

# Approche bayésienne « cachée » et approche fréquentiste « ambiguë » dans les livres de cours français de Première S et ES

Pablo Carranza, Alain Kuzniak

Equipe Didirem. Université Paris 7  
carranza@math.jussieu.fr, alain.kuzniak@orleans-tours.iufm.fr

**Résumé.** Grâce à une étude comparative de deux livres de cours sur le thème des statistiques, nous esquissons l'espace de travail potentiel existant en classe de première en France (11<sup>e</sup> année). L'approche statistique implicite met clairement en évidence une tendance très fortement probabiliste et calculatoire en classe scientifique et une tendance plus confuse en classe économique où joue ce que nous appelons l'intrication des signifiés de la probabilité.

## 1 Introduction

Une première étude épistémologique nous a permis d'identifier une forme duale de la signification de la probabilité qu'expriment d'une part l'approche fréquentiste et d'autre part l'interprétation bayésienne. Cette dualité de signifié peut, en se manifestant dans un même problème, induire des erreurs d'interprétation dans la démarche inférentielle.

Les auteurs des programmes français actuels de lycée ont choisi de ne retenir que l'approche fréquentiste pour présenter la notion de probabilité sans aucune prise en compte de la notion bayésienne. Cependant, nous avons pu vérifier Carranza & Kuzniak, 2006 que la dualité des signifiés se manifestait de manière implicite et *cachée* dans l'enseignement. Pour mettre en évidence ce phénomène, nous avons étudié certains exercices proposés dans les livres de cours et aussi dans les sujets d'examen. Nous souhaitons revenir d'une manière plus systématique sur cette question en étudiant de manière exhaustive les exercices posés dans des livres de cours de Première ES et S (les élèves de ces classes ont entre 16 et 17 ans).

Après avoir rapidement présenté ce que sont pour nous les deux interprétations de la probabilité, nous décrirons les variables que nous avons retenues pour décrire le contenu des exercices de façon à analyser la présence d'un éventuel signifié de la probabilité et son mode de traitement proposé par les livres de cours.

Plus précisément, nous cherchons à savoir si les exercices proposés dans les livres de cours restent dans un contexte fréquentiste ou si, comme nous le pensons, ils s'en échappent et dans ce cas quel type de signifié apparaît de manière cachée ou non.

## 2 La dualité de signifié

Le concept de probabilité admet plusieurs signifiés et ceci, comme l'a montré Hacking, (Hacking, 2002, page 37) depuis son émergence à la fin du 17<sup>e</sup> siècle. Ces diverses interprétations sont réunies en deux grands groupes se rattachant à un des deux courants inférentiels, l'un fréquentiste et l'autre bayésien que nous avons retenus pour notre travail.

### Notion fréquentiste

Pour un certain type d'expériences susceptibles de se reproduire sous les mêmes conditions (au moins mentalement), la fréquence d'apparition d'un événement donné se stabilise progressivement lorsque le nombre de réalisations croît considérablement. Dans ce cas, la probabilité est la tendance du système à produire un événement donné. Cette première notion trouve sa justification dans la loi faible des grands nombres.

### Notion Bayésienne

Dans cette autre approche, la probabilité représente une mesure de crédibilité sur une proposition donnée compte tenu des informations disponibles. La probabilité n'est ainsi pas une caractéristique de l'objet mais la mesure de la crédibilité accordée par un sujet à une proposition donnée. Cette fois la justification de cette probabilité s'appuie sur la notion de probabilité conditionnelle et sur le théorème de Bayes (Bayes, 1763) : les probabilités attribuées initialement à un événement peuvent différer d'un observateur à l'autre mais elles

évolueront ensuite de manière convergente en intégrant les nouvelles informations selon les critères donnés par la formule de Bayes.

### L'intrication des signifiés

Les deux interprétations de la probabilité sont souvent étroitement reliées et enchevêtrées dans de nombreuses situations où elles apparaissent simultanément. Pour désigner ce phénomène, nous parlerons d'*intrication* des signifiés.

