



HAL
open science

Fonctionnement didactique de la simulation en statistique.

Jean-Claude Oriol, Jean-Claude Regnier

► **To cite this version:**

Jean-Claude Oriol, Jean-Claude Regnier. Fonctionnement didactique de la simulation en statistique.. 35e Journées de la Société Française de Statistique, Jun 2003, Lyon, France. pp.743-754. halshs-00407555

HAL Id: halshs-00407555

<https://shs.hal.science/halshs-00407555v1>

Submitted on 26 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.

Fonctionnement didactique de la simulation en statistique

Exemple de l'enseignement du concept d'intervalle de confiance

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*

*** Département des Sciences de l'éducation
Université Lumière Lyon II
16 Quai Claude Bernard
F69007 Lyon
Laboratoire Interdisciplinaire de Recherche sur les Curricula et les Apprentissages
Jean-Claude.regnier@univ-lyon2.fr*

RÉSUMÉ. Le fonctionnement didactique de la simulation dans l'acquisition d'un concept statistique est l'objet de cette communication. Notre propos concerne précisément la notion d'«intervalle de confiance», rencontrée par des étudiants de deuxième année de DUT (Diplôme Universitaire de Technologie). La particularité du travail présenté réside dans le fait que ce sont les étudiants qui construisent l'outil de simulation, à partir de consignes données par l'enseignant. Nous présentons la séquence didactique mise en place, les principaux obstacles rencontrés et la mesure des apprentissages réalisés par les étudiants..

ABSTRACT. The didactic functioning of simulation in learning a statistical concept is studied in this paper. We more precisely deal with the confidence interval notion, which is studied in particular by DUT (Diplôme Universitaire de Technologie) second year students in university. The particularity of the presented piece of work lies in the fact that the students themselves build the simulation tool on their teacher's orders. The established didactic sequence, the main encountered obstacles and the measurement of the results obtained by the learning students will be presented.

MOTS-CLÉS : didactique de la statistique, apprentissage, simulation, intervalles de confiance.

KEYWORDS: statistical education, knowledge, simulation, confidence intervals.

1. Introduction

Notre propos s'ancre sur la prise en compte de la place de la simulation dans la pratique de l'enseignement de la statistique et dans la champ théorique de la didactique de la statistique. La simulation ressort dans nombre d'ouvrages de référence, de directives officielles, de commentaires des programmes comme un outil privilégié qui doit être mis au service des apprenants dans les situations didactiques visant des notions du champ de la statistique. Ainsi de nombreux logiciels de simulation plus ou moins sophistiqués, élaborés à l'intention des étudiants et des enseignants sont apparus. Cette panoplie va des langages dédiés jusqu'aux outils généralistes. À ce jour le langage Java permet de produire des Applets qui offrent des possibilités considérables. Et même si notre propos n'est pas de développer ici cette idée, nous pensons que cet intérêt pour les simulations et cette insistance à les intégrer dans des séquences didactiques concernant les apprentissages statistiques sont d'une certaine manière significatifs d'une spécificité endogène au champ de la statistique.

2. Quelques distinctions concernant les simulations utilisées en statistique

Comme le signale Jamie D. Mills (Mills, 2002) on peut distinguer dans la littérature quatre manières d'envisager l'intégration de simulations dans l'apprentissage de la statistique : *" Four definitions of computer simulation methods were described in the literature reviewed for this article. One definition involved students writing their own programs (using SAS PROC IML®, say), setting up a model for a problem and investigating diagnostics for the model in seeking possible violations of assumptions. A second definition allowed students to experience similar advantages using a random number generator in Excel® or MINITAB®. Using Excel® or MINITAB®, the commands to generate the random samples and perform experiments on the model are mostly window-driven. Third, many instructors used some combination of the first and second definitions, by providing program templates that allowed students to change parameters during the experiments (commonly in SAS® or SPSS®). Finally, a fourth definition involved using commercial software packages designed exclusively for simulation purposes (for example, the "Samplings Distribution" program). The literature reviewed in this paper included journal articles that utilized all four operational definitions, although the majority of the authors reported using the latter three. "*

3. Le contexte de l'IUT Lumière et de la statistique dans le département GEA

3.1. L'alternance à l'IUT Lumière

Depuis sa création en 1992 l'IUT Lumière est le seul IUT dans lequel la formation des étudiants est complètement réalisée en alternance. Cette formation est ainsi

construite : en première année les étudiants suivent un stage de sept semaines dans une entreprise pendant la période avril-mai, la deuxième année est réellement en alternance sur la base d'un contrat d'apprentissage. Deux modalités sont en place en GEA selon l'option choisie en deuxième année par les étudiants 2 jours + 3 jours, 7 semaines + 7 semaines, tandis qu'une autre gestion du temps, à savoir 15 jours à l'IUT + 15 jours en entreprise, est organisée dans les trois autres départements. Dans tous les cas, l'année universitaire s'achève par la tenue des jurys et la délivrance des diplômes au mois de septembre. Cette forme d'organisation des études est usuellement désignée par «système 1+1» . On trouvera des compléments détaillés sur l'organisation des cursus dans une communication de Paul Rousset, directeur de l'IUT depuis sa création (Rousset, 1997), et plus particulièrement sur le département GEA dans un article d'Isabelle Barth (Barth, 2001) qui a dirigé le département GEA pendant plusieurs années.

3.2. L'organisation de l'enseignement de la statistique dans le département GEA

Le temps passé en entreprise est modulable entre la première et la deuxième année. Il en résulte que l'équipe enseignante a pondéré différemment les cours de statistique dans chacune des deux années. En première année, une quarantaine d'heures est consacrée à l'étude de la statistique descriptive. Cet enseignement donne lieu à deux contrôles sous la forme d'épreuves individuelles en temps limité et à une évaluation à partir d'une enquête impliquant plusieurs disciplines (communication, marketing, statistique, etc.) et réalisée en groupes. En deuxième année, seize heures sont consacrées à travailler à l'étude de la statistique inférentielle, en particulier sur les notions de lois de probabilité, d'échantillonnage, d'estimation, d'intervalle de confiance

4. Le cas particulier du concept d'intervalle de confiance

4.1 le cadre théorique

Le point de vue de la formation en alternance nous renvoie à la question fondamentale : Comment apprenons-nous ?

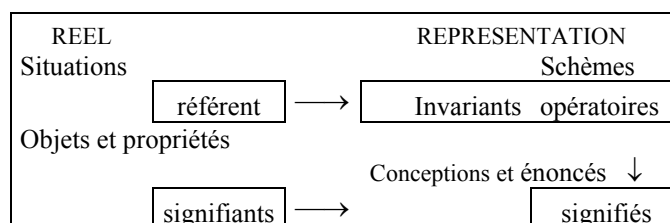
La théorie des champs conceptuels de Gérard Vergnaud (Vergnaud 1991 p.135) répond de manière satisfaisante pour ce qui touche aux apprentissages en statistique : « *La théorie des champs conceptuels est une théorie cognitive, qui vise à fournir un cadre cohérent et quelques principes de base pour l'étude du développement et l'apprentissage des compétences complexes, notamment celles qui relèvent des sciences et des techniques.(...) [et] pour la compréhension des filiations et des ruptures entre connaissances* ».

La conceptualisation, élément central de l'apprentissage, prend appui sur le concept conçu comme « un triplet de trois ensembles référence, signifié et signifiant :

- l'ensemble des situations qui donnent du sens au concept (la référence),
- l'ensemble des invariants sur lesquels repose l'opérationnalité des schèmes (le signifié)
- l'ensemble des formes langagières et non langagières qui permettent de représenter symboliquement le concept, ses propriétés, les procédures de traitement (le signifiant) »

Suivons Gérard Vergnaud (Vergnaud 1994, p.180) qui pour affiner « progressivement la définition d'un schème,..[dit]... d'abord que c'est une totalité dynamique fonctionnelle, c'est-à-dire quelque chose qui fonctionne comme une unité ; en second lieu que c'est une organisation invariante de la conduite pour une classe de situations données (l'algorithme est un cas particulier du schème) ; et en troisième lieu qu'un schème est composé de quatre catégories d'éléments : 1)des buts, intentions et anticipations ; 2)des règles d'action; 3)des invariants opératoires; 4)des possibilités d'inférences en situation. »

Revenant sur ce qu'il appelle lui-même le « triangle des linguistes », Vergnaud (Vergnaud, 1994, p. 189-190) propose le modèle suivant :



Ainsi la notion de champ conceptuel en tant qu'ensemble des situations renvoyant à l'idée de procédure, permet de situer la simulation en statistique dans sa finalité qui est de fournir à l'apprenant un éclairage sur le signifié (invariants opératoires)

Par ailleurs l'approche développée par Jean-Claude Régnier (Régnier 1988) intégrant l'apprentissage fondé sur le tâtonnement expérimental de l'apprenant offre une perspective pour analyse didactique de la simulation et relation à la résolution de problème.

4.2 Quelques travaux concernant les intervalles de confiance et la simulation

L'utilisation de la simulation pour l'enseignement de l'inférence statistique a fait l'objet de développements et d'expériences didactiques. Citons [Kennedy, Olinsky, and Schumacher \(1990\)](#) utilisant MINITAB®, [Marasinghe et al. \(1996\)](#) développant un module et un logiciel particulier, [Hesterberg \(1998\)](#) qui après diverses

expérimentations recommande un langage interactif et plus particulièrement S-PLUS®, ou chez [Pedazhur \(1997\)](#) défendant le point de vue d'une compréhension plus intuitive des notions. Enfin Batanero et Godino (Batanero, Godino 2001 p.7-10) observent que " *Puesta que la media de la muestra varia de una muestra a otra, los intervalos de confianza variarán de una muestra a otra (lo mismo ocurre con la propoción). Lo que nos dice el coeficiente de confianza es que en un porcentaje dado de muestras, el verdadero valor del parámetro estará incluido en el intervalo. "*

5. La situation proprement dite

5.1 Le choix d'Excel

Notre pratique pédagogique vise à intégrer le plus tôt possible l'outil informatique comme instrument canonique d'une pratique de la statistique. Et en matière d'outil informatique, le tableur est privilégié. Plusieurs raisons expliquent ce choix surprenant en comparaison de la puissance de certains logiciels de traitements statistiques. Les cours concernent des étudiants de premier cycle " généralistes ". Le tableur est dans ce cadre utilisé par différentes disciplines (comptabilité, gestion, mathématiques, mathématiques financières) et l'investissement sur un autre logiciel serait peu rentable.

5.2 La situation problème et la construction de la simulation

La séance didactique prend appui sur une bonne maîtrise du logiciel Excel par les étudiants et fait suite à une séance d'une heure sur les intervalles de confiance. Les consignes sont les suivantes : « *Vous devez construire un classeur Excel, sans utiliser le langage VBA, qui contient les opérations permettant de tirer au hasard 10000 nombres compris entre 0 et 1, de calculer leur moyenne. Après quoi qui permet par tirage aléatoire, d'obtenir 10 échantillons, chacun de trente nombres, de calculer l'intervalle de confiance pour la moyenne selon un niveau choisi par menu entre 90% et 99% et d'afficher les intervalles correspondants à chaque échantillon* ». Nous ajoutons que « *Cet énoncé subira des modifications dans une heure avec des consignes modifiées .»*. Par binôme, les étudiants produisent un classeur dont la feuille de visualisation a, aux variations individuelles près, l'aspect suivant :

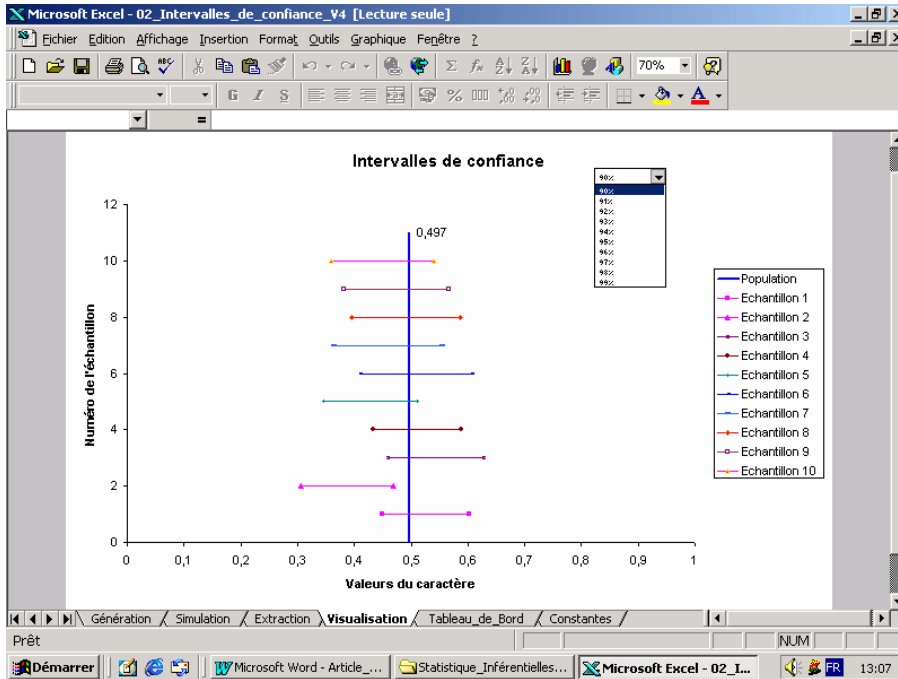


Figure 1 La simulation d'un tirage de 10 échantillons de 30 individus chacun sur une population de 10000

Les variations de l'énoncé concernent : le changement ou non de la population chaque fois que l'on tire 10 nouveaux échantillons, le choix de la taille de la population (<65536), le choix de la taille des échantillons, le nombre d'échantillons. Ce qui donne le " tableau de bord " suivant :

