeurs auraient désiré que le Frère du Défunt, qui eût été tout à fait capable d'achever cette œuvre, comblât la lacune. Mais ils ne voulurent pas, en lui demandant ce travail, importuner un homme pris par de nombreuses autres tâches. Et moi-même qui, ils le savaient, avais jadis donné dans ma Dissertation Inaugurale des Exemples de cet Art appliqué au Droit, et à qui ils avaient pensé confier cette tâche, je ne pouvais l'entreprendre car j'étais absent, en séjour à l'étranger. Sollicité de nouveau à mon retour dans ma patrie, je refusai ce travail. Je suis jeune, je suis dépourvu de la longue pratique et de la longue expérience qui sont nécessaires pour traiter ce sujet ; je sentais donc que je ne serais pas à la hauteur de la tâche ; je jugeais aisément qu'il ne serait pas donné satisfaction au Lecteur et surtout que le reste perdrait de son prix si j'y ajoutais seulement des banalités vraiment rebattues. C'est pourquoi j'ai conseillé que ce Traité, qui était déjà imprimé en grande partie, fût communiqué au public dans l'état où l'Auteur l'avait laissé. Mais pour qu'une chose aussi utile que l'application du calcul des probabilités à l'économie et à la politique ne soit pas négligée, nous invitons l'Illustre Auteur du livre en Français Essai d'analyse sur les jeux de hasard, ainsi que le célèbre De Moivre, qui ont été l'un et l'autre publié il y a peu d'excellents travaux sur cet art, à vouloir bien prendre eux-mêmes la responsabilité de cette tâche et le temps de communiquer au public leurs remarquables trouvailles. Nous espérons toutefois que les généralités que l'Auteur expose dans les cinq chapitres de la dernière Partie ne seront pas d'une médiocre utilité au Lecteur attentif pour l'éclaircissement de certaines questions. Voilà comment nous avons jugé bon de préfacer ce Traité. Les Editeurs ont ajouté les Propositions de l'Auteur sur les Séries Infinies, sujet qu'il avait traité autrefois à fond en cinq Mémoires mais comme de nombreux exemplaires en ont été recherchés jusqu'à présent en vain chez nos libraires, les éditeurs les ont réimprimés ensemble à la suite de ce Traité. S'y ajoute également, puisque le sujet en est voisin, la Lettre en Français de l'Auteur sous le titre Lettre à un amy, etc... Il faut encore signaler que le tableau des variations du vers de Bauhusius Tot Tibi etc ..., trouvé au milieu des papiers de l'Auteur a été inséré par le Correcteur, par amitié envers les Curieux ; ainsi n'y a-t-il pas à tenir compte des derniers mots de la page 78. Les autres erreurs, peu nombreuses à la vérité, qui ont échappé au Correcteur, nous les avons notées à la fin du livre. Nous prions le Lecteur de vouloir bien les corriger.

Document 12

Extrait de « L'ART DE CONJECTURER » (4ème partie) traitant de l'Usage et l'Application de la Doctrine précédente aux affaires Politiques, Morales et Economiques, Jacques BERNOULLI.

Traduction : M. MEUNIER - LALANDE.

On en est ainsi venu à ce point que pour former, selon les règles, des conjectures sur n'importe quelle chose, il est seulement requis d'une part, que le nombre des cas soit soigneusement déterminé, et d'autre part, que soit défini combien les uns peuvent arriver plus facilement que les autres. Mais, c'est ici, enfin, que surgit une difficulté, nous semble-t-il : cela peut se voir à peine dans quelques très rares cas et ne se produit presque pas en dehors des jeux de dés que leurs premiers inventeurs ont pris soin d'organiser en vue de se ménager l'égalité de telle sorte que fussent assurés et connus les nombres des cas qui doivent entraîner le gain ou la perte, et de telle sorte que tous ces cas pussent arriver avec une égale facilité. En effet, lorsqu'il s'agit de tous les autres résultats dépendant pour la plupart soit de l'œuvre de la nature soit de l'arbitraire des hommes, cela n'a pas du tout lieu. Ainsi, par exemple, les nombres des cas sont connus lorsqu'il s'agit des dés car pour chacun des dés ils sont manifestement aussi nombreux que les bases, et sont tous également enclins à échoir ; à cause de la similitude des bases et du poids égal des dés il n'y a point de raison pour qu'une des bases soit plus encline à échoir que l'autre comme cela arriverait si les bases de forme dissemblable, ou si le dé était constitué d'un côté d'une matière plus lourde que l'autre.

Ainsi, sont connus de même les nombres des cas pour que sorte de l'urne un bulletin blanc ou noir, et on sait que tous sont également possibles ; puisque sont évidemment déterminés et connus les nombres des bulletins de chaque espèce, on ne voit aucune raison pour que celui-ci ou celui-là doive sortir plutôt que n'importe quel autre. Mais qui donc parmi les mortels définira, par exemple, le nombre des maladies, qui sont autant de cas, qui ont le pouvoir d'envahir les innombrables parties du corps humain à l'âge qu'on voudra, et qui ont le pouvoir de nous apporter la mort ? Qui définira combien est plus facile à celle-ci qu'à celle-là, la peste ou l'hydropisie, l'hydropisie ou la fièvre, d'anéantir un homme, en sorte qu'à partir de là puisse être formée une conjecture sur un état futur de vie ou de mort ? Qui encore recensera les cas innombrables des changements auquel l'air est soumis chaque jour, en sorte qu'on puisse à partir de là conjecturer ce que sera son état après un mois, sans parler d'une année ? En outre, qui aurait sur la nature de l'esprit humain, ou sur l'admirable machine de notre corps une vue suffisante pour oser déterminer dans les jeux, qui dépendent en totalité ou en partie de la finesse de celui-là ou de l'agilité de celui-ci, les cas qui peuvent donner la victoire ou l'échec à celui-ci ou celui-là des concurrents ? Car ces faits dépendent de causes tout à fait cachées, et destinées de plus par l'indénombrable variété des assemblages à se jouer éternellement de notre recherche : il serait donc absolument d'un insensé de vouloir connaître quelque chose de cette matière.

Mais à la vérité ici s'offre à nous un autre chemin pour obtenir ce que nous cherchons. Ce qu'il n'est pas donné d'obtenir a priori l'est du moins A POSTERIORI, c'est-à-dire qu'il sera possible de l'extraire en observant souvent l'issue d'exemples semblables ; car on doit présumer que par la suite, chaque fait peut arriver et ne pas arriver dans le même nombre de cas qu'il avait été constaté auparavant dans un état de choses semblable qu'il arrivait ou n'arrivait pas. En effet, si, par exemple, après avoir fait autrefois l'expérience sur trois cents hommes de l'âge et de la complexion qu'à aujourd'hui Titius, on a observé que deux cents d'entre-eux ont trouvé la mort avant la fin de la décennie, tandis que le reste a poursuivi sa vie au-delà, on pourrait conclure avec assez de certitude qu'il y a deux fois plus de cas pour que Titius doivent acquitter son tribut à la nature pendant la prochaine décennie plutôt que pour qu'il puisse franchir cette limite. De même si on a été attentif depuis de nombreuses années passées aux conditions atmosphériques et si on a noté combien de fois elles ont été bonnes ou pluvieuses ; ou si quelqu'un a assisté très souvent au luttes de deux concurrents, s'il a vue combien de fois celui-ci ou celui-là est sorti vainqueur, par là-même il aura découvert le rapport qu'ont probablement entre eux les nombres des cas qui permettent aux mêmes

évènements, dans des circonstances semblables aux précédentes, d'arriver ou de ne pas arriver ultérieurement.

Cette manière empirique de déterminer par expérience les nombres des cas n'est ni neuve, ni insolite ...

C'est la même que tous observent constamment dans la pratique quotidienne. Enfin, il ne peut échapper à personne, que pour juger par ce moyen de quelque événement, il ne suffira pas d'avoir fait choix d'une ou de deux expériences, mais qu'il serait requis une grande quantité d'expériences : tout être des plus stupides, par je ne sais quel instinct naturel, par lui-même et sans le guide d'aucun enseignement (chose absolument admirable) tient pour évident que, plus on aura recueilli de nombreuses observations de ce genre, moins grand sera le danger de s'écarter du but. Or, bien que cela soit naturellement connu de tous, la démonstration qui permet de le tirer des principes de l'analyse n'est pas du tout répandue, et par la suite, il nous incombe d'en traiter en cet endroit, endroit où cependant j'estimerais que je ferais trop peu si je m'en tenais à démontrer seulement ce que personne n'ignore. Il reste alors à examiner par la suite quelque chose que peut-être personne n'a jusqu'à maintenant rencontré même en y pensant. Il reste assurément à chercher si, en augmentant ainsi le nombre des observations, nous augmentons continuellement la probabilité d'atteindre le rapport réel entre les nombres des cas qui font qu'un événement peut arriver et ceux qui font qu'il ne peut arriver, de sorte que cette probabilité dépasse enfin un degré quelconque donné de certitude ; il reste à chercher si le Problème, pour ainsi dire a son Asymptote, c'est-à-dire s'il existe un degré de certitude qu'il n'est jamais possible de dépasser, de quelque manière qu'on multiplie les observations ; par exemple, quand nous ne pouvons jamais obtenir de certitude au-delà de la moitié, ou de $\frac{2}{3}$ *oude* $\frac{3}{4}$ *nous avons découvert le vrai rapport des cas. Un exemple rendra clair ce que je voudrais dire. Je suppose que dans une urne, à ton insu, soit placées trois milles pierres blanches et deux milles pierres noires ; je suppose que pour connaître leur nombre par expérience tu tires une pierre après l'autre (en remplaçant cependant chaque fois la pierre que tu as tirées avant de choisir la suivante, pour que le nombre des pierres ne diminue pas dans l'urne tu observes combien de fois sort une pierre blanche et combien de fois une noire. On demande si tu peux le faire tant de fois qu'il devienne dix fois, cent fois, mille fois, etc ... plus probable (c'est-à-dire qu'il devienne moralement certain) que le nombre de fois où tu choisis une pierre blanche et le nombre de fois où tu choisis une pierre noire soit dans ce même rapport sesquialtère (rapport égal à* $\frac{3}{2}$*) où se flattent d'être les nombres eux-mêmes des pierres ou des cas, plutôt que dans tout autre rapport différent de celui-ci. Car, si cela ne se produisait pas, j'avoue que c'en serait fait de notre effort pour rechercher expérimentalement les nombres des cas. Mais, si nous l'obtenons et si nous acquérons enfin par ce moyen la certitude morale (et je montrerai dans le chapitre suivant que cela aussi se produit réellement), nous aurons trouvé A POSTERIORI les nombres des cas presque comme s'ils nous étaient connus a priori ; assurément, dans la pratique de la vie civile, où le moralement certain est tenu pour absolument certain, en vertu de l'Ax. 9, Ch. II*[1] *aide largement pour régler nos conjectures dans n'importe quel domaine non moins scientifiquement que dans les jeux de hasard : en effet, si à la place de l'urne nous mettions l'air, par exemple, ou le corps humain, qui contiennent en eux l'aliment des variations atmosphériques et des maladies, comme l'urne contient les pierres nous pourrions en tout cas, par le même procédé déterminer grâce à l'observation combien plus facilement peut arriver dans ces sujets tel ou tel événement*[3].

[1] « Parce qu'il est rarement permis d'atteindre une certitude absolue, la nécessité et l'usage veulent que ce qui est seulement moralement certain soit tenu pour absolument certain ».

Par exemple, c'est au juge de décider si c'est à partir de $\frac{99}{100}$ ou de $\frac{999}{1000}$ etc... de certitude qu'on peut parler de certitude absolue.

[2] Le texte se prolonge par une rigoureuse démonstration de la loi des grands nombres à partir de la considération et de la comparaison entre eux des termes du binôme de NEWTON.

Proposition Principale

Il s'ensuit enfin la proposition elle-même, pour laquelle tout cela a été formulé, mais dont la démonstration se fait maintenant par la seule application des lemmes préparatoires à l'objet présent. Pour éviter la fatigue d'une circonlocution, j'appellerai « féconds » ou « fertiles » ceux dans lesquels un événement ne peut se produire : de même, j'appellerai expériences « fécondes » ou « fertiles » celles pour lesquelles on constate qu'un des cas fertiles peut survenir, et « inféconds » ou « stériles » celles pour lesquelles on observe qu'un des cas stériles se produit. Soit donc le nombre de cas fertiles au nombre de cas stériles précisément ou approximativement dans le rapport $\frac{r}{s}$ *et qu'il soit en conséquence, au nombre de tous dans le rapport* $\frac{r}{r+s}$ *ou* $\frac{r}{t}$, *rapport qu'encadrent les limites* $\frac{r+1}{t}$ et $\frac{r-1}{t}$. *Il faut montrer que l'on peut concevoir des expériences en un nombre tel qu'il soit plus vraisemblable d'autant de fois que l'on veut (soit c) que le nombre des observations tombe à l'intérieur de ces limites plutôt qu'en dehors, c'est-à-dire que le nombre des observations fertiles soit au nombre de toutes les observations dans un rapport ni plus grand que* $\frac{r+1}{t}$, *ni plus petit que* $\frac{r-1}{t}$. *Mais, dans les applications numériques, il va de soi que plus les nombres r, s, et t prennent des valeurs élevées en conservant les mêmes rapports, plus les limites* $\frac{r+1}{r}$ *et* $\frac{r-1}{t}$ *qui encadrent le rapport* $\frac{r}{t}$ *peuvent être étroitement resserrées. Ainsi, si le rapport* $\frac{r}{s}$ *entre les nombres de cas, déterminé expérimentalement, était par exemple sesquialtère, pour r et s je ne pose pas 3 et 2, mais 30 et 20, ou 300 et 200 etc. et il suffirait d'avoir posé r = 30, s = 20, et t = r + s = 50, pour que les limites deviennent* $\frac{r+1}{t}=\frac{31}{50}$ *et* $\frac{r-1}{t}=\frac{29}{50}$ *et qu'on pose en outre* $c = 1000$ *on aura ainsi d'après ce qui a été écrit dans la Scholies pour terme à*

Gauche :

$$m > \frac{Lc(s-1)}{L(r+1)-Lr} = \frac{4.2787536}{142405} < 301$$

$$nt = mt + \frac{mst - st}{r+1} < 24728$$

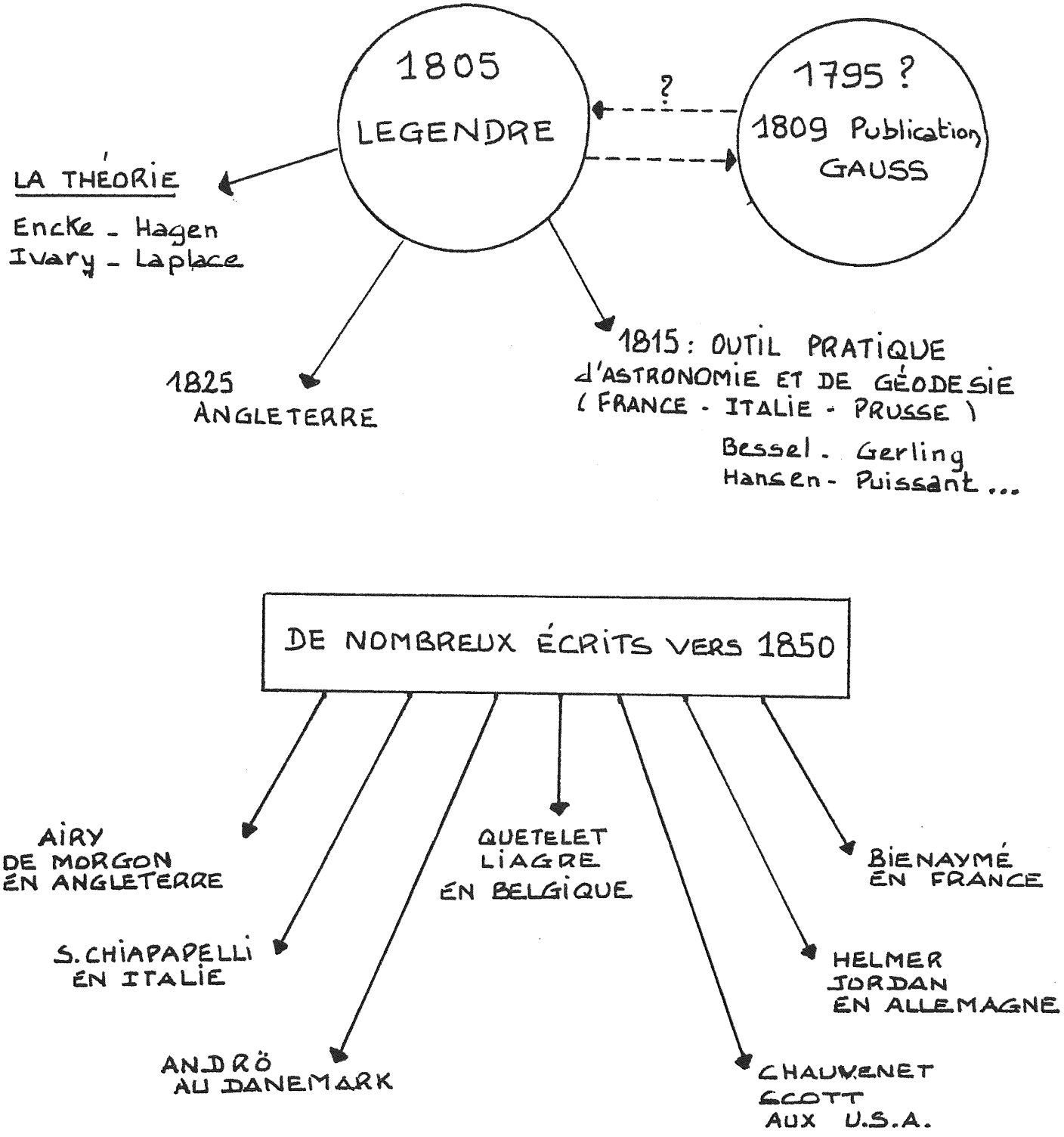
Droite :

$$m > \frac{Lc(r-1)}{L(s+1)-Ls} = \frac{4.4623980}{211893} < 211$$

$$nt = mt + \frac{mrt - rt}{s+1} = 25550$$

De là, on déduit, grâce à ce qui a été démontré ici, qu'ayant fait 25550 expériences, il est vraisemblable de bien plus de mille fois que le rapport du nombre des observations fertiles au nombre de toutes sera compris entre les limites $\frac{31}{50}$ *et* $\frac{29}{50}$ *plutôt qu'en dehors. Et, par le même raisonnement, on connaîtra ces limites lorsqu'on pose c = 10000 ou c = 100000, etc. et ainsi il sera plus probable de plus de 10000 fois, si on fait 31258 expériences, et de plus de 100000 fois si on prend 36966, etc., et ainsi de suite à l'infini, sans cesser n'est-ce pas d'ajouter 5708 expériences nouvelles à partir de 25550. De là, enfin, semble découler cette chose extraordinaire que, si les observations de tous les événements se poursuivaient à travers l'éternité entière, (la probabilité débouchant à la fin sur la certitude parfaite), nous saisirions que toute chose dans le monde arrive par des raisons certaines (68) et par une loi constante du retour des choses ; à tel point que, même dans les choses les plus accidentelles, et les plus fortuites, nous sommes tenus de reconnaître une quasi nécessité et pour ainsi dire une fatalité ; je ne sais si Platon lui-même n'a pas voulu suggérer cette fatalité, dans son dogme du retour universel des choses, selon lequel il a prédit que toute chose reviendrait à son état antérieur après d'innombrables siècles.*

Document 13


1805
LEGENDRE
?
1795 ?
1809 Publication
GAUSS
LA THÉORIE
Encke - Hagen
Ivary - Laplace
1825
ANGLETERRE
1815 : OUTIL PRATIQUE
d'ASTRONOMIE ET DE GÉODESIE
(FRANCE - ITALIE - PRUSSE)
Bessel . Gerling
Hansen - Puissant ...
DE NOMBREUX ÉCRITS VERS 1850
AIRY
DE MORGON
EN ANGLETERRE
S. CHIAPARELLI
EN ITALIE
ANDRÖ
AU DANEMARK
QUETELET
LIAGRE
EN BELGIQUE
CHAUVENET
SCOTT
AUX U.S.A.
HELMER
JORDAN
EN ALLEMAGNE
BIENAYMÉ
EN FRANCE

Document 14

Laplace revient sur la loi des erreurs et écrit en 1781

Maintenant, on peut entendre une infinité de choses différentes par le *milieu* ou le *résultat moyen* d'un nombre quelconque d'observations, suivant que l'on assujettit ce résultat à telle ou telle condition. Par exemple, on peut exiger que ce milieu soit tel que la somme des erreurs à craindre en *plus* soit égale à la somme des erreurs à craindre en *moins;* on peut exiger que la somme des erreurs à craindre en plus, multipliées par leurs probabilités respectives, soit égale à la somme des erreurs à craindre en moins, multipliées par leurs probabilités respectives. On peut encore assujettir ce milieu à être le point où il est le plus probable que doit tomber le véritable instant du phénomène, comme M. Daniel Bernoulli l'a fait dans les Mémoires cités : en général, on peut imposer une infinité d'autres conditions semblables qui donneront chacune un milieu différent; mais elles ne sont pas toutes arbitraires. Il en est une qui tient à la nature du problème et qui doit servir à fixer le milieu qu'il faut choisir entre plusieurs observations : cette condition est que, en fixant à ce point l'instant du phénomène, l'erreur qui en résulte soit un minimum;

Table 5.7. Quetelet's analysis of 1817 data on the heights of French conscripts.

(1) Height (in meters)	(2) Number of men	(3) Relative freq.	(4) Cumulative freq.	(5) Rank	(6) Fitted rank	(7) Prob.	(8) Fitted no. of men	(9) Difference (2)–(8)
Less than 1.570	28,620	0.2862	0.5000	—	—	—	26,345	+2,275
1.570–1.597	11,580	0.1158	0.2138	9	10.0	0.23655	13,182	−1,602
1.597–1.624	13,990	0.0980	0.0980	4	4.2	0.10473	14,502	−512
		0.0419	0.0419	1.6	1.6	0.04029		
1.624–1.651	14,410	0.1441	0.1860	7.5	7.4	0.18011	13,982	+428
1.651–1.678	11,410	0.1141	0.3001	13.5	13.2	0.29814	11,803	−393
1.678–1.705	8,780	0.0878	0.3879	19	19.0	0.38539	8,725	+55
1.705–1.732	5,530	0.0553	0.4432	25	24.8	0.44166	5,627	−97
1.732–1.759	3,190	0.0319	0.4751	31	30.6	0.47355	3,189	+1
Above 1.759	2,490	0.0249	0.5000	—	—	—	2,645	−155
	100,000						100,000	

Sources: Quetelet (1846, pp. 401–402; 1849, pp. 277–278).

Note: The entries in columns 5 and 7 are derived from Quetelet's table (Figure 5.5). The seventh and eighth entries in columns 8 and 9 are here corrected from the original, where they read 5,527 and 3,187 in column 8 and +3 and +3 in column 9. Quetelet apparently obtained the data from Villermé (1829), who cited as his source an 1817 "brochure" by Hargenvilliers titled "Recherches et considerations sur la formation et le recrutement de l'armée en France."

Table 5.8. Relative frequency distribution of the heights of 9,002 French conscripts from the department of Doubs, 1851–1860.

Height	Relative frequency
Below 4′10″	577
4′10″ to 4′11″	637
4′11″ to 5′0″	1,116
5′0″ to 5′1″	1,766
5′1″ to 5′2″	1,457
5′2″ to 5′3″	1,777
5′3″ to 5′4″	1,313
5′4″ to 5′5″	820
5′5″ to 5′6″	291
5′6″ to 5′7″	153
5′7″ to 5′8″	64
5′8″ to 5′9″	17
5′9″ to 5′10″	9
5′10″ to 5′11″	3
	10,000

Source: Bertillon (1863, p. 238).

Note: Livi (1896) found that the apparent drop in the class 5′1″ to 5′2″ was due to Bertillon's incorrect conversion of measurements grouped by centimeters to measurements grouped by inches.

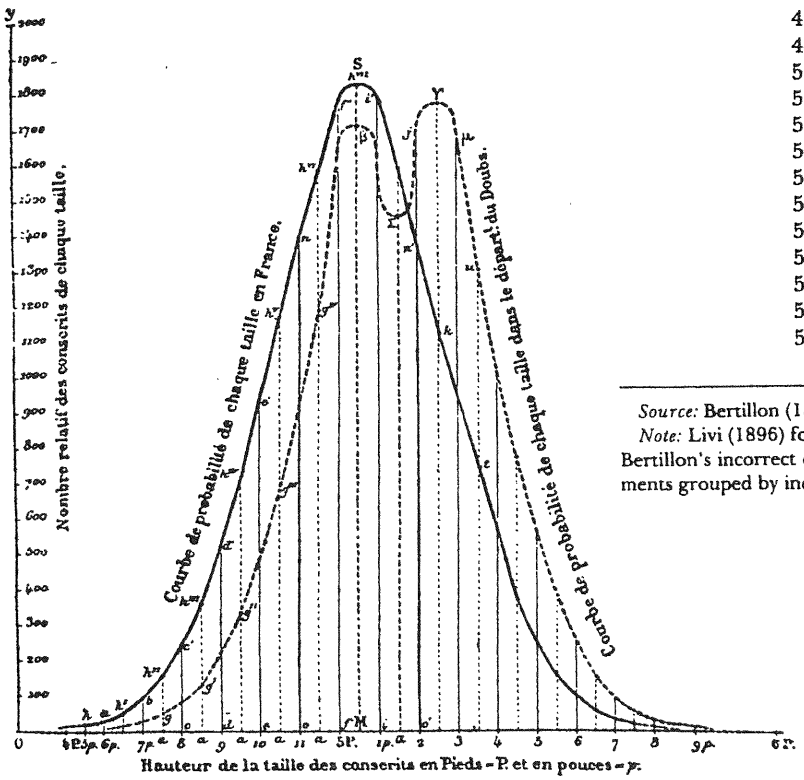


Figure 5.6. Adolphe Bertillon's depiction of the curve of heights of conscripts for all of France (solid line) *and for the department of Doubs* (dotted line), *purporting to show a bimodal distribution. (From Bertillon, 1876, p. 306.)*

Table 5.6. Distribution of heights and chest circumferences of 5,732 Scottish militia men.

Height (inches)	Number of men with chest circumference (inches) of—																Total no. of men in height class
	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	
64–65	1	7	31	69	108	154	142	118	66	17	6	3	0	0	0	0	722
66–67	1	9	30	78	170	343	442	337	231	124	34	12	3	1	0	0	1,815
68–69	1	2	16	34	91	187	341	436	367	292	126	70	13	3	2	0	1,981
70–71	0	1	4	7	31	62	117	153	209	148	102	40	16	7	0	0	897
72–73	0	0	0	1	9	7	20	38	62	65	45	43	18	7	1	1	317
Total no. of men in chest-size class	3	19	81	189	409	753	1,062	1,082	935	646	313	168	50	18	3	1	5,732

Source: Edinburgh Medical and Surgical Journal (1817, pp. 260–264).

MESURES de la poitrine.	NOMBRE d'hommes.	NOMBRE proportionnel.	PROBABILITÉ d'après l'observation.	RANG dans la table.	RANG d'après le calcul.	PROBABILITÉ d'après la table.	NOMBRE d'observations calculé.
Pouces.							
33	3	5	0,5000			0,5000	7
34	18	31	0,4995	52	50	0,4993	29
35	81	141	0,4964	42,5	42,5	0,4964	110
36	185	322	0,4823	33,5	34,5	0,4854	323
37	420	732	0,4501	26,0	26,5	0,4531	732
38	749	1305	0,3769	18,0	18,5	0,3799	1333
39	1073	1867	0,2464	10,5	10,5	0,2466	1838
			0.0597	2,5	2,5	0,0628	
40	1079	1882	0,1285	5,5	5,5	0,1359	1987
41	934	1628	0,2913	13	13,5	0,3034	1675
42	658	1148	0,4001	21	21,5	0,4130	1096
43	370	645	0,4706	30	29,5	0,4090	560
44	92	160	0,4866	35	37,5	0,4911	221
45	50	87	0,4953	41	45,5	0,4980	69
46	21	38	0,4991	49,5	53,5	0,4996	16
47	4	7	0,4998	56	61,8	0,4999	3
48	1	2	0,5000			0,5000	1
	5738	1,0000					1,0000

Figure 5.3. Quetelet's analysis fitting a normal distribution to data on the chest circumferences of Scottish soldiers. Column 1 gives the chest circumference in inches; columns 2 and 3 give the frequency and relative frequency distributions for 5738 individuals, columns 4–7 give the details of Quetelet's calculations (see text), and column 8 gives the fitted relative frequency distribution. (From Quetelet, 1846, p. 400.)

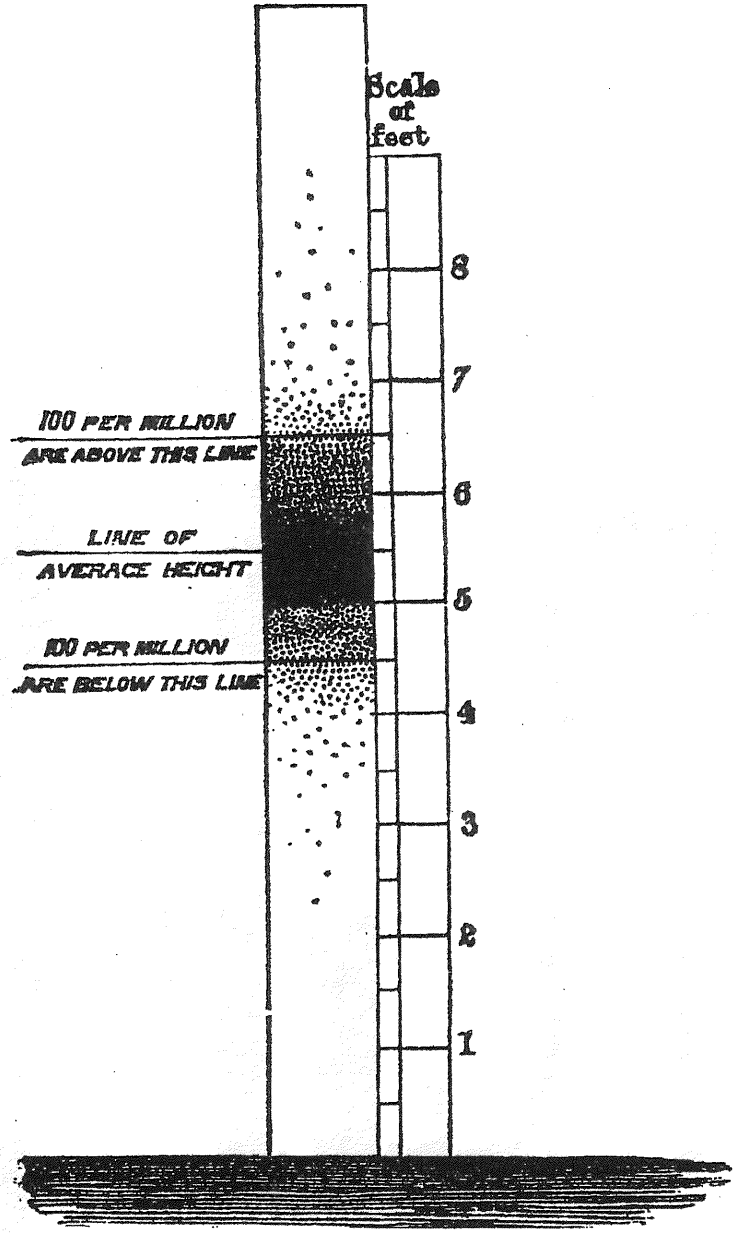


Figure 8.1. Galton's illustration of the "law of deviation from an average," showing the heights of a million hypothetical men. (From Galton, 1869, p. 28.)

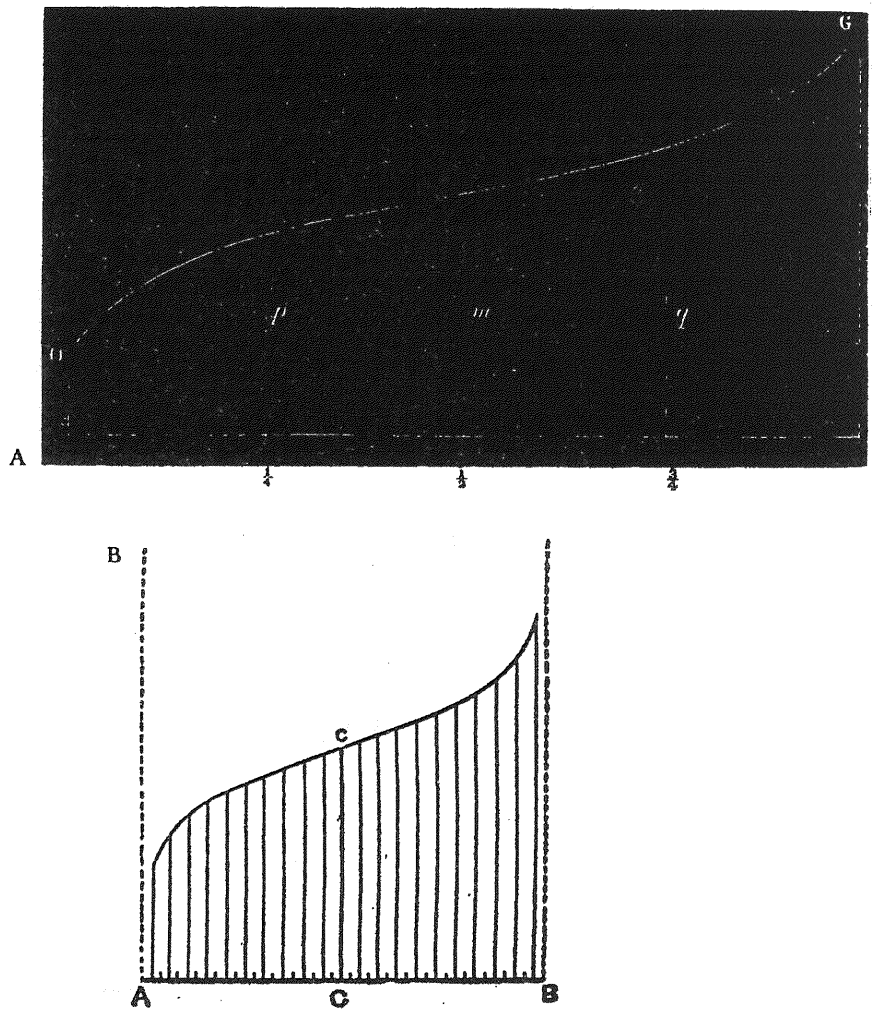


Figure 8.2. Two of Galton's renditions of the ogive. (A) The earlier, from 1875, shows the median and the quartiles (Galton, 1875). (B) The later drawing shows spikes with heights representing 21 equally spaced ideal data values (Galton, 1883, p. 51).

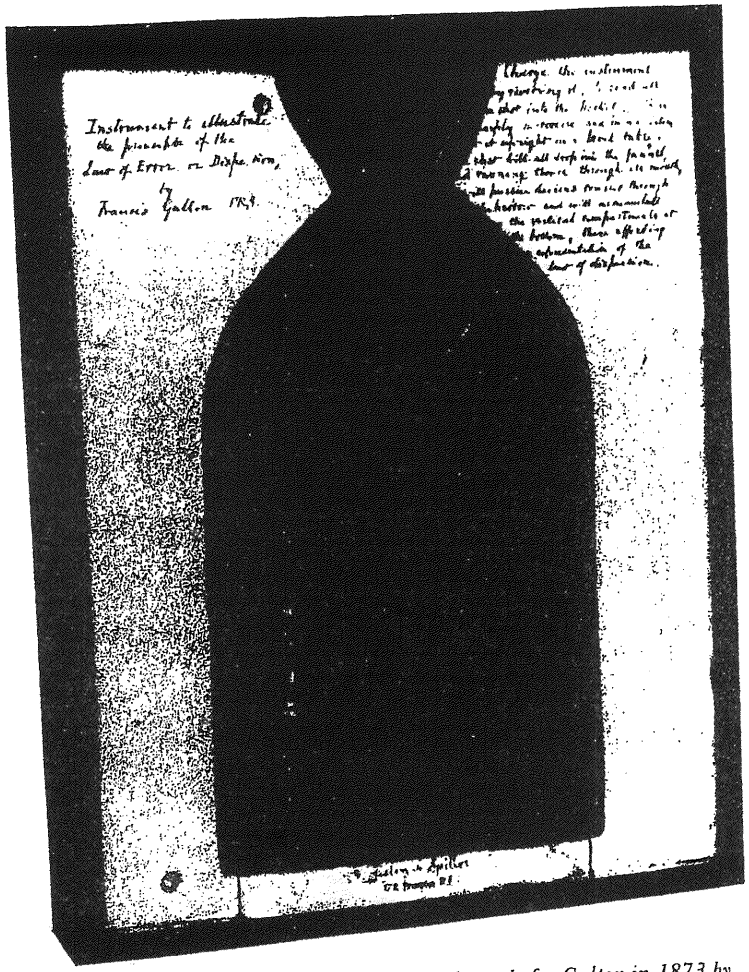


Figure 8.4. The original quincunx, apparently made for Galton in 1873 by Tisley & Spiller. Although it once had an opening at the top through which the shot could be poured, the top is now sealed with the shot inside. The glass has become cloudy with lead dust over the years. The caption, in Galton's handwriting, reads:

Instrument to illustrate
the principle of the
Law of Error or Dispersion
by
Francis Galton F.R.S.

Charge the instrument by reversing it, to send all the shot into the pocket. Then sharply re-reverse and immediately set it upright on a level table. The shot will all drop into the funnel, and running thence through its mouth, will pursue devious courses through the harrow and will accumulate in the vertical compartments at the bottom, there af-

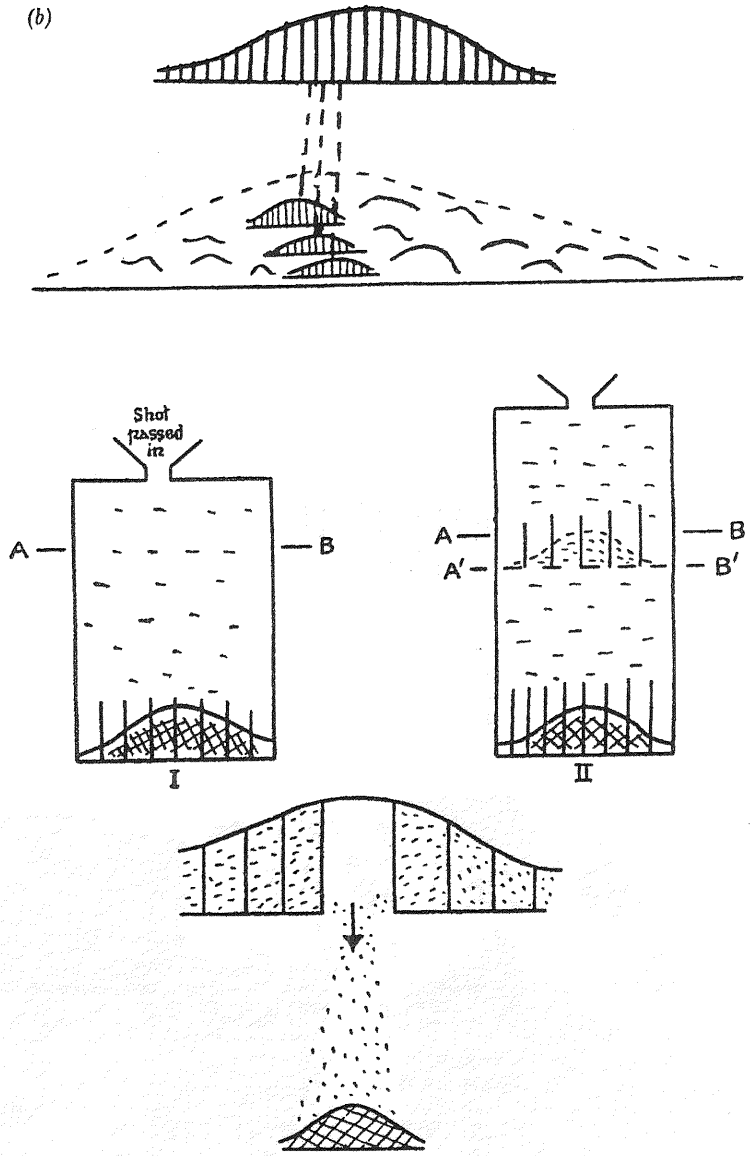


Figure 8.5. (a) Galton's explanation of the working of the two-stage version of the quincunx in a 12 January 1877 letter to his cousin, George Darwin. (From Galton Archives.) (b) Drawings by Karl Pearson, based on Galton's original letter. (From Pearson, 1914–1930, vol. 3B, pp. 465–466.)

Annexe 1

Gauss et la méthode des moindres carrés.

Au XVIII^{ème}, les astronomes et les géodésiens ne cessent d'être confronté aux questions suivantes :

*Comment combiner les observations effectuées dans des conditions différentes, afin d'obtenir les **meilleures** estimations possibles de plusieurs grandeurs astronomiques ou terrestres, liées entre elles par une relation linéaire ? (Il est toujours possible de se ramener à un tel cas en utilisant la formule de Taylor).

*Comment estimer le degré de confiance que méritent les estimations faites ?

A partir de 1801, Gauss s'intéresse à l'astronomie puis vers 1820, il se tourne vers la géodésie. Ses travaux ne sont pas seulement théoriques ; il utilise des instruments d'observation à l'observatoire de Göttingen et lors de la triangulation du Hanovre, il exécute lui-même les longs calculs qui résultent du traitement des observations, comme l'atteste sa correspondance avec Schumacher, Bessel, Jacobi, Gerling, Encke...

Gauss prétend qu'il utilisait déjà la méthode des moindres carrés en 1795. Il est certain, qu'à l'aide de cette méthode, il détermine en 1801, à la surprise générale, l'orbite de la planète Cérès découverte par Guiseppe Piazzi à la veille du nouvel an 1801 et permet ainsi à Von Zack de retrouver la planète le 7 décembre 1801 dans la position indiquée par ses calculs. Il expose cette méthode seulement en 1809 dans *Theoria Motus*. L'objet de cet ouvrage est d'étudier les relations entre les différentes positions d'un corps céleste ; Bertrand n'en traduit que la fin sous le titre *Exposition de la méthode des moindres carrés.*[1] Gauss justifie sa méthode par l'axiome selon lequel la moyenne arithmétique des valeurs observées d'une certaine quantité est la valeur la plus probable de cette quantité. Il est conduit à la loi des erreurs dite loi de Gauss $\varphi(x) = \frac{h}{\sqrt{\pi}} \exp(-h^2x^2)$ où h est le paramètre de précision. **Il établit ainsi un lien entre la méthode des moindres carrés et la théorie des probabilités.** Une querelle de priorité s'en suit entre Gauss et Legendre[2]. En effet, en 1805 dans un appendice à *Nouvelles méthodes pour la*

---

[1]Une traduction complète de *Theoria Motus* est faite par E.Dubois. Elle est publiée à Paris en 1864.
[2]A ce sujet, on peut consulter S.M. Stigler,*The History of statistics* , 1986 ou Kendall, M.G and R.L.Plackett, *Studies in the History of Statistics and Probability*, vol.2 Londres, 1977

détermination des Orbites des Comètes, Legendre est amené à résoudre un système surdéterminé. Il résout ce problème par la méthode consistant à minimiser la somme des carrés des erreurs. Il présente cette méthode comme une technique commode pour choisir au mieux des estimations ; la présentation est de type algébrique et la justification est la suivante : "*Par ce moyen il s'établit entre les erreurs une sorte d'équilibre qui, empêchant les extrêmes de prévaloir, est très propre à faire connaître l'état du système le plus proche de la vérité*".

En 1812, dans sa *Théorie analytique des Probabilités*, Laplace remarque que rien ne prouve la règle du milieu arithmétique et donc la règle de la méthode des moindres carrés. Il en donne une justification, s'appuyant uniquement sur le calcul des probabilités, cherchant à rendre minimum ce qu'il appelle "*l'erreur moyenne à craindre*" : $\int |\Delta| \varphi(\Delta)d\Delta$. Il précise : "*si l'on considère un grand nombre d'observations, ... ce choix devient indépendant de la loi des erreurs et .. l'Analyse conduit alors directement aux résultats de cette méthode des moindres carrés.*" De nos jours, nous savons que l'indépendance vis-à-vis de la loi des erreurs résulte du théorème central limite.

Peut-être influencé par Laplace, Gauss en 1821 dans *Theoria Combinationus* établit la méthode des moindres carrés sur de nouvelles bases. Il écrit : " *La méthode des moindres carrés fournit les combinaisons les plus avantageuses des observations ... et cela quelle que soit la loi des probabilités des erreurs et, quel que soit le nombre des observations, pourvu que l'on adopte pour l'erreur moyenne, non pas la définition de Laplace, mais celle que nous avons donné*" c'est-à-dire celle qui correspond actuellement à l'écart-type de la loi des erreurs, supposée centrée. Gauss introduit aussi la notion de poids d'une observation et de la précision d'une observation.[3]

Pour la résolution des grands systèmes formés par les équations normales, Gauss met en place une démarche évoquant pour nous la méthode d'élimination ou du pivot de Gauss. Cette méthode est en particulier développée dans le mémoire sur la planète Pallas. Il sera aussi amené à proposer des méthodes itératives comme celle connue sous le nom de méthode de Gauss-Seidel.[4]

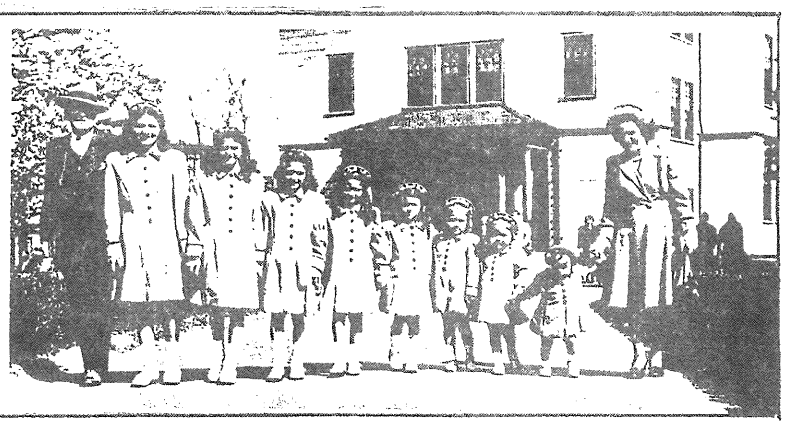
Marie Françoise Jozeau

---

[3]On peut consulter, par exemple, J.L.Chabert, *Gauss et la méthode des moindres carrés*, in Revue d'histoire des sciences, 1989.

[4]Voir à ce sujet une lettre de Gauss à Gerling datant de 1823, dont une traduction et une étude est proposée par A.Michel-Pajus in *Histoire d'Algorithmes* , p.335, Paris, Belin, 1994

II - NOUVEAUX PROGRAMMES, NOUVELLE APPROCHE



What an amazing coincidence. . . . All the daughters have the same outfits!

NOUVEAU PROGRAMME, NOUVELLE DEMARCHE

Noël BASCOU

Equipe Statistique de l'IREM de Montpellier

Résumé : *L'article propose des éléments de réflexion justifiant la* formation, *dès le lycée, au raisonnement probabiliste et statistique. Il présente ensuite les principales étapes d'une démarche statistique et propose quelques activités de statistique pour la classe, s'intégrant dans les nouveaux programmes.*

Dans les nouveaux programmes de seconde (2000) puis de première et de terminale, les statistiques prennent une part importante.

1. Un enseignement important, pourquoi ?

... L'idée que nous puissions être soumis à des aléas et des risques est spontanément évacuée et refoulée ; l'idée que nous ayons droit à un niveau nul de risque est au contraire diffusée et accréditée ; l'incertitude des connaissances est tout autant ignorée. Les Français les plus cultivés se montrent souvent incapables de construire un raisonnement correct intégrant la prise en compte d'aléas, dont il existe des mesures fiables en termes de fréquences ... C'est une culture qui fait défaut.

Or, une culture s'acquiert lentement, et cela d'autant plus qu'on y est exposé plus tard dans la vie et plus rarement dans la société où l'on travaille. Le système éducatif a une responsabilité évidente en vue de la résorption d'une carence et d'un handicap aussi répandus ...

... Il s'agit de familiariser progressivement les élèves avec un pan important d'une vraie culture moderne, de créer des réflexes intellectuels et un état d'esprit. Il doit y avoir des occasions de la sixième à la terminale.

Quel corps d'enseignants devrait en assurer la charge principale ? Celui des professeurs de mathématiques, semble-t-il, car les autres seraient, sans doute, maladroits pour enseigner le support technique de cette culture... Tout cela est simple pour un mathématicien, difficile pour d'autres ...

Les professeurs de mathématiques ont sans doute besoin de bons exemples montrant l'intervention des concepts et des raisonnements probabilistes ou inductifs dans les diverses disciplines des mondes physique et humain.

Extrait d'une communication d'Edmond MALINVAUD, professeur honoraire au Collège de France : « La formation des collégiens et lycéens aux raisonnements probabiliste et statistique ».

2. Les principales étapes dans la démarche statistique

Bien préciser la problématique, définir un modèle statistique possible, et les inconnues du problème (paramètres, dépendance stochastique, famille de loi ...), définir le **protocole expérimental** (particulier à chaque domaine d'application, en concertation avec le spécialiste de la discipline concernée) : ces notions ne seront introduites qu'en post-bac.

Résumer les données (calculs de grandeurs caractéristiques de l'échantillon : caractéristiques de position, de dispersion, ...), accompagnées de graphiques (camemberts, diagrammes en bâtons, histogrammes, courbes et nuages de points, ...) : **c'est la Statistique Descriptive.**

Appliquer des méthodes statistiques, à partir de la connaissance des modèles probabilistes usuels, pour **décider** : valeurs possibles du (ou des) paramètres (**estimation**, ponctuelle, ou par intervalle de confiance), **validation** d'un modèle, ou plus généralement d'une hypothèse statistique (test) : **c'est la Statistique Inférentielle**.

Revenir à l'expérience pour interpréter la décision, éventuellement prendre un décision (d'ordre économique, médical, écologique,...) ou modifier les hypothèses faites a priori sur le modèle statistique (notamment s'il est non validé) et poursuivre l'expérimentation, ou encore proposer des valeurs de "prédiction" ...

3. Nouveau chapitre du programme de seconde : la statistique

Voici le nouveau programme de seconde, applicable à la rentrée de septembre 2000, publication du B.O. hors série n°6 du 12/08/99 :

En seconde le travail sera centré sur :

– la réflexion conduisant au choix de résumés numériques d'une série statistique quantitative ;

– la notion de fluctuation d'échantillonnage vue ici sous l'aspect élémentaire de la variabilité de la distribution des fréquences ;

– la simulation à l'aide du générateur aléatoire d'une calculatrice. La simulation remplaçant l'expérimentation permet, avec une grande économie de moyens, d'observer des résultats associés à la réalisation d'un très grand nombre d'expériences. On verra ici la diversité des situations simulables à partir d'une liste de chiffres.

L'enseignant traitera des données en nombre suffisant pour que cela justifie une étude statistique ; il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts.

La notion de fluctuation d'échantillonnage et de simulation ne doit pas faire l'objet d'un cours. L'élève pourra se faire un « cahier de statistique » où il consignera une grande partie des traitements de données et des expériences de simulation qu'il fait, des raisons qui conduisent à faire des simulations ou traiter des données, l'observation et la synthèse de ses propres expériences et de celles de sa classe. Ce cahier sera complété en première et terminale et pourra faire partie des procédures d'évaluation annuelle.

En classe de première et de terminale, dans toutes les filières, on réfléchira sur la synthèse des données à l'aide du couple moyenne, écart-type qui sera vu à propos de phénomènes aléatoires gaussiens et par moyenne ou médiane et intervalle inter-quartile sinon. On amorcera une réflexion sur le problème de recueil des données et la notion de preuve statistique ; on fera un lien entre statistique et probabilité. L'enseignement de la statistique sera présent dans toutes les filières mais sous des formes diverses.

CONTENUS	CAPACITES ATTENDUES	COMMENTAIRES
Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de seconde à l'étendue).	Utiliser les propriétés de linéarité de la moyenne d'une série statistique. Calculer la moyenne d'une série à partir des moyennes de sous-groupes. Calcul de la moyenne à partir de la distribution des fréquences.	L'objectif est de faire réfléchir les élèves sur la nature des données traitées, et de s'appuyer sur des représentations graphiques pour justifier un choix de résumé. On peut commencer à utiliser le symbole. On commentera quelques cas où la médiane et la moyenne diffèrent sensiblement. On remarquera que la médiane d'une série ne peut se déduire de la médiane de sous-séries. Le calcul de la médiane nécessite de trier les données, ce qui pose des problèmes de nature algorithmique.
Définition de la distribution des fréquences d'une série prenant un petit nombre de valeurs et de la fréquence d'un événement. Simulation et fluctuation d'échantillonnage.	Concevoir et mettre en œuvre des simulations simples à partir d'échantillons de chiffres au hasard.	La touche « Random » d'une calculatrice pourra être présentée comme une procédure qui, chaque fois qu'on l'actionne, fournit une liste de n chiffres (composant la partie décimale du nombre affiché). Si on appelle la procédure un très grand nombre de fois, la suite produite sera sans ordre ni périodicité et les fréquences des dix chiffres seront sensiblement égales. Chaque élève produira des simulations de taille n (n allant de 10 à 100 suivant les cas) à partir de sa calculatrice ; ces simulations pourront être regroupées en une simulation ou plusieurs simulations de taille N, après avoir constaté la variabilité des résultats de chacune d'elles. L'enseignant donnera alors éventuellement les résultats de simulations de même taille N préparées à l'avance et obtenues à partir de simulations sur ordinateurs.

THEMES D'ETUDE DE STATISTIQUE

Pour chacun des chapitres, le professeur choisira, pour l'ensemble des élèves ou pour certains seulement en fonction de leurs centres d'intérêt, un ou plusieurs thèmes d'étude dans la liste ci-dessous.

– Simulations d'un sondage ; à l'issue de nombreuses simulations, pour des échantillons de taille variable, on pourra introduire la notion de fourchette de sondage, sans justification théorique. La notion de niveau de confiance 0,95 de la fourchette peut être introduite en terme de « chances » (il y a 95 chances sur 100 pour que la fourchette contienne la proportion que l'on cherche à estimer) ; on pourra utiliser les formules des fourchettes aux niveaux 0,95, 0,90 et 0,99 pour une proportion observée voisine de 0,5 afin de voir qu'on perd en précision ce qu'on gagne en niveau de confiance. On incitera les élèves à connaître l'approximation usuelle de la fourchette au niveau de confiance 0,95, issue d'un sondage sur n individus (n>30) dans le cas où la proportion observée $\hat{p}$ est comprise entre 0,3 et 0,7, à savoir : $\left[\hat{p} - 1/\sqrt{n}; \hat{p} + 1\sqrt{n}\right]$.

– Simulations de jeux de pile ou face : distribution des fréquences du nombre maximum de coups consécutifs égaux dans une simulation de 100 ou 200 lancers de pièce équilibrée ; distribution de fréquences du gain sur un jeu d'au plus dix parties où on joue en doublant la mise (ou en la triplant) tant qu'on n'a pas gagné. On pourra aussi faire directement l'expérience avec des pièces pour bien faire sentir la notion de simulation...

– Simulation du lancer de deux dés identiques et distribution de la somme des faces. On pourra aussi faire directement l'expérience avec des dés pour bien faire sentir la notion de simulation...

– Simulations de promenades aléatoires sur des solides ou des lignes polygonales, fluctuation du temps et estimation du temps mis pour traverser un cube, ou pour aller d'un sommet donné à un autre sommet donné d'une ligne polygonale.

– Simulation de naissances : distribution du nombre d'enfants par famille d'au plus quatre enfants lorsqu'on s'arrête au premier garçon, en admettant que pour chaque naissance, il y a autant de chances que ce soit un garçon ou une fille.

Dans ces nouveaux programmes, le chapitre sur les statistiques prend une place importante, de par sa proportion, par rapport aux autres chapitres, mais aussi par l'introduction de thèmes nouveaux, qui sont la simulation et la fluctuation d'échantillonnage.

Il faut cependant remarquer que la partie statistique descriptive, même si elle n'introduit pas de concepts nouveaux, est accompagnée de commentaires qui précisent l'esprit dans lequel il faut les enseigner :" *L'objectif est de faire réfléchir les élèves sur la nature des données traitées, et de s'appuyer sur des représentations graphiques pour justifier un choix de résumé."*

Nous pensons que la partie "calcul" attachée aux statistiques, si elle est nécessaire, n'est pas la plus importante.

- Un calcul de moyenne arithmétique, s'il permet de voir des propriétés comme la linéarité ou est l'occasion d'introduire le signe somme, n'est pas suffisant pour donner tout son sens à cette moyenne. Cependant la comparaison sur certaines séries de la médiane et de la moyenne arithmétique, permettra de donner plus de sens à ces paramètres. De la même façon, il peut être utile de montrer aux élèves que la moyenne n'est pas toujours la moyenne arithmétique, mais peut aussi être la moyenne harmonique, ou une autre.
- Une représentation graphique ne prendra son intérêt que si elle est comparée à une autre, ou si son choix est justifié. De même, il faut être conscient des limites quant à l'interprétation de données à partir d'un graphique.

Nous proposons quelques activités qui permettent de réfléchir sur des données statistiques, sur la pertinence de certains résumés, sur l'interprétation de certains paramètres.

Activité 1

Interprétation de données statistiques
ou comment faire dire aux chiffres ce que l'on a envie qu'ils disent

Exercice 1

1300 candidats se présentent à un examen qui comporte deux options.
On observe les résultats suivants :

Reçus	Garçons	Filles
Option A	400/500	90/100
Option B	23/200	85/500
Total	423/700	175/600

- Les filles ont-elles mieux réussi a l'examen?

Remarque : *La question ouverte doit permettre, après les calculs de pourcentages de réussite, un débat sur les résultats, apparemment contradictoires, entre le pourcentage de réussite global, puis par option.*

Exercice 2

Nombre d'élèves dans les grandes écoles :

Origine	1950	1993
Classe populaire	25%	9%
Classe intellectuelle	60%	80%

- Que penser de l'évolution ?
- Quelles sont vos conclusions si on tient compte du tableau suivant ?

Répartition dans la population :

% de la population	1950	1993
Classe populaire	80%	60%
Classe intellectuelle	5%	20%

1950 : Pop Q_1 = 25/80

Int Q_2 = 60/5 Q_2/Q_1 =40

1993 : Pop Q_1 = 9/60

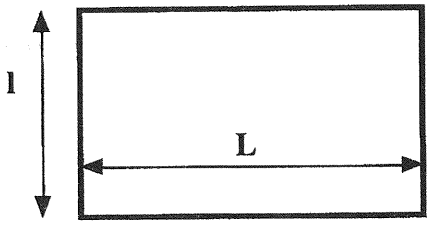
Int Q_2 = 80/20 Q_2/Q_1 =25 Il y a amélioration.

Activité 2

Le mot "moyenne" ne correspond pas toujours à la moyenne arithmétique

Exercice 1

On considère un rectangle de longueur L de largeur l :



Quel est la longueur du coté "c" d'un carré :

- Si le carré a même périmètre que le rectangle ?
- Si le carré a même aire que le rectangle ?
- Si le carré a même diagonale que le rectangle ?

Exercice 2

On considère deux carrés :
le premier de 2 cm de côté,
le deuxième de 4 cm de côté.

- Quel serait le "carré moyen" ?

Remarque : *Cette activité peut permettre de réfléchir sur la notion de moyenne, qui dans certains cas ne peut être définie.*

Activité 3

Comment employer les différents paramètres de position

Exercice 1

Trois élèves André, Jean et Paul ont obtenu les mêmes notes , à savoir :

2 - 2 - 6 - 8 - 10 - 12 - 12 - 12 - 17

André rentre chez lui et dit "*En moyenne j'ai 12*"

Jean rentre chez lui et dit "la *moyenne de mes notes est 9*"

André rentre chez lui et dit "*ma note moyenne est10*"

- Que pensez-vous des réponses des trois élèves ?

Exercice 2

Un cycliste gravit un col à la vitesse de 15 km/h.

Il redescend par le même chemin à la vitesse de 45 km/h.

- Quelle est sa vitesse moyenne sur l'ensemble du trajet?
- Quelle devrait être sa vitesse de descente s'il veut une vitesse moyenne sur l'ensemble du trajet de 30 km/h ?

Exercice 3

La population d'une ville est de 80000 habitants au 1/1/1998

Elle augmente de 6% en 1998.

Elle augmente de 10% en 1999.

- Quelle est la population au 1/1/2000 ?
- Quel est le taux moyen d'augmentation sur les deux années ?

Activité 4

Les limites des résumés statistiques

Exercice 1

On considère la série de notes suivante :

1-2-2-3-4-4-5-7-8-9-10-11-12-13-14-15-16-17-18-18-19

- Calculer la moyenne et la médiane
- Calculer la moyenne et la médiane de la série formée des 10 premières valeurs.
- Calculer la moyenne et la médiane de la série formée des 11 dernières valeurs.
- A partir des résultats des deux séries précédentes, peut-on retrouver la moyenne et la médiane de la série initiale?

Exercice 2

Deux élèves E_1 et E_2 ont obtenus les notes suivantes :

E_1 : 2 - 6 - 8- 10 - 12 - 14 - 18.

E_2 : 7 - 8 - 9 - 10 - 11 - 12 - 13.

- Pour les deux séries comparer la médiane et la moyenne.
- Par quel critère peut-on distinguer les deux séries de notes ?

Activité 5

Quelle exploitation peut-on faire d'un diagramme

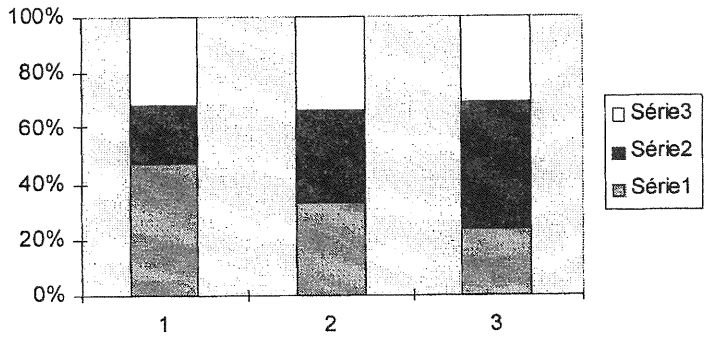
1. Le service, le revers, ou le coup droit pour 5 joueurs :

	Services gagnants	Coups droits	Revers
Joueur 1	25	40	32
Joueur 2	46	32	18
Joueur 3	17	45	30
Joueur 4	29	28	32
Joueur 5	14	17	45

- Donner une représentation graphique sous les trois formes suivantes :
  - Diagrammes en bâtons "juxtaposés".
  - Diagrammes en bâtons "cumulés".
  - Diagrammes en bâtons "ramenés à 100%".
- Quelles informations pouvez-vous tirer de chaque graphique?

Pour 3 joueurs, 1, 2, 3, on dispose du diagramme suivant (ramené a 100%)

- La série 1 correspond aux "services gagnants".
- La série 2 correspond aux "coups droits"
- La série 3 correspond aux "revers".



- Que peut-on dire de ces trois joueurs ?
- Quelles précisions supplémentaires peut-on apporter si on sait de plus qu'ils ont tous trois gagné 20 balles sur 100 au service ?

Remarque : *L'objectif de l'activité est de faire prendre conscience aux élèves des limites d'un graphique quant à ses interprétations et aux conclusions que l'on peut en tirer.*

LES ENJEUX DE L'ENSEIGNEMENT DE LA STATISTIQUE ET DES PROBABILITES AUJOURD'HUI

Michel HENRY

IREM de Franche-Comté

Résumé : *Les programmes de statistique des années 2000 des classes de seconde, premières et terminales proposent de faire un pas en avant dans la compréhension par les élèves des démarches statistiques. Cependant, ils ne prennent pas suffisamment en charge l'objectif d'un enseignement conjoint et coordonné de la statistique et des probabilités, réduisant celles-ci à un appui minimum et non théorique aux démarches de l'observation statistique dans le cas d'épreuves répétées. Malgré cela, l'objectif d'un enseignement de la modélisation probabiliste est clairement affirmé, pour des situations restreintes à des exemples issus de jeux de hasard où l'équiprobabilité peut être clairement postulée. Cette option novatrice ne va pas sans poser de nombreux problèmes didactiques dont les réponses ne sont pas évidentes. La mise en œuvre des situations de classes proposées en accompagnement des programmes, ainsi que les travaux de recherche menés notamment au sein des IREM, contribueront à faire avancer la réflexion. Cet article tente de faire le point sur les composantes épistémologiques et didactiques des questions posées par cet enseignement.*

1. Leur place dans la production scientifique contemporaine

La théorie des probabilités occupe aujourd'hui une place essentielle dans les mathématiques contemporaines où les publications de recherche occupent une place prépondérante. L'outil statistique est de plus en plus performant, même surprenant : prises de décisions, contrôles fins, gains de production importants, sondages, connaissances inaccessibles directement, poids de différents facteurs dans les comportements de systèmes complexes (analyse factorielle). Le développement de la statistique théorique se nourrit des problématiques issues de la réalité et bénéficie des avancées de la recherche en probabilités : utilisation des savoirs probabilistes en vue des applications (analyse, théorie des nombres, codage), production de théorèmes. On a assisté au XX^{ème} siècle, surtout à partir de 1950, à un foisonnement de la théorie statistique. Certains résultats enseignés aujourd'hui étaient secret militaire il y a 30 ans.

Ces connaissances et techniques sont-elles accessibles à l'éducation au-delà des pratiques professionnelles souvent empiriques ? Quelles transpositions didactiques, à quels niveaux ? Quels rapports entre connaissance empirique ou pratiques empiriques en technologie et savoirs théoriques dans l'enseignement ? La mise à jour dans l'enseignement de la statistique devient aussi importante que la maîtrise de l'outil informatique dans les années 90. Ces deux domaines ne sont d'ailleurs pas indépendants et la puissance des NTIC[1] contemporaines permet un enseignement de la statistique complètement rénové.

[1] Nouvelles Technologies de l'Information et de la Communication.

2. Statistique ou Probabilités ?

2.1. La Statistique

On a tendance à appeler "statistique" un vaste domaine à la fois de connaissances empiriques et théoriques, de techniques de maîtrise de l'information et d'aides à la prise de décisions relatives à de vastes populations. En globalisant de la sorte on oublie ce qui fait la spécificité de la statistique comme outil de description de la réalité, on tend à confondre l'aléatoire avec le déterminé, même si ce dernier est trop vaste ou complexe pour permettre une approche directe. La pensée probabiliste reste essentielle pour la compréhension des phénomènes socio-économiques par exemple, pour comprendre aussi les fondements théoriques de la décision statistique. La formation ne peut pas faire l'impasse sur la dualité entre statistique et probabilités.

La statistique, traditionnellement science des populations, est constituée de grands domaines de savoirs et de techniques. Très schématiquement, j'en dégage deux :

- Des outils de la description des populations, qui donnent lieu à diverses pratiques empiriques, issues d'une pratique sociale et pas nécessairement d'une compréhension scientifique : les programmes des collèges, et en partie de la seconde, introduisent un vocabulaire et des modes de représentations élémentaires relatifs à « l'organisation et la gestion des données ». D'autres outils beaucoup plus sophistiqués et performants comme l'analyse des données ou l'analyse en composantes principales utilisent des connaissances mathématiques de haut niveau pour obtenir des informations plus difficiles à atteindre sur des populations connues exhaustivement.
- L'inférence statistique qui permet de dégager la meilleure information d'une connaissance partielle d'une population statistique, qui donne des outils performants de traitement de cette information et qui peut indiquer la confiance à attribuer à une décision. Toute déclaration au sujet d'une population est entachée d'un risque de se tromper dû aux aléas de l'échantillonnage, c'est-à-dire d'une probabilité que les informations dégagées de cette connaissance partielle soient trop éloignées de la réalité de la population étudiée.

2.2. Les probabilités

La notion de probabilité est une notion abstraite au caractère dual : calcul des chances (calcul a priori) et approche fréquentiste (valeur expérimentale a posteriori). La théorie des probabilités ouvre un champ conceptuel spécifique. Elle a en vue la modélisation mathématique des situations de hasard (expériences aléatoires), et donne lieu à des développements théoriques de haut niveau. Le savoir probabiliste est une connaissance de type mathématique, organisée en théorie : concepts de base, définitions, axiomes, propriétés abstraites et théorèmes. Ces connaissances sont conçues pour opérer sur des situations modèles afin de déterminer les évolutions de processus où le hasard intervient.

La notion de probabilité est une construction de l'esprit humain, qui a fondamentalement un caractère abstrait. Selon les approches, différentes appréhensions sont possibles :

- D'une part, ce que PASCAL appelait la « géométrie du hasard » : la probabilité est conçue comme calculable a priori par dénombrement des issues « favorables », avant la réalisation d'une épreuve, en tenant compte des symétries autorisant des hypothèses d'équiprobabilité posées a priori.
- D'autre part l'appréhension fréquentiste validée par BERNOULLI, qui explique qu'on peut avoir une approche de la probabilité (au sens précédent) d'un événement par une mesure de la

fréquence stabilisée des réalisations de cet événement au cours de la répétition, un grand nombre de fois, de l'épreuve.

Ces deux approches vont se nourrir l'une l'autre et vont poser du point de vue didactique des questions de coordination et de relation dialectique. La question de la relation entre cet outil théorique et l'étude des situations où le hasard intervient dans un contexte concret se pose aujourd'hui au niveau de l'enseignement secondaire.

La question de la modélisation est donc fondamentalement posée.

Voici deux extraits pour illustrer ces propos :

- **Pierre Simon LAPLACE** dont l'œuvre (la *théorie analytique des probabilités, 1812*) me le fait considérer comme « l'Euclide des probabilités », a fait la synthèse de tous les travaux produits depuis 150 ans et a poussé la théorie aussi loin que le lui permettaient les outils mathématiques de l'époque, qu'il a contribué à développer. Dans sa préface, qui constitue *l'essai philosophique sur les probabilités,* il écrit :

  "*La probabilité est relative en partie à notre ignorance, en partie à nos connaissances*".

C'est une prise de position épistémologique : la probabilité dans les applications serait ainsi de nature fondamentalement **subjective**. Il ajoute :

"*La théorie des hasards consiste à réduire tous les événements du même genre à un certain nombre de cas également possibles, c'est-à-dire tels que nous soyons également indécis sur leur existence*".

Et il conclut à la définition de la probabilité, son *premier principe*, comme :

"*Le rapport du nombre des cas favorables à celui de tous les cas possibles.*"

- **Alfred RENYI**, probabiliste hongrois, auteur de manuels universitaires dans les années 60, écrit :

"*Nous appellerons probabilité d'un événement le nombre autour duquel oscille la fréquence relative de l'événement considéré*".

Cette définition est totalement différente de celle de LAPLACE. Comment gérer cela dans l'enseignement secondaire ? Il précise, adoptant ainsi une position **objectiviste** :

"*Nous considérons la probabilité comme une valeur indépendante de l'observateur qui indique approximativement* [car considérée comme une mesure de grandeur] *avec quelle fréquence l'événement considéré se produira au cours d'une longue série d'épreuves*".

Ce point de vue fonde donc sa « définition » sur une longue série d'épreuves réalisées et non sur une seule expérience à venir. RENYI précise :

"*Une théorie mathématique des probabilités* [entendez celle de KOLMOGOROV] *ne s'occupe pas de jugements subjectifs. Elle concerne les probabilités objectives, qui peuvent être mesurées comme des grandeurs physiques. La définition ci-dessus n'est pas une définition mathématique, c'est une description du substrat concret du concept de probabilité.*"

L'enseignement des probabilités doit intégrer la dualité de ce concept et rester ouvert aux deux appréhensions :

- subjectiviste reposant sur l'évaluation a priori « des chances » par le probabiliste (géométrie du hasard), soit qu'il postule l'équiprobabilité quelque part (point de vue laplacien), soit qu'il fonde ses calculs sur le choix d'une loi qu'en tant que spécialiste il pressent et contrôle (démarche bayésienne),
- objectiviste se basant a posteriori sur l'expérience renouvelée et sa description statistique, pour estimer une probabilité par une fréquence stabilisée (théorème de BERNOULLI), ou ajuster une répartition par une fonction de fréquences cumulées (théorème de GLIVENKO-CANTELLI), appliquant un principe de « loi des grands nombres » (approche fréquentiste).

Profitons d'avoir invoqué la loi des grands nombres de BERNOULLI pour faire une petite remarque historique. Ce théorème dit que :

la probabilité que la fréquence d'une issue observée dans une expérience « Succès-Echec » répétée s'écarte de la probabilité théorique de cette issue de plus d'un ε donné, tend vers 0 quand le nombre d'expériences tend vers l'infini.

Ce résultat crée la dualité fondamentale de la notion de probabilité. Jacques BERNOULLI démontre ce théorème à l'entrée du XVIII^{ème} siècle, d'une manière éblouissante, introduisant le langage des limites bien avant Cauchy, après avoir constaté les limites (!) d'un raisonnement à base de calcul infinitésimal.

3. Enjeux et défis pour l'enseignement secondaire

Le problème didactique dans le secondaire est donc de faire le lien entre l'**observation statistique** de la réalité et les résultats issus d'une **théorie mathématique** dans laquelle la probabilité est introduite, selon les choix de transposition :

- par une gestion combinatoire de la définition de LAPLACE ;
- ou bien, ainsi que l'indique RENYI, comme fréquence stabilisée ;
- ou encore, selon la théorie de KOLMOGOROV enseignée à l'université, comme mesure positive bornée définie sur un ensemble abstrait.

3.1. Questions didactiques posées par l'enseignement des probabilités

A l'université, particulièrement en second cycle, l'enseignement des probabilités ne soulève pas de difficulté, c'est une théorie mathématique comme une autre.

Dans le secondaire, cet enseignement est plus délicat. L'objectif est la maîtrise de concepts de base en vue de leur opérationnalité. On est alors amené à distinguer l'observation du réel (fréquence stabilisée) de la notion de probabilité. Celle-ci est introduite depuis 1991 par l'observation de la stabilisation des fréquences, mais est définie autrement : c'est la somme des probabilités des événements élémentaires qui constituent l'événement considéré.

Le problème est de gérer la distance entre l'introduction d'une notion pour sa compréhension opératoire et la définition théorique qu'on va en donner.

Autre problème didactique que posait le programme de 1991 : la probabilité peut-elle être donnée comme valeur limite de la suite des fréquences observées ? Ce serait une erreur profonde du point de vue épistémologique, qui produit comme obstacle majeur la confusion entre réalité et modèle théorique.

Ainsi, dans la tête des élèves, le mot « probabilité » revêt plusieurs significations. C'est, suivant les niveaux :

- un nombre relatif de chances (appréciation subjective du réel : « quelles chances j'ai de... »),
- un rapport déterminé a priori (nombre de cas favorables sur nombres de cas possibles),
- une valeur limite, estimée à partir de l'observation des fréquences d'un événement dans une épreuve répétée,
- une application (distribution ou loi de probabilités) définie sur les parties d'un espace abstrait (l'univers Ω) qui donne comme image un nombre compris entre 0 et 1,
- une mesure sur un ensemble numérique, définie par une densité (loi) par rapport à la mesure uniforme.

Comment préparer les élèves à cette diversité de sens donnés au même concept ? Comment les relier aux problèmes concrets posés dans les applications ? Comment intégrer les investigations à la fois pratiques (données statistiques) et théoriques (risque d'une décision) dans une compréhension probabiliste dont l'exemple prototypique est la notion d'intervalle de confiance, subrepticement introduite dans les thèmes d'étude de la classe de seconde de l'an 2000 ?

3.2. Quels rapports entre Statistique et Probabilités ?

La symbolique mathématique est particulièrement adaptée à la description statistique (variables indicées, représentations graphiques), et à la manipulation des résumés statistiques (moyennes, médianes, dispersion ...). Les mathématiques contemporaines (combinatoire, analyse, algèbre linéaire...) sont performantes pour, à partir de données complexes et très nombreuses, extraire des informations inaccessibles directement : corrélations, études de tendances, analyse factorielle ou en composantes principales, analyse des données... Mais ces techniques sont actuellement hors de portée de l'enseignement secondaire.

Pour étudier une population inaccessible globalement (trop vaste, trop complexe, trop coûteux...), on procède par étude d'échantillons (pas nécessairement représentatifs !), prélevés **au hasard** (avec quel générateur de hasard ?) dans la population. Ce hasard est la condition pour pouvoir appliquer les résultats théoriques du calcul des probabilités et obtenir des théorèmes sur les comportements des échantillons en fonction des caractéristiques de la population mère et dégager la possibilité d'**inférences**.

Car il n'y a pas de probabilités s'il n'y a pas de hasard !

A partir de là, les connaissances probabilistes vont être omniprésentes dans le traitement des échantillons aléatoires : lois théoriques pour ajuster des distributions de caractères quantitatifs, connaissance des propriétés des résumés statistiques, comportements asymptotiques, théorèmes limites pour valider des inférences, **nature probabiliste** de toute affirmation concernant la population.

Sur le plan historique, la (ou les) statistique(s) a précédé toute idée ou concept de probabilité. On faisait des statistiques à Babylone, chez les Égyptiens, chez les Romains... Les outils de la

description statistique ont dû être développés au moyen âge. Au XIII^{ème} siècle, avec les transports maritimes, on a dû créer des compagnies d'assurance intéressées par les données statistiques et motivées par la prévision. Mais on n'avait pas de mathématiques pour y voir clair dans ces recueils de données.

Au collège, en France, ces outils de la description statistique servent aux élèves à disposer d'un vocabulaire et de méthodes de représentation qui leur permettent d'interpréter des observations sur de vastes populations. Mais, comme au XIII^{ème} siècle, il n'y a pas de hasard au collège, c'est une question complètement exclue des programmes actuels, donc pas de probabilité. Cela n'empêche pas les élèves d'avoir une certaine pratique du hasard, car ils vivent continuellement des situations où il intervient.

Si l'on souhaite que l'outil statistique permette de porter des jugements sur des populations inaccessibles globalement, il faut se familiariser avec l'échantillonnage aléatoire consistant en prélèvements au hasard dans cette population, afin de faire fonctionner les connaissances de nature probabiliste dont on dispose. Le lien entre statistique et probabilités va se situer à cet endroit : mieux connaître les populations statistiques inaccessibles globalement ou systématiquement, seulement connues au travers d'échantillons aléatoires. C'est ce caractère aléatoire qui, par applications de résultats probabilistes, permet de prendre les décisions les plus performantes.

Jusqu'à présent, toute étude de l'inférence statistique était exclue de l'enseignement secondaire (avant le bac) où l'on avait d'un côté des outils de description statistique et, d'un autre côté, de façon complètement déconnectée, une définition théorique de la probabilité (somme de probabilités élémentaires), susceptible de fonctionner dans des situations artificielles d'événements qui se combinent logiquement et simplement pour obtenir des résultats issus de calculs combinatoires. Dans ce contexte, il n'y a aucune raison de rencontrer les applications à la réalité et de faire la connexion entre statistique et probabilités. Mais, aujourd'hui, se pose la question de l'introduction de cette pratique sociale dans l'enseignement, de maîtriser des situations d'inférence statistique et de décision à propos de populations. Se pose alors un nouveau problème didactique complètement occulté jusqu'à présent : établir cette liaison entre statistique et probabilités.

C'est la tentative du programme de seconde de l'an 2000 et des programmes de premières publiés en août 2000 et ceux de terminales en préparation. Le lecteur comprendra que je suis totalement acquis à cet objectif-là. Cela ne m'empêche pas d'émettre sur le plan didactique de nombreuses remarques ou critiques à l'encontre de certains choix de transposition et de formulations.

3.3. Un exemple historique de lien entre statistique et probabilité

Mais, avant de passer à ces remarques, à titre de TP, il m'a semblé intéressant de proposer l'exemple des *Tables de mortalité et espérance de vie* de Christiaan HUYGENS. Ce travail des frères HUYGENS[2], notamment Christiaan, bien connu comme physicien, est le premier exemple historique du passage de l'observation statistique à la création d'un concept probabiliste.

---

[2] Correspondance des frères HUYGENS (1669), cf. Bernard PARZYSZ : atelier APMEP B19 de Marseille (1997), bulletin APMEP n° 416.

Il faut dire ici quand est née la notion de probabilité : c'est en 1654 dans la correspondance entre PASCAL et FERMAT[3]. Christian HUYGENS fréquentait PASCAL et, en 1657, il publiait un petit traité[4] d'interprétation des principes que PASCAL avait avancés dans cette correspondance. La notion d'espérance mathématique est notamment présente explicitement chez HUYGENS. En 1669, HUYGENS, qui a donc un peu maîtrisé la notion d'espérance mathématique, va chercher à conceptualiser et à théoriser l'observation des tables de mortalité de John GRAUNT qui avaient été dressées à Londres durant le siècle précédent.

Huygens pose un axiome :

"*Dans un jeu, la chance qu'on a de gagner quelque chose a une valeur telle que si l'on possède cette valeur on peut se procurer la même chance par un jeu équitable* "

Puis il énonce des « théorèmes » :

- Théorème 1 : "*Avoir des chances égales d'obtenir a ou b me vaut (a+b) / 2*".

- Théorème 3 : "*Avoir p chances d'obtenir a et q chances d'obtenir b, les chances étant équivalentes, me vaut (pa + qb) / (p+q).*"

C'est la formule de l'espérance mathématique.

Disposant de ces définitions théoriques, HUYGENS les applique à l'étude des tables de mortalité :

"*À Londres, de 100 personnes conçues, il en meurt* :

36 au bout de 6 ans,	9 entre 26 et 36 ans,	3 entre 56 et 66 ans
24 entre 6 et 16 ans,	6 entre 36 et 46 ans,	2 entre 66 et 76 ans
15 entre 16 et 26 ans,	4 entre 46 et 56 ans,	1 entre 76 et 86 ans."

C'est l'observation statistique. HUYGENS l'interprète en disant :

"*qui gagerait qu'un enfant conçu vivrait jusqu'à 6 ans peut mettre 64 contre 36*".

Il probabilise la situation :

"*donc un enfant conçu a 36 chances pour vivre 3 ans* (centre de la classe [0 ; 6]),

24 chances pour vivre 11 ans, 15 chances pour vivre 21 ans, etc."

Et il conclut :

"*Donc par ma règle des jeux de hasard, il faut multiplier chaque nombre des chances par les ans qu'elles donnent et diviser la somme des produits par la somme de toutes les chances.*"

Ce quotient est l'espérance de vie.

Résumons ces quelques objectifs didactiques qui me semblent se dégager des remarques précédentes, pour un enseignement rénové, du collège à l'université :

- maîtriser les outils de la description statistique,
- pouvoir analyser et décrire des situations familières où le hasard intervient,
- introduire assez tôt la notion de probabilité dans ses différentes appréhensions,
- comprendre et mesurer les effets du hasard dans les fluctuations d'échantillonnage,
- traduire en termes probabilistes des situations simples de la réalité (modélisation),

---

[3] C'est très rare qu'on puisse désigner aussi précisément l'acte de naissance d'un concept mathématique.

[4] *De ratiociinis in ludo aleae.*

- introduire aux principes et méthodes des investigations statistiques,
- donner des outils théoriques (concepts de base et raisonnements),

en vue de valider les démarches en statistique ou de porter des appréciations critiques sur les interprétations de leurs résultats (sondages par exemple), ces différents objectifs peuvent être résumés en un seul :

Renforcer les liens entre méthodes statistiques et théorie des probabilités.

4. Un état des lieux dans l'enseignement secondaire français

4.1. La situation avant la réforme des années 2000

Jusqu'à présent, en France, on peut remarquer un grand cloisonnement entre la description statistique et l'initiation aux probabilités.

- En collège, on aborde successivement :
  - l'organisation et la gestion des données,
  - les notions de populations, caractères, séries statistiques, séries classées,
  - les outils de la description: effectifs, fréquences, représentations graphiques et diagrammes,
  - les résumés statistiques, médianes, moyennes et une idée de la dispersion.

Il n'y a pas de hasard au collège.

- Au lycée (1991-1999), on avait la progression :
  - En seconde : synthèse des outils de la description statistique.
  - En première : notion de probabilité, appréhension fréquentiste et définition théorique. Probabilité de l'événement contraire et de la réunion de deux événements incompatibles. Cas de l'équiprobabilité.
  - En terminale : probabilités conditionnelles, arbres probabilistes, variables aléatoires et notion de loi d'une variable aléatoire.

Pas d'applications concrètes, pas de liens avec l'observation statistique.

4.2. Évolution des programmes[5]

Le calcul des probabilités a été introduit dans l'enseignement secondaire en 1954 dans la série TE (technique économique), il passe dans la série classique D en 1960 :

- 1960 : *Calcul des probabilités par dénombrement des cas, combinatoire dans les situations simples d'équiprobabilité (jeux de hasard).*

---

[5] Cf Bernard PARZYSZ, Les probabilités et la statistique dans le secondaire d'hier à aujourd'hui, *Enseigner les probabilités au lycée*, livre de la Commission Inter-IREM Statistique et Probabilités, 1997 (cf. bibliographie en fin de brochure).

On remplace un problème de probabilité par un problème de combinatoire (c'est une mine impressionnante d'exercices du fait de la complexité de certaines situations de dénombrements). Pas de notion de loi.

- 1971 : *Espaces probabilisés, Ω fini et répartition uniforme* pour se placer encore dans les cas d'équiprobabilité, *combinatoire.*

On explicite le modèle probabiliste abstrait (Mathématiques modernes), mais en réalité, il n'y a pas de changement de point de vue, seulement l'introduction d'une structure mathématique et du formalisme pour se ramener encore à l'équiprobabilité dans les exercices.

- 1982 : *Expérience aléatoire. Ensemble des issues, calcul des événements, situations d'équiprobabilité.*

Un virage est négocié avec l'introduction du mot « expérience aléatoire », à la base de l'enseignement des probabilités. Pour un probabiliste, c'est un terme primitif qu'on ne définit pas. Mais, pour parler d'expérience aléatoire aux élèves du lycée, on est placé devant un problème de présentation et l'on s'aperçoit que ce n'est pas simple du tout. Malgré cette approche un peu appliquée, on en reste au calcul des probabilités par situations d'équiprobabilité.

- 1991 : *Expérience aléatoire, description, répétition, interprétation probabiliste de la stabilisation des fréquences. Définitions et propriétés de base* (en première).

Probabilité conditionnelle, notion de loi (en terminale).

Malgré l'approche fréquentiste préconisée pour introduire la notion, pas d'applications, sinon dans des exercices habillés de combinatoire ou d'application directe interne des formules de définition, pas de lien avec l'observation statistique. Les mots « hasard » et « expérience aléatoire » avaient été introduits en 1986 dans les programmes dans leur sens commun. En 1991, l'expérience aléatoire va intervenir comme outil d'introduction à la notion de probabilité. "*On étudiera la stabilisation des fréquences lors de la répétition d'une expérience aléatoire* ". La définition générale de la probabilité est donnée ensuite, sans limitation aux situations d'équiprobabilité.

Parenthèse historique : LAPLACE, dans son premier principe, posait :

"la probabilité est le rapport du nombre de cas favorables au nombre de tous les cas possibles."

Mais dans son deuxième principe, il écrit :

> "*Mais cela suppose les divers cas également possibles. S'ils ne le sont pas, on déterminera d'abord leurs possibilités respectives dont la juste appréciation est un des points les plus délicats de la théorie des hasards. Alors la probabilité sera la somme des possibilités de chaque cas favorable*".

LAPLACE donne donc la définition générale, celle que propose le programme de première de 1991. Remarquons sa difficulté d'expression, utilisant tour à tour les mots « probabilité » et « possibilité ». Cette définition donne l'impression (pas entièrement fausse) d'une circularité, définissant la probabilité par la possibilité. Ma thèse est que, sur le plan didactique, on ne peut sortir de ce cercle vicieux que par la notion de modèle : les probabilités élémentaires (les possibilités) sont des valeurs issues d'hypothèses de modèle, d'appréciations éventuellement subjectives ou d'estimations fréquentistes, et interviennent comme données dans le modèle probabiliste construit, validé ou non par ses applications.

Mais LAPLACE pose aussi le problème de l'inférence, il suggère en gros :

Si vous n'êtes pas en mesure de calculer a priori la probabilité par une réduction à des cas également possibles, vous devez la mesurer concrètement, par l'expérience. En quoi votre mesure va effectivement être proche de la valeur théorique que vous devez utiliser ? *"C'est un des problèmes les plus délicats de la théorie des hasards."* affirme -t-il dans son deuxième principe. C'est le problème de l'estimation statistique que LAPLACE n'ignore pas mais dont il dit qu'il est très difficile (problème résolu seulement au XX^{ème} siècle).

Fermons cette parenthèse historique pour revenir aux programmes actuels.

5. Le programme de seconde de l'an 2000

5.1. L'option d'une présentation perceptive et statistique des fluctuations d'échantillonnage

Le programme de seconde de l'an 2000, dans sa deuxième partie, est caractérisé par une approche purement expérimentale des fluctuations d'échantillonnage. Les élèves sont confrontés au fait que quand on observe un certain nombre de fois les résultats d'une expérience aléatoire répétée, on n'observe pas tout à fait la même chose à chaque fois : le hasard intervient. Ce programme propose l'introduction du hasard directement avec l'observation des distributions de fréquences d'un ensemble d'issues dans des expériences répétées par la comparaison d'échantillons. Mais le programme ne propose pas de faire du hasard et de la notion d'expérience aléatoire l'objet d'un travail explicite particulier.

Cette démarche n'est pas évidente pour tous les élèves. Je cite très souvent cette remarque qu'un élève de première nous a faite dans une classe, et qui n'en a pas démordu. On disait :

- "*On recommence la même expérience aléatoire*" (jet d'un dé).
- "*Mais non*", a-t-il rétorqué, "*vous n'avez pas recommencé la même expérience* !"
- "*Mais si, regarde, on jette le même dé, de la même manière...*"

À ce moment-là, on pensait que cet élève demandait de préciser les conditions de l'expérience, mais il s'agissait en fait de la confusion entre la réalité et le modèle, ce que l'élève pointait immédiatement :

- "*Mais non, vous n'avez pas recommencé la même expérience, la preuve, c'est que vous n'obtenez pas le même résultat* !".

À partir de cet exemple, vous comprenez que si vous ne parlez pas d'expérience générique (de modèle), vous ne pourrez pas empêcher cet élève de penser qu'on ne peut pas refaire la même expérience aléatoire. Cette remarque nous renvoie à la proposition faite en 93 pour le programme de la section L qui n'a pas abouti :

> *"Il s'agit d'aborder la notion de probabilité à partir de la fréquence mais on a choisi d'affiner l'explicitation du processus de modélisation... L'objet de cette partie de la formation est donc de faire découvrir en s'appuyant sur l'expérimentation numérique quelques notions qualitatives et quantitatives liées à la modélisation mathématique des phénomènes aléatoires. L'objectif est de dégager de façon qualitative, éventuellement critique, quelques notions probabilistes.*

Les notions probabilistes étaient donc appelées comme outils de cette modélisation des phénomènes aléatoires. Ce projet reste un projet sur lequel on peut encore travailler.

5.2. Objectifs, intentions du programme de Seconde, et commentaires

J'en viens à quelques remarques d'ordre didactique sur ce programme de seconde.

- **La partie statistique descriptive**

L'ancien programme de seconde proposait une synthèse des notions et du vocabulaire de base de cette statistique descriptive, prolongée dans certaines séries de premières et terminales par les outils d'étude de statistiques à deux caractères conjoints. L'acquisition de ces notions et leur utilisation critique sont essentielles pour la formation des futurs citoyens : lecture de tableaux, de graphiques, d'analyses économiques, sociologiques... Or, il apparaît que cet apprentissage n'est pas spontané, il s'appuie sur une bonne compréhension des pourcentages et de la proportionnalité, ainsi que sur une certaine pratique des articulations logiques (disjonction, conjonction, négation) mettant en jeu un peu du vocabulaire des ensembles. Il nous semblerait dommage de faire l'impasse sur cette partie des connaissances de base, qui sont loin d'être maîtrisées à l'entrée en seconde.

Le nouveau programme n'a pas ces ambitions, il se limite à une consolidation à mon avis insuffisante des outils de la description statistique (moyennes, médianes et classes modales), notamment sur la notion de dispersion, trop abusivement réduite à l'étendue d'une série statistique. L'introduction de l'intervalle interquartile et de sa représentation par boîtes à pattes, pourtant à la portée de la main, est renvoyée en première.

- **Échantillonnage par simulation informatique et réalité**

L'intention affichée dans la deuxième partie du programme est véritablement une **approche de type expérimental** de l'investigation statistique. Pour cela, on donne des outils : définition des **fréquences**, observation d'échantillons et des **fluctuations d'échantillonnage**, introduction de l'informatique pour se donner les moyens d'avoir des échantillons de taille suffisante pour observer quelque chose.

Jusqu'à présent, sans ordinateur, on arrivait à exhiber concrètement le phénomène de stabilisation d'une fréquence, quand tous les élèves de la classe s'y mettaient, on finissait par avoir des données en assez grand nombre. Mais, pour étudier plus globalement les comportements des distributions de fréquences, lorsqu'on réalise un millier d'expériences pour chaque échantillon, là l'ordinateur devient incontournable.

On pourrait dire beaucoup de choses sur le terme de « **simulation** » qui est utilisé dans le programme. Il l'est peut-être improprement si cette « simulation » se limite à générer des données sans rapport avec une expérience concrète dont on aurait implanté le modèle en informatique. Laissons cette question de côté. Mais, cette remarque ne disqualifie pas l'utilisation de l'ordinateur, qui peut être considéré comme un générateur puissant de hasard, même si ce n'en est pas un en réalité. D'ailleurs il est expliqué pour les enseignants que, certes, l'ordinateur ne produit pas de hasard proprement dit, mais qu'en réalité il donne des nombres « pseudo-aléatoires » en quantité faramineuse, imprévisibles et correctement équirépartis, ce qui permet de les considérer comme la manifestation du hasard. L'ordinateur est donc considéré comme une boîte noire génératrice de hasard.

Du point de vue pédagogique, la question est plus précisément : comment va-t-on manipuler l'aléatoire qu'il produit pour singer (plutôt que simuler) des situations de lancers de dés, de

tirages d'urnes etc. ? Se pose alors la question didactique du rapport avec la réalité. Comment interpréter les situations concrètes, y compris au niveau du collège, quand par exemple les élèves vont au super marché regarder les files d'attente aux caisses, ou qu'ils écoutent un bulletin météo ? Comment prendre en compte ce contact avec l'aléatoire dans la réalité ? Regarder l'aléatoire quand on « simule » le jet d'un dé (par principe non pipé) n'est pas plus porteur d'applications que l'ancien calcul des probabilités basé sur la définition de Laplace.

- **Introduction à l'aléatoire et à l'échantillonnage**

Concernant l'expérience de l'aléatoire, on peut remarquer qu'il est grand temps en classe de seconde de travailler sur un sujet qui semble tabou au collège : l'aléatoire dont les élèves sont familiers depuis longtemps dans leur vie quotidienne. Des expérimentations montrent que la variabilité des issues d'une expérience aléatoire est facilement acceptée, et dans les situations simples de « succès-échec », peut être modélisée par des tirages dans une urne de BERNOULLI dont le statut théorique ne semble pas poser de problème.

Mais, le programme propose d'introduire à l'aléatoire dans des situations complexes de répétitions multipliées d'expériences aléatoires à plusieurs issues, dont l'ensemble des observations constituent un « **échantillonnage** », lui-même répété pour mettre en évidence les « **fluctuations** » que présentent ces divers échantillons recueillis. Cela en excluant toute introduction explicite de l'idée de probabilité, interprétée implicitement sous le vocable flou de « chances », ou travestie sous le terme de « fréquence théorique » dans le document d'accompagnement.

La notion de « fluctuation » ne me pose pas de problème : c'est la mise en évidence de l'intervention du hasard dans la détermination des issues d'une expérience aléatoire. Par contre, le terme d'usage courant d' « échantillon » (prélèvement, qui peut éventuellement être aléatoire, dans une population statistique) ne correspond pas rigoureusement au sens introduit dans la locution « fluctuation d'échantillonnage » adoptée dans le programme. Celle-ci vise l'ensemble des données recueillies (au moyen éventuel d'une simulation informatique) à l'issue de la répétition d'une même expérience aléatoire. De quelles populations sont alors extraits ces « échantillons » ? Le raisonnement traditionnel des statisticiens considérant une donnée statistique comme la réalisation fictive d'un prélèvement au hasard dans une population (qui ici reste à préciser), n'est pas immédiat, ni pour les élèves, ni pour leurs professeurs en général. Il y a là une difficulté didactique, qui n'est certes pas insurmontable, mais qui montre que ce premier contact avec l'aléatoire en classe n'est pas des plus simples. Aborder l'aléatoire par l'observation des fluctuations d'échantillonnage me semble donc ambitieux pour un démarrage. A l'étape actuelle, en attendant que les futurs programmes de collèges intègrent cette familiarisation avec l'aléatoire, ne vaudrait-il pas mieux se limiter, en seconde, à l'observation des fréquences de réalisation **d'un événement** dans plusieurs séries d'épreuves, comme d'ailleurs les programmes des classes de première de 1991 le proposaient ?

Cela amène à s'interroger sur les conditions didactiques permettant de travailler sur la simulation d'expériences aléatoires **sans introduire la notion de probabilité** en tant que concept théorique. Quel lien en effet établir entre une simulation et l'expérience réelle sans passer par la référence au **modèle théorique** commun ? Je partage entièrement la critique faite à l'encontre des anciens programmes de premières sur le risque de faire passer la probabilité pour une fréquence limite. Or, n'est-ce pas multiplier ce risque que de proposer l'étude et même la comparaison des distributions de fréquences d'événements associés à une épreuve répétée, comme préfigurant la « distribution théorique » ou « loi » de probabilité, dans la démarche proposée par le nouveau programme de première ?

Je pense qu'il conviendrait de limiter la complexité de ces premiers pas : les travaux en didactique des probabilités montrent que la notion même d'**expérience aléatoire** nécessite un certain travail, faute duquel des confusions rédhibitoires sont monnaie courante. De ce point de

vue, il me semble essentiel de bien séparer dans le vocabulaire ce qui relève de l'**observation statistique** de ce qui permet la description d'un **modèle probabiliste interprétatif**. Quel sens donner par exemple à la locution « événement associé à une série statistique » ? Au passage, il ne me semble pas judicieux de réduire l'outil statistique à la seule description des issues d'une épreuve aléatoire répétée.

Le deuxième point marquant de cette partie du programme est l'introduction de la notion de **distribution des fréquences** des issues (puis d'un événement) associées à l'expérience, distributions variables d'une série d'épreuves à l'autre, sans que les élèves puissent disposer des outils de comparaison de ces distributions pour en tirer des conclusions transférables. Le professeur ne peut s'en tenir alors qu'à une **observation naturaliste** sans pouvoir faire appel à une connaissance de nature théorique (probabiliste) pour dégager une généralisation. Notons que le programme exclut toute étude de stabilisation des fréquences alors que, incontournable dans l'environnement informatique de ce cadre expérimental, elle donnerait du sens à cette observation et préparerait l'introduction de la notion de probabilité, repoussée en Première.

Nous sommes en présence d'un vrai problème didactique, rencontré par ailleurs en physique ou en économie :

- Comment introduire un regard de type scientifique sur le réel quand les connaissances déjà installées ne permettent pas de formuler en termes clairs et raisonnés des faits observés afin de les transformer en phénomènes ?
- Quel sens peut prendre une activité de simulation en l'absence de la notion de modèle ? Quelle signification et statut ont alors les nombres pseudo-aléatoires issus de la touche "Random" ?
- Comment éviter la « **leçon de choses** » à ce niveau déjà avancé des études secondaires, et comment présenter cette partie du cours de statistique aux élèves en termes de savoirs, condition didactique de sa pérennité ? Bref, peut-on délivrer un enseignement limité à l'expérimental et à sa description ?

Il y a là un vide didactique ingérable pour le professeur s'il ne peut faire appel à des concepts théoriques élémentaires lui permettant le **saut épistémologique** qui seul emmène les élèves vers **l'acquisition d'un savoir** au-delà d'une **révélation perceptive**.

Par exemple, comment éviter le mot « probabilité » quand on jette un dé ? Comment échapper à la présentation de la notion « d'expérience aléatoire » pour théoriser les situations de la réalité explorées ? Les expressions « hypothèses de modèle » (même si l'on s'en tient à l'équiprobabilité), « univers », « événements », « issues », « distribution de probabilité » sont autant d'outils théoriques qui permettent l'approche scientifique et le transfert de cette perception à des classes de situations. Je n'ai pas de réponse toute faite, les expérimentations en cours pourront donner des éléments d'appréciation pour, je l'espère, nuancer ces réserves.

Pour résumer, je partage pleinement l'appréciation portée par les membres de la commission de réflexion sur l'enseignement des mathématiques, présidée par J.P. KAHANE, qu'

> *"un bon enseignement de la statistique ne peut se concevoir sans une bonne maîtrise des probabilités. Mais de manière plus précise, c'est la dialectique « statistique-probabilités » qui permet de donner un sens profond aux belles activités proposées"* (communiqué n° 4 du 27/11/99).

- **Dérapages dans les thèmes d'étude**

Les objectifs du programme restent donc au niveau du perceptif, voire du rapport naïf à la réalité complexe, marquée par le refus du concept théorique de probabilité. D'où au niveau des thèmes d'étude des impasses didactiques contournées par cette locution dénuée de sens : "les chances de se tromper quand"... , avec cette ambiguïté portée par le terme de « chances », signifiant tantôt issues (favorables ou observables), tantôt probabilité. D'où la réduction des exemples proposés à des jeux de dés, pièces, cartes ... et à des situations simples d'équiprobabilité.

Le premier thème propose une introduction élémentaire aux sondages, ce qui vise un fait massif de nos sociétés développées. Il paraît évident que les élèves des lycées doivent pouvoir disposer des notions de base leur permettant d'avoir des éléments d'appréciation, face à un déferlement médiatique. La notion de « **fourchette de sondage** » peut être présentée comme un encadrement d'une mesure entachée d'erreur en physique, mais cela pose le problème du statut de la grandeur à mesurer. Cet encadrement est associé a priori à une certaine probabilité d'être correct, ou, si l'on veut éviter ce point de vue, à un certain **risque** d'observer une certaine proportion des fréquences recueillies en dehors de cet intervalle. Le parachutage de la formule $[p - 1/\sqrt{n} ; p + 1/\sqrt{n}]$, impossible à justifier au niveau des élèves (ce qui ne serait pas un drame) ainsi qu'au niveau des connaissances de base des enseignants (ce qui pose plus de problèmes), ne permet pas de cerner clairement sa signification et les conditions de son application.

Le problème didactique est : comment introduire une initiation aux sondages (pourtant souhaitable) et à leur interprétation un peu rigoureuse sans la notion d'**intervalle de confiance** et de la probabilité-risque qui l'accompagne : $P(F_n - \varepsilon < p < F_n + \varepsilon) = 1 - \alpha$?

Cette remarque nous ramène à la nécessité de la notion de probabilité pour clarifier le statut des fourchettes de sondage et leur connexion avec la notion d'intervalle de confiance enseignée dans certains BTS ou DUT, ainsi que dans les cours de Statistique de second cycle universitaire. Vu l'expérience acquise en post-bac, je reste réservé quant à la possibilité de faire bien comprendre ces questions aux élèves de seconde (indifférenciée).

Les autres thèmes d'étude proposés (promenades aléatoires) semblent avoir provoqué de l'incompréhension chez les enseignants. Intéressants par leurs prolongements possibles, ils me semblent plutôt convenir pour l'illustration de modèles probabilistes à un niveau post-bac que pour un enseignement de la statistique.

6. Remarques sur le programme de Première S et sur le projet de programme de Terminale

Les programmes de la filière S, après ceux de la série L, ont été soumis au débat au premier trimestre de 1999-2000. Les programmes des classes de première ont été adoptés par le CNP, à nouveau "*sans réserve*" et publiés au B.O. du 31 août. La Commission Inter-Irem *Statistique et Probabilités* avait fait part de son appréciation critique le 30 janvier 2000. L'essentiel des remarques qui suivent s'inspirent de cette appréciation. La partie *Contenus* du programme de première S est présentée en deux parties que je reprends d'abord succinctement, avec quelques premiers commentaires.

1 - La partie statistique complète celle du programme de seconde par l'introduction de paramètres de dispersion : variance et écart type d'une part, intervalle inter-quartile et « diagrammes en boîte » d'autre part. Dans la *mise en œuvre,* le programme suggère de comparer par l'observation expérimentale les comportements respectifs de ces deux paramètres statistiques et leur robustesse, quand les données varient sensiblement. En les associant aux paramètres de position correspondants que sont respectivement la moyenne et la médiane, les *commentaires* limitent l'objectif de l'étude descriptive d'une série statistique à ces deux couples de résumés. L'interprétation des paramètres empiriques comme estimateurs possibles des paramètres correspondants dans une population échantillonnée est donc hors programme.

2 - La partie probabilités des *contenus* s'ouvre directement par la *Définition d'une loi de probabilité sur un ensemble fini* (lequel ?).

Cette définition ne fait pas l'objet d'une proposition dans la *mise en œuvre,* mais fait l'objet d'une explicitation précise dans le document d'accompagnement (celui daté du 2 octobre 2000). Elle paraît de nature pragmatique, rapportée aux propriétés des *distributions de fréquences*

introduites en seconde, via *un énoncé vulgarisé* (sic) *de la loi des grands nombres.* Cet énoncé est donné dans les *commentaires.* Il va bien au-delà de l'énoncé traditionnel du théorème de BERNOULLI[6], car il englobe la convergence, repérée empiriquement et présentée sous la forme vague de " *se rapprochent* ", des distributions de fréquences vers la "*loi de probabilité*", puis des "*moyennes vers l'espérance et des variances empiriques vers les variances théoriques*".

Le programme indique ensuite : *espérance, variance, écart type d'une loi*, ce qui suppose pas mal d'implicites (notamment l'hypothèse que l'on se place dans un ensemble d'issues, non seulement numériques, mais dont le caractère quantitatif a du sens), sur lesquels je n'insisterai pas, ces notions étant reprises ensuite à propos des *lois de variables aléatoires*, d'où une certaine redondance, expliquée dans le document d'accompagnement du programme de terminale S en préparation. L'étude systématique de lois élémentaires (BERNOULLI et binomiales, par exemple) n'est d'ailleurs pas proposée en première, elle est renvoyée en terminale (projet du 8 janvier 2001).

Les propriétés de base des probabilités sont introduites avec le concept d' « événement », sans indication sur les définitions à institutionnaliser, dans une progression qui semble difficile à gérer sur le plan didactique, allant de la compréhension globale de la notion de loi de probabilité à celle plus élémentaire de probabilité d'un événement ! L'équiprobabilité est ensuite présentée comme un cas particulier, alors qu'elle est, dans une optique laplacienne, imposée sur le modèle de base représentatif des issues, éventuellement transformé par une variable aléatoire.

Le dernier paragraphe concerne la modélisation d'expériences aléatoires. Remarquons d'abord que la notion fondamentale d'*expérience aléatoire* est introduite à cet endroit sans faire l'objet d'indications particulières, limitée aux expériences de référence (les exemples donnés relèvent tous de l'équiprobabilité). Dans les commentaires, le programme précise : "*On indiquera que simuler une expérience consiste à simuler un modèle de cette expérience*", ouvrant ainsi la possibilité sur ce qu'il me semble essentiel : le travail sur la modélisation. Les modalités de mise en œuvre proposent la simulation d'une loi de probabilité image par une v.a. d'une loi équirépartie, ce qui suppose un sérieux recul théorique de la part des enseignants. La question du lien de cet enseignement avec l'aléatoire dans la réalité reste donc posée, avec celle de son articulation avec la partie statistique.

3 - Quelques commentaires épistémologiques et didactiques

- Sur la notion de probabilité

Une première constatation confirme la tendance relevée dans le programme de seconde : tout travail sur la notion de probabilité semble être exclu. L'introduction préconisée par le programme de 1991 est absente des objectifs affichés. Cependant, le statut théorique de la probabilité comme objet mathématique inscrit dans un modèle descriptif de la réalité, sous certaines hypothèses, est indiqué dans le document d'accompagnement. Ce document donne à l'intention des enseignants les définitions axiomatiques de la probabilité d'un événement ("*somme des probabilités des éléments qui le composent*") et de la loi d'une variable aléatoire (loi P' sur E' donnée par $P'(x') = P(T=x')$ où $x' \in E'$, image de E par la v. a. T).

Par contre, les axiomes fondateurs de cet objet mathématique (très simplement la sous-additivité) ne sont pas proposés aux élèves en tant que tels, ses propriétés découlant de remarques faites sur celles des fréquences d'une distribution, ce qui invalide la compréhension du statut théorique de

---

[6] Théorème de BERNOULLI (non vulgarisé) : *Dans une épreuve répétée de Bernoulli, la suite des fréquences de succès converge en probabilité vers la probabilité de ce succès*, i.e. : $\forall \varepsilon > 0$, $P(|F_n - p| < \varepsilon) \rightarrow 1$ quand $n \rightarrow \infty$.

Conséquence : pour ε assez petit (précision) et n assez grand (taille de l'échantillon), $]F_n - \varepsilon, F_n + \varepsilon[$ est un intervalle de confiance pour p de niveau de confiance $P(|F_n - p| < \varepsilon) = 1 - \alpha$.

cette notion et sa transformation en concept abstrait. Cette ambiguïté de statut épistémologique rend problématique, me semble-t-il, l'insertion de l'objet « probabilité » dans des démonstrations mathématiques.

- Sur la notion d'expérience aléatoire

De même la notion d'expérience aléatoire n'est abordée que d'un point de vue pragmatique, au travers d'exemples limités (jeux de hasard). Elle ne peut jouer pour les élèves le rôle de concept primitif autorisant une modélisation mathématique sous la forme d'un univers Ω. Bien sûr, il n'est pas question de spécifier Ω dans chaque exemple, mais l'univers joue un rôle théorique irremplaçable dans la compréhension de la formalisation de la notion d'événement. Bref, dans le programme de première, les bases de la théorie probabiliste, et par là même d'une connaissance de type scientifique, ne font pas l'objet d'un travail explicite. Dans ces conditions, le processus de modélisation qui à juste titre est mis en avant explicitement, semble devoir être limité aux "*expériences de référence amenant à une probabilité équirépartie*", c'est à dire à des situations limitées et contextualisées qui pourraient être ultérieurement sources d'obstacles didactiques.

- Sur l'équiprobabilité

Au delà de ce type d'obstacle, la limitation aux situations d'équiprobabilité porte en elle une question épistémologique de fond, déjà soulignée par J. BERNOULLI dans *Ars Conjectandi* et indiquée par P. S. LAPLACE dans son deuxième principe. Les situations réelles relèvent-elles nécessairement de l'équiprobabilité quelque part ? Évidemment non. D'où la limitation aux jeux de hasard, "*définis par leurs inventeurs pour se ménager l'équité*" selon BERNOULLI, du champ d'application des connaissances introduites dans ce programme. Cette limitation est-elle pertinente aujourd'hui où les problèmes de fiabilité, de files d'attentes, de modélisation d'une complexité (économique ou météorologique par exemple) sont l'objet même de cet apprentissage ?

Le programme de 1991 embrassait cette dimension sans difficulté, laissant aux hypothèses de modèle le soin de déterminer les probabilités en jeu. Les situations d'équiprobabilité étaient présentées comme des cas particuliers, même si cette hypothèse a l'avantage de permettre les calculs a priori dans un cadre combinatoire. La conséquence didactique prévisible du choix de transposition actuel est d'induire à nouveau chez les élèves l'idée que le calcul des probabilités n'est pas applicable à la réalité, se réduisant à d'aimables situations de dés ou urnes dans lesquelles on est conduit à des problèmes incontrôlables de dénombrement des cas.

- Sur la notion de loi de probabilité

La notion de loi de probabilité est proposée, sans que celle de probabilité fasse l'objet d'un travail préparatoire. On peut s'en étonner. Cette notion est présentée dans le document d'accompagnement comme une structure formelle jouissant des propriétés des distributions de fréquences, qui, au travers d'un énoncé "*vulgarisé*" de la loi des grands nombres, détermineraient cette loi limite. Outre le fait qu'un tel énoncé, même vulgarisé, ne me semble pas présentable en première, sauf à prendre de grandes libertés avec la rigueur mathématique, il me semble porter le risque d'assimiler ces probabilités à des fréquences limites, ce qui revient à confondre la réalité observable avec le modèle théorique dans lequel on obtient des théorèmes, malgré les mises au point judicieuses du document d'accompagnement.

Cette notion de loi de probabilité, qui est essentielle effectivement, mérite plus qu'une simple assimilation à une distribution de probabilités sur Ω. Comme pour la notion de fonction, introduite en seconde, un tableau de valeurs ne suffit pas pour rendre compte de la cohérence interne qui lie ces valeurs pour en faire des signifiants numériques des phénomènes modélisés. Pour donner du sens à cette notion, qu'il me semble préférable d'associer systématiquement et intrinsèquement à celle de variable aléatoire, il convient de la rapporter à des situations concrètes. On pourrait, par exemple, opter pour l'introduction es qualité de situations binomiales (qu'il faudrait anticiper par celles de BERNOULLI), et peut-être l'accompagner d'autres exemples

comme celui de la loi géométrique, très présente dans la pratique de tests industriels, malgré le cardinal théoriquement infini de son ensemble référentiel.

- Sur le projet de programme de terminale

Le projet d'introduire en terminale S des lois continues semble ambitieux, mais pas inaccessible, modulo la formation des enseignants nécessaire. Mais, suite au point de vue adopté en première, comment pourra-t-on expliquer qu'une variable aléatoire, transporteuse d'une telle loi de probabilité, puisse être définie sur un Ω fini (muni de plus d'une distribution d'équiprobabilité) et néanmoins prendre un ensemble continu de valeurs ($\mathbb{R}$ ou un intervalle [a, b]) ?

Les lois uniformes et gaussiennes (il nous semble préférable de les appeler « normales », dénomination de K. PEARSON, ne serait-ce que pour rendre à leurs inventeurs LAPLACE, MOIVRE et peut-être BERNOULLI ce qui leur revient en dimension 1) sont présentées par des définitions purement mathématiques en termes de calcul d'aires, sans que la relation puisse être faite avec la notion de loi d'une variable introduite par ailleurs, et sans que leurs domaines d'applications concrètes puissent être explicités.

On peut s'étonner à cet endroit que le programme propose de recourir aux anciennes tables de valeurs de la fonction de répartition de la loi normale, alors que les calculatrices ou logiciels contemporains les donnent sans difficulté.

Ces propositions, ambitieuses pour les connaissances en jeu, mais insuffisantes quant aux savoirs probabilistes fondamentaux qu'une introduction intuitive et superficielle ne saurait rendre opératoires, sont cependant séduisantes pour qui partage l'importance d'un enseignement modernisé de la statistique et des probabilités. Il reviendra aux enseignants d'expérimenter leur mise en œuvre dans la classe pour pouvoir pousser d'avantage l'élaboration d'une didactique du couple statistique-probabilités, intéressant l'ensemble des classes de collège et lycée.

7. Vers l'enseignement de la modélisation : quelle formation des enseignants ?

7.1. Former à la modélisation probabiliste

La nécessité d'un véritable décloisonnement entre statistique et probabilités se dégage clairement des remarques qui précèdent. Il n'est pas, à mon sens, vraiment réalisé avec la progression proposée par les programmes de seconde et de première et les objectifs annoncés pour celui de terminale. Ce décloisonnement suppose en fait une réforme globale, du collège au lycée puis à l'université. Intégrant les outils de la description statistique dans l'étude des grandes populations, cet enseignement introduirait les probabilités dans un processus de modélisation de situations de la réalité.

Cette approche expérimentale serait donc éclairée par l'introduction parallèle d'outils de description et de concepts théoriques simples :

- descriptions de populations statistiques, caractères et opérations logiques,
- situations où le hasard intervient et descriptions d'expériences aléatoires,
- prélèvements au hasard dans une population et expériences de Bernoulli (succès-échec),
- répétition de telles expériences, stabilisation des fréquences. Urne de BERNOULLI et pré-probabilité,

- modéliser le lancer d'une punaise (par exemple) : du modèle pseudo-concret au modèle mathématique,
- situations variées et modèles d'urnes, choix de modèles,
- schéma binomial, situations de sondages, principes de l'inférence statistique (nature probabiliste d'une décision statistique),
- distribution de fréquences et modèle probabiliste, le théorème de BERNOULLI,
- générateurs aléatoires, distribution uniforme (discrète puis continue),
- caractères quantitatifs et variables aléatoires, lois de variables discrètes,
- résumés statistiques et moments d'une loi, théorème de la statistique,
- simulation informatique et modélisation. Probabilités géométriques (jeu du Franc-Carreau par exemple), modèles à lois continues,
- conflits de modèles et paradoxes (D'ALEMBERT, BERTRAND, ...),
- modéliser pour simuler. Résolution de problème par simulation : files d'attentes...

Ces objectifs n'ont pas de caractère exhaustif, ils ne sont pas opérationnalisés et sont de l'ordre de l'utopie. Ils montrent néanmoins quelles pourraient être les étapes d'un enseignement intégré de la statistique et des probabilités, vers lequel la réforme des programmes des années 2000 a tenté un premier pas.

7.2. Transposition didactique et formation des enseignants

Je conclurai ces quelques remarques d'ordre didactique, par des préoccupations de nature à la fois psychologiques et institutionnelles.

Nous avons déjà connu (période des « maths modernes ») cette sorte de volontarisme prétendant pousser les enseignants à compléter leur formation par l'obligation de s'adapter à une réforme des programmes. Pour louable qu'il soit, il se heurte aux réalités : on n'enseigne vraiment que ce qu'on a profondément compris. Il faut que l'objet d'enseignement puisse vivre dans une relation didactique contrainte par une culture scolaire dont l'évolution est lente et complexe. Combien de professeurs de seconde auront compris les objectifs du programme au point d'en faire leur affaire personnelle ? Combien sauront aménager ces objectifs avec les contraintes horaires, en fonction de leur propre conception de l'enseignement des mathématiques ? Il y aura à vaincre des résistances au niveau même du rapport au savoir. De nombreuses réunions d'information ont été tenues, de nombreux documents riches et variés ont été proposés pour accompagner les nouveaux programmes. Ces efforts devraient être prolongés par la mise en place de stages de formation continue, si l'organisation de celle-ci pouvait être améliorée, notamment au niveau des universités d'été.

L'enjeu profond des programmes des années 2000, qu'au fond je partage, est de prendre en charge la dimension expérimentale des outils mathématiques, et d'en tirer une didactique. La meilleure méthode n'est pas forcément de bousculer les pratiques enseignantes au point de déstabiliser un système de transposition. Pour tout ce qui touche à l'enseignement, on ne peut ignorer les contraintes didactiques et la prise en compte des travaux de recherche sur le sujet ne peut être négligée. C'est le message de prudence que je désire faire passer tout en assurant de la mobilisation des collègues de la commission inter-IREM *Statistique et Probabilités* pour contribuer à la réussite de cet enseignement : suivi d'expérimentations, stages de formation continue, édition de documents en liaison avec l'APMEP...

STATISTIQUES ET PROBABILITES : UN POINT DE VUE DIDACTIQUE

Michel ROCHE - Michel SECO - Claudine VERGNE

Equipe Didactique et équipe Statistique
IREM de Montpellier

Résumé : *Dans une première partie, à la lumière d'études didactiques portant sur l'enseignement des statistiques et probabilités, nous tentons d'analyser les nouveaux programmes de lycée (rentrée 2000), leurs enjeux, les difficultés, les avantages de la nouvelle approche proposée.*

Dans une deuxième partie, nous rappelons rapidement les différentes conceptions du hasard proposées par les philosophes et scientifiques, confrontés à ce difficile problème.

La troisième partie décrit les résultats d'un questionnaire posé aux élèves de deux classes de seconde dans le but de faire émerger leurs conceptions " spontanées " du hasard.

1. Un point de vue didactique

1.1. Ruptures

La statistique, omniprésente dans tous les domaines touchant à la gestion des sociétés (sciences humaines, économie, industrie, biologie, médecine,...), est restée longtemps absente des programmes du secondaire en France. Son introduction, en 1986 au collège et en 1990 en seconde, s'est faite sans grande conviction : « on » garde ce chapitre pour la fin de l'année, s'il reste du temps et puis, comme il n'en reste pas, « on » se dit que ce n'est pas trop grave, les élèves le verront l'an prochain. Et s'ils ne le voient pas, après tout ce n'est pas une grande perte, car ce n'est pas vraiment des mathématiques !

Comme le dit justement J.C. DUPERRET, dans le premier article de cette brochure, il est certes difficile de donner un intérêt à un enseignement réduit à quelques recettes, quelques calculs, quelques constructions de graphiques : ça, ce n'est pas vraiment des statistiques !

La pensée statistique est autrement plus riche et complexe :

- à partir d'un grand nombre de données, accepter de les résumer au risque de perdre des informations, proposer des modèles mathématiques, souvent complexes, et dégager des lois permettant des décrire ou de prévoir le comportement de populations ;
- choisir un échantillon au hasard dans une population, à partir de l'étude mathématique de celui-ci déduire des résultats « probablement justes », et s'appuyer sur ces résultats pour prendre des décisions avec risque d'erreur consenti !
- Prendre le hasard comme axiome, l'intégrer, le modéliser et même l'utiliser comme un outil.

Quand on a toujours fonctionné dans des modèles prédéfinis et structurés, appris qu'une propriété est soit vraie soit fausse, qu'il faut envisager tous les cas, qu'on ne doit rien laisser *au hasard*, qu'on doit démontrer ce qu'il fallait démontrer... l'on risque d'être désorienté pour appréhender et enseigner ce domaine totalement nouveau. On le voit, pour les professeurs de mathématiques nourris au lait du formalisme dès leur plus jeune âge, la rupture est profonde !

Les nouveaux programmes de seconde (rentrée 2000) incitent, de plus, à aborder les statistiques dans une démarche de type expérimental :

- manipulations, observations , simulations, formulation naïve de questions, conjectures en seconde pour « faire connaissance » avec l'aléatoire ;
- formalisation, élaboration du concept de probabilité et étude théorique en première et terminale : l'approche théorique permet de modéliser, interpréter les phénomènes observés en seconde puis d'en prévoir d'autres et d'élargir les champs d'étude.

Il y a là un bouleversement des pratiques. D'une part sur le fond : pas de dogmatisme, de présentation a priori, et d'autre part sur la forme : comment exploiter les observations faites, sur quel savoir les faire déboucher, à quelle institutionnalisation vont-elles donner lieu ? Le programme repousse en première l'introduction d'un langage adapté et invite à utiliser en seconde, pour cette première rencontre avec l'aléatoire, le langage de tous les jours.

Dans ce travail réparti sur deux ans, le **cahier de statistique** prend tout son sens : sur le recueil d'expérimentations et d'observations, consignées dans ce cahier, le professeur et les élèves de première pourront s'appuyer pour préciser le vocabulaire et introduire les concepts de probabilité, de modélisation. Pour réussir cette transition, une concertation entre le professeur de seconde et celui de première et mieux, entre tous les professeurs d'une équipe, paraît souhaitable pour établir une liste de travaux communs aux classes de seconde d'un établissement et assurer ainsi une cohérence dans la progression.

1.2. Première rencontre avec l'aléatoire

Cette première rencontre est prévue en seconde par l'observation des fluctuations d'échantillonnage obtenues par répétitions d'expériences aléatoires comportant un nombre fini d'issues. Le document d'accompagnement du nouveau programme de seconde prévoit d'observer que :

- les distributions de fréquences fluctuent d'un échantillon à l'autre ;
- la moyenne des distributions de fréquences fluctue d'un échantillon à l'autre ;
- la médiane des distributions de fréquences fluctue d'un échantillon à l'autre ;
- l'ampleur des fluctuations des distributions de fréquence diminue lorsque la taille de l'échantillon augmente.

Cela fait, somme toute, beaucoup d'éléments à observer et il peut être prudent, comme le suggère M. HENRY[1], de se limiter dans un premier temps à l'observation des fréquences d'**un** événement dans plusieurs séries d'épreuves.

L'évitement systématique de mots et d'éléments théoriques nouveaux comporte des risques. En parlant de chance au lieu de probabilité, quelle part de subjectivité, d'affectivité laisse-t-on passer dans les représentations qui vont s'installer chez l'élève ? On pourra se référer à ce propos à la thèse de M. ZAKI [ZAKI, 1990] qui décrit, de manière très intéressante, comment plusieurs binômes d'étudiants, travaillant de façon pertinente sur un questionnaire relatif à une situation probabiliste et à sa simulation, produisent, tout à coup, toutes sortes de disfonctionnements lorsqu'ils arrivent à la question du calcul de « l'espérance » mathématique, ce mot, très connoté, semblant perturber la rationalité des étudiants.

Il n'est pas évident que la notion « d'expériences identiques et indépendantes » soit intuitive et soit abordée sans soulever de problèmes, comme l'affirme le document d'accompagnement du

---

[1] M. HENRY, conférence « *Les enjeux de l'enseignement de la statistique et des probabilités aujourd'hui* », séminaire de didactique des mathématiques. DIDIREM. Paris VII : 22/11/2000. Voir aussi la contribution de M. HENRY dans cette brochure.

GTD (Groupe Technique Disciplinaire). Identiques parce qu'on ne sait pas différencier les conditions initiales de chacune ? Parce qu'on ne peut pas prévoir le résultat ? Pas si identiques que cela puisqu'elles n'ont pas le même résultat !

Il doit être précisé ce que le mathématicien entend par expérience aléatoire, celle-ci est du côté de la modélisation, elle s'appuie sur le concept de hasard « objectif », mathématisé, qui ne correspond pas nécessairement à la représentation de l'élève (cf. 2). Ces questions ne doivent pas être laissées sous silence, d'autant que, si la notion de modélisation et d'expériences de référence (jet d'une pièce, jet d'un dé, jet de deux dés, tirage dans une urne ...) est bien reprise dans les programmes de première, il n'y est pas proposé de suite à la « première rencontre » avec le hasard.

1.3. Statistique et informatique

L'enseignement des statistiques est présenté dans les programmes comme le lieu privilégié pour l'utilisation de l'outil informatique ; il est vrai que l'outil informatique permet de gérer des séries statistiques importantes, permet de produire en un clin d'œil des graphiques, multiplie les possibilités d'expérimenter et, par son aspect attractif, peut favoriser l'engagement des élèves dans des conjectures et recherches de réponses. Mais attention, la maîtrise de l'outil informatique n'est en aucun cas l'enjeu de l'enseignement des statistiques. Dans le pire des cas, celui où les élèves n'auraient pas de calculatrices adaptées, on pourrait les faire réfléchir sur des résumés de séries statistiques et des graphiques donnés avec les énoncés, les travaux sur la linéarité de la moyenne se justifieraient d'autant mieux, les simulations, commencées « à la main », pourraient être complétées par des résultats obtenus par le professeur à l'aide d'un tableur.

Il paraît nécessaire de préciser à des adolescents qui, pour beaucoup, se plongent dans des jeux virtuels et peuvent avoir de mondes simulés une représentation ludique et fantastique, quel est le statut des simulations obtenues par les moyens informatiques. Préciser sur quels principes est basé un programme de simulation d'une expérience est nécessaire pour éliminer tout caractère magique des résultats ; c'est par ailleurs une façon de s'engager dans la démarche de modélisation. Les livres scolaires semblent avoir pris en compte ce problème et proposent généralement, en parallèle avec le programme, à rentrer dans la calculatrice pour effectuer la simulation d'une expérience, un commentaire précisant le sens de chaque ligne du programme. La multiplicité et la qualité des graphiques obtenus par ordinateur ne doit pas faire perdre de vue que leur interprétation n'est pas toujours facile, il est important de passer le temps nécessaire à bien préciser ce qu'ils représentent et quels renseignements on en tire.

1.4. Simulation et tables de chiffres au hasard

Dans la partie « conception et mise en œuvre de simulations simples à partir d'échantillons de chiffres au hasard » du programme de seconde, l'important n'est pas la production de ces chiffres mais bien l'utilisation que l'on va en faire, dont la charge doit incomber à l'élève : il y a là l'occasion de laisser s'exprimer son inventivité et de le faire travailler sur le concept même de modélisation .

Il peut être formateur de faire simuler des expériences différentes avec la même table de chiffres pour bien faire sentir en quoi l'identification de différences (ou similitudes) dans les situations va induire des différences (ou similitudes) dans l'utilisation de l'outil « table de chiffres au hasard » et par suite des différences (ou similitudes) dans la nature des résultats obtenus. Si des résultats diffèrent, ce ne sera pas à cause de la variabilité des chiffres obtenus mais bien à cause de modélisations différentes. Et si deux expériences sont simulées de la même manière, c'est que peut-être elles correspondent au même modèle et qu'alors pour les étudier on peut se ramener à une même expérience de référence.

Il serait dommage à ce niveau de court-circuiter l'étape de conceptualisation et d'utiliser trop vite les nombreux programmes « clé en main » proposés par les livres pour simuler telle ou telle

situation. Par contre, l'utilisation de la touche Random et l'imprévisibilité de l'apparition des chiffres permet de bien reproduire le caractère aléatoire des phénomènes étudiés.

La simulation peut certes contribuer à l'enrichissement des conceptions probabilistes des élèves et en particulier, la production répétée d'échantillons, grâce à la simulation, peut efficacement engager les élèves dans une conception fréquentiste des probabilités et leur éviter une confusion (décrite par ALARCON [ALARCON, 199] et ZAKI [ZAKI, 1990]) entre épreuve (résultat d'une expérience) et événement, entre possible et probable. Mais l'enrichissement est lié nécessairement aux questions que l'élève se posera tout en effectuant des simulations et aux réponses qui lui seront apportées. Un renvoi de toute interprétation en première risque d'annihiler les avantages de la simulation d'expériences.

On a reproché au nouveau programme de seconde de passer sous silence le fait que simuler, c'est choisir un modèle de l'expérience et simuler ce modèle ; les exemples proposés concernent des situations pour lesquelles le vécu culturel des élèves fournit des modèles implicites (jet d'une pièce, d'un dé, naissances,....). On peut toutefois faire très facilement toucher du doigt le problème en demandant aux élèves de simuler 100 lancers d'une punaise avec une table de chiffres : c'est impossible ! Cette situation peut amener des débats intéressants et être l'occasion de lever un certain nombre d'implicites. Il faudra renvoyer en fin de compte le problème à la classe de première... (via le cahier de statistique) !

1.5. Statistiques et mathématiques

En s'engageant dans l'enseignement des statistiques, on n'aura aucun mal à y trouver matière à analyse, réflexion, anticipation, calculs, démonstrations, autant d'activités que l'on peut qualifier de mathématiques (cf. en annexe quelques exemples d'exercices du niveau de seconde).

En abordant la dimension mathématique des statistiques, on sera amené à produire des démonstrations, pour lesquelles il faudra s'engager dans le registre des écritures statistiques où foisonnent les notations, les formules utilisant des symboles nouveaux (par exemple : $\Sigma\, kx_i = k\, \Sigma\, x_i$, $\overline{x+y} = \overline{x} + \overline{y}$). Ce genre d'écriture ne peut faire sens que s'il est patiemment introduit et travaillé ; un minimum d'expertise n'est envisageable que si le temps passé à cet apprentissage est suffisant. Pour pouvoir, en cycle terminal, aborder des travaux de type calcul et démonstration qui contribueront à conférer aux statistiques un statut de mathématiques authentiques, il n'est pas trop tôt de commencer à préparer les élèves en seconde. Cette difficulté, de nature algébrique, ne semble pas avoir été prise en compte par les programmes. On trouve seulement dans la partie commentaires en seconde la proposition de « l'on pourra utiliser le symbole Σ », suggestion très peu suivie par les livres d'ailleurs, et il n'est plus fait mention d'apprentissage dans les programmes de première, ni en S, ni en ES.

1. 6. Quels avantages peut-il y avoir à enseigner les statistiques ?

C'est une occasion, en quelque sorte, d'actualiser notre enseignement. Non pas pour jouer le jeu du pseudo-concret, dont les limites ont été analysées ailleurs. Mais si nous avons pour objectif de développer l'esprit critique de nos élèves, de les préparer de façon crédible à une sortie sur la vie, cela ne peut se faire sans tenir compte des réalités du monde contemporain et donc sans une formation à l'outil et aux méthodes statistiques.

La statistique s'accommode mal du dogmatisme : pour une série donnée qu'il faut résumer afin d'en obtenir des renseignements, il y a différents résumés possibles qui déboucheront peut-être sur des interprétations différentes ; pour une expérience donnée, il peut y avoir différentes modélisations ; pour une situation donnée, plusieurs décisions sont possibles. L'absence d'une réponse-type dont la vérité est de toute façon détenue à l'avance par le professeur peut libérer certains élèves de l'image qu'ils ont d'une matière trop rigide et inaccessible.

Grâce à la simulation, l'élève a un outil nouveau qui lui permet de vérifier ses conjectures ; certaines questions qu'il se pose peuvent ne plus être des questions au professeur, mais des questions qu'il va essayer de prendre en compte lui-même et pour lesquelles il pourra obtenir des éléments de réponse (sans certitude) grâce à la simulation. La simulation peut ainsi devenir une aide à la dévolution de problèmes à l'élève. Cette hypothèse semble confirmée par le bilan de la mise en œuvre anticipée du programme de seconde [GTD, 2000]. On peut à ce propos faire un parallèle avec l'utilisation en géométrie d'un logiciel comme CABRI, qui constitue un outil d'aide à la conjecture.

Pour certaines de ces conjectures, seuls les éléments théoriques abordés en première et terminale permettront d'obtenir une réponse ferme et / ou des interprétations des phénomènes : cela doit contribuer à faire sentir à l'élève l'efficacité, la force, la raison d'être d'une théorie mathématique et ne peut que contribuer à terme à sa formation culturelle au sens large.

L'absence de réponses a priori implique que, pour des décisions prises ou des estimations faites, il faudra justifier ses choix, convaincre ses pairs, tenir compte des arguments d'autrui ; c'est l'occasion d'instaurer dans la classe des situations de dialogue, de débat, où les élèves vont parler de leurs travaux et s'impliquer. Cela peut déboucher sur des changements dans les pratiques d'enseignement pour arriver à créer des situations où l'élève soit un des acteurs de la construction de son savoir.

1.7. Conclusion

L'introduction d'une statistique plus moderne et de la notion d'aléatoire en seconde paraît un objectif intéressant qui, s'il est mené à bien, peut entraîner des changements dans les pratiques d'enseignement et d'apprentissage. Certains didacticiens proposent même d'introduire l'aléatoire dès le collège[2]. Mais il y a risque à introduire des contenus nouveaux dans les programmes sans que les enjeux soient bien perçus par tous et sans que les professeurs soient tout à fait préparés.

Dans le cas présent, des difficultés sont à craindre au niveau de la seconde : comment aborder les contenus proposés, comment les institutionnaliser ? Mais c'est surtout le « passage de relais » entre la seconde et la première, qui, laissé dans le flou, risque de se faire difficilement : rien ne garantit que les observations et les simulations menées à bien en seconde soient réinvesties pour installer la théorie des probabilités en première et donner du sens aux contenus proposés.

De fait, ce nouveau programme se présente un peu comme un pari, et on peut se demander s'il sera tenu ...

2. Le hasard, une question problématique et didactique

Avec le nouveau programme de statistique de la classe de seconde, les élèves vont rencontrer le hasard de façon officielle. Au collège, certes, la statistique descriptive étudie avec des outils appropriés, des situations où intervient la variabilité, mais la notion de fréquence, au sens de l'observation répétée de la valeur d'un caractère dans une série d'expériences du même type, n'y est pas abordée.

C'est précisément par l'observation des fluctuations des fréquences des issues d'une expérience aléatoire que les élèves sont censés comprendre et mesurer les effets du hasard.

L'objectif affiché par les programmes est de construire un état d'esprit statistique, autrement dit passer d'affirmations uniquement basées sur l'expérience personnelle ou sur la simple opinion à des affirmations de type probabilistes acceptant une incertitude contrôlée. Le hasard dont il est

---

[2] Cf. « Quelle place pour l'aléatoire au collège », 2001, Repères-IREM n°42.

question dans les fluctuations d'échantillonnage et les situations de jeux n'est que l'une de ses facettes. Les simulations d'expériences aléatoires à l'aide de tables de chiffes au hasard font intervenir un hasard « aseptisé » dans des situations bien souvent choisies de façon ad-hoc.

Un regard sur les différentes conceptions du hasard peut éclairer l'enseignant sur les difficultés qui ne manqueront pas de surgir lors de l'enseignement de cette notion, dont l'élève a déjà des représentations personnelles.

2.1. L'expérience courante

L'expérience courante de la vie, fait éprouver le hasard au commun des individus, par l'inattendu, les rencontres fortuites, ou la réalisation d'événements non souhaités. Le sentiment d'une impuissance à prévoir le futur conduit alors les hommes à invoquer le destin. Le destin est soit considéré comme l'inévitable, décidé par des puissances surnaturelles, soit appréhendé comme l'impossibilité d'agir sur la nature. Depuis toujours les hommes tentent d'apprivoiser ce hasard par des moyens divinatoires, que sont par exemple les rites et les pythies. Dans cette conception subjective du hasard, la chance qualifie le dénouement heureux d'événements imprévisibles, c'est un caractère intrinsèque à l'homme.

2.2. La philosophie

La philosophie tente de préciser les sens du hasard ou de la contingence : ce caractère des événements futurs d'être ou de n'être pas. Les hommes peuvent être tenus pour responsables des événements portant la marque du hasard. La liberté humaine, parce qu'elle donne le libre choix aux actions des hommes et à leurs interférences, peut conduire à la réalisation d'événements inattendus. Les passions incontrôlées, la démence vont à l'encontre d'un déterminisme historique.

2.3. La " nature "

La nature, quant à elle, ne suit pas que des lois rigoureuses, elle est un mélange de nécessité et d'événements fortuits. La variabilité des phénomènes naturels est une source de hasard, indépendante des hommes. Lorsque les décisions des hommes concernent la nature, il peut alors se produire ce que Sartre appelle des contre-finalités. Les effets produits par l'action humaine sont contraires aux effets attendus. Le hasard reflète ainsi l'ignorance que nous avons des chaînes causales de la nature. Les exemples qui vont dans ce sens ne manquent malheureusement pas. Aussi une activité des hommes est-elle la prévision des risques, afin de mettre en place des techniques de prévention , d'assurances et de précautions.

L'idée, mécaniste, selon laquelle l'effet est proportionné à la cause à conduit à considérer un autre aspect du hasard. Un fait insignifiant peut engendrer des conséquences imprévisibles et disproportionnées. On qualifie un tel enchaînement « d'effet boule de neige ». PASCAL déjà y faisait allusion lorsqu'il indique que, si CROMWELL n'avait pas eu un calcul mal placé, à un moment déterminé de l'histoire, celle-ci en eut été changée. De même Ernest RENAN se demande ce qu'aurait été l'avenir de l'Europe, si la trajectoire du boulet de canon qui tua l'empereur Gustave ADOLPHE avait été légèrement différente.

Antoine Augustin COURNOT (1801-1877), demande de considérer que le hasard est : « l'idée de rencontre de faits rationnellement indépendants les uns des autres, rencontre qui n'est elle même qu'un pur fait, auquel on ne peut assigner de loi ni de raison ». Ainsi l'univers est présenté comme une infinité de séries, sans début ni fin, pouvant coexister et pouvant se croiser fortuitement pour donner naissances à d'autres séries indépendantes. Il précise aussi que la

notion de hasard a son fondement dans la nature et n'est pas seulement relative à la faiblesse de l'esprit humain.

2.4. La science

Les scientifiques sont eux aussi confrontés au hasard. Leur approche ne peut être qu'objective. Le déterminisme affiché des lois de la physique classique, glorifié par celles de l'attraction universelle, n'est pas en contradiction avec la contingence. Pierre Simon DE LAPLACE (1749-1827) à la suite de Jacques BERNOULLI (1654-1705) proclame un déterminisme universel, pur et dur. Tout événement, si modeste soit-il, a une cause. Mais, le découvreur de la loi qui porte son nom n'ignorait évidemment pas l'effet de causes irrégulières sur la détermination d'un événement. C'est l'ignorance que nous avons de toutes les causes, si infimes soient-elles, qui impose une approche probabiliste et non pas une composante intrinsèque des lois de la nature. Le hasard, dans ces conditions, n'explique rien.

Benoît MANDELBROT, l'inventeur des fractales, réfléchit au problème du hasard. Il propose de qualifier de bénin le hasard maîtrisé, celui soumis aux lois connues des probabilités et en particulier à celle de la puissante loi des erreurs : la loi normale de LAPLACE-GAUSS. Il donne le nom de hasard sauvage à celui qui préside à l'évolution de systèmes dont les fluctuations ne se laissent pas contrôler, en particulier, par la loi des grands nombres. Il prend pour exemple les fluctuations de la bourse.

Henri POINCARE (1854-1912) et Jacques HADAMARD (1865-1963) ont étudié des systèmes tellement sensibles aux conditions initiales, qu'ils peuvent évoluer vers des états fort différents de l'état non perturbé. L'interdiction faite à l'expérimentateur de connaître avec précision les conditions initiales, rend illusoires les prévisions certaines à long terme, introduisant ainsi une autre dimension du hasard. Un exemple familier de tel système est celui des conditions atmosphériques qui conduisent les ingénieurs météorologistes à assortir la prévision du temps d'indices de confiance.

Avec la mécanique quantique, enfin, le hasard est l'essence de la théorie. On se souviendra que la relation d'incertitude d'HEISENBERG, concernant les variables conjuguées position et quantité de mouvement déterminant l'évolution d'un système de particules, interdit d'accroître la précision de la mesure de l'une sans augmenter d'autant l'imprécision de l'autre. L'état initial ne peut donc être connu de façon à prévoir l'évolution du système selon un point de vue déterministe et seules des probabilités seront mesurables.

Ce rapide compte-rendu de différentes conceptions du hasard et de son traitement dans des paradigmes très différents montre la complexité de la notion. L'évolution des espèces dans lesquelles le hasard joue un rôle tout aussi important que le principe de sélection naturelle de DARWIN, les théories des systèmes dissipatifs d'Ilya PRIGOGINE qui, loin de l'équilibre, peuvent néanmoins être stables, les théories du chaos et de la percolation, etc... montrent que les problèmes concernant le hasard sont toujours d'actualité.

Pour avoir une idée de certaines représentations du hasard chez les élèves, nous avons proposé, en classe, un questionnaire à deux classes de seconde. Les questions proposées s'alimentent à quelques-unes des différentes conceptions du hasard décrites ci-dessus. On a choisi la forme alternative pour les questions avec l'objectif de forcer les élèves à un minimum de réflexion. Ceux-ci n'étaient pas tenus de fournir une seule réponse, ce qui a permis à beaucoup d'entre eux de nuancer leur point de vue tout en le structurant.

Le texte du questionnaire et le dépouillement des réponses font l'objet du chapitre suivant. Les commentaires accompagnant chaque question tentent de brosser, à grands traits, la synthèse des réponses, afin que le lecteur puisse se faire une idée rapide des faits significatifs à prendre en compte.

3. Questionnaire

3.1. Le texte

En quelques phrases exprimer ce que vous pensez des situations suivantes :

1 - Le Pdg d'une marque de voitures voudrait connaître les couleurs préférées des conducteurs d'automobiles. Il charge son ingénieur des ventes de trouver une solution à ce problème. Celui-ci a une idée. Il se procure auprès des services des cartes grises le fichier des conducteurs. Il tire 1000 noms au hasard et envoie à chacun un questionnaire. Il obtient les 1000 réponses et propose alors à son Pdg de choisir les couleurs de la future production conformément aux préférences exprimées. Que pensez vous de sa méthode ?

a) Elle est bonne car :
b) Elle n'est pas bonne car :

2 - Un jour de grand vent une tuile tombe sur la tête de monsieur Dupont qui se rendait chez le boulanger. Cet accident est-il dû au hasard ?

a) Oui car :
b) Non car :

3 - Une vilaine guêpe ayant réussi à pénétrer dans une voiture, elle piqua la main du conducteur. Sursautant, le chauffeur fit une embardée qui amena son véhicule en travers de la route. Un camion venant en sens inverse ne put éviter le choc. Les trois autres voitures qui suivaient le camion ne purent freiner à temps et se percutèrent. Circulant en vélo et distrait par ce carambolage, je me retrouvai dans le fossé. Le hasard est-il responsable de cet enchaînement ?

a) Oui car :
b) Non car :

4 - J'ai jeté 99 fois de suite une pièce non truquée. J'ai obtenu 99 piles. Je parie ma chemise qu'au centième coup, je vais obtenir face, et vous ?

a) Je parie car :
b) Je ne parie pas car :

5 - Jean joue les nombres 1, 2, 3, 4, 5, 6 au loto. Pierre lui dit qu'il n'a aucune chance de gagner avec cette série. Etes-vous d'accord ?

a) Je suis d'accord car :
b) Je ne suis pas d'accord car :

6 - Il sera un jour possible de prévoir avec certitude le temps qu'il fera en France un mois à l'avance.

a) Oui car :
b) Non car :

7 - On entend souvent dire ces temps derniers que le risque zéro n'existe pas, êtes-vous de cet avis ?

a) Je suis de cet avis car :
b) Je ne suis pas de cet avis car :

8 - Après une trajectoire d'au moins 150 m dans les airs, la balle du golfeur atterrit à 1 m du drapeau, roule sur le gazon et finit par tomber dans le trou. Quel coup exceptionnel ! Est-il dû uniquement à l'adresse du champion ?

a) Oui car :
b) Non car :

3.2. Dépouillement des réponses au questionnaire

Résultats exprimés en pourcentage du nombre de réponses fournies et non du nombre d'élèves questionnés ; par suite, si un élève a répondu plusieurs fois on ne dépasse pas les 100%.

Question 1

Oui	67%	Non	16%	Autre	5%
Donne une indication	38%	Echantillon insuffisant	68%		
Choix au hasard	24%	Echantillon mal construit	31%		
Avis direct	13%	Individus trop différents	18%		
Suffisant	13%	Plusieurs avis			
Avis partagé		Echantillon particulier			
		Approximatif			

Commentaires pour les oui

Des réponses indiquent que cette méthode va donner un renseignement utile sur les goûts des acheteurs. S'adresser directement au consommateur est une bonne chose.

Des réponses font référence à l'effectif de l'échantillon et invoquent le choix au hasard comme un point positif de la méthode.

Commentaires pour les non

Les arguments contre la méthode du sondage se fondent sur la non représentativité de l'échantillon, soit par le nombre trop faible de sondés, soit par le fait que le choix se faisant au hasard , il ne peut tenir compte de la diversité de la population des acheteurs.

Question 2

Oui	73%	Non	26%	Autre	2%
Coïncidence	14%	Vent	40%		
Imprévisibilité	8%	Cause mécanique	25%		
Autre que Dupont	20%	Indépendance			
Fatalité, destin, malchance	42%	Entretien du toit	30%		
Dupont ne peut pas prévoir		Faute à Dupond			

Commentaires pour les oui

Des réponses semblent, sous des formulations pas toujours explicitent, s'appuyer sur l'indépendance des deux phénomènes : le détachement de la tuile et le passage de Monsieur Dupont et de leur rencontre fortuite. Les autres réponses oui portent l'empreinte d'un hasard subjectif, en particulier Dupont est considéré comme une personne unique.

Commentaires pour les non

Les causes physiques sont reconnues responsables de la chute de la tuile, le passage de Monsieur Dupont étant secondaire. La responsabilité du propriétaire du toit est engagée : mauvais entretien. Une réponse implique Monsieur Dupont : il aurait dû rester chez lui, un jour de grand vent. Un élève affirme que le hasard n'existe pas.

Question 3

Oui	44%	Non	52%	Autre	4%
Imprévisibilité	23%	Guêpe fautive	35%		
Malchance		Fatalité	5%		
Coïncidence		Voitures	8%		
Hasard responsable	35%	Distraction	10%		
Inévitable		Faute au conducteur	30%		
		Cycliste	15%		

Commentaires pour les oui

Certaines réponses décrivent l'enchaînement comme un effet boule de neige, sans engager la responsabilité des acteurs. D'autres réponses mettent en cause le hasard dans le fait que la guêpe pique le conducteur, mais engagent la responsabilité des acteurs pour la suite. D'autres réponses encore mettent en cause le hasard dans la rareté d'un tel enchaînement.

Commentaires pour les non

La majorité des réponses met le hasard hors de cause, le carambolage est de la responsabilité des acteurs. Et certaines réponses mettent particulièrement le cycliste en cause.

Question 4

Oui	28%	Non	63%	Autre	9%
Chance	52%	Hasard	85%		
Pièce truquée	21%	Loi des séries			

Commentaires pour les oui

Des réponses oui sont très subjectives, les élèves se mettent dans la situation d'un joueur qui aurait de la chance. Des élèves remarquent que pile est trop souvent sorti.

Commentaires pour les non

Des réponses invoquent clairement le fait qu'il y a toujours une chance sur deux pour le coup suivant. D'autres que comme le pile sort, il va continuer à sortir.

Commentaires pour les réponses autres

Manifestement la question est déroutante et les réponses en dehors du sujet.

Question 5

Oui	27%	Non	73%
Choix des nombres	100%	Chance	18%
		Hasard	45%
		Faible probabilité	27%

Commentaires pour les oui

Les nombres donnés semblent bien particuliers, l'absence de nombres avec des dizaines est un argument fort.

Commentaires pour les non

Des réponses reconnaissent que les nombres sont particuliers, mais qu'avec de la chance cette combinaison, peu probable peut arriver. Des réponses attribuent au hasard pur le tirage des nombres et à l'égalité des chances pour tous les tirages.

Question 6

Oui	47%	Non	53%
Technologie	100%	imprévisibilité	66%
		Long terme	18%
		Complexité	
		Mise en doute	
		Limite de la technologie	15%

Commentaires pour les oui

Les réponses oui ont un argument unique, celui des progrès de la technologie qui pourront un jour conduire à des prévisions certaines ; ou presque.

Commentaires pour les non

Pour la plupart des non, le temps reste un phénomène dépendant du hasard, on ne peut le prévoir et à plus forte raison à long terme.

Question 7

Oui	78%	Non	15%	autre	7%
Part de risque imprévu	77%	Précautions	90%		
Elimination impossible des risques	23%	Le risque n'existe pas toujours			

Commentaires pour les oui

Le risque est permanent, imprévisible ; tapis, en quelque sorte prêt à surgir. Des réponses pour dire qu'il est impossible de tout prévoir.

Commentaires pour les non

Les réponses font allusion à la possibilité, dans l'avenir, de juguler tous les risques en prenant les précautions nécessaires. Une seule réponse envisage qu'après tout il n'y a pas toujours un risque à l'action entreprise. Le risque est considéré négativement, pour eux par exemple, gagner au loto n'est pas un risque.

Question 8

Oui	28%	Non	70%	Autre	2%
Adresse, professionnalisme	100%	Conditions extérieures	32%	Intervention d'une fée	100%
		Chance	60%		
		Elimination impossible des risques			

Commentaires pour les oui

Les réponses invoquent l'adresse, l'entraînement du joueur, mais la majorité ajoutent cependant un bémol en ajoutant un zeste de chance.

Commentaires pour les non

La chance est l'argument majoritaire. Des réponses font allusion à des épreuves répétées qui ne seraient pas gagnantes. La chance est cependant aidée par des conditions extérieures favorables : le vent, le terrain.

4. Conclusion

Un tel questionnaire et des réponses élaborées dans un temps limité ne peuvent pas, évidemment, faire accéder de façon approfondie et totalement fiable aux représentations des élèves, mais, cependant, on peut en tirer certains enseignements. Comme le montre le vocabulaire employé pour répondre aux questions et les explications données, globalement les élèves interrogés rencontrent les différentes conceptions du hasard décrites dans le chapitre 2. Pourtant on remarque une grande hétérogénéité face aux situations proposées.

Les réponses aux questions 2 et 3 réunissent : *coïncidence, imprévisibilité, fatalité, destin, malchance, hasard, inévitable, fatalité*, une bonne palette pour l'expression d'un hasard subjectif. Certains élèves, cependant, sans doute « déterministes », recherchent une cause à ces évènements. Le recours au sondage est une pratique familière, mais le choix aléatoire des individus n'est pas pour certains une méthode fiable pour prendre une bonne décision. Certains ont une confiance religieuse en la science qui répondra à toutes les incertitudes, on prévoira le temps, on mesurera tout et la balle du golfeur expérimenté ne sera plus soumise aux aléas.

Les questions 4 et 5 font apparaître les modèles mathématiques des jeux de hasard, on peut lire : *la probabilité que la grille gagnante soit 1, 2, 3, 4, 5, 6 est faible, on a une chance sur deux d'obtenir pile*. Mais aussi la loi des séries... *puisque pile sort, ça va continuer*. Le mot chance est employé dans de multiples sens, certains ne croient pas au hasard, il faudra pourtant les « convaincre » qu'il existe un hasard mathématique et que la chance est une probabilité.

On peut alors poser la question suivante : la première rencontre officielle avec le hasard dans l'institution scolaire n'est elle pas par trop réductrice ? Les opinions des élèves sur le hasard, acquises en acte, ne peuvent-elles pas s'ériger en obstacle épistémologique et par là compromettre l'acquisition d'un esprit statistique et probabiliste adéquat aux contenus de l'enseignement de ces deux matières ? Ne faut-il pas tendre une oreille plus attentive aux déclarations de Gaston BACHELARD : « la science, dans son besoin d'achèvement, comme dans son principe s'oppose absolument à l'opinion... l'opinion pense mal. On ne peut rien fonder sur l'opinion : il faut d'abord la détruire, elle est le premier obstacle à surmonter » ?

P.S

Nous avons pris connaissance, après avoir rédigé ce modeste travail, de l'ouvrage de Dominique LAHANIER-REUTER (cf. bibliographie ci-après) : « Conceptions du hasard et enseignement des probabilités et des statistiques ». L'auteur étudie, avec toute la rigueur et la profondeur de la recherche scientifique, la problématique que nous avons effleurée. La lecture de ce livre est indispensable à qui veut approfondir le sujet.

Bibliographie

PREMIERE PARTIE

ALARCON J., 1995, « *L'appréhension des situations probabilistes chez les élèves de 12-14 ans : résultat d'une enquête proposée à des élèves de 4ème et de 5ème* ». Thèse de 3ème cycle, Université Louis Pasteur, Strasbourg.

BATANERO Carmen et al., 1997, « *Effect of the implicit combinatorial model on combinatorial reasonning in secondary school pupils* ». Educational Studies in Mathematics, pp. 181-199, vol 32, n°2.

CASTRO Cesar Saenz, 1998, « *Teaching probability for conceptual change* ». Educational Studies in Mathematics, pp. 233-254, vol 35, n°3.

FISCHBEIN E. et al., 1997, « *Schemata and intuitions in combinatorial reasonning* ». Educational Studies in Mathematic s, pp. 27-47, vol 34, n°1.

GIRARD, HEHRY, PICHARD, 2001, « *Quelle place pour l'aléatoire au collège ?* ». Repères-IREM n° 42.

GTD, 2000, « *Bilan de la mise en œuvre anticipée durant l'année scolaire 1999-2000 du programme de mathématiques de la classe de Seconde* ». Documents d'accompagnement des programmes de Seconde, Première ES, Première S.

HENRY M., 2000, « *Perspectives pour l'enseignement de la statistique et des probabilités* ». Gazette de la SMF.

JONES Graham A. et al., 1997, : « *A framwork for assessing and nurturing young childrens thinking in probability* ». Educational Studies in Mathematics, pp. 101-125, vol 32, n°2.

MAURY S., 1986, « *Contribution à l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problèmes* ». Thèse, Université des sciences et techniques du Languedoc, Montpellier.

ZAKI M., 1990, « *Traitements de Problèmes de Probabilité en Situation de Simulation* ». Thèse, Université Louis Pasteur, Strasbourg.

DEUXIEME PARTIE

RUELLE Davis, 1997, Encyclopédie Philosophique Universelle, pp. 1185-1192. PUF 3° édition.

Encyclopédie Universalis : articles *probabilités* et *statistiques*.

Pour la Science : Dossier spécial avril 1996.

Tangente : *Spécial Probabilités* n°47. Décembre 1995 - Janvier 1996.

Tangente : *Faut-il se fier aux statistiques ?* n°77. Octobre - Novembre 2000.

TROISIEME PARTIE

LAHANIER-REUTER D., 1999, « *Conceptions du hasard et enseignement des probabilités et statistiques* ». Education et formation, recherche scientifique. PUF.

Voir aussi références en partie IV de cette brochure.

ANNEXE

Proposition d'exercices pour la classe de seconde

1. Dans ma classe de 26 élèves, la moyenne des notes (sur 20) à un examen est 10,1. Puis-je raisonnablement penser que la moitié de la classe environ a été reçue ?

2. Pour une série de valeurs, la moyenne est 126, la médiane est 115, et l'étendue 15.

Que deviennent ces paramètres si toutes les valeurs augmentent de 5 ?

Que deviennent ces paramètres si toutes les valeurs diminuent de 10% ?

3. Pour la série des 50 valeurs suivantes, la moyenne est 438,96. Compléter le tableau.

430	431	435	442		451
12	5		10	8	6

4. Démontrer que : *si une série admet dans un repère orthogonal un diagramme en bâton symétrique par rapport à une droite parallèle à l'axe des ordonnées, alors la moyenne de la série est égale à la médiane.*

La réciproque est-elle vraie ?

5. A l'aide d'une table de chiffres au hasard, simuler 10 lancers de 4 dés équilibrés.

6. A l'aide d'une table de chiffres au hasard, simuler un tirage du loto.

7. Dans une urne, il y a 71 boules rouges et 23 boules noires. A l'aide d'une table de chiffres au hasard, simuler 10 tirages au hasard de boules dans l'urne.

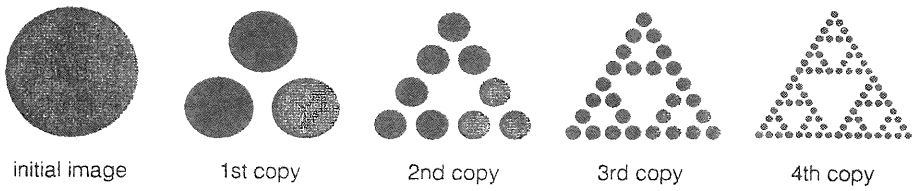
8. Un sondage est effectué auprès d'une population qui doit se prononcer pour ou contre une réforme. 61% des personnes interrogées se sont prononcées « pour ». Quelle conclusion peut-on en tirer pour l'ensemble de la population si :

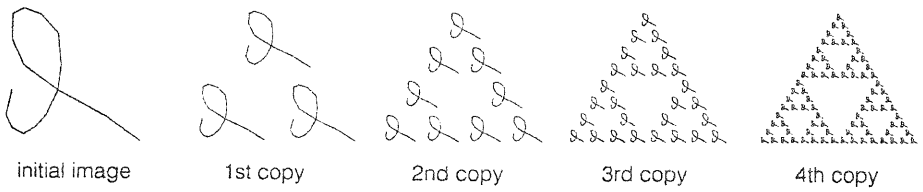
a) l'échantillon est représentatif et constitué de 940 personnes ?

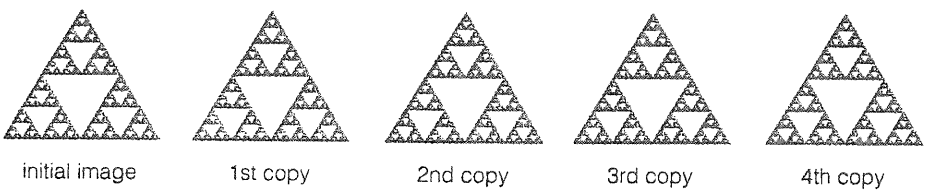
b) l'échantillon est représentatif et constitué de 3500 personnes ?



III - ACTIVITES POUR LA CLASSE ET PROLONGEMENTS MATHÉMATIQUES







SIMULATION ET MODELISATION : ETUDE D'UN EXEMPLE

Joëlle FONTANA – Maryse NOGUES

Equipe Analyse de l'IREM de Montpellier

Résumé : *Un des objectifs de l'enseignement des statistiques au lycée est de faire prendre conscience des fluctuations d'échantillonnage et de procéder à des simulations d'expériences à l'aide d'outils informatiques. L'étude d'un problème, apparemment très ludique, permet d'observer comment ces simulations sont liées à nos schémas de représentation et influencent la « simplification » que l'on opère en élaborant un modèle. Ce problème permet ainsi d'illustrer les nécessaires « allers-retours » entre expérience, modèle, simulation afin de pouvoir construire un modèle qui pourrait être considéré comme valide, c'est-à-dire en adéquation avec ce qu'il est censé représenter.*

1. De nouveaux programmes, de nouveaux problèmes

Le nouveau programme de seconde applicable depuis la rentrée 2000[1] développe largement l'étude des statistiques. Nous reprenons ci-dessous quelques extraits du document d'accompagnement des programmes qui nous semblent mettre en évidence les nouveautés introduites dans l'enseignement des statistiques au lycée.

> *Le programme s'inspire largement du contenu des programmes antérieurs* ***sauf en statistiques****... Les choix, traduits en termes de programme pour la classe de seconde, sont guidés par les perspectives suivantes pour le lycée :*
> *- acquérir une expérience de l'aléatoire ;*
> *- comprendre ce qu'est une question statistique et le type de réponse que l'on peut proposer ;*
> *- voir dans un cas simple ce qu'est un modèle probabiliste.*
>
> *...L'esprit statistique naît lorsqu'on prend conscience de* ***la fluctuation d'échantillonnage...***

Il fait une large place à la démarche expérimentale et à l'utilisation des outils informatiques :

> ***L'outil informatique*** *donne la possibilité d'une* ***démarche quasi expérimentale*** *dans le champ des nombres... favorisant une approche plus active et plus impliquante. Il élargit considérablement les possibilités* ***d'observation et de manipulation****...*

Dans ce cadre, il promeut les méthodes de simulation d'évènements aléatoires :

> *A propos de simulation ..., il s'agit de faire comprendre à chaque élève ce que simuler veut dire.....Parallèlement, on utilisera la touche* ***random*** *comme générateur de chiffres au hasard ... Simuler une expérience de référence consiste*

[1] Ce programme a été publié au B.O n°6 du 12 Août 1999. On peut également en prendre connaissance actuellement sur le site du CNDP : http://www.cndp.fr/lycee/maths ainsi que du document d'accompagnement des programmes dont la version définitive est datée d'octobre 2000.

ainsi à produire une liste de résultats qui pourraient être ceux que l'on obtiendrait par la réalisation pratique de cette expérience.

Formellement, ***simuler une expérience, c'est choisir un modèle pour cette expérience... puis produire une liste de données...***

Il évoque les prolongements en classe de première et terminale :

Les probabilités : elles seront introduites et travaillées en première ou terminale...

Ce nouveau programme appelle ainsi de nouveaux problèmes, de nouvelles façons d'aborder les statistiques en seconde. C'est dans cette perspective que l'équipe analyse de l'IREM de Montpellier a travaillé sur un problème que nous pourrions énoncer ainsi :

« Quelle est la probabilité, en coupant au hasard un spaghetti en trois morceaux, de pouvoir construire un triangle avec les 3 morceaux obtenus ? »

Ce problème a fait l'objet d'un atelier animé par Joëlle FONTANA et René BERNARD au stage de formation des formateurs de l'IREM de Montpellier à Boisseron en Juin 2000. Il a aussi été présenté, par ces mêmes formateurs, au colloque francophone, EM2000, de Grenoble en Juillet 2000[2]. Lors de ces différents ateliers, les participants ont réalisé eux-mêmes l'expérience : découpe de 30 spaghettis en 3, comptage des triangles obtenus et calcul puis regroupement des fréquences, ils ont aussi essayé de simuler à l'aide d'une calculatrice ou d'un tableur l'expérience.

Dans tous les cas, un écart important a été observé entre :

- d'une part les fréquences obtenues expérimentalement, qui se stabilisaient approximativement autour de 0,7 ;
- d'autre part, les résultats obtenus par simulations, pour lesquels de plus, les différentes simulations produites ne permettaient pas d'obtenir une stabilisation des fréquences vers une seule valeur... , mais deux valeurs : 0,25 et 0,20.

Ainsi cette expérience, très ludique, a permis de mettre en évidence les différents problèmes liés aux rapports entre expérimentation et simulation. Simulation qui repose, même si parfois cela est implicite, sur un modèle. Dans un premier temps, le type de modèle utilisé ne correspond pas nécessairement à la réalité observée et c'est donc par ajustements successifs que peut être élaboré un modèle qui rendra compte « au mieux » de l'expérience réelle. Même si une question se pose : « quelle réalité observe-t-on et celle-ci ne dépend-elle pas du protocole expérimental ? »

2. Expérience, et simulations

La diversité des résultats observés amène assez naturellement à se poser la question : « mais comment introduit-on le hasard dans le choix des découpes ? »

En interrogeant les participants sur leur façon de procéder, différentes possibilités sont évoquées :

- on choisit, avant de couper, l'emplacement des deux coupures que l'on va réaliser (situation qui est dans la pratique rare) ;

- on coupe un premier morceau, et l'on recoupe le second (beaucoup plus classique) ;

---

[2] On peut aussi trouver une rédaction du problème pour les élèves dans le livre de seconde collection Fractale paru en 2000.

- on coupe un premier morceau et on recoupe le plus grand des deux morceaux obtenus ;
- on coupe un premier morceau et on recoupe le plus petit des deux morceaux obtenus.

Ces différentes façons de faire ne relèvent pas du même modèle et, de la même façon, les simulations proposées par les participants ne relevaient pas toutes du même modèle, même si certains invariants pouvaient être mis en évidence. Nous avons donc réalisé diverses simulations possibles à l'aide d'un tableur, en les associant à leur modèle respectif. Celui-ci nous a permis, à l'aide d'une représentation graphique d'obtenir une fréquence (ou probabilité théorique), si l'on considère que la probabilité d'obtenir un point dans une zone donnée est proportionnelle à l'aire de cette zone.

Simulation 1

On considère que couper un spaghetti en trois, c'est découper un segment, que l'on peut supposer de longueur 1, en trois morceaux, en plaçant au hasard et indépendamment deux points M et N d'abscisses x et y sur ce segment. Cette situation correspond à la première façon de couper au hasard le spaghetti en 3 mentionnée précédemment.



Il s'agit donc de trouver 2 nombres x et y compris entre 0 et 1 qui sont les abscisses des 2 coupures et qui déterminent les longueurs des 3 morceaux, notées a, b et c. On obtient :

$$a = \inf(x, y) \; ; \; b = \max(x, y) - \min(x, y) \text{ et } c = 1 - (a + b).$$

Les morceaux permettant la construction d'un triangle sont donc les morceaux de longueur a, b et c satisfaisant les trois inégalités triangulaires et $a + b + c = 1$, ce qui conduit à la condition :

$$\max(a, b, c) < \frac{1}{2}.$$

Avec un tableur, on peut procéder comme ci-dessous :

A	B	C	D	E	F	G
Première coupure	**Deuxième coupure**	**Premier morceau**	**Deuxième morceau**	**Troisième morceau**	**Max des 3 morceaux**	**test**
ALEA()	ALEA()	MIN(A;B)	MAX(A;B)-MIN(A;B)	1-(C+D)	MAX(C;D;E)	SI(F<0.5;1;0)
0,17116123	0,62735719	0,17116123	0,456195958	0,372642809	0,456195958	1
0,92721917	0,43075206	0,43075206	0,496467101	0,072780834	0,496467101	1
0,82647913	0,06289745	0,06289745	0,76358168	0,173520866	0,76358168	0

On a observé pour 8 séries de 500 simulations les fréquences suivantes :

0,25 ; 0,23 ; 0,25 ; 0,25 ; 0,26 ; 0,27 ; 0,24 ; 0,26.

La fréquence moyenne pour 4000 simulations est ainsi de 0,251.

Une étude théorique graphique, ou géométrique, du modèle lié à cette simulation, qui de façon implicite sous-entend une équirépartition des coupures, peut permettre de comprendre pourquoi les fréquences se stabilisent autour de 0,25.

Une approche graphique :

x et y sont deux nombres compris entre 0 et 1. L'ensemble des points M (x ; y) possibles est donc, dans un repère orthonormal, le carré de côté 1 construit sur les unités.

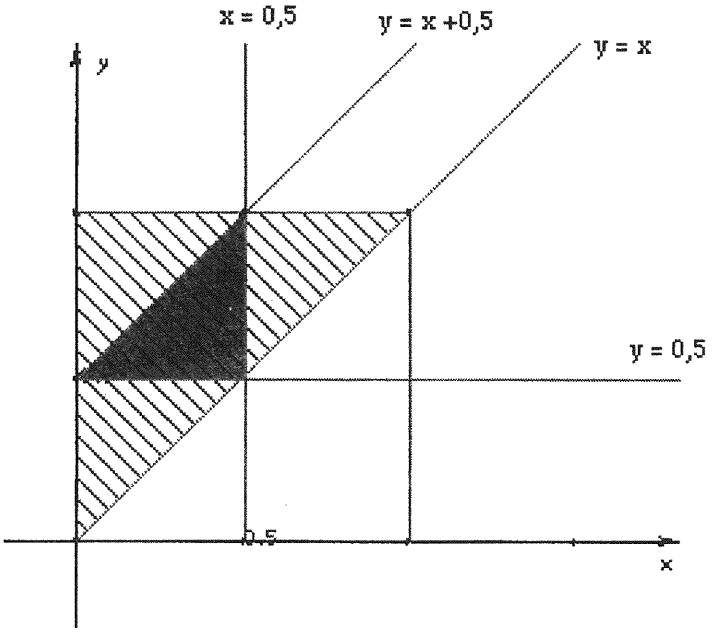
En raison de la symétrie du problème, on peut supposer que x est le plus petit des deux nombres ($x < y$). On obtient ainsi comme ensemble de points possibles le triangle "supérieur" du carré obtenu.

La maximum des trois longueurs qui correspondent aux trois nombres x, y - x et 1 – y doit être inférieur à $\frac{1}{2}$, chacun d'eux doit donc l'être. On a ainsi trois inéquations :

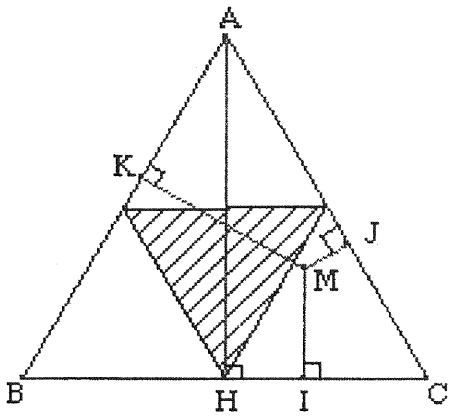
$$x < \frac{1}{2} ; y < x + \frac{1}{2} ; \text{ et } y > \frac{1}{2}.$$

Par régionnement du plan, on obtient le triangle coloré en bleu dont l'aire est égale à $\frac{1}{8}$. Le rapport entre l'aire du triangle coloré et celle du triangle hachuré étant lui de $\frac{1}{4}$.

On trouve une valeur dont s'approchent les distributions de fréquences obtenues par les simulations réalisées avec un tableur.



Une approche géométrique du problème



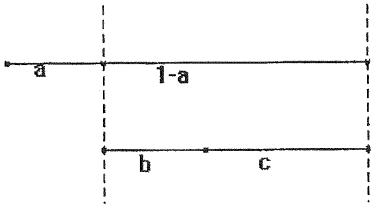
ABC est un triangle équilatéral, M un point quelconque à l'intérieur du triangle. Une considération rapide des aires des triangles MBC, MAB et MAC ainsi que celle de ABC permet d'obtenir : MI + MJ + MK = AH.

Couper un spaghetti de longueur AH correspond donc à placer un point M de façon quelconque dans le triangle. On reporte ensuite les 3 longueurs MI, MJ puis MK sur la longueur AH pour obtenir les découpes.

Et ces trois longueurs permettront de construire un triangle si chacune d'elle est inférieure à la moitié de AH. Ainsi, le point M doit se trouver dans le triangle des "milieux" dont l'aire est bien égale au quart de l'aire du triangle ABC.

Simulation 2

On peut aussi considérer que couper un spaghetti en trois, supposé toujours de longueur 1, c'est couper un premier morceau de longueur a, $a \in [0 ; 1]$, puis un second morceau de longueur $b \in [0 ; 1 - a]$, la longueur du troisième morceau est alors donnée par $c = 1 - (a + b)$. Cette situation correspond à la deuxième façon de couper au hasard le spaghetti qui a été mentionnée.



Pour vérifier que les trois segment obtenus forment un triangle, le test reste le même :

$$\max(a, b, c) < \frac{1}{2}.$$

Avec un tableur, on peut procéder comme ci-dessous :

A	B	C	D	E
Premier morceau	**Deuxième morceau**	**Troisième morceau**	**Le max des 3**	**Test**
ALEA()	(1-A)*ALEA()	1-(A+B)	MAX(A;B;C)	SI(D<0.5;1;0)
0,427001379	0,453740714	0,119257907	0,453740714	0
0,171678282	0,218457105	0,609864613	0,609864613	1
0,348929263	0,029698656	0,621372081	0,621372081	0

On observe pour 8 séries de 500 simulations les fréquences suivantes :

0,194 ; 0,192 ; 0,208 ; 0,24 ; 0,21 ; 0,206 ; 0,208 ; 0,18.

Soit une fréquence moyenne de 0,204 pour 4000 simulations.

Approche graphique :

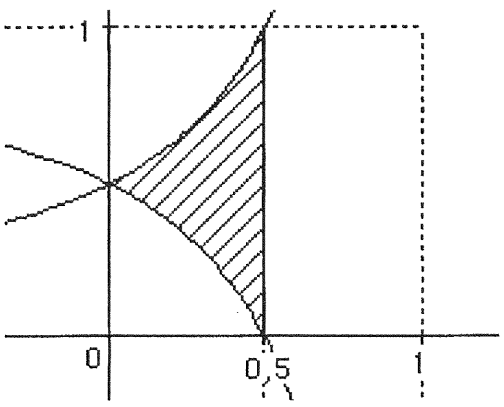
Soit x le premier nombre aléatoire ($x \in [0 ; 1]$) et y le second ($y \in [0 ; 1]$), les longueurs des segments sont $a = x$, $b = (1 - x)\, y$ et $c = 1 - (a + b)$.

La construction du triangle est possible lorsque les trois segments ont des longueurs inférieures à un demi, soit :

$$x < \frac{1}{2} \quad ; \quad (1 - x)\, y < \frac{1}{2} \quad \text{et} \quad x + (1 - x)\, y > \frac{1}{2}$$

où encore : $x < \frac{1}{2}$; $y < \frac{1}{2\,(1 - x)}$ et $y > \frac{1 - 2x}{2\,(1 - x)}$

C'est à dire lorsque les couples (x ; y) sont les coordonnées des points de la zone hachurée ci-dessous. Un petit calcul intégral permet d'établir que l'aire du domaine est égale à $\ln 2 - 0{,}5$, qui à 10^{-3} près vaut 0,193.



Simulation 3

Une autre façon de procéder qui a été donnée considère que couper un spaghetti en trois, c'est en couper un morceau puis recouper le plus grand des 2 morceaux.