La principale source de cette intrication s'appuie sur le *principe fréquentiste* (Figure 1) où l'attente à long terme en tant que seule source d'information du système est considérée comme la raison fondant le degré de certitude (Droesbeke, Fine & Saporta, 2002, page 25 ; Gärdenfors et al., 1988, page 110 ; Hacking & Dufour, 2004, page 149) ainsi, lorsque on lance un dé non pipé, dont la fréquence d'apparition du trois est  $1/6$ , on pense avec une certitude de  $1/6$  que le résultat du lancer à venir sera un trois.

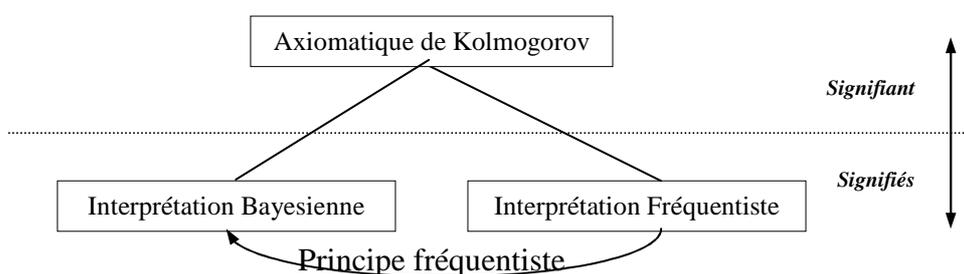


Fig. 1

### 3 La dimension cachée

Dans les programmes actuels de l'enseignement français de lycée, la probabilité n'est officiellement présentée que comme fréquentiste. Nous posons comme hypothèse que la dualité de signifié est une caractéristique inhérente à la probabilité et qu'elle se manifeste nécessairement dans l'enseignement même si officiellement il l'ignore.

Dans ce texte, nous nous interrogeons plus particulièrement sur les éventuelles manifestations de cette dualité dans les exercices proposés dans les livres de cours de Première ES et S. Nous avons choisi ces niveaux de programme car la probabilité y est introduite en relation avec la statistique ce qui nous permet d'observer le jeu entre les trois pôles que nous avons articulés dans la figure 1. La classe de ES est une section à dominante économique et sociale, les élèves ne suivront pas pour la plupart des études scientifiques par la suite à la différence de ceux qui suivent la classe de S qui est à dominante scientifique. Il est donc intéressant de voir si l'orientation différente des élèves va faire prendre en compte de manière différente la question de la dualité des signifiés et sa relation avec le modèle ensembliste de Kolmogorof.

Nous avons retenu pour notre étude un certain nombre de variables, les unes, d'ordre pédagogique, prennent en compte des points classiques dans la gestion d'un exercice comme son rôle dans les apprentissages, sa longueur, sa place dans le chapitre. Mais nous avons aussi considéré un certain nombre de variables qui sont plus étroitement liées à l'articulation entre signifiant et signifié et aussi de manière encore plus spécifique à l'interprétation en termes fréquentiste ou bayésien.

L'analyse statistique implicite (Gras et al., 1996 ; Gras & Bailleul, 2000) nous permet de dégager certaines règles d'inférence partielles du type « *si a alors presque b* » dans les exercices de ces deux livres de cours. Ainsi nous pourrions examiner des règles telles que « si un problème réunit des conditions pour une interprétation bayésienne de la probabilité alors une interprétation est demandée » ou « si une interprétation est requise alors il s'agit d'un problème fréquentiste ». La validation ou non de ces règles nous permettra de décrire l'espace de travail disponible dans les exercices du livre pour certaines interprétations de la probabilité. Il s'agit ainsi d'analyser la compatibilité ou la cohérence entre la notion de probabilité à institutionnaliser et un ensemble de situations disponibles lui donnant le sens visé.

En plus de cette étude originale de l'espace de travail disponible, cette recherche s'intègre dans toutes celles qui portent sur les difficultés éprouvées par les élèves sur le plan interprétatif d'une démarche statistique et dont

la source est la dualité de signifié (Batanero, Godino & Roa, 2004; Gärdenfors et al., 1988; Régnier & Oriol, 2001). Les variables *a priori* constitutives d'un problème ne suffisent pas à déterminer une interprétation donnée et la gestion de la situation par le professeur en classe est aussi essentielle pour déterminer l'importance de l'intrication des signifiés de la dualité. Dans notre recherche, nous avons pris en compte cette dimension en réalisant et en observant des séances de classe mettant en jeu différentes interprétations probabilistes. Nous avons choisi de nous concentrer ici sur l'espace de travail disponible que proposent les livres de cours aux élèves et aux professeurs pour lequel une étude implicite nous avait paru *a priori* intéressante.

#### 4 Les variables étudiées

Nous avons observé deux livres édités en 2001 : le livre de cours de première ES est édité par Bréal (Breal ES par la suite) et celui de Première S par Nathan (Nathan S).

Le manuel Breal ES propose pour le chapitre probabilité, trente et un exercices (page 98 à 103) répartis en trois sections. « Maîtriser le cours » : Exercices 1 à 12. « Savoir-faire fondamentaux » : Exercices 13 à 19. « S'entraîner » : Exercices 20 à 31.

Pour le manuel Nathan S, le nombre d'exercices est beaucoup plus important et les soixante-treize exercices (page 224 à 234) sont répartis en quatre sections : « Maîtriser le cours » : Exercice 1 à 20. « Pour apprendre à chercher » : Exercices 21 à 25. « Pour progresser » : Exercice 26 à 67. « Problèmes de synthèse » : Exercice 68 à 73.

Nous avons considéré 21 variables, à valeurs binaires (0 ou 1). Nous allons expliciter les critères à satisfaire pour que les variables prennent la valeur 1.

##### *Variables concernant la longueur du texte de l'exercice A :*

A la suite d'une première étude statistique quantitative visant à réunir les conditions requises pour l'application de la méthode d'analyse statistique implicite, nous avons retenu deux classes à effectifs supérieurs à 10 :

Codage	Etiquette	Description
A1	Lignes(1-10)	L'exercice a 10 lignes au plus
A2	Lignes(11-40)	L'exercice a plus de 10 lignes

##### *Variables concernant le nombre de questions B :*

En relation avec la variable précédente, le nombre des questions posées à l'élève est pris en compte par cette variable.

Codage	Etiquette	Description
B1	Questions(1-2)	L'exercice a moins de trois questions
B2	Questions((3-4)	L'exercice a trois ou quatre questions
B3	Questions((5-plus)	L'exercice a plus de quatre questions

Avec ces deux variables, nous souhaitons voir la relation entre la notion la probabilité en jeu et la taille des exercices.

##### *Variables concernant le contexte de l'exercice E :*

Cette catégorie nous renseigne sur le type de contexte proposé à l'élève lors des exercices. S'agit-il notamment d'un contexte issu des supports traditionnels du calcul probabiliste, du contexte de la classe ou d'une situation de travail dans la société. Ou alors, note-t-on une absence de contexte ou une référence à d'autres domaines mathématiques. Le regroupement est dû à des raisons d'effectifs.

E1	Contexte(Urnes-Jeux)	Urnes, jeux, etc.
E2	Contexte(Classe-Quotidien)	Référence à l'environnement de l'élève ou d'un travail
E3	Contexte(Sans ou Maths)	Absence de contexte ou contexte mathématique

##### *Variables portant sur les hypothèses du modèle F :*

Nous cherchons, à partir du statut donné aux hypothèses, à caractériser le raisonnement mis en œuvre dans le problème, pour ainsi émettre des hypothèses sur le signifié potentiel de la probabilité. Si l'on probabilise :

-dans le sens déductif (des hypothèses vers la série infinie), on sera en présence potentiellement d'une interprétation fréquentiste.

-dans le sens inductif (des hypothèses vers un élément de la série), alors il s'agira plutôt d'une probabilité bayésienne avec un possible apport du principe fréquentiste (voir Figure 1).

-dans le sens d'une abduction au sens de Peirce (de quelques évidences vers des hypothèses plausibles), alors c'est la probabilité bayésienne qui est en germe.

En classe de Première aucun outil inférentiel n'est enseigné, on ne doit s'attendre à probabiliser sur l'hypothèse d'un modèle.

F1	Hypothèses(Admises)	Hypothèses du modèle sont admises, l'ensemble des événements possibles est spécifié
F2	Hypothèses(AdmisesFréquent ou P(A))	Hypothèses admises par la proportion dans un échantillon important ou explicitement fournies
F3	Hypothèses(ADecider)	Hypothèses à décider à partir de la proportion d'un petit échantillon...

*Variables concernant la demande du calcul de probabilité H :*

Nous nous intéressons ici au statut de la notion de probabilité en tant qu'objet mathématique : le signifiant P(A) est-il demandé (explicitement ou implicitement) ou non. En croisant ces variables avec l'éventuelle demande d'interprétation (J) nous verrons les liens entre signifiant et signifié dans les exercices.

H1	CalculProbabilité(Oui)	Calcul de P est explicitement ou implicitement demandé
H2	CalculProbabilité(Non)	Calcul de P n'est pas demandé

*Variables concernant l'explicitation d'un contexte fréquentiste ou non I :*

Avec ces variables nous cherchons à renforcer l'identification de la notion de probabilité potentiellement en jeu en la confrontant aux variables F (nature des hypothèses) et K (nature de l'événement). Dans les exercices des livres de cours, nous avons repéré le phénomène suivant : en général, l'énoncé précise les éléments de l'ensemble des cas possibles puis, une action hypothétique est proposée (« on tire au hasard une carte... ») avec pour fonction de donner un contexte d'incertitude et aussi de fournir des indices sur la cardinalité de l'événement en question. Enfin, une demande de probabilité est explicitée (Calculer la probabilité que la carte tirée soit...). Lorsqu'il n'y a aucune référence à la série mais un élément de la série (la carte tirée), nous sommes dans un contexte bayésien de la probabilité. Cependant, quelques exercices proposent de vérifier les calculs effectués grâce à des simulations ou à des répétitions de l'expérimentation. Cette fois nous sommes en présence encore une fois de la manifestation du principe fréquentiste.

I1	RepetitionExper(Oui)	Mention explicite ou implicite à la répétition de l'expérience aléatoire
I2	RepetitionExper(Non)	Aucune mention

*Variables concernant la demande d'interprétation J :*

Cette variable confrontée aux variables I (répétition), F (nature des hypothèses) et K (nature des événements) nous informe sur la notion potentielle de la probabilité dans un exercice et du lien avec son signifiant (variables H).

J1	InterprProbabilité(Oui)	L'interprétation de P est explicitement ou implicitement demandée
J2	InterprProbabilité(Non)	Pas d'interprétation de P demandée

*Variables concernant la nature de l'événement à probabiliser K :*

Elles visent observer particulièrement la nature de l'événement A dans l'application P(A).

-A est une série infinie d'événements susceptibles de se reproduire : on est en présence de la notion fréquentiste.

-A est un événement dont la reproductibilité est ou bien impossible ou bien elle conduirait à modifier les valeurs données à la probabilité.

-A est un cas générique élément d'une série éventuellement infinie. Dans le cas fréquentiste cela n'a pas de sens de probabiliser sur un élément, pour nous, demander par exemple la probabilité de tirer un six lors du *prochain* lancer avec un dé équilibré relève de l'approche bayésienne et non fréquentiste.

K1	NatureA(Serie)	L'événement sur lequel on probabilise est la tendance à long terme
K2	NatureA(CasUnique)	
K3	NatureA(Cas Generique)	
K4	NatureA(Ambiguë)	

Nous avons pensé que la description d'un événement déjà arrivé permettrait de donner une mesure de certitude en fonction de l'information disponible. Ceci nous a conduit à considérer la temporalité de l'événement. Cependant la très faible occurrence de la variable passé nous a conduit à la supprimer de l'étude : la plupart des exercices décrivent l'événement en utilisant un temps du présent ou un infinitif impersonnel.

## 5 Etude des données

### Un travail intensif sur le modèle probabiliste en première S

Nous allons commencer notre présentation des résultats avec le livre Nathan qui se prête sans doute mieux à une étude statistique compte tenu de la grande abondance des exercices. Avant de commenter nos résultats, voici l'arbre cohésitif.

Grâce à une analyse au seuil de 0.90, nous avons pu retenir quatre ensembles qui vont structurer de manière significative l'ensemble des exercices.

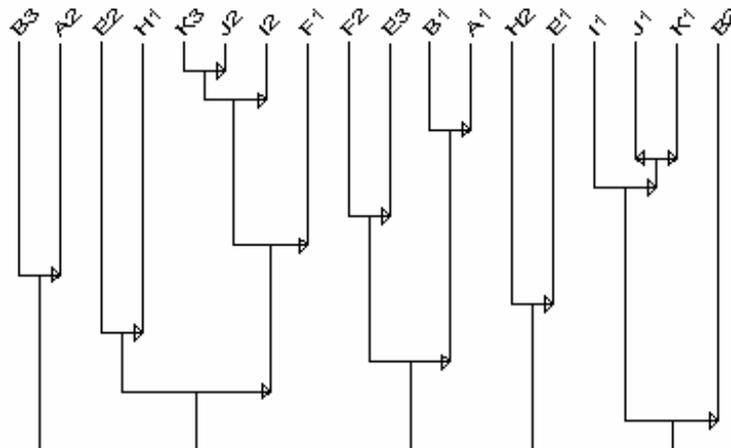


Fig. 2 Arbre cohésitif Nathan 1S

**Exercices génériques.** Le premier ensemble [K3,J2,I2,F1,(H1)] représente près des deux tiers des exercices donnés dans ce manuel et rend compte de l'exercice type auquel est confronté l'élève de cette classe. Il s'agit de raisonner sur un cas générique (K3) sans que la répétition de l'expérience soit suggérée et sans qu'aucune interprétation du résultat ne soit demandée. En complément, notons que ces exercices sont balisés par la donnée fréquente de l'ensemble des événements (E1). Les exercices prenant appui sur un contexte de classe ou de travail impliquent également la branche H1 qui demande l'explicitation d'un calcul de la probabilité.

Ainsi, l'ensemble le plus typique des exercices proposés dans cette classe se situe clairement dans le modèle probabiliste. Celui-ci est donné d'entrée et aucun lien n'est demandé avec une situation à l'origine du modèle.

**Exercices fréquentistes.** La classe [K1,I1,J1] est très homogène dans cet ouvrage et elle présente des exercices basés sur une suite (K1) avec une répétition de l'expérience (I1) et une demande d'interprétation (J1). Ces exercices peu nombreux (en fait six) correspondent à la demande du programme qui souhaite que la probabilité soit introduite à partir de la fréquence.

**Exercices d'entraînement.** Notons aussi un petit ensemble d'exercices (une dizaine) [F2,E3,B1,A1] exercices courts sans contexte et avec la donnée de la probabilité  $P(A)$ . Il s'agit d'exercices d'application sur la notion de probabilité.

**Exercices sur urnes et jeux.** La classe [H1,E1] montre seulement que les exercices avec jeux et urnes forment une entité autonome mais on sait par ailleurs qu'elle est fortement connotée là encore à l'idée de modèle. Ainsi, dans cet ouvrage qui respecte le programme, quelques exercices travaillent sur le sens fréquentiste de la probabilité mais l'ensemble apparaît comme un travail intensif dans un modèle qui privilégie le signifiant probabiliste. Si l'on suit l'approche de Peirce, ce travail intensif sur un signifiant laissé à la charge de l'élève (l'interprétant) débouchera sur un signifié profondément dépendant du jeu sur le signifiant. On peut attendre à ce que l'élève ne résume les probabilités qu'à un jeu dans le modèle laissant en jachère tout travail sur les signifiés dont nous avons parlé. Ce point sera clarifié par notre étude au niveau des BTS où nous devons nous attendre à retrouver une faible progression dans la compréhension des phénomènes liés au hasard avec des conceptions initiales assez proches de celles que Lahanier-Reuter (1999) a pu dégager chez des élèves plus jeunes.

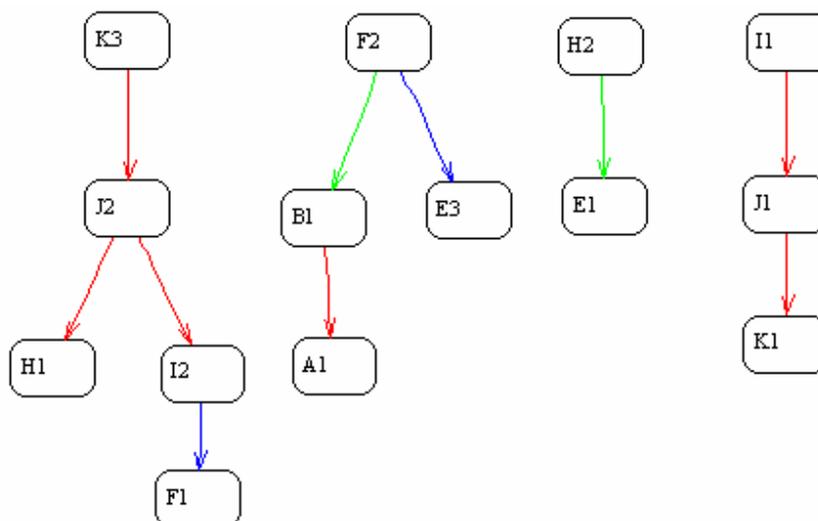


Fig. 3 Graphes Nathan S (seuil 0,9)

### Un espace de travail « flou » en première ES

Donnons maintenant notre analyse de l'ouvrage de première ES.

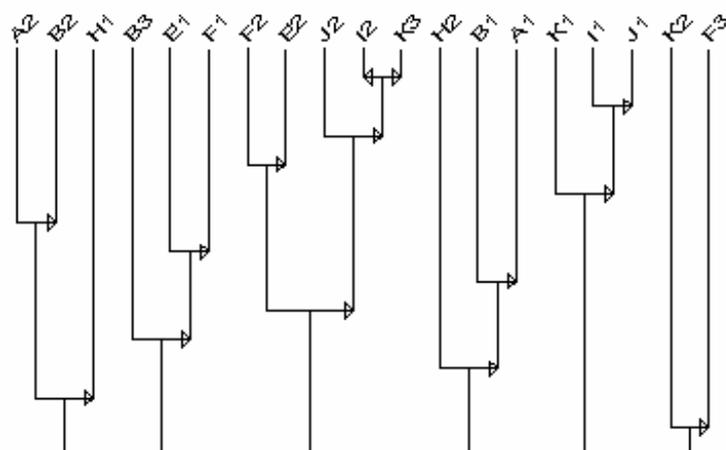


Fig. 4 Arbre cohésif Bréal ES

L'arbre cohésif confirme bien l'impression que nous a donné cet ouvrage assez atypique qui propose un nombre restreint d'exercices mais d'une très grande variété. La seule présentation de ces exercices ne permet pas aisément de dégager une idée dominante : l'influence du professeur sera déterminante en supposant qu'il ait lui-même une idée nette des objectifs qu'il poursuit.

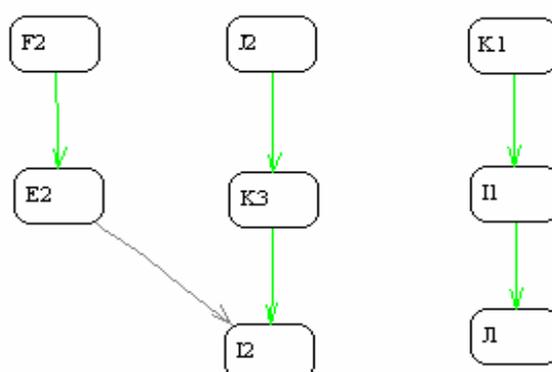


Fig. 5 Graphes Bréal ES (seuil 0.90)

En retenant les graphes implicatifs au même seuil que pour l'ouvrage de première S, nous retrouvons, mais de manière moins marquée certaines des classes précédente. C'est ainsi que nous retrouvons la classe fréquentiste [K1,I1,J1] qui s'appuie sur des répétitions d'expériences, cependant l'homogénéité est moins grande et un certain nombre d'interprétation ne sont pas demandées dans le contexte de détermination d'une série, ce qui n'était pas le cas précédemment. Nous retrouvons ainsi des exercices génériques [J2,K3,I2] mais de façon moins marqué qu'en 1S

La dernière classe, relié à la précédente par I2 (pas d'interprétation demandée) mais au seuil de 0.82 est la classe [F2,E2]. Elle est nouvelle et spécifique de la section ES, il s'agit de proposer des exercices dans un contexte quotidien (classe ou travail) : ils sont relativement très nombreux (plus de la moitié des exercices) et s'appuient prioritairement sur la donnée de la probabilité. Tout se passe comme si, du fait que cet enseignement s'adressait à des élèves qui ne deviendront pas des scientifiques, il était nécessaire d'ancrer davantage l'enseignement des probabilités dans une interaction avec la réalité source du modèle. Cependant, d'une certaine façon, les auteurs se trouvent pris dans une contradiction car il ne peuvent interpréter correctement dans une approche fréquentiste la plupart des problèmes qu'ils donnent et qui relèvent nettement de la deuxième approche du signifié.

## Limites de l'étude

Comme nous l'avons signalé en présentant les variables étudiées (partie 4) le nombre d'exercices du chapitre probabilité du livre de cours Bréal de Première ES est significativement inférieur au celui du Nathan de Première S (respectivement trente un et soixante treize). Un faible nombre d'exercices d'un certain type a rapidement une forte influence sur les conclusions que nous pouvons tirer sur le livre de cours Bréal de Première ES. L'étude des deux manuels est exhaustive mais on ne peut en tirer une généralisation sur la comparaison entre livres de cours de ES et S. Cette comparaison supposerait l'étude des collections de livres utilisées dans ces filières. Ce n'est pas ce que nous avons envisagé dans cet article qui présente une étude locale destinée à s'insérer dans un cadre plus global.

## 6 Conclusion

En travaillant de manière exhaustive les exercices proposés dans deux livres de cours de première, nous espérons dresser une carte de l'espace de travail probabiliste potentiel des élèves de première. Nos résultats confirment la priorité qui est donnée dans l'enseignement français à l'axe qui va de l'interprétation fréquentiste à l'axiomatique de Kolmogorov mais dans sa forme molle.

Les élèves travaillent ensuite entièrement dans ce modèle en manipulant des signifiants qui finissent par acquérir un signifié assez éloigné des interprétations fondamentales qui ont donné lieu à l'émergence de la probabilité et qui donne aussi du sens à son usage dans le monde réel. Il faut noter cependant une différence sensible entre ces livres de cours de la classe scientifique S et la classe économique ES, ce dernier s'engage moins dans la voie ensembliste et se trouve donc davantage confrontée de manière plus ouverte à question de l'intrication des signifiés. Dans tous les cas, mais de manière un peu différente, nous voyons là, en germe, naître toutes les questions futures posées à l'enseignement des statistiques par la difficile adéquation que les élèves pourront réaliser entre réalité et modèle dans la résolution des problèmes qui seront issus de la réalité.

## Références

- Batanero, C., Godino, J. D. & Roa, R. (2004), Training teachers to teach probability. *Journal of statistical education*, 12(1).
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances.
- Carranza, P. & Kuzniak, A. (2006). Dualité de la notion de probabilité et enseignement de la statistique au Lycée en France. In *EMF 2006*, Sherbrooke.
- Droesbeke, J.-J., Fine, J. & Saporta, G. (2002), *Méthodes bayésiennes en statistique*, Paris: Sfds.
- Gärdenfors, P., Sahlin, N.-E., Ramsey, F., Luce, D., Raiffa, H., Savage, L., Jeffrey, R., Kyburg, H., Hacking, I., Hansson, B., McClennen, E., Kahneman, D., Tversky, A., Machina, M., Ellsberg, D. & Rabinowicz, W. (1988), *Decision, Probability and Utility*: Cambridge University Press.
- Gras, R., Ag Almouloud, S., Bailleul, M., Larher, A., Polo, M., Ratsimba-Rajohn, H. & Totahasina, A. (1996), *L'Implication Statistique*: La Pensée Sauvage.
- Gras, R. & Bailleul, M. (2000). La Fouille dans les données par la méthode d'analyse statistique implicative. ARDM, Caen.
- Hacking, I. (2002), *L'émergence de la probabilité*: Seuil.
- Hacking, I. & Dufour, M. (2004), *L'ouverture au probable*, Paris: Armand Colin.
- Régnier, J.-C. & Oriol, J.-C. (2001). Fonctionnement didactique de la simulation en statistique. In *Journées de Statistique Lyon 2003* (ed. S. F. d. Statistique). Sfds, Lyon, France.

## Summary

Based on a comparative study of two textbooks on the topic of statistics, we sketch the potential working space existing at Year 11 in France. A strong calculative and formal probabilistic approach in the scientific section and a more vague tendency in the economic section are highlighted by a statistical implicative analysis. In the second case, what we call the intricacy of probability's signifiers seems to have a real importance.