

JEAN VAN BUGGENHAUT

**Questionnaires booléens : schémas d'implications et degrés de cohésion**

*Mathématiques et sciences humaines*, tome 98 (1987), p. 9-20

[http://www.numdam.org/item?id=MSH\\_1987\\_\\_98\\_\\_9\\_0](http://www.numdam.org/item?id=MSH_1987__98__9_0)

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1987, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

**QUESTIONNAIRES BOOLEENS : SCHEMAS D'IMPLICATIONS ET  
DEGRES DE COHESION.**

JEAN VAN BUGGENHAUT.

**INTRODUCTION**

Afin de généraliser d'une manière rationnelle l'analyse hiérarchique classique ( par les échelles de Guttman ) de questionnaires booléens , C.Flament (1) a montré qu'une approche mathématique basée sur la recherche d'un idéal d'une algèbre de Boole et de son algèbre quotient permettait de caractériser le protocole à analyser et d'en extraire divers schémas d'implications dont l'interprétation socio-psychologique peut se révéler relativement aisée.

La caractérisation de l'idéal booléen du protocole se ramenant à une recherche algorithmique de "projections canoniques ultimes" (ou PCU), l'augmentation en capacité et vitesse des micro- ordinateurs actuels met cette recherche à la portée de tous et permet d'élargir l'éventail des possibilités d'analyse.

Pour les définitions, les algorithmes et le traitement mathématique nous renvoyons le lecteur à l'ouvrage de C.Flament (1).

## 1. LA RECHERCHE DES PCU D'UN PROTOCOLE BOOLEEN.

Un questionnaire à  $n$  questions binaires comporte  $2^n$  patrons de réponses possibles, en supposant que les questions sont ordonnées, les patrons de réponses peuvent se représenter de diverses façons e.a. :

-en associant un couple de lettres  $(a,a')$ ;  $(b,b')$ ; ... ;  $(n,n')$  à chaque question. Un patron de réponse sera alors constitué d'une suite ordonnée de  $n$  lettres distinctes :  $ab'cde'...n$

-en associant à chaque question une suite de  $n$  chiffres booléens  $(0,1)$  : les réponses (négatives ou positives) aux questions :  $10110...1$

-en associant à chaque patron le nombre entier naturel  $(x \leq 2^n - 1)$  dont la représentation binaire correspond à la suite précédente.

- ...

Ainsi pour un questionnaire à 7 questions, un même patron pourra être indifféremment représenté par  $(abc'd'efg)$  ou  $(1100111)$  ou par  $(103)$ .

Considérons l'effectif de chacun de ces patrons et associons à tout entier  $\beta$  un seuil de dichotomisation :

tout patron dont l'effectif est inférieur ou égal à ce seuil  $\beta$  sera considéré comme négligeable et assimilé à un effectif nul.

Le protocole sera dichotomisé par le seuil  $\beta$  en deux sous-ensembles: celui des patrons non négligeables (à effectif supérieur à  $\beta$ ) et celui des patrons négligeables.

Soit  $R_\beta$  l'ensemble des patrons négligeables pour le seuil  $\beta$  .

Considérons un sous-ensemble de  $k$  questions ( $k \leq n$ ) du questionnaire, nous appellerons sous-patrons de longueur  $k$  les différents patrons de réponses à ce sous-questionnaire.

Observons qu'un sous-patron de longueur  $k$  est compris dans  $2^{(n-k)}$  patrons différents.

A tout seuil de dichotomisation  $\beta$  nous associerons un ensemble de PCU (projections canoniques ultimes) : l'ensemble maximal de sous-patrons de longueur minimale tels que tous les patrons qui les contiennent appartiennent à  $R_\beta$ .

Ainsi un protocole d'un questionnaire à 7 questions contiendra une PCU de longueur 3 de la forme (abd') lorsque l'ensemble des patrons négligeables  $R_\beta$  contient les  $2^4=16$  patrons dont (abd') est un sous-patron.

Cette propriété n'est pas suffisante pour caractériser les PCU d'un protocole dichotomisé. Il faudra de plus que la longueur du sous-patron soit minimale et que l'ensemble  $R_\beta$  puisse être engendré par l'ensemble des PCU.

## **2. IMPLICATIONS ASSOCIEES A UNE PCU.**

La détermination des PCU d'un protocole dichotomisé va permettre de construire des schémas d'implications entre les réponses.

Les implications associées dépendront fortement de la longueur de la PