PUBLICATIONS MATHÉMATIQUES ET INFORMATIQUES DE RENNES

CAMILO CHARRON

Une approche bayésienne de l'analyse implicative. Un exemple sur la catégorisation de problèmes relatifs aux fractions

Publications de l'Institut de recherche mathématiques de Rennes, 1994-1995, fascicule 3 « Fascicule de didactique des mathématiques et de l'E.I.A.O. », , exp. n° 8, p. 1-28

<http://www.numdam.org/item?id=PSMIR_1994-1995___3_A7_0>

Une approche bayésienne de l'analyse implicative. Un exemple sur la catégorisation de problèmes relatifs aux fractions.

Camilo Charron [1]
LaPsyDEE, GDR sciences cognitives,
CNRS (URA1353), Université Paris V

RESUME : L'analyse implicative bayésienne est exposée à partir d'un exemple d'étude des processus de construction de catégories de problèmes. La méthode permet de quantifier descriptivement et inférentiellement une dépendance orientée entre deux variables binaires ($a \rightarrow b$). Une utilisation systématique et approfondie du coefficient H de Loevinger est proposée pour caractériser les données recueillies sur un échantillon de sujets. Les conclusions descriptives sont ensuite étendues à une population parente par une inférence directe sur le paramètre η correspondant à la statistique H employée. Cette procédure proche de l'intuition permet de garder à l'esprit les questions posées, tout en offrant des moyens rigoureux d'y répondre.

Mots-clés : variable binaire, ampleur d'implication, description, inférence bayésienne.

INTRODUCTION

Le contact prolongé avec des enfants et adolescents amène l'observateur éclairé à une connaissance intuitive des différentes voies développementales. Dans sa quête d'une meilleure compréhension de ces phénomènes, le psychologue se dote d'un arsenal d'épreuves et d'un dispositif expérimental. Il retranscrit les performances d'un échantillon d'enfants dans un protocole chiffré. Comment extraire la quintessence des processus testés ? Il est possible de relever les dépendances orientées entre deux observables a et b. Cette relation appelée implication ($a \rightarrow b$) apparaît dans les cas où les sujets qui présentent la modalité a présentent en général la modalité b, alors que l'inverse n'est pas nécessairement vrai. Comment induire

---

[1] Cet article doit beaucoup aux travaux de Régis Gras sur l'analyse implicative qui m'ont rapidement passionné et servi de source d'inspiration ; j'en remercie vivement son auteur.

Mon travail méthodologique s'est également appuyé d'une part sur les travaux de Bruno Lecoutre et d'autre part sur ceux de Jean Marc Bernard.

J'exprime ma reconnaissance à l'un et à l'autre, Bruno Lecoutre dont par ailleurs, les précieux conseils m'ont aidé à progresser tout au long de l'élaboration de ce travail et dans certaines parties expérimentales, et Jean Marc Bernard pour ses suggestions et la relecture de cet article. Enfin merci à tous les deux qui m'ont chacun fabriqué « sur mesure » des programmes informatiques de travail.

l'existence et la nature de la relation d'implication pour l'ensemble des enfants dont on n'a pu recueillir les observations ? La méthode d'analyse implicative de Gras (1979, 1993) infère à partir des données la structure la plus crédible des liaisons orientées dans la population parente, sans toutefois en fournir l'importance numérique. Le classique indice de Loevinger (1947, 1948) quant à lui, livre une mesure précise de la proportion des liaisons orientées, mais il ne renvoie qu'aux données recueillies. La validité des résultats d'une recherche dépend grandement de leur degré de généralisation. Il importe alors que la méthode d'inférence choisie autorise au niveau inductif les mêmes possibilités de jugement que celles offertes par les indices descriptifs.

Or l'approche bayésienne (Rouanet, Lépine, Hollender, 1978 ; Lecoutre, 1984 ; Bernard, 1983, 1991 ; Bernardo et Smith, 1994) offre en théorie la possibilité de faire de l'inférence à partir de n'importe quel indice descriptif. Le lent essor des techniques bayésiennes s'explique en partie par les énormes difficultés que posent les calculs et par l'attachement trop mécanique des chercheurs à des méthodes bien connues comme la traditionnelle analyse de la variance. Avec la diffusion de puissants instruments informatiques, l'utilisation de procédures bayésiennes devient plus facilement envisageable. J'ai eu alors l'idée de les appliquer pour la première fois à l'indice de Loevinger et de repenser sous l'angle bayésien les notions d'implication et de graphe implicatif développées par Gras (1979, 1993). La construction d'une nouvelle méthode d'analyse implicative a été entreprise. En l'état, la méthode élaborée exploite, dans son étape descriptive (de caractérisation des données relevées), les propriétés de l'indice de Loevinger dans une plus large mesure que ne le fait l'usage communément répandu et elle généralise, dans sa phase inférentielle, l'ensemble des conclusions tirées descriptivement.

Le propos du présent article est d'exposer l'analyse implicative bayésienne à travers un exemple sur l'étude de la catégorisation de problèmes relatifs aux fractions chez l'adolescent. La première partie présentera la problématique en psychologie du développement et les résultats issus de méthodes usuelles d'analyse, la deuxième partie développera la méthode statistique elle-même, d'abord dans sa dimension descriptive, puis dans ses aspects inférentiels. A cette occasion, quelques résultats de la méthode de Gras (1979, 1993) seront présentés avant l'introduction de l'approche bayésienne. Enfin dans le troisième volet, les données de l'exemple abordé seront soumises à une analyse implicative bayésienne puis à des analyses statistiques complémentaires (études des procédures de résolution, des patrons individuels de réponse) qui aideront à l'interprétation. La conclusion récapitulera les résultats de la recherche exposée et rappellera les principaux traits de

l'analyse implicative bayésienne. Elle évoquera également les perspectives de développements de la méthode.

Une remarque clos cette introduction. Les outils mathématiques utilisés dans cette méthode proviennent de différentes théories. Certains sont classiques et très connus, d'autres le sont moins. Quelques résultats sont complètement nouveaux. Dans tous les cas, au sein de chaque paragraphe, se trouvent les références où le lecteur pourra trouver plus de détails ainsi que les démonstrations de base.

1. POSITION DU PROBLEME EN PSYCHOLOGIE DU DEVELOPPEMENT : LA CATEGORISATION DES PROBLEMES RELATIFS AUX FRACTIONS

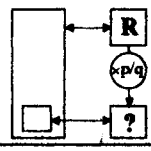
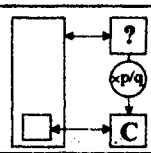
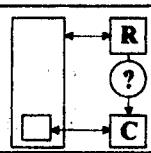
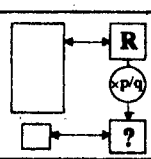
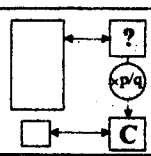
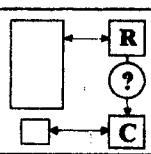
1.1. Questions posées en psychologie et méthode expérimentale

La prédiction testée dans cette recherche (Charron, 1995 a et b) est que lors de la résolution d'exercices mathématiques, les sujets construisent implicitement des catégories de problèmes. L'intérêt de cette prédiction est d'appréhender la catégorisation comme un processus dynamique d'adaptation à une situation complexe. En effet les diverses théories relatives à la catégorisation proposent des modèles de représentations cognitives d'objets statiques, manufacturés ou naturels, stockées en mémoire et indépendantes des traitements cognitifs. Or le critère de catégorisation exploré ici se fonde avant tout sur les traitements cognitifs réalisés en situation : je définis une catégorie mentale de problèmes comme *un ensemble de problèmes possédant des attributs communs et différenciés faisant appel pour un même individu à une même organisation invariante de la conduite.* Au-delà de l'existence de ces catégories de problèmes, c'est l'organisation de la construction des représentations qui est mise à l'épreuve dans ce travail. Est-elle strictement hiérarchisée ou suit-elle différents chemins avec des nécessités développementales ?

Pour tenter de répondre à l'ensemble de ces questions, 18 problèmes relatifs à l'usage de la fraction ont été donnés à 165 sujets : 55 élèves de CM2 (d'âge moyen : 10 ans, 11 mois), 55 élèves de 5ème (âge moyen : 12 ans, 10 mois), et 55 élèves de 3ème (âge moyen : 15 ans, 3 mois). Dans les exercices proposés, une fraction (donnée ou à calculer) reliait une quantité à une autre. Trois types de tâches ont été distinguées (calcul de la fraction, OF ; calcul de la quantité comparée, QC ; calcul du référent, QR) dans deux cas différents (la fraction exprime un rapport Partie-Tout, PT ; ou un rapport Partie-Partie, PP). Pour chaque sorte de problème,

trois habillages ont été présentés (parts de gâteau, clients dans un restaurant, arbres d'une forêt). Le croisement des facteurs « tâche », « rapport » et « habillage » a conduit à l'élaboration de 18 énoncés très proches les uns des autres linguistiquement. Des exemples de problèmes sont présentés dans le tableau I.

TABLEAU I. – *Exemples d'énoncés pour chaque type d'épreuves.*

	Recherche de la Quantité Comparée dans un rapport Partie-Tout (QC/PT)
	« Dans un bois, il y a 80 arbres. Les 4/5 des arbres sont des marronniers. Trouve le nombre de marronniers qui sont dans le bois. »
	Recherche de la Quantité de Référence dans un rapport Partie-Tout (QR/PT)
	« Au restaurant, 30 clients ont fini de manger, c'est à dire les 3/5 de tous les clients du restaurant. Trouve le nombre de clients du restaurant. »
	Recherche de l'Opérateur Fractionnaire dans un rapport Partie-Tout (OF/PT)
	« Un grand gâteau contient 90 parts, 36 parts ont été mangées. Quelle fraction représente celles qui ont été mangées ? Trouve la fraction la plus simple possible (c'est-à-dire une fraction irréductible). »
	Recherche de la Quantité Comparée dans un rapport Partie-Partie (QC/PP)
	« Au restaurant, 70 personnes prennent de la viande, les autres prennent du poisson. Le nombre de personnes qui prennent du poisson représente les 2/5 du nombre de personnes qui prennent de la viande. Trouve le nombre de personnes qui prennent du poisson. »
	Recherche de la Quantité de Référence dans un rapport Partie-Partie (QR/PP)
	« Dans une forêt de cèdres et de sapins, il y a 60 cèdres, c'est à dire les 3/5 des sapins. Trouve le nombre de sapins. »
	Recherche de l'Opérateur Fractionnaire dans un rapport Partie-Partie (OF/PP)
	« Dans un gâteau, 49 parts sont décorées, 14 ne le sont pas. Quelle fraction représente celles qui ne le sont pas par rapport aux parts décorées ? Trouve la fraction la plus simple possible (c'est à dire une fraction irréductible). »

1.2. Résultats généraux

1.2.1. Analyses des performances

Deux analyses statistiques préliminaires mettent en évidence le processus de construction des catégories de problèmes. En premier lieu l'analyse fiducio-bayésienne des performances (Lecoutre, 1984 ; logiciel PAC) confirme l'impact des variables contextuelles. A tout âge le rapport Partie-Tout est mieux réussi que le rapport Partie-Partie. La recherche de

la quantité comparée est mieux réussie que la recherche de l'opérateur fractionnaire qui elle-même est mieux réussie que la recherche de la quantité de référence. Par ailleurs pour toutes les épreuves, un effet de l'âge a été observé et l'effet de l'habillage du problème est négligeable.

1.2.2. Analyses des procédures

En second lieu l'analyse des procédures révèle qu'il existe une majorité de stratégies inadaptées de résolution (plus de 90 %) et peu de procédures justes (deux ou trois par type de problèmes). 1) Une classification ascendante hiérarchique (effectuée sous ADDAD), regroupant les problèmes en fonction des procédures observées, fait ressortir, à un item près, les 6 catégories d'épreuves correspondant au croisement des facteurs tâche et type de rapport. Ce résultat est identique aux trois âges. 2) Au niveau intra-individuel les procédures stables (reproduites au moins deux fois sur trois pour les mêmes sortes d'épreuves correspondant à ces six classes) sont majoritaires, mais elles croissent avec l'âge. Ces résultats ne sont pas triviaux car ils portent sur l'ensemble des procédures. Y compris lorsque la stratégie de résolution est erronée, elle est remployée plusieurs fois pour un même type de problème.

L'ensemble de ces faits montre que certains ensembles de problèmes diffèrent du point de vue des performances (effets des facteurs rapport et type de questions) et que les individus tiennent pour équivalents ces mêmes problèmes (réutilisation d'une même stratégie), y compris lorsqu'ils ne savent pas les résoudre ! Ces faits expérimentaux tendent à confirmer l'existence des catégories mentales postulées. Du point de vue développemental, il importe alors de savoir comment se construisent ces représentations. Quelles sont les différentes voies de développement possible ? Quelles sont les nécessités cognitives par lesquelles passe la construction de ces catégories ?

2. METHODE STATISTIQUE

2.1. Etude descriptive du problème

Ces questions amènent à étudier la nature des liaisons existant entre les performances observées aux différents types de catégories pour chaque âge. Par exemple est-ce que la réussite à une catégorie donnée A implique la réussite à une catégorie B ? Est-ce que l'échec à A entraîne la réussite à B ? Les données constituent un groupe de n observations qui peuvent

être soit des « succès ou présences » pour A et B notés *a* et *b*, soit des « échecs ou absences » notés *a'* et *b'*. L'étude bivariée des données conduit à la construction de tableaux 2 x 2 des fréquences observées. On rappelle que les fréquences sont les effectifs de chaque case rapportés à l'effectif total soit $f_{ij} = n_{ij}/ n$, et que leur somme vaut 1.

	b	b'	
a	f_{ab}	$f_{ab'}$	f_a
a'	$f_{a'b}$	$f_{a'b'}$	$f_{a'}$
	f_b	$f_{b'}$	1

2.1.1. Carré moyen de contingence Phi-deux

Classiquement, pour quantifier globalement la liaison entre deux variables le calcul d'un coefficient de contingence est de rigueur. Le plus usité est le Phi-deux ou carré moyen de contingence. Il est égal à la variance des taux de liaison pondérés par les fréquences-produits. Un taux de liaison est l'écart de la fréquence conjointe à la fréquence-produit divisé par la fréquence-produit. Il indique pour une case donnée l'écart entre la fréquence observée (conjointe) et la fréquence qu'il y aurait eu s'il n'y avait pas eu de liaison (donnée par la fréquence-produit).

$$\text{Le taux de liaison vaut } t^{ab} = \frac{f_{ab} - f_a f_b}{f_a f_b}$$

$(t^{ab}, f_a f_b)$ est un protocole centré (de moyenne 0) de variance Phi-deux :

$$\text{Phi}^2 = \sum f_a f_b (t^{ab})^2$$

Dans le cas du tableau 2×2 on a la formule suivante :

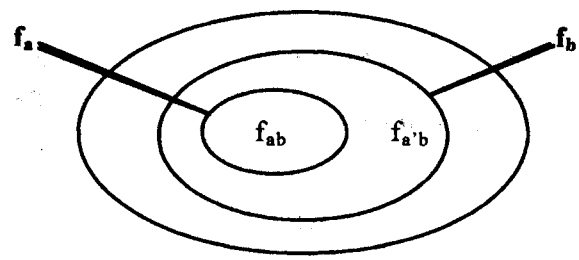
$$\text{Phi}^2 = fafb'\left(\frac{fafb'-fab'}{fafb'}\right)^2 + fa'fb\left(\frac{fa'fb-fa'b}{fa'fb}\right)^2 + fafb\left(\frac{fafb-fab}{fafb}\right)^2 + fa'fb'\left(\frac{fa'fb'-fa'b'}{fa'fb'}\right)^2$$

Plus la valeur du Phi-deux est proche de 1 plus la liaison est déclarée élevée, c'est-à-dire que la nature des performances obtenues à A varie en fonction de la nature des performances obtenues à B. Plus la valeur avoisine 0 plus on considère qu'il y a indépendance entre A et B ou que la nature des performances de A ne varie pas suivant la nature des performances de B (Rouanet, Le Roux, Bert, 1987).

Or il m'importe ici de connaître l'orientation de la liaison, plus précisément les relations d'implication entre les modalités de A et B.

2.1.2. Définition de l'implication statistique

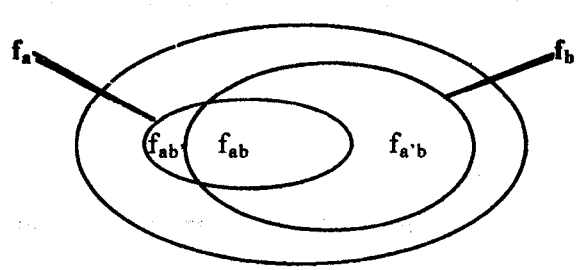
On dit que la réussite à A implique la réussite à B, notée $a \rightarrow b$, si les sujets qui réussissent A sont ceux qui réussissent en général B alors que la réciproque n'est pas nécessairement vraie. On dit qu'il y a une implication stricte ou logique, notée $a \Rightarrow b$, si les sujets qui ont réussi A sont exactement inclus dans les sujets qui ont réussi B, ce qui revient à dire qu'il n'y a aucun sujet qui a réussi A en ayant échoué B ($f_{ab'} = 0$). L'implication stricte peut être représentée par le diagramme suivant :



ce qui revient au tableau croisé suivant avec $f_{ab'} = 0$

	b	b'	
a	f_{ab}	0	f_a
a'	$f_{a'b}$	$f_{a'b'}$	$f_{a'}$
	f_b	$f_{b'}$	1

S'il existe une faible proportion (en regard des autres cases) de sujets qui contredisent l'implication stricte $a \Rightarrow b$, c'est-à-dire qui réussissent A en échouant B, il y a une quasi implication ou implication (au sens statistique) que l'on peut représenter par le diagramme suivant :



Pour la suite de l'exposé, le terme « implication » désignera la quasi-implication et le terme « implication stricte » renverra au cas d'implication logique. Le coefficient H de Loevinger[2] (1947, 1948) permet de mesurer le degré d'implication $a \rightarrow b$ reflété par le degré

[2] Cet indice a été utilisé par J. Loevinger (1947, 1948) dans une perspective d'évaluation psychométrique, pour mesurer l'homogenéité relative d'un item au sein d'un ensemble d'épreuves totalement ordonnées.

d'inclusion de f_a dans f_b (selon le terme de Hildebrand et al, 1977) ou par le degré de « petitesse » de la fréquence $f_{ab'}$ située dans la « case d'erreur » qui contredit l'implication stricte. Cet indice peut être exprimé sous l'une des deux formes :

$$H(a,b) = 1 - \frac{f_{ab'}}{f_a f_{b'}}$$

$$(\text{ou encore } H(a,b) = 1 - (nn_{ab'})/(n_a n_{b'}))$$

$$H(a,b) = 1 - \frac{f_{ab'}}{(f_{ab} + f_{ab'})(f_{a'b'} + f_{ab'})}$$

Il est aussi égal à l'opposé du taux de liaison, qu'on pourrait appeler " taux de répulsion ", associé à la case contredisant l'implication : $H(a,b) = -t^{ab'} = -\frac{f_{ab'} - f_a f_{b'}}{f_a f_{b'}}$

$$\text{ou à } H(a,b) = \frac{f_b}{f_{b'}} t^{ab}$$

Cet indice varie de moins l'infini à un. Pour les valeurs inférieures ou égales à 0 l'implication n'existe pas. Lorsque l'indice vaut 0, A et B sont indépendants. Plus la statistique se rapproche de 1 et plus l'implication est élevée, et il y a une implication stricte pour une valeur de 1. Dans la pratique, on définit un seuil, appelé borne supérieure d'implication négligeable (que je note $h_{nég}$), en dessous duquel on estimera que la valeur de l'indice est négligeable. A l'inverse, une borne inférieure d'implication notable est également définie (que je note h_{not}), au-dessus de laquelle on qualifiera l'implication de notable ou d'ampleur importante. Une valeur comprise entre $h_{nég}$ et h_{not} est qualifiée de médiane. Dans l'exemple abordé plus bas, en regard de ma problématique, j'ai choisi respectivement les valeurs de .20 et .60 pour $h_{nég}$ et h_{not} Il est loisible bien entendu de choisir d'autres bornes.

□ Un exemple : les données fictives

Prenons trois tables de contingence correspondant à des implications $a \rightarrow b$ de tailles différentes :

	b	b'
a	6	5
a'	4	6

$a \rightarrow b$ (négligeable)

	b	b'
a	6	2
a'	6	7

$a \rightarrow b$ (médiane)

	b	b'
a	8	0
a'	5	8

$a \Rightarrow b$ (notable)

Pour chacune, l'indice de Loevinger, H(a,b) quantifiant la relation d'implication de *a* vers *b*, vaut respectivement .13, .42 et 1. La première relation est négligeable car inférieure à .20. Une lecture directe des données montre que 5 sujets sur 21 soit 24% des individus contredisent une relation d'implication stricte. Cette proportion, au regard des proportions des autres cases, semble trop élevée pour pouvoir dire qu'il y ait une relation de dépendance de *a* vers *b*. La seconde relation est médiane, 2 individus sur 21 soit 10% contredisent l'implication stricte. La troisième implication est notable et même stricte : aucun sujet ne contredit l'implication logique.

2.1.3. Généralisation de l'usage de l'indice de Loevinger à l'ensemble des cases de la table de contingence

Usuellement l'indice de Loevinger est employé de façon restrictive aux implications entre réussites ($a \rightarrow b$). Or c'est un indice plus spécifique que le Phi-deux, lié à la fréquence d'une case d'erreur. Il est généralisable à l'ensemble des cases de la table de contingence, ce qui revient également à considérer les implications de réussites à échecs. Ainsi dans le tri croisé on peut s'intéresser aux relations suivantes :

- implications directes (d'une modalité de A vers une modalité de B)

  $a \rightarrow b$ (i.e. $b' \rightarrow a'$), la réussite à A implique la réussite à B,

  $a \rightarrow b'$ (i.e. $b \rightarrow a'$), la réussite à A implique l'échec à B, ou la réussite à A *exclut* la réussite à B, ce qui signifie qu'au moins une des deux épreuves est généralement échouée (*exclusion négative*).

- implications réciproques (d'une modalité de B vers une modalité de A)

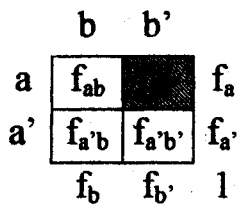
  $b \rightarrow a$ (i.e. $a' \rightarrow b'$), la réussite à B implique la réussite à A,

  $b' \rightarrow a$ (i.e. $a' \rightarrow b$), l'échec à B implique la réussite à A, ou l'échec à B *exclut* l'échec à A, ce qui veut dire qu'au moins une des deux épreuves est généralement réussie (*exclusion positive*).

Ceci revient à prendre successivement comme case d'erreur (case hachurée) les 4 cases du tableau de contingence[3]

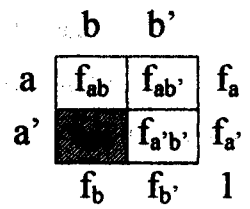
[3] Les calculs des indices de Loevinger sont facilement réalisables avec AIB (logiciel d'Analyses Implicatives Bayésienne, 1995) disponible gratuitement sur simple demande au LaPsyDEE, 46 rue Saint-Jacques 75005 Paris.

Pour les utilisateurs de DS3 le calcul est également aisé : dans le menu DESC demander le calcul des taux de liaison (commande TXL puis « entrée ») et prendre l'opposé des valeurs affichées. DS3 permet de saisir et de traiter les données. Les tableaux de données fabriqués sous DS3 sont réutilisables par l'AIB.



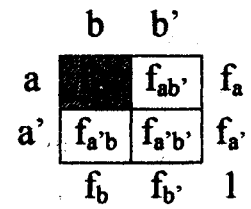
$a \rightarrow b$ ie $b' \rightarrow a'$

$$H(a,b) = 1 - \frac{f_{ab'}}{f_a f_{b'}}$$



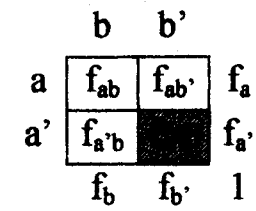
$b \rightarrow a$ i.e. $a' \rightarrow b'$

$$H(b,a) = 1 - \frac{f_{a'b}}{f_{a'} f_b}$$



$a \rightarrow b'$ ie $b \rightarrow a'$

$$H(a,b') = 1 - \frac{f_{ab}}{f_a f_b}$$



$b' \rightarrow a$ i.e. $a' \rightarrow b$

$$H(b',a) = 1 - \frac{f_{a'b'}}{f_{a'} f_{b'}}$$

2.1.4. Quelques propriétés ou relations concernant l'indice de Loevinger :

1) L'indice H de Loevinger est une statistique descriptive, c'est-à-dire qui dépend exclusivement des fréquences. Il reste invariant dans toute dilatation des effectifs : si les effectifs de chaque case sont multipliés par un même nombre, la valeur de l'indice reste la même.

2) Le signe d'un des coefficients H détermine le signe des trois autres :

si $H(a,b) > 0$ alors $H(a,b') < 0$, $H(b,a) > 0$ et $H(b',a) < 0$.

En effet $H(a,b) > 0 \Leftrightarrow H(b,a) > 0$, et $H(a,b') > 0 \Leftrightarrow H(b',a) > 0$.

3) On déduit des résultats précédents que l'implication statistique, quantifiée par l'indice de Loevinger, généralise certaines propriétés propres à l'implication logique, le célèbre modus ponendo ponens (se reporter à Hottois, 1989, pour les lois logiques) :

- $a \rightarrow b \Leftrightarrow b' \rightarrow a'$ et $a \rightarrow b' \Leftrightarrow b' \rightarrow a$ (contraposition).
- Si $a \rightarrow b$ et $b \rightarrow a$ alors $a \leftrightarrow b$ (équivalence).
- Si $a \rightarrow b'$ et $b' \rightarrow a$ alors on a $a \vee b$ (disjonction).

4) Dans un tableau 2×2, il existe la relation suivante entre le Phi-deux et l'indice de Loevinger : $Phi2 = f_a f_{b'} H(a,b)^2 + f_b f_{a'} H(b,a)^2 + f_a f_b H(a,b')^2 + f_{b'} f_{a'} H(b',a)^2$. La démonstration est immédiate dès lors que l'on note que $Phi2 = \sum f_a f_b (t^{ab})^2$ et que $H(a,b) = -t^{ab'}$ (se reporter aux paragraphes 2.1.1. et 2.1.2.).

5) Notons enfin que l'indice de Loevinger constitue un cas particulier pour un tableau 2×2, de l'indice Del s'appliquant aux tables de contingence de toutes tailles (J.M. Bernard communication personnelle, cf. Hildebrand, Laing, Rosenthal, 1977, pour un exposé détaillé).

2.1.5. Synthèse des configurations possibles et interprétations statistiques

En s'intéressant conjointement aux implications directes et réciproques, il est possible d'obtenir une des correspondances présentées ci-dessous dans le tableau II.

TABLEAU II. – *Synthèse des relations observables entre deux épreuves et interprétations statistiques.*

		$b \rightarrow a$ i.e. $a' \rightarrow b'$		
		$H(b,a) > h_{not}$	$h_{nég} < h(b,a) < h_{not}$	$H(b,a) < h_{nég}$
$a \rightarrow b$ i.e. $b \rightarrow a$	$H(a,b) > h_{not}$	**Equivalence**[4]	**Implication avec tendance à l'équivalence**	**Implication**
	$h_{nég} < H(a,b) < h_{not}$	**Implication avec tendance à l'équivalence**	**Tendance à l'équivalence**	**Tendance à l'implication**
	$H(a,b) < h_{nég}$	**Implication**	**Tendance à l'implication**	**Indépendance**[5]

		$b' \rightarrow a$ i.e. $a' \rightarrow b$		
		$H(b',a) > h_{not}$	$h_{nég} < h(b',a) < h_{not}$	$H(b',a) < h_{nég}$
$a \rightarrow b'$ i.e. $b' \rightarrow a$	$H(a,b') > h_{not}$	**Disjonction**	**Exclusion négative avec tendance à la disjonction**	**Exclusion négative**
	$h_{nég} < H(a,b') < h_{not}$	**Exclusion positive avec tendance à la disjonction**	**Tendance à la disjonction**	**Tendance à l'exclusion négative**
	$H(a,b') < h_{nég}$	**Exclusion positive**	**Tendance à l'exclusion positive**	**Indépendance**[5]

Chacune des cases révèle une proposition d'interprétation statistique relative aux bornes $h_{nég}$ et h_{not} choisies. Ces interprétations ne suivent pas une logique classique stricte, mais s'en inspirent pour proposer le « modèle » qui décrirait le mieux la structure des données. Comme le suggère la lecture du tableau, quelques précautions sont à prendre dans l'application de l'indice de Loevinger. En effet compte tenu de la propriété 3 (2.1.4.), il est nécessaire de prendre en compte simultanément les relations directes et réciproques pour expliciter intégralement le lien entre deux modalités. Pourtant il semblerait que cette démarche rigoureuse, et mathématiquement fondée, n'ait jamais été employée auparavant.

L'application exhaustive de l'indice de Loevinger permet descriptivement de répondre aux questions posées. Cette statistique aurait été très adaptée à mon étude si j'avais disposé de

[4] Une pratique répandue consiste à quantifier l'équivalence à partir d'une utilisation conjointe de l'indice de Loevinger et du coefficient Phi (la racine carrée du Phi-deux présenté au paragraphe 2.1.1.) : on conclue à l'équivalence lorsque les deux statistiques sont proches de 1 (Bideaud, Lautrey, 1983). Si cette procédure est comparable à une mesure des implications directe et réciproque, pour des valeurs élevées de l'indice de Loevinger, elle ne permet pas d'identifier les tendances à l'équivalence lorsque les valeurs de l'indice de Loevinger sont d'une moindre ampleur. L'équivalence $a \leftrightarrow b$ pourrait être quantifiée globalement par l'indice Del appliqué aux deux cases d'erreur a'b et ab' (Hildebrand, Laing, Rosenthal, 1977).

[5] Le terme d'indépendance est employé ici dans un sens plus fort que l'indépendance mesurée par le Phi-deux. Cette indépendance implique que $Phi^2 < h_{nég}^2$ alors que la réciproque n'est pas forcément vraie.

tous les sujets. Cependant je ne détiens que les mesures prélevées sur un échantillon tiré au hasard et il me faut en induire les conclusions que je pourrais porter sur la population parente dont il serait représentatif. Dans la partie suivante sur l'inférence, nous allons voir comment il est possible de généraliser les conclusions descriptives.

2.2. Extension inductive des analyses

2.2.1. Méthode de Gras

Soient A et B deux parties aléatoires, issues d'une population parente de taille N, qui ont réussi A et B. On notera A' et B' les échecs. Dans la perspective de Gras, on compare l'effectif $n_{ab'}$ à celui qu'aurait l'intersection $N_{AB'}$, dans une hypothèse d'indépendance entre A et B. Si l'effectif observé $n_{ab'}$, représentant les sujets qui contredisent l'implication stricte $a \Rightarrow b$, est invraisemblablement petit par rapport à celui qu'on peut attendre de la distribution des cardinaux $N_{AB'}$, on acceptera l'existence d'une quasi implication $a \rightarrow b$. Soient n_a, n_b, $n_{ab'}$, $n_{b'}$, et n, les effectifs respectifs des ensembles a, b, ab', b' et le nombre total de sujets, Gras (1979) montre que , pour un certain modèle de tirage et sous l'hypothèse H_0 d'indépendance, la variable aléatoire $N_{AB'}$, de réalisation $n_{ab'}$, suit une loi de Poisson de paramètre $n_a\, n_{b'}/\, n$. En centrant et en réduisant cette variable Gras obtient la variable : $Q(a,b') = (N_{AB'} - \frac{n_a n_{b'}}{n}) \Big/ \sqrt{\frac{n_a n_{b'}}{n}}$,

de réalisation $q(a,b') = (n_{ab'} - \frac{n_a n_{b'}}{n}) \Big/ \sqrt{\frac{n_a n_{b'}}{n}}$, qui dans l'hypothèse d'indépendance lorsque $n_a n_{b'}/\, n > 3$ suit une loi Normale centrée réduite. La vraisemblance de $a \rightarrow b$ est mesurée par le complément à 1 de la mesure de la probabilité déduite de la comparaison de l'effectif aléatoire $N_{AB'}$, à celui de $n_{ab'}$. L'implication $a \rightarrow b$ est admissible au niveau de confiance 1- α si et seulement si

$$\phi(a,b') = 1 - \Pr[Q(a,b') \leq q(a,b')] \geq 1 - \alpha$$

Dans la pratique, le niveau de confiance retenu est de .95. L'indice inductif de Gras q(a,b') entretient la relation suivante avec l'indice de Loevinger (Gras, 1993) :

$$q(a,b') = -\sqrt{\frac{n_a n_{b'}}{n}} . H(a,b)$$

□ Un exemple : les données fictives et l'inférence selon la méthode de Gras

Reprenons l'exemple fictif abordé précédemment au paragraphe 2.1.2. Les trois tables de contingence correspondant à des implications $a \rightarrow b$ de tailles différentes :

	b	b'
a	6	
a'	4	6

$a \rightarrow b$ (négligeable)

	b	b'
a	6	
a'	6	7

$a \rightarrow b$ (médiane)

	b	b'
a	8	
a'	8	5

$a \Rightarrow b$ (notable)

On rapelle que pour chacune, l'indice de Loevinger, H(a,b) vaut respectivement .13, 42 et 1. La première relation est négligeable la seconde est médiane et la troisième notable. Pour chaque tri-croisé, différentes dilatations des effectifs ainsi que les indices de Gras correspondant ont été affichés dans le tableau III. Ces indices peuvent être calculés à l'aide du logiciel CHIC[6] .

TABLEAU III. – *Calculs des indices de Loevinger et de Gras pour les différentes dilatations des données*

	Ampleur négligeable		Ampleur médiane		Ampleur notable	
Dilatations des effectifs	H(a,b)	φ(a,b')	H(a,b)	φ(a,b')	H(a,b)	φ(a,b')
× 1	.13	.62	.42	.78	1	.92
× 2	.13	.67	.42	.86	1	
× 3	.13	.71	.42	.91	1	.99
× 4	.13	.74	.42		1	
× 10	.13	.84	.42	.99	1	
× 20	.13	.92	.42		1	
× 25	.13		.42		1	

On constate à la lecture de l'exemple que plus il y a de sujets plus la probabilité livrée par l'indice de Gras augmente. De même lorsque l'ampleur croît pour un effectif donné, la probabilité augmente. L'indice de Gras croît en fonction de l'ampleur de l'implication et de la taille des effectifs. Cette propriété mathématique entraîne les deux conséquences suivantes :

- Les valeurs .94, 95 et .97 incluses dans les cases hachurées, bien qu'à peu près semblables, renvoient à des relations descriptivement de tailles différentes. Il est impossible de discriminer l'ampleur des implications dans la population parente sur la base de ces seuls

[6] Le Logiciel CHIC (Classification Hiérarchique Implicative et Cohésitive) élaboré par l'équipe de Gras (1993) permet d'effectuer des classifications ascendantes hiérarchiques selon la méthode de Lerman, des analyses implicatives inductives selon la technique de Gras ainsi que des analyses cohésitives. En analyse implicative, il n'effectue que les calculs inférentiels relatifs aux implications directes, et seules les relations entre réussites sont abordées.

indices. Cette distinction est pourtant essentielle car elle amène à trancher entre des conclusions radicalement différentes (implication négligeable, médiane, ou notable).

- Pour les probabilités faibles (inférieures au seuil de significativité) comment savoir si l'absence de significativité provient d'une implication faible ou d'un manque de données (effectifs faibles) ? Dans ces conditions, il est difficile de se prononcer sur l'absence d'implication dans la population de référence.

Etant donné une valeur observée $n_{ab'}$, et une loi du paramètre $N_{AB'}$ sous l'hypothèse d'indépendance de a et b (d'absence d'implication $a \rightarrow b$), l'indice de Gras donne une probabilité de $n_{ab'}$ si $N_{AB'}$, que je note $Pr(n_{ab'} \mid N_{AB'})$ *si l'hypothèse d'indépendance est vraie.* Cette probabilité peut être interprétée comme la probabilité d'existence de l'implication dans la population parente. Cependant son principal inconvénient est de ne pas permettre d'estimer à partir des données, la valeur que l'implication aurait pris si on avait pu la mesurer sur tous les sujets de la population parente, à savoir la probabilité inverse $Pr(N_{AB'} \mid n_{ab'})$. Or il faut éviter le risque d'erreur d'interprétation qui consisterait à confondre l'intensité d'implication de Gras, qui mesure une probabilité d'existence, avec l'ampleur elle-même de l'implication[7]. Les méthodes bayésiennes permettent de juguler ce biais de raisonnement, car elles offrent la possibilité de tirer des conclusions inductives directes sur le paramètre correspondant à n'importe quel indice descriptif, et par là même, la possibilité de porter un jugement sur l'ampleur d'une mesure dans la population parente.

2.2.2. Inférence bayésienne sur les fréquences

Les travaux de Bernard (1983, 1991) fournissent une méthode bayésienne pour faire de l'inférence sur une ou plusieurs fréquences. Cette approche permet de répondre à la question suivante : étant donné une fréquence observée f_{obs} sur un échantillon tiré au hasard et de taille finie, que peut-on dire de la fréquence parente φ de la population parente de taille inconnue ? Pour la suite de l'exposé on appellera f_{obs}, une fréquence ou proportion observée, et φ la fréquence vraie dans la population parente. On obtient ces probabilités à l'aide de la distribution bayésienne *a posteriori*, (c'est-à-dire après le recueil des données) qui probabilise les valeurs possibles des fréquences parentes φ. Cette distribution est construite en appliquant le théorème de Bayes qui permet de dériver la *distribution a posteriori* (ou *finale*) à partir des distributions d'échantillonnage et initiale (ou à *priori*) :

---

[7] Ce biais d'interprétation est fréquent en analyse de la variance où le seuil est confondu avec l'importance elle-même de l'effet, y compris parmi des chercheurs chevronnés (Lecoutre M.P., 1991).

$$P(\varphi \mid f) = K \,.\, P(f \mid \varphi).\, P(\varphi)$$

ou $P(\varphi \mid f)$ est la densité de la distribution *a posteriori* ou *finale* pour la valeur φ considérée ;
$P(f \mid \varphi)$ est la densité de la distribution d'échantillonnage pour la réalisation observée f_{obs} conditionnellement à la valeur considérée de φ ;
$P(\varphi)$ est la densité de la distribution *initiale* pour la valeur φ considérée ;
K est une constante qui assure que la probabilité totale vaut 1.

La distribution initiale est une Dirichlet de paramètres α_{ab}, $\alpha_{ab'}$, $\alpha_{a'b}$, $\alpha_{a'b'}$, (effectifs de l'initiale) généralisation de la loi Bêta qui est plus connue (cf. Wilks, 1962, p.177-182, où Bernardo et Smith, 1994, p.134, pour leur définition) ; elle a pour densité :

$$h(\varphi_{ab}, \varphi_{ab'}, \varphi_{a'b}, \varphi_{a'b'}) = K.\ \varphi_{ab}^{\alpha_{ab}-1} \,.\, \varphi_{ab'}^{\alpha_{ab'}-1} \,.\, \varphi_{a'b}^{\alpha_{a'b}-1} \,.\, \varphi_{a'b'}^{\alpha_{a'b'}-1},$$

qui se note $h(\varphi_{ab}, \varphi_{ab'}, \varphi_{a'b}, \varphi_{a'b'}) \sim D_4(\alpha_{ab}, \alpha_{ab'}, \alpha_{a'b}, \alpha_{a'b'})$.

avec

$$K = \Gamma(\alpha_{ab}+\alpha_{ab'}+\alpha_{a'b}+\alpha_{a'b'})/[\Gamma(\alpha_{ab}).\Gamma(\alpha_{ab'}).\Gamma(\alpha_{a'b}).\Gamma(\alpha_{a'b'})]$$

où $\Gamma(g)$ est la généralisation de la fonction factorielle pour les nombres réels . Il est à rappeler que si $g \in \mathbb{N}$, $\Gamma(g) = (g-1)!$.

La distribution finale est également une Dirichlet de paramètres ($n_{ab} + \alpha_{ab}$, $n_{ab'} + \alpha_{ab'}$, $n_{a'b} + \alpha_{a'b}$, $n_{a'b'} + \alpha_{a'b'}$). Elle combine les données (les f_{obs} dans n) et la distribution initiale, ou information à priori, sur les valeurs de φ. En d'autres termes, la distribution initiale fournit la possibilité d'intégrer des informations disponibles avant expérience : traduire des essais antérieurs, ou encore par exemple, exprimer l'opinion d'un comité de spécialistes dans le domaine étudié (Lecoutre et al, 1995). Aussi comme la distribution finale peut être fortement influencée par la distribution initiale, le choix de cette dernière est crucial.

Parmi les multiples possibilités pour la distribution initiale, J.M. Bernard (1991, et à paraître) distingue la *distribution uniforme* de Bayes-Laplace obtenue pour $\alpha_{ab} = \alpha_{ab'} = \alpha_{a'b} = \alpha_{a'b'} = 1$, et la *distribution « zéro »* de Haldane (1948) obtenue pour $\alpha_{ab} = \alpha_{ab'} = \alpha_{a'b} = \alpha_{a'b'} = 0$. On peut penser que toute idée qui formaliserait l'ignorance initiale conduirait à une distribution intermédiaire entre ces deux dernières. J'adopte la *solution bayésienne standard* retenue par Bernard (1991, et à paraître) qui consiste à prendre $\alpha_{ab} = \alpha_{ab'} = \alpha_{a'b} = \alpha_{a'b'} = .25$, c'est-à-dire qu'un « poids » total de 1 réparti de façon symétrique entre les différentes cases.

□ Un exemple : le développement précoce de la logique.

Dans une expérience tirée des travaux de Houdé et Charron (1994, 1995 a et b) l'objectif est de montrer que les critères de logicité focalisés exclusivement sur la quantification extensionnelle (réussite à l'épreuve piagetienne de quantification de l'inclusion, en présence de 10 marguerites et deux roses : « Y a-t- il plus de marguerites ou plus de fleurs ? »), et sur divers « pièges » ne sont pas suffisant pour circonscrire la rationalité en développement. Les performances d'enfants de 5 à 8 ans ont été mises en relation, en logique extensionnelle (réussite (E+) ou échec (E-) à l'épreuve de quantification de l'inclusion) et en logique intensionnelle (réussite (I+) ou échec (I-) à la verbalisation d'un englobement de signification transitif en présence d'un matériel concret « les lions sont des animaux d'Afrique, les animaux d'Afrique sont des animaux, *donc* les lions sont des animaux »). La prédiction testée est que lors du niveau d'échec en logique extensionnelle, les épreuves de logique intensionnelle seront déjà réussies. On s'attend donc à observer une absence de liaison entre les performances aux deux types de logique, que les procédures inductives usuelles (tests de Fisher, du KHI ² ...) ne sont pas à même de détecter (Rouanet et al, 1991). Les analyses pour 48 sujets conduisent au tri croisé suivant :

	E+	E-
I+	0,40	0,40
I-	0,10	0,10

Pour se prononcer en faveur de l'hypothèse, on cherchera à montrer que dans la population parente la différence DELTA = φ_{E+I+} - φ_{E-I+} est négligeable. Descriptivement on a d_{obs}= f_{E+I+} - f_{E-I+} = 0 et inductivement, $\varphi 1$-$\varphi 2$ ~ Bêta(19,25 ; 5,25) - Bêta(19,25 ; 5,25). On calcule la probabilité de l'énoncé généralisant correspondant : prob(DELTA<.15) = .90. On conclue qu'à la garantie .90, l'écart entre les deux fréquences ne dépasse pas .15, ce qui revient à dire qu'il y a 90% de chances pour que cette différence soit inférieure à .15 si on la mesurait dans la population parente (les calculs ont été effectués à l'aide du logiciel IBF2XK, avec des forces initiales standards, on trouvera le mode d'emploi dans Bernard, 1986). Cette méthode permet ici de confirmer l'hypothèse formulée.

2.2.3. Application de l'inférence bayésienne aux analyses implicatives

On rappelle que H(a,b) est l'indice dérivé à partir des f_{obs}. On notera η(a,b)son paramètre :

$$\eta(a,b) = 1 - \varphi_{ab'} / [(\varphi_{ab} + \varphi_{ab'}).(\varphi_{a'b'} + \varphi_{ab'})]$$

Il est possible de déduire la loi de probabilité de $\eta(a,b)$ par transformation de variable et intégration numérique, par une méthode de simulation puis d'intégration appelée Gibbs' sampling (Bernard, communication personnelle), ou en se ramenant à un calcul sur des distributions connues. Chacune de ces techniques amène à effectuer des calculs sophistiqués, qui cependant, restent surmontables avec l'appui de quelques logiciels informatiques. Pour des raisons de brièveté de l'exposé, je n'exposerai que la troisième méthode qui a déjà été employée par Lecoutre et al (1995) pour des différences de fréquences.

Cette méthode de calcul originale repose sur des propriétés des distributions de Dirichlet qui permettent d'exprimer les distributions marginales (cf. Bernardo et Smith, 1994, p.135, Bernard, à paraître) :

1) soit la Dirichlet $D_k(\varphi \mid \alpha) = K\ \varphi_1^{\alpha_1 - 1} \ldots \varphi_k^{\alpha_k - 1}$,

ou $K = \Gamma(\sum_i^{k+1}{}_{=1} \alpha_i) / \prod_i^{k+1}{}_{=1} \Gamma(\alpha_i)$, la distribution marginale de $\varphi^{(m)} = (\varphi_1, \ldots, \varphi_m)$, $m < k$, est la Dirichlet $p(\varphi^{(m)}) = D_m(\varphi^{(m)} \mid \alpha_1, \ldots, \alpha_m, \sum_j^{k+1}{}_{=m+1} \alpha_j)$

2) étant donné $\varphi_{m+1}, \ldots, \varphi_k$, la distribution conditionnelle, de $\varphi'_i = \varphi_i / (1 - \sum_j^k{}_{=m+1} \varphi_j)$, $i = 1, \ldots, m$ est une Dirichlet, $D_m(\varphi'_1, \ldots, \varphi'_m \mid \alpha_1, \ldots, \alpha_m, \alpha_{k+1})$.

3) en particulier $p(\varphi'_i \mid \varphi_{m+1}, \ldots, \varphi_k) = \text{Bêta}(\varphi'_i \mid \alpha_i, \sum_j^m{}_{=1} \alpha_j + \alpha_{j+1} - \alpha_i)$, $i = 1, \ldots, m$.

La méthode de calcul consiste à effectuer un changement de variable qui permet, à partir des propriétés précédentes, d'écrire la loi du paramètre avec les distributions marginales :

$z = \varphi_{ab'}$

$y = \varphi_{a'b'} / (1 - \varphi_{ab'}) = \varphi_{a'b'} / (1 - z)$

$x = \varphi_{ab} / (1 - \varphi_{ab'} - \varphi_{a'b'}) = \varphi_{ab} / [(1 - y).(1 - z)]$

Les propriétés énoncées plus haut permettent de déduire les distributions suivantes avec indépendance entre les variables x, y et z et les paramètres $\alpha'_{AB} = n_{AB} + \alpha_{AB}$:

$x \sim \text{Bêta}(\alpha'_{ab}, \alpha'_{a'b})$

$y \sim \text{Bêta}(\alpha'_{a'b'}, \alpha'_{ab} + \alpha'_{a'b})$

$z \sim \text{Bêta}(\alpha'_{ab'}, \alpha'_{a'b'} + \alpha'_{ab} + \alpha'_{a'b})$

que l'on peut calculer à l'aide du logiciel Le Bayésien (Lecoutre et Poitevineau, 1995) fonctionnant sous Windows.

par ailleurs on a :

$\varphi_{ab'} = z$

$\varphi_{a'b'} = y(1-z)$

$\varphi_{ab} = x(1-y).(1-z)$

et donc $\eta(a,b) = 1 - z / [\,[z+x(1-y).(1-z)].[z+y(1-z)]\,]$.

Ceci permet aisément de déduire la limite supérieure ou inférieure à $\eta(a,b)$. Du fait de l'indépendance des paramètres x, y, z, on peut facilement calculer leur distribution conjointe d'où la distribution du paramètre η.

La méthode exposée permet de calculer la probabilité, ou garantie bayésienne, que l'indice de Loevinger vrai (de la population parente dont l'échantillon est représentatif), noté $\eta(a,b)$ soit supérieur ou inférieur à certaines valeurs choisies. L'utilisateur pourra utiliser le programme informatique AIB (Charron, Lecoutre et Poitevineau, 1995) pour effectuer l'intégralité de ces calculs. Pour la suite de l'exposé, les notations H et η désigneront de façon générique la statistique descriptive et le paramètre correspondant ; H(a,b) et $\eta(a,b)$ seront employés pour le cas $a \rightarrow b$.

□ Un exemple : les données fictives traitées sous l'angle bayésien.

Reprenons l'exemple fictif des paragraphes 2.1.1. et 2.2.2. auquel on y adjoint des analyses implicatives bayésiennes. Dans ces dernières et pour chaque cas, est calculée dans le tableau IV la probabilité (ou garantie bayésienne) que le paramètre $\eta(a,b)$ soit inférieur, compris ou supérieur aux bornes choisies à l'étape descriptive (2.1.2.).

TABLEAU IV. – *Calculs des énoncés bayésiens pour les différentes dilatations des données fictives.*

	Ampleur négligeable			Ampleur médiane			Ampleur notable		
Dilatations des effectifs	H(a,b)	φ(a,b')	prob $\eta(a,b)<.20$	H(a,b)	φ(a,b')	prob $.20<\eta(a,b)<.60$	H(a,b)	φ(a,b')	prob $\eta(a,b)>.60$
× 1	.13	.62	.65	.42	.78	.50	1	.92	.91
× 2	.13	.67	.69	.42	.86	.66	1	.97	.97
× 3	.13	.71	.73	.42	.91	.75	1	.99	.99
× 4	.13	.74	.76	.42	.95	.81	1		
× 10	.13	.84	.86	.42	.99	.96	1		
× 20	.13	.92	.93	.42		.99	1		
× 25	.13	.94	.95	.42			1		

La lecture du tableau révèle que les garanties bayésiennes augmentent en fonction des effectifs lorsque la borne ou les bornes auxquelles est comparé le paramètre $\eta(a,b)$ sont maintenues constantes. En lisant successivement les cases hachurées on conclura

respectivement : il y a une probabilité de 93% que l'indice de Loevinger vrai prenne une valeur inférieure à .20 dans la population parente, il y a une probabilité de 96% que la valeur de l'indice vrai soit comprise entre .20 et .60, et il y a une probabilité de 97% que le paramètre vaille au moins .60 dans la population de référence. En revanche pour des probabilité inférieures à .90 on dira que la garantie bayésienne est insuffisante pour se prononcer sur l'ampleur de l'implication vraie avec la borne choisie : il faudrait plus de données pour conclure. Cette configuration est appelée cas d'ignorance. Dans l'exemple où l'implication est descriptivement négligeable, il peut apparaître paradoxal que la probabilité d'existence (mesurée par l'indice de Gras) croisse en même temps que la garantie bayésienne d'implication vraie négligeable. Cette contradiction peut être levée si l'on conçoit qu'une relation d'implication existe (au sens où elle est positive) bien qu'elle soit de faible ampleur (négligeable). Mais cette information ne ressort pas à la lecture de l'indice de Gras. La méthode d'analyse implicative bayésienne permet quant à elle de répondre aux questions posées sur l'ampleur des implications, en faisant de l'inférence directe sur le paramètre η à partir des données observées. Elle livre $\Pr(\eta \mid H)$. Elle offre au moins les trois possibilités suivantes :

- se prononcer sur l'*existence* de l'implication dans la population de référence lorsque $\eta > 0$;

- caractériser l'ampleur de l'implication vraie qui est *négligeable* quand $\eta < h_{nég}$ (borne jugée négligeable), *médiane* lorsque $h_{nég} < \eta < h_{not}$ (h_{not} étant la valeur estimée notable), et *notable* pour $\eta > h_{not}$.

- identifier les *cas d'ignorance* pour une borne donnée lorsque la garantie bayésienne est inférieure à .90 ;

L'étape inductive bayésienne permet ainsi de déterminer le degré de généralisabilité des propriétés observées sur les données à l'étape descriptive (se reporter au paragraphe 2.1.5.).

3. RETOUR A L'EXEMPLE SUR LA CATEGORISATION DES PROBLEMES

3.1. Analyse implicative bayésienne des données

On rappelle que les catégories correspondent aux 6 types de problèmes posés, présentés chacun sous trois habillages différents. Afin de simplifier l'exposé pour la suite des

analyses, le critère de réussite retenu à une catégorie est la réussite à deux épreuves sur trois. Les analyses avec les critères de 1 sur 3 ou de 3 sur 3 ont également été réalisées mais elles ne changent pas sensiblement les conclusions. Les calculs des indices d'implication ont été réalisés pour chacun des tris croisés des épreuves prises deux à deux à chaque âge, et représentées par un graphe (inspiré du graphe implicatif de Gras, 1993). D'une façon générale le nombre maximum d'implications observables est de x^2-x où x représente le nombre de variables. La figure 1 présente les graphes implicatifs qui reflètent l'image des liaisons orientées définies sur l'ensemble des six épreuves et correspondant à chaque niveau scolaire. Seules les implications descriptivement non négligeables (supérieures à .20) ont été relevées (Charron, soumis). Pour les énoncés bayésiens la garantie usuelle de .90 a été adoptée. Les flèches dénuées d'énoncé bayésien correspondent à un état d'ignorance dans la population parente.

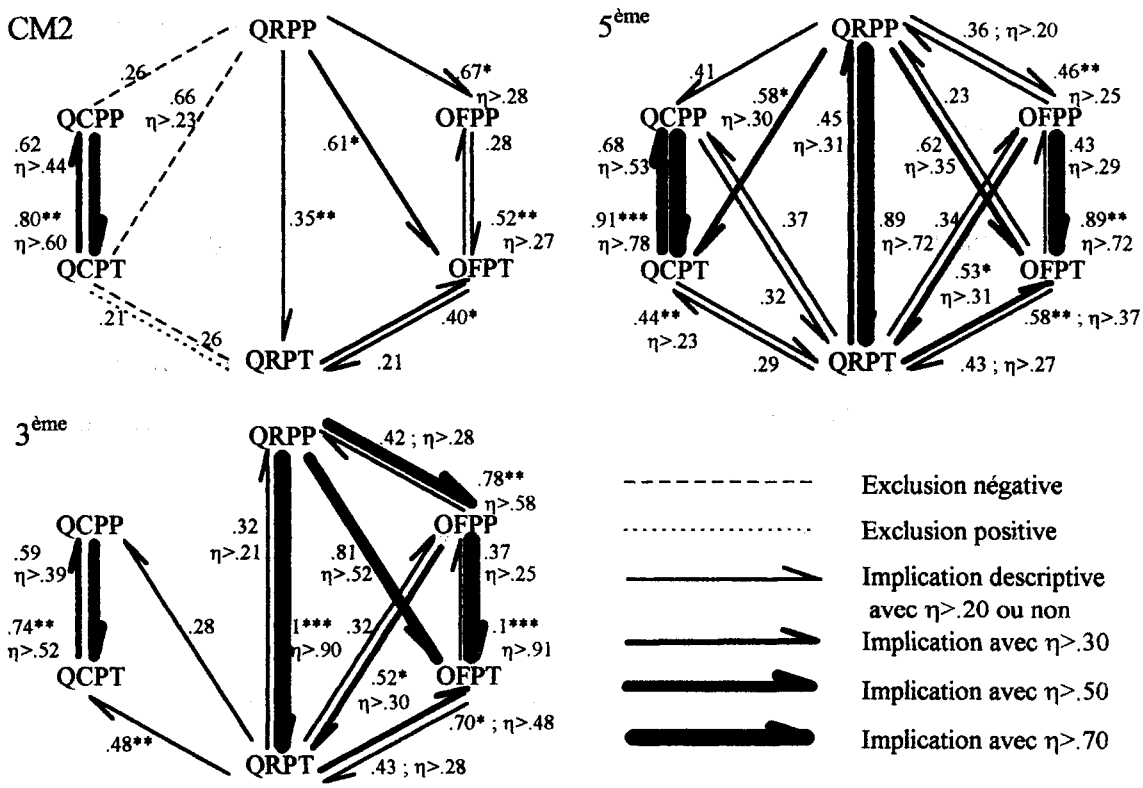


Fig. 1. – Graphes implicatifs représentant les relations orientées définies à chaque âge pour les six épreuves. Les flèches indiquent les implications entre réussites et les lignes pointillées les implications de réussite à échec. En regard des traits sont indiquées les valeurs descriptives des indices de Loevinger ainsi que les énoncés bayésiens pour lesquels une garantie élevée a été obtenue. Les astéristiques indiquent le degré de significativité calculé selon la méthode de Gras : * p .50, ** p > .95, *** p > .99.

On constate à la lecture de la figure 1 que les trois graphes présentent une configuration générale à peu près analogue, avec un réseau d'implication légèrement plus riche en 5ème et des exclusions qui n'apparaissent qu'en CM2. Parmi ces dernières,

l'implication de réussite à QCPT *vs* échec à QRPP est non négligeable dans la population parente, ce qui signifie que pour une majorité des individus lorsqu'on réussit le premier problème on échoue généralement le second. En 5ème cette relation est remplacée par une implication de QRPP *vs* QCPT, alors qu'en 3ème aucune relation n'a été observée.

La lecture comparative du tissu implicatif des trois graphes montre qu'aux trois âges l'implication des items PP *vs* PT est plus forte que l'implication réciproque PT *vs* PP. Les sujets qui ont réussi aux problèmes Partie-Partie (PP) réussissent le plus souvent les problèmes Partie-Tout (PT), alors que le phénomène inverse est plus rare. Localement l'ampleur des implications directes PP *vs* PT augmente avec l'âge (indices vrais supérieurs respectivement à .60, .78, .52 pour QC ; cas d'ignorance puis .71, .90, pour QR ; et supérieurs à .27, .72, .91 pour OF) alors que les implications réciproques PT *vs* PP restent de tailles comparables (dans le même ordre indices vrais supérieurs à .44, .53, .39 ; ignorance en CM2, .31, .21 ; ignorance puis .29, .25). D'une façon générale ces relations directes et réciproques révèlent des implications PT *vs* PP avec tendance à l'équivalence entre les épreuves PT et PP pour chaque tâche (QC, QR, et OF).

Il est à noter que les épreuves QR impliquent les épreuves OF aux trois âges. Les ampleurs de ces liaisons augmentent avec l'âge (ignorance sur les indices vrais en CM2, supérieurs à .25 et .37 en 5ème, et à .58 et .48 en 3ème). Enfin en 5ème et en 3ème, il existe des implications médianes de OFPP à QRPT, et de QRPT à QCPT (indices vrais toujours supérieurs à .20 avec absence de conclusion inductive pour les relations réciproques). L'implication de QRPP à OFPT médiane en 5ème (supérieure à .35) devient plus importante en 3ème (supérieure à .52).

Aucune conclusion inductive d'implication négligeable n'a pu être tirée pour les indices observés inférieurs à .20. Il faudrait plus de données pour se prononcer à ce propos. Néanmoins les analyses effectuées, tous âges confondus, révèlent qu'à un niveau général les implications non tracées (inférieure à .20) sont négligeables (les probabilités η<.20 sont pour la plupart au moins égales à . 90).

3.2. Analyses des procédures

Afin de mieux comprendre les phénomènes d'exclusion relevés chez les CM2 au paragraphe précédent, une nouvelle analyse des procédures a été menée. Les occurences des démarches qui sont potentiellement justes pour une tâche donnée, et qui sont parfois réutilisées par les mêmes sujets de façon inadaptée pour d'autres, ont été relevées en

pourcentage, pour les épreuves QC et QR et à chaque âge. Les réponses qui sont présentées dans le tableau V sont au nombre de 420 pour les CM2, 386 pour les 5èmes et 413 pour les 3èmes.

TABLEAU V. – *Fréquences d'utilisation des procédures potentiellement justes selon les catégories QC, QR et l'âge. La lettre c désigne la quantité comparée, R la quantité de référence, p/q le rapport, et n représente l'entier naturel donné par l'énoncé (C ou R suivant la question posée).*

	CM2				5ème				3ème			
Procédures	QCPT	QCPP	QRPT	QRPP	QCPT	QCPP	QRPT	QRPP	QCPT	QCPP	QRPT	QRPP
n×p/q	27,6	24	11,4	19,7	29,5	27,2	7,4	12	28,14	25,35	2,5	12,9
n×q/p	0,02	0,02	0,076	0,04	-	0,005	13,9	0,09	0,07	0,01	18,1	0,09
n+n(q-p)/p	-	-	2	-	-	-	-	2	-	-	1	3,3

Les cases hachurées contiennent les fréquences d'utilisation adaptées des procédures. On remarque à la lecture du tableau V que les conduites pertinentes sont les plus nombreuses pour les rapports Partie-Tout que pour les relations Partie-partie. D'autre part, on observe une généralisation abusive du raisonnement n×p/q, juste pour QCPT, à des tâches qui n'en justifient pas l'usage. Cette généralisation abusive tend à être plus fréquente pour la catégorie QRPP et diminue avec l'élévation du niveau scolaire. Elle pourrait expliquer les relations d'exclusion relevées, en particulier la réussite à QCPT qui exclut la réussite à QRPP.

3.3. Analyses des patrons individuels de réponse

Par ailleurs comme les indices retenus précédemment sont inférieurs à un (et donc quantifient des quasi implications), il existe une certaine frange de la population dont les performances contredisent les implications strictes. Il est possible de circonscrire ces observations, par l'étude des patrons individuels de réponses. La suite des réussites ou des échecs observés (selon le critère de réussite à au moins deux des trois habillages) pour les épreuves a été relevée pour chaque individu (Charron, 1995a). 152 patrons de réussites - soit 92%- sont des patrons qui vérifient les relations d'implication entre les épreuves qui avaient été présentées dans les graphes implicatifs. Les 8% des patrons qui ne confirment pas ces relations restent très marginaux – et c'est là une information importante – car aucun d'entre eux pris isolément n'est observé plus de deux fois. Parmi les patrons compatibles 13 séries présentées par 80% des élèves sont apparues plus de quatre fois. La figure 2 fait ressortir la hiérarchie des performances et les relations d'ordre partiel qui émanent de ces patrons. Le nombre d'occurence est annoté près des séries.

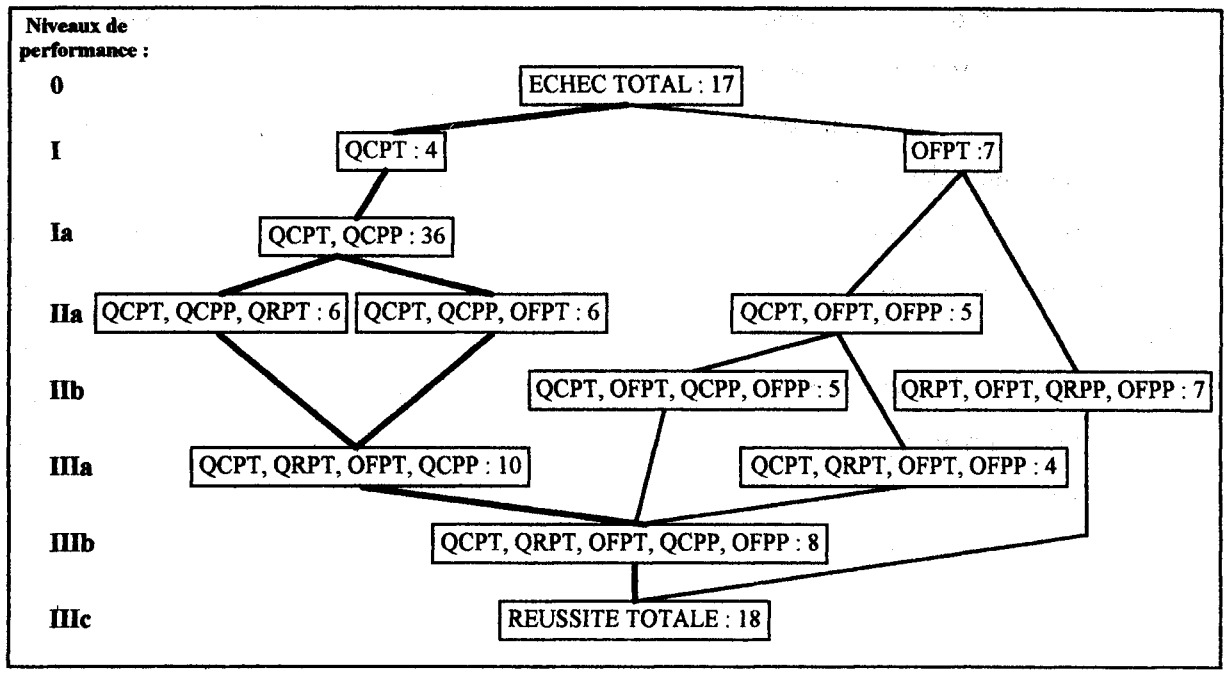


Fig. 2. – La notation des épreuves est la même que pour le tableau I et la figure 1. Le classement des performances est le suivant : Le niveau 0 regroupe les patrons «d'échec» total ; le niveau I réunit tous les patrons qui comptent au moins une réussite à l'une des trois tâches quel que soit le rapport ; le niveau II rassemble les profils détenant la réussite à deux tâches distinctes pour au moins un des rapports ; le niveau III est atteint par les individus qui résolvent les trois tâches pour au moins un des deux rapports. A l'intérieur de ce découpage général des performances, des distinctions supplémentaires ont été définies : la lettre «a» indique pour les trois niveaux qu'une tâche donnée a été réussie dans les deux rapports ; la lettre «b» signifie pour les niveaux II et III que deux tâches ont été chacune réussies dans les deux rapports ; enfin la lettre «c» signe les réussites totales du niveau III.

L'observation de cette figure permet conformément aux attentes, de cerner les étapes possibles du développement des savoirs opératifs se rapportant aux catégories de situations. Tous les chemins débutent par les épreuves QCPT ou OFPT qui étaient impliquées par les autres épreuves dans les graphes implicatifs. Les épreuves sont d'abord réussies dans un rapport Partie-Tout et ensuite dans un rapport Partie-Partie. La concentration relative des effectifs rend particulièrement saillante l'une des voies développementales (tracée en gras). Cette dernière, qui contient 105 patrons, pourrait représenter le cheminement conceptuel de 60% des enfants. L'ensemble des évolutions passe par la maîtrise des épreuves une à une, suivant certains ordres qui apparaissent comme nécessaires. En particulier les épreuves QCPT et OFPT sont des catégories pré-requises pour la suite du développement.

3.4. Conclusions relatives aux données

En accord avec la première partie des prédictions, il apparaît que les sujets construisent implicitement des catégories de problèmes pendant la résolution des épreuves. Conformément à la deuxième partie des attentes, la construction des catégories peut suivre

plusieurs voies développementales dont la structure a été mise en évidence par la méthode exposée. Au CM2, pour une partie des enfants, la réussite à la catégorie QCPT exclut la réussite à QRPP. L'étude des procédures a révélé que cette exclusion s'explique par l'utilisation aux deux épreuves d'un même schème adapté pour l'une et dangereux pour l'autre (en référence au modèle de Pascual-Leone, 1988). Ce phénomène de surgénéralisation d'une procédure disparaît aux âges suivants, ce qui laisse à penser que les élèves apprennent à limiter la portée du schème aux catégories pertinentes, en inhibant son utilisation inadaptée. Ce résultat semble compatible avec l'hypothèse que la compétition de schèmes contribue à la construction des catégories de problèmes (Charron, 1995 c). D'une façon plus générale les apprentissages suivent l'ordre suivant : il faut avoir réussi les épreuves Partie-Tout pour réussir les tâches Partie-Partie. L'ampleur de ces implications croît au cours du développement. En tenant compte des stratégies employées, je reprends à mon compte l'interprétation de Gras (Lerman, Gras, et Rostam, 1981) pour les implications $a \rightarrow b$: « le processus mental nécessaire à la solution de B est implicite pour résoudre A ; en d'autres termes, c'est un des composants de la solution de A ». Ainsi l'accroissement des implications directes de Partie-Partie à Partie-Tout, en présence d'implications réciproques (tendance à l'équivalence) peut s'interpréter comme le reflet d'une construction d'un invariant opératoire qui intègrerait les six épreuves au sein de catégories surordonnées QC, QR, OF correspondant aux trois types de questions posées.

4. CONCLUSIONS

La méthode présentée permet de quantifier descriptivement et inférentiellement les implications entre deux variables binaires. Deux étapes dans la procédure d'analyse des résultats ont été distinguées. En premier lieu, une démarche descriptive vise à décrire les données recueillies. A cette étape je propose d'étendre l'utilisation de l'indice de Loevinger à l'ensemble des cases de la table de contingence. Cette pratique permet de caractériser systématiquement la nature et le sens des liaisons bivariées, et de répondre à la question : « observe-t-on une implication, une équivalence, une exclusion, une disjonction ou une indépendance ? ». La recherche statistique de disjonctions me paraît d'autant plus importante qu'à ma connaissance cette approche n'a jamais été exploitée en psychologie. En détectant l'incompatibilité de deux modalités, elle pourrait constituer un moyen pour révéler des systèmes ou processus cognitifs qui permettent la résolution de certains items mais qui dans le même temps, constituent des schèmes « dangereux » faisant obstacle à la réussite d'autres épreuves. D'autre part pour caractériser chaque type de relation, il faut comparer les indices

de Loevinger observés à des valeurs-repères d'implication négligeable , médiane ou notable. La responsabilité du choix de ces valeurs incombe au psychologue seul. Ce choix peut se faire à partir de critères externes aux données, liés à la problématique ou à une norme couramment usitée, ou bien à partir de références internes aux données sans tenir compte des informations extérieures. La critique parfois faite à cette démarche est que cette décision est subjective donc non scientifique. Je ne le crois pas. La grande richesse des indices descriptifs est d'offrir au psychologue un moyen contrôlable, vérifiable, et répétable d'exercer sa sensibilité interprétative. La précaution à prendre, pour éviter de traduire les données dans le sens attendu, consiste à définir les valeurs de référence avant de débuter l'analyse.

Le second volet de la méthode, a pour objet d'inférer à partir des relations bivariées observées, l'ampleur des liaisons orientées existant dans la population dont l'échantillon est représentatif. A ce niveau, l'approche bayésienne standard restitue la souplesse de jugement que l'expérimentateur possède à l'étape descriptive. En induisant la valeur du paramètre directement à partir des données, elle permet de répondre de façon naturelle aux questions posées. Cette démarche se différencie du test d'hypothèse ou d'une comparaison des données au paramètre sous l'hypothèse d'indépendance. En particulier elle permet de discriminer les cas où l'implication n'existe pas, des cas d'ignorance dans lesquels les données sont insuffisantes pour répondre. Cette spécificité constitue un contrôle rigoureux pour éviter toute conclusion spéculative sur la population parente.

L'analyse implicative bayésienne a été appliquée ici à l'étude de la catégorisation de problèmes relatifs aux fractions. Elle a conduit a révéler la structure de systèmes cognitifs en développement, ainsi que des biais de raisonnement. Mes travaux méthodologiques actuels réalisés en commun avec J.M. Bernard, prennent appui sur l'indice Del (Hildebrand, Laing, Rosenthal, 1977) ou sur le concept de classification hiérarchique cohésitive de Gras (1993) pour explorer les possibilités d'analyses implicatives bayésiennes multivariées, applicables aux tables de correspondance de plus de quatre cases. Si la méthode est généralisable à une grande variété de protocoles et de domaines (en psychologie, en économie ou en biologie...) les interprétations quant à la signification psychologique des implications restent propres au sujet étudié. Comme dans toute procédure d'étude statistique, l'utilisateur doit attribuer un sens aux chiffres calculés, en s'appuyant notamment sur des analyses qualitatives. Il est possible d'imaginer des situations où l'implication revêtira une signification de causalité, ou d'ordre chronologique entre deux modalités. L'outil statistique aide à pondérer les significations qui précèdent, génèrent et guident toute étude. Il permet l'extraction d'un *métal précieux*, la réponse aux questions posées à partir du *minerais* fourni : les données.

Références bibliographiques :

AIB, (1995) *logiciel d'Analyses Implicatives Bayésiennes*, créé par C. Charron, module d'inférence bayésienne programmé par B. Lecoutre et J. Poitevineau, LaPsyDEE, CNRS, Université de Paris V, version 1.0.

ADDAD, (1989) Bibliothèque de logiciels pour l'analyse factorielle des données, *Association pour le Développement et la Diffusion de l'Analyse des Données*, Paris.

Bernard J.M., (1983) *Inférence bayésienne sur les fréquences dans le cas de données structurées : méthodes exactes et approchées.* Thèse de doctorat de troisième cycle, Paris : Université René Descartes.

Bernard J.M., (1986) Méthode d'inférence bayésienne sur les fréquences, *Informatique et Sciences Humaines.* 68-69, p. 89-133.

Bernard J.M., (1991) Inférence bayésienne et prédictive sur les fréquences, in Rouanet H., Lecoutre M.P., Bert M.C., Lecoutre B., Bernard J.M., *L'inférence statistique dans la démarche du chercheur* , Peter Lang, p. 121-153.

Bernard J.M., (à paraître) Bayesian Interpretation of Frequentist Procedures for a Bernouilli Process, *American Statistican.*

Bernardo J.M., Smith A.F., (1994) *Bayesian Theory,* New York. Wiley.

Bideaud J., Lautrey J., (1983) De la résolution empirique à la résolution logique du problème d'inclusion : Évolution des réponses en fonction de l'âge et des situations expérimentales, *Cahiers de Psychologie Cognitive*, 3, p. 295-326.

Charron C., (1995a) Individual Variations in the Construction of Categories of Problems Involving Fractions, *Psychological Mathematical Education*, Onasbruck Germany, p.7-10.

Charron C., (1995b soumis) Categorization of problems and conceptualization of fractions in adolescents, *European Journal of Psychology of Education.*

Charron C., (1995c) Adolescent cognitive conflict and scheme activation/inhibition processes during problem solving involving fractions, *Actes de colloques : International Conference on Conflict in adolescents ICA* Ghent, Belgique, p.18.

C.H.I.C., (1993) *Classification Hiérarchique Implicative et Cohésitive* : logiciel d'analyse de données, IRMAR (Institut de Recherche Mathématique de Rennes), Université de Rennes I.

DS3, (1992) *Logiciel pour le traitement informatique et statistique des données et son enseignement*, crée par D. Corroyer, Brunoy : APETISD, version 3.10.

Gras R., (1979) *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, thèse d'état, Université de Rennes I.

Gras R., Larher A., (1993) L'implication statistique, une nouvelle méthode d'analyse de données, *Mathématiques Informatique et Sciences Humaines*, 120.

Haldane J.B.S., (1948) The Precision of Observed Values of Small Frequencies, in *Biometrika*, 33, p. 222-225.

Hildebrand D.K., Laing J.D., Rosenthal H., (1977) *Prediction analysis of cross classifications*, New York : Wiley. 312 p.

Hottois G., (1989) *Penser la logique, Une introduction technique, théorique et philosophique à la logique formelle*, Bruxelles, De Boeck, 273 p.

Houdé O., Charron C., (1994) Categorization and intensional logic in 6- to 9- year-olds children, *Actes de Colloques*, Amsterdam ISSBD, p. 403.

Houdé O., Charron C., (1995a) Catégorisation et logique intensionnelle chez l'enfant, *L'Année Psychologique*, *95*, p. 63-86.

Houdé O., Charron C.,(1995b révisions en cours) Logic of meaning and development of taxonomic knowledge, *Cognitive Development*.

Le Bayésien, (1995) *Bibliothèque de logiciels d'analyses bayésiennes* crée par Lecoutre B., et Poitevineau J., version 1.0., C.I.S.I.A., Saint-Mandé (France).

Lecoutre B., (1984) L'analyse bayésienne des comparaisons, Presses Universitaires de Lille.

Lecoutre B., Derzko G., Grouin J.M., (1995) Bayesian predictive approach for inference about proportions, *Statistics in medecine*, Vol 14, p. 1057-1063.

Lecoutre M.P., (1991) Et...le point de vue des chercheurs ? Quelques éléments de réflexion, in Rouanet H., Lecoutre M.P., Bert M.C., Lecoutre B., Bernard J.M., *L'inférence statistique dans la démarche du chercheur* , Peter Lang, p. 47-74.

Lerman I.C., Gras R., Rostam H., (1981) Elaboration et évaluation d'un indice d'implication pour des données binaires, *Mathématiques et sciences humaines*, 74, p. 5-35.

Loevinger J., (1947) A systematic approach to the construction and evaluation of tests of ability, *Psychological Monographs*, 61, 4.

Loevinger J., (1948) The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis, *Psychological Bulletin*, 45, p. 507-529.

PAC, (1994) *Programme d'Analyse des Comparaisons*, conçu par Lecoutre B., et Poitevineau J.M., CISIA (Centre Internationnal de Statistique et d'Informatique Appliquées) Saint-Mandé, France.

Pascual-Leone J., (1988) Organismic processes for neo-Piagetian theories : a dialectical causal account of cognitive development, in Demetriou A., (Ed), *The neo-Piagetian theories of cognitive development : Toward an integration*, Amsterdam, North-Holland, p. 25-64.

Rouanet H., Lépine D., Holender D., (1978) Model Acceptability and Use of Bayes-Fiducial Methods for Validing Models, in Requin J. (Ed) *Performance VII*, Hillsdale : Lawrence Erlbaum Associates, p. 687-701.

Rouanet H., Le Roux B., Bert M.C., (1987) *Statistique en sciences humaines : procédures naturelles*, Paris : Dunod Bordas.

Wilks S.S., (1962) *Mathematical Statistics*, New York : Wiley.

Camilo Charron,
allocataire de recherche.
LaPsyDEE, 46 rue Saint-Jacques
75005 Paris.