De la même façon que précédemment, on peut alors procéder à une simulation à l'aide d'un tableur comme indiqué ci-dessous :

A	B	C	D	E	F
Nombre entre 0 et 1	**Premier morceau**	**Deuxième morceau**	**Troisième morceau**	**Le max des 3**	**Test**
ALEA()	SI(A<0.5;A;1-A)	(1-B)*ALEA()	1-(B+C)	MAX(B;C;D)	SI(E<0.5;1;0)
0,7318545	0,2681455	0,552091211	0,179763289	0,552091211	0
0,907699282	0,092300718	0,658971228	0,248728055	0,658971228	0
0,057691851	0,057691851	0,756667873	0,185640276	0,756667873	0

On observe pour 8 séries de 500 simulations les résultats suivants :

0,364 ; 0,354 ; 0,366 ; 0,392 ; 0,394 ; 0,388 ; 0,396 ; 0,39.

Soit une fréquence moyenne de 0,381 pour 4000 simulations.

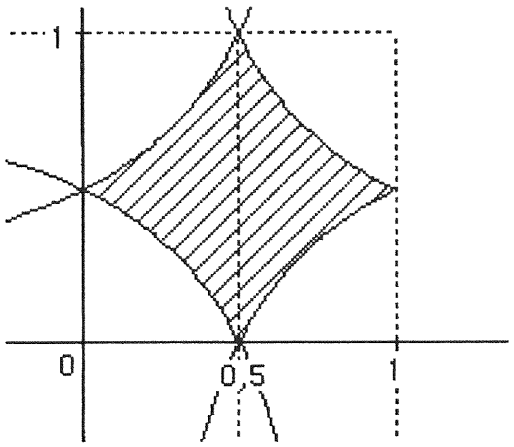
Approche graphique :

Lorsque le premier nombre aléatoire x obtenu est inférieur à $\frac{1}{2}$, la situation est identique à celle de la simulation 2. En appelant y le second nombre aléatoire, les trois morceaux ont pour longueur : a = x, b = (1 – x)y et c = 1 – (a + b). Ce qui conduit au même système d'inéquation.

Lorsqu'il est supérieur, on a : a = 1 - x ; b = xy et c = 1 – (a + b). Il est alors possible d'obtenir un triangle lorsque $x > \frac{1}{2}$; $y < \frac{1}{2x}$ et $y > 1 - \frac{1}{2x}$.

Les points du domaine hachurés de la figure ci-dessous, ont donc des coordonnées (x ; y) qui permettent de construire un triangle. Par symétrie avec le cas précédent, cette aire est :

2 (ln 2 – 0,5) ce qui en valeur approché à 10^{-3} près donne : 0,386.



Simulation 4

Il ne s'agit pas ici exactement d'une simulation mais d'observer le cas qui consisterait pour couper un spaghetti en trois, à couper un morceau puis recouper le plus petit des 2 morceaux.

Dans ce cas, on n'obtiendra jamais de triangle. En effet, en coupant d'abord en deux le spaghetti, l'un des 2 morceaux a une longueur nécessairement supérieure à la moitié de celle du spaghetti (à moins du cas très étrange, où on aurait égalité...), on recoupe ensuite le plus petit morceau. On n'aura donc jamais le maximum des trois longueurs inférieur à la moitié de celle du spaghetti... Ce procédé ne correspond donc pas à une pratique usuelle observée lors de l'expérience, à moins que par jeu les expérimentateurs souhaitent qu'il ne soit jamais possible d'obtenir un triangle.

3. « Ajuster » un modèle

Les différentes simulations effectuées précédemment ne permettent pas d'approcher, et de loin, les résultats expérimentaux obtenus. Diverses raisons sont envisageables, l'une d'elles nous semble être qu'il n'est pas matériellement très réalisable de couper un spaghetti à moins de 3 cm de ses extrémités.

Plaçons nous alors dans le cas où couper le spaghetti en trois consiste à couper un morceau et ensuite recouper le plus grand des deux morceaux, ce qui a semblé être la pratique la plus courante. On va effectuer une simulation en obtenant, de la même façon que dans la simulation 3, deux nombres aléatoires x et y compris entre 0 et 1. Un spaghetti ayant une longueur de 24 cm environ, si l'on ramène cette longueur à l'unité, le premier nombre x obtenu ne pourra correspondre à une découpe « réelle » que s'il est compris entre $\frac{1}{8}$ et $\frac{7}{8}$.

Selon que $x < \frac{1}{2}$ (respectivement $x > \frac{1}{2}$), le premier morceau a une longueur a égale à x (respectivement 1 – x). On recoupe le deuxième morceau de longueur 1 – a à l'aide d'un nombre aléatoire y et (1 – a) y représente la longueur b du deuxième morceau. Mais cette longueur b ne peut correspondre à une expérience « réelle » que si elle est comprise entre $\frac{1}{8}$ et $\frac{7}{8}$.

On calcule alors la troisième longueur c et on applique le test : max (a, b, c) $< \frac{1}{2}$.

Avec un tableur, il est donc nécessaire d'introduire, outre les tests portant sur x et b, un test permettant de compter le nombre de cas « valides ».

x	test : 1/8<x<7/8	premier morceau : a	Deuxième morceau : (1-a)y = b	test : 1/8<b<7/8	Troisième morceaux c	nombre de cas valides	Max (a;b;c)	test
ALEA()	SI(1/8<A;SI (A<7/8;A;-10);-10)	SI(B<0,5;B; 1-B)	(1-C) * ALEA()	SI(1/8<D;SI (D<7/8;D;-10);-10)	1 - (C+E)	SI(F<1;1;0)	MAX (C;E;F)	SI (H<0,5;1;0)
0,02431597	-10	-10	0,20297682	0,20297682	10,7970232	0	10,7970232	0
0,43272211	0,43272211	0,43272211	0,55318897	0,55318897	0,01408892	1	0,55318897	0
0,29685293	0,29685293	0,29685293	0,42466687	0,42466687	0,2784802	1	0,42466687	1
0,62102075	0,62102075	0,37897925	0,15907705	0,15907705	0,4619437	1	0,4619437	1
0,90139942	-10	-10	9,40542877	-10	21	0	21	0
0,54914113	0,54914113	0,45085887	0,17923429	0,17923429	0,36990684	1	0,45085887	1
0,6896084	0,6896084	0,3103916	0,58991741	0,58991741	0,09969099	1	0,58991741	0
0,47050858	0,47050858	0,47050858	0,04031014	-10	10,5294914	0	10,5294914	0
0,88331367	-10	-10	2,5837216	-10	21	0	21	0
0,00494389	-10	-10	8,41750082	-10	21	0	21	0
0,50910426	0,50910426	0,49089574	0,06324865	-10	10,5091043	0	10,5091043	0

On a ainsi réalisé 8 séries de 1000 simulations :

									Total
Nombre de cas valides	612	595	630	617	599	611	604	607	4875
Nombre de triangles	323	322	362	354	337	336	333	347	2714
Fréquence	0,5277	0,5411	0,5746	0,5737	0,5626	0,5499	0,551	0,5716	0,557

Pour 8000 simulations, le nombre de cas « valides », ou correspondant à ce que nous avons considéré comme possible réellement, est de 4875, et la fréquence d'obtention d'un triangle est alors de 0,557. On peut également observer que la fréquence du nombre de cas valides pour les 8000 simulations est de 4875/8000 = 0,609.

Approche graphique :

Le couple (x ; y) de nombres appartenant à l'intervalle [o ;1] satisfait cette fois-ci aux conditions supplémentaires :

- Pour x : $\frac{1}{8} < x < \frac{7}{8}$.

- Pour y, si $\frac{1}{8} < x < \frac{1}{2}$ alors a = 1 – x et on doit avoir $\frac{1}{8} < b = (1 - x)\, y < \frac{7}{8}$, d'où :

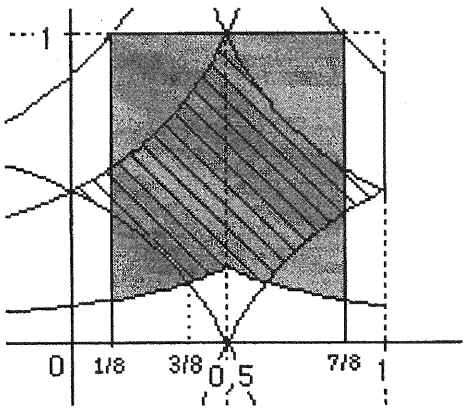
$$\frac{1}{8(1-x)} < y < \frac{7}{8(1-x)}.$$

Il suffit de compléter par symétrie pour x supérieur à $\frac{1}{2}$ et les couples (x ; y) qui conviennent correspondent aux coordonnées des points de la zone colorée. Zone dont l'aire est en valeur approchée à 10^{-3} près de 0,610.

Pour obtenir les couples qui permettent d'obtenir des triangles, on reporte les contraintes déjà observées lors de la simulation 3, soit :

$$\frac{1 - 2x}{2(1 - x)} < y < \frac{1}{2(1 - x)}.$$

La zone hachurée correspond à ces points. Son aire est de 0,340 en valeur approchée et la fréquence théorique d'obtention des triangles est donc de 0,558 environ.



Le modèle proposé ici a donc permis d'ajuster, même si les résultats ne sont pas encore concluants, la fréquence théorique et l'observation expérimentale.

4. Conclusion

L'étude de ce problème nous a amené à préciser les relations et interactions entre expérience, simulation et modèle en nous confrontant aux divers aspects liés à l'étude statistique d'une expérience aléatoire.

1. L'expérience

L'imprécision du protocole expérimental : « couper au hasard un spaghetti en 3 » a permis de s'interroger sur la signification, pour cette expérience, du hasard. D'une part des contraintes matérielles influencent les découpes, d'autre part diverses façons de faire peuvent se manifester. On peut bien sûr essayer de réduire le nombre de paramètres, objectifs ou subjectifs, susceptibles d'intervenir dans l'expérience, en précisant les modalités de son déroulement. Par exemple dans le cas présent, on aurait pu préciser : on coupe le spaghetti en deux, puis on recoupe le plus grand des deux morceaux. Mais vouloir observer et analyser un phénomène aléatoire c'est malgré tout essayer de prendre en compte divers paramètres même si ceux-ci sont a priori difficiles à prendre en compte...

2. Le processus de modélisation

Son objectif est de représenter et simplifier grâce à un symbolisme mathématique des situations réelles afin de pouvoir effectuer des calculs, pour dégager des propriétés ou prévoir des évolutions. Ces choix de simplification sont arbitraires et sont acceptés par la communauté qui travaille sur le problème, si celle-ci considère qu'ils ne modifient pas fondamentalement la situation.

Dans le cas de l'expérience étudiée, tous les modèles considèrent que le spaghetti est un segment de longueur 1. Première simplification qui peut paraître naturelle à des enseignants de mathématiques, mais qui fera certainement l'objet dans une classe de discussion avec des élèves non spécialistes à priori. Ce choix n'est pas indépendant de la simulation envisagée à l'aide d'un outil informatique qui permet d'obtenir des nombres aléatoires entre 0 et 1, puisque la deuxième simplification consiste à dire que les deux coupures correspondent au choix de 2 nombres aléatoires x et y compris entre 0 et 1 (confer simulation 1). L'outil (ou théorie) mathématique qui permet ensuite de réaliser des calculs étant l'équirépartition des nombres sur les intervalles considérés. L'utilisation de surfaces associées facilite ensuite les calculs.

Les modèles des simulations suivantes proposent quelques rectificatifs à ces premières simplifications en modifiant par exemple l'utilisation du second nombre aléatoire y demandé, ou en modifiant l'intervalle auquel appartienne ces nombres (simulation 5) mais dans tous les cas l'équirépartition est un invariant. On aurait pu, par exemple, essayé d'imaginer un modèle avec une fréquence d'apparition du nombre x déterminant la première coupure ayant un profil continu mais en cloche entre 0 et 1.

3. La simulation

La simulation permet d'obtenir grâce à l'outil informatique des séries importantes, mais les distributions de fréquences obtenues peuvent ne pas approcher du tout l'expérience réelle. Ces simulations ne pourront donc permettre de valider le modèle inévitablement associé.

4. La validation du modèle

De façon générale, nous reprendrons ici l'expression de Nicolas BOULEAU [BOULEAU, 1999] pour qui un modèle est un simulacre utile et partisan, partisan car « au sein de multiples possibilités d'expression et de représentation, ils sont le choix d'un parti ». Ainsi la question de la validité du modèle englobe divers registres. Nous pouvons nous limiter dans un premier temps à la question de l'adéquation du modèle et de ce qu'il est censé représenter. Pour notre exemple, la tentative d'ajustement (paragraphe 3) permet peut-être de rendre un peu mieux compte de l'expérience, mais quels sont les critères d'adéquations qui peuvent être retenus ? Ceci est encore une autre question, pour d'autres problèmes et d'autres analyses.

En définitive, ce type de problème nous semble permettre de créer, dans une classe de seconde de lycée, un débat sur les conditions de réalisation d'une expérience, sur la réalisation concrète d'une simulation à l'aide d'outils informatiques (tableurs ou calculatrices) qui suppose des choix simplificateurs auxquels les élèves devront se confronter et sur les écarts entre expérience réelle et simulation. On peut enfin citer d'autres problèmes de même type qui permettront de soulever ces mêmes questions : « comment partager un spaghetti en quatre pour pouvoir construire un quadrilatère ? en six pour pouvoir construire une pyramide ? », on change d'espace, de modèle mais entre expérience, simulation et modélisation les mêmes interactions ont lieu.

BIBLIOGRAPHIE

ALDON G., FEURLY-REYNAUD J., 1994, « *Modélisation en probabilité au lycée* ». IREM de Lyon.

BERNARD R., 1996, « *Observer et agir, mathématiques en seconde générale et technologique* ». IREM de Montpellier.

BERNARD R., FAURE C., NOGUES M., NOUAZE Y., TROUCHE L., 1998, « *Pour une prise en compte des calculatrices symboliques en analyse au lycée* ». IREM de Montpellier.

BOULEAU N., 1999, « *Philosophies des mathématiques et de la modélisation* ». Editions l'Harmattan.

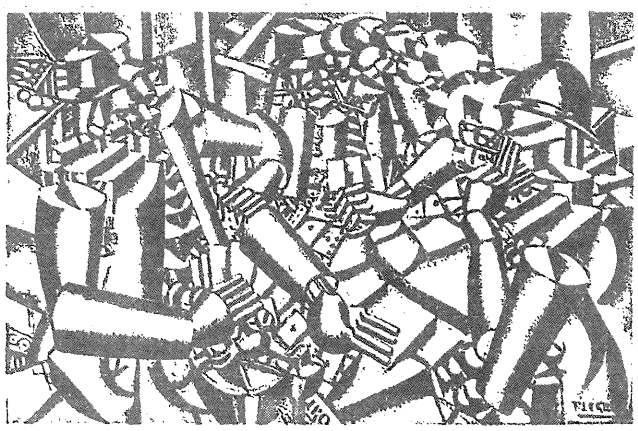
GIRARD J.-C., 1999, « *Le professeur de mathématiques doit-il enseigner la modélisation ?* ». Repères-IREM, n° 36, pp. 7-14, Topiques Editions.

HENRY M., 1999, « *L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie* ». Repères-IREM, n° 36, pp. 15-34, Topiques Editions.

HUGUET F., 1999, « *Simulation et modélisation d'un problème de bouteilles* ». , Repères-IREM, n° 36, pp. 35-42, Topiques Editions.

HAMMERSLEY L.M. & HANDSCOMB D.C., « *Les méthodes de Monte-Carlo* ». Monographies DUNOD.

TROUCHE L., 1998, « *Expérimenter et prouver : Faire des mathématiques au lycée avec des calculatrices symboliques* ». IREM de Montpellier.



Soldiers Playing at Cards (1917) by Fernand Léger.

LE JEU DE FRANC-CARREAU
EXPERIMENTATION ET SIMULATION

Des fréquences observées à la fréquence théorique

Marie-Claire COMBES - Michel LACAGE - Jean-Marc RAVIER
François ROUX - Jacques SALLES

Equipe Intégration des Outils Informatiques
IREM de Montpellier

Résumé : *Le jeu de franc-carreau consiste à lancer une pièce au hasard sur un damier ; on gagne si la pièce ne rencontre aucune des lignes du quadrillage (on dit alors que la pièce est à franc-carreau) . Une idée d'activité pour la classe, avec tous les documents d'appui.*

Les « chances » de gagner dépendent bien évidemment de la taille de la pièce et des carreaux qui constituent le damier.

La présentation du jeu est simple, l'expérimentation en classe (ou mieux en demi-classe) facile à réaliser, avec un matériel réduit (voir le document professeur dans les pages suivantes). Chaque élève dispose alors d'un échantillon de cette expérience et peut calculer la fréquence de franc-carreau qui lui est associée ; la collecte des résultats du groupe conduit à observer la fluctuation d'échantillonnage.

Les outils informatiques permettent de réaliser une simulation de cette expérience présentant un lien très étroit avec celle-ci. Grâce à cette simulation, on peut disposer d'échantillons de grande taille et observer que les fréquences se stabilisent. Mais existe-t-il une fréquence « théorique » de l'événement franc-carreau, et quelle peut-elle être ? La réponse repose entièrement sur le choix du modèle qui a servi à cette simulation. C'est alors qu'une animation pertinente devient un **révélateur du modèle** (voir le document professeur et l'annexe). Dans ce sens, ce thème d'étude respecte « le choix pédagogique du programme de statistiques de seconde d'aller de l'observation vers la conceptualisation » recommandé dans le document d'accompagnement de ce programme.

La démarche adoptée pour la présentation en classe de ce thème est celle propre au dispositif d'enseignement autour d'un fichier rétroprojetable développé par le groupe Intégration des Outils Informatiques de l'IREM de Montpellier.

- **Un document élève** est distribué en début de séance ; il permet à chacun de réaliser une expérimentation et de noter les fréquences associées à chacun des échantillons ainsi constitués dans la classe.
- **Deux figures animées**, réalisées sur ordinateur avec le logiciel Cabri-Géomètre, ainsi qu'**un fichier Excel**, sont ouverts en classe, **en rétroprojection** (à l'aide d'un vidéoprojecteur, d'une tablette couplée à l'ordinateur ou d'un grand écran de télévision)[1] et pilotés par le professeur. Celui-ci régule en outre le dialogue et note les conjectures.

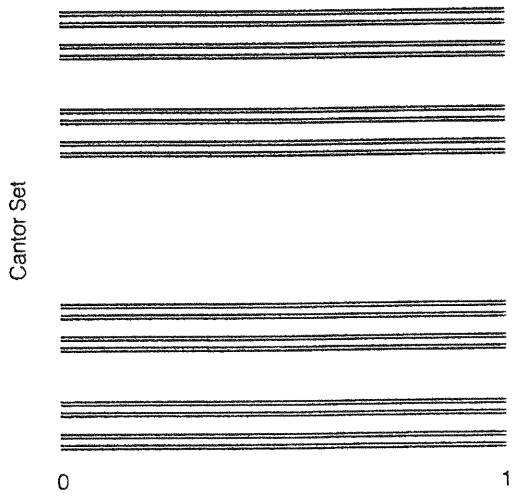
[1] La rétroprojection constitue un choix d'utilisation en classe de l'outil informatique retenu pour ses atouts, parmi lesquels : favoriser le dialogue dans la classe, créer des situations de débat, exploiter l'interactivité permise par la présence d'objets mobiles dans le logiciel Cabri ; on trouvera un développement de cette démarche dans le rapport

- **Une fiche technique** (page suivante) décrit en détail les figures animées.
- **Un document professeur** apporte des compléments sur l'utilisation des fichiers informatiques et le déroulement de la séance.

Le lecteur pourra enfin trouver en annexe quelques écrans saisis lors de la phase de rétroprojection. Les trois fichiers informatiques nécessaires à la simulation peuvent être téléchargés sur le site http://www.irem.univ-montp2.fr, groupe Intégration des Outils Informatiques.

Deux programmes de simulation écrits pour les calculatrices Texas Instruments TI 83 permettent à chaque élève (pour les établissements où l'on dispose d'un parc de calculatrices) de réaliser un échantillon de grande taille, puis, après le recueil de tous les résultats, de représenter graphiquement l'intervalle de confiance au niveau 95% (par exemple) et de dénombrer les échantillons qui sortent de cet intervalle.

Cette partie fera l'objet d'un complément de publication ultérieur (consulter le site ci-dessus).



Type : Simulation d'une expérience, aide à la conjecture

Nom : francarreau 8c.fig, francarreau.fig et francarreau.xls

Intention : Proposer une expérience et un exemple de simulation de cette expérience.
Calculer la fréquence de succès à l'issue de n lancers ($n \leq 999$).
Prendre conscience de la fluctuation d'échantillonnage.
Concevoir la fréquence théorique (probabilité) de succès associée à la modélisation choisie.

Description et actions possibles :

A. La figure **[francarreau 8c.fig]** présente un damier de 8 carreaux de 5 cm de côté au-dessus duquel apparaît un curseur à trois positions.

En position gauche, la pièce est matérialisée par un disque de rayon 1 cm, dont le centre est un point situé à l'intérieur du damier (annexe, écran 1). En sélectionnant le ***nombre*** d'essais, et en pressant la touche - (moins), la pièce est relancée (écrans 2 et 3).

En position médiane, le curseur fait apparaître l'ensemble de tous les centres possibles (pour un certain protocole d'expérimentation, voir document professeur) ; en position droite s'affiche l'ensemble des centres dans les cas de franc-carreau. La présentation en classe de ces deux affichages ne doit intervenir qu'en fin de séance, après l'utilisation de la figure [francarreau.fig].

B. La figure [**francarreau.fig**] comporte au bas un curseur à deux positions ; ouverte en **position gauche**, elle présente un unique carreau de 5 cm de côté. Sur la droite, une table à deux colonnes : essai et franc-carreau (annexe, écran 5). En sélectionnant le ***nombre*** d'essais, et en pressant la touche - (moins), la pièce est relancée (écrans 6 et 7).

Une simulation de plusieurs lancers de la pièce (jusqu'à 999) peut être réalisée grâce à l'animation du nombre d'essais ; les résultats (nombres d'essais et de franc-carreaux) s'affichent dans la table ; leur importation dans le fichier [**francarreau.xls**] permet de lire la fréquence de franc-carreau à l'issue de la simulation effectuée (voir les écrans commentés 8 à 11).

C. Revenant à la figure Cabri [**francarreau.fig**], la **position droite** du curseur fait apparaître deux carreaux dans lesquels s'affiche à l'identique le résultat d'un essai (bord et centre de la pièce) : annexe écran 12.

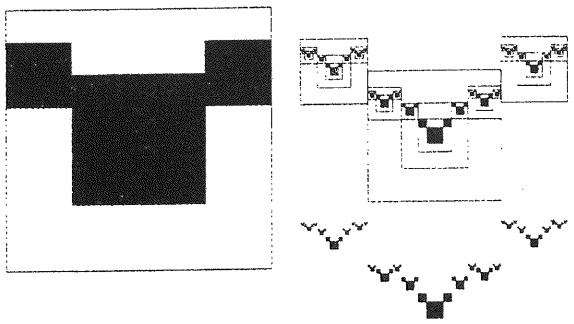
Description et actions possibles (suite) :

Une simulation d'un grand nombre de lancers (il n'y a pas ici de limitation) obtenue par l'animation du nombre d'essais, après avoir mis les centres des deux pièces en mode trace, permet de **suivre l'évolution de la fréquence** de franc-carreau (rapport des nombres de points affichés dans les deux carreaux) lorsque le nombre de lancers augmente.
La densification des deux ensembles de centres permet de conjecturer leur forme (carrée), ce qui engage à en apporter la preuve et à déterminer les dimensions (annexe, écrans 13 à 15).

L'exploitation de l'animation et et d'un débat sur la notion de *hasard* dans le lacer de la pièce sur le damier permettent d'introduire la fréquence théorique (probabilité) de l'événement franc-carreau associé au modèle choisi pour la simulation (voir le document professeur).

D. Le retour à la figure [**francarreau 8c.fig**], avec le curseur placé en position médiane, puis droite (voir A. ci-dessus), offre la possibilité d'évoquer l'éventuelle indépendance de cette fréquence théorique avec le nombre de carreaux du damier, la taille des carreaux et de la pièce demeurant inchangée (annexe, écrans 16 et 17 et document professeur).

Scénario possible : Voir pages suivantes.



Prérequis Connaître la définition de la fréquence d'un événement.

Matériel nécessaire

Une vingtaine d'exemplaires du damier (35 carrés de 5 cm de côté) de format A3 (à préparer soi-même, ou à imprimer à partir de deux exemplaires du document Word [demidamier.doc]) et autant de pièces de 10 centimes (de Franc !).

Activité *papier-crayon*

Chaque élève, lors d'une séance en demi-classe, reçoit avec la fiche élève un damier et une pièce de 10 centimes. Lorsque chacun dispose d'un échantillon de 20 expériences et a calculé la fréquence de franc-carreau pour son échantillon, les résultats sont collectés et représentés sur un même graphique ; la fréquence de franc-carreau sur l'ensemble des échantillons est calculée. On observe la **fluctuation d'échantillonnage**.

Rétroprojection

La figure [francarreau 8c.fig] est ouverte dans le logiciel Cabri, le curseur en position gauche. On reconnaît un damier analogue à celui que les élèves ont utilisé lors de l'expérimentation : les dimensions de la pièce et des carreaux sont identiques, seul le nombre de carreaux diffère. Il en est de même dans la figure [francarreau.fig] où le damier se réduit à un unique carreau. Le passage de l'expérimentation (par chaque élève) à deux simulations de cette expérimentation (sur ordinateur, l'une avec un damier à huit carreaux, l'autre sur un unique carreau), n'a rien d'immédiat ; on veillera à poser la question du bien-fondé de ces simulations. Quelques éléments de réponse seront discutés en classe à la fin de la partie rétroprojection (voir ci-dessous).

Le déroulement de la rétroprojection est noté dans la fiche technique (description et actions possibles) et dans l'annexe.

Au cours de la rétroprojection, les élèves sont amenés à débattre de ce que l'on entend par « ***lancer la pièce au hasard sur le damier*** ». Faire un choix de modèle consiste à donner ***un*** sens mathématique à cette phrase ; ici, le choix du modèle est celui de l'**équiprobabilité** des points du damier : tous ces points (pixels sur l'ordinateur ou la calculatrice) ont la même « chance » de devenir le centre de la pièce lancée. L'animation des lancers simulés de la pièce suivant ce modèle aide à concevoir une fréquence théorique de l'événement franc-carreau : rapport des aires de deux carrés.

Il convient également de préciser le protocole expérimental, qui n'a volontairement pas été explicité sur le document élève. En particulier se pose la question d'acceptabilité des lancers. On amène les élèves à présenter leur choix lors de l'expérimentation ; peut-être ont-ils rejeté l'essai lorsque la pièce roule au sol, ou même dès qu'elle quitte la feuille sur laquelle est représenté le damier ; ont-ils retenu l'essai dans le cas où la pièce touche le quadrillage, son centre étant situé à l'extérieur du damier ? Le travail mené lors de ce débat va permettre de

choisir une condition d'acceptabilité des lancers qui rende la probabilité d'obtenir franc-carreau indépendante du nombre de carreaux du damier. Lors de l'expérimentation initiale (damier papier et pièce de 10 centimes), il peut paraître naturel de rejeter les essais lorsque la pièce est entièrement à l'extérieur du damier. Si l'on s'en tient à ces seuls cas de rejet des essais, passer d'un damier de 35 carreaux (papier) à celui de 8 carreaux (simulé dans [francarreau 8c.fig]) ou à celui à carreau unique ([francarreau.fig]) modifie la probabilité d'obtenir franc-carreau. En effet, si l'ensemble F des centres des pièces à franc-carreau est inchangé, (cet ensemble « des cas favorables », réunion de carrés de 3 cm de côté, apparaît dans la figure [francarreau 8c.fig], curseur en position droite), par contre l'ensemble E des centres des pièces pour tous les essais acceptés (ensemble « des cas possibles ») est le damier tout entier, bordé d'une bande de 1 cm de côté ! Dans ce cas les quotients « $\frac{\text{nombre d'éléments de F}}{\text{nombre d'éléments de E}}$ » (nombres de pixels lors de la simulation sur ordinateur ou calculatrice) ou « $\frac{\text{aire de F}}{\text{aire de E}}$ » ne sont pas indépendants du nombre de carreaux du damier. Cette invariance est obtenue si l'on convient de rejeter tous les essais pour lesquels le centre de la pièce est situé à l'extérieur du damier (l'ensemble « des cas possibles » est alors celui montré par la position médiane du curseur de la figure [francarreau 8c.fig]. Un tel choix justifie et éclaire, a posteriori, l'hypothèse évoquée lors du passage de l'expérimentation à la simulation dans [francarreau 8c.fig] et nécessairement formulée lors de l'ouverture de [francarreau.fig] : *l'on ne change pas les « chances » d'obtenir franc-carreau en lançant la pièce sur un damier constitué d'un seul carreau.*

Complément

Une simulation sur calculatrice graphique, à l'aide de deux programmes, a été préparée et peut être réalisée par les élèves ; ce choix est facilité si l'établissement dispose d'un parc d'une vingtaine de calculatrices de même type.

Vocabulaire à préciser à l'issue de ce thème d'étude

L'expérience aléatoire est ici le lancer au hasard d'une pièce de 1 cm de rayon sur un damier à un carreau de 5 cm de côté, en convenant de rejeter les lancers pour lesquels le centre de la pièce est à l'extérieur du carré.

Le modèle retenu est l'équiprobabilité des points du carré.

La simulation du modèle consiste à choisir deux nombres aléatoires x et y compris entre 0 et 5 et à représenter le point de coordonnées (x,y) dans le repère (A,B,D) (si ABCD désigne le carré).

document-élève

LE JEU DE FRANC-CARREAU

Pour jouer au jeu de franc-carreau, on dispose d'un damier constitué de carreaux de forme carrée de 5 cm de côté et d'une pièce de 10 centimes de Franc, dont le rayon est 1 cm.

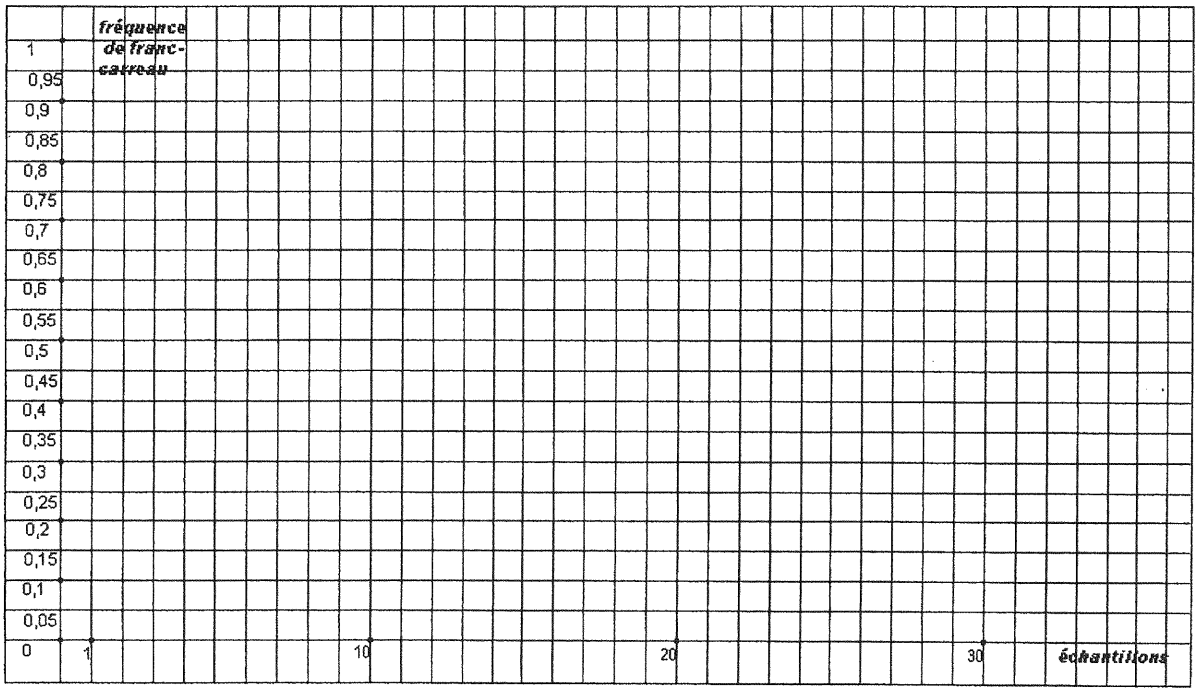
On lance la pièce au hasard sur le damier.

On gagne si la pièce ne chevauche pas les lignes du quadrillage, on dit alors que la pièce est à *franc-carreau*.

1. Réalisez 20 expériences (un lancer de la pièce constitue une expérience) ; notez le nombre de franc-carreaux. L'ensemble de ces 20 expériences constitue un échantillon. Calculez la fréquence de franc-carreau associée à cet échantillon.

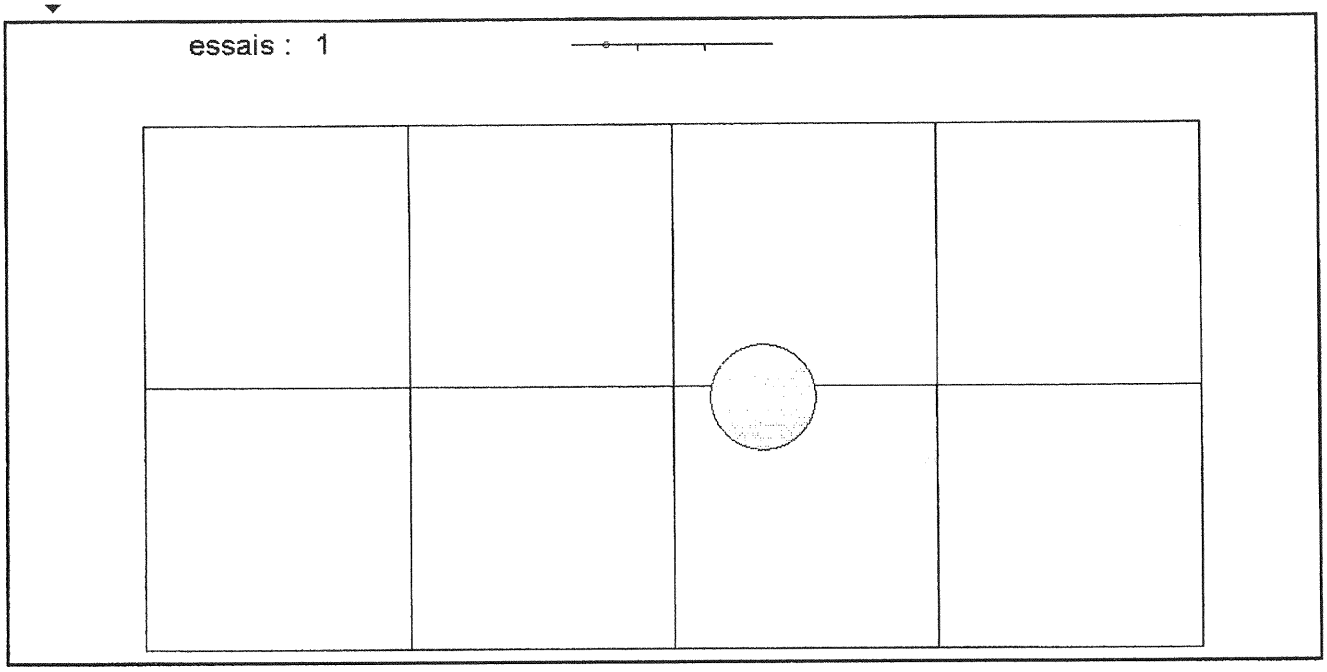
En collectant les résultats de tous les élèves, on dispose alors d'un certain nombre d'échantillons et de fréquences.

2. Sur le graphique ci-dessous, marquez d'un point la fréquence de chaque échantillon (dont le votre), calculez la fréquence f de franc-carreau sur l'ensemble des échantillons et tracez la droite d'équation $y = f$.



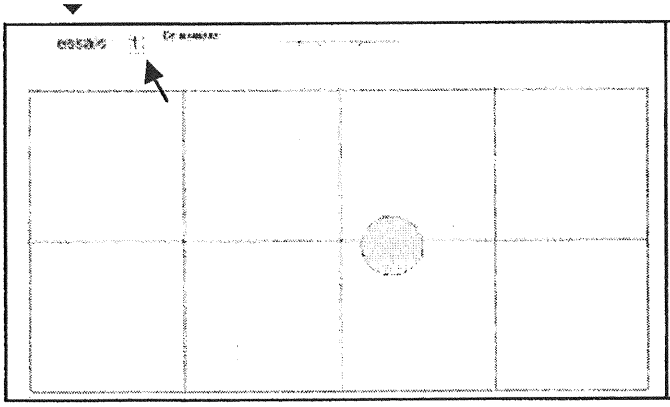
ANNEXE

Écran 1 **[francarreau 8c.fig]** En **position gauche**, le curseur affiche le damier et le résultat d'un essai.

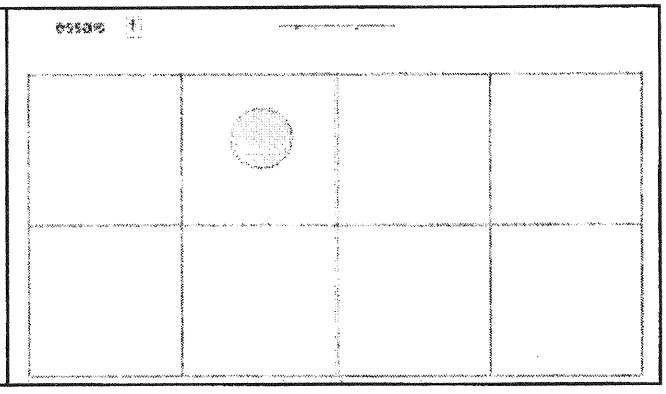


En sélectionnant le ***nombre*** d'essais, et en pressant la touche = (moins),
la pièce est relancée (elle est à franc-carreau sur l'écran 3).

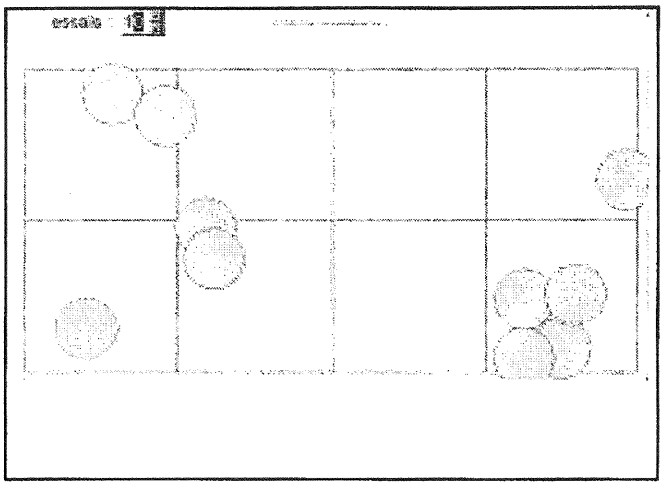
Écran 2



Écran 3



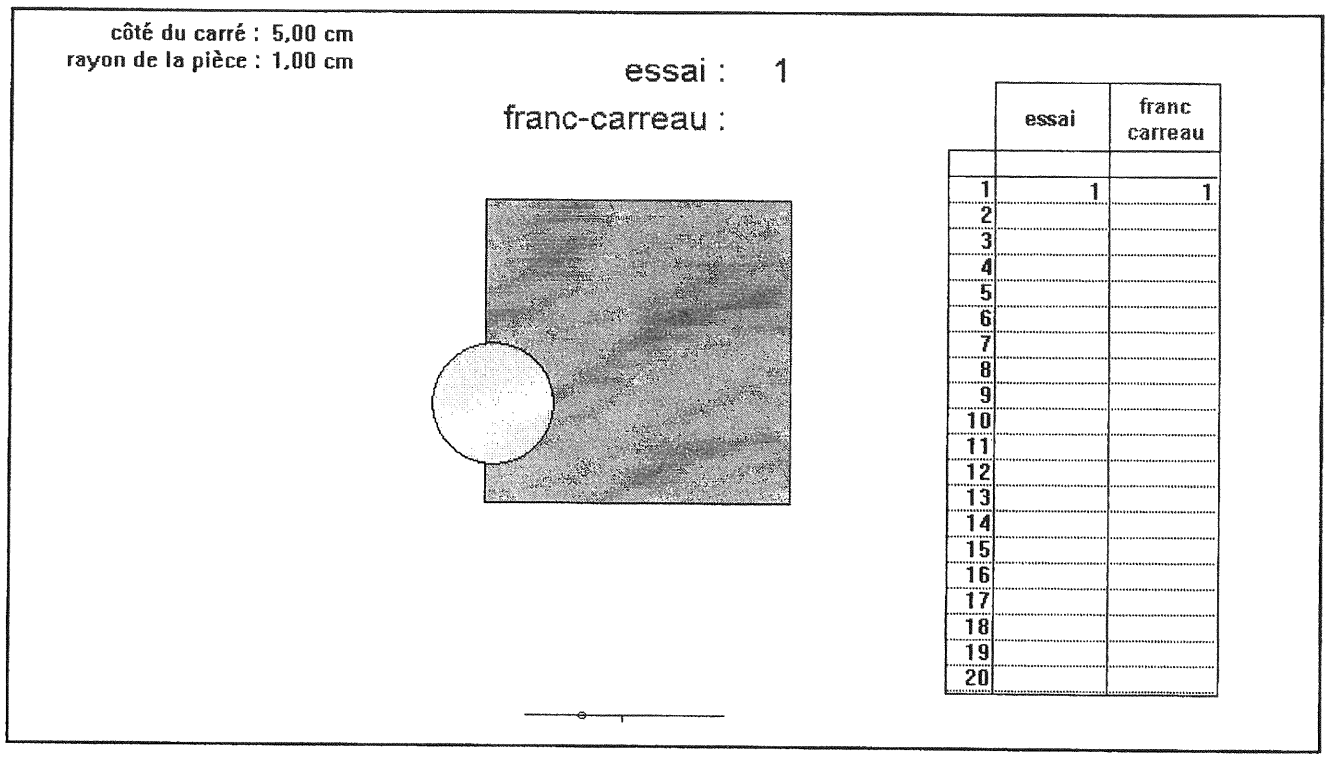
Écran 4



La pièce étant mise en mode **trace** (sélectionner "ce cercle"), on peut effectuer plusieurs essais (10 par exemple sur l'écran 4), en double cliquant sur le ***nombre*** d'essais puis en cliquant plusieurs fois sur la flèche ▲ . On peut enfin observer le nombre de franc-carreaux obtenus.

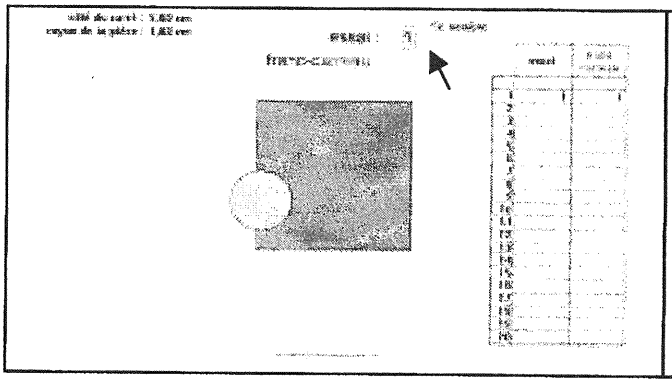
N.B. L'usage de cette fonctionnalité est délicat ; le déplacement de la souris provoque en effet une actualisation du nombre d'essais et donc l'arrivée intempestive de nouvelles pièces ; si l'on souhaite s'en servir, il convient de laisser la souris sur le compteur d'essais sans la déplacer.

Écran 5 **[francarreau.fig]** En **position gauche**, le curseur affiche un carreau et le résultat d'un essai.

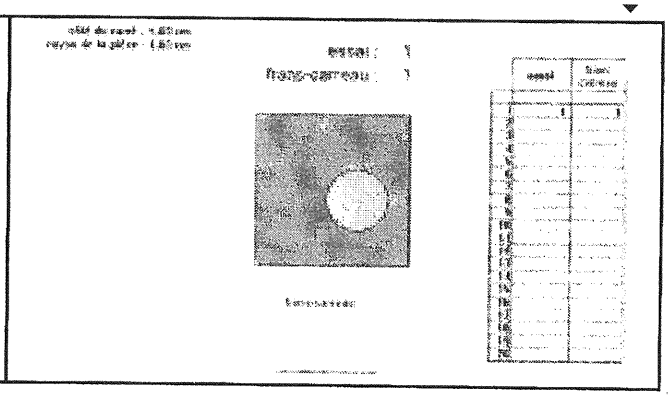


En sélectionnant le ***nombre*** d'essais, et en pressant la touche = (moins), la pièce est relancée (elle est à franc-carreau sur l'écran 7).[2]

Écran 6



Écran 7



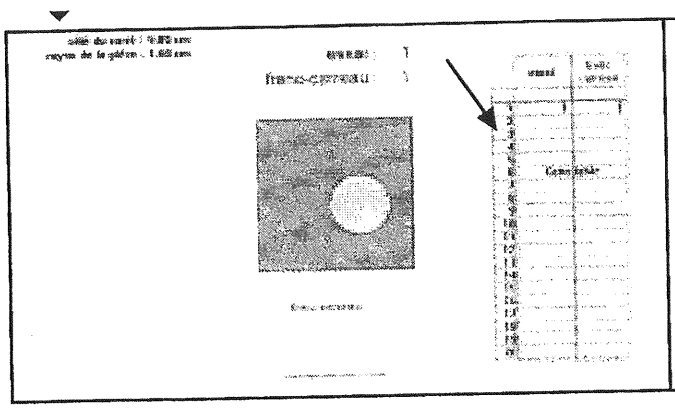
[2] N.B. Le résultat de chaque simulation du lancer d'une pièce apparaît à l'écran, à droite du texte « franc-carreau » : 1 s'affiche en cas de succès, ou rien dans le cas contraire. La table, située sur la droite de l'écran, sert à collecter les résultats lors d'une animation (écrans 8 et suivants) ; elle ne s'actualise pas à chaque simulation (touche = (moins), le ***nombre*** d'essais étant sélectionné). On peut observer que la première ligne des colonnes « **essai** » et « **franc-carreau** » de la table comporte toujours deux chiffres 1 : ils servent à initialiser la table pour la collecte des résultats. Bien évidemment ce premier essai n'a rien d'aléatoire : il a fallu placer la pièce à franc-carreau afin d'initialiser la colonne correspondante de la table ; ce premier essai ne sera donc pas pris en compte pour le calcul de la fréquence de franc-carreau dans [franc-carreau.xls].

L'animation du ***nombre*** d'essais permet de simuler jusqu'à 999 lancers de la pièce :

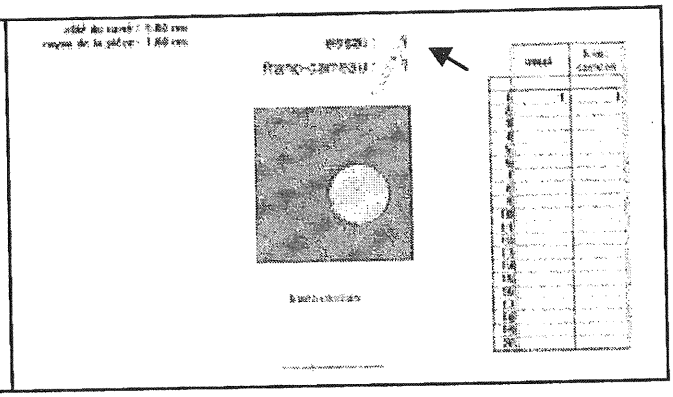
1. sélectionner "**cette table**" (et pas "ce polygone") ;
2. cliquer sur le bouton "animation", sélectionner le ***nombre*** d'essais et tendre le ressort.

N.B. La réalisation des 999 essais demande près d'une minute.

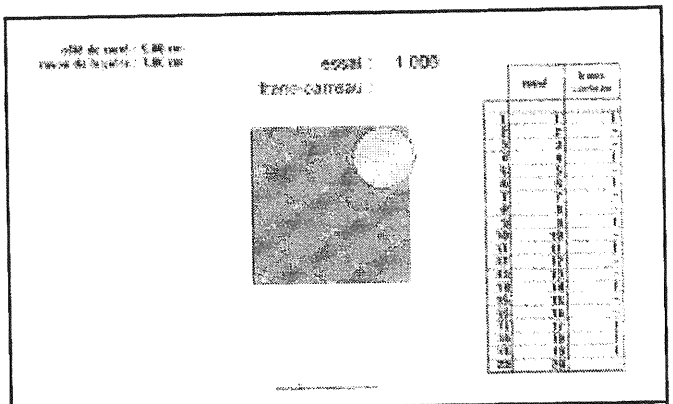
Écran 8



Écran 9



Écran 10



Dans Cabri, la taille maximum de la table est 999 ; en pratique, on peut arrêter l'animation lorsque le compteur atteint ou dépasse 1000 : seuls les 999 premiers résultats seront retenus. A l'écran s'affichent les 20 premiers, mais tous sont mémorisés [3].

Chaque fois que la pièce est à franc-carreau, 1 s'affiche dans la colonne correspondante de la table ; ainsi, sur l'écran 10 le premier essai était à franc-carreau ainsi que les quatrième, sixième, neuvième, etc.

On sélectionne alors la table puis on la copie (Ctrl C).

On ouvre [**francarreau.xls**] et on la colle (Ctrl V).

La fréquence de franc-carreau à l'issue de ces 999 essais est calculée dans la cellule G3, le premier n'étant pas comptabilisé (voir la note 2).

Écran 11

essais	franc-carreau oui : 1
1	1
2	--
3	--
4	1
5	--
6	1
7	--
8	--
9	1
10	--
11	1
12	--
13	--
14	1
15	--

nombre total d'essais	nombre de franc-carreaux	pourcentage de franc-carreaux
998	367	36,77355%

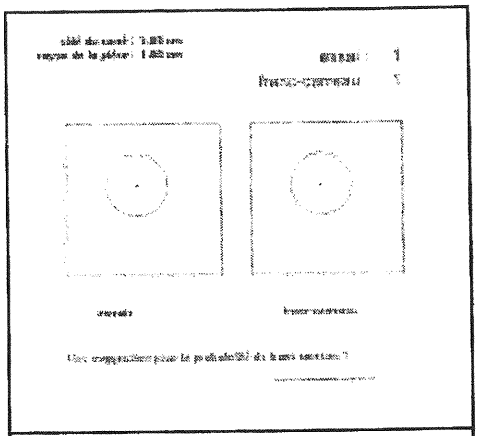
*Remarque : pour une utilisation nouvelle ou ultérieure, il peut être commode de remettre la table de [francarreau.xls] dans l'état initial ; il suffit pour cela d'effacer le contenu des cellules recopiées depuis Cabri (**sélectionner** les cellules **B3 à C1001** à l'aide de la souris, des touches « majuscule », « page inférieure » et des flèches de direction du clavier puis **effacer le contenu**). On peut aussi, très simplement, fermer le classeur [francarreau.xls] sans enregistrer !*

---

[3] Si le compteur s'incrémente de 10 (et non de 1 comme on pourrait l'attendre), on peut, sans interrompre l'animation, attendre que le compteur atteigne 10000 (au lieu de 1000). Pour retrouver une incrémentation de 1, il convient, en double cliquant sur le **nombre** d'essais, de placer le curseur à droite du chiffre des unités (et non à gauche).

[francarreau.fig]

Écran 12



La **position droite** du curseur fait apparaître deux carreaux dans lesquels s'affiche à l'identique le résultat d'un essai (bord et centre de la pièce). Pour bien montrer cette duplication, on peut effectuer quelques lancers (sélectionner le **nombre** d'essais et presser la touche = (moins)).

On place alors les centres des deux pièces en mode **trace** et on lance l'animation du nombre d'essais : 50 essais dont 15 franc-carreaux sur l'écran 13.

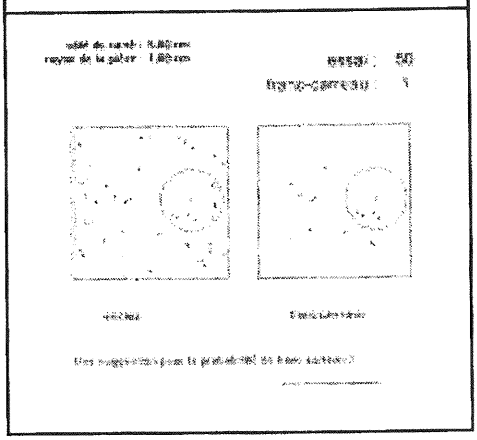
Remarques :

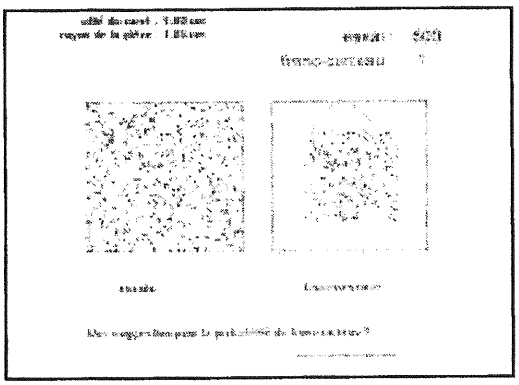
1) Il peut arriver que le centre de l'une des pièces n'apparaisse pas : cela se produit normalement chaque fois que l'animation est interrompue sur un lancer qui n'est pas à franc-carreau. Afin de le faire réapparaître, ce qui est nécessaire pour vérifier s'il est en mode ***trace****, il suffit de réaliser quelques lancers (sélectionner le* ***nombre*** *d'essais et presser la touche = (moins)) jusqu'à l'obtention de franc-carreau.*

2) L'animation peut être interrompue autant de fois qu'on le veut (d'un clic de souris à n'importe quel endroit de la figure), pour prendre conscience par exemple de l'interprétation graphique de la fréquence de franc-carreau après un certain nombre d'essais (rapport des nombres de points affichés dans les deux carreaux).

L'animation ***se poursuit*** *en reprenant l'outil "animation".*

Écran 13



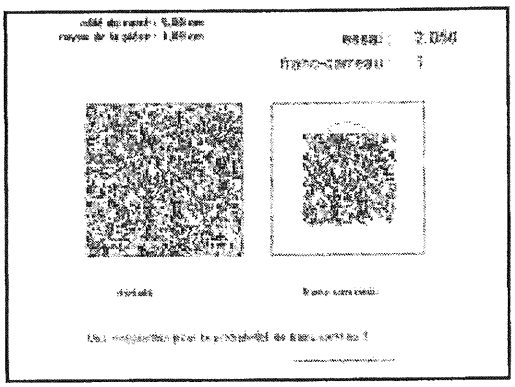


Écran 14

500 essais

2000 essais environ

Écran 15



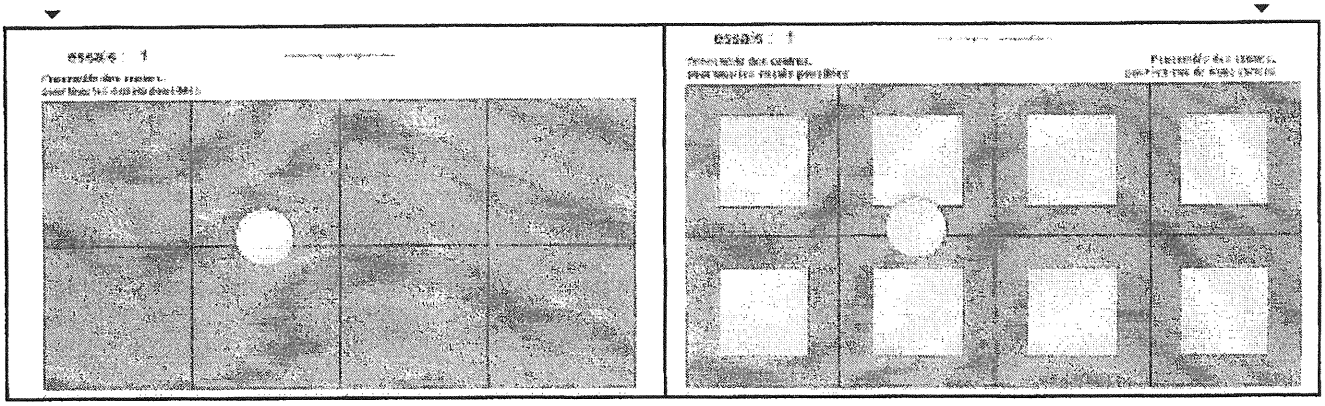
Retour à la figure **[francarreau 8c.fig]**

Curseur en **position médiane** : en bleu (ou gris), l'ensemble des centres, pour tous les essais possibles

Écran 16

Curseur en **position droite** : en jaune (ou gris clair), l'ensemble des centres, pour tous les cas de franc-carreau

Écran 17



OBSERVER L'ALEATOIRE

Alain BERNARD

Equipe Statistique
IREM de Montpellier

On enseigne une manière autant qu'une matière, et dans cette manière ne s'expriment pas seulement des individualités mais la communauté tout entière, dans ses élans, ses résistances et ses contradictions, dont se dessinent alors l'image et les projets.

Willy ROZENBAUM
Extrait du livre :
" La vie est une maladie sexuellement transmissible et constamment mortelle ".

Résumé : *Une petite réflexion historique sur l'opposition entre POINCARE tenant de l'observation et HILBERT tenant de la logique suivie de :*

1) Quelques expériences aléatoires utilisables en seconde : pile ou face, jeu de cartes, pour observer des fluctuations d'échantillonnage ; expérimentation de la loi de BENFORD.

2) Expériences aléatoires avec calculatrice et programmes. Exploitation des résultats avec un intervalle de confiance à 95%.

Un pas vers l'observation, le désordre, la vie.

1. Introduction

Un nouveau chapitre en seconde, et quel chapitre, la statistique ! Et les choses se compliquent encore un peu, car si la première partie du chapitre est assez classique avec de la statistique descriptive - diagrammes divers, médiane, moyenne, étendue - la seconde partie du chapitre traite de fluctuation d'échantillonnage, réclame la réalisation d'expériences aléatoires, d'observations, de modèles et de simulations.

1.1. Première question : la statistique fait-elle partie des mathématiques ?

Car peu d'entre nous ont appris ce genre de choses dans un cursus classique de mathématiques et les premiers sondages effectués en classe et auprès de collègues montrent qu'un nombre important de jeunes quittent le collège sans avoir fait des statistiques ou en en ayant fait très peu.

Premiers éléments de réponses.

- **Elément 1 :** les statisticiens à l'université font partie du département de mathématiques.
- **Elément 2 :** les statistiques usent (et abusent?) des nombres et des courbes. Il est ainsi difficile d'exclure leur travail du domaine des mathématiques.

Alors, où se situe le problème ?

Peut-être dans le fait qu'à partir de données numériques, les statisticiens commencent par un traitement de leurs données sous forme de tableaux, de diagrammes pour créer des images lisibles et simples de ces mêmes données. Et pourtant, tant que les programmes de

mathématiques ont traité de statistiques descriptives, les professeurs ont accepté de les enseigner et considéré qu'ils faisaient des mathématiques.

Alors, le problème se situe bien dans les expériences aléatoires et les observations. En effet, pas de démonstrations ni de théorèmes à espérer dans ce chapitre en seconde. Consciemment ou inconsciemment, les enseignants de maths sont formés dans un moule unique défini par Hilbert puis l'équipe Bourbaki et les mathématiques sont pour nous caractérisées par **l'implication logique permettant de démontrer et de donner des résultats généraux et certains.** (Le modèle (tiens !) remonte en fait à Euclide sans modifications notoires)

Modèle euclidien : math = Axiomes, définitions et théorèmes.

Sans remonter aux sources grecques déclinées depuis 23 siècles, peut-être sommes-nous en train de rejouer, très modestement, la querelle qui opposa au début du siècle Poincaré et Hilbert. Tous deux étaient considérés comme les maîtres de leur art et cependant, leurs vues sur les mathématiques étaient très différentes, quasiment contradictoires au niveau de leurs approches, et en fait complémentaires dans leurs démarches scientifiques. Une petite explication de cette querelle peut nous éclairer sur les statistiques en seconde.

POINCARE était un intuitionniste et HILBERT était un logicien.

1.2. Le point de vue de POINCARE

Il écrivait « ***L'expérience est la source unique de la vérité ; elle seule peut nous apprendre quelque chose de nouveau*** ». Bien entendu, comme nous le verrons en statistique, les expériences et les observations ne suffisent pas. POINCARE poursuit « *Une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison* ». Clairement, selon POINCARE, nous devons commencer par une expérience, puis observer, puis construire une science.

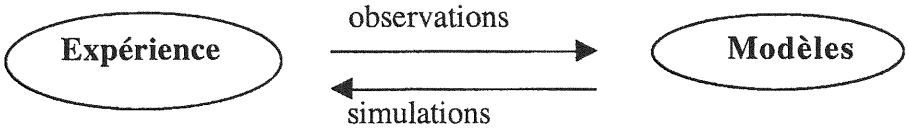
Que ce soit la baignoire d'ARCHIMEDE, la pomme de NEWTON ou la gare qui s'arrête devant le train dans lequel se trouve EINSTEIN, l'observation est première et ... insuffisante. ARCHIMEDE n'était pas le premier à prendre un bain, NEWTON n'était pas le premier à voir tomber une pomme, EINSTEIN... Mais ils ont eu une idée et sont parvenus à construire une science, un modèle de ce qu'ils voyaient.

1.3. Le point de vue d'HILBERT

Il ne semble s'intéresser « qu'au point de vue logique ». Quand il écrit « Les fondements de la géométrie », il prend trois objets premiers -point, droite, plan- qu'il dit pouvoir remplacer par une table, une chaise et une chope de bière (ce qu'il ne fait pas !!), puis il déduit une suite logique de propositions par usage de l'implication logique. POINCARE lui reproche alors : « *Le fondement de cette première proposition, il ne s'en occupe pas* ». Les axiomes sont énoncés « *mais on ne sait pas d'où ils sortent* ». A la suite d'HILBERT, les bourbakistes de DIEUDONNE reconstruiront les mathématiques utiles (!) sans se préoccuper des probabilités et encore moins des statistiques. Ce qui peut faire dire aujourd'hui que les statistiques ne font pas partie des mathématiques de BOURBAKI. Et pourtant le chapitre de statistique de seconde est sûrement le seul chapitre du programme de mathématiques dont tous les élèves se serviront un jour dans leur vie comme dans leur travail.

1.4. La méthode statistique (au lycée)

Le programme demande une extension de la méthode habituelle dans laquelle l'enseignant occulte une part importante de la démarche scientifique, la part d'expérimentations, d'essais, d'erreurs et d'observations qui amènent au modèle. Nous avons en seconde une occasion, peut-être unique, de construire une science à partir d'observations.



En espérant que ce court exposé liminaire a satisfait votre côté logique, revenons à la démarche scientifique qui commence par l'expérience, ici l'expérience aléatoire.

2. Les expériences aléatoires

« *Toute pensée émet un coup de dés* » (Stéphane MALLARME).

Extrait des commentaires des programmes :

> **L'enseignant traitera des données en nombres suffisant pour que cela justifie une étude statistique ; il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts.**
>
> **La notion de fluctuation d'échantillonnage et de simulation ne doit pas faire l'objet d'un cours. L'élève pourra se faire un « cahier de statistique » où il consignera une grande partie des traitements de données et des expériences de simulation qu'il fait, des raisons qui conduisent à faire des simulations ou traiter des données, l'observation et la synthèse de ses propres expériences et de celles de la classe.**

Tout est dit ou presque.

Conformément au programme, nous devons :

- expérimenter avec des objets matériels,
- expérimenter avec la touche " Random " de la calculatrice.

Nous y adjoindrons :

- Expérimenter avec des programmes.

A) Expériences aléatoires directes

Expérience aléatoire n° 1 : pile-face, ou 1-2, ou fille-garçon, ou ... toute expérience binaire que l'on répète, c'est-à-dire, schéma de BERNOULLI.

On découpe 10 cartons de 3cm sur 3cm sur lesquels on écrit p-f ou 1-2 ou F-G.

(Autre version peut-être plus motivante pour les élèves : chaque élève met dans un chapeau un carton avec son nom et l'on va essayer de retrouver la proportion de garçons et de filles dans la classe par des tirages d'échantillons aléatoires de taille 10, par exemple.)

Intuition : 1°) Chaque carton a 1 chance sur 2 de tomber sur la face 1.

2°) Sur 10 cartons, on va obtenir (environ ?) 5 fois " 1 " et 5 fois " 2 ".

On effectue 20 lancers des 10 cartons.

On obtient : 8-2, 8-2, 5-5,4-6, 3-7, 4-6, 5-5, 4-6, 6-4, 7-3, 5-5, 5-5, 5-5, 7-3, 4-6, 6-4, 5-5, 4-6, 3-7.

Traitement mathématique ou statistique :

- Tracer un diagramme représentatif.
- Calculer la différence pour chaque lancer
- Quelle est la différence moyenne ?
- Combien de faces " 1 " obtenues au total des 200 lancers ?
- Pourcentage ou fréquence de la face " 1 ".

2ème série de lancers avec 20 cartons :

12-8, 7-13, 15-5, 12-8, 7-13, 10-10, 7-13, 8-12, 10-10, 8-12.

3ème série :

11-9, 8-12, 10-10, 9-11, 8-12, 9-11, 12-8, 8-12, 9-11, 11-9.

Expérience aléatoire n° 2 : Prendre un jeu de 52 cartes.

On connaît la proportion de cartes du jeu de 32 cartes. Elle est de 32 / 52 bien sûr. La retrouve-t-on par des tirages aléatoires ?

On bat les cartes et l'on tire 10 cartes au hasard (ou les 10 premières cartes ?). Et l'on recommence l'expérience 20 fois : battre puis tirer.

1ère personne :

6-4	6-4	5-5	6-4	7-3
4-6	3-7	5-5	6-4	5-5
6-4	5-5	5-5	6-4	3-7
6-4	5-5	5-5	5-5	5-5

Note : 6-4 signifie : le tirage aléatoire a donné 6 cartes du jeu de 32 et 4 autres cartes.

2ème personne :

6-4	6-4	7-3	6-4	6-4
8-2	6-4	5-5	7-3	6-4
4-6	8-2	7-3	3-7	7-3
4-6	7-3	4-6	6-4	6-4

Noter la fluctuation d'échantillonnage .

Cette expérience permet-elle de retrouver la proportion connue ?

Le résultat est-il éloigné ?

Se rapproche-t-on de 32 / 52 quand on augmente le nombre d'échantillons ?

Vocabulaire :

Expérience = essai, épreuve = connaissance acquise par la pratique, par l'observation.

Fluctuation = flotter = mouvement alternatif comparable à l'agitation des flots (d'après le dictionnaire).

Echantillon = morceau d'étoffe = petite **quantité** d'un produit destinée à en faire connaître la **qualité**.

Fraction représentative d'une population ou d'un ensemble statistique

(Pas ici ou l'échantillon est aléatoire)

Aléa = Evénement imprévisible.

Remarque : ces deux expériences ont l'avantage d'être silencieuses par rapport aux lancers de dés ou de pièces.

Expérience aléatoire n° 3 : L'expérience est peu intéressante quand on connaît le résultat.

On retire maintenant 10 cartes du jeu sans les regarder et l'on recommence l'expérience. Trouver ou estimer la proportion de cartes du jeu de 32 cartes se trouvant dans les 42 cartes restantes.

Il s'agit d'une version intermédiaire de la « punaise ». Quand on a un type de punaises, on ne saura jamais quelle est la probabilité qu'elle retombe sur la tête quand on la lance tandis qu'avec les cartes, on peut vérifier dans le jeu si on a trouvé un résultat proche ou même exact.

Expérience aléatoire n° 4 : Le problème est de trouver une méthode de génération de chiffres aléatoires.

Les calculatrices fournissent des chiffres pseudo-aléatoires parfaitement déterminés.

(Remarque : dès le début du 21ème siècle, les calculatrices devraient donner des nombres vraiment aléatoires fournis par le " bruit thermique " des bécanes.)

Comment résoudre ce problème ?

Idée : pour générer des chiffres aléatoires, on pourrait prendre un journal d'économie et prendre le premier chiffre non-nul à gauche de tous les nombres.

Allons-y !

Bourse du 14 Avril 2000. Indice SBF 120

On prend le premier chiffre à gauche des 120 valeurs qui s'y trouvent.

Chiffre	1	2	3	4	5	6	7	8	9
fréquence	33	20	15	18	5	11	5	8	5

Y a un problème !

Donc on recommence avec les variations en % depuis le 31/12 jusqu'au 14/4

chiffre	1	2	3	4	5	6	7	8	9
fréquence	50	18	15	10	9	5	4	8	3

Y a toujours un problème ! Mais....

Pour un observateur avisé, il y a aussi une ressemblance entre les deux résultats obtenus. Forte occurrence du chiffre 1 puis décroissance assez régulière dans les deux cas.

Remarque : le mot « fréquence » sera utilisé pour les effectifs d'une classe comme pour le pourcentage de cette classe. La tradition veut qu'on utilise plutôt le mot effectif pour dénombrer les éléments d'une classe.

La loi de BENFORD

BENFORD avait remarqué qu'une table de logarithme ne donne pas des chiffres aléatoires quand on prend le premier chiffre non nul à gauche des logarithmes qui s'y trouvent. Cette observation faite, il a cherché s'il n'y avait pas une loi.

Problème : trouver une méthode pour générer des nombres aléatoires.

Solution : prendre un journal, ici Le Monde du 23/4/99, et en extraire 1000 nombres dont on conserve le 1er chiffre non nul à gauche.

chiffre	1	2	3	4	5	6	7	8	9
fréquence	0,322	0,151	0,108	0,099	0,073	0,081	0,055	0,065	0,046

PERDU. L'observation est têtue.

On s'attendait à trouver la même fréquence de 1/9ème pour chaque chiffre, or on trouve une autre répartition.

De plus, l'observation montre que les trois échantillons donnent des résultats voisins !

Conclusion 1 : la méthode utilisée ne donne pas des chiffres au hasard.

Conclusion 2 : L'expérience est bien nécessaire pour valider ou invalider un modèle ou une idée.

Conclusion 3 : Comme nous sommes déterministes par formation jusqu'à la déformation, nous allons chercher une « loi » bien que les expériences ne donnent pas les mêmes résultats.

Et BENFORD trouva à partir de l'idée suivante : si une loi existe, la répartition doit être inchangée si on multiplie les nombres par une constante (penser au taux de change entre deux monnaies).

Les calculs donnent effectivement une répartition théorique proche de l'expérience laquelle valide le modèle de BENFORD.

Loi de BENFORD

En base 10, la probabilité que le premier chiffre à gauche soit i compris entre 1 et 9 est :

$$p_i = \log(1+ 1/i)$$

chiffre	1	2	3	4	5	6	7	8	9
fréquence	30,1%	17,6	12,5	9,7	8,0	6,7	5,8	5,1	4,6

Pour tester une comptabilité, c'est simple : prendre 100 ou 200 nombres et regarder la répartition des premiers chiffres à gauche. Il doit y avoir environ 30 % de 1 et aussi environ 50 % de 1 ou 2, sinon il y a des chances que la comptabilité soit trafiquée.

Une remarque sur le terme de loi.

Aux difficultés de l'expérience près, un physicien doit retrouver toujours le même résultat quand il recommence une expérience.

Par contre, une loi de probabilité est un modèle éloigné de la réalité et quand on répète une expérience en statistique, on n'espère pas retrouver le même résultat. C'est la fluctuation d'échantillonnage que l'on constate sur tous les exemples donnés.

Une première conclusion

Ai-je montré **qu'expérimenter puis observer** sont les actes premiers, au moins au sens de POINCARE, de l'activité mathématique ?

Certes, nous n'avons pas encore de science, mais expérimenter permet de réfléchir et d'obtenir certains résultats.

Dès lors, **observer l'aléatoire est l'acte premier des probabilités**. Bien sûr, observer l'aléatoire est insuffisant et il faut dans un deuxième temps **construire le ou les modèles** susceptibles de correspondre au problème étudié

Dans un troisième temps, nous aurons à **simuler les modèles** pour tester ou vérifier leur bien-fondé.

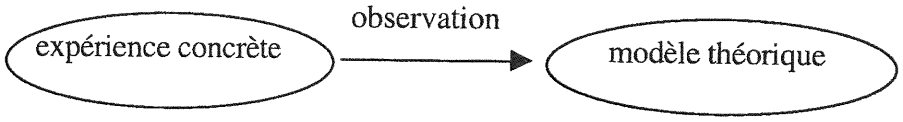
B) Observer l'aléatoire avec une calculatrice programmée

Pour la première fois, les programmes de mathématiques font une place non négligeable à l'observation. Jusqu'à présent, les maths paraissaient ou étaient présentées comme une activité basée sur la logique seule et exclusive. « Faire » des maths, pour la grande majorité des gens, consiste exclusivement à démontrer en utilisant des implications logiques. Dans le schéma de présentation des nouveaux programmes paru au B.O.E.N, l'observation occupe une place première. « Faire » des maths est, en premier lieu, observer. Alors, peut-on réhabiliter l'observation après tant d'années de primauté, voire d'exclusivité, donnée à la logique ? Là est une des questions posées par les nouveaux programmes de seconde.

Pour observer, il faut expérimenter et l'expérience change du même coup de statut. Jusqu'à présent l'élève faisait un grand nombre d'exercices et acquérait ainsi de ... l'expérience, laquelle lui permettait de résoudre un exercice similaire proposé en devoir pour évaluer sa capacité à reproduire un raisonnement.

Dans l'approche prônée dans le chapitre de statistique, l'observation doit être à la source de l'activité intellectuelle. Ce qu'elle est effectivement dans la démarche scientifique. Un ordinateur n'observe pas. Il n'est pas intelligent. Par contre, un être humain observe et c'est seulement après avoir observé qu'il déduit une ou plusieurs possibilités de solutions excluant ainsi d'autres possibilités ne semblant pas mener à une solution.

L'observation nourrit l'intelligence en vue de construire un ou plusieurs modèles. Dès la seconde, le chapitre de statistiques demande de faire fonctionner cette méthode.



Alors observons.

Expérience aléatoire n° 5 : Le jeu de pile ou face

Reprenons l'expérience la plus simple, le jeu de pile ou face. L'expérience n'a que deux issues, et l'on répète n fois l'expérience.

Méthode 1 : expérience " à la main " ou avec le cerveau, une feuille, et un crayon.

Ecrire une suite de 100 lettres p ou f , essayant par là même de " simuler " 100 lancers d'une pièce équilibrée.

pppffpfppffpfppfppfpfpfpfpfpppfpfpfpppfffffpppfppfpfpfpppppfpfpfpfpfpfpppf...etc

Une image peut-être donnée par deux lignes.

F f f f f f f f

P p p p p p p p p p p p p .....

On peut observer :

1) La fréquence des " p " sous forme de nombre de piles ou sous forme de pourcentage.

2) Le nombre de " changements " p ⟶ f ou f ⟶ p.

3) La longueur moyenne des séquences de 0 et des séquences de 1.

4) Le rang de la première apparition de 6 piles consécutifs.

5) Autre ?

Apprendre à observer n'est pas facile et choisir des critères mathématiques non plus.

Méthode 2 : la même expérience réalisée ou " simulée " avec une calculatrice.

Court-circuitons le cerveau le temps d'appuyer dix fois sur la touche Random de la calculatrice pour obtenir dix nombres aléatoires (ou presque).

Le codage : les chiffres pairs sont des côtés face et les chiffres impairs sont les côtés pile.

Cent expériences de pile ou face :

pppfpfppfpppffpfpfffpffpfffppffpfffppppp....etc

La même image :

F f f f f f f f f f f f f f f f f p

P p p p p p p p p p p p p p p f f etc.....

Effectuer le même traitement mathématique et comparer avec la série donnée par le cerveau. Qu'observe-t-on?

Quelques données (Claudine ROBERT, Toulouse, Mars 2000)

Probabilité d'avoir au moins six données consécutives égales dans une série de N pile-face.

N = 100 ==> proba = 0.80

N = 200 ==> proba = 0.96

Le rang qui clôt la première apparition de six données consécutives identiques vaut en moyenne 62.

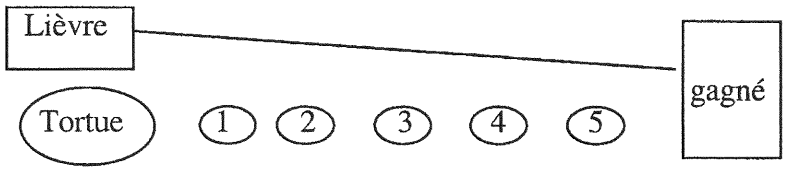
La probabilité d'avoir au moins cinq données consécutives identiques dans une série de N pile-face: N = 50 ==> proba = 0.82

N = 100 ==> proba = 0.97

N = 200 ==> proba = 0.999

Expérience aléatoire n° 6 : Le lièvre et la tortue

Un jeu simple et facile à modifier avec un modèle un peu compliqué mais abordable.



On lance un dé ou l'on simule avec la touche Random.

Déroulement d'une partie : si le 6 sort, le lièvre a gagné.

Sinon, la tortue avance d'une case.

On continue jusqu'à ce qu'il y ait un gagnant.

Quelle est la situation la plus enviable ?

Simulation à la calculatrice

On sort des nombres aléatoires avec la touche Random dont on ne conserve que les chiffres de 1 à 6. (Est-ce légitime? Pourquoi?)

5526 56 256 6 316 6 356 1226 214352 133316 16

245142 36 14556 546 246 434212 6 4256 6 142516 56

12446 36 424523

Le lièvre L a gagné 21 fois sur 25, soit dans 84 % des jeux

Une deuxième simulation donne le lièvre gagnant dans 18 cas sur 25 soit 72 % des jeux.

Une troisième simulation donne le lièvre ganant dans 35 jeux sur 50 soit 70 % des jeux.

Question : l'esprit du programme veut qu'on utilise des résultats dont on admet la construction ou la démonstration.

On peut construire un modèle : on note x les numéros 1 à 5 où la tortue avance d'une case.

T gagne <===> x et x et x et x et x et x

Les chances que la tortue T gagne se calculent en remplaçant les " et " par une multiplication.

et " x " par la probabilité 5/6 d'obtenir une face de 1 à 5.

La tortue a une probabilité égale à $(5/6)^6 = 0{,}335$ soit 33,5%.

Les calculs donnent la victoire au lièvre L dans 66,5 % des jeux.

Le jeu n'est pas équitable. Comparer avec les résultats obtenus dans les expériences.

Exercice : Simuler un jeu dans lequel la tortue gagne après 5 coups seulement.

La première simulation faite plus haut donne : 6 victoires à la tortue.

La deuxième simulation donne : 9 victoires à la tortue.

Le jeu ne semble toujours pas équitable

Exercice : Evolution du jeu

Simuler un jeu où la tortue gagne après 4 coups, 3 coups.

Essayer quand le lièvre a une case intermédiaire nécessitant l'obtention de 2 N°6.

Etc.....

C) Expériences aléatoires avec une calculatrice programmée

(Hors programme !!!)

Comme évoqué plus haut, l'ordinateur ne réfléchit pas, mais il faut savoir l'utiliser pour les tâches répétitives. Il sait aussi faire certaines observations... qu'on lui a dit de faire.

Expérience aléatoire n° 7 :

Reprenons le jeu de pile ou face où l'on va mesurer la fréquence de « face » obtenue lors de nombreux lancers.

Le programme (fourni par Casio) : nom -PF-

```
0 → S ↙
" NB LANCERS ": ? →L ↙
FOR 1 → I TO L ↙
INT( 2Ran # ) → X ↙
If X = 0 ↙
Then 1+S → S ↙
If end ↙
Next ↙
" FREQ FACE " : S ÷ L
```

Ce programme donne donc la fréquence de faces obtenue lors d'un tirage aléatoire de L lancers d'une pièce équilibrée.

La boucle est classique de FOR à NEXT

L'instruction INT(2Ran #) → X ↙ calcule la partie entière de 2Ran#

Ran# est entre 0 et 1 donc 2Ran# est entre 0 et 2. L'entier est rangé dans X et X a 1 chance sur 2 de valoir 0. Le " If " incrémente donc S dans ce cas.

Une première expérimentation : 20 séries de 100 lancers

0,51	0,53	0,52	0,47
0,59	0,51	0,46	0,54
0,48	0,51	0,51	0,44
0,45	0,47	0,5	0,54
0,5	0,66	0,46	0,52

100 lancers donnent entre 44 et 66 piles, ce qui est bien autour du 50 attendu mais curieusement décentré. L'observation, toujours l'observation !

La moyenne:50,85 confirme le décentrage mais rien d'inquiétant. On " savait " bien qu'on n'allait pas trouver 50 et le résultat réel est proche de 50.

Deuxième expérimentation : 20 séries de 500 lancers.

0,516	0,520	0,504	0,506
0,514	0,530	0,514	0,484
0,520	0,520	0,488	0,536
0,580	0,482	0,524	0,528
0,528	0,498	0,500	0,484

500 lancers : les fréquences obtenues s'approchent de 0,5, soit entre 0,482 et 0,580.

Résultat toujours décentré au-dessus de 0,5.

Moyenne : 513,8

Cela devient bizarre.

Troisième expérimentation : 10 séries de 10 000 lancers.

Nous apprécions la rapidité de la calculatrice - quelques minutes suffisent - et le silence du dé.

0,5192	0,5037
0,5144	0,5163
0,5092	0,5161
0,5133	0,5136
0,5160	0,5073

Observations :

1) Dans chaque série, on a plus de 5000 faces!?

2) Le résultat est dans chaque série proche de 5100.

3) Calculons la moyenne sur les 100 000 lancers. On trouve 51291 faces.

4) On sait (intuition ?) qu'on a peu de chances de trouver 50 000 mais on devrait trouver certains résultats de séries inférieurs à 5000 et des résultats supérieurs à 5000. Ce n'est pas le cas.

Conclure

LE MODELE

Pour la conclusion, nous devons nous adresser à la partie « science » ou à la maison « science » en construction comme aurait dit POINCARE. Le thème sur les intervalles de confiance est là pour cela sans pour autant nous donner une certitude.

a) Pour un échantillon (une série) de N = 10 000 lancers, calculons l'intervalle de confiance à 95% :

La probabilité d'obtenir Face lors d'un lancer est F = 0,5

Le nombre de lancers est N = 10 000, donc $\sqrt{N} = \sqrt{10\,000} = 100$.

L'intervalle de confiance est donné par $[F - 1/\sqrt{N} , F + 1/\sqrt{N}]$

Dans notre cas : $[\ 0,5 - \frac{1}{100} , 0,5 + \frac{1}{100}\ ]$ soit [0,49 ; 0,51] .Le modèle dit : Il y a 95% de chances que l'échantillon donne un résultat dans l'intervalle [0,49 , 0,51]. Observons nos 10 échantillons. 7 séries sont en dehors de cet intervalle de confiance.

b) Regroupons nos 10 échantillons en 1 série de N' = 100 000 lancers.

$1/\sqrt{N'} = 0,003$ soit 0,3%

Il y a 95% de chances de trouver une fréquence comprise entre 0,497 et 0,503.

Cette fois, nous sommes loin du compte puisque la fréquence de faces est 0,5129

Au risque de 5%, nous pouvons dire que les nombres aléatoires fournis par cette calculatrice ne sont pas ... aléatoires. Nous le savions. Aucune machine ne fournit des nombres parfaitement aléatoires. Chaque génération de nombres est déterminée par un algorithme parfaitement déterminé mais, qui est sensé donner des nombres répartis entre 0 et 1. Nous n'en sommes pas loin mais...

Expérience aléatoire 8 :

Une autre observation sur le jeu de pile ou face se fait sur les séquences de 1 consécutifs. Casio donne le programme suivant :

Nom: -PPPPPP-

Int(2 Ran#) → A ↙

Int(2 Ran#) → B ↙

Int(2 Ran#) → C ↙

Int(2 Ran#) → D ↙

Int(2 Ran#) → E ↙

Int(2 Ran#) → F ↙

6 → I ↙

Lbl 1 ↙

10 000 A + 10 000 B + 1000 C +100 D + 10 E + F → X ↙

If X = 111 111 ↙

Then " NB LANCERS POUR PPPPPP " ↙

I ↙

Else B → A : C → B : D → C : E → D : F → E ↙

Int (2 Ran#) → F ↙

I + 1 → I ↙

Goto 1 ↙

If end

Expérimentation : le nombre de lancers pour obtenir 6 P consécutifs (ou 6 numéros 1) est :

127	43	385	14	287	296	218	261	58	31
83	13	136	290	57	24	186	103	62	564
153	117	170	6	239	102	65	160	264	42
229	153	198	130	217	132	79	14	115	104
116	600	33	318	84	47	250	90	106	130

Observations :

1) Très grande **fluctuation** des résultats de 6 à 600.

2) Nombre d'expériences plus limité. Nous avons seulement 50 expériences. Donc plus de prudence dans les " déductions ". Un programme plus complet pourrait tourner seul sur un plus grand nombre d'expériences mais là, il faut relancer la calculatrice après chaque expérience en notant le résultat sur une feuille.

3) Chaque colonne donne cependant des moyennes assez voisines

188 146 177 118 141

La moyenne sur 50 expériences est de 154.

4) Si on avait cherché la première séquence 111 111 ou 000 000, on aurait, selon le modèle, obtenu une moyenne $x = \frac{154}{2} = 77$

5) L'intervalle de confiance à 95% est maintenant : $\sqrt{50} \approx 7$ donc $1 / \sqrt{50} \approx 0{,}14$ ou 14%.

L'intervalle de confiance est grand quand N = 50 est petit.

En prenant la recherche de la première série de 6 résultats consécutifs identiques (des 0 ou des 1), l'intervalle de confiance à 95% est [66, 88].

Curieusement les données de C.ROBERT (mars 2000, Toulouse) sont : le rang qui clôt la première apparition de 6 données consécutives vaut en moyenne 62 !

Nous n'en sommes pas loin, mais nous sommes quand même dans le risque 5%.

Il faut refaire des expériences. Une seconde série de 50 expériences donnera 100 expériences et un intervalle de confiance de 10%.

Une série de 200 expériences- pas trop longue à faire - donnera un intervalle de confiance de 7%.

Voici les 50 expériences suivantes :

35	73	94	215	424	139	343	22	326	93
166	79	373	14	55	207	108	108	53	86
51	41	116	49	10	67	9	213	16	81
127	222	27	809	25	550	17	114	119	231
12	177	162	8	218	366	329	18	149	60

Qu'en déduisez-vous ? Quelle confiance avez-vous dans vos résultats ?

Quelles sont vos certitudes ?

Encore 50 expériences :

150	54	13	31	229
99	241	163	215	254
226	146	169	452	241
68	109	104	362	29
133	132	305	148	264
247	89	59	258	436
154	215	189	91	10
231	29	8	34	31
76	318	23	142	36
12	148	65	123	17

<u>Expérience aléatoire 9</u> : le dé à 5 faces

Le programme proposé par Casio est le suivant :

```
Nom:        - DE 5-
            " NB LANCERS " : ? → L ↙
            Seq ( 0, X, 1, 5, 1) → List2 ↙
            For 1 → I to L ↙
            Int ( 5 Ran# + 1 ) → X ↙
            1 + List2 [ X ] → List2 [ X ] ↙
```

Next ↙

List 2

Expérimentons :

50 lancers

	série 1	série 2	série 3	série 4	série 5
Face1	10	10	14	12	6
Face2	14	9	9	9	12
Face3	5	9	8	11	10
Face4	11	14	12	7	10
Face5	10	8	7	11	12

500 lancers

	série 1	série 2	série 3	série 4	série 5
Face1	98	108	119	95	113
Face2	91	108	96	118	119
Face3	120	98	88	101	89
Face4	96	98	106	93	91
Face5	95	88	91	93	88

5000 lancers : un peu plus long.

	série 1	série 2	série 3	série 4	série 5
Face1	1092	985			
Face2	1050	1023			
Face3	978	966			
Face4	895	1042			
Face5	985	980			

3. Esquisse d'une conclusion

Voici d'abord quelques lignes extraites du livre de Robert GERMINET, directeur de l'Ecole des Mines de Nantes : **l'apprentissage de l'incertain.**

Page 23 : « *les exigences de la formation d'ingénieurs*

Première exigence : acquisition des connaissances fondamentales.....

Deuxième exigence, inédite celle-là : les ingénieurs doivent cultiver leur faculté d'observation et acquérir le sens du réel, par le regard comme par l'écoute. *Aujourd'hui, cette capacité d'écoute, de regard, de compréhension fine du besoin est considérée comme la capacité d'adaptation à des situations évolutives, qualité indispensable dans un monde*

chaotique et mouvant. Les élèves n'y sont pas habitués. ***Tout le système d'enseignement, depuis les classes du primaire[1] jusqu'aux classes d'ingénieurs, donne trop peu de place à l'observation.***

Troisième exigence, tout à fait classique : « la capacité de conceptualisation de l'ingénieur ».

Page 28 : « *Chemin faisant, les mathématiques - qui, il faut aussi le noter, offrent un mode de correction bien pratique aux enseignants - sont devenues l'outil de sélection absolu des concours. Et cet absolutisme a pesé sur la culture, la mentalité des élèves et de leurs professeurs, car* ***les mathématiques*** *- sauf quand elles s'inscrivent dans le champ de la recherche -* ***ne laissent pas une place suffisante à l'observation, au désordre, à l'inattendu, en un mot à la vie****. Les élèves ont grandi à l'abri de ce cocon intellectuel, à l'ombre des théorèmes porteurs de vérités harmonieuses et bien ordonnées. Plutôt que d'être entraînés à se frotter à la réalité,* ***plutôt que d'observer avant d'imaginer un modèle et une expérimentation possible destinés à confirmer ou infirmer la théorie, tout s'est passé comme si les jeunes ingénieurs n'avaient été éduqués qu'à procéder par certitudes, par expressions abstraites sans souci d'une confrontation au réel*** ».

Le programme de seconde précise bien que la simulation et la modélisation font partie intégrante des programmes.

Ni les expérimentations, ni les modèles ne devraient être livrés prêts à l'emploi !

Voilà ce que demande le programme de seconde en statistique. Faire entrer dans l'enseignement l'observation, la modélisation et la simulation pour effectuer une démarche scientifique complète.



[1] Nuancer cette affirmation car l'observation a une grande place au primaire et même au collège. Les résistances croissent ensuite au lycée et dans le supérieur.

UNE APPROCHE DE LA MODELISATION AU TRAVERS DE LA LOI DE POISSON

Ali GANNOUN - Nicolas SABY

Equipe Liaison lycée-université
IREM de Montpellier

Résumé : *Nous nous proposons de montrer comment des problèmes concrets ont amené des modèles théoriques très utilisés. L'exemple que nous utilisons pour ce propos est la "Loi de POISSON".*
La "loi de POISSON" a une efficacité redoutable pour modéliser les phénomènes rares. Le but que l'on veut atteindre est ici de comprendre comment cette simplification abstraite qu'est la loi de POISSON permet d'avoir des renseignements sur la réalité que l'on étudie et quels sont ces phénomènes rares qui sont modélisés par un processus de POISSON.

1. Introduction

Les probabilités et statistiques sont très adaptées pour initier à des activités de modélisation puisque leur origine est intimement liée à ce souci de modéliser des phénomènes réels de hasard par des modèles mathématiques dans lesquels on puisse mesurer ce " hasard ". La démarche sous-entendue dans une activité de type probabilité ou statistique devrait être en trois temps : hypothèses, modélisation mathématique, confrontation des résultats du modèle avec l'expérience.

Le rôle des probabilités dans cette activité est la façon de traiter le modèle mathématique. Le rôle des statistiques concerne plutôt le troisième temps de confrontation des résultats avec l'expérience. Le premier temps, à savoir le choix des hypothèses et la modélisation du problème, ne trouve pas toujours sa place dans un enseignement de probabilités et statistiques et l'on considère souvent qu'il relève de la discipline qui utilisera ces modèles théoriques.

Il nous semble important de pouvoir montrer comment les problèmes (concrets) ont amené ces modèles théoriques. C'est ce que nous proposons de faire sur l'exemple de la " Loi de POISSON " ou des " Processus de POISSON ".

On trouve couramment que la " loi de POISSON " est l'une des plus importantes loi discrète et qu'elle a une efficacité redoutable pour modéliser les phénomènes rares. Ce type de justification mérite d'être éclairer si l'on veut réellement se placer dans une démarche de modélisation. Le but que l'on veut atteindre est ici de comprendre comment cette simplification abstraite qu'est la loi de POISSON permet d'avoir des renseignements sur la réalité que l'on étudie et quels sont ces phénomènes rares qui sont modélisés par un processus de POISSON.

Nous allons tenter de présenter dans ce texte comment apparaît la loi de POISSON comme limite de loi binomiale et comment si on interprète la loi de POISSON comme un processus ponctuel se déroulant sur la droite réelle, il modélise des phénomènes rares dans un sens que l'on précisera. On illustrera notre propos par des exemples variés. L'exemple initial qui sera utilisé comme motivation de notre texte sera celui du problème des voitures arrivant à un péage d'autoroute. La modélisation de ce problème sera l'occasion de montrer quelles sont les hypothèses faites qui justifient l'utilisation de la loi de POISSON et quelles en sont les limites.

2. La loi binomiale

Cette loi est utilisée dans tous les phénomènes discrets où l'on veut compter le nombre de succès parmi n épreuves identiques indépendantes comme le jeu de pile ou face. Cela signifie que l'on est en présence d'une épreuve qui a deux éventualités que l'usage dénomme " succès" et " échec ". L'éventualité " succès " a une probabilité p d'être réalisée, l'éventualité " échec " a la probabilité $q=1-p$ d'être réalisée.

On réalise alors n épreuves identiques. La probabilité d'obtenir ainsi k succès et $n-k$ échecs est $C_n^k p^k q^{n-k}$. La preuve est seulement un problème de dénombrement. Notons Ω l'ensemble des éventualités succès, échec. L'événement k succès et $n-k$ échecs est le sous-ensemble de Ω^n constitué des éléments $(a_1,...,a_n)$ où k d'entre eux sont des succès et $n-k$ des échecs. La probabilité de chacun de ces éléments est $p^k q^{n-k}$. Comme il y a C_n^k tels éléments, la probabilité d'obtenir k succès et $n-k$ échecs est $C_n^k p^k q^{n-k}$.

Dans certains cas, on s'intéresse à la réalisation d'un grand nombre d'expériences et le calcul effectif des $C_n^k p^k q^{n-k}$ devient très fastidieux à cause de la complexité des calculs des coefficients binomiaux. Il est alors utile de trouver des approximations de ces quantités des expressions plus calculables. C'est ce que réalise la " loi de POISSON " et que nous allons illustrer par notre problème des arrivées de voitures à un péage d'autoroute.

2.1. Problème

On s'intéresse donc aux arrivées de voitures à un péage d'autoroute et on aimerait modéliser cette situation afin de pouvoir considérer la probabilité d'un événement comme l'arrivée de 10 voitures dans un intervalle de 5 minutes.

2.2. Modélisation

Afin de pouvoir proposer un modèle de ce problème dans lequel la question prenne un sens mathématique, on est amené à faire des hypothèses sur le trafic routier.

Il est très probable dans notre exemple qu'il y a plus d'arrivées de voitures au péage entre 16:00 et 18:00 qu'entre 3:00 et 4:00 et cela change très certainement la probabilité cherchée.

Hypothèse 1 : On considère un intervalle de temps tel que le nombre moyen d'arrivées de voitures par minute est constant égal à λ. Dans un intervalle de 5 minutes, il arrive donc en moyenne 5λ véhicules.

Hypothèse 2 : Les nombres d'arrivées de véhicules dans deux intervalles disjoints sont indépendants : il n'y a pas d'influence de l'un sur l'autre. C'est une hypothèse de fluidité du trafic.

Voyons maintenant comment une troisième hypothèse s'impose pour modéliser le problème par une distribution binomiale.

On découpe notre intervalle de temps considéré (par exemple [16:00, 18:00]) en n sous-intervalles de même longueur. Si les intervalles sont suffisamment petits, on aimerait pouvoir

dire que l'arrivée de plus de 2 véhicules au péage est négligeable devant celles de 0 ou 1 véhicule. C'est l'hypothèse 3.

Hypothèse 3 : Si la longueur de l'intervalle est suffisamment petite, la probabilité qu'il arrive plus de 2 véhicules est négligeable devant celles de l'arrivée de 0 ou 1 véhicule. C'est une nouvelle hypothèse sur le comportement du trafic. Elle signifie que l'on suppose qu'il n'y a pas en quantités importantes de paquets de véhicules.

Le choix de notre modèle est alors le suivant : On a une suite de n intervalles de même longueur et dans chaque intervalle, on fait une expérience de BERNOULLI où le succès est l'arrivée d'un véhicule au péage (on a ici interprété l'hypothèse 3 que la probabilité d'arrivée de plus de 2 véhicule est nulle).

La probabilité p du succès dans un intervalle est calculée grâce à l'hypothèse 1 de taux moyen λ. En moyenne, dans l'intervalle considéré de longueur t que l'on a découpé en n morceaux, il y a λt arrivées de véhicules et l'on peut espérer np arrivées puisque la probabilité d'arrivée dans un intervalle est p et qu'il y a n intervalles. Ces quatre quantités sont donc reliées par l'équation :

$$\lambda t = np, p = \frac{\lambda t}{n}.$$

On en déduit alors que la probabilité d'arrivée de k véhicules dans notre intervalle de longueur t que l'on note $P(N_{to,to+t} = k)$ vaut $C_n^k p^k q^{n-k}$.

Dans la pratique, si notre intervalle a été découpé en beaucoup de morceaux, le calcul des C_n^k devient fastidieux et l'on cherche à approcher $C_n^k p^k q^{n-k}$ par une autre expression plus calculable.

Par exemple pour k=0, on trouve :

(1) $$P(N_{to,to+t} = 0) = C_n^0 p^0 (1-p)^n$$

(2) $$= (1 - \frac{\lambda t}{n})^n$$

qui est approché par $e^{-\lambda t}$. Une formule de TAYLOR estime l'erreur si n est grand.

De même :

(3) $$P(N_{to,to+t} = 1) = C_n^1 p^1 (1-p)^{n-1}$$

(4) $$= np(1-p)^{n-1}$$

(5) $$= \lambda t(1 - \frac{\lambda t}{n})^{n-1}$$

qui est approchée par $\lambda t e^{-\lambda t}$.

On peut montrer par récurrence que $P(N_{to,to+t} = k)$ est approchée $\frac{(\lambda t)^k}{k!} e^{-\lambda t}$.

La distribution $\frac{(\lambda t)^k}{k!} e^{-\lambda t}$ est appelée distribution de POISSON de paramètre λt. La discussion précédente montre qu'elle est limite d'une distribution binomiale sous la condition suivante : np a une limite lorsque n tend vers l'infini. Cela signifie que la probabilité p est petite, c'est-à-dire que l'événement est rare. C'est ici que s'explique pourquoi la loi de POISSON modélise des phénomènes rares.

La formule de TAYLOR assure que, au plus n est grand, au plus l'approximation de BERNOULLI par POISSON est bonne.

2.3. Fiabilité du modèle

Il y a en fait deux types d'approximations dans ce problème. Le premier est dans la modélisation du problème par une distribution binomiale où l'on a vu que l'on faisait trois hypothèses dont une où l'on négligeait complètement une éventualité. Le deuxième type est lui interne au modèle puisqu'on approche la loi binomiale par une nouvelle expression baptisée loi de POISSON où le contrôle que l'on peut faire est interne aux mathématiques.

3. Exemples

La loi de POISSON peut être utilisée dans deux types de situations. La première est comme approximation effective de la loi binomiale car n est grand. La deuxième car on sait que n est grand ou ne peut pas être interprété ou n'a plus vraiment de sens (comme dans le problème de péage de voiture) parce qu'on considère que l'on est effectivement dans une situation de limite de loi binomiale.

1. Un dactylographe fait en moyenne une faute tous les 1000 mots. On suppose qu'il écrit un livre de 100 mots par pages. On note N_{100} le nombre de fautes pour une page. On s'intéresse à la probabilité de $N_{100} = j$.
2. On fait une pâte à cookies pour faire 500 cookies et on met dans cette pâte 600 raisins. Après la confection des cookies, on s'intéresse à la probabilité qu'un cookie contienne k grains de raisins.
3. Instant de désintégration dans un corps radioactif.
4. Appels à un standard téléphonique.

4. Bibliographie

DELMAS J.-P., 1993, *Introduction aux probabilités*, Ellipses.

FELLER W. , An Introduction to Probability Theory and Its Application, John Wiley Sons, Inc.

GRINSTEAD Ch., SNELL J. L., Introduction to Probability Theory.

http://www.dartmouth.edu/~chance

http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

LES NOMBRES PSEUDO-ALEATOIRES

Michel JANVIER

Equipe Informatique
IREM de Montpellier

Résumé : *Les calculettes contiennent des générateurs de nombres pseudo-aléatoires, que l'on fait intervenir par la touche Random. Les tableurs, comme Excel, permettent de générer des nombres aléatoires par la fonction Alea. Tous les logiciels de calcul formel ont aussi cette fonctionnalité. Ces nombres pseudo-aléatoires sont indispensables pour la simulation d'expériences aléatoires. Nous nous proposons d'expliquer une méthode utilisée pour engendrer ces nombres. Nous constaterons à ce propos qu'on emploie de beaux résultats d'arithmétique connus à l'avant dernier siècle. Nous contrôlerons la fiabilité des résultats obtenus et nous utiliserons ces nombres pour calculer une valeur approchée de π.Néanmoins on montrera, en utilisant l'inégalité de Bienaymé-Tchebichev, que pour atteindre une précision satisfaisante, il faut un grand nombre de répétitions de la même expérience aléatoire. Les nombres pseudo-aléatoires permettront la simulation d'expériences aléatoires mais pour obtenir la distribution empirique de probabilité d'une variable aléatoire il faudra de très nombreuses simulations, ce qui exige l'utilisation d'ordinateurs suffisamment rapides.*

1 Production de nombres pseudo-aléatoires

Tous les logiciels de calcul, comme les tableurs ou les logiciels de calculs formels, possèdent des générateurs de nombres aléatoires. Par exemple, dans SWP, Maple nous fournit dix nombres pseudo-aléatoires, répartis suivant une loi uniforme dans l'intervalle [0,1] :

0.95105, 0.14649, 0.42939, 0.52543, 0.02726, 0.21976, 0.067598, 0.84547, 0.67647, 0.45375, 0.64403, 0.92062

(Par Maple, Statistics, Random Numbers, puis dans la boite de dialogue qui s'ouvre choisir le nombre de nombres désirés, la distribution de probabilité, ici "uniform", les bornes de l'intervalle dans lequel doivent se trouver les nombres). Pour obtenir le même résultat avec Mathematica on utiliserait Table[Random[Real,{0,1}],{10}]. La principale méthode employée pour engendrer une suite de nombres aléatoires consiste à considérer un grand nombre premier p puis à construire une suite $(x_n)_n$ d'éléments de Z/pZ définie par itérations d'une fonction $f : Z/pZ \to Z/pZ$ de le forme $f(x) = (ax + b)_p$, en notant $(x)_p$

la classe d'équivalence de $x \in Z$ dans Z/pZ, la suite $(x_n)_n$ est définie par :

$$\begin{cases} x_0 &= (s)_p \\ x_n &= f(x_n - 1) \end{cases}$$

s s'appelle le germe de la suite $(x_n)_n$. Si y_n désigne l'unique entier de l'intervalle [0, $p-1$], représentant la classe $(x_n)_p$, dans la congruence modulo p, le réel $z_n = \frac{y_n}{p} \in [0, 1]$.

Dans tous les exemples qui vont suivre la programmation des calculs sera donnée en Mathematica.

Exemple 1 : *Commençons par un exemple peu réaliste, en prenant* $p = 7$, $s = 0, a = 3, b = 2$.

```
a=3 ; b= 2;(* introduction des valeurs de a et b *)
f[x_]:= a x+ b (* définition de la fonction f *) ;
g[x_, p_Integer] := Mod[f[x],p] (* indique modulo quel entier p il faut calculer *) ;
NestList[g[#,7] &,0,6] (* On fixe l'entier p à la valeur 7, le germe à 0,
et on itère 6 fois *)
{0,2,1,5,3,4,0}
```

On obtient six valeurs différentes, puis on retrouve le germe $s = 0$.

Comme Z/pZ est fini, la suite $(x_n)_n$ est nécessairement périodique. Essayons le germe $s = 6$.

```
NestList[g[#,7]&,6,6]
{6,6,6,6,6,6,6}
```

On a obtenu une suite stationnaire. $(6)_7$ est un point fixe de f. Il nous faut donc éviter de prendre comme germe un point fixe. Il est facile de trouver le point fixe de f, puisque Z/pZ est un corps.

```
Solve[a x + b ==x && Modulus == 7,x]
{{Modulus ->7,x -> 6}}
```

Remarque 1 : *En Mathematica, les équations se présentent à l'aide de la fonction booléenne Equal, dont la forme infixe est ==. On ajoute ici l'équation Modulus == 7, pour contraindre Mathematica à calculer dans* $Z/7Z$. *&& est la forme infixe de la fonction booléenne And. Elle permet ici d'introduire un système d'équations, que l'on aurait aussi bien pu présenter comme une liste d'équations. La fonction Solve résout le système d'équations proposées, sous la forme de règles de substitutions : Ainsi x -> 6, qui indique qu'il faut substituer 6 à* x *dans l'expression Equal[ax+b,x] pour obtenir la valeur "True". 6 est bien une solution de l'équation que l'on voulait résoudre.*

Exemple 2 : *Prenons un exemple plus réaliste avec* $p = 97, a = 37, b = 57$ *et* $s = 77$

```
a=37; b= 57;
list1 = NestList[g[#,97]&,77,96]
{77,93,6,85,1,94,43,96,20,21,58,69,88,15,30,3,71,65,37,68,51,4,
```

11,76,56,92,66,74,79,70,28,26,49,27,86,38,8,62,23,35,91,29,63,
60,46,13,53,78,33,17,7,25,12,16,67,14,90,89,52,41,22,95,80,10,
39,45,73,42,59,9,2,34,54,18,44,36,31,40,82,84,61,83,24,72,5,48,
87,75,19,81,47,50,64,0,57,32,77}

Y-a-t-il des répétitions dans cette liste?

Length[Union[Drop[list1,1]]]

96

Drop[List1,1] retire à la liste précédente son premier élément 77. Union classe ensuite les éléments de la liste par ordre croissant en supprimant les répétitions. Puis Length donne le nombre d'éléments de la suite ainsi obtenue : il y en a 96. On a obtenu 96 éléments distincts de $Z/97Z$. Les voici rangés par ordre croissant :

Union[Drop[list1,1]]

{0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,
27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,
50,51,52,53,54,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,
74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96}

55 n'apparaît pas dans cette liste. Vérifions que 55 est le point fixe de f.

Solve[a x + b ==x && Modulus == 97,x]

{{Modulus -> 97,x -> 55}}

Drop[List1,1] fournit donc 96 éléments distincts de $Z/97Z$, dans un ordre qu'il est difficile de prévoir quand on ne connaît pas le secret de sa fabrication. On en tire une suite de 96 nombres réels compris entre 0 et 1.

list2 = Drop[list1,1]/97 //N

{0.958763,0.0618557,0.876289,0.0103093,0.969072,0.443299,0.989691,
0.206186,0.216495,0.597938,0.71134,0.907216,0.154639,0.309278,
0.0309278,0.731959,0.670103,0.381443,0.701031,0.525773,0.0412371,
0.113402,0.783505,0.57732,0.948454,0.680412,0.762887,0.814433,
0.721649,0.28866,0.268041,0.505155,0.278351,0.886598,0.391753,
0.0824742,0.639175,0.237113,0.360825,0.938144,0.298969,0.649485,
0.618557,0.474227,0.134021,0.546392,0.804124,0.340206,0.175258,
0.0721649,0.257732,0.123711,0.164948,0.690722,0.14433,0.927835,
0.917526,0.536082,0.42268,0.226804,0.979381,0.824742,0.103093,
0.402062,0.463918,0.752577,0.43299,0.608247,0.0927835,0.0206186,
0.350515,0.556701,0.185567,0.453608,0.371134,0.319588,0.412371,
0.845361,0.865979,0.628866,0.85567,0.247423,0.742268,0.0515464,
0.494845,0.896907,0.773196,0.195876,0.835052,0.484536,0.515464,
0.659794,0,0.587629,0.329897,0.793814}

Remarque 2 : *Pour diviser chacun des termes de Drop[list1,1] par 97, il suffit d'écrire Drop[list1,1]/97. On dit que la division est une opération "listable". N[a] fournit une valeur approchée de a. On utilise ici la fonction N sous sa forme postfixe //N. N est également "listable".*

Représentons les points (n, x_n) pour $1 \leq n \leq 96$.

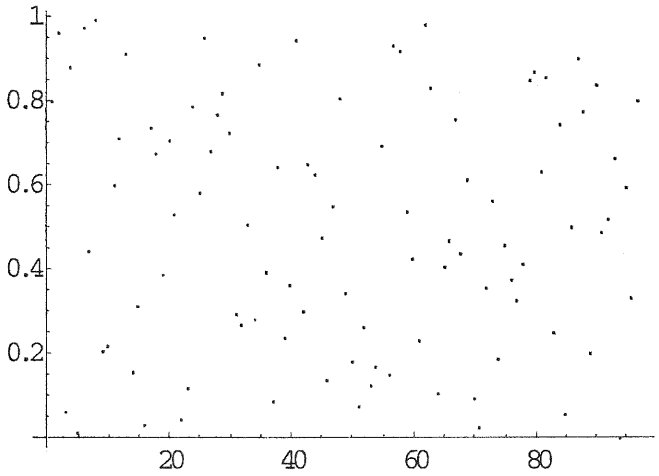


Figure 1 : Représentation de 96 points répartis aléatoirement

Suivant le choix de a on obtient pas nécessairement $p-1$ éléments distincts de Z/pZ. Prenons, par exemple $a = 64, b = 57$.

```
a=64; b= 57;
list2 = NestList[g[#,97]&,77,96]
Union[Drop[list2,1]]
Length[Union[Drop[list2,1]] ]
{77,38,64,79,69,11,82,67,77,38,64,79,69,11,82,67,77,38,64,79,69,11,
82,67,77,38,64,79,69,11,82,67,77,38,64,79,69,11,82,67,77,38,64,79,69,
11,82,67,77,38,64,79,69,11,82,67,77,38,64,79,69,11,82,67,77,38,64,79,
69,11,82,67,77,38,64,79,69,11,82,67,77,38,64,79,69,11,82,67,77,38,64,
79,69,11,82,67,77}
{11,38,64,67,69,77,79,82}
8
```

On a obtenu que huit éléments distincts. On est donc amené à trouver une caractérisation des éléments $(a)_p$ de Z/pZ tels que la suite $(x_n)_n$ fournisse $p-1$ éléments distincts de Z/pZ.

Proposition 1 : *Soit y le point fixe de f. Pour tout $n \in N^*$, $f^n(x) = (a^n(x-y) + x_0)_p$*
La relation est vraie pour $n = 1$. Par définition

$$y = (ay + b)_p$$

Par ailleurs

$$f(x) = (ax + b)_p$$

En soustrayant $f(x) - y = (a(x - y))_p$, et par suite

$$f(x) = (a(x - y) + y)_p$$

Supposons, la relation établie à l'ordre n, alors

$$\begin{aligned} f^{n+1}(x) &= f(f^n(x)) \\ &= (a(a^n(x - y)) + y)_p \\ &= (a^{n+1}(x - y) + y)_p \end{aligned}$$

La relation est encore vraie à l'ordre $n + 1$.

On a établi par récurrence la relation annoncée pour tout $n \in N^*$.

Corollaire 2 : *Les périodes de la suite $(x_n)_n$ sont les entiers q tels que $(a^q)_p = 1_p$.*

En effet, si $(a^n)_p = 1_p$, $f^n(x) = x - y + y = x$. Inversement si q est une période de la suite $(x_n)_n$, pour tout $n \in N$, $x_{n+q} = x_n$ et $f^{n+q}(s) = f^n(s)$, où s est le germe de la suite $(x_n)_n$, donc $(a^{n+q}(s - y))_p = (a^n(s - y))_p$, comme s n'est pas le point fixe de f, $s - y$ est non nul, et $(a^q)_p = 1_p$.

2 Petit théorème de Fermat

Rappelons la définition de l'ordre d'un élément dans Z/pZ.

Définition 1 : *L'ordre d'un élément a dans $(Z/pZ, \times)$ est le plus petit entier n tel que $a^n = 1_p$.*

Corollaire 3 : *La plus petite période de la suite $(x_n)_n$ est la période de a_p*

Nous allons établir que l'ordre d'un élément a_p dans $(Z/pZ, *)$ est un diviseur de $p - 1$. Pour cela nous aurons besoin du petit théorème de Fermat , selon lequel $(a^{p-1})_p = 1_p$.

Théorème 4 : *(petit théorème de Fermat). Dans Z/pZ, $(a^{p-1})_p = 1_p$*

Lemme 5 : *Dans Z/pZ, $((a + b)^p)_p = (a^p)_p + (b^p)_p$.*

Comme Z/pZ est un corps commutatif, on peut appliquer la formule du binôme de Newton

$$((a + b)^p)_p = \sum_{k=0}^{p} C_p^k (a^k)_p (b^{p-k})_p$$

$C_p^k = p(p - 1)...(p - k + 1)$ est divisible par p pour $1 \leq k \leq p - 1$. Donc $C_p^k (a^k)_p (b^{p-k})_p = (0)_p$, pour $1 \leq k \leq p - 1$, et $((a + b)^p)_p = (a^p)_p + (b^p)_p$.

Lemme 6 : *Dans Z/pZ, pour $n \geq 2$, $(x_1 + x_2 + ... + x_n)^P = x_1^p + x_2^p + ... + x_n^p$*

D'après le lemme précédent, la formule est vraie pour $n = 2$.

Supposons la formule vraie à l'ordre n, alors

$$\begin{aligned}(x_1+x_2+...+x_{n+1})^p &= ((x_1+x_2+...+x_n)+x_{n+1})^p \\ &= (x_1+x_2+...+x_n)^p + x_{n+1}^p\end{aligned}$$

grâce au lemme précédent. En appliquant l'hypothèse de récurrence,

$$(x_1+x_2+...+x_{n+1})^p = x_1^p+x_2^p+...+x_n^p+x_{n+1}^p$$

On a établi la formule par récurrence pour tout entier $n \geq 2$.

Soit a un entier qui vérifie $1 \leq a \leq p-1$, en prenant $x_i = 1$, pour $1 \leq i \leq a$, le lemme précédent donne

$$a^p = (1+1+...+1)^p = 1^p+1^p+...+1^p = a1^p = a$$

et comme a est inversible dans Z/pZ, $a^{p-1} = 1$, ce qui démontre le petit théorème de Fermat.

Proposition 7 : *L'ordre d'un élément dans Z/pZ est un diviseur de $p-1$.*

Soit n l'ordre de a, dans Z/pZ. $n \leq p-1$, car $a^{p-1} = 1$. Soit q et r le quotient et le reste de $p-1$, dans la division euclidienne de $p-1$ par n :

$$p-1 = qn+r,\ 0 \leq r < n$$

Alors $1 = a^{p-1} = a^{qn+r} = (a^n)^q a^r = 1^q a^r = a^r$. Si r n'était pas nul, on aurait $a^r = 1$, avec $0 < r < n$, ce qui contredirait la définition de n. Par suite $r = 0$, et $p-1 = qn$, n est un diviseur de $p-1$.

En conséquence, si l'on veut obtenir une suite $(x_n)_n$ dont le plus petite période soit égale à $p-1$, il nous faut rechercher les éléments a_p de Z/pZ d'ordre $p-1$, de tels éléments s'appellent des racines primitives de p.

Définition 2 : *Si a est un élément d'ordre $p-1$ de Z/pZ, on dit que a et une racine primitive de p.*

Heureusement, les racines primitives de p ne sont pas rares, grâce à un résultat de Legendre.

Théorème 8 : *(Legendre) Soit p un nombre premier impair. Il existe $\Phi(p-1)$ racines primitives de 1, modulo p, où Φ désigne la fonction d'Euler.*

Rappelons la définition de la fonction Φ d'Euler.

3 La fonction Φ d'Euler

Définition 3 : *Pour tout entier naturel n on note $\Phi(n)$ le nombre d'entiers naturels non nuls inférieurs à n et premiers avec n.*

Exemple 3 : $\Phi(1) = 1, \Phi(10) = 4$. $\Phi(p) = p - 1$, *si p est un nombre premier*

On démontre dans tous les traités d'arithmétique le lemme suivant :

Lemme 9 : *Si $n = m_1 m_2 ... m_p$ où les entiers m_i et m_j sont premiers entre eux pour toute paire d'entiers $\{i, j\}$ de $\{1, 2, ..., p\}$*

$$\Phi(n) = \Phi(m_1)\Phi(m_2)...\Phi(m_p)$$

En utilisant ce lemme, Euler a démontré en 1760 le théorème suivant :

Théorème 10 : *(Euler, 1760) Soit n un entier naturel et $n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_k^{\alpha_k}$ sa décomposition en produit de facteurs premiers. Alors*

$$\Phi(n) = \Phi(p_1^{\alpha_1})\Phi(p_2^{\alpha_2})...\Phi(p_k^{\alpha_k})$$

Par ailleurs, on démontre que si p est un nombre premier

$$\Phi(p^\alpha) = p^{\alpha-1}(p-1) = p^\alpha(1 - \frac{1}{p})$$

En conséquence, si $n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_k^{\alpha_k}$,

$$\begin{aligned} \Phi(n) &= p_1^{\alpha_1 - 1}(p_1 - 1)p_2^{\alpha_2 - 1}(p_2 - 1)...p_k^{\alpha_k}(p_k - 1) \\ &= n(1 - \frac{1}{p_1})(1 - \frac{1}{p_2})...(1 - \frac{1}{p_k}) \end{aligned}$$

Exemple 4 : $\Phi(2560) = \Phi(2^9.5) = 2560(1 - \frac{1}{2})(1 - \frac{1}{5})$.

On a alors comme conséquence du lemme, un résultat assez surprenant, dont nous allons avoir besoin.

Théorème 11 : *Si n est un entier positif, alors $n = \sum_{d|n} \Phi(d)$. (La sommation s'effectue sur l'ensemble des diviseurs de n).*

Nous allons raisonner par récurrence sur n. Il est clair que la propriété est vraie si $n = 1$. Supposons la propriété vraie pour tout entier $n \leq N - 1$, et démontrons la dans le cas où $n = N$. Soit p un diviseur premier de N. Posons $N = p^\alpha M$ où M et p sont premiers entre eux. Lorsque d parcourt l'ensemble des diviseurs de M, les diviseurs de n sont les nombres de la forme $d, pd, p^2 d, ..., p^\alpha d$. Il s'en suit que

$$\sum_{\delta|N} \Phi(\delta) = \sum_{d|M} [\Phi(d) + \Phi(pd) + ... + \Phi(p^\alpha d)]$$

D'après le lemme, pour chaque entier i, vérifiant $1 \leq i \leq \alpha$

$$\Phi(p^i d) = \Phi(p^i)\Phi(d)$$

Il en résulte que

$$\begin{aligned}\sum_{\delta|N}\Phi(\delta) &= \sum_{d|M}\Phi(d)[1+(p-1)+p(p-1)+...+p^{\alpha-1}(p-1)]\\ &= \sum_{d|M}\Phi(d)p^{\alpha}\\ &= p^{\alpha}\sum_{d|M}\Phi(d)\end{aligned}$$

Par hypothèse de récurrence, $\sum_{d|M}\Phi(d)=M$, et $\sum_{\delta|N}\Phi(\delta)=p^{\alpha}M=N$.

Avant d'établir le théorème de Legendre, il nous faut énoncer et démontrer le théorème de Lagrange sur les polynômes à coefficients dans Z/pZ, qui ressemble au théorème de d'Alembert sur les polynômes à coefficients complexes.

4 Théorème de Lagrange

Théorème 12 : *(Lagrange, 1768) Soit p un nombre premier, et soit $f(x)=a_nx^n+a_{n-1}x^{n-1}+...+a_0$ un polynôme à coefficients dans Z/pZ, de degré n. Alors l'équation $f(x)\equiv 0 \bmod(p)$, admet au plus n solutions dans Z/pZ.*

Pour établir ce théorème, nous commencerons par la démonstration de résultats, analogues à ceux que l'on établit pour les polynômes à coefficients réels ou complexes.

Théorème 13 : *Soit $f(x)=a_nx^n+a_{n-1}x^{n-1}+...+a_0$, un polynôme à coefficients dans Z/pZ et c_1 une solution dans Z/pZ de l'équation $f(x)=0$. On peut alors écrire que $f(x)=(x-c_1)f_1(x)$, où $f_1(x)$ est un polynôme à coefficients dans Z/pZ et de degré $n-1$.*

En effet si c_1 est une solution de l'équation $f(x)=0$, $f(x)=f(x)-f(c_1)$ dans l'anneau $Z/pZ[X]$. Mais $f(x)-f(c_1)=\sum_{i=1}^{n}a_i(x^i-c_1^i)$, la sommation pouvant débuter à $i=1$. Pour $1\leq i\leq n$,

$$x^i-c_1^i=(x-c_1)[\sum_{k=0}^{i-1}x^{i-1-k}c_1^k]$$

Par suite

$$f(x)-f(c_1)=(x-c_1)\sum_{i=1}^{n}a_i[\sum_{k=0}^{i-1}x^{i-1-k}c_1^k]=(x-c_1)f_1(x)$$

où $f_1(x)$ est un polynôme à coefficients dans Z/pZ, de degré $n-1$, son coefficient de degré $n-1$ étant égal à a_n.

Corollaire 14 : *Soit $c_1, c_2, ..., c_r$ r solutions distinctes de $f(x) = 0$ dans Z/pZ, où p est un nombre premier. Alors*

$$f(x) = (x - c_1)(x - c_2)...(x - c_r)f_r(x)$$

où $f_r(x)$ est un polynôme à coefficients dans Z/pZ, de degré $n - r$.

D'après le théorème que l'on vient d'établir $f(x) = (x - c_1)f_1(x)$ où $f_1(x)$ est un polynôme à coefficients dans Z/pZ. Comme c_2 est aussi une solution de l'équation $f(x) = 0$, $(c_2 - c_1)f_1(c_2) = 0$. Comme c_1 est distinct de c_2, $c_1 - c_2$ est non nul et Z/pZ étant un corps, $f_1(c_2) = 0$. En appliquant le théorème à f_1, on conclue à l'existence d'un polynôme $f_2(x)$, à coefficients dans Z/pZ, de degré $n - 2$, tel que $f_1(x) = (x - c_2)f_2(x)$. On achève facilement la démonstration du corollaire par un raisonnement par récurrence sur le nombre r de racines de f considérées.

Nous sommes maintenant en mesure de démontrer le théorème de Lagrange.

Soit $f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_0$, un polynôme à coefficients dans Z/pZ de degré n. Supposons que $f(x)$ ait $n + 1$ racines distinctes dans Z/pZ : $c_1, c_2, ..., c_n, c_{n+1}$. En appliquant le corollaire précédent aux n premières racines, on montre qu'il existe un polynôme de degré 0, $f_n(x)$, à coefficients dans Z/pZ tel que

$$f(x) = (x - c_1)(x - c_2)...(x - c_n)f_n(x)$$

$f_n(x)$ est donc un polynôme constant. Posons $f_n(x) = b$ On voit immédiatement que $b = a_n$. c_{n+1} étant une racine de $f(x)$, on a la relation

$$b(c_{n+1} - c_1)(c_{n+1} - c_2)...(c_{n+1} - c_n) = 0$$

Comme pour tout i, vérifiant $1 \leq i \leq n$, $c_{n+1} - c_i$ est non nul et que cette relation est écrite dans le corps Z/pZ on en conclue que $b = a_n = 0$, ce qui contredit le fait que $f(x)$ soit un polynôme de degré n. Par suite $f(x)$ ne saurait avoir plus de n racines distinctes.

Remarque 3 : *Le fait que p soit premier est essentiel. Par exemple l'équation $2x - 2 = 0$ admet deux solutions dans $Z/4Z$, à savoir $x = (1)_4$ et $x = (3)_4$, bien que $2x - 2$ soit un polynôme du premier degré.*

Il nous faut encore établir deux corollaires du théorème de Lagrange.

Corollaire 15 : *Soit p un nombre premier, alors dans $Z/pZ[X]$*

$$x^{p-1} - 1_p = (x - 1_p)(x - 2_p)...(x - (p - 1)_p)$$

En effet d'après le petit théorème de Fermat, $x^{p-1} = 1_p$ pour tout $x \in Z/pZ$. Par suite

$$x^{p-1} - 1_p = (x - 1_p)(x - 2_p)...(x - (p - 1)_p)f_p(x)$$

où $f_p(x)$ est de degré 0. Naturellement $f_p(x) = 1_p$.

Nous en arrivons enfin au corollaire du théorème de Lagrange, dont nous aurons besoin pour démontrer le théorème de Legendre.

Corollaire 16 : *Soit p un nombre premier et d un diviseur de $p-1$. Le polynôme $x^d - 1_p$ a exactement d racines dans Z/pZ.*

Comme d divise $p-1$, il existe un entier q tel que $p-1 = qd$. Alors, en posant $X = x^d$, on a

$$x^{p-1} - 1_p = X^q - 1_p = (X - 1_p)(X^{q-1} + X^{q-2} + ... + X + 1) = (x^d - 1)g(x)$$

où $g(x)$ est un polynôme de degré $p - 1 - d$. Considérons maintenant les équations

$$\begin{aligned} x^d - 1_p &= 0 \\ g(x) &= 0 \end{aligned}$$

dans Z/pZ. D'après le théorème de Lagrange, la première équation a au plus d solutions et la seconde a au plus $p - 1 - d$ solutions. Comme p est un nombre premier, toute solution u, dans Z/pZ, de l'équation $x^{p-1} - 1_p = 0$, est soit une solution de $x^d - 1_p = 0$, soit une solution de $g(x) = 0$. Si l'équation $x^d - 1_p = 0$ avait moins de d solutions dans Z/pZ, comme l'équation $x^{p-1} - 1_p = 0$ a exactement $p - 1$ solutions dans Z/pZ l'équation $g(x) = 0$ aurait plus de $p - 1 - d$ solutions, ce qui est impossible à cause du théorème de Lagrange. Par suite l'équation $x^d - 1$ a exactement d solutions dans Z/pZ.

Nous allons maintenant démontrer le théorème de Legendre.

5 Théorème de Legendre

Démontrons maintenant le théorème de Legendre, selon lequel si p est un nombre premier, alors il existe $\Phi(p-1)$ racines primitives de p.

Notons $\Psi(e)$, le nombre de solutions de l'équation $x^{p-1} - 1_p = 0$, dans Z/pZ d'ordre e exactement (au sens de l'ordre d'un élément dans Z/pZ). Nous nous proposons d'établir que $\Psi(e) = \Phi(e)$.

Cette égalité est évidente si $e = 1$, car alors $x = 1_p$ et $\Psi(e) = 1$, donc $\Psi(e) = \Phi(e) = 1$.

On a vu que si a est un élément de Z/pZ d'ordre e, alors e divise $p-1$. On peut donc restreindre son attention aux diviseurs de $p-1$. Soit d un tel diviseur supposons que l'on ait établi l'égalité $\Psi(\delta) = \Phi(\delta)$ pour tout diviseur δ de $p-1$, inférieur à d. D'après le corollaire du théorème de Lagrange, on sait que l'équation $x^d - 1_p = 0$ a exactement d solutions. Chacune de ces solutions est évidemment une solution, dont l'ordre est un diviseur de d, de l'équation $x^{p-1} - 1_p = 0$. Réciproquement toute solution de l'équation $x^{p-1} - 1_p = 0$, dont l'ordre divise d, est par définition même de l'ordre, une solution de l'équation $x^d - 1_p = 0$. Par conséquent $d = \sum_{\delta|d} \Psi(\delta)$. Par hypothèse de récurrence, $\Psi(\delta) = \Phi(\delta)$ si $\delta|d$ et $\delta < d$. On peut donc réécrire la sommation précédente sous la forme

$$d = \sum_{\delta|d} \Psi(\delta) = \sum_{\delta|d} \Phi(\delta) - \Phi(d) + \Psi(d)$$

Mais on a vu que pour tout entier naturel non nul $t = \sum_{\tau|t} \Phi(\tau)$. Par suite $d = d - \Phi(d) + \Psi(d)$, c'est à dire que $\Phi(d) = \Psi(d)$.

Le théorème de Legendre s'en déduit immédiatement en prenant $e = p - 1$. On obtient $\Psi(p-1) = \Phi(p-1)$, comme le dit le théorème.

Exemple 5 : *Si* $p = 11$, *alors* $1^1 = 2^{10} \equiv 3^5 \equiv 4^5 \equiv 5^5 \equiv 6^{10} \equiv 7^{10} \equiv 8^{10} \equiv 9^5 \equiv 10^2 \mod(p)$. *Ainsi* $2, 6, 7$ *et* 8 *sont les* $4 = \Phi(10) = 10(1 - \frac{1}{2})(1 - \frac{1}{5})$ *racines primitives de* 1_{11} *dans* $Z/11Z$.

Le théorème de Legendre nous indique que les racines primitives de 1_p ne sont pas rares, pour un nombre premier donné p. Par contre il n'existe pas d'algorithme permettant de les trouver. Gauss disait que "leur distribution était un grand mystère" et Euler ne connaissait aucune règle pour en trouver une. On peut évidemment procéder par essai successifs pour des nombres premiers pas trop grand et se faire aider par des ordinateurs.

6 Racines primitives de p, lorsque p est premier

On choisira un très grand nombre premier p , par exemple supérieur à 100 000, de manière en prenant pour a une racine primitive de p, à obtenir une suite $(x_n)_n$ dont la période soit grande.

Prime[10000]

104729

p = Prime[n] est le n-ième nombre premier. Décomposons $p - 1$ en produit de facteurs premiers à l'aide de la fonction FactorInteger.

FactorInteger[104729-1]

{{2,3},{13,1},{19,1},{53,1}}

104 728 =2^3.13.19.53. Calculons $\Phi(104728)$

phi = 104728 (1-1/2)(1-1/13)(1-1/19)(1-1/53)

44928

En prenant un entier au hasard entre 1 et 104728 on a donc, grâce au théorème de Legendre une probabilité égale à 44928/104728, de tomber sur un entier a tel que $(a)_{104729}$ soit une racine primitive de 104729. Essayons 3.

L'ordre de 3 est un diviseur de 104728. Les diviseurs de 104728 sont de la forme $2^\alpha 13^\beta 19^\gamma 53^\delta$, avec $0 \leq \alpha \leq 3$, $0 \leq \beta \leq 1$, $0 \leq \gamma \leq 1$, $0 \leq \delta \leq 1$.Il y en a donc $4.2.2.2 = 32$. Construisons la liste des diviseurs de 104728.

list3 =
Distribute[{{0,1,2,3},{0,1},{0,1},{0,1}},List,List]

{{0,0,0,0},{0,0,0,1},{0,0,1,0},{0,0,1,1},{0,1,0,0},{0,1,0,1},{0,1,1,0},
{0,1,1,1},{1,0,0,0},{1,0,0,1},{1,0,1,0},{1,0,1,1},{1,1,0,0},{1,1,0,1},
{1,1,1,0},{1,1,1,1},{2,0,0,0},{2,0,0,1},{2,0,1,0},{2,0,1,1},{2,1,0,0},
{2,1,0,1},{2,1,1,0},{2,1,1,1},{3,0,0,0},{3,0,0,1},{3,0,1,0},{3,0,1,1},
{3,1,0,0},{3,1,0,1},{3,1,1,0},{3,1,1,1}}

On a construit d'abord l'ensemble produit A = {0,1,2,3}× {0,1} × {0,1}× {0,1}. On construit ensuite l'application q de cet ensemble dans l'ensemble des diviseurs de 104728 qui à (x, y, z, t) associe $2^x.13^y.19^z53^t$. On applique ensuite cette application à chacun des éléments de A, ce qui donne la liste, notée div, des diviseurs de 104728.

p[x_, y_, z_, t_] := $2^x.13^y.19^z53^t$
q[u_List] := Apply[p, u]
div = Map[q, list3]
{1,53,19,1007,13,689,247,13091,2,106,38,2014,26,1378,494,26182,4,
212,76,4028,52,2756,988,52364,8,424,152,8056,104,5512,1976,104728}

Elevons $(3)_{104729}$ à toutes les puissances $(3^k)_{104729}$ lorsque k parcourt l'ensemble des diviseurs de 104728.

r[a_, b_] := Mod[a^b, 104729]
Map[r[3, #]&, div]
{3,102446,83754,33964,23388,79730,59571,70765,9,80368,88825,68090,
103706,32058,66605,68090,81,63807,16681,104728,103968,9687,10314,
104728,6561,98103, 95537,1,55476,785,78661,1}
On constate que 3 n'est pas d'ordre 104728, mais d'ordre 8056. Vérifions.

r[3,8056]

1

Faisons d'autres essais.

Map[r[5,#]&, div]

{5,23894,76154,9176,86630,22401,52755,1,25,45457,63341,101589,
86218,48162,21579,1,625,35679,19020,15074,88562,40352,28107,1,
76438,10046,26434,68275, 73034,62141,32602,1}

Pas de chance 5 est d'ordre 13091. Vérifions.

r[5,13091]

1

Après quelques essais on trouve que 12 est d'ordre 104728.

Map[r[12,#]&, div]

{12,81998,96799,3329,6932,70998,45158,33964,144,70204,47500,
85696,86742,4505,66605,68090,20736,54876,73153,102207,24288,
82328,10314,104728,69151,102439, 23696,76744,73216,48162,
78661,1}

On prendra donc $a = 12$, et $b = 3$.
Cherchons alors le point fixe de f.

Solve[12 x + 3 ==x && Modulus ==104729,x]

{{Modulus -> 104729,x -> 38083}}

Vérifions que la suite $(x_n)_n$ obtenue en prenant $f(x) = (12x + 3)_{104729}$ a bien pour plus petite période 104728.

a=12 ; b= 3;
list4= NestList[g[#,104729]&,0,104728] ;
Length[Union[Drop[list4,1]]]

104728

Mais les ordinateurs calculent en base 2. Un bon moyen d'accélérer les calculs est donc de se placer dans l'anneau $Z/2^{16}Z$. Dans ce cas pour un nombre rationnel compris entre 0 et 1, la division par 2^{16} consistera simplement à déplacer la virgule de 16 places vers la droite. Malheureusement, il n'y a pas d'entiers a tel que $(a)_{2^{16}}$ soit d'ordre $2^{16}-1$.

7 Racine primitive de a, entier non premier

Les éléments $(a)_n$ de $(Z/nZ)^*$, inversibles pour la multiplication sont ceux pour lesquels a est premier avec n. En effet si a est premier avec n, d'après l'identité de Bezout, il existe des entiers λ et μ tels que

$$\lambda a + \mu n = 1$$

Par suite, $(\lambda)_n(a)_n = 1_n$ et $(a)_n$ est un élément inversible pour la multiplication dans $(Z/nZ)^*$. Inversement, si $(a)_n$ est inversible poiur la multiplication dans $(Z/nZ)^*$ il existe un entier relatif μ tel que $\lambda a - 1 = \mu n$, soit $\lambda a + (-\mu)n = 1$, ce qui prouve que a et n sont premiers entre eux. Par suite le nombre d'éléments du groupe des éléments inversibles pour la multiplication dans $(Z/nZ)^*$ est $\Phi(n)$. Soit $(a)_n$ un élément inversible pour la multiplication dans $(Z/nZ)^*$, l'ordre k du sous-groupe cyclique engendré par $(a)_n$ est un diviseur de l'ordre du groupe des éléments inversibles de $(Z/nZ)^*$. Il existe donc un entier m tel que $k.m = \Phi(n)$, et $a_n^{\Phi(n)} = a_n^{km} = (a_n^k)^m = 1_n^m = 1_n$. On a ainsi démontré un théorème du à Euler.

Théorème 17 : *(Euler, 1760) Si a et n sont premiers entre eux, alors $a^{\Phi(n)} \equiv 1 \bmod(n)$.*

Définition 4 : *Soit n un entier supérieur ou égal à 2, et a un entier premier avec n. Si l'ordre de a_n est $\Phi(n)$, on dit que a est une racine primitive de n.*

Exemple 6 : *8 n'a pas de racine primitive. En effet* $\Phi(8) = 8(1-\frac{1}{2}) = 4$ *et* $1^1 \equiv 3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \bmod(8)$

15 n'a pas de racine primitive. En effet $\Phi(15) = 15(1-\frac{1}{3})(1-\frac{1}{5}) = 8$ *et* $1^1 \equiv 2^4 \equiv 4^2 \equiv 7^4 \equiv 8^4 \equiv 11^2 \equiv 13^4 \equiv 14^2 \equiv 1 \bmod(15)$.

27 a une racine primitive, 2 en est une : $\Phi(27) = 27(1-\frac{1}{3}) = 18$.

Table[Mod[2^k, 27], {k, 1, 18}]

{2,4,8,16,5,10,20,13,26,25,23,19,11,22,17,7,14,1}

Tout cela semble très chaotique. $\Phi(2^{16}) = 2^{15}(2-1) = 2^{15} = 32768$. Par suite le groupe multiplicatif des éléments inversibles de $Z/2^{16}Z$ est d'ordre 2^{15}, et n'a pas d'élément d'ordre 2^{16}. Il n'a pas non plus d'élément d'ordre 2^{15} : 2^{15} n'a pas de racine primitive.

Théorème 18 : *Soit $n = 2^k$. Pour $k = 1$ et 2, n a une racine primitive. Pour $k \geq 3$, n n'a pas de racine primitive.*

1 est une racine primitive de 2, tandis que 3 est une racine primitive de 4 : $\Phi(4) = 2^1(2-1) = 2$, et $3^2 \equiv 1 \bmod(4)$. Si a est une racine primitive de 2^k, a est impair. Soit $k \geq 3$, $\Phi(k) = 2^{k-1}(2-1) = 2^{k-1}$,nous allons prouver que si a est impair, $a^{2^{k-2}} \equiv 1 \bmod(2^k)$. Si $k = 3$, l'exemple ci-dessus a montré que $8 = 2^3$ n'avait pas de racine primitive et que $a^2 \equiv 1 \bmod(2^3)$, si a est impair, avec $\Phi(8) = 4 = 2^2$. Supposons que pour $3 \leq l$, on ait $a^{2^{l-2}} \equiv 1 \bmod(2^l)$. Il existe alors un entier relatif b tel que $a^{2^{l-2}} = 1 + b2^l$. En élevant au carré les deux membres de cette expression, on obtient

$$a^{2^{l-1}} = (1 + b2^l)^2 = 1 + 2b.2^l + b^2 2^{2l}$$

et $a^{2^{l-1}} \equiv 1 \bmod(2^{l+1})$. La propriété est démontrée pour $n = l + 1$. La proposition est établie par récurrence.

Gauss a démontré en 1801 le théorème suivant :

Théorème 19 : *Soit n un entier supérieur à 1, alors n n'admet de racine primitive que si $n = 2, 4, p^k$ ou $2p^k$, où p est un nombre premier impair.*

Dans le cas où $n = 2^k$ avec $k \geq 3$, tout entier impair a, vérifie $a^{2^{l-2}} \equiv 1 \bmod(2^l)$,comme nous l'avons vu dans la démonstration du théorème 18. Pour $a = 3$, on vérifie que 3 est d'ordre 2^{k-2}, dans le groupe multiplicatif de $Z/2^kZ$. En particulier vérifions à l'aide de Mathematica que 3 est d'ordre 2^{14} dans le groupe $Z/2^{16}Z$.

```
list5 = Table[Mod[3^k, 2^16], {k, 2^14}];
2^14
Length[Union[list5] ])
16384
16384
```

On a fait calculer la liste des puissances 3^k de 3 dans $Z/2^{16}Z$, pour k compris entre 1 et 2^{14}, puis on supprime les éventuelles répétitions dans cette liste en lui appliquant la fonction Union. On remarque qu'il n'y en a pas puisque cette liste a bien 2^{14} = 16 384 éléments. 3 est bien d'ordre 2^{14} dans $Z/2^{16}Z$.

On peut donc construire la suite $(x_n)_n$ d'éléments de $Z/2^{16}Z$ en posant $x_0 = s$ et $x_n = f(x_{n-1})$ où $f(x) = ax + b$ en prenant $a = 3$ et par exemple $b = 1$. Alors $s = 0$ n'est pas un point fixe. Calculons les 100 premiers termes de cette liste.

```
a = 3 ;
b = 1; list6 = NestList[g[#, 2^16]&, 0, 99]
{0,1,4,13,40,121,364,1093,3280,9841,29524,23037,3576,10729,32188,
31029,27552,17121,51364,23021,3528,10585,31756,29733,23664,5457,
16372,49117,16280,48841,15452,46357,8000,24001,6468,19405,58216,
43577,65196,64517,62480,56369,38036,48573,14648,43945,764,2293,
6880,20641,61924,54701,33032,33561,35148,39909,54192,31505,28980,
21405,64216,61577,53660,29909,24192,7041,21124,63373,59048,46073,
```

7148,21445,64336,61937,54740,33149,33912,36201,43068,63669,59936,
48737,15140,45421,5192,15577,46732,9125,27376,16593,49780,18269,
54808,33353,34524,38037,48576,14657,43972,845}

On en déduit 100 nombres aléatoires compris entre 0et 1.

list7 = list6/2^{16} // N

{0,0.0000152588,0.0000610352,0.000198364,0.000610352,0.00184631,
0.0055542,0.0166779,0.0500488,0.150162,0.4505,0.351517,0.0545654,
0.163712,0.49115,0.473465,0.42041,0.261246,0.783752,0.351273,
0.053833,0.161514,0.484558,0.45369,0.361084,0.0832672,0.249817,
0.749466,0.248413,0.745255,0.235779,0.707352,0.12207,0.366226,
0.0986938,0.296097,0.888306,0.664932,0.994812,0.984451,0.953369,
0.860123,0.580383,0.741165,0.223511,0.670547,0.0116577,0.0349884,
0.10498,0.314957,0.944885,0.834671,0.504028,0.5121,0.536316,
0.608963,0.826904,0.480728,0.4422,0.326614,0.979858,0.93959,
0.818787,0.456375,0.369141,0.107437,0.322327,0.966995,0.901001,
0.703018,0.10907,0.327225,0.981689,0.945084,0.835266,0.505814,
0.517456,0.552383,0.657166,0.971512,0.914551,0.743668,0.231018,
0.693069,0.0792236,0.237686,0.713074,0.139236,0.417725,0.253189,
0.759583,0.278763,0.836304,0.508926,0.526794,0.580399,0.741211,
0.223648,0.670959,0.0128937}

Vérifions que ces nombres sont biens répartis selon une loi uniforme. Pour cela, on les regroupe en classes d'égale amplitude, à savoir 0,1 dans l'intervalle[0,1].Puis on représente les effectifs de ces classes par un diagramme en bâtons.

```
<<Statistics'DataManipulation'
list8=BinCounts[list7,{0,1,0.1}]
```

{16,8,11,9,10,10,6,10,7,12}

On dénombre les effectifs de chacune des classes par la fonction BinCounts, du "package" Statistics'DataManipulation', qu'il faut préalablement charger. On calcule les centres des classes précédentes.

```
centres=Table[0.05+k 0.1, {k,0,9}]
```

{0.05,0.15,0.25,0.35,0.45,0.55,0.65,0.75,0.85,0.95}

On représente enfin par un diagramme en bâtons les effectifs de la série list7, une fois classée.

```
<<Graphics'Graphics'
BarChart[Transpose[{list8, centres}]]
```

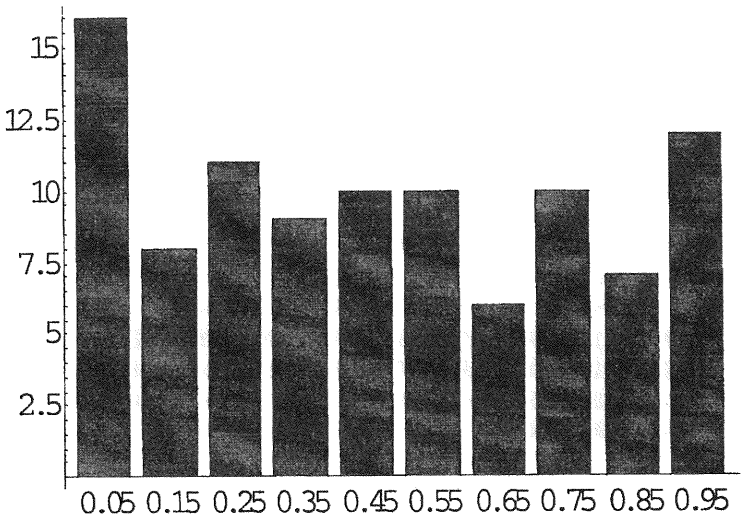


Figure 2 : Répartition de 100 nombres pseudo-aléatoires en classes d'amplitude 0,1

Remarque 4 : *Pour construire un diagramme en bâtons, il nous faut disposer de la fonction BarChart, qui se trouve dans le "package" Graphics'Graphics', qu'il faut donc charger préalablement. Transpose vise à constituer une liste de couples dont le premier élément est l'effectif de la classe et le second le centre de cette classe*

La première classe a un effectif nettement plus élevé que celui des autres classes, cela s'explique sans doute par la trop petite valeur donnée à a, à savoir 3. Il faut donc prendre une valeur de a plus centrale. Essayons $a = 25733$. Il nous faut d'abord trouver l'ordre de $(25733)_{2^{16}}$ dans $Z/2^{16}Z$.

Mod[$25733^{2^{14}}$, 2^{16}]

1

L'ordre de 25733 est donc un diviseur de 2^{14}. Les diviseurs de 2^{14} sont les nombres de la forme 2^k avec $0 \leq k \leq 14$. Calculons donc $(25733^{2^k})_{2^{16}}$ pour $1 \leq k \leq 14$.

Table[Mod[25733^{2^k}, 2^{16}], {k, 1, 14}]

{11545,52337,18913,6081,16257,48897,32257,64513,63489,61441,
57345,49153,32769,1}

$(25733)_{2^{14}}$ est bien d'ordre 2^{14}dans le groupe multiplicatif de $Z/2^{16}Z$. On prend a =$(25733)_{2^{16}}$, $b = (13849)_{2^{16}}$, $s = 0$.

a = 25733;
b = 13849;
list9 = NestList[g[#, 2^{16}]&, 0, 99]

{0,13849,5398,49799,572,53061,55538,29651,52920,31665,40206,

18015,57716,42845,32106,50731,63088,65097,54790,48951,2476,
27765,17122,15747,22312,7649,40958,36111,23268,32397,2394,
14811,53472,14969,56054,4071,46364,15781,45266,6963,17304,
46097,26350,42943,63572,2493,6474,16779,36688,60073,9190,
46231,1164,17109,8898,3299,37896,15937,61918,38511,47556,
18669,44346,56635,12736,3801,45270,44359,63484,31749,39090,
4755,18552,48241,16590,22815,40756,15389,50474,3307,47152,
42761,33222,63991,36716,59701,4770,11331,24808,11937,21438,
61391,43172,58189,24858,53403,8864,45881,38582,39591}

list10 = list9/2^{16} // N

{0,0.211319,0.0823669,0.759872,0.00872803,0.809647,0.847443,
0.452438,0.807495,0.48317,0.613495,0.274887,0.880676,0.653763,
0.489899,0.774094,0.962646,0.993301,0.836029,0.746933,
0.0377808,0.42366,0.261261,0.24028,0.340454,0.116714,0.624969,
0.55101,0.355042,0.494339,0.0365295,0.225998,0.815918,0.228409,
0.855316,0.0621185,0.707458,0.240799,0.690704,0.106247,
0.264038,0.703384,0.402069,0.655258,0.970032,0.0380402,
0.0987854,0.256027,0.559814,0.916641,0.140228,0.705429,
0.0177612,0.261063,0.135773,0.0503387,0.578247,0.243179,
0.944794,0.587631,0.725647,0.284866,0.676666,0.864182,0.194336,
0.0579987,0.690765,0.676865,0.968689,0.484451,0.596466,
0.0725555,0.283081,0.736099,0.253143,0.348129,0.621887,
0.234818,0.770172,0.0504608,0.719482,0.652481,0.506927,
0.976425,0.560242,0.910965,0.0727844,0.172897,0.37854,0.182144,
0.327118,0.936752,0.658752,0.887894,0.379303,0.814865,0.135254,
0.700089,0.588715.0.604111}

list11 = BinCounts[list10, {0,1,0.1}]

{13,8,15,6,7,8,12,11,10,9}

BarChart[Transpose[{list11, centres}]]

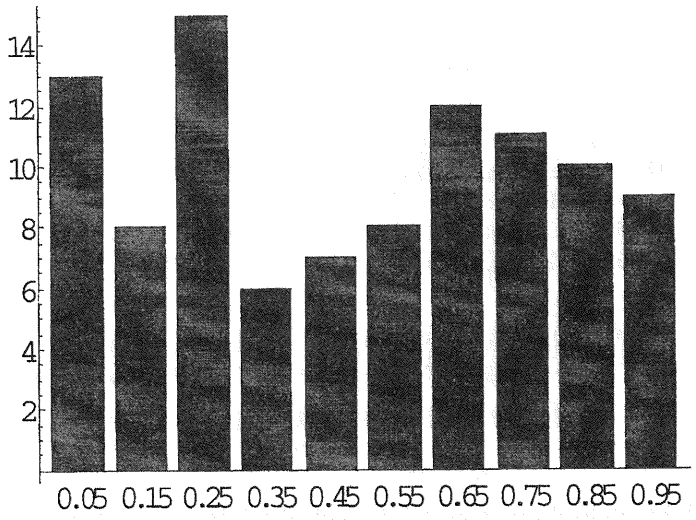


Figure 3 : Nouvelle répartition de nombres pseudo-aléatoires en classes d'amplitude 0,1.

L'échantillon n'est pas assez grand pour juger. Reprenons les calculs avec un échatillon de taille 10000, sans afficher les résultats intermédiaires.

```
list12 = NestList[g[#, 2^16]&, 0, 9999]
list13 = list12 /2^16 // N ;
list14 = BinCounts[list13, {0, 1, 0.1}] ;
BarChart[Transpose[{list14, centres}]]
```

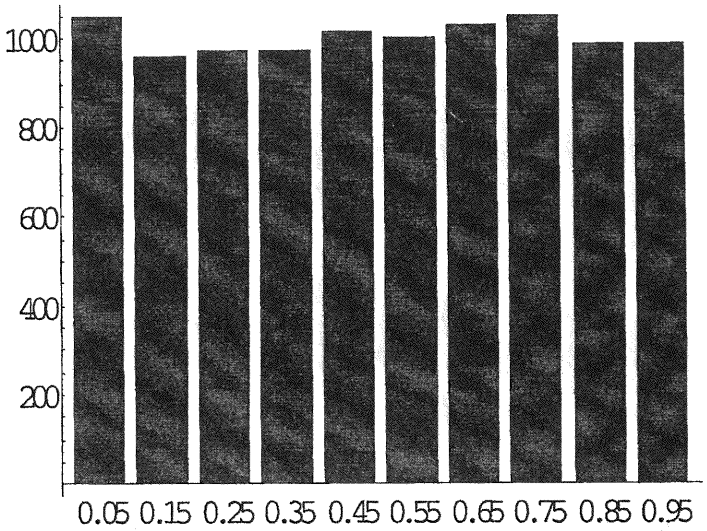


Figure 4 : Répartition de 10 000 nombres aléatoires en classes d'amplitude 0,1

8 Calcul d'une valeur approchée de π

Maintenant que nous faisons confiance aux générateurs de nombres aléatoires, utilisons les pour calculer une valeur approchée de π.

On dessine sur le sol un disque de rayon 1, inscrit dans un carré de côté de longueur 2. Il pleut. En faisant l'hypothèse que les gouttes de pluie se répartissent uniformément sur la surface du carré, l'aire du cercle étant égale à π, et celle du carré à 4, la probabilité qu'une goutte d'eau tombe à l'intérieur du disque est égale à $\frac{\pi}{4}$.

Faisons une représentation graphique de cette expérience. Chaque goutte de pluie sera représentée par un point rouge, le cercle est dessiné en vert, et les bords du carré en noir.

```
g1 = Graphics[
{RGBColor[1,0,0],AbsolutePointSize[4],
Point /@ Table[{Random[Real,{0,1}],Random[Real,{0,1}]},{100}]},
{AspectRatio -> Automatic}]
g2=Graphics[{RGBColor[0,1,0],Circle[{0.5,0.5},0.5] },{
AspectRatio -> Automatic}]
g3 = Graphics[Line[{{0,0},{1,0},{1,1},{0,1},{0,0}}]]
Show[g1, g2, g3]
```

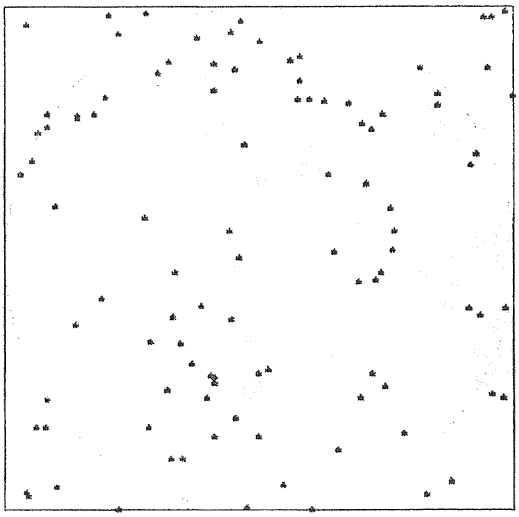


Figure 5 : Répartition de 100 gouttes de pluie à l'intérieur d'un carré

On génère une liste de couples de nombres aléatoires. Puis on transforme chacun de ces couples de réels en un point graphique par la primitive graphique Point. On applique point à chacun des éléments de la liste précédente par la fonction Map, qui apparaît ici sous sa forme infixe /@. La fonction Graphics rassemble alors tous ces objets en un seul graphique g1. L'option AspectRatio -> Automatic oblige Mathematica à utiliser la même échelle sur les deux axes. Les directives graphiques, placées en tête, s'imposent aux objets graphiques qui suivent. Ce sont ici RGBColor[1,0,0] et AbsolutePointSize[4] qui imposent la couleur rouge et la taille 4 pour les points. On construit de même un cercle et une ligne brisée pour représenter le carré. Enfin la fonction Show permet d'afficher simultanément les trois graphiques ainsi construits.

Qu'obtient-on pour vingt mille gouttes d'eau ?

```
list15 =
Table[{Random[Real, {0, 1}], Random[Real, {0, 1}]}, {20000}];
h1[x_, y_] :=(x - 1/2)^2 + (y - 1/2)^2
h2[u_List] := Apply[h1, u]
h3[u_List] := If[h2[u] <= 1/4,1, 0]
list16 = Map[h3, list15] ;
list17 = Select[list16, (# == 1)&] ;
4 (Length[list17]/Length[list16]) // N
```

3.1414

Que pensez-vous de ce résultat ?

Il serait sans doute utile d'indiquer la précision des résultats obtenus. En utilisant

l'inégalité de Bienaymé-Tchebychev, il est possible de calculer le nombre de gouttes de pluies nécessaires pour obtenir π par cette méthode avec une précision donnée.

En effet, si on se limite au premier quadrant, comme on vient de la faire dans la calcul avec Mathematica. On considère l'expérience aléatoire qui consiste à faire tomber n gouttes d'eau sur ce premier quadrant, en faisant l'hypothèse que les gouttes d'eau se répartissent uniformément sur le carré $[0,1]^2$. Soit Ω l'ensemble des issues de cette expérience aléatoire et (Ω, T, P) l'espace probabilisé associé. On peut considérer que la i-ème goutte d'eau est la réalisation d'une variable aléatoire $Z_i = (X_i, Y_i)$. Si n est le nombre de gouttes d'eau $1 \leq i \leq n$. Les variables aléatoires Z_i suivent une loi uniforme sur $[0,1]^2$ et elles sont mutuellement indépendantes. Le quart de disque contenu dans le domaine $[0,1]^2$ est le domaine D défini par

$$D = \{(x,y) \in [0,1]^2 : x^2 + y^2 \leq 1\}.$$

On considère la variable aléatoire T_i définie sur Ω par $T_i(\omega) = 1$ si $Z_i(\omega) \in D$ et $T_i(\omega) = 0$ sinon. Soit alors la variable $F_n = \frac{1}{n}\sum_{i=1}^{n} T_i$. $F_n(\omega)$ est la fréquence de réalisation de l'événement "la goutte d'eau est tombée dans le disque". On a

$$\begin{aligned} P([T_i &= 1]) = \frac{aire(D)}{aire([0,1]^2)} = \frac{\pi}{4} \\ P([T_i &= 0]) = 1 - \frac{\pi}{4} \end{aligned}$$

Les variables aléatoires T_i sont donc des variables aléatoires de Bernoulli, mutuellement indépendantes. Par suite $\sum_{i=1}^{n} T_i$ suit une loi binômiale $\mathcal{B}(n, \frac{\pi}{4})$, donc $E(\sum_{i=1}^{n} T_i) = n\frac{\pi}{4}$ et $V(\sum_{i=1}^{n} T_i) = n\frac{\pi}{4}(1 - \frac{\pi}{4})$. On en déduit que

$$E(F_n) = \frac{\pi}{4} \text{ et que } V(F_n) = \frac{\frac{\pi}{4}(1 - \frac{\pi}{4})}{n}$$

On voit que $\lim_{n \to +\infty} V(F_n) = 0$. Par suite, d'après la loi forte des grands nombres la convergence de (F_n) vers $\frac{\pi}{4}$ est presque sûre et à fortiori (F_n) converge en probabilité.

L'inégalité de Bienaymé-Tchebichev

$$P([|F_n - \frac{\pi}{4}| \geq \varepsilon]) \leq \frac{Var(F_n)}{\varepsilon^2}$$

écrite pour $\varepsilon = 0,01$ devient

$$P([|F_n - \frac{\pi}{4}| \geq \varepsilon]) \leq \frac{\frac{\pi}{4}(1 - \frac{\pi}{4})}{n(0,01)^2}$$

Pour que $P([|F_n - \frac{\pi}{4}| \geq 0,01])$ soit inférieur à $0,005$ il suffit que

$$\frac{\frac{\pi}{4}(1 - \frac{\pi}{4})}{n(0,01)^2} \leq 0,005$$

c'est à dire que

$$n \geq \frac{\frac{\pi}{4}(1-\frac{\pi}{4})}{0,005(0,01)^2} = \frac{\frac{\pi}{4}(1-\frac{\pi}{4})}{5.10^{-7}} = \frac{\pi(3-\pi)}{8.10^{-6}}$$

Comme $3,1 \leq \pi \leq 3,2$, $3-\pi \leq 0.1$, $\pi(3-\pi) \leq 0,1 \times 3,2 = 0,32$. Il suffit donc que

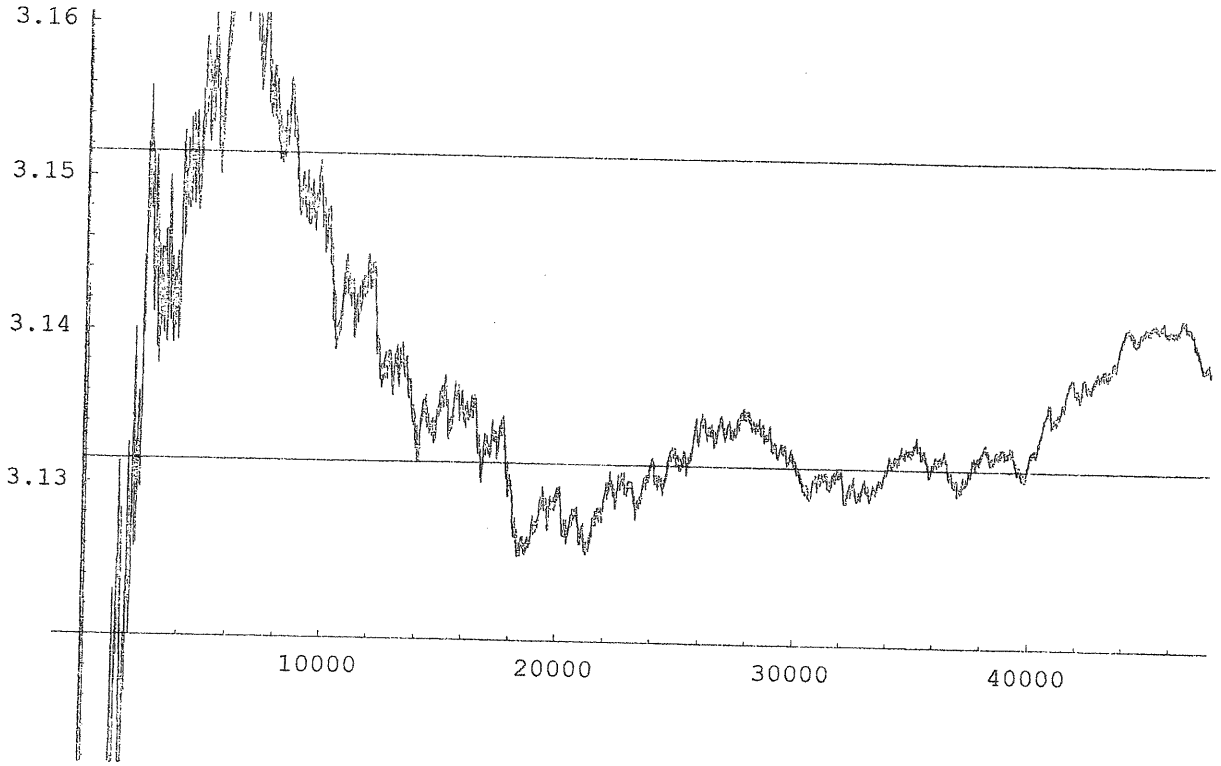
$$n \geq \frac{0.32}{8.10^{-6}} = 4.10^4 = 40\,000$$

On voit que l'on ait assuré d'obtenir une précision de 10^{-2}, si on utilise 40 000 gouttes d'eau, ce qui est considérable. Cette méthode de calcul, exige donc l'utilisation d'ordinateurs suffisamment rapides. Il est exclu de l'utiliser avec des calculettes.

On pourra utiliser la génération de nombres pseudo-aléatoires pour obtenir la distribution empirique de certaines variables aléatoires dont on ignore la distribution de probabilité. Mais l'emploi de cette méthode exige la répétition d'un grand nombre de simulations de la même expérience aléatoire et là aussi on devra utiliser des machines rapides. La vitesse de calcul des processeurs augmentant rapidement, on peut espérer pouvoir pratiquer cette méthode dans les lycées convenablement équipés.

On peut utiliser Mathematica pour illustrer la convergence en probabilités de $(4F_n)$ vers π.

```
u[1] = h3[{Random[Real,{0,1}],Random[Real,{0,1}]}] ;
u[n_] := u[n] = u[n-1] + h3[{Random[Real,{0,1}],Random[Real,{0,1}]}]
v[n_] := u[n]/n
Table[u[n],{n,1,3000}] ;
list16 = Table[4 v[n],{n,1,3000}] //N ;
g4 = ListPlot[list16, PlotStyle -> RGBColor[1,0,0], PlotJoined -> True]
g5 = Plot[{\[Pi] - 0.01,\[Pi] + O.O1},{x,0, 3000},
PlotStyle -> RGBColor[0,1,0]] ;
Show[g4,g5]
```



Cette technique d'intégration fait partie d'une famille de méthodes probabilistes, appelées Méthode de Monte Carlo, utilisées dans le calcul automatique sur ordinateur. Ces méthodes peuvent servir à calculer comme on vient de le voir des aires ou des intégrales. Elles ont été employées an physique nucléaire, pour la simulation d'un neutron dans un matériau fissile, dans la physique des corps poreux, en recherche opérationnelle dans les problèmes de trafic sur des réseaux ayant de nombreux noeuds et arcs, en chimie, pour des travaux concernant les très longs polymères.

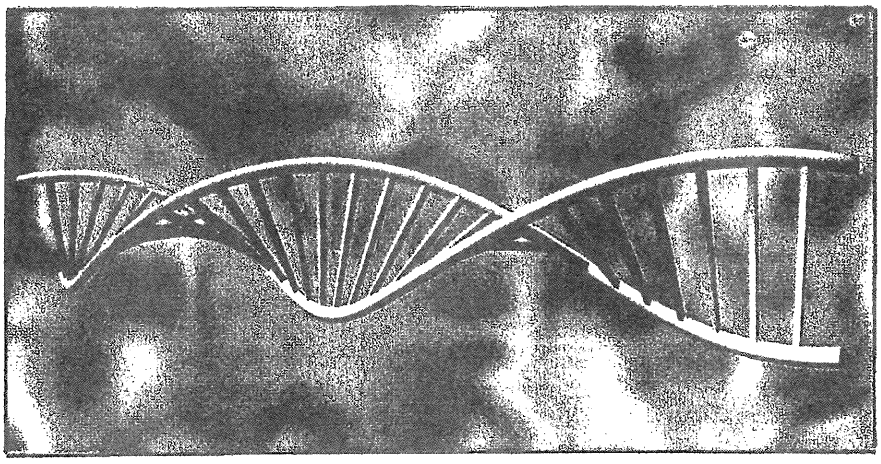
Bibliographie

R.B.J.T. ALLENBY, E.J. REFERN *Introduction to number theory with computing* Edward Arnold, a division of Holder & Stoughton (1989)

G. SAPORTA, *Probabilités, Analyse des données et statistique,* Editions Technip (1990)

M.HAUSNER, *Discrete Mathematics,* Saunders College Publishing, a Harcourt Brace Jovanovitch College Publisher (1989)

IV - DES REFERENCES POUR ALLER PLUS LOIN



BIBLIOGRAPHIE D'APPUI POUR UN ENSEIGNEMENT DES STATISTIQUES

Marie-Claire COMBES - Mireille SAUTER - Alain BERNARD

IREM de Montpellier

Résumé : *Beaucoup de publications ont vu le jour récemment, consacrées à l'enseignement des probabilités et des statistiques. Nous présentons d'abord les productions du réseau des IREM, à travers les articles de la revue nationale Repères-IREM et des publications des IREM des différentes Académies. On y trouvera à la fois des éléments de réflexion sur les nouveaux programmes et des propositions d'activités pour les classes. Nous donnons ensuite une présentation de la revue Tangente consacrée aux statistiques.*

1. Articles parus dans la revue REPERES-IREM

LABROUE F., 1991, *Statistique et technologie*. Repères-IREM n°2, pp. 30-50.

Résumé : *Cet article donne, à partir de situations concrètes de fabrication de matériel, des exemples d'activités permettant de pratiquer le programme de statistiques du collège à la classe de seconde.*

DUPERRET J.-C., 1992, *Mode, moyenne, médiane au collège : pourquoi ?* Repères-IREM n°6, pp. 5-19.

Résumé : *L'article met en évidence l'intérêt mathématique d'une introduction des statistiques au collège.*

ARNAUD R., 1992, *Peut-on commencer le 1er trimestre de seconde par les statistiques ?* Repères-IREM n°6, pp. 20-26.

Résumé : *Observer, mesurer, collecter les mesures, représenter graphiquement les données, synthétiser par des caractéristiques de position ou de dispersion : l'article développe les raisons d'introduire les activités mathématiques en seconde à partir des statistiques.*

HENRY A., HENRY M., 1992, *L'enseignement des probabilités dans le programme de première*. Repères-IREM n°6, pp. 27-52.

Résumé : *Ce sont les changements de programmes de 1991 qui ont incité à l'écriture de cet article organisé en deux parties : "épistémologie et didactique" et "expérimentation du nouveau programme en 1ère S".*

Dans la partie "épistémologie et didactique", il présente l'esprit du nouveau programme de Première en le replaçant dans la continuité des programmes de Collège et de Seconde et met en évidence l'option "fréquentiste" pour introduire la notion de probabilité, en opposition à la présentation traditionnelle dans le cadre de l'équiprobabilité, désignée par Pascal par les termes de "géométrie du hasard". S'appuyant sur le texte du programme : "l'objectif est d'entraîner les élèves à décrire quelques expériences aléatoires simples", il insiste sur la nécessité pour les élèves d'acquérir les concepts d'expérience aléatoire et d'événement. Une digression épistémologique aborde rapidement différentes approches du hasard. La description d'une expérience aléatoire associée à une série statistique (exemple qui sera repris dans l'expérimentation de la deuxième partie) débouche sur la notion d'événement et l'utilisation pratique du langage des ensembles pour modéliser les combinaisons logiques d'événements. La notion de probabilité est dégagée de manière très intuitive du modèle d'urne de Bernoulli auquel la description de l'expérience aléatoire avait conduit. Cette première partie souligne enfin les difficultés conceptuelles auxquelles se heurte l'objectif du programme d'introduire la probabilité d'un événement par l'observation de la fréquence de réalisation de cet événement quand on répète l'expérience un grand nombre de fois. Cependant, l'observation de la stabilité relative des

fréquences est une activité intéressante pour donner confiance aux élèves dans l'approche fréquentiste.

La partie "expérimentation du nouveau programme en 1ère S" relate 3 séances de travail de deux heures dans une classe et l'évaluation qui a suivi : test individuel pour faire émerger les conceptions spontanées sur des phénomènes aléatoires, exemples d'expériences aléatoires qui aboutissent à la définition intuitive de la probabilité des événements considérés, observation des fréquences d'un événement attaché à une expérience qu'on répète un grand nombre de fois.

PARZYSZ B., 1993, *Des statistiques aux probabilités, exploitons les arbres.* Repères-IREM n°10, pp. 91-104.

Résumé : *Dans le cadre des programmes de première, appliqués à partir de septembre1991, et comprenant un enseignement de probabilités alors nouveau à ce niveau du cursus scolaire, l'auteur étudie le problème de la modélisation en utilisant comme outils les arbres. Il montre bien la différence entre arbre de possibles et arbre probabilisé, puis aborde la loi binomiale et les probabilités conditionnelles.*

HENRY M., 1993, Paradoxes et lois de probabilité. Repères-IREM n°13, pp. 5-16.

Résumé : *A partir d'un jeu télévisé au cours duquel le candidat doit choisir "au hasard" la porte derrière laquelle se trouve le cadeau qu'il souhaite remporter, les auteurs sont amenés à modéliser la situation et faire un certain nombre de remarques didactiques, notamment à propos des paradoxes et des stratégies.*

HENRY M., 1994, *L'enseignement du calcul des probabilités dans le second degré. Perspectives historiques, épistémologiques et didactiques.* Repères-IREM n°14, pp. 69-104.

Résumé : *Cet article présente les enjeux de l'enseignement des probabilités en première et terminale, dans le cadre des programmes de 1991 et 1993. Après une brève description des objectifs du programme de première, il en présente une analyse didactique montrant l'importance de distinguer fréquence stabilisée et probabilité. Il repose la question de l'introduction de cette notion à partir des connaissances des élèves, notamment du lien entre pourcentage et probabilité, et débouche sur la nécessité d'un travail de fond sur la notion d'expérience aléatoire et de sa modélisation.*

Dans une deuxième partie, l'article retrace les événements essentiels de l'histoire des probabilités, pour montrer comment la dualité "géométrie du hasard" - "définition fréquentiste" a traversé cette histoire et se retrouve naturellement au centre de la problématique de leur enseignement.

Dans une troisième partie, l'article montre ce que pourrait être un enseignement de la modélisation des notions d'expériences aléatoire et de probabilité, permettant le calcul a priori sous l'hypothèse d'équiprobabilité aussi bien que l'estimation expérimentale des probabilités en jeu. La notion de probabilité conditionnelle est enfin présentée de manière cohérente avec l'option fréquentiste et utilisée pour éclairer un problème paradoxal : le jeu des deux chèvres et de la voiture.

DUPERRET J.-C., 1995, *L'apprenti fréquentiste.* Repères-IREM n°21, pp. 95-108.

Résumé : *L'auteur propose un protocole d'observation des nouveaux programmes, permettant d'en corriger plus rapidement certains effets pervers (mesure de la fracture constituée par les nouveaux programmes, réduction de la fracture en s'appuyant sur la formation continue, analyse à long terme de l'effet des changements). Il analyse dans un deuxième temps la fracture à l'aide d'une parabole, illustrant les conflits entre le pôle « géomètre » (reposant sur la rigueur et le maniement d'un arsenal complet de théorèmes) et le pôle « fréquentiste », reposant sur l'observation, la recherche de régularités et de modèles, l'apprivoisement du hasard.*

ROUSSET-BERT S., DUPUIS C., 1996, *Arbres et tableaux de probabilité : analyse en termes de registres de représentation.* Repères-IREM n°22, pp. 51-71.

Résumé : *L'article recense les différents registres de représentation sémiotique utilisés dans les manuels scolaires pour l'enseignement des probabilités : les arbres, les tableaux, les écritures symboliques et la langue naturelle. Il analyse plus particulièrement ensuite les codes de construction et de lecture des arbres et insiste sur la difficulté, et la nécessité, de passer d'un registre à un autre pour construire le sens des objets mathématiques que l'on veut enseigner.*

GIRARD J.-C., 1996, *Pourquoi il ne faut pas laisser de côté les chapitres de statistiques au collège.* Repères-IREM n°23, pp. 5-18.

Résumé : *L'article part d'un constat : l'enseignement des statistiques est systématiquement négligé en collège. Après avoir évoqué des explications possibles (les professeurs maîtrisent moins cette partie du programme et surtout sont peu convaincus de son intérêt), il plaide pour un renouvellement de cet enseignement, avec trois arguments : c'est l'occasion de faire des graphiques réfléchis, c'est l'occasion de faire des calculs motivés, c'est l'occasion d'introduire une réflexion sur le hasard et la variabilité des résultats.*

Le n° 32 de la revue Repères est spécialement consacré aux probabilités et à la statistique.

PICHARD J.-F., 1998, *Approche épistémologique et diverses conceptions de la probabilité.* Repères-IREM n°32, pp. 5-24.

Résumé de l'auteur : *J'indique dans cet article quelques-unes des conceptions et interprétations de la probabilité, et leur place dans l'évolution historique de la théorie. La diversité d'interprétations est la source des difficultés d'une définition de la probabilité lorsqu'on sort du cadre axiomatique et que l'on veut utiliser cette théorie pour modéliser la réalité.*

La probabilité est la notion de base de la théorie probabiliste. On comprendra mieux ses diverses interprétations en la replaçant dans son évolution historique et dans le cadre général probabiliste et statistique, dont une esquisse est donnée dans la frise historique qui figure à la suite de l'article.

GANDIT M., HELMSTETTER C., 1998, *Expérimenter, simuler en classe.* Repères-IREM n°32, pp. 25-42.

Résumé :*Les auteurs évoquent les objectifs de l'enseignement des probabilités et les difficultés rencontrées. Ils proposent une démarche pour la classe (réalisation de l'expérience, simulation par différentes techniques (table de nombres au hasard, calculatrice ou ordinateur, modélisation, lien entre résultats empiriques et théorie), à partir de deux exemples d'activités.*

RAYMONDAUD H., HENRY M., 1998, *Attention ! Un modèle peut en cacher un autre.* Repères-IREM n°32, pp. 43-62.

Résumé :*L'article critique vivement les avatars de certains sujets de Bac, habillés par des situations pseudo-concrètes. Il se termine par un cri d'alarme « Auteurs de sujets d'exercices : suivez le contenu et l'esprit des programmes, par respect pour les élèves et par pitié pour leurs enseignants... ».*

CHRETIEN C., GAUD D., 1998, *Qu'est-ce que le hasard ? Comment le mathématiser ?* Repères-IREM n°32, pp. 81-110.

Résumé :*L'article présente une réflexion épistémologique sur le « hasard », appuyé sur l'histoire des idées. Cette réflexion débouche sur la proposition d'activités pour les classes (exemple de situations, choix de textes mathématiques et/ou philosophiques.*

BUHLER M., 1998, *Un problème de dés en terminale.* Repères-IREM n°32, pp. 111-120.

Résumé :*L'auteur présente sept méthodes expérimentées en classe pour traiter le même problème de dé.*

GIRARD J.-C., 1998, *A bas la moyenne ! A propos des paramètres de tendance centrale et de dispersion d'une série statistique*. Repères-IREM n°33, pp. 97-114.

Résumé :*L'auteur propose plusieurs interprétations de la moyenne et plusieurs moyennes possibles « classiques » ou moins classiques, pour définir la tendance centrale d'une série statistique.*

GIRARD J.-C., 1999, *Le professeur de mathématiques doit-il enseigner la modélisation ?* Repères-IREM n°36, pp. 7-14.

Résumé : *L'auteur pose la question du rapport entre les mathématiques et la réalité. "Doit-on entraîner nos élèves à critiquer les résultats que donnent les mathématiques, à se poser la question de la pertinence de certaines applications des mathématiques et pas seulement les entraîner à exécuter docilement certains programmes de calculs dans des situations construites pour que l'on puisse les faire ?... Doit-il enseigner la modélisation ou seulement des modèles ?".*

Après avoir défini ce qu'il entend par modèle, il montre quelques exemples "où l'on fait de la modélisation sans le dire dans l'enseignement", puis quelques exemples d'essais de modélisation dans l'enseignement primaire et secondaire.

HUGUET F., 1999, *Simulation et modélisation d'un problème de bouteilles*. Repères-IREM n°36, pp. 35-42.

Résumé : *L'auteur décrit et analyse une activité dans des classes. Cette activité est basée sur une situation concernant la fabrication de bouteilles en verre à partir de verre liquide. Ce type d'activités est abordé dans les pays anglo-saxons dès l'école élémentaire.*

L'auteur, reprenant des idées de Engel, a expérimenté, en France, avec des étudiants. Il décrit les méthodes de simulation utilisées avec ses étudiants, les résultats de ces expérimentations, puis développe deux approches de modélisation du problème, puis des remarques sur ces approches. Il termine en rappelant une activité proposée à des enseignants qui montrait que "le hasard est composé de certains paramètres que l'on maîtrise et d'autres que l'on ne maîtrise pas".

HENRY M., 1999, *L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie*. Repères-IREM n°36, pp. 15-34.

Résumé : *La première partie constitue un rappel historique dans lequel l'auteur décrit les différentes conceptions des probabilités. La conception laplacienne des probabilités qui fut la base de l'introduction des probabilités dans l'enseignement dans les années soixante, cette conception est liée à la combinatoire : "Les connaissances théoriques de l'époque ne permettaient pas de concevoir une autre approche théorique que celle de géométrie du hasard". Puis la conception fréquentiste, qui fut en particulier développée par Renyi dans un manuel universitaire qui a fait référence à partir des années soixante, qui est l'approche choisie dans les programmes de lycée depuis 1991, et qui s'applique aux expériences aléatoires que l'on peut répéter un assez grand nombre de fois pour pouvoir estimer les probabilités en jeu. Cette conception s'appuie sur la définition de Kolmogorov, c'est-à-dire une mesure abstraite qui modélise les issues possibles d'une expérience aléatoire. Cependant cette approche semble exclure les situations non reproductibles. L'auteur montre l'analogie, du point de vue épistémologique, entre les probabilités et la géométrie, toutes deux issues de la pratique sociale.*

Dans la deuxième et la troisième partie, l'auteur expose les choix des programmes actuels de l'enseignement secondaire (depuis 1991) et leurs conséquences didactiques. Ce domaine de modélisation devant s'appuyer sur l'appréhension des situations aléatoires, l'auteur préconise un enseignement qui commencerait au moins dès le début du collège.

Dans la quatrième partie il définit ce qu'il entend par "modèle" et "modélisation" en s'appuyant sur des exemples et en décrivant ce que sont ou peuvent être des séquences d'enseignement.

La cinquième partie est un parallèle entre l'enseignement des probabilités et celui de la géométrie, du point de vue des fondations, de la modélisation, du regard théorique, de la formalisation mathématique et de la réponse au problème posé.

Enfin dans la sixième partie, l'auteur s'intéresse aux objectifs de l'enseignement des probabilités et en dégage des perspectives didactiques.

LEFORT J., 1999, *Fiabilité du lanceur Ariane.* Repères-IREM n°36, pp. 77-83.

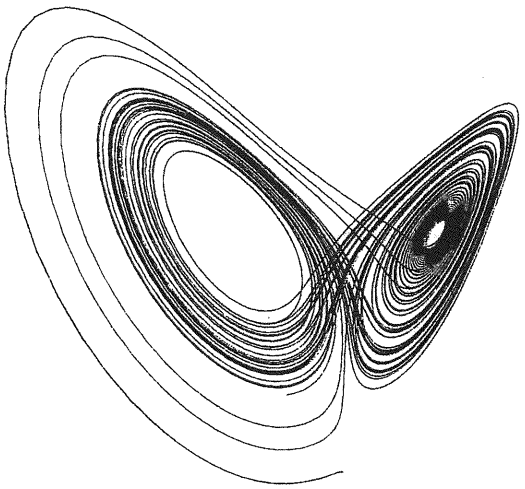
Résumé : *Intrigué par des articles de presse donnant un taux de réussite de 96,2% aux lancements de la fusée Ariane, et ceci après le centième lancement, l'auteur a cherché plus d'informations et s'est intéressé à la façon d'évaluer le taux de réussite d'un prochain lancement.*

BOURGUET M., 2000, *Roulette russe et très basses fréquences aux bacs de techniciens.* Repères-IREM n°38, pp. 80-83.

Résumé : *L'article critique un sujet de mathématiques proposé en juin 1999 aux sections Génie Electrique et Génie Electronique du baccalauréat de techniciens : problème de probabilités s'appuyant sur une situation d'électricité : exemple de "mise en scène pseudo-concrète" qui ne conduit qu'à des absurdités. On propose en effet aux candidats d'imaginer des branchements au hasard, au mépris des plus élémentaires règles de sécurité et de bon sens, et d'étudier la variable aléatoire "vitesse de rotation du moteur"...*

GIRARD J.-C., HENRY M., PICHARD J.-F., 2001, *Quelle place pour l'aléatoire au collège ?* Repères-IREM n°42, pp.27-43.

Résumé : *Cet article constitue une sensibilisation aux enjeux épistémologiques et didactiques de l'introduction de l'aléatoire au collège, en vue de préparer un enseignement des statistiques et des probabilités qui puisse reposer sur une approche expérimentale des notions de base. A l'intention des enseignants en collège et en classe de seconde, il présente un exemple historique concernant la liaison entre observation statistique et probabilités: le jeu de Franc-Carreau (cf. aussi dans cette brochure l'article page 133).*



Lorenz attractor

2. Brochures des IREM

On trouvera, dans Repères-IREM n°32 (juillet 1998), une bibliographie générale sur l'enseignement des probabilités et des statistiques. Parmi celle-ci, nous avons distingué en particulier certaines brochures IREM.

BLOCH C., BOUZAT M., DUCHENNE O., GUICHARD J-P., THIENARD J.-C., 1994, *Thèmes pour l'enseignement de la statistique et des probabilités*. IREM de Poitiers.

Résumé : *Cette brochure traite du calcul des probabilités et des statistiques sous un aspect essentiellement culturel. Elle comprend quatre parties indépendantes.*

1. Bref historique de la notion de probabilités.

Cette partie a pour but de donner quelques repères sur le développement du calcul des probabilités en mettant en évidence :

- *les problèmes qui en ont été à l'origine,*
- *les problématiques qui ont motivé son développement,*
- *la dialectique très féconde qui unit le calcul des probabilités et les sciences sociales à leurs débuts,*
- *elle aborde également les étapes qui ont conduit à l'intégration de la pensée et des méthodes probabilistes à la pensée scientifique globale.*

2. Naissance de la sociologie et pensée statistique.

L'objectif principal de cette partie réside dans la mise en évidence du fait que la sociologie n'a pu naître que parce qu'avait été développé préalablement un concept de vérité statistique et des procédures pour son utilisation, comme le montre très clairement les extraits des textes de Durkheim proposés à l'étude.

3. Le loto.

Cette partie a pour but de faire appréhender un phénomène social de grande ampleur : le loto. Quelques éléments simples de calcul de probabilité permettent à chacun de réfléchir à ses chances de gain, aux stratégies de jeux proposées et de s'interroger sur la règle de distribution des gains.

Elle débouche sur une réflexion sur les caractéristiques de la pensée rationnelle, scientifique opposée aux types de pensée mythique, magico-religieux...

4. Statistique, probabilités et problèmes de société.

L'objet de cette partie est :

a) de mettre en évidence et d'illustrer la prégnance de la pensée probabiliste dans la pensée contemporaine ;

b) de faire mesurer l'intérêt du concept de vérité statistique, en particulier par rapport aux problèmes d'opinion ;

c) de faire prendre connaissance des appareils de mesure statistiques dont nos sociétés se sont dotées pour se connaître : démographiquement, économiquement, sociologiquement, politiquement ;

d) de faire fonctionner sur des exemples variés quelques concepts de probabilités et de statistiques.

IREM de Bourgogne, 2000, *Statistiques en seconde... et un peu après*. IREM de Poitiers.

Résumé : *la brochure a trois objectifs :*

1. Informer sur la statistique inférentielle, par une approche expérimentale pouvant déboucher sur des activités à proposer aux élèves (en seconde, première et terminale) et par une approche théorique de l'estimation ;

2. Donner du sens aux statistiques, aux représentations en usage et aux caractéristiques de position et de dispersion ;

3. Proposer des activités mathématiques à partir de situations statistiques.

Elle est divsée en deux grandes parties :

- *la première partie propose des activités, en distinguant ce qui peut être proposé aux élèves et ce qui relève de l'information du professeur ;*

- *la deuxième partie, à destination du professeur, propose des compléments théoriques sur l'estimation et les tests d'hypothèse.*

Commission Inter-IREM « Lycées techniques » avec le soutien de la DESCO, 2000, *Simulation et statistiques en seconde*. IREM Paris-Nord.

Présentation par les auteurs : *L'objectif de cette brochure est de fournir aux enseignants des Documents-Ressource à propos des différents thèmes d'étude de la partie statistique du programme de seconde (rentrée 2000).*

Ces documents sont de quatre types :

1. Des compléments de Probabilités et Statistique inférentielle permettant d'avoir le recul théorique nécessaire, par rapport aux activités proposées aux élèves. On y reste à un niveau simple, renvoyant le lecteur désireux d'approfondir, à des références d'ouvrages universitaires.

2. Des travaux pratiques avec calculatrice, fondés sur des programmes de simulation, donnés pour les marques CASIO et TEXAS INSTRUMENTS. Ils sont prévus pour une demi-classe, voire en petits groupes, où plusieurs thèmes peuvent être traités simultanément, avec une relative autonomie.

3. Des travaux pratiques sur EXEL, permettant de profiter des possibilités de calcul et d'illustration de l'ordinateur, tout en offrant une initiation au tableur de Microsoft. Entièrement guidés, les élèves commencent avec un classeur vierge Excel (95 - 97 ou 2000). Ces T.P. sont prévus pour être traités en salle informatique, au maximum à deux élèves par poste.

4. Des développements sous EXEL : programmes "prêts à l'emploi", disponibles sur Excel 97 ou 2000 et ne nécessitant aucune connaissance du logiciel. Il s'agit d'illustrations, soit à utiliser en cours (où le professeur est le manipulateur), soit en réseau (où chaque élève expérimente à son gré).

Le CD-ROM qui accompagne cette brochure, lisible sur PC, contient, outre les développements sous Excel 97 ou 2000, les énoncés des travaux pratiques (calculatrices et tableur), lisibles sous Word 97 ou supérieur. Il vous sera possible d'adapter ces T.P. à vos besoins ou les faire évoluer.

Pour plus d'informations, consulter le site :
http://www-irem.univParis13.fr/dossierpublications/DesPages/

Commission Inter-IREM « Lycées techniques », 2000, *Simulation d'expériences aléatoires. Une expérimentation du hasard de la première au B.T.S. sur calculatrice et ordinateur.* IREM de Paris-Nord.

Résumé : *Les instructions officielles insistent sur la nature expérimentale qu'il est souhaitable de donner à l'enseignement des statistiques et probabilités.*
Dans cet objectif, il s'agit de lier la notion de probabilité à celle de fréquence observée après la répétition un grand nombre de fois d'une expérience. La simulation permet, à peu de frais, d'y parvenir, donnant ainsi plus de sens aux concepts théoriques.
1. Cette brochure s'organise autour :
** d'activités sur calculatrices proposées sous forme de fiches de travaux pratiques photocopiables, contenant des programmes à saisir par les élèves,*
** de développements sur Excel, conçus pour être des produits finis, directement utilisables pour illustrer la leçon.*
2. Deux stratégies sont utilisées selon les objectifs visés et les contraintes matérielles : "Le hasard en direct" et "la simulation globale".

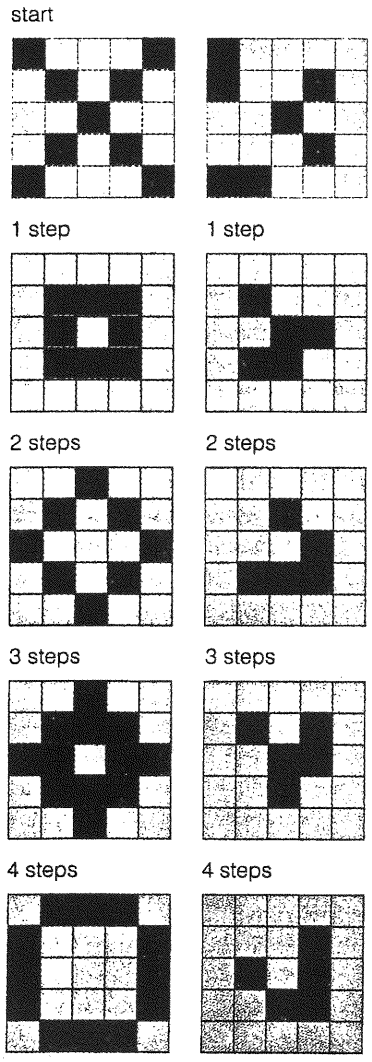
Pour plus d'informations, consulter le site :
http://www-irem.univParis13.fr/dossierpublications/DesPages/

Commission Inter-IREM « Statistique et probabilités », 1997, *Enseigner les probabilités au lycée, ouvertures statistiques, enjeux épistémologiques, questions didactiques et idées d'activités.*

Présentation des auteurs : *Cet ouvrage tente de répondre à quelques questions de fond que les enseignants rencontrent quand ils désirent que les objets probabilistes introduits (expérience aléatoire, événement, probabilité, variable aléatoire, indépendance) prennent du sens pour leurs élèves.*

Nous avons eu aussi le souci de placer les connaissances probabilistes naissantes dans leur perspective historique, plusieurs articles font intervenir les acteurs fondateurs avec quelques problèmes que l'histoire a immortalisés.

Destiné aux professeurs de mathématiques, ce livre se devait aussi de proposer des idées d'activités, de TP ou d'exemples, pour couvrir le champ des programmes des premières et des terminales. Quelques prolongements font entrevoir l'horizon qui servira de cadre probabiliste dans les classes de techniciens supérieurs ou à l'université. De ce point de vue, les débouchés vers la statistique n'ont pas été oubliés.



3. Revue Tangente : « Faut-il se fier aux statistiques ? »

Il existe trois types de mensonges :
les mensonges simples, les sacrés mensonges
et les statistiques.
Mark TWAIN(page17)

Un chapitre nouveau du programme de seconde fait beaucoup parler de lui : la statistique ou les statistiques comme vous voudrez. La revue " Tangente "[1], toujours proche de l'actualité, propose un numéro " complet pour comprendre les statistiques ", et j'ajouterai prendre un peu de hauteur ou un peu de distance. La culture statistique manque beaucoup dans notre pays où la certitude fondée sur une démonstration ne laisse que peu de place à une branche des mathématiques établie sur des mesures incertaines, des intervalles de confiance à 95% et des seuils de risques.

En un numéro courageux, " Tangente " tente - et réussit assez bien - de nous montrer l'importance des statistiques dans les sciences et dans la société. Une « brève histoire des outils statistiques » ouvre le numéro. On y découvre que les statistiques sont à la fois une science très ancienne car les recensements sont vieux comme les nombres et très récente, encore plus récente que les probabilités. La moyenne arithmétique est attribuée à l'astronome Tycho BRAHE au $16^{ème}$ siècle et l'usage d'une moyenne pondérée à COTES en 1722. Le français DUTOT ayant inventé un « indice des prix » en 1738.

Puis, un article traite du danger de classer les hôpitaux à partir de pourcentages et de moyennes. Comment peut-on classer des hôpitaux au dixième de pour-cent près ? Il y a bien là quelque aberration et les programmes de seconde nous demandent bel et bien de faire réfléchir les élèves sur l'usage que l'on fait des nombres en statistiques.

La moyenne est encore évoquée par Francis CASIRO dans « les pièges de la moyenne », cet estimateur de la « vraie » valeur. Mais au fait, quelle moyenne : harmonique, géométrique, arithmétique, quadratique ou toute autre moyenne abstraite ? On doit pouvoir tirer de cet article un thème d'étude car il y a beaucoup de moyennes ce qui est d'autant plus curieux que les statisticiens considèrent ,ou du moins considéraient, la moyenne comme la « vraie » valeur de la série.

L'article « Du recensement au sondage » de Marie-josé PESTEL nous mène des soucis de recensement des Incas aux difficultés des Américains avec les sondages lors des élections présidentielles de 1936. Déjà ! Naissance du célèbre Institut Gallup et de la très efficace méthode des quotas jusqu'à ce que .... Car représenter des dizaines de millions d'électeurs par un échantillon de 1000 personnes ne sera jamais une science exacte même si l'intervalle de confiance ne dépend que de la taille de l'échantillon et pas de celui de la population.

Un bref article sur « les méandres de la théorie de l'estimation » par Daniel TEMAM doit être là pour rassurer, mais il est trop court car très théorique et on ne peut entrer dans le sujet qui mérite plus de développement ce qui n'est pas le sujet de la revue.

Nous devons reprendre notre passionnant parcours pour retrouver MENDEL qui nous attend avec ses petits pois tout ridés et de bien surprenants résultats. Et encore parcourir deux articles sur les statistiques dans les assurances-vie dont les primes sont bien sûr calculées sur les tables de... mortalité.

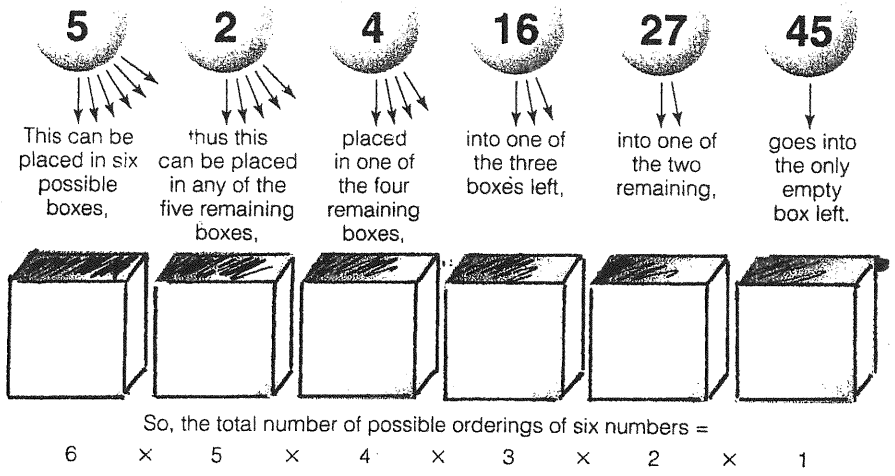
Enfin, ouvrez votre boîte à lettres et vous y découvrirez ... la loi de BENFORD ! " Tangente ", là, tu exagères ! Le physicien américain Franck BENFORD reprit au début du siècle une remarque de Simon NEWCOMB concernant le fait que les tables de logarithmes étaient beaucoup

[1] Revue Tangente : Editions Archimède 5, rue Jean Grandel 951000 ARGENTEUIL.

plus usées et défraîchies au début vers les chiffres 1 et 2 que vers la fin. BENFORD confirme la loi donnée empiriquement par NEWCOMB et accumule 20 000 observations pour en avoir le coeur net. Aujourd'hui, la loi de BENFORD est en train d'acquérir une célébrité certaine en France.

En ajoutant quelques jeux et énigmes, " Tangente ", en 50 pages, propose une foule de petites histoires et de pistes utilisables avec les élèves et surtout pas trop scolaires. Car ne l'oubliez pas, une des raisons des statistiques en seconde est de nous rapprocher de la réalité et de la vie alors « Tangente » est une excellente source d'anecdotes et de jeux pour ce faire.

Si vous en doutiez encore, les auteurs montrent bien qu'il y a beaucoup de maths en statistiques et beaucoup de statistiques dans notre société, donc par ... transitivité (c'est quoi ?) beaucoup de maths dans notre vie.



DES RESSOURCES SUR L'INTERNET POUR ENSEIGNER LA STATISTIQUE

Michel JANVIER
Equipe Informatique
IREM de Montpellier

Résumé : *Où trouver des ressources pour l'enseignement de la statistique ? Bien sur dans les bibliothèques. Mais un enseignement moderne de cette discipline suppose le recours à des simulations. Pour présenter des simulations à ces élèves, à moins d'apprendre à programmer en Java ou en Mathematica, le professeur de mathématiques peut faire appel aux très nombreuses applets Java que l'on trouve sur Internet. Il faudra également se procurer des données brutes. Nous indiquons une liste de sites où l'on en trouve et nous expliquons comment les importer pour les traiter par exemple dans un tableur. Mais les tableurs n'ont pas été conçus à l'origine pour faire des analyses statistiques. Il faut donc parfois ruser avec eux pour obtenir ce que l'on veut ou les compléter avec des additifs. Or on trouve sur le réseau d'excellent logiciels de calcul statistiques gratuits. Certains sont des clones de logiciels commerciaux, d'autres sont très conviviaux et peuvent facilement être utilisés par les élèves. On trouve également sur Internet des cours de statistiques en ligne. Enfin pour ceux qui ne seraient pas satisfaits de notre marché sur le réseau, nous indiquons des listes de liens, constituées par des sociétés savantes, des organismes commerciaux ou des universitaires.*

L'introduction d'un enseignement de la statistique appliquée dans l'enseignement secondaire inquiète les enseignants qui estiment ne pas avoir été formé à cette discipline lors de leurs études universitaires. Ils ont appris un peu de calcul des probabilités, comme application de la théorie de la mesure, à l'occasion d'un certificat de calcul intégral par exemple. Mais peu d'entre eux ont appris la statistique descriptive, supposée sans doute être une science infuse, et la plupart n'ont aucune notion de statistique inférentielle. Comment est réalisé un sondage d'opinion, quelles sont les techniques statistiques utilisées en contrôle de qualité, tout cela est superbement ignoré dans le cursus des futurs professeurs de mathématiques. Pourtant dans l'apprentissage de la citoyenneté il faudrait apprendre aux électeurs à juger de la confiance qu'ils doivent accorder aux sondages. Les élèves des sections de techniciens supérieurs devront être initiés aux techniques de contrôle de qualité, par des professeurs qui ne les ont jamais pratiquées.

Il existe un obstacle culturel peut-être plus grave encore que ces manques de connaissances pratiques. Un professeur devrait pouvoir acquérir ces connaissances par lui-même, mais la formation d'un mathématicien, par sa rigueur et sa propension à l'axiomatique, le prédispose à une conception déterministe du monde. Pour enseigner les statistiques appliquées et le calcul des probabilités, il lui faudra intégrer à sa réflexion les concepts de variabilité des résultats, de marges d'erreur, de risques d'erreur, de hasard.

La statistique grâce au développement des logiciels de calcul n'est plus le domaine où l'on faisait des calculs rébarbatifs. En contre partie, il faudra savoir programmer des simulations et des animations. Il faudra savoir interpréter les résultats obtenus et choisir la méthode de calcul et les représentations graphiques adaptées à la situation étudiée. Souvent, il

faudra travailler avec le chercheur ou le professionnel qui a fourni les données. La statistique oblige à l'interdisciplinarité.

Où trouver les données à traiter ? Quels logiciels de calcul et de représentations graphiques utiliser ? Faut-il s'initier à la programmation de logiciels professionnels, souvent compliqués et très onéreux ? Où trouver la documentation indispensable, ne serait-ce que les tables des distributions usuelles de probabilité ? Où trouver des cours pour combler les lacunes de sa formation ?

Nous ne prétendons pas donner des réponses à toutes ces questions. Simplement nous voudrions montrer ici que le réseau Internet offre des possibilités. Il devrait permettre de mettre en place une nouvelle forme d'enseignement intégrant les nouvelles technologies. On y trouve des bases de données, que l'on peut télécharger pour les traiter ensuite sur un tableur. On peut récupérer des logiciels gratuits. Certains sont très simples d'emploi et peuvent être utilisés en classe sans nécessiter un apprentissage lourd. On y trouve aussi des logiciels de niveau professionnel, pour des calculs plus spécialisés. On peut même faire traiter ses données en ligne par un serveur éloigné. On peut intégrer dans des pages de sites web, des applets Java permettant de faire des calculs. On trouve également des cours de statistique complets, plus ou moins interactifs, fournis par exemple par des fabricants de logiciels professionnels ou par des organismes d'enseignement à distance, comme le centre régional associé du Languedoc-Roussillon du Conservatoire National des Arts et Métiers.

Une grande partie des adresses signalées m'ont été indiquées par G.Saporta et se trouvent dans ses publications (G.Saporta, 2000).

1. Les logiciels de calcul gratuits

- **Statsoft**

Statsoft est une société américaine qui commercialise un logiciel très employé par les professionnels : Statistica. On peut télécharger gratuitement le module de statistiques élémentaires de Statistica, en français, à l'adresse suivante http://www.statsoft.com/french/download.fr. Ce module est largement suffisant pour couvrir les besoins des élèves de lycée. Cela prendra une vingtaine de minutes à 28.8 Kbps pour effectuer le téléchargement mais on disposera alors d'une vaste gamme de fichiers d'exemples pour réaliser des analyses et créer différents graphiques. Mais on ne pourra pas imprimer les graphiques obtenus.

Surtout on disposera *d'un manuel de statistique* particulièrement intéressant. On peut télécharger ce manuel directement à l'adresse suivante http://www.statsoft.com/textbook/stathome.html . On y trouve des présentations de l'analyse de la variance, de l'analyse des correspondances, de l'analyse factorielle, des cartes de contrôle de qualité, ... mais dans le chapitre sur les concepts élémentaires, dans l'introduction, les néophytes apprendront ce que sont les variables, les variables indépendantes, les relations entre variables, le niveau de signification et le rôle de la distribution normale. Dans la partie concernant les statistiques de base, ils trouveront des explications claires sur la moyenne et les intervalles de confiance, sur les corrélations, les tableaux de contingence et les tests d'hypothèse.

Voici un exemple qui est cité dans ce manuel et qui illustre bien la démarche très pragmatique employée. « Il s'agit du rapport entre le nombre de bébés de sexe féminin et le nombre de bébés de sexe masculin. Considérons l'exemple suivant tiré de recherches sur le raisonnement statistique (Nisbett, et al., 1987). Il y a deux hôpitaux : dans le premier 120 bébés naissent chaque jour, et dans l'autre seulement 12. En moyenne le rapport du nombre de bébés de sexe féminin et du nombre de bébés de sexe masculin est proche de 1. Cependant, un jour, dans l'un de ces hôpitaux il est né deux fois plus de filles

que de garçons. Dans quel hôpital était-il plus vraisemblable que cela se soit produit ? La réponse est évidente pour un statisticien, mais pas pour une personne non avertie. Il est plus vraisemblable que cela se soit produit dans le plus petit des hôpitaux. La raison en est, pour parler d'une manière technique, que l'écart type d'une variable aléatoire égale à une proportion calculée sur un échantillon tiré d'une population mère, décroît lorsque la taille de l'échantillon augmente.

Quand un rapport est proche de 1, dans un échantillon, ce résultat ne pourra être considéré comme significatif que sur un grand échantillon. Par contre quand un rapport est élevé dans la population, il suffira d'un échantillon assez petit pour obtenir un résultat significatif. Considérons l'exemple suivant : si une pièce est légèrement asymétrique et donnera face plus fréquemment que pile, par exemple dans 60% des cas, alors 10 lancers ne seront pas suffisants pour convaincre quiconque que la pièce n'est pas symétrique. Est-ce que 10 lancers sont insuffisants pour prouver quelque chose ? Non. Imaginons que la pièce soit suffisamment asymétrique pour que, quelle que soit la manière dont elle est lancée, elle donne toujours face. Alors si on lance dix fois la pièce et qu'à chaque fois elle donne face, la majorité des gens sera persuadée qu'il y quelque chose d'anormal concernant cette pièce. Autrement dit, on pourra considérer qu'il est très vraisemblable que dans une population d'une infinité de lancers on obtiendra un plus grand nombre de fois face que pile. Un petit échantillon permet ici de conclure. »

Ce manuel comporte également un *glossaire* de tous les termes statistiques usuels, une *bibliographie* très importante, et aussi une *table des distributions usuelles* des lois de probabilité. Ces tables peuvent encore servir, bien que les logiciels permettent de calculer les valeurs des fonctions de répartition des lois usuelles.

- **Vista**

Vista, "the visual statistics system", est un logiciel gratuit créé par Forest Young, professeur de psychométrie à l'Université de Caroline du Nord. On peut le télécharger à l'adresse suivante : http://cons-dev.univ-lyon1.fr/Enseignement/Vista/VistaF.html

On trouvera, incorporées au logiciel, un grand nombre de données classées selon le traitement à leur faire subir. Naturellement on peut également traiter ses propres données qui doivent être saisies comme dans un tableur. L'utilisation de Vista est très intuitive.

Voici un exemple qui concerne une étude de l'effet des couloirs pour deux roues sur les conducteurs de voitures et les conducteurs de deux roues. Ce fichier comporte deux variables, « Travelspace », la distance entre les couloirs pour deux roues et la ligne médiane de la route, et « Separation », la distance entre un cycliste et une voiture le doublant. On obtient immédiatement les graphiques suivants : nuage de points, boîtes de dispersion, représentation des quantiles, histogrammes. On peut obtenir des explications sur la signification des graphiques obtenus : « La représentation des quantiles permet de décider de la normalité des distributions des variables. Dans ces graphiques la ligne brisée représente la distribution de la variable et la ligne droite la distribution normale. Si la ligne brisée est grossièrement linéaire, c'est à dire si elle suit approximativement la ligne droite, la distribution de la variable peut être approchée par une distribution normale. » De manière plus savante les QQ-Plots ou ajustement par quantiles servent à contrôler l'ajustement d'un échantillon avec une loi de probabilité, choisie comme modèle. On représente les points de coordonnées $(x_{(i)}, Q(i/n))$, où les $x_{(i)}$ sont les statistiques d'ordre de l'échantillon et Q la fonction quantile théorique. Si le modèle est bon, les points doivent être proches de la première bissectrice.

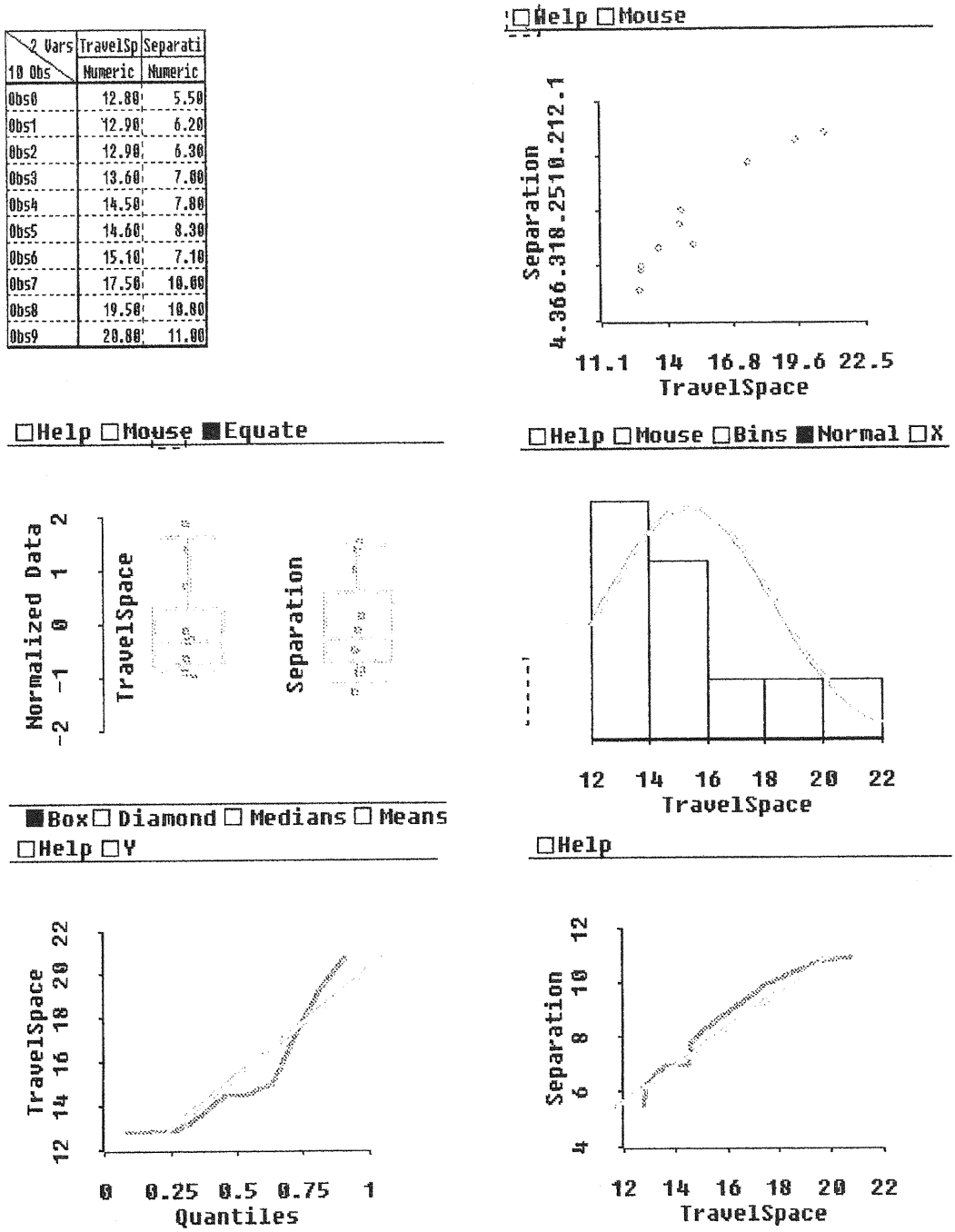


Figure 1 : Etude de la distance entre les véhicules à deux roues circulant sur un couloir aménagé et les véhicules à quatre roues

Le logiciel est disponible en anglais, en français et en espagnol. Il contient la statistique élémentaire mais aussi de l'analyse multivariée, avec des sujets de pointe comme la régression robuste, les biplots, l'analyse des correspondances, etc.

- **R**

Les utilisateurs plus avancés pourront télécharger un clone gratuit de S, logiciel professionnel très répandu. R est du à Ross Ihaka et Robert Gentleman, connus comme R et R, du département de statistique de l'université d'Auckland.

Le téléchargement se fait à partir de http://www.ci.tuwien.ac.at/R/ .

On peut alors obtenir des graphiques comme des boites à moustaches avec entailles qui représentent des intervalles de confiance à 95% pour la médiane. La longueur des moustaches est 1,5 fois l'intervalle interquartile. Dans le cas où les données peuvent être modélisées en utilisant une loi de Gauss, les extrémités des moustaches, au niveau théorique, sont voisines du premier et du 99° décile. Ces diagrammes permettent alors de détecter les données exceptionnelles. On peut comparer plusieurs groupes et obtenir des représentations du type de celle qui est donnée ci-dessous.

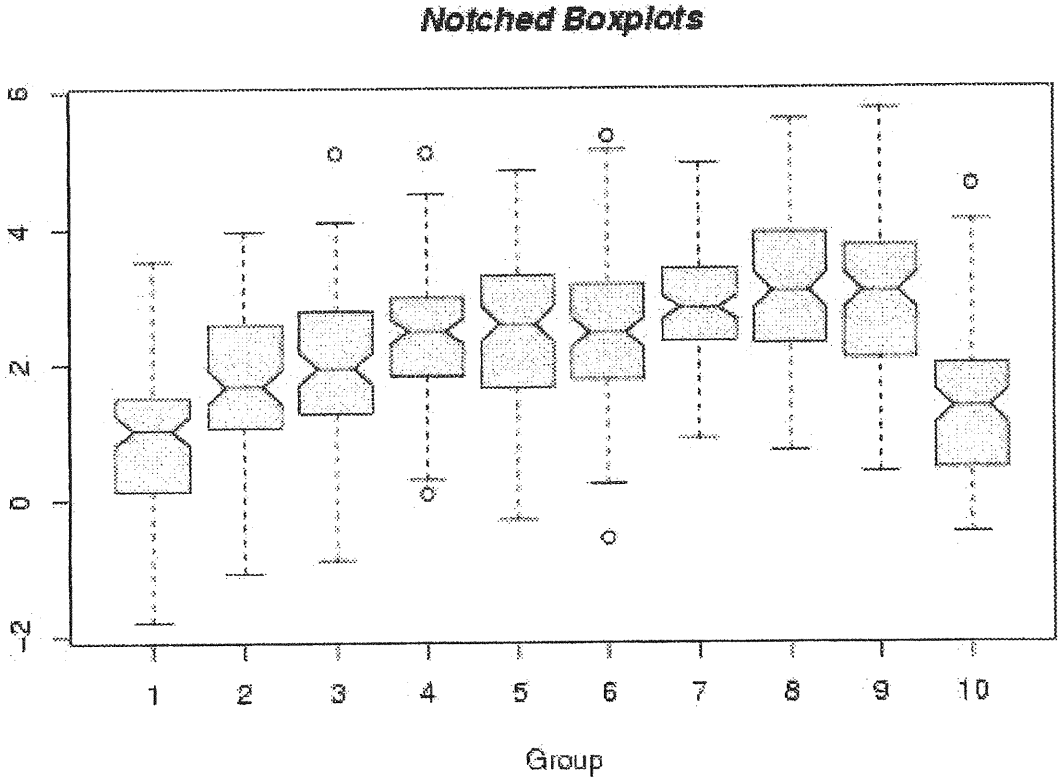


Figure 2 : Boîtes à moustaches à entailles obtenues avec R

A partir d'une étude comportant quatre variables, la longueur et la largeur des sépales, la longueur et la largeur des pétales de 150 iris, on obtient les 12 nuages de points représentant les couples de variables, ce qui permet d'étudier la corrélation entre ces variables. On peut évidemment obtenir la matrice des corrélations.

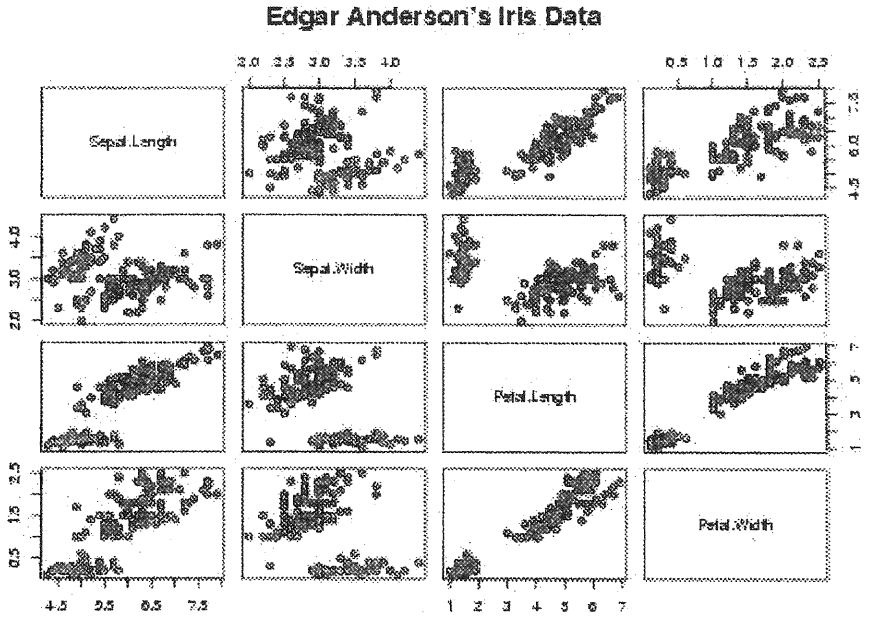


Figure 3 : Largeur et longueur des pétales et ces sépales de 150 iris

On peut également faire faire les calculs à distance sur le serveur de l'université du Montana en faisant appel à Rweb à l'adresse suivante : http://www.math.montana.edu/Rweb/ . Rweb est une interface développée par Jeff Banfield. Elle fonctionne sur la plupart des navigateurs, mais exige la connaissance du langage R (ou celle de Splus). On tape le code, on clique sur un bouton et une page est retournée avec les résultats (analyse et graphiques). Il existe aussi des modules Rweb destinés aux étudiants débutants. Les étudiants n'ont pas besoin de savoir écrire des programmes en R. Ils choisissent un ensemble de données (ou entrent leur propres données), ils sélectionnent le type d'analyse souhaité et les options de l'analyse. Les analyses implémentées sont la régression, les résumés statistiques, l'analyse de la variance, l'analyse des tableaux de contingence, les calculs de distribution de probabilité. Les résultats renvoyés par le serveur de l'université du Montana peuvent être imprimés ou copiés dans des traitements de texte.

Voici un exemple concernant l'étude de l'évaporation journalière du sol en fonction de la température moyenne de l'air. L'étude porte sur 46 jours allant du 5 juin 1964 au 21 juillet de cette même année. La variable avat désigne la température moyenne de l'air et la variable evap la somme de l'évaporation journalière du sol. On obtient, grâce à Rweb, par quelques clics de souris les résultats suivants :

```
    Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -59.82275   14.18847  -4.216 0.000122
avat          0.49605    0.07404   6.699 3.16e-08

    Residual standard error: 10.42 on 44 degrees of freedom
    Multiple R-Squared: 0.505,      Adjusted R-squared: 0.4937
```

F-statistic: 44.88 on 1 and 44 degrees of freedom, p-value: 3.157e-08

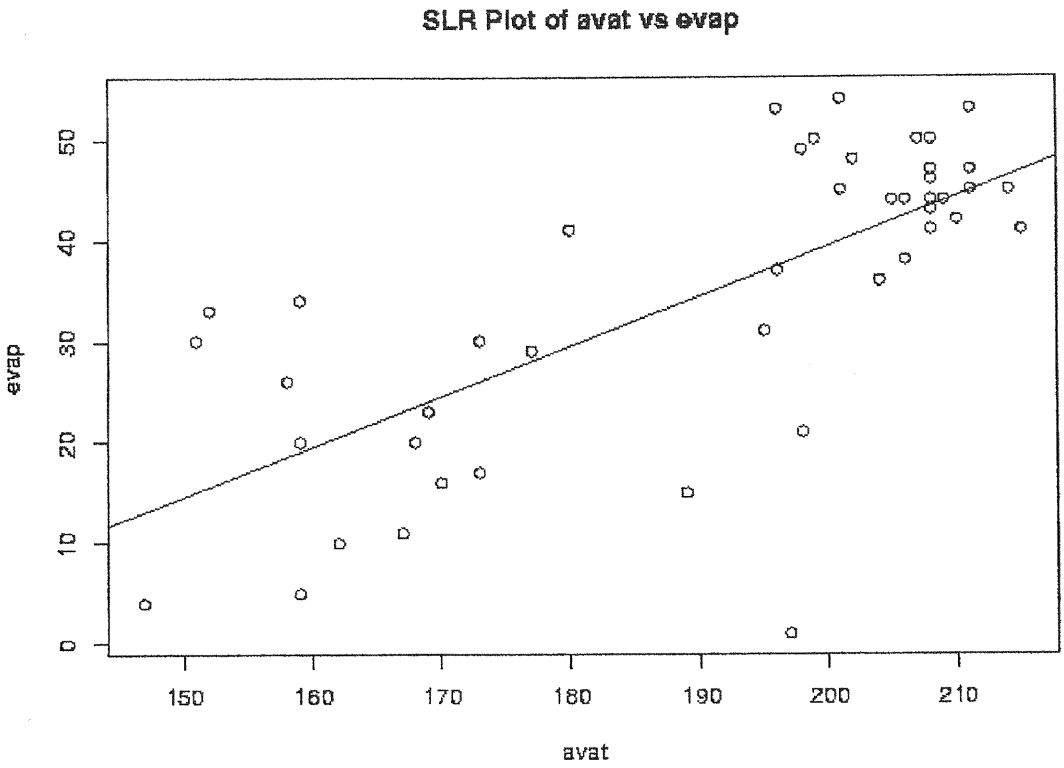


Figure 3 : Evaporation du sol en fonction de la température moyenne de l'air

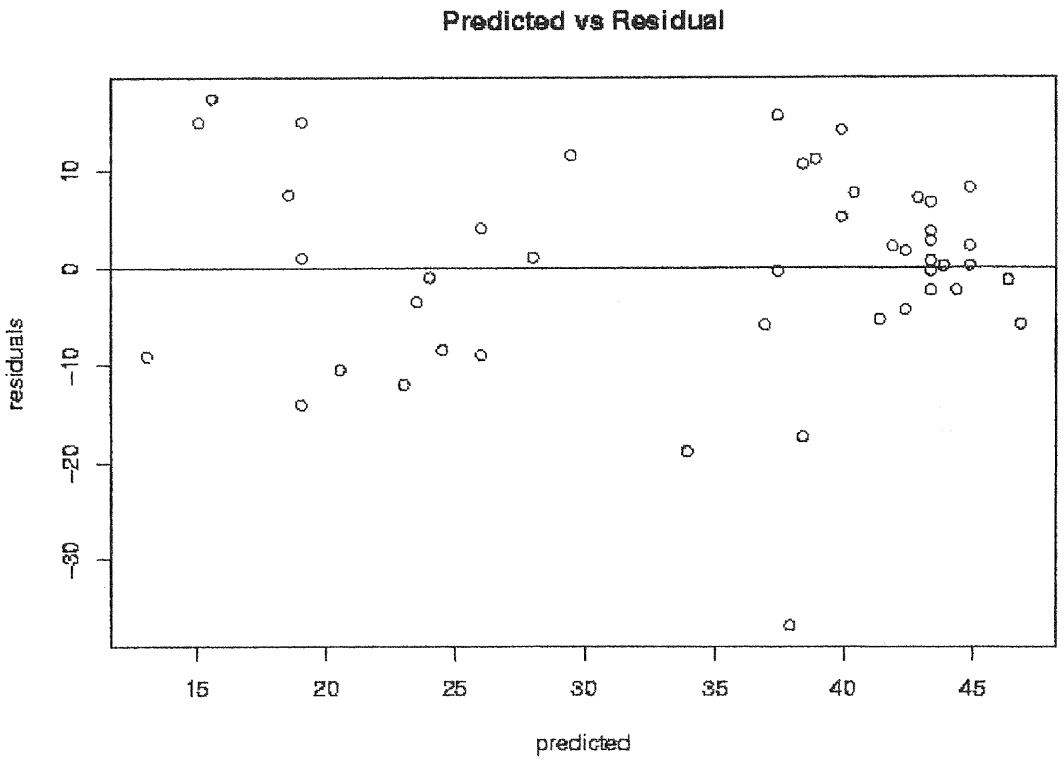


Figure 4 : Nuage des résidus dans la régression précédente

On voit que le coefficient de détermination R = 0,505 et que les points du nuage des résidus semblent être répartis aléatoirement. Le graphique suivant semble confirmer que la distribution des résidus est normale.

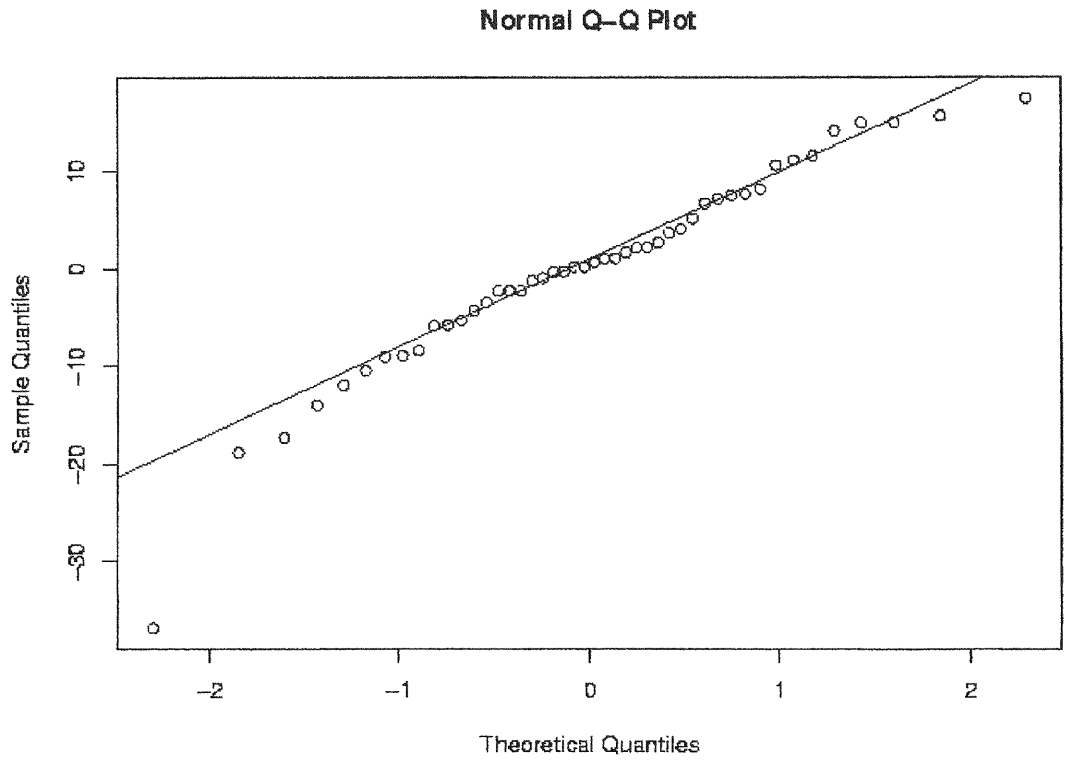


Figure 5 : Ajustement par quantiles de la distribution des résidus à la loi normale.

- **WebStat**

Le projet Webstat a été développé par l'Université de Caroline du Sud. On a accès à ses ressources à l'adresse suivante http://www.stat.sc.edu/webstat/ W. West, membre du projet, cité par G. Saporta, (Saporta, 2000) déclare : « Bien souvent les statisticiens développent des procédures de calcul dans des langages comme Splus, SAS, MiniTab, etc. Les étudiants et autres utilisateurs potentiels peuvent ne pas avoir accès à ces programmes. En employant Java et le World Wide Web, WebStat atteint une audience plus large que celle de n'importe quel autre logiciel statistique.»

Voici un fichier traité par WebStat. Il concerne une étude de JF Fraunieri, concernant la consommation de cigarettes et les cancers, et ses variations géographiques suivant les Etats aux U.S.A. Une première variable est la quantité de cigarette fumée par tête dans les 43 états et le district de Columbia, en 1960, et le taux de mortalité pour 1000 habitants par cancer du poumon. Voici les résultats que l'on obtient en quelques clics :

Simple linear regression results:

Independent variable: CIG
Dependent variable: LUNG
Sample size: 44

Correlation coefficient: 0.6974
(See fitted line plot in Graphics Panel.)
Residuals stored in column var7

Estimate of sigma: 3.0660706

Parameter	Estimate	Std. Err.	DF	Tstat	Pval
Intercept	6.4716864	2.1406693	42	3.023207	0.0021
CIG	0.52907795	0.083895095	42	6.3064227	0

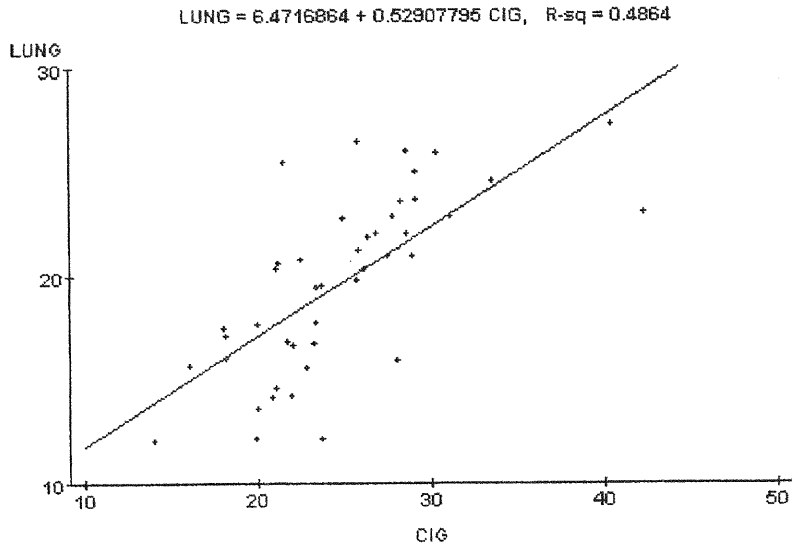


Figure 6 : Cancer du poumon et consommation de cigarettes aux U.S.A. en 1960
Ici le coefficient de détermination est égal à 0,4864.

On peut facilement entrer ces propres données dans WebStat, comme dans un tableur. Un autre avantage c'est que l'on peut inclure Webstat dans des pages web. Sur le site de WebStat on trouera les balises à placer à cet effet dans les fichiers html.

- **XploRe**

Xplore a été développé par l'université Humboldt de Berlin pour l'analyse des données. Il utilise un langage orienté objet. On peut charger une version de démonstration de capacités limitées à l'adresse suivante http://www.xplore-stat.de/ . On trouvera sur ce site de nombreux manuels à télécharger, au format PDF, pour guider l'apprentissage du langage. On peut également communiquer avec Xplore-server et envoyer ses données à Berlin où elles seront traitées et l'analyse vous sera retournée.

- **XLStat**

Il s'agit ici d'un additif au tableur Excel de Microsoft, permettant de faire des analyses poussées tout en restant dans le tableur. On le trouve à l'adresse suivante http://www.xlstat.com/indexfr.html . C'est un additif programmé par Thierry Fahmy, payant, mais à un prix modique.

Voici un exemple. Imaginez que vous travaillez dans une usine et que vous disposez de mesures relatives à un processus de fabrication qui est déjà bien connu. Vous savez que l'intensité Y du phénomène varie en fonction du temps suivant la relation

$$Y = F(t) = 1 + a \exp(-t/b) + \cos(2 \pi t / c)$$

Vous avez observé l'évolution du phénomène sur une période de 45 minutes et vous voulez adapter un modèle aux données, ce qui signifie que vous voulez estimer les paramètres a,b et c. XLSTAT fournit une macro, nommée GenFit, qui va le permettre. En représentant les données on voit qu'il y a une période de 20 minutes. On peut donc dores et déjà dire que c sera proche de 20. On ne peut rien dire sur les paramètres a et b. On donne néanmoins une première estimation de a et b pour éviter à GenFit d'en trouver une lui-même et éventuellement des bornes supérieures et inférieures pour les paramètres à estimer. Il faut ensuite entrer les fonctions à prendre en compte pour l'ajustement, indiquer quelles sont les variables explicatives. On obtient alors l'estimation des paramètres, les écart types, R^2 et les résidus du modèle.

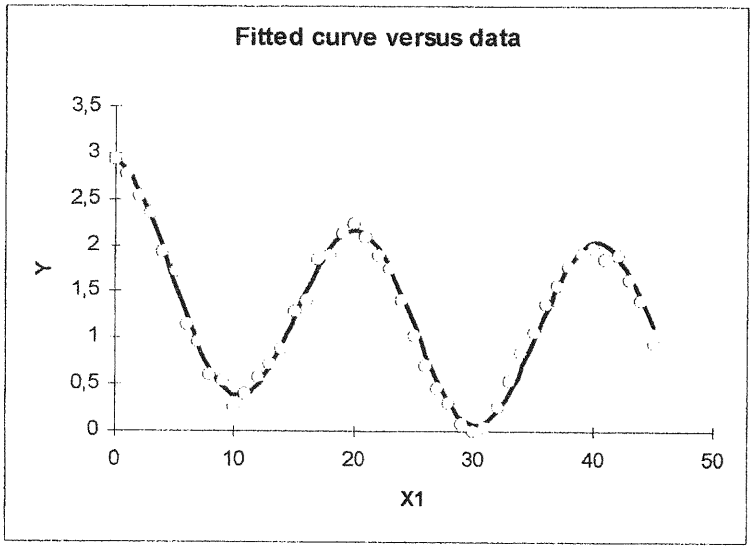


Figure 7 : ajustement d'une courbe à des données

Observations	X1	Y (Data)	Y (Predicted)	Residuals
Obs 1	0	2,93	2,93203341	-0,00203341
Obs 2	1	2,78	2,80456988	-0,02456988
Obs 3	2	2,54	2,59207245	-0,05207245
Obs 4	3	2,35	2,30746553	0,04253447
Obs 5	4	1,94	1,97126967	-0,03126967
Obs 6	5	1,72	1,60956829	0,11043171
Obs 7	6	1,15	1,25144259	-0,10144259
Obs 8	7	0,96	0,92612172	0,03387828
Obs 9	8	0,6	0,66012259	-0,06012259
Obs 10	9	0,55	0,4746547	0,0753453
Obs 11	10	0,27	0,38354001	-0,11354001

Obs 12	11	0,4	0,39184816	0,00815184
Obs 13	12	0,59	0,49537822	0,09462178
Obs 14	13	0,73	0,68103696	0,04896304
Obs 15	14	0,88	0,92807665	-0,04807665
Obs 16	15	1,28	1,21007317	0,06992683
Obs 17	16	1,41	1,49745323	-0,08745323
Obs 18	17	1,86	1,76032708	0,09967292
Obs 19	18	1,9	1,97135325	-0,07135325
Obs 20	19	2,14	2,10835916	0,03164084
Obs 21	20	2,23	2,15646454	0,07353546
Obs 22	21	2,09	2,10950304	-0,01950304
Obs 23	22	1,89	1,97060473	-0,08060473
Obs 24	23	1,74	1,75188326	-0,01188326
Obs 25	24	1,41	1,47325768	-0,06325768
Obs 26	25	1,03	1,16052263	-0,13052263
Obs 27	26	0,7	0,84285267	-0,14285267
Obs 28	27	0,46	0,54998154	-0,08998154
Obs 29	28	0,3	0,30932804	-0,00932804
Obs 30	29	0,09	0,14334583	-0,05334583
Obs 31	30	0	0,06735247	-0,06735247
Obs 32	31	0,05	0,08804704	-0,03804704
Obs 33	32	0,26	0,20285928	0,05714072
Obs 34	33	0,55	0,40019282	0,14980718
Obs 35	34	0,84	0,6605392	0,1794608
Obs 36	35	1,06	0,95835511	0,10164489
Obs 37	36	1,36	1,26452178	0,09547822
Obs 38	37	1,57	1,54914954	0,02085046
Obs 39	38	1,77	1,78445685	-0,01445685
Obs 40	39	1,91	1,94744643	-0,03744643
Obs 41	40	1,98	2,02212022	-0,04212022
Obs 42	41	1,86	2,00101968	-0,14101968
Obs 43	42	1,9	1,88594298	0,01405702
Obs 44	43	1,62	1,68776969	-0,06776969
Obs 45	44	1,41	1,42541007	-0,01541007
Obs 46	45	0,95	1,12398016	-0,17398016

Tableau des observations, des valeurs prédites par le modèle et les résidus.

Les habitués d'Excel reconnaîtrons les tableaux et les graphiques que l'on obtient dans ce tableur. Notons aussi qu'Excel dispose dans sa version 2000, d'un *utilitaire d'analyse*, qui n'est pas automatiquement chargé lorsqu'on installe le logiciel, mais que l'on peut charger ultérieurement. Cet utilitaire permet enfin de construire des histogrammes sans faire des acrobaties en passant d'abord par la construction de nuages de points ! Il permet de faire de l'analyse de la variance, d'étudier les moyennes mobiles, de faire des régressions multiples, des échantillonnages et des tests, construire des intervalles de confiance.

2. Les applets Java

On peut également trouver sur le réseau de nombreux applets Java, pour réaliser des simulations. Les sites où l'on en trouve sont assez nombreux. Citons en quelques-uns.

- **SEL et SMEL**

SEL signifie statistique en ligne, on trouve ce site à l'adresse suivante : http://www.inrialpes/sel , SMEL signifie statistique médicale en ligne. Ce sont des sites miroir, le deuxième fournissant en plus des données médicales. L'adresse de SMEL est http://www.math-info.univ-paris5.fr/ .

On trouve sur ces sites des *simulations de tirages d'échantillons* distribués suivant une des lois suivantes : loi normale, loi log-normale, loi exponentielle, loi de Weibull, loi binomiale, loi uniforme, loi de Poisson ou loi géométrique. On choisit une loi, on simule un échantillon de cette loi, les caractéristiques empiriques de l'échantillon (histogramme ou diagramme en bâtons, diagrammes en boîtes, fonction de répartition, fonction quantile empirique) apparaissent en bleu avec les mêmes échelles que les caractéristiques théoriques en rouge. On peut alors observer la convergence quand la taille des échantillons augmente.

On y trouve également une *simulation à propos du théorème central-limite*. On choisit une loi et un nombre n. La loi étudiée est celle de la somme centrée réduite d'un échantillon de taille n de la loi de départ. Cette loi est simulée de façon répétée et un histogramme est superposé à la densité de la loi normale $N(0,1)$.

Un autre applet permet la *comparaison d'estimateurs*. La loi simulée est la loi uniforme sur l'intervalle $[0,\theta]$, où θ est un paramètre à choisir. Un échantillon $(X_1,\ldots X_n)$ de cette loi est simulée et cinq estimateurs de θ sont calculés à partir de cet échantillon. Tous les cinq sont des estimateurs convergents de θ : T_1 est la moyenne géométrique des X_i multipliée par e, T_2 est la médiane des X_i multipliée par 2, T_3 est la moyenne arithmétique des X_i multipliée par 2, T_4 est le maximum des X_i multiplié par n, T_5 est le maximum des X_i multiplié par $n/(n-1)$. Il est intéressant de pouvoir étudier ainsi les propriétés des cinq estimateurs de θ.

On trouvera également des applets mettant en évidence *le caractère biaisé ou non biaisé de certains estimateurs de la variance*, et bien d'autres choses.

Sur ces sites on trouvera un *lexique* des termes usuels de statistique et un *cours de statistique*, rédigé par B. Ycart, dans la plus pure tradition de rigueur de l'école mathématique française, avec les notations adoptées par la communauté des statisticiens.

- **VESTAC**

Visualisation of and Experimentation with STAtistical Concepts, site de l'université de Louvain : http://www.kuleuven.ac.be/ucs/java/index.html . On trouve sur ce site des applets Java qui s'affichent en plein écran, avec des boutons dont les étiquettes ne sont écrites ni en français, ni en allemand, ni en néerlandais, mais en anglais, langue neutre. On peut jouer très facilement sur les paramètres. On pourra regarder l'applet consacré aux *intervalles de confiance pour la moyenne*, qui permet de bien comprendre l'influence du taux de confiance et celle de la taille de l'échantillon sur la largeur de l'intervalle de confiance.

- **GASP**

Globally Accessible Statistical Procedures à l'adresse suivante : http://www.stat.sc.edu/rsrch/gasp/ . On y trouvera sous la rubrique « Educational Procedures » une présentation intéressante du *théorème de la limite centrée*, et une démonstration interactive de la *notion de puissance d'un test d'hypothèses.*

On trouvera également sur ce site un applet très intéressant sur la construction des *histogrammes*. A partir de données sur la durée des éruptions du geyser Old Faithfull, du parc national de Yellowstone, aux Etats Unis, il s'agit de faire comprendre aux étudiants comment la largeur des classes ou le nombre de classes affecte un histogramme. De plus

vous pourrez incorporer cet applet dans vos propres pages web, pour représenter vos données.

Sous la rubrique procédures d'analyse de données, on pourra consulter le Statiscope. On obtient alors un calculateur des densités et des fonctions de répartition des lois usuelles, ainsi qu'un logiciel de représentation graphique de ces fonctions, et enfin un générateur de nombres pseudo-aléatoires distribués selon la loi que l'on veut, et qui vous sont livrés par courrier électronique. Moyennant une licence ce logiciel peut être intégré à vos pages Web.

- **WISE**

Web Interface for Statistics Education de la Claremont Graduate University. Sur son site http://acad.cgu.edu/wise/appletsf.shtml on trouve trois applets particulièrement intéressants sur la *distribution d'échantillonnage d'une moyenne*, sur l'*introduction aux tests d'hypothèses* et enfin sur la *notion de puissance d'un test et sa relation avec la taille de l'échantillon et le taux de risque.*

On est renvoyé à d'autres sites offrant d'autres applets. Signalons un applet reproduisant une *planche de Galton* pour la simulation de la loi binomiale. On peut y accéder directement à l'adresse : http://www.stattucino.com/bernie/clst/Galton.html .

Signalons également un applet sur la *régression* permettant d'ajouter un point aux données et voir ainsi son influence sur la droite de régression. On y accède directement par http://www.stat.sc.edu/~west/javahtml/Regression.html.

- **Duke University**

La Duke University propose une liste d'applets à l'adresse http://www.stat.duke.edu/sites/java.html. Signalons un applet sur le *mouvement brownien* auquel on peut accéder directement par http://www.stat.unm.edu/~charlie/Stoch/brown.html et une autre pour illustrer la convergence de la loi binomiale vers la loi normale, suivant le théorème de de Moivre Laplace que l'on trouve directement à l'adresse suivante : http://www.ruf.rice.edu/~lane/stat-sim/binom-demo.html .

- **Champaign-Urbana Web University Statistics Program**

Champaign-Urbana Web University Statistics Program propose à l'adresse suivante : http://neyman.stat.uiuc.edu:80/~stat100/cuwu/index.html des jeux, comme le *poker*, permettant de compter le nombre de mains comportant des doubles, des quintes, etc. sur 1000 jeux et de comparer le résultat aux probabilités théoriques ; pour *le coefficient de corrélation*, un applet propose plusieurs nuages de points et des coefficients de corrélation à associer aux nuages. On trouve également sur ce site des applets pour analyser les données et un certain nombre de fichiers de données à analyser.

- **Rice Virtual Lab in Statistics**

Le Rice Virtual Lab in Statistics de la Rice University que l'on trouve à l'adresse suivante http://www.ruf.ricr.edu/~lane/stat_sim/index.html fournit des applets Java de simulation, accompagnés de commentaires explicatifs et d'exercices pour tirer partie de ces ressources. On y trouve des applets pour comparer deux valeurs centrales *la moyenne et la médiane.* Dans une applet dédiée à la comparaison de deux séries, à l'aide de boîtes de dispersion : on mesure le temps que vous mettez à cliquer sur un bouton de taille réduite et sur un bouton plus grand. Une applet particulièrement intéressante sur les

distributions d'échantillonnage permet de changer la distribution de la population mère, la taille de l'échantillon, le nombre des échantillons, de comparer la distribution d'échantillonnage à une distribution normale, de changer d'estimateur (moyenne, médiane, étendue, variance biaisée ou non biaisée). Les exercices associés permettent de faire réfléchir sur la notion de biais d'un estimateur, d'efficacité, sur le théorème de la limité centrée. Plusieurs applets permettent d'étudier la notion de *coefficient de corrélation* en faisant varier l'étendue du domaine des variables. Une autre applet permet d'étudier la *décomposition de la variance* dans un ajustement à un modèle linéaire.

- **Virtual Laboratories in Probability and Statistics**

de Kyle Siegrist de l'Université d'Alabama à Hunstville, dont l'adresse est http://www.math.uah.edu/stat/ . Ce laboratoire virtuel se présente sous la forme d'un cours regroupant des notions de combinatoire, de probabilités, de statistique inférentielle et la présentation de différents modèles probabilistes. Ce cours est illustré par des simulations et enrichi par des exercices. En combinatoire on trouvera une *planche de Galton* pour illustrer les coefficients du binôme de Newton. En probabilités on pourra simuler le *lancer de n dés* et comparer les distributions théoriques ou distributions empiriques de la somme ou de la moyenne des résultats des *n* dés ou de leur valeur maximum ou minimum. On pourra *simuler les lois usuelles*, construire un *nuage de points interactif*, pour illustrer la notion de coefficient de corrélation. On trouvera également des applets sur *les statistiques d'ordre*, les *processus de Poisson, les marches aléatoires, la ruine des joueurs (Red and black game)*, et le *problème du dépouillement des scrutins (ballot experiment)*.

- **JAVA Applets par B.Narasimhan de Stanford University**

A l'adresse http://www-stat.stanford.edu/~naras/jsm/ on trouvera quelques applets de B.Narasimhan, notamment un applet illustrant *la convergence de la loi de Student vers la loi normale*, lorsque le nombre de degrés de liberté croît.

- **SurfStat JAVA Applet Hotlist**

A l'adresse www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html, on trouve quelques applets illustrant, par exemple, les *paramètres centraux* d'une distribution statistique, et la *régression linéaire*.

- **STATLETS**

est une collection de plus de 50 Java applets pour analyser des données sur Internet, intranets locaux ou sur son PC. L'adresse à laquelle on peut les trouver est : http://www.sgcorp.com/statlets.htm

On peut travailler en ligne ou télécharger les 50 applets gratuitement dans leur version académique. Mais dans ce cas les fichiers traités ne doivent pas dépasser 50 observations et 8 variables. Les domaines traités sont la statistique descriptive, l'estimation, la théorie des tests, la régression, l'analyse de la variance, le contrôle de qualité et les séries temporelles.

3. Les bases de données.

La recherche de données brutes est loin d'être aisée. La grande totalité des sites des grands organismes traitant des statistiques, comme l'INSEE, en France, ne fournissent pas de données brutes mais des tableaux et des graphiques où les données sont déjà analysées. Le

professeur peut néanmoins en tirer partie pour faire construire de nouveaux graphiques et comparer des situations ou montrer des évolutions. On peut citer le travail remarquable accompli par des professeurs de sciences économiques et sociales de l'académie de Montpellier, à partir du *tableau de l'économie française* de l'INSEE. Ce travail a fait l'objet d'un CD-Rom édité par le CRDP de Montpellier. Les données sont commentées et un mini-tableur permet de les traiter. Les tableaux fournis dans ce CD-Rom peuvent être importés dans Excel.

Quand on trouve des tableaux de données, encore faut-il pouvoir ensuite les traiter soit dans un tableur, soit dans un logiciel de traitement statistique. On évitera les fichiers PDF qui ne peuvent être qu'imprimés. Certains sites proposent des fichiers qui sont directement au format d'Excel, d'autres des fichiers au format html, qu'Excel sait lire, ou enfin des fichiers au format txt, qu'Excel sait également lire pourvu qu'ils s'agissent de tableaux. On n'évitera pas, parfois, d'avoir à changer le point en virgule, comme séparateur entre la partie entière et la partie décimale d'un nombre.

- **Organisation des Nations Unies pour le Développement**

On trouvera une mine de données économiques, démographiques, culturelles sur le site de cet organisme : http://www.undp.org. Il publie chaque année un rapport sur un thème différent concernant le développement économique. On trouvera la totalité du rapport pour l'année 2000 à l'adresse suivante : http://www.undp.org/hdr2000/home-fr.html au format PDF. Dans ce rapport on trouvera les définitions des indices IDH, indice de développement humain, ISDH, indice sexo-spécifique de développement humain, IPF, indice de participation des femmes, dans les notes techniques du rapport. On pourra télécharger tous les tableaux de données à partir de http://www.undp.org/hdro/statistics.html On obtiendra ainsi sur son disque dur un fichier compressé au format zip, de 1229Ko, qui une fois décompressé fournira 34 tableaux au format Excel.

- **OnuSida**

On trouve sur le site d'OnuSida, des statistiques en format html, sur la pandémie du Sida : http://www.unaids.org/wac/2000/wad00/files/WAD_Epidemic_report_f.htm

Le même rapport existe au format PDF, plus facile à lire mais les données ne pourront pas alors être importées dans un tableur.

- **Organisation mondiale de la santé**

On accède directement aux statistiques de l'organisation mondiale de la santé par http://www.who.int/whosis/ . On obtient alors un grand nombre de fichiers concernant l'espérance de vie, les différentes maladies, la couverture vaccinale, etc. sous la forme de fichiers Excel que l'on peut télécharger ou ouvrir directement si ce tableur figure dans votre système. On obtient aussi de magnifiques cartes statistiques, comme cette carte sur l'espérance de vie.

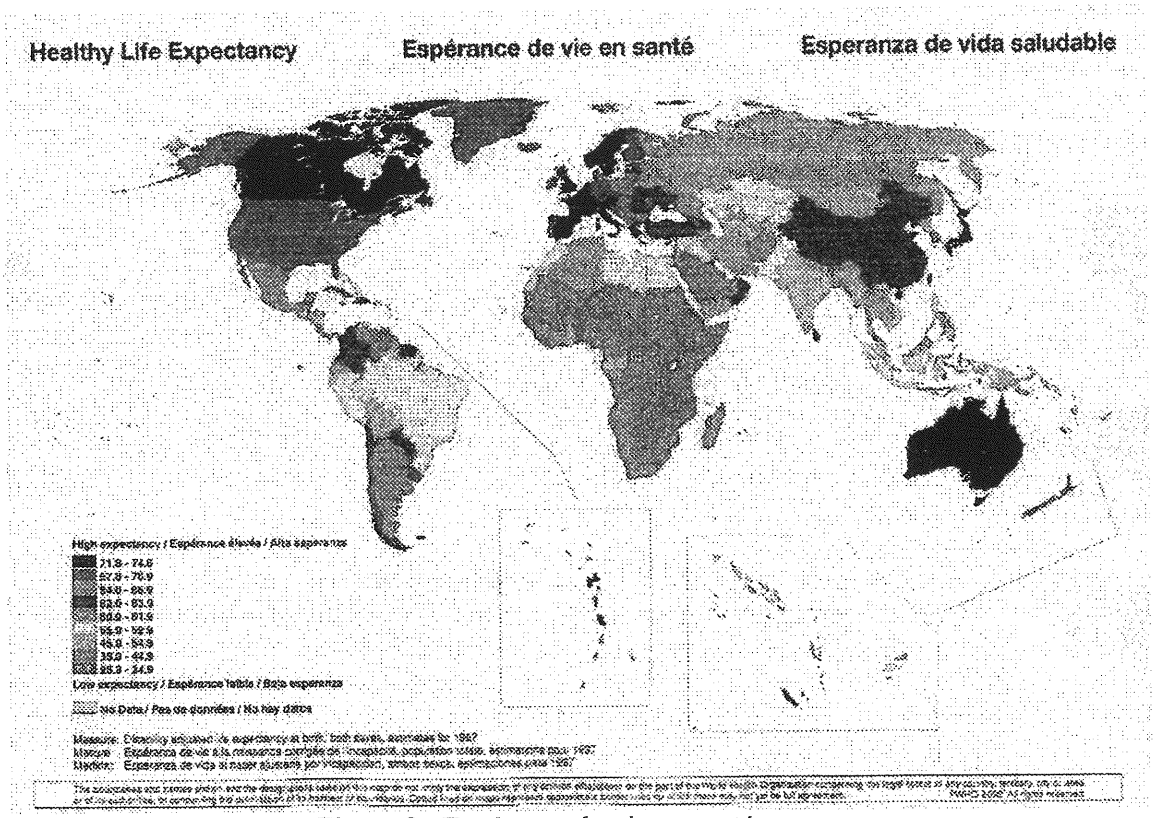


Figure 8 : Espérance de vie en santé

- **Le recensement de 1999 en France**

Les données concernant le recensement de la population en France, effectué en 1999 se trouve sur le site http://www.recensement.insee.fr/ . On y trouve les évolutions démographiques par zones géographiques.

- **INSEE**

Par http://www.insee.fr/fr/region/accueil_region.htm on accède aux données régionales en France, mais elles sont en général payantes. Quelques données sont fournies gratuitement à titre d'exemples.

- **Statistiques des communautés européennes**

A l'adresse http://europa.eu.int/comm/eurostat/ on accède à un service des communautés européennes fournissant des statistiques économiques. Malheureusement la plupart des données sont payantes.

- **Institut National des Etudes Démographiques (INED)**

A l'adresse http://www.ined.fr/ on trouvera des tableaux statistiques sur la démographie de tous les pays, au format htm, que l'on peut facilement importer dans Excel ou des graphiques au format gif, que l'on peut insérer dans des documents crées à l'aide de traitements de textes, comme cette pyramide des âges.

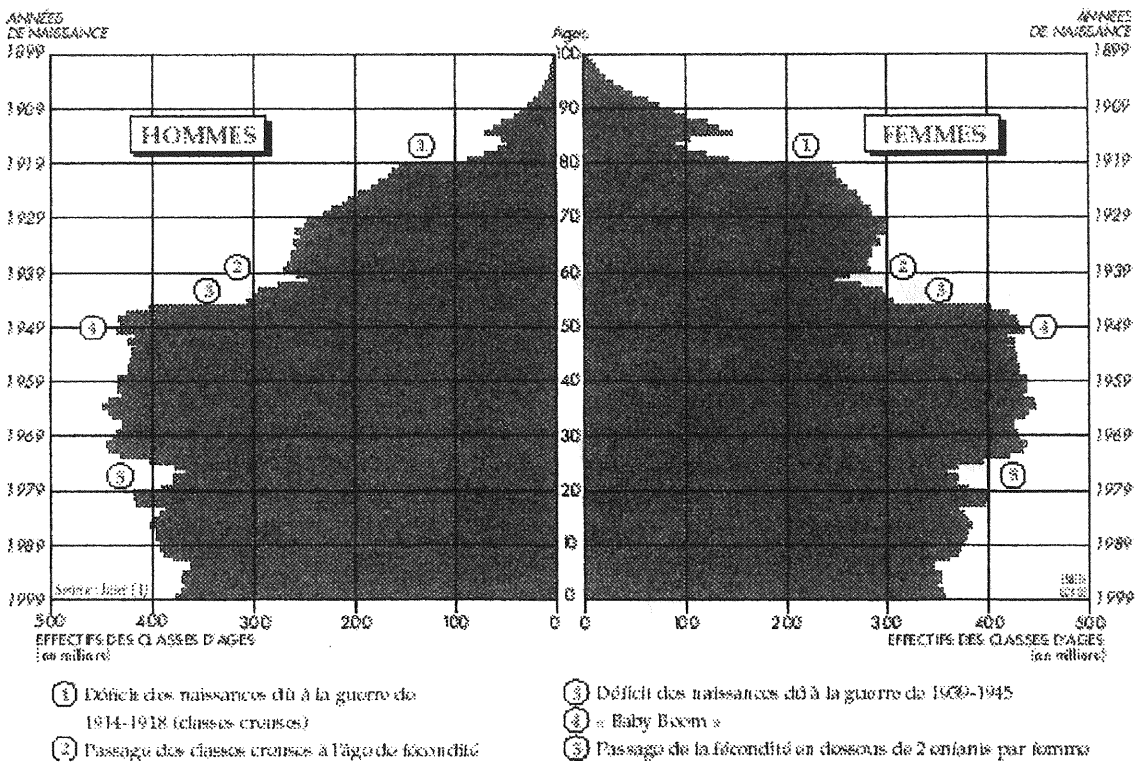


Figure 9 : Pyramide des âges en France au premier janvier 2000

- **Organisation de coopération et de développement économique (OCDE)**

Le site de cet organisme (http://www.oecd.org/) fournit de nombreux renseignements, mais ses publications sont payantes. On peut néanmoins se procurer un certain nombre de données gratuitement au format PDF.

- **Division statistique de l'ONU**

Cet organisme fournit un certain nombre de données démographiques et économiques gratuitement, au format htm, ou format gif, comme cette carte indiquant la population des principales villes de l'océanie.

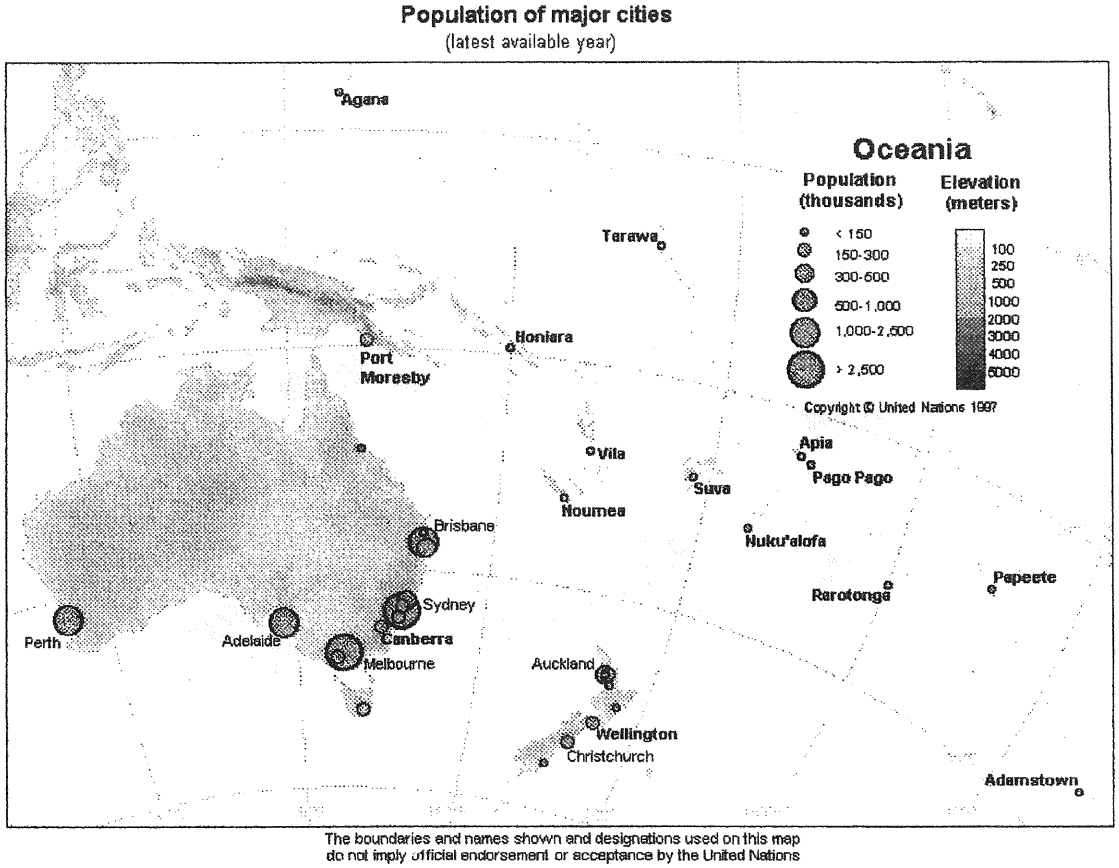


Figure 10 : Population des principales villes de l'Océanie

- **Statistique du Canada**

On trouvera sur le site http://www.Statcan.ca/ de nombreuses données concernant la géographie, la santé, l'économie, le climat, l'environnement du Canada, mais aussi de pays voisin. Les tableaux au format htm sont importables dans les tableurs.

- **Statistique médicale en ligne (SMEL)**

Sur ce site, on trouvera quelques donnée brutes médicales.

- **The Data And Story Library (DASL)**

Les enseignants ont besoin d'exemples pour illustrer les concepts de statistique. Un bon exemple rendra une leçon plus vivante et plus pertinente. Le projet DASL (prononcer dazzle), conçu par des enseignants de l'université Cornell, à New York, a pour but de permettre aux professeurs de trouver des données utiles pour leur enseignement. Les données que l'on trouve à l'adresse http://lib.stat.cmu.edu/DASL/DataArchive.html sont classées suivant les sujets ou suivant les méthodes statistiques utilisées pour les traiter. Les données peuvent être téléchargées et pourront être traitées par la plupart des programmes de statistiques. Avec Excel il faudra savoir utiliser les fonctions sur les chaînes de caractères pour transformer les tableaux obtenus en tableaux numériques.

On aura ainsi les données recueillies par Hubble pour étudier la relation entre la distance et la vitesse radiale des nébuleuses extra-galactiques, mais aussi des explications concernant cette expérience et la signification de la constante de Hubble. On aura les données relatives au procès engagé par la municipalité de New York contre la Brink, dont

les employés, chargés de relever les pièces dans les parcmètres, étaient soupçonnés de détournement. On aura des données sur l'évolution des dimensions des crânes des habitants de l'Egypte ancienne, de –4000 à – 150 ans avant JC, prouvant qu'à la population locale s'était ajoutée des vagues d'immigrants.

- **Chance**

Le but de Chance est de mieux informer les étudiants, et en faire des lecteurs critiques des informations utilisant les probabilités et la statistique. La rubrique « Chance News » est une lettre mensuelle qui fournit des articles tirés de journaux et suggèrent des débats à mener en classe, On trouve également sur ce site une collection de conférences, que l'on peut écouter avec l'aide du plug-in RealPlayer. Le conférencier apparaît dans une fenêtre et ses transparents dans une autre. On appréciera la conférence de Howard Warner, où il traite avec beaucoup d'humour du sujet : « Comment mal présenter des données » (Attention, la conférence dure une heure et demie). On trouve sur le site un certain nombre de données et surtout des liens vers d'autres sites fournissant des données. Comme le sujet est inépuisable, nous vous laissons le plaisir d'aller butiner sur les sites que propose Chance.

- **Les instituts de sondage.**

On trouve sur les sites des instituts de sondage les résultats de leurs enquêtes, concernant à la fois la politique, les côtes de popularité des personnalités, les phénomènes de société, les questions sociales et économiques. Voici leurs adresses :

- **BVA** [illegible]
- **IPSOS** www.canalip[illegible]
- **IFOP** www.ifop.fr/homeifop.htm
- **CSA** www.csa-tmo.fr
- **SOFRES** www.sofres.com

4. Les manuels électroniques et les cours sur Internet

Des manuels électroniques sous forme de Cd-Rom avec des liens hypertextes existent déjà depuis plusieurs années comme "Techniques de la Statistique" dans la collection CNAM-Media diffusée par Jériko (F.Kazi-Aoual, M.Janvier, 1998), ou en italien "Introduzione alla Statistica Descrittiva" de A.Di Ciaccio et S.Borra diffusé par McGraw Hill. La nouvelle génération sur Internet fournit des liens Web et rend possible l'utilisation de ressources distantes comme des applets Java.

- **Multivariate statistics, concepts, models and applications**

de David W. Stockburger (1998) de Southwest Missouri State University http://www.psychstat.smsu.edu/multibook/mlt00.htm où on trouve un cours sur l'analyse discriminante, la classification; la régression multiple, l'analyse de variance et les plans d'expériences. Certains calculs proposés peuvent être faits à l'aide d'applets Java. On y trouvera également une abondante liste de liens vers des sites proposant des données.

- **Le dictionnaire de StatSoft**

Nous avons déjà mentionné ce dictionnaire téléchargeable gratuitement diffusé par StatSoft Inc (http://www.statsoft.com/textbook/stathome.html).

- **Statistics. The study of stability in variation**

Le manuel électronique de l'Université de Californie à Los Angeles, développé par J. de Leeuw (1997) (http://www.stat.ucla.edu/textbook/) a l'ambition de fournir un manuel gratuit accessible à tous sur Internet, indépendamment du niveau de l'étudiant. C'est un produit interactif, utilisant largement les graphiques et couvrant un large spectre depuis l'analyse d'une seule variable à l'analyse multidimensionnelle. Le cours offre un lexique, une boîte à outils et des démonstrations écrites en Xlisp-Stat .

- **SurfStat**

SurfStat (http://www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html) qui a démarré en 1994 avec *« l'intention courageuse mais naïve de mettre des notes de cours existantes en ligne en hypertexte. Il s'est ensuite enrichi d'un glossaire, d'exercices interactifs, de fonctions JavaScript à la place des tables statistiques, et un ensemble d'applets Java pour illustrer les concepts statistiques avec des graphiques dynamiques. Surfstat est la ressource de base pour les étudiants en statistique élémentaire à l'Université de Newcastle, Australie ».* Le cours comprend les rubriques suivantes :

i. Résumer et présenter les données
ii. Variations et probabilités
iii. Recueil de données
iv. Statistiques inférentielles
v. Contrôle de production

On appréciera la manière très pragmatique et illustrée de nombreux exemples pratiquée sous les rubriques sur le recueil de données et la statistique inférentielle. Les professeurs pourront y trouver matière à réflexion pour préparer leurs cours.

- **Virtual Laboratories in Probability and statistics**

Le projet "Virtual Laboratories in Probability and statistics" a été mené à bien par Kyle Siegrist de l'Université d'Alabama, à Huntsville (http://www.math.uah.edu/~stat/). Il est divisé en chapitres comme dans un cours traditionnel : calcul des probabilités, statistique inférentielle, modèles particuliers. Le chapitres comportent les définitions de notions et de nombreux exercices qui visent à faire retrouver pas à pas les différents résultats théoriques. Certains de ces exercices consistent en des question suggérées par des animations.

On pourra consulter le chapitre sur les modèles particuliers, où l'on trouvera des explications théoriques, illustrées par des simulations, sous la forme d'applets Java. Les modèles suivants sont présentés : *processus de Poisson, marches aléatoires, propagation du feu, dépouillement d'un scrutin*, différents jeux de hasard (*roulette, baccarat, poker, keno*,etc.), des modèles géométriques (*paradoxe de Bertrand, aiguille de Buffon, construction d'un triangle à partir d'un segment découpé en trois segments*)

- **Web Interface for Statistics Education (WISE)**

WISE "Web Interface for Statistics Education" (http://acad.cgu.edu/wise/) du Claremont College concerne plutôt les sciences sociales. WISE est conçu comme un complément à l'enseignement présentiel pour aider à la présentation de sujets que les enseignants ont généralement du mal à faire passer en classe, mais sert aussi d'auto apprentissage et de moyens de révision pour les étudiants. Il comporte plusieurs didacticiels : distributions d'échantillonnage, la puissance d'un test, une présentation du théorème de la limite centrée, la logique des tests d'hypothèse, choisir correctement un test statistique, comparaison de deux cultures. Dans ces didacticiels l'étudiant est aidé pas à pas dans sa démarche, par une analyse asses fine de ses réponses.

5. L'enseignement à distance

En dépit de leur qualité, les produits précédents sont surtout destinés à un usage en classe ou pour de l'auto apprentissage. De manière quelque peu surprenante il semble que les Universités à distance traditionnelles ne font encore guère usage des ressources d'Internet, à part la communication par courrier électronique et la possibilité de télécharger des documents papier.

- **Cours multimedia de la FernÜniversität de Hagen**

Une exception est le cours multimédia (en allemand) développé par H.J.Mittag de la FernÜniversität de Hagen (http://www.fernuni-hagen.de/STATISTIK). Cette université a développé l'enseignement à distance et en particulier un cours de *statistique descriptive et d'analyse des données.* Il est à la disposition des écoles secondaires en Westphalie pour préparer l'enseignement de la statistique. On peut télécharger une version de démonstration, à condition de disposer d'une connexion à haut débit à Internet (3,25 minutes de téléchargement à 15 Ko/s)! On verra alors des présentations de notions à l'aide de Real Vidéo, avec des animations et des commentaires parlés (en allemand). On aura ainsi une présentation des *boîtes à moustaches*, un applet Java sur *la concentration et l'indice de Gini.*

- **Le projet** St@tnet

Il a également semblé à la chaire de statistique appliquée du Conservatoire National des Arts et Métiers, et au centre associé du Languedoc-Roussillon qu'il y avait un fort potentiel pour une combinaison de ressources Internet en statistique et une organisation d'enseignement à distance.

St@tnet est un projet d'enseignement à distance sur Internet, mené à bien par le centre régional associé du Languedoc-Roussillon du Conservatoire National des Arts et Métiers. Il a été cofinancé par le programme UVF de l'AUPELF-UREF. Il correspond à la demi valeur Techniques de la Statistique et comprend six modules :

Module 1 : Statistique Descriptive
Module 2 : Calcul des probabilités et Analyse combinatoire
Module 3 : Variables aléatoires
Module 4 : Echantillonnage et Estimation
Module 5 : Ajustement linéaire
Module 6 : Tests

Les modules 1, 3, 4 et 6 ont été conçu par F.Kazi-Aoual, les modules 2 et 5 par M. Janvier, ils ont été réalisés par M.Marco, les animations on été confiées à V.Guijarro. Ce travail a été réalisé dans le cadre de l'Ecole d'Agronomie de Montpellier, avec la caution de G.Caraux et dans le cadre de la Chaire de Statistique appliquée du Conservatoire National des Arts et Métiers, dirigée par G.Saporta. A.Brethon et E.Verdoire ont eu la charge de la logistique. Le projet devrait être opérationnel fin 2001.

La partie déjà réalisée est consultable actuellement à l'adresse http://www.cnam.agropolis.fr/statnet. Chaque module est décomposé en quatre parties : présentation, phase de dévolution pour présenter une problématique ; développement où les concepts sont présentés et illustrés, l'étudiant est constamment sollicité, par des questions auxquelles il doit répondre ; résumé, qui est la phase d'institutionalisation des

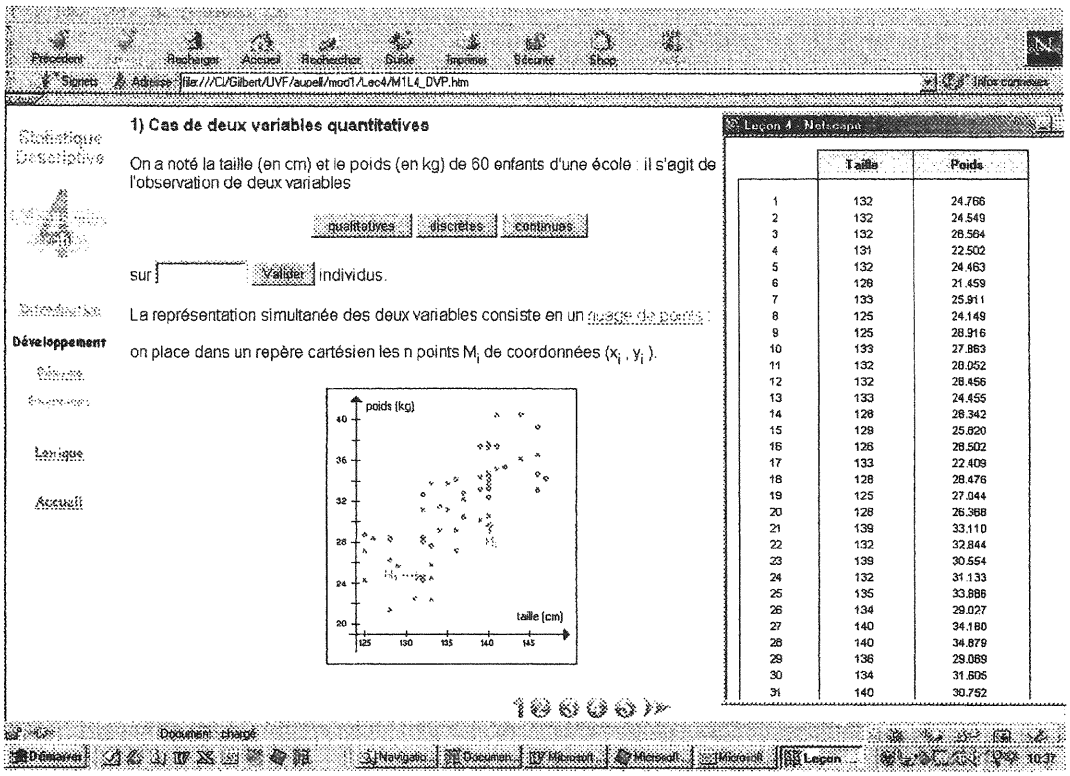


Figure 11 : St@tnet, leçon sur la régression

connaissances ; enfin exercices pour vérifier l'acquisition des concepts et des procédures visées par le module.

On trouve également un glossaire, des utilitaires de statistique et des liens vers des bases de données ou d'autres sites de statistique. Le cours sur Internet a repris en partie le contenu du CD-Rom « CNAM-Media », distribué par Jeriko, écrit en Tool-Book,((F.Kazi-Aoual, M.Janvier, 1998), mais le portage sur Internet a nécessité une réécriture complète en utilisant des langages comme HTML, Java ou des utilitaires comme Flash ou DreamWeaver.

Les auditeurs qui s'inscrivent à cet enseignement à distance ont accès à un tutorat, avec devoirs, forum et courrier électronique. Ils ont la possibilité de passer l'examen dans leur centre régional.

6. Une petite liste de liens

Pour les lecteurs qui n'auraient pas encore trouvé leur bonheur dans les ressources que nous avons signalé nous donnons maintenant une liste de liens où ils pourront tenter leur chance. Comme il est impossible de tenir à jour en temps réel de telles listes, celles que nous indiquons ci-dessous sont nécessairement incomplètes et contiennent probablement des liens périmés. Ces listes sont maintenues soit par des sociétés savantes, soit par des sociétés privées soit par des universitaires.

- **International Association for Statistical Computing**
http://www.stat.unipg.it/iasc/

- **American Statistical Association**
http://www.amstat.org

- **Société Française de Statistique**
http://www.sfds.asso.fr.

• **Centro Interuniversitario di Ricerca per la Didattica delle Discipline Statistiche (CIRDIS)**
http://www.stat.unipg.it/cirdis/, site italien

• **Guide to the Web for Statisticians**
http://www.maths.uq.oz.au/~gks/webguide/, site australien

• **Statlib**
http://lib.stat.cmu.edu/, site de l'université Carnegie Mellon .

• **STATSNET**
http://www.statsnet.net/, site privé.

• **York University**
http://www.math.yorku.ca/SCS/StatResource.html, site tenu par Mike Friendly.

• **Juha Puranen**
d'Helsinki (http://noppa5.pc.helsinki.fi/links.html) qui tient une des listes les plus complètes pour l'enseignement de la statistique avec 19 pages de liens!

• **David Stockburger**
fournit avec son manuel électronique d'Analyse Multidimensionnelle une importante liste de liens avec de brefs commentaires (http://www.psychstat.smsu.edu/scripts/dws148f/statisticsresourcesmain.asp).

• **Robin H. Lock**
L'article interactif de Robin H. Lock du programme TISE (Technology in Statistics Education) (http://it.stlawu.edu/~rlock/tise98) est également une source précieuse .

Conclusion

On a pu constater la grande quantité de ressources disponibles sur Internet. Nous avons certainement laissé de côté des sites dignes d'intérêt que nous n'avons pas visité. Il faut prendre garde de ne pas devenir accroc au réseau. Plus sérieusement, on peut déplorer le petit nombre de sites francophones. Si l'on peut sans grandes difficultés mettre au point des applets Java, il est déjà beaucoup plus difficile de réaliser un véritable logiciel d'analyse statistique. La réalisation d'un didacticiel suppose des moyens humains considérables. On peut craindre la domination de ceux qui auront le plus de moyens. Les francophones ont de gros efforts à faire pour améliorer leur présence sur le réseau.

Espérons que la mise sur Internet de simulations, de présentations de notions permettront aux enseignants de confronter leurs pédagogies, et ainsi de se perfectionner pour le plus grand bénéfice des étudiants.

Bibliographie

De Leeuw J. (1997) *The UCLA Statistics Textbooks and Modules*, **51ième session de l'Institut International de Statistique**, Istanbul, 55-59

Janvier M. (1999) *Statistique descriptive, avec ou sans tableur*, Dunod

Lejeune M. (1998): *Statistique et Internet*, Journée organisée par la Société Française de Statistique, 11 rue P.et M.Curie, F 75005 Paris

Lock, Robin H. (1998) : *WWW resources for teaching statistics*, http://it.stlawu.edu/~rlock/tise98

Mittag H.J. (1999): *Emphasizing Visualization and Activities in Teaching Introductory Statistics by Interactive Multimedia*, **52ème Session de l'Institut International de Statistique**, Helsinki, Finlande, 10-18 août 1999

Morin, A., and Saporta, G. and Spinakis, A. (1998): *Computer Assisted Training in Statistics: Internet and Multimedia: a Survey of Existing Tools*, **in Proceedings of NTTS'98**, Sorrento, Italy (eds. P.Nanopoulos, P.Garonna, C.Lauro), vol 2, 333-338.

Saporta G.(1999): *Teaching Statistics with Internet: a Survey of Available Resources and the St@tNet Project*, **52ème Session de l'Institut International de Statistique**, Helsinki, Finlande, 10-18 août 1999

Saporta G.(2000) : *Enseigner la statistique sur Internet; présentation de St@tNet* **VII° Journées Internationales de Technologies**, Beyrouth, 2-5 mai 2000.

Velleman, D.S. and Moore, D.(1996) : *Multimedia for teaching statistics : promises and pitfalls*, **The American Statistician** , 3 , 217-225

West, R.W., Ogden, R.T: (1998): *Interactive Demonstrations for Statistics Education on the World Wide Web*, **Journal of Statistics Education**, http://www.amstat.org/publications/jse/v6n3/west.html

CD-ROM

F.Kazi-Aoual, M.Janvier (1998) *Techniques de la Statistique : Statistique descriptive, Calcul des probabilités, Variables aléatoires, Echantillonnage, Estimation*, ouvrage collectif CNAM/CRA Languedoc-Roussillon, AGRO de Montpellier, édité par Jeriko, collection CNAM-Média.

Tableau de l'EconomieFrançaise (1999), CRDP du Languedoc-Roussillon, INSEE.

L'IREM, CENTRE DE RESSOURCES

Luc TROUCHE

Directeur de l'IREM.

Résumé : *La politique d'édition de l'IREM de Montpellier n'est qu'un des volets d'une activité visant à mettre à la disposition des enseignants des ressources utiles pour leur enseignement, leur formation et leurs recherches. Cette activité comporte trois volets : l'animation d'un centre de documentation, la construction d'un site Internet et la publication de documents.*

1) Un centre de documentation

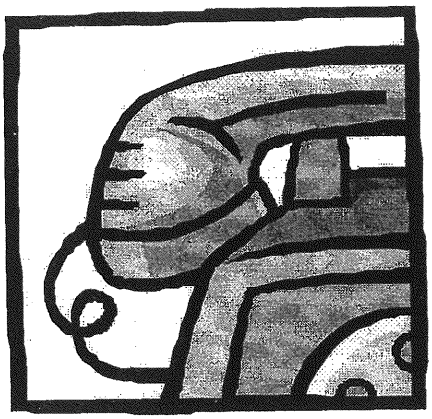
Tous les enseignants de mathématiques et les étudiants de l'IUFM ont la possibilité d'emprunter ou de consulter durant les heures d'ouverture (lundi à vendredi, 9h-12h et 13h30-17h) une documentation riche et variée sur toutes les recherches (en particulier, celles émanant du réseau des IREM) concernant l'enseignement des mathématiques au niveau national et international. La base de données de Publimath (http://publimath.irem.univ-mrs.fr/), permettant une recherche par thème ou mot-clé, est accessible à partir d'un ordinateur de l'IREM.

Il est possible de consulter les revues suivantes : le BOEN, La lettre du monde de l'éducation, La lettre de l'éducation, Le monde de l'éducation, le bulletin de l'APMEP, Educational Studies in Mathematics, For the learning of Mathematics, International Journal of Computer for Mathematical Learning, Grand N, Petit x, Plot, Quadrature, Tangente, Hypercube, Recherche en Didactique des Mathématiques, Repères.

Il est aussi possible d'emprunter des films, en particulier ceux de la Cité des Sciences : « L'esprit des mathématiques », « Le village des mathématiciens », « Tambour, que dis-tu ? », « D'une nature fractale », « L'ironie du sort », « La nouvelle étoile du berger ».

2) Un site Internet

Son adresse : http://irem-univ-montp2.fr. On y trouve des informations sur les équipes de recherche de l'IREM, mais aussi des activités extraites de publications, les productions des G.E.R. (groupes d'étude et de recherche), des expérimentations pédagogiques en collèges et lycées, des informations sur la formation continue dans l'Académie (les stages suggérés dans Epsilon peuvent y être retrouvés), des sites à visiter (en particulier ceux des autres IREM, ceux de l'IUFM ou du Rectorat), des logiciels à télécharger.



On y trouve aussi des informations sur l'actualité mathématique (conférences et colloques). Prochaine échéance importante pour l'IREM de Montpellier : un colloque national sur l'enseignement de la géométrie en collège, les 21, 22 et 23 juin à la Grande-Motte.

3) Un centre d'édition et de diffusion de documents

Les recherches entreprises à l'IREM débouchent sur des productions de documents, qui sont mis en forme, imprimés et diffusés.

On trouvera ci-après une présentation des brochures les plus récentes, la liste de l'ensemble des publications disponibles ainsi qu'un bon de commande.

Le réseau national des IREM publie une revue, Repères-IREM.

Abonnements (quatre numéros par an) : Etablissements : 250F Particuliers : 200F

Bulletin d'abonnement à renvoyer à :

Topiques-Editions, 3 Place Jeanne d'Arc, 57000 Metz

Sommaire du numéro de janvier 2001 :

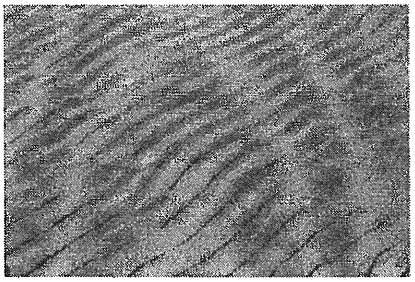
- *Quelques problèmes pour donner du sens au calcul littéral.*
- *Quelle place pour l'aléatoire au collège ?*
- *Utilisation du « et » et du « ou » en mathématiques.*
- *Derrière la droite, l'hyperplan.*
- *Sur la géométrie élémentaire du triangle dans le plan complexe.*

PUBLICATIONS RÉCENTES

Des statistiques à la pensée statistique

IREM de Montpellier, février 2001 (100F + frais de port)

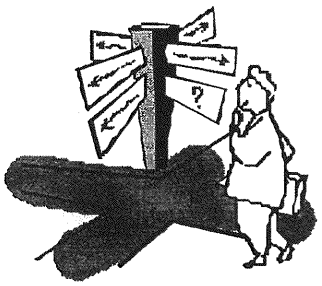
Cette brochure présente un ensemble de ressources relatives au nouveau programme de statistiques en lycée : une présentation générale des programmes, un point de vue plus général sur « la pensée statistique », un point de vue historique, didactique, mathématique, ainsi que des activités pour les classes et des points de repères bibliographiques et sur Internet.



Fonctions de l'écrit dans la classe de mathématiques

BRONNER A., PELLEQUER S., novembre 2000 (50F + frais de port)

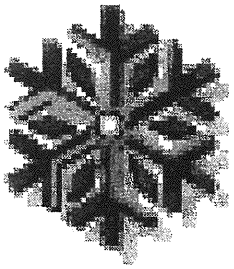
Cette brochure présente une étude didactique sur la place et le rôle des différents types d'écrits dans l'enseignement et l'apprentissage des mathématiques en classe de troisième. L'étude propose des éléments de réponse à ces questions à travers des ingénieries développées sur deux thèmes mathématiques : la racine carrée et la géométrie, en lien avec l'apprentissage de la démonstration.



L'algèbre au lycée et au collège

Coordonné par l'équipe didactique, octobre 200 (50F + frais de port)

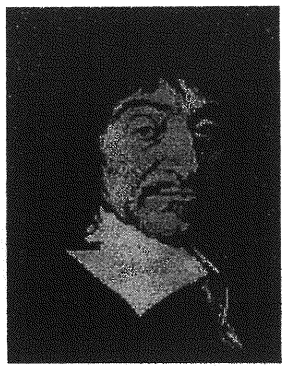
Cette brochure contient les conférences et les ateliers des journées de formation de formateurs que l'IREM de Montpellier organise chaque année.
Elle tente de répondre aux questions relatives à la nature de l'algèbre et de son enseignement : l'algèbre est-elle un langage ? Quelles sont les compétences algébriques à construire chez un élève ? Quels sont les apports des nouvelles technologies : tableurs, logiciels de calcul formel, Internet ? Quelle évolution de l'enseignement de l'algèbre peut-on envisager dans les nouveaux environnements de calcul ?



Les unités de méthodologie en DEUG

BELHAJ D., BONAFE F., CUER D., FAURE C., GANNOUN A., KIEFFER F., SABY N. & TISSERON C., octobre 2000 (50F + frais de port)

Cette brochure contient les propositions de l'équipe IREM « liaison lycée-université » pour un renouvellement de l'enseignement en DEUG. Ce renouvellement s'inscrit dans le cadre de la mise en place de la réforme des premiers cycles universitaires, à la rentrée 2000 à Montpellier 2. La réflexion porte principalement sur les questions de méthodologie : faut-il enseigner les méthodes en mathématiques ? Quels types de méthodes et comment ?
Les propositions de renouvellement s'accompagnent d'idées d'activités sous forme de TPE.

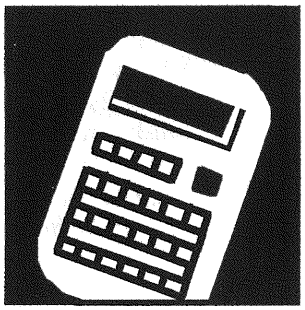


Pour une prise en compte des calculatrices symboliques en lycée

BERNARD R., FAURE C., NOGUES M., NOUAZE Y. et TROUCHE L., juin 1998 (70F + frais de port)

Pour permettre une intégration raisonnée des calculatrices graphiques et symboliques dans le cours de mathématiques, il est nécessaire de repenser celui-ci, aussi bien en termes de contenu que de pratique.
Cette brochure propose des pistes en ce sens :

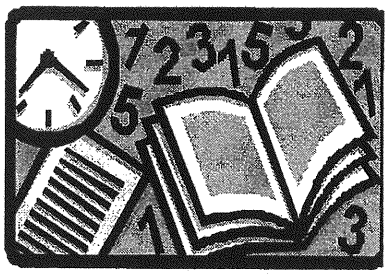
- présentation de la structure de l'outil ;
- analyse des effets de la calculatrice, dans ses différentes applications, sur les objets mathématiques représentés ;
- présentation de dispositifs de travail possibles dans la classe ;
- présentation de différentes leçons du programme d'analyse de lycée, adaptée à ces nouveaux environnements.



Pour mieux connaître les décimaux

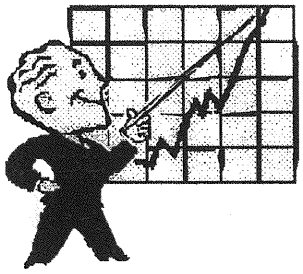
Equipe didactique, janvier 1999 (25F + frais de port)

Ce fascicule contient l'élaboration et le déroulement, dans une classe de CM2, de cinq séances sur le thème de la recherche des dimensions de rectangles de périmètre donné.
L'objectif des auteurs est de montrer qu'il est possible, sous certaines conditions, d'organiser, dans une classe de ce niveau, une véritable activité de résolution de problème. Les lecteurs pourront y trouver des idées pour organiser de telles situations de recherche, et de quoi alimenter une réflexion sur leurs pratiques.



Analyse de données et proportionnalité à l'aide d'un tableur au collège et au lycée

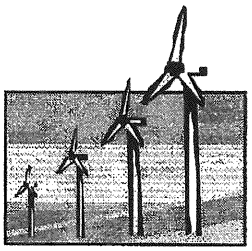
BELLAY M., COUDERC G., GUILHAUMOU D., JANVIER M. & SALA M., novembre 1998 (50F + frais de port)



Ce document présente un certain nombre de notions statistiques usuelles à l'attention des enseignants des collèges et des lycées professionnels. Il propose pour chacune de ces notions des activités utilisant le tableur.
Les enseignants y trouveront des propositions pour présenter des notions et des fichiers statistiques permettant de mettre en œuvre les concepts proposés à l'aide d'une exploitation informatique.La brochure est accompagnée d'une disquette contenant 11 fichiers, chacun possédant une feuille d'énoncé, une feuille de correction et éventuellement une feuille graphique.

Enseigner les transformations

Equipe géométrie, avril 1999 (50F + frais de port)



Ce document a été réalisé à partir d'une réflexion sur l'enseignement des transformations et d'expérimentations réalisées en classe.
Une première partie est consacrée aux isométries (programme du collège), introduites par le mouvement puis à partir d'algorithmes de construction. Cette première partie est complétée par un travail avec le logiciel "Cabri-Géomètre". Une deuxième partie, consacrée à la classification des isométries, se termine par un panorama des transformations du plan. Tous les chapitres sont accompagnés de fiches d'exercices utilisables en classe.

Histoire de constructions... Logarithmes, nombres réels

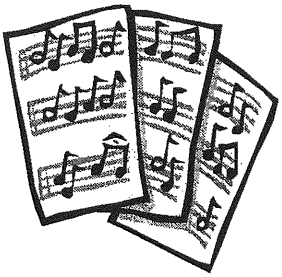
JABOEUF F., LALANDE F. & RAVEL D., juillet 1999 (50F + frais de port)



Une partie de l'ouvrage est consacrée à l'invention des logarithmes, l'autre aux diverses constructions des nombres réels, en rappelant dans les deux cas les conditions historiques de leur apparition.

Activités pour le cycle central : des nombres et des lettres,de l'observation au raisonnement

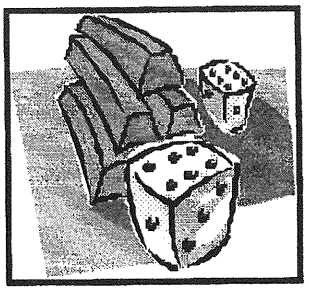
Equipe « nouveaux programmes de collège », juin 1999 (50F + frais de port)



Ce document concerne des activités sur les notions qui paraissent essentielles dans les programmes de 1997. Le choix des auteurs s'est porté sur deux thèmes, "démonstration" et "calcul littéral", qui sont des thèmes transversaux des programmes. Chacune des activités proposées a été utilisée en classe par les enseignants qui ont rédigé ce document. Pour chacune d'entre elles, sont mentionnés les objectifs qu'elle vise, le détail de son organisation en classe et ses prolongements possibles.

Fragments d'arithmétique

BERNARD R., BRIANT N., FAURE C., FONTANA J., NOGUES M. & TROUCHE L., septembre 1999 (70F + frais de port)



A la suite de la réintroduction de l'arithmétique en classe de Terminale S (spécialité mathématique), notre équipe a été amenée à reprendre une réflexion déjà engagée sur cette partie des mathématiques lors de ses stages de formation. Nous reprenons ici les différents thèmes qui ont été abordés lors de ces stages : divisibilité, rapports discret/continu, nombres rationnels, avec des idées d'applications pratiques (cryptographie).

Topologie : synthèse de cours et démonstrations

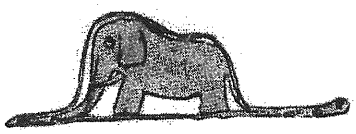
DUCOS A., novembre 1999 (70F + frais de port)



L'ouvrage propose une synthèse des notions de topologie générale et métrique accompagnée de seize démonstrations de théorèmes.
L'ensemble est commenté et étayé de nombreux exemples pour faciliter l'assimilation des concepts.

Expérimenter et prouver : faire des mathématiques avec des calculatrices symboliques

TROUCHE L. et 37 élèves d'une classe de TS, mai 1998 (100F + frais de port)



Une classe de terminale S a été dotée de calculatrices comportant un logiciel de calcul symbolique et un logiciel de géométrie. Le professeur et les 37 élèves ont contribué à cet ouvrage, comportant :

- des TP accompagnés de pistes de résolution ;
- des défis mathématiques, avec les différentes solutions proposées par les élèves ;
- les sujets des devoirs surveillés, avec les problèmes spécifiques posés par l'évaluation des travaux avec calculatrices ;
- l'évaluation que les élèves eux-mêmes font de cette année expérimentale.

Fiche signalétique de l'ouvrage

Titre : DES STATISTIQUES A LA PENSEE STATISTIQUE

Auteurs : BASCOU N. - BERNARD A. - COMBES M.C. - DUPERRET J.C. - FONTANA J. - GANNOUN A. - HENRY M. - JANVIER M. - JOZEAU M.-F. - LACAGE M. - NOGUES M. - RAVIER J.-M. - ROCHE M. - SABY N. - SALLES J. - SAUTER M. - SECO M. - TROUCHE L. - VERGNE C.
Coordonnée par l'IREM de Montpellier.

Editeur : IREM de Montpellier.

Mots-clés : MODELISATION - PROBABILITES - SIMULATION - STATISTIQUE.

Résumé : Les articles de cette brochure sont des points d'appui pour approfondir la réflexion sur l'enseignement des statistiques et des probabilités. Ils viennent du réseau des IREM (Jean-François DUPERRET, de l'IREM de Reims, Michel HENRY, de l'IREM de Besançon et Marie-Françoise JOZEAU, de l'IREM de Paris 7). Ils viennent aussi de l'ensemble des équipes de l'IREM de Montpellier : les équipes « statistiques » bien sûr, mais aussi « analyse », « didactique », « informatique » et « intégration des logiciels » ont illustré à leur façon les nouveaux programmes d'enseignement, à partir de leur propre domaine de recherche.

Cette brochure comporte quatre grandes parties :

- deux points de vue généraux sur l'enseignement des statistiques, épistémologiques et historique ;
- une présentation des nouveaux programmes, suivie de deux points de vue didactiques ;
- des idées d'activités pour la classe avec des prolongements mathématiques ;
- des références bibliographiques pour aller plus loin.

En matière d'enseignement de la statistique et des probabilités, les besoins de formation continue et la nécessité de la recherche didactique sont patents. Cette brochure est une contribution à cette formation et à cette recherche.

Nbre de pages : 222

N° ISBN : 2-909916-42-1

Prix public : 100F