Population aléatoire à chaque nouveau tirage de 10 échantillons	Oui
Taille de la population	N = 10000
Taille de chaque échantillon	n = 30
Niveau de confiance	90%
Nombre d'échantillons	10

Figure 2 Tableau de bord de la simulation

6. Conclusion

Les simulations «prêtes à être admirées» emmènent le plus souvent les étudiants à une «bovarysation» des notions étudiées. Ici, en revanche, la construction par les apprenants, eux-mêmes, de l'outil s'apparente aux méthodes d'apprentissage où les premiers travaux des apprentis consistaient à fabriquer leurs propres outils (l'équerre de l'outilleur par exemple) et mettant en scène le tâtonnement expérimental producteur d'apprentissages. Parmi les problèmes en suspens reste le contrôle par les étudiants du générateur de nombres aléatoires d'Excel. A la lumière de la théorie des champs conceptuels et du modèle d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant, nous présenterons nos résultats issus de l'expérimentation et de l'observation, en particulier ceux liés à l'effet des variables d'énoncé du problème sur l'activité des étudiants et sur leur performances. Nous prolongerons nos hypothèses au comment cette activité de simulation se contextualise dans la pratique d'enquêtes par questionnaires réalisées par les étudiants dans la suite de cet enseignement

7. Bibliographie

- Albert, J. H. (1993), "Teaching Bayesian Statistics Using Sampling Methods and MINITAB," *The American Statistician*, 47, 182-191.
- Antoine F., Grootaers D., Tilman F., (1988), *Manuel de la formation en alternance*, Lyon, Chronique sociale : Bruxelles, Vie ouvrière.
- Bachelard G., (1938), *La formation de l'esprit scientifique*, Paris Vrin.
- Barth I., (2001), « L'utilisation de la situation de travail dans la formation d'un cursus diplômant, cas du DUT GEA de l'IUT Lumière », Deuxième rencontre des acteurs de la formation technologique d'Europe et d'Amérique Latine, Pointe à Pitre.
- Batareno C. et Godino J., (2001), Análisis de Datos y su Didáctica, Universidad de Granada
- Combes M.-C., (1996), coordination de *L'alternance : enjeux et débats*, Paris, La documentation française.
- Hesterberg, T. C. (1998), "Simulation and Bootstrapping for Teaching Statistics," *American Statistical Association Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, 44-52.
- Kennedy, K., Olinsky, A., and Schumacher, P. (1990), "Using Simulation as an Integrated Teaching Tool in the Mathematics Classroom," *Education*, 111, 275-296.
- 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.
- Malglaive G., Weber A., (1982), « Théorie et pratique : approche critique de l'alternance en pédagogie », *Revue française de pédagogie*, n°61, pp. 17-27.

- Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342-351.
- Mills, J. D., (2002), Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education* Volume 10, Number 1
- Oriol J-C, (1999), "L'évaluation en alternance", in Actes du colloque sur l'alternance, IUT Lumière, Lyon.
- Oriol J-C, (2002), "Réaliser une enquête par questionnaires : un outil didactique pour la statistique inférentielle à l'Université", in 3^{ème} Rencontre Europe-Amérique-Latine sur la formation technologique et professionnelle, La Havane, Cuba.
- Pedhazur, E. J. (1997), *Multiple Regression in Behavioral Research: Explanation and Prediction* (3rd ed.), Orlando: Harcourt Brace College.
- Raynal F., Reunier A., (1997), *Pédagogie : dictionnaire des concepts clés*, Paris, ESF Editeur, 408 p.
- 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
- Régnier, JC, (2000a) L'alternance: des évidences (en) aux questions. *Forum, revue de la recherche en travail social*, n°93, pp 38-47 ISSN 0988.6486
- Régnier, JC, (2000b) 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.
- Revuz A., (1980), *Est-il impossible d'enseigner les mathématiques ?*, Paris Puf.
- Ronan C., (1988), *Histoire mondiale des sciences*, Paris Seuil, 704 p.
- Rousset P., (1997), «L'IUT Lumière : une expérience d'IUT en alternance», communication présentée dans l'atelier formations professionnelles du secteur tertiaire et alternance aux journées d'études sur la professionnalisation dans l'enseignement supérieur.
- Vergnaud, G., (1991) La théorie des champs conceptuels, *Recherches en Didactique des mathématiques*, 10/2.3, Grenoble, La Pensée Sauvage Editeurs pp. 133-169.
- Vergnaud, G., (1994) Le rôle de l'enseignant à la lumière des concepts de schème et de champ conceptuel, in M. Artigues, R. Gras, C. Laborde, P. Tavnnot (EDS), *Vingt ans de didactiques des mathématiques en France*, Grenoble, La Pensée Sauvage Editeurs, pp. 177-191.