La loi de Benford et la pratique d'enquêtes : un exemple de situations didactiques en statistique

Jean-Claude Oriol, Jean-Claude Regnier

▶ **To cite this version:**

Jean-Claude Oriol, Jean-Claude Regnier. La loi de Benford et la pratique d'enquêtes : un exemple de situations didactiques en statistique. XXXVIèmes JOURNéES DE STATISTIQUE -SFDS, May 2004, Montpellier, France. halshs-00408259

HAL Id: halshs-00408259
https://shs.hal.science/halshs-00408259v1

Submitted on 29 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LA LOI DE DE BENFORD ET LA PRATIQUE D'ENQUÊTES : UN EXEMPLE DE SITUATIONS DIDACTIQUES EN STATISTIQUE

Jean-Claude Oriol* & Jean-Claude Régnier**

*Institut Universitaire de Technologie Lumière
Université Lumière, Lyon II F-69676 BRON Cedex
Centre d'Etudes, de Recherche et de Recherche-Action (CERRAL)
Jean-Claude.Oriol@univ-lyon2.fr

** Institut des Sciences et Pratiques d'Education et de Formation ISPEF
Université Lumière Lyon II 16 Quai Claude Bernard F69007 Lyon
EA648 "Savoirs, Diversité et Professionalisation"
Jean-Claude.regnier@univ-lyon2.fr

Résumé
Dans cet article, nous décrivons et analysons une activité d'enseignement de la statistique destinée à des étudiants de deuxième année d'IUT (Institut Universitaire de Technologie) que nous avons centrée à la fois sur la loi de Benford et sur une pratique d'enquêtes par questionnaire. Par cette analyse, nous essayons aussi d'expliciter cette activité par l'approche de la théorie des situations didactiques développée par G. Brousseau.
***MOTS CLÉS** : didactique de la statistique, apprentissage, théorie des situations didactiques*

Abstract
In this article we describe and we analyze practices in statistic education for students in second year in university (precisely in Institut universitaire de technologie) between Benford probability distribution and investigations and survey made by students. In this situation, we research the relations between the genesis of statistic and the theory of didactic situations of Guy Brousseau.

***KEYWORDS** : statistical education, learning, theory of didactic situations*

Dans l'activité d'enseignement de la statistique comme dans celui des mathématiques, l'intention de l'enseignant reste de faire apprendre quelque chose, ce qui est le sens de la situation didactique qu'il organise. Pour qu'un apprentissage puisse être visé, il est clair que l'étudiant ne doit pas connaître la réponse au problème posé. A la différence des situations didactiques telles qu'elles sont définies en didactique des mathématiques, l'enseignant peut lui aussi ne pas connaître la réponse pertinente au problème sur lequel est centré une situation didactique en statistique. En effet les questions auxquelles se confrontent l'apprenant peuvent ne pas être énoncées explicitement *a priori* et leur formulation est laissée à la charge de ce dernier. Pour explorer le fonctionnement didactique, nous avons proposé deux situations qui, bien que ne mettant pas en jeu les mêmes objets : la loi de Benford et une enquête par questionnaire, se fondent sur des données construites par une procédure statistique explicitée par l'étudiant. Le but est alors d'exploiter ces données statistiques pour inférer une loi théorique susceptible de bien modéliser la loi empirique de la variable statistique du phénomène étudié.
En fait l'apprenant est alors confronté à une prise de décision en situation incertaine, c'est à dire à mobiliser le raisonnement statistique étayé par des outils connus des statisticiens mais objets d'apprentissage pour l'étudiant.

1. Petit retour sur une étrange distribution de probabilité : la loi de Benford

1.1. Présentation de la loi de Benford

Un « étrange » phénomène [1] [2] a été observé depuis plus d'un siècle : si l'on «collecte » des nombres au hasard les premiers chiffres de ces nombres n'ont pas la même probabilité d'apparition ; dans des domaines aussi différents que les cours de la bourse, les prix présents dans des catalogues, les constantes en physique, on retrouve le même phénomène : il y a plus de nombres commençant par le chiffre 1 que par le chiffre 2, etc., et

les fréquences sont en ordre décroissant jusqu'au chiffre 9 (si l'on est en base dix). Dans un article de 1881, Simon Newcomb exposait ses remarques à propos d'un phénomène portant sur l'usage des tables de logarithmes qui le conduisait à penser que la distribution de fréquences de l'usage des volumes et des pages n'était pas uniforme. Il proposa alors une loi de probabilité de la variable K donnant le premier chiffre significatif d'un nombre pris dans un ensemble quelconque ; elle était régie par la formule suivante : $\mathrm{Prob}\{K = k\} = \log_{10}(1 + \frac{1}{k})$ avec k=1,…,9 en base 10. En 1938, Franck Benford [3] aboutit aux mêmes observations et popularisa ce qui devint alors la loi de Benford.

1.2. Situation didactique centrée du la loi de Benford

Nous avons donné aux étudiants un échantillon de catalogues récoltés dans les divers magasins de commerces voisins de l'IUT. Nous leur avons demandé ce qu'ils pensaient de la fréquence du premier chiffre des nombres contenus dans les catalogues. Quelques-uns ont spontanément conjecturé l'équiprobabilité de l'apparition. Cependant d'autres ont considéré que la question ainsi posée induisait l'idée qu' « il devait y avoir quelque chose à trouver ». Partant de là, nous avons engagé collectivement les étudiants dans la saisie sous tableur Excel. Par deux ils ont saisi une partie des données et puis tous les fichiers ont été mis en commun. Ensuite, ils ont établi une représentation graphique de la distribution qui semblait alors bien contredire l'hypothèse d'équiprobabilité. Il s'agissait alors de décider en recourant à un raisonnement statistique étayé. Les étudiants ont alors proposé de recourir à la procédure du test d'adéquation (Khi deux à 9 ddl) à partir de laquelle ils ont conclu au rejet de l'hypothèse d'équiprobabilité au seuil de 0.01.

2. Une enquête par questionnaire réellement conduite par les étudiants.

Partant du constat paradoxal d'un certain désintérêt chez les étudiants du département de GEA à l'égard de l'enseignement de la statistique, nous avons engagé ceux-ci dans une tâche de conduite d'une enquête par questionnaire en vraie grandeur. Cela a requis une organisation pédagogique adaptée compte tenu du fait que les études universitaires dans cet IUT sont conduites intégralement dans un système en alternance. Notons que d'autres projets d'enquêtes institutionnalisées avec des partenaires extérieurs à l'IUT sont présents dans d'autres départements de l'IUT, c'est le cas en particulier dans le département GLT (Gestion Logistique et Transport) [4] et dans le département STiD (Statistique et traitement informatique des données).
Pour revenir au travail qui nous intéresse, officiellement, dans leur cursus d'étude, les étudiants sont chargés de réaliser une enquête par questionnaire. Cette tâche est assumée par groupe de deux étudiants. En raison du fait que le temps passé en entreprise est à géométrie variable entre première et deuxième année, la durée de l'enseignement de la statistique est ainsi répartie. En première année, une quarantaine d'heures est consacrée à l'étude de la statistique descriptive. Une part de la validation est réalisée par une évaluation basée sur la réalisation d'une enquête par questionnaire concentrée sur plusieurs disciplines (communication, marketing, statistique, etc.) et réalisée en groupes. En deuxième année, vingt heures sont consacrées à travailler sur les notions de lois de probabilités, de tests statistiques d'hypothèses, d'échantillonnages, d'estimations, d'intervalles de confiance. Nous avons pris le parti de consacrer les quatre dernières heures à une tâche didactique centrée sur la réalisation d'une d'enquête par questionnaire.

2.1. Organisation de la situation didactique centrée sur l'enquête par questionnaire

Les consignes suivantes données aux étudiants, regroupés par groupes de 4 au maximum, régissent la tâche à réaliser :

Phase 1

- *Déterminer un objet d'intérêt dont l'exploration requiert de manière pertinente une enquête par questionnaire.*
- *Déterminer une (ou plusieurs) question qui explicite ce que vous souhaiteriez connaître sur cet objet.*
- *Déterminer la population concernée par cet objet qui constituera la population de votre étude.*

Le projet rédigé qui ressort de cette première phase fait l'objet d'un contrôle par l'enseignant et requiert son accord pour la poursuite.

Phase 2

- *Construire un questionnaire adéquat à la situation d'étude et constitué de questions pertinentes.*
- *Expliciter par anticipation les traitements statistiques qui seront mis en œuvre.*

- *Mettre en forme le questionnaire sous traitement de texte.*
- *Expliciter le protocole de passation.*

Le questionnaire et le protocole de passation issus de cette seconde phase font encore l'objet d'un contrôle par l'enseignant et son accord est requis pour la poursuite.

Phase 3

- *Faire passer le questionnaire pour la construction des données.*
- *Saisir les données sous un tableur.*
- *Débattre avec l'enseignant sur les premières hypothèses avancées.*

Phase 4

- *Réaliser un rapport construit et rédigé d'environ dix pages, relatant cette étude et ses résultats qui doit être remis à une date précise.*

Ce rapport fait l'objet de l'attribution d'une note par l'enseignant.

2.2. Exemples de quelques thèmes abordés dans l'enquête

Les thèmes sont librement choisis par les étudiants parmi leurs centres d'intérêt. Il est à noter qu'aucun thème n'a jamais été refusé. Parmi les nombreux thèmes proposés, nous pouvons citer :

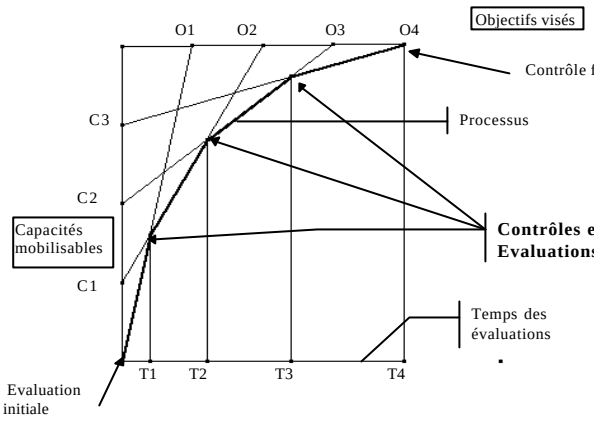
- *Le budget des étudiants : Quelle est la part du budget des étudiants accordée aux vacances et quelles sont leurs destinations ? Quelle est la distribution du budget global mensuel suivant les postes : logement, transport, alimentation, sport et loisirs, habillement [de la population des étudiants de Lyon 2] Origine et répartition du budget des étudiants de Lyon2.*
- *Les étudiants et le sport : comportement des étudiants face au sport.*
- *Les étudiants et le temps libre : que font les étudiants de leur temps libre ?*
- *Influence des médias sur les choix politiques.*
- *Mode de transport des étudiants de Lyon2 en rapport avec diverses variables : sexe, budget, etc.*
- *Influence de l'âge sur le mode d'habitation.*
- *Comportement tabagique des étudiants et pratique du sport.*
- *Les étudiants et les discothèques.*

2.3. Quelques remarques immédiates issues des enquêtes réalisées.

Population étudiée : pour des raisons de commodité la population étudiée est celle des étudiants de l'université Lyon2 et quelquefois celle de l'IUT Lumière.
Types de thèmes : nous pouvons remarquer qu'il y a eu très peu d'enquêtes centrées sur des questions d'ordre politique et même aucune sur la santé.
Traitement des données : l'anticipation concernant ce traitement est pratiquement toujours présente.
Nature et type des variables : un questionnement n'est présent que dans la moitié des rapports.
Compétences collatérales : la maîtrise des outils « traitement de textes » et « tableur » s'est avérée largement suffisante.
Échantillons : Beaucoup de groupes prévoient d'interroger plus de personnes que nécessaire afin d'avoir un peu de marge.
Données aberrantes : Peu de groupes s'interrogent sur la possibilité d'apparition de données aberrantes.
Représentativité : Quelques groupes s'inquiètent de ne pas avoir dans leur échantillon 50% de garçons (en fait la population des étudiants de Lyon2 est composée d'une forte majorité d'étudiantes…).

3. La régulation du parcours d'apprentissage

Nous insistons sur l'importance des phases de dialogue à l'intérieur du groupe d'étudiants mais aussi entre ce groupe et l'enseignant. Ces moments durant lesquels peut apparaître le processus du conflit socio-cognitif, sont autant de moments d'évaluation formatrice à visée régulatrice qui donnent l'occasion aux étudiants de fixer de nouveaux objectifs, donnant à leur parcours d'apprentissage une forme déjà explicitée [5] que nous schématisons ainsi :



Parcours d'apprentissage (Oriol,1999)

De cette schématisation, il ressort que les moments indiqués par les points « Contrôles et évaluation » sont d'une extrême importance. En effet, à chaque étape, les objectifs se sont déplacés. Cet échange apprenants-enseignant permet de faire le point dans le temps et dans l'espace du processus d'apprentissage, de se fixer des nouveaux objectifs, de nouvelles directions, de prendre conscience d'une manière métacognitive, du chemin déjà parcouru. C'est par cette vision cinétique du parcours d'apprentissage que l'apprenant peut affronter les obstacles. Nous avons ici ce que nous avons développé à propos de la morphogenèse des parcours d'apprentissage [5] (Oriol, 1999).

Dans cette approche, nous retrouvons l'importance du processus de tâtonnement expérimental [6] (Régnier 1998) et ceux d'auto-évaluation et d'autocorrection dans une perspective développée par Régnier [7] (Régnier 2000). Nous avons introduit des outils autocorrectifs et auto-évaluatifs mis au point par Régnier [8] (Régnier, 1994, 1995) qui donne l'occasion d'étayer le développement de l'autonomie des étudiants par rapport aux apprentissages.

4. L'apport de la théorie des situations didactiques de Brousseau à la didactique de la statistique

La théorie des situations didactiques a été développée par Guy Brousseau [9] (Brousseau 1974) dès les années 70 à partir de travaux conduits sur des questions relatives à l'enseignement et l'apprentissage des mathématiques à l'école primaire constituant ainsi les fondements du champ de la didactique des mathématiques. Pour ce qui concerne la didactique de la statistique [10], il a abordé la question des situations didactiques en statistique [11] (Brousseau 2003) en revenant sur une expérience ancienne connue sous le nom de « la bouteille ». Dans cette expérience didactique, les élèves dans une classe de l'école primaire sont confrontés à la situation problème suivante : une bouteille opaque contenant cinq billes blanches ou noires leur est présentée. Un dispositif permet en renversant la bouteille de faire apparaître une bille dans le bouchon transparent. On demande aux élèves quelle est la composition de l'ensemble des billes contenu dans la bouteille.

Sur la plan didactique, notre question est alors de savoir dans quelle mesure les deux situations didactiques présentées à savoir celle centrée sur loi de Benford et celle centrée sur une question mettant en jeu la pratique d'une enquête statistique par questionnaire, relèvent de la théorie des situations didactiques.

Pour Brousseau, parlant à l'école d'été de didactique des mathématiques [11] (Brousseau 2003) de situations didactiques en statistique :

« Une situation est fondamentale du premier type si elle vise à fournir un modèle qui, par le jeu de ses variables et de leurs limitations, peut convenir à n'importe quelle situation où cette notion intervient. Exemple : une situation statistique est une situation où un sujet essaie de reconnaître un objet dans une collection de situations fluctuantes et/ou avec un répertoire incertain.

Une situation est fondamentale du deuxième type si elle vise à servir de référence, à représenter symboliquement au besoin, ce qui est essentiel dans les objets et dans leurs relations, de façon à pouvoir y rattacher des situations effectives par « le sens », par des « représentations » ou par des transformations diverses. Exemple : celle qui est finalement présente dans l'esprit des élèves à la fin du processus « comment se convaincre de la vérité d'un fait qui reste incertain en répétant des prises d'informations ».

Une situation est fondamentale du troisième type si elle peut engendrer un processus qui aboutit à la connaissance de la notion par le jeu des questions qu'elle conduit à se poser, et des réponses qu'elle appelle. La situation « combien de boules blanches dans cette bouteille » appartient à ce type. Elle aboutit à la connaissance

d'une situation du deuxième type. »
Bien évidemment la situation d'enquête par questionnaire effectuée par les étudiants relève du troisième type de situations pointées par Brousseau alors que le travail sur la loi de Benford, dans la mesure où un prolongement pourrait être la découverte de la loi appartient au deuxième type de situations. Mais ces deux situations ne doivent pas nous faire perdre de vue combien les obstacles à l'apprentissage de la statistique restent nombreux. Un travail minutieux de recherche reste à poursuivre pour identifier les plus fondamentaux. Pour Brousseau certains de ces obstacles sont d'ailleurs à rechercher dans la situation scolaire ou universitaire même *« …les difficultés proviendraient des représentations (au sens de la sociologie) que se font ces institutions de la façon de pratiquer, d'apprendre ou d'enseigner ce savoir en le reliant aux mathématiques…»* [11] (Brousseau, 2003). Pour surmonter ce type d'obstacle, un travail délicat convoquant plusieurs disciplines dans une fragile catalyse est à mener, pour permettre au plus grand nombre d'étudiants un réel apprentissage de la statistique.

Conclusion

On le voit, la théorie des situations didactiques appliquée à la statistique apporte un éclairage pour une meilleure compréhension des situations problèmes à mettre en place avec les étudiants. Les apprentissages ne doivent pas se limiter aux conditionnements issus d'une suite d'exercices d'entraînement qui, s'ils sont propices à la préparation immédiate des contrôles, ne peuvent à eux seuls développer l'esprit statistique fondé sur le questionnement inhérent à toute étude statistique.
Dans cette perspective, le « inquirere », ou « chercher à découvrir » d'où vient le mot « enquête », trouve ainsi une place centrale dans l'enseignement de la statistique qui ne peut être coupé de la pratique même de la statistique.

5. Bibliographie

[1] Le premier chiffre significatif fait sa loi, *La Recherche* 01/1999
[2] http://www.math-info.univ-paris5.fr/smel/articles/benford/benford.html *Le premier chiffre à gauche*, Roman Mélanie & Robert Claudine
[3] Benford F. (1938), Proceedings of the American Philosophical Society, 78 , 551,
[4] Le Nir M., Oriol JC, (2000), L'enquête : élément central de l'enseignement de statistique en première année GLT in CNR'IUT 2000 Recherche et innovation, Presses Universitaires d'Orléans, pages 157 à 169.
[5] Oriol J-C, (1999), "L'évaluation en alternance", in Actes du colloque sur l'alternance, IUT Lumière, Lyon.
[6] Régnier, JC, (1988) Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant, Annales de Didactique et de Sciences Cognitives, séminaire de Didactique des Mathématiques de ULP Strasbourg , pp 255-279
[7] Régnier, JC, (2000) Auto-évaluation et autocorrection dans l'enseignement des mathématiques et de la statistique. Note de synthèse pour l'obtention de l'Habilitation à Diriger des Recherches. Université Marc Bloch (13 décembre 2000) 240 p.
[8] Régnier, JC, (1994, 1995) *Statistique : documents autocorrectifs et auto-évaluatifs d'aide à l'apprentissage*, livret n°1, 2, 3, 4, Lyon : APPRENDRE-Sciences de l'éducation Collection «outils de base pour la recherche en éducation», Université Lumière,
[9] Brousseau Guy, G. (1974). "L'enseignement des probabilités à l'école élémentaire" in Compte-rendu de la XXVIe rencontre de la C.I.E.A.E.M. IREM de Bordeaux.
[10] Brousseau G., Brousseau N., Warfield Virginia, (2002) "An experiment on the teaching of statistics and probability" Journal of Mathematical Behavior, 20 (2002) 363-441
[11] Brousseau G., (2003), "Cours de l'école d'été de didactique des mathématiques », ARDM.