MATHÉMATIQUES ET SCIENCES HUMAINES

JEAN-MARC BERNARD
CAMILO CHARRON

L'analyse implicative bayésienne, une méthode pour l'étude des dépendances orientées. II : modèle logique sur un tableau de contingence

Mathématiques et sciences humaines, tome 135 (1996), p. 5-18
<http://www.numdam.org/item?id=MSH_1996__135__5_0>

Math. Inf. Sci. hum. (34e année, n° 135, *1996*, p. 5–18)

L'ANALYSE IMPLICATIVE BAYÉSIENNE, UNE MÉTHODE POUR L'ÉTUDE DES DÉPENDANCES ORIENTÉES. II : MODÈLE LOGIQUE SUR UN TABLEAU DE CONTINGENCE *

Jean-Marc BERNARD[1], Camilo CHARRON[2]

RÉSUMÉ — *Dans Bernard & Charron (1996), nous avons proposé une nouvelle méthode, l'Analyse Implicative Bayésienne (AIB), pour l'étude des dépendances orientées entre deux variables binaires, méthode qui permet de conclure en terme de quasi-implication entre modalités des variables. Nous étendons ici cette méthode au cas d'un tableau de contingence $A \times B$ quelconque avec le problème de la mesure du degré de quasi-adéquation des données à un modèle logique donné. Au niveau descriptif, la méthode repose sur l'indice "Del", proposé par Hildebrand et al. (1977), qui généralise l'indice H de Loevinger. L'étape inductive est ici aussi envisagée dans le cadre bayésien.*

SUMMARY — Bayesian Implicative Analysis, a method for the study of oriented dependencies. II : Logical model on a contingency table.
In Bernard & Charron (1996) we proposed a new method, the Bayesian Implicative Analysis, for the study of oriented dependencies between two binary variables, method which enables to conclude in terms of quasi-implication between modalities of the variables. In this article, we extend the method to the general case of an $A \times B$ contingency table, with the problem of measuring the degree of quasi-conformity of the data to a given logical model. At the descriptive level, the method is based on the "Del" index, proposed by Hildebrand et al. (1977), which generalises Loevinger's index H. Here again, the inductive step is envisaged within the Bayesian framework.

1. INTRODUCTION

Dans un précédent article (Bernard, Charron, 1996), que nous abrégerons par la suite en "BC-96", nous avons étudié les dépendances orientées entre deux variables binaires A_2 et B_2 ; le problème de base que nous avons traité est celui consistant à juger de l'adéquation des données à un modèle implicatif ($a \Longrightarrow b$), de façon à conclure, lorsque les données le permettent, à une "quasi-implication" ($a \longrightarrow b$). La méthode proposée, l'*Analyse Implicative Bayésienne* (ou *AIB*), permet, selon les cas, une conclusion de quasi-implication, de quasi-équivalence, ou encore de quasi-indépendance. Au niveau descriptif,

* Nous remercions Henry Rouanet et Denis Corroyer pour leurs remarques sur une version antérieure de cet article, ainsi que Jacques Lautrey pour avoir attiré notre attention sur le problème que nous traitons ici.

[1]Laboratoire de Psychologie Cognitive, CNRS ER 125, Université Paris 8, 2 rue de la Liberté, 93526 Saint-Denis Cedex.

[2]LaPsyDEE, CNRS URA 1353, Université Paris 5.

cette méthode repose sur les quatre indices H de Loevinger associés à chacune des cases du tableau 2×2, et, au niveau inductif, sur l'utilisation de l'approche bayésienne.

Dans l'article présent, notre propos concernera les situations où la question d'intéret porte sur la dépendance orientée entre deux variables catégorisées A et B ayant un nombre quelconque de modalités, avec le problème plus général de la mesure de l'*adéquation des données à un modèle logique* défini par un ensemble d'implications. Notre démarche procédera à nouveau en deux temps : étape descriptive qui conduit à une propriété des données, propriété dont l'étape inductive permet d'évaluer la généralisabilité à la population dont les données sont extraites.

L'étape descriptive repose sur l'indice "*Del*" proposé par Hildebrand, Laing & Rosenthal (1977) ; l'indice *Del* généralise l'indice H de Loevinger (1947, 1948) ainsi que l'indice "*Kappa*" de Cohen (1960) ; il mesure le degré de "quasi-adéquation" (en bref "q-adéquation") de données catégorisées bivariées à un modèle logique d'intérêt reliant les deux variables.

En ce qui concerne l'aspect inductif, l'inférence relative à l'indice *Del* a été moins étudiée que celle portant sur l'indice de Loevinger. En particulier, aucune approche fondée sur des tests conditionnels ne semble avoit été développée. A notre connaissance, seuls Hildebrand *et al.* (1997) ont proposé une méthode fréquentiste asymptotique, mais elle se heurte aux mêmes difficultés que celles évoquées dans BC-96 : du fait qu'il s'agit d'une méthode approchée fondée sur un argument asymptotique, sa validité est discutable lorsque les données sont peu nombreuses, cas où l'inférence est la plus nécessaire, ou lorsque les données sont proches du modèle logique, cas où l'inférence est véritablement utile. D'où l'alternative que nous proposons d'utiliser l'approche bayésienne, exempte de ces limitations, et qu'il est possible de mettre en oeuvre pratiquement grâce à une méthode approchée de calcul par *échantillonnage-MC* ("MC" pour "Monte-Carlo"), présentée dans BC-96 et décrite en détail dans Bernard (soumis).

2. NOTATIONS

Rappelons brièvement les notations utilisées dans BC-96. Etant données deux variables catégorisées A et B, on note a (resp. b) une modalité quelconque de A (resp. B). On note n_{ab} l'effectif observé de la case ab du tableau de contingence $A \times B$ (et n_{AB} l'ensemble de ces effectifs), f_{ab} la fréquence conjointe correspondante (et f_{AB} l'ensemble de ces fréquences) et f_a et f_b les fréquences marginales associées[3].

Dans BC-96, l'analyse implicative pour un tableau 2×2 est fondée sur les *taux de liaison* entre modalités, indices locaux d'écart à l'indépendance[4]. Les taux de liaison sont également les indices de base pour l'article présent. On rappelle que le taux de liaison t^{ab} entre a et b vaut

$$t^{ab} = \frac{f_{ab} - f_a f_b}{f_a f_b}, \tag{1}$$

avec comme propriétés essentielles : $\forall ab, t^{ab} \geq -1$, et $t^{ab} = -1 \Longleftrightarrow f_{ab} = 0$.

Une autre propriété importante est celle du *moyennage par regroupement* : si on effectue un codage A' et B' des variables A et B (application de A vers A' et de B

[3]Les fréquences marginales sont toutes supposées non-nulles.

[4]Nous renvoyons à BC-96, Rouanet, Le Roux & Bert (1987) et Rouanet & Le Roux (1993) pour le détail des propriétés des taux de liaison.

vers B'), les nouveaux taux de liaison $t^{a'b'}$ s'obtiennent par moyennage pondéré des t^{ab} des cases ab regroupées, avec comme poids les fréquences-produits correspondantes $f_a f_b$:

$$t_{a'b'} = \frac{\sum_{a \subset a', b \subset b'} f_a f_b t^{ab}}{\sum_{a \subset a', b \subset b'} f_a f_b} \tag{2}$$

3. ANALYSE DESCRIPTIVE : INDICE *Del* POUR UN MODÈLE LOGIQUE $\mathcal{M}$ SUR UN TABLEAU $A \times B$

Pour deux variables binaires $A_2 = \{a, a'\}$ et $B_2 = \{b, b'\}$, l'implication stricte "$a \Longrightarrow b$" constitue un modèle dans lequel la case ab' est "interdite" : la case ab' constitue ainsi une *case d'erreur* pour le *modèle logique* $\mathcal{M} = (a \Longrightarrow b)$. Dans BC-96, nous avons eu recours à l'indice de Loevinger, $H_{ab} = -t^{ab'}$, comme mesure du degré d'adéquation des données à ce modèle : H_{ab} est un indice descriptif du degré de quasi-implication $a \longrightarrow b$.

Considérons maintenant un tableau de contingence relatif à deux variables A et B ayant des nombres quelconques de modalités. Considérons également une hypothèse de recherche qui s'exprime comme un modèle logique $\mathcal{M}$, constitué d'une conjonction d'implications. Chaque implication du modèle se traduit par une ou plusieurs cases d'erreur dans le tableau $A \times B$. Tout modèle $\mathcal{M}$ peut ainsi aussi bien être défini soit par un énoncé logique, soit par l'ensemble $\mathcal{E}$ des cases d'erreur associées. On notera e_{ab} la variable indicatrice de $ab \in \mathcal{E}$, c'est-à-dire que $e_{ab} = 1$ si $ab \in \mathcal{E}$ et $e_{ab} = 0$ sinon[5].

A titre d'illustration, nous prendrons un exemple de données relatives à l'investigation expérimentale des stades de Piaget, extrait de Jamison (1977) et que appelerons ici simplement l'exemple "Stades" : dans cette recherche 101 enfants ont été soumis à deux épreuves, sériation des longueurs (A) et inclusion des longueurs (B), chaque épreuve se soldant par la catégorisation de l'enfant en trois niveaux de performance : "pré-opératoire" ($a1$ et $b1$), "transition" ($a2$ et $b2$) ou "opératoire" ($a3$ et $b3$). Le tableau 1 donne les effectifs observés n_{AB}.

Tableau 1: *Exemple "Stades". Effectifs observés n_{AB} sur 101 enfants classés selon le niveau atteint pour la sériation des longueurs (A) et l'inclusion des longueurs (B), d'après Jamison (1977, p. 248).*

		Inclusion		
		$b1$	$b2$	$b3$
Sériation	$a1$	14	0	0
	$a2$	15	5	2
	$a3$	19	20	26

Dans ce contexte, une hypothèse psychologique d'intérêt est celle selon laquelle il n'est possible d'atteindre le niveau i pour B que si on a déjà atteint au moins le niveau i pour

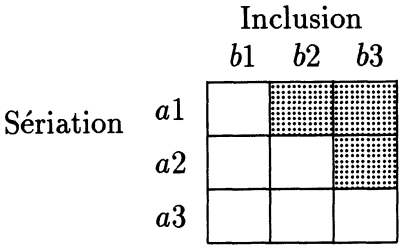
[5]On peut parler de *modèle-vide* lorsque $\mathcal{E} = \emptyset$, de *modèle-ligne* (resp. *modèle-colonne*) lorsque $\mathcal{E}$ est constitué uniquement de une ou plusieurs lignes (resp. colonnes) de $A \times B$, et de *modèle-propre* dans les autres cas. Seuls les modèles-propres qui expriment véritablement l'idée d'une dépendance orientée entre A et B nous concerneront par la suite.

A. Ainsi, le modèle logique d'intérêt peut être décrit par les deux implications suivantes[6] :

$$\mathcal{M} \quad = \quad b3 \Longrightarrow a3 \quad et \quad b2 \Longrightarrow (a2 \; ou \; a3) \tag{3}$$

L'ensemble des cases d'erreur $\mathcal{E}$ associé au modèle $\mathcal{M}$ donné en (3) est figuré par des cases grisées dans le tableau 2 : $\mathcal{E} = \{a1b2, a1b3, a2b3\}$.

Tableau 2: *Exemple "Stades". Ensemble $\mathcal{E}$ des cases d'erreur (cases grisées) associées au modèle $\mathcal{M}$ donné à l'équation (3).*



Hildebrand *et al.* (1977) ont proposé un indice général, appelé "*Del*", pour mesurer le degré de quasi-adéquation des données à un modèle logique $\mathcal{M}$ quelconque. Cet indice, que nous noterons $d_{\mathcal{M}}$, peut se définir comme l'opposé de la moyenne pondérée des taux de liaison des cases d'erreur, les poids étant les fréquences-produits associées[7] :

$$\begin{aligned} d_{\mathcal{M}} &= -Moy_{ab\in\mathcal{E}}(t^{ab}, f_a f_b) \\ &= -\frac{\sum_{ab\in\mathcal{E}} f_a f_b t^{ab}}{\sum_{ab\in\mathcal{E}} f_a f_b} \end{aligned} \tag{4}$$

En particulier, pour un tableau 2×2 et pour un modèle $\mathcal{M}$ constitué de la seule implication $a \Longrightarrow b$, il y a une seule case d'erreur ab', et on retrouve alors l'indice de Loevinger H_{ab} :

$$d_{a\Longrightarrow b} = H_{ab} = -t^{ab'} \tag{5}$$

L'indice $d_{\mathcal{M}}$ apparaît ainsi comme une généralisation de l'indice de Loevinger H. Un indice H mesure le degré de q-adéquation des données à un modèle logique élémentaire (une case d'erreur unique) ; pour un modèle logique complexe $\mathcal{M}$ (plusieurs cases d'erreur), l'indice $d_{\mathcal{M}}$ est une moyenne pondérée des indices H associés à chacune des cases d'erreur du modèle[8]. Le tableau 3 fournit, pour l'exemple "Stades", les indices H associés à chaque case de $A \times B$, considérée comme une case d'erreur ; dans chaque case on a également indiqué le poids associé $f_a f_b$. A partir de ce tableau et des cases d'erreur associées au modèle $\mathcal{M}$ (cf. tableau 2), on calcule aisément $d_{\mathcal{M}}$:

$$d_{\mathcal{M}} = \frac{(0.034 \times 1) + (0.038 \times 1) + (0.060 \times 0.672)}{0.034 + 0.038 + 0.060} = 0.851 \tag{6}$$

[6]Cette hypothèse conduit à ce que Jamison (1977) appelle le "Wohlwill's divergent-decalage pattern of intertask relation".

[7]L'indice *Del* est noté $\nabla_{\mathcal{M}}$ par Hildebrand *et al.* (1977). Cet ouvrage est en grande partie consacré à l'indice *Del* ; nous y renvoyons le lecteur pour des précisions et des propriétés importantes, mais non essentielles pour les besoins de cet article.

[8]En adoptant la définition (4), notre présentation de l'indice *Del* met l'accent sur ses liens avec l'indice de Loevinger. Mais Hildebrand *et al.* (1977) proposent une autre interprétation de cet indice en tant qu'une "*proportionate reduction in error (PRE) measure*". Si, à partir de la variable A, on cherche à prédire la variable B, la prise en compte du modèle $\mathcal{M}$ se traduira par une certaine proportion moyenne p_1 d'erreurs de prédiction, alors que sa non-prise en compte se traduira par une proportion d'erreurs p_0 ; l'indice $d_{\mathcal{M}}$ peut s'exprimer comme $(p_1 - p_0)/p_0$ et mesure ainsi la réduction relative d'erreurs de prédiction induite par la prise en compte du modèle.

Tableau 3: *Exemple "Stades". Indices de Loevinger H associés à chaque case d'erreur possible ab, i.e.* $-t^{ab}$. *En bas de chaque case figure le poids correspondant, i.e.* $f_a f_b$.

	$b1$	$b2$	$b3$
$a1$	-1.104 0.066	1.000 0.034	1.000 0.038
$a2$	-0.435 0.104	0.082 0.054	0.672 0.060
$a3$	0.385 0.306	-0.243 0.159	-0.443 0.178

3.1. Valeurs possibles de l'indice *Del*

Etant défini comme moyenne pondérée de quantités comprises entre $-\infty$ et 1, l'indice $d_{\mathcal{M}}$ peut varier lui-aussi entre $-\infty$ et l'unité[9]. L'indice $d_{\mathcal{M}}$ vaut 0 lorsqu'il y a indépendance stricte entre A et B ($A \perp\!\!\!\perp B$), mais, hors du cas d'un tableau 2×2, la réciproque n'est pas vraie. Une valeur $d_{\mathcal{M}} > 0$ indique un écart à l'indépendance dans la direction du modèle $\mathcal{M}$; par contre une valeur $d_{\mathcal{M}} = 0$ signifie seulement que l'éventuel écart à l'indépendance n'a pas lieu en direction du modèle, et non pas qu'il y a indépendance : en effet, lorsque les taux de liaisons sont non tous nuls, il est possible d'obtenir une moyenne nulle pour les taux de liaison d'un sous-ensemble strict des cases de $A \times B$.

La valeur maximale, $d_{\mathcal{M}} = 1$, est obtenue lorsque les taux de liaison t^{ab} des cases d'erreur valent tous -1, c'est-à-dire lorsque les fréquences f_{ab} des cases d'erreur sont toutes nulles, c'est-à-dire encore lorsque le modèle est strictement vérifié. La valeur trouvée ici, $d_{\mathcal{M}} = 0.851$, est relativement proche de 1 et indique donc un degré élevé de q-adéquation des données au modèle.

3.2. Valeurs-repères pour l'indice *Del*

Relativement à deux valeurs-repères d_{tend} et d_{quasi}, on parlera d'*absence de q-adéquation* (si $d_{\mathcal{M}} < d_{tend}$), de *tendance à la q-adéquation* si ($d_{tend} \leq d_{\mathcal{M}} < d_{quasi}$) ou de *q-adéquation* (si $d_{\mathcal{M}} \geq d_{quasi}$) des données au modèle, conformément aux termes déjà adoptés pour qualifier les q-implications dans BC-96. Etant défini comme une moyenne pondérée d'indices H de Loevinger, l'indice *Del* doit être apprécié avec les mêmes critères que ceux adoptés pour ce premier ; ainsi on recourra de préférence aux mêmes valeurs-repères que celles proposées dans BC-96 pour l'indice H : $d_{tend} = 0.40$ et $d_{quasi} = 0.60$. Rappelons néanmoins que ces valeurs-repères ont seulement un caractère indicatif.

Comme alternative, on peut également considérer une grille de valeurs-repères par rapport auxquelles on situera l'indice *Del*, par exemple : 0, 0.20, 0.40, 0.60, 0.80 et 1.

3.3. Modèles logiques équivalents et indice *Del*

L'indice *Del* possède une propriété d'invariance particulièrement appréciable. Considérons un codage A' et B' des variables A et B qui *est compatible* avec la structure de l'ensemble des cases d'erreur $\mathcal{E}$, c'est-à-dire que chaque nouvelle case $a'b'$ correspond au regroupement

[9]Pour un modèle-vide l'indice $d_{\mathcal{M}}$ est indéfini ; pour un modèle-ligne ou un modèle-colonne il vaut 0 quelles que soient les données, du fait que le protocole pondéré des taux de liaison est centré en lignes et en colonnes. L'indice $d_{\mathcal{M}}$ n'a donc d'intérêt que pour les modèles-propres (pour ces notions, voir note 5).

de cases ab ayant toutes même $e_{ab} = e'_{a'b'}$. Pour un tel codage, le modèle $\mathcal{M}$ sur $A \times B$ est logiquement équivalent au modèle $\mathcal{M}'$ sur $A' \times B'$. On montre qu'alors on a $d_{\mathcal{M}} = d_{\mathcal{M}'}$. Cette propriété d'invariance découle directement de la propriété de moyennage par regroupement des taux de liaison (équation (2)) et de la définition de l'indice *Del* comme moyenne pondérée de taux de liaison (équation (4)).

3.4. Modèle d'équivalence, indice κ ("*Kappa*")

L'indice κ ("*Kappa*") introduit par Cohen (1960) permet de mesurer l'accord entre deux juges ayant chacun séparément affecté des observations à un ensemble pré-déterminé de classes. Les observations peuvent alors être rangées dans un tableau croisé $A \times B$, où la classe attribuée par un juge constitue la variable A, et celle attribuée par l'autre la variable B. Les deux variables A et B sont alors *homologues*, la modalité a_i correspondant à la modalité b_i. Un accord parfait entre juges se traduit par l'absence d'observations dans les cases extérieures à la diagonale, c'est-à-dire les cases $a_i b_j$ avec $i \neq j$.

Autrement dit, l'accord parfait constitue un modèle logique $\mathcal{M}$ qu'on peut exprimer par $A \Longleftrightarrow B$, *i.e.* $\forall i, a_i \Longleftrightarrow b_i$. L'indice κ n'est autre que l'indice *Del* appliqué à ce modèle d'équivalence :

$$\kappa \;=\; d_{A \Longleftrightarrow B}. \tag{7}$$

3.5. Q-adéquation au modèle $a \Longleftrightarrow b$ et q-équivalence pour un tableau 2×2

Pour un tableau 2×2, nous avons introduit dans BC-96 la relation de "q-équivalence de degré h" qui, en utilisant l'indice *Del* introduit ici, peut s'exprimer comme :

$$a \overset{h}{\longleftrightarrow} b \quad ssi \quad d_{a \Longrightarrow b} \geq h \;\; et \;\; d_{b \Longrightarrow a} \geq h. \tag{8}$$

En d'autres termes, pour un degré h fixé, il y a q-équivalence entre a et b lorsque les indices d associés aux deux modèles logiques élémentaires composants $a \Longrightarrow b$ et $b \Longrightarrow a$ dépassent chacun h.

Mais, dans la ligne de l'article présent, on peut également se poser la question de la q-adéquation au modèle logique $a \Longleftrightarrow b$ dont le degré $d_{a \Longleftrightarrow b}$ se calcule comme moyenne pondérée de $d_{a \Longrightarrow b}$ et $d_{b \Longrightarrow a}$. Pour un degré h fixé, on parlera de q-adéquation de degré h au modèle $a \Longleftrightarrow b$ lorsque $d_{a \Longleftrightarrow b} \geq h$.

D'après ces deux définitions, il est clair que la q-équivalence entre a et b de degré h implique la q-adéquation au modèle $a \Longleftrightarrow b$ de degré h, mais que la réciproque n'est pas vraie : $a \overset{h}{\longleftrightarrow} b$ signifie que $d_{a \Longrightarrow b}$ et $d_{b \Longrightarrow a}$ sont tous deux supérieurs ou égaux à h, alors que $d_{a \Longleftrightarrow b} \geq h$ indique seulement que leur moyenne est supérieure ou égale à h. La notion de q-équivalence entre a et b proposée dans BC-96 est donc plus forte que la notion de q-adéquation au modèle $a \Longleftrightarrow b$.

3.6. Intérêt et limites de l'indice *Del*

Le point précédent indique que, pour le problème de l'adéquation à un modèle logique complexe (plusieurs cases d'erreur), deux approches peuvent être envisagées : dans la ligne de la définition de la q-équivalence, on peut requérir une condition sur chaque case d'erreur prise isolément, alors que, dans la ligne de l'indice moyen *Del*, on imposera seulement une condition globale.

Cette remarque permet de percevoir l'intérêt et les limites de l'indice *Del*, qui sont, en somme, ceux que présente tout indice moyen. Il permet de quantifier par une seule valeur numérique l'adéquation des données à un modèle logique éventuellement complexe, à partir des valeurs des indices de Loevinger relatifs aux q-implications élémentaires composantes. Mais, sauf dans le cas extrême où cet indice vaut 1, le résultat global qu'il fournit n'indique rien sur les écarts entre ces valeurs : certaines des q-implications élémentaires peuvent être de degré élevé et d'autres de faible degré. Ainsi, lorsqu'on a recours à l'indice global *Del* et même si sa valeur est élevée, il sera important d'examiner dans le détail les indices de Loevinger composants, afin de déterminer si les éventuelles cases "déviantes" n'amènent pas à reconsidérer ou à amender le modèle initial.

4. ANALYSE INDUCTIVE

Nous renvoyons à BC-96 pour des considérations générales sur l'inférence sur les données catégorisées et sur les deux approches envisageables : approches fréquentiste et bayésienne. Nous nous placerons toujours ici dans le cadre d'un modèle d'échantillonnage multinomial : les fréquences observées f_{AB} sont le résultat du tirage aléatoire d'un échantillon de taille n dans une population infinie caractérisée par les fréquences parentes φ_{AB}.

A l'indice descriptif observé $d_{\mathcal{M}}$, calculé en fonction des fréquences observées f_{AB} selon l'équation (4), correspond le paramètre parent inconnu $\delta_{\mathcal{M}}$, qu'on calculerait de façon analogue en fonction des fréquences parentes φ_{AB}. L'étape inductive a pour but de se prononcer sur $\delta_{\mathcal{M}}$ au vu des données observées.

4.1. Inférence fréquentiste sur l'indice *Del*

Hildebrand *et al.* (1977) ont étudié l'inférence relative à $\delta_{\mathcal{M}}$ dans le cadre fréquentiste, en proposant l'approximation asymptotique de la variance d'échantillonnage de $d_{\mathcal{M}}$ suivante (cf. Hildebrand *et al.*, 1977, équation (7), p. 200) :

$$Var(d_{\mathcal{M}}) \;=\; \frac{1}{n-1}\left[\sum_{ab} f_{ab}c_{ab}^2 - \Big(\sum_{ab} f_{ab}c_{ab}\Big)^2\right] \qquad (9)$$

où les coefficients c_{ab} sont donnés par :

$$c_{ab} \;=\; \frac{e_{ab} - (1-d_{\mathcal{M}})(\sum_a e_{ab}f_a + \sum_b e_{ab}f_b)}{\sum_{ab} e_{ab}f_a f_b},$$

les e_{ab} étant, rappelons-le, les variables indicatrices des cases d'erreur.

Pour l'exemple "Stades" nous avons trouvé descriptivement $d_{\mathcal{M}} = 0.851$. L'intervalle de confiance approché (approximation normale) pour $\delta_{\mathcal{M}}$ déduit de l'équation (9), à la garantie $\gamma = 0.95$, est :

$$IC \;=\; [0.654; 1.048] \qquad (10)$$

De façon analogue à ce nous avions constaté dans BC-96 pour certaines des méthodes asymptotiques proposées pour l'indice de Loevinger, l'intervalle de confiance obtenu ici couvre des valeurs impossibles pour $\delta_{\mathcal{M}}$, constat qui entache fortement, dans le cas présent, la validité de cette méthode approchée. De façon générale, cette méthode est en fait d'un intérêt limité puisque sa validité nécessite que les données soient en nombre suffisant et que l'indice observé $d_{\mathcal{M}}$ ne soit pas trop proche de sa borne maximale 1. Malheureusement,

à notre connaissance, aucune méthode fréquentiste alternative, qu'elle soit basée sur une meilleure approximation ou qu'elle fasse intervenir des tests conditionnels, n'a été proposée jusqu'à présent[10].

4.2. Inférence bayésienne sur l'indice *Del*

L'approche bayésienne de l'inférence permet de lever les limitations précédentes. Rappelons brièvement les éléments de cette approche qui sont présentés plus en détail dans BC-96, section 4.3 (voir aussi Bernard, 1991). A partir d'une *distribution initiale* de Dirichlet $Di(\alpha_{AB})$ (avec $\forall a, b,\ \alpha_{ab} \geq 0$) sur φ_{AB}, on est conduit à une *distribution finale* de Dirichlet $Di(\alpha_{AB} + n_{AB})$ qui intègre l'information apportée par les données. De cette distribution finale globale sur φ_{AB}, on peut déduire la distribution finale de tout paramètre dérivé des φ_{AB}, et en particulier du paramètre $d_{\mathcal{M}}$.

La méthode d'échantillonnage-MC (Bernard, soumis) permet de résoudre le problème technique délicat du calcul des distributions pertinentes et des énoncés probabilistes associés. Dans cette méthode, on approche la distribution finale sur φ_{AB} par un échantillon aléatoire de celle-ci. Le degré d'approximation est contrôlable par le nombre d'échantillons-MC extraits ; dans l'article présent, tous les calculs ont été réalisés avec $N = 10^5$ échantillons-MC, nombre qui procure une bonne qualité d'approximation.

L'unique point de choix de l'approche bayésienne est celui des *forces initiales* α_{AB}. Dans la perspective d'analyse des données, nous avons privilégié deux solutions pour ce choix : *(i)* l'idée de *zone d'ignorance* qui consiste à considérer l'ensemble des distributions initiales définies par la contrainte $\sum_{ab} \alpha_{ab} = 1$ (voir aussi Bernard, 1996 et Walley, 1996) ; et *(ii)* la notion de *distribution finale standard* (ou simplement *distribution standard*) obtenue avec les forces initiales $\alpha_{ab} = 1/K$, avec $K = A \times B$ (solution initialement proposée par Perks, 1947).
Nous préconisons l'utilisation de la solution standard, qui fournit un résultat unique, avec le recours conjoint à la zone d'ignorance pour l'évaluation de la sensibilité de ce résultat au choix des forces initiales[11].

Considérons à nouveau l'exemple "Stades" (cf. tableau 1) avec le modèle $\mathcal{M}$ donné à l'équation (3). La première étape consiste à déterminer la distribution standard sur le vecteur de $K = 9$ fréquences parentes φ_{AB} ; celle-ci est donnée par :

$$\varphi_{AB} \sim Di(14 + \tfrac{1}{9},\ 0 + \tfrac{1}{9},\ 0 + \tfrac{1}{9},\ 15 + \tfrac{1}{9},\ 5 + \tfrac{1}{9},\ 2 + \tfrac{1}{9},\ 19 + \tfrac{1}{9},\ 20 + \tfrac{1}{9},\ 26 + \tfrac{1}{9})$$

La distribution standard de $\delta_{\mathcal{M}}$, approchée par échantillonnage-MC dans cette distribution standard globale, est donnée à la figure 1. Elle est assez dissymétrique du fait d'un indice observé élevé, $d_{\mathcal{M}} = 0.851$, proche de la borne maximum 1 pour $\delta_{\mathcal{M}}$. Ses caractéristiques de centralité et dispersion sont : $Moy = 0.830$ et $Ety = 0.102$.

L'indice observé présente la propriété $\mathcal{P}(f_{AB}) = (d_{\mathcal{M}} > 0.60)$. La probabilité pour que cette propriété soit vérifiée dans la population se déduit de cette distribution, et vaut : $Prob(\delta_{\mathcal{M}} > 0.60) = 0.969$. Ainsi, pour la valeur-repère $d_{quasi} = 0.60$, on peut conclure à une q-adéquation des données au modèle $\mathcal{M}$ avec une bonne garantie, 0.969.

[10] On peut supposer que la raison de l'absence de tests conditionnels est la difficulté de déterminer une statistique ancillaire pour $\delta_{\mathcal{M}}$ par laquelle on puisse conditionner.

[11] Outre cette mesure de sensibilité, la notion de zone d'ignorance présente l'avantage de conduire à des résultats invariants pour tout recodage des variables compatible avec le modèle $\mathcal{M}$ (cf. 3.3) ; ceci n'est pas le cas de la solution standard qu'il faut spécifier à un niveau donné de codage des variables.

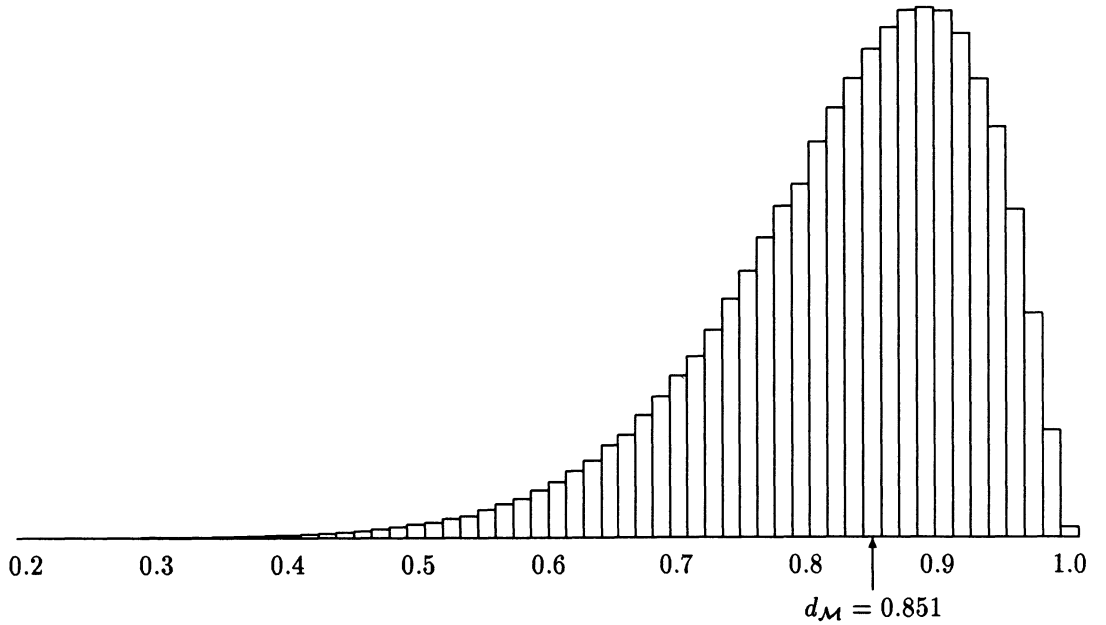


Figure 1: *Exemple "Stades". Distribution bayésienne standard de* $\delta_{\mathcal{M}}$ *(approchée par un échantillon-MC de taille* $N = 10^5$*).*

L'indice observé $d_{\mathcal{M}} = 0.851$ présente aussi la propriété ($d_{\mathcal{M}} > 0.80$), mais cette propriété plus forte ne peut être généralisée avec une garantie suffisante puisqu'on trouve : $Prob(\delta_{\mathcal{M}} > 0.80) = 0.675$.

Pour offrir un point de comparaison entre l'approche bayésienne standard et l'approche fréquentiste de la section 4.1, on peut calculer l'intervalle de crédibilité standard pour $\delta_{\mathcal{M}}$, à la garantie 0.95 (intervalle symétrique),

$$ICR = [0.584; 0.972], \tag{11}$$

c'est-à-dire qu'on a : $Prob(0.584 < \delta_{\mathcal{M}} < 0.972) = 0.95$. Cet intervalle est de largeur sensiblement égale à l'intervalle de confiance correspondant donné en (10), mais est décalé à gauche, de sorte qu'il ne couvre pas de valeurs impossibles pour $\delta_{\mathcal{M}}$. Ceci est une propriété générale de l'approche bayésienne proposée ici ; cette approche peut ainsi s'appliquer aussi bien avec des données peu nombreuses qu'avec des données pour lesquelles l'indice observé $d_{\mathcal{M}}$ est proche de 1.

Les énoncés probabilistes précédents ont été obtenus avec la solution standard. Pour chacun d'eux, on peut faire varier les forces initiales dans la zone d'ignorance et calculer chaque fois la probabilité associée, d'où, en fin de compte, l'obtention d'un intervalle de probabilités par énoncé. Pour parvenir à cet objectif, nous proposons une procédure approchée simplifiée qui consiste à étudier seulement deux points extrêmes de la zone d'ignorance, l'un où les forces finales $n_{AB} + \alpha_{AB}$ sont les plus en faveur du modèle, l'autre où elles sont le plus en sa défaveur[12].

Pour l'énoncé $\delta_{\mathcal{M}} > 0.60$, on trouve ainsi une probabilité qui varie de 0.933 à 0.982 ; on peut donc conclure que, quel que soit le choix des forces initiales dans la zone d'ignorance, on a $Prob(\delta_{\mathcal{M}} > 0.60) > 0.933$. Avec une valeur-repère moins élevée, par exemple 0.50, on aura une garantie minimum plus grande : pour tout point de la zone d'ignorance, on a $Prob(\delta_{\mathcal{M}} > 0.50) > 0.982$.

[12]Précisément, on maximise (ou on minimise) selon α_{AB}, l'indice $d_{\mathcal{M}}$ calculé non pas à partir des n_{AB} mais à partir des $n_{AB} + \alpha_{AB}$.

Pour avoir une idée globale de l'effet des forces initiales (dans la zone d'ignorance) sur la distribution finale, on peut calculer les intervalles de crédibilité à la garantie 0.95 correspondant aux deux points extrêmes, qui sont respectivement : [0.522; 0.952] et [0.622; 0.981].
La sensibilité des résultats à la distribution initiale n'est pas insignifiante, mais reste modérée; en tout état de cause, pour la valeur-repère 0.50, on peut conclure à la q-adéquation des données au modèle $\mathcal{M}$ avec une garantie d'au moins 0.982.

5. APPLICATION À L'EXEMPLE FRACTIONS : MODÈLE IMPLICATIF ENTRE "PARTIE-PARTIE" ET "PARTIE-TOUT"

Nous reprendrons l'exemple des données "Fractions" déjà présenté et analysé en partie dans BC-96 (voir également Charron, 1995 et sous presse). Rappelons que les données concernent 165 enfants et adolescents qui doivent résoudre une série de problèmes impliquant des fractions ; chaque enfant passe six types de problèmes obtenus en croisant "type de tâche" — calcul de la fraction "OF", de la quantité comparée "QC" ou de la quantité de référence "QR" — et deux "situations" — la fraction (donnée ou à calculer) exprime un rapport Partie-Partie "PP" ou un rapport Partie-Tout "PT".

Dans BC-96, nous avons abordé l'étude de la structure implicative globale reliant les six épreuves binaires QCPT, QCPP, QRPT, QRPP, OFPT et OFPP, en proposant un résumé des relations d'implication entre chaque paire d'épreuves. Une hypothèse générale de cette recherche est que la réussite aux situations Parties-Tout est nécessaire à la réussite aux situations Partie-Partie. Lors de cette première analyse, nous avons effectivement trouvé de nombreuses q-implications de degré élevé de "réussite à la situation PP" vers la "réussite à la situation PT" correspondante.

Cependant, cette première analyse constitue seulement une juxtaposition d'analyses spécifiques réalisées tâche par tâche. La seconde analyse envisagée ici a pour but de fournir une réponse plus globale à l'hypothèse d'implication des tâches Partie-Partie vers les tâches Partie-Tout. Pour cela, nous considèrerons le modèle logique $\mathcal{M}$ suivant :

> "La réussite à une tâche donnée dans la situation PP implique la réussite à cette même tâche dans la situation PT, et ceci, quelle que soit la tâche (QC, QR ou OF)."

Pour juger de l'adéquation des données à ce modèle, construisons deux nouvelles variables catégorisées : la variable "PP" correspond au pattern de réussite/échec aux trois épreuves QCPP, QRPP et OFPP, et comprend donc 8 ($= 2^3$) modalités ; la variable "PT", à 8 modalités également, correspond au pattern de réussite/échec aux trois épreuves QCPT, QRPT et OFPT. Le tableau 4 donne, en effectifs, la répartition des 165 élèves selon ces deux variables, en indiquant également les cases d'erreur associées au modèle $\mathcal{M}$.

La grande majorité des observations tombent dans une des 27 cases autorisées par le modèle (153 sur 165) ; parmi ces 27 cases autorisées, seules 4 n'ont pas été observées. Parmi les 37 cases d'erreur possibles, 9 ne sont en fait pas vides et ont recueilli seulement 12 observations en tout. Qualitativement, ces premiers résultats vont dans le sens du modèle d'intérêt. Pour quantifier descriptivement l'adéquation des données au modèle, on calcule l'indice *Del* observé et on trouve :

$$d_{\mathcal{M}} = 0.802.$$

Tableau 4: *Répartition des 165 élèves selon les variables PP et PT. Pour chacune, les modalités notées e1, e2 et e3 indiquent la réussite aux tâches QC, QR et OF respectivement ; les modalités e12, e13 et e23 indiquent la réussite aux deux tâches correspondantes ; enfin les modalités* 0 *et "e123" indiquent respectivement l'échec ou la réussite aux trois tâches. Les cases d'erreur associées au modèle $\mathcal{M}$ sont soit vides (lorsque non-observées) soit grisées (lorsqu'observées).*

PP \ PT	0	e1	e2	e3	e12	e13	e23	e123
0	17	4	2	7	0	1	2	1
e1	2	36			6	6		10
e2	1		0		0		0	1
e3	1			3		5	1	4
e12					4			3
e13		2	1	1	1	5		8
e23							7	2
e123						2	1	18

Cette valeur suffisamment proche de 1, indique une q-adéquation au modèle descriptivement élevée. Si on analyse en détail le calcul conduisant à l'indice global $d_{\mathcal{M}}$, on constate que les indices H associés aux neuf cases d'erreur non vides varient de -2.93 à 0.74, ceux associés aux cases d'erreur vides valant nécessairement toujours 1 ; les deux seules cases d'erreur pour lesquelles l'indice H est négatif (ce qui va à l'opposé du modèle) sont les cases (PP = e2, PT = 0) et (PP = e13, PT = e3) qui ne comportent chacune qu'une seule observation et dont les poids (les fréquences-produits) sont petits.

La figure 2 donne la distribution bayésienne standard de l'indice parent $\delta_{\mathcal{M}}$, obtenue avec un échantillon-MC de taille $N = 10^5$. On notera la moindre dissymétrie de cette distribution par rapport à celle de la figure 1 ; ceci découle du fait qu'ici l'indice observé est moins proche de la borne maximum 1 et que les données sont plus nombreuses.

De cette distribution, on déduit par exemple les énoncés inductifs suivants :

$$\begin{aligned} Prob(\delta_{\mathcal{M}} > 0.60) &= 0.9995 \\ Prob(\delta_{\mathcal{M}} > 0.705) &= 0.95 \\ Prob(\delta_{\mathcal{M}} \in [0.686; 0.885]) &= 0.95 \end{aligned}$$

Le premier énoncé permet de conclure inductivement que, pour la valeur-repère $d_{quasi} = 0.60$, il y a q-adéquation des données au modèle $\mathcal{M}$ avec une forte garantie, 0.9995. En se fixant une garantie plus petite, 0.95, on peut même conclure à la q-adéquation au degré 0.705. Le dernier énoncé donne l'intervalle de crédibilité standard pour $\delta_{\mathcal{M}}$ à la garantie 0.95.

La sensibilité au choix de la distribution initiale est ici faible : à l'intérieur de la zone d'ignorance, l'énoncé $\delta_{\mathcal{M}} > 0.60$ a toujours une garantie supérieure à 0.999, et les intervalles de crédibilité associés aux deux points extrêmes de cette zone sont [0.676; 0.880] et [0.698; 0.893].

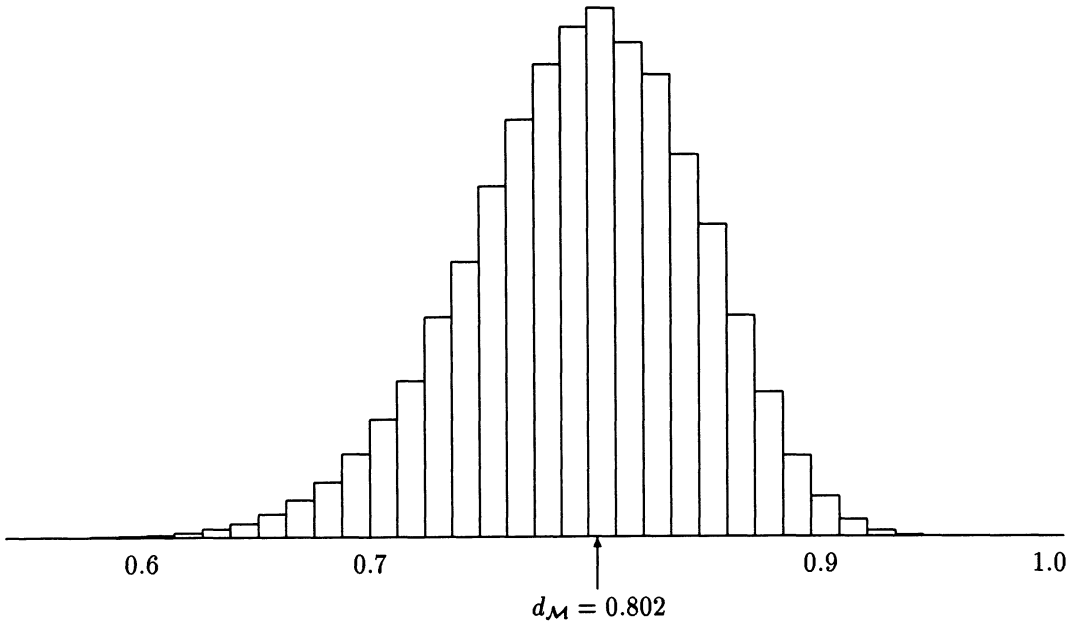


Figure 2: *Exemple "Fractions". Distribution bayésienne standard de $\delta_{\mathcal{M}}$ (approchée par un échantillon-MC de taille $N = 10^5$).*

Un autre modèle concurrent, $\mathcal{M}'$, que l'on pourrait également envisager est celui selon lequel la réussite à *toutes* les tâches PT est une condition nécessaire à la réussite à une *quelconque* tâche PP. Pour ce modèle, toutes les cases hors de la ligne (PP = 0) et hors de la colonne (PT = e123) sont des cases d'erreur[13]. Pour ce modèle alternatif, on trouve descriptivement $d_{\mathcal{M}'} = 0.093$ et inductivement $Prob(\delta_{\mathcal{M}'} < 0.15) = 0.990$ et $Prob(\delta_{\mathcal{M}'} < 0.20) > 0.999$ (résultats obtenus avec une distribution standard sur PP×PT). Pour la valeur-repère $d_{tend} = 0.20$, on peut ainsi conclure descriptivement et inductivement à une absence de q-adéquation des données à ce modèle. Cette conclusion signifie que l'écart à l'indépendance dans la direction de $\mathcal{M}'$ est faible, mais non pas qu'on est proche de l'indépendance, comme l'étude du premier modèle $\mathcal{M}$ l'a attesté.

6. ASPECTS INFORMATIQUES

Les résultats descriptifs et bayésiens donnés dans cet article ont été obtenus à l'aide du logiciel `BayDel` développé par les auteurs. Ce logiciel permet de mettre en oeuvre l'inférence bayésienne (par échantillonnage-MC) sur l'indice *Del* pour un modèle logique quelconque portant sur un tableau bivarié. Toute information concernant ce logiciel et ses modalités de diffusion peut être obtenue auprès des auteurs.

7. DISCUSSION

Hildebrand *et al.* (1977) introduisaient leur travail en avançant que "... une grande partie de la recherche antérieure concernant les variables qualitatives a porté sur la mesure du degré avec lequel les données s'écartent de l'indépendance statistique (quel que soit la direction de cet écart), plutôt que ... sur l'évaluation des performances prédictives d'une théorie particulière de l'expérimentateur" (p. 6). Ce constat reste encore aujourd'hui en grande partie fondé, ne serait-ce que par l'importance, parfois exclusive, accordée dans

[13]Compte-tenu de la répartition des cases d'erreur, l'analyse du modèle $\mathcal{M}'$ se ramène en fait à l'étude d'une q-implication dans un tableau 2×2 en recodant la variable PP en (0, autre) et PT en (e123, autre), et en regroupant (par sommation) les effectifs et les forces initiales correspondantes ; voir section 3.3.

les publications au test du Khi-2, même s'il doit être nuancé par le développement de méthodes descriptives fines comme celle de la mesure des attractions et répulsions par les taux de liaison (Rouanet, Le Roux & Bert, 1987 ; Rouanet, Le Roux, 1993) ou celle de l'analyse des correspondances.

L'intérêt de l'indice *Del* proposé par Hildebrand *et al.* (1977) est de répondre précisément à ce besoin puisqu'il mesure l'adéquation des données à un modèle logique *particulier* d'intérêt et permet ainsi d'étudier les *dépendances orientées* entre variables. La caractérisation de l'indice *Del* comme moyenne pondérée d'indices de Loevinger (et donc de taux de liaison) est essentielle pour établir des jonctions entre l'approche de ces auteurs et les autres méthodes descriptives pour les tableaux de contingence existantes.

Dans le même ouvrage, ces auteurs ont également énoncé divers critères qu'une mesure de "qualité de prédiction" telle que *Del* doit satisfaire et, parmi ceux-ci, celui de se prêter à une estimation précise même avec des données peu nombreuses. Malheureusement, ils reconnaissent que l'approche fréquentiste asymptotique qu'ils proposent pour l'inférence sur l'indice *Del* ne satisfait que très imparfaitement ce critère et regrettent l'absence, en 1977, d'une solution bayésienne. L'approche bayésienne que nous avons présentée dans BC-96 et dans l'article présent permet effectivement de résoudre cette difficulté puisqu'elle peut s'appliquer aussi bien avec des données extrêmes qu'avec des données en faible nombre. Bien que disponible en théorie depuis plusieurs décennies, cette approche n'a pu voir sa mise en oeuvre pratique que grâce à l'émergence récente de l'idée d'approcher les distributions bayésiennes complexes à l'aide de méthodes de type Monte-Carlo, telle que la méthode d'échantillonnage-MC utilisée ici.

Un tableau de fréquences bivarié $A \times B$ possède $AB-1$ degrés de liberté. Mais, dans cet espace de tous les tableaux possibles de dimension $AB-1$, le sous-espace des tableaux vérifiant l'indépendance est seulement de dimension $A+B-2$. Il reste ainsi $(A-1)(B-1)$ dimensions selon lesquelles un tableau peut s'écarter de l'indépendance, nombre d'autant plus grand que l'est la taille du tableau elle-même. Les divers modèles logiques possibles (une ou plusieurs cases vides) constituent les extrêmes de ces multiples formes possibles de dépendance.

Pour un tableau 2×2, il nous a été possible dans BC-96 de dresser une *carte implicative* qui figure l'ensemble des modèles logiques envisageables et de proposer une méthodologie qui permet, en situant descriptivement et inductivement les données sur cette carte, d'identifier le modèle pour lequel les données militent.

Pour le cas d'un tableau $A \times B$ étudié ici, nous nous sommes centrés sur le problème de la mise à l'épreuve d'un (ou de quelques) modèle(s) d'intérêt. Le nombre élevé de modèles logiques envisageables rend en effet plus délicat un processus de sélection du "meilleur" modèle résumé, notamment parce que cette sélection ne peut se faire sur la seule base de l'indice *Del*. Il faudrait également prendre en compte d'autres indices, introduits également par Hildebrand *et al.* (1977), de "portée" ("scope") et de "précision" du modèle qui permettent d'opérationaliser l'idée qu'un modèle peu restrictif (peu de cases d'erreur) doit être "pénalisé" par rapport à un modèle qui l'est plus (beaucoup de cases d'erreur). Ceci constitue une des directions de recherche dans laquelle l'article présent pourrait être prolongé.

Une autre direction d'extension de notre travail consisterait à généraliser l'analyse implicative bayésienne au cas multidimensionnel, c'est-à-dire à plus de deux variables catégorisées.

BIBLIOGRAPHIE

BERNARD, J.-M. (1991), "Inférence Bayésienne et Prédictive sur les Fréquences" dans *L'Inférence Statistique dans la Démarche du Chercheur*, par H. Rouanet, M.-P. Lecoutre, M.-C. Bert, B. Lecoutre, J.-M. Bernard, European University Studies, Series 6, Psychology, Berne : Peter Lang, pp. 121–153.

BERNARD, J.-M. (1996), "Bayesian Interpretation of Frequentist Procedures for a Bernoulli Process", *The American Statistician*, 50, No. 1, 7–13.

BERNARD, J.-M. (soumis), "Sample-based Approximate Methods for the Bayesian Analysis of Multinomial Data".

BERNARD, J.-M., CHARRON, C. (1996), "L'Analyse Implicative Bayésienne : une méthode pour l'étude des dépendances orientées. I : Données binaires", *Mathématiques, Informatique et Sciences humaines*, 134, 5–38.

CHARRON, C. (1995), "Individual Variations in the Construction of Categories of Problems Involving Fractions", Actes de *Psychological Mathematical Education 1995*, Osnabruck, Germany, 7–10.

CHARRON, C. (sous presse), "Categorization of Problems and Conceptualization of Fractions in Adolescents", *European Journal of Psychology of Education*.

COHEN, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, 20, 37–46.

HILDEBRAND, D. K., LAING, J. D., ROSENTHAL, H. (1977), *Prediction Analysis of Cross Classifications*, New-York : Wiley.

JAMISON, W. (1977), "Developmental Inter-Relationships Among Concrete Operational Tasks : An Investigation of Piaget's Stage Concept", *Journal of Experimental Child Psychology*, 24, 235–253.

LOEVINGER, J. (1947), "A systematic Approach to the Construction and Evaluation of Tests of Ability", *Psychological Monographs*, 61, No. 4, 1–49.

LOEVINGER, J. (1948), "The Technic of Homogeneous Tests Compared with some Aspects of Scale Analysis and Factor Analysis", *Psychological Bulletin*, 45, 507–530.

PERKS, F. J. A. (1947), "Some Observations on Inverse Probability Including a New Indifference Rule (with discussion)", *Journal of the Institute of Actuaries*, 73, 285–334.

ROUANET, H., LE ROUX, B. (1993), *Analyse des Données Multidimensionnelles*, Paris : Dunod.

ROUANET, H., LE ROUX, B., BERT, M.-C. (1987), *Statistique en Sciences Humaines : Procédures Naturelles*, Paris : Dunod.

WALLEY, P. (1996), "Inferences from Multinomial Data : Learning about a Bag of Marbles (with discussion)", *Journal of the Royal Statistical Society*, Ser. B., 58, 3–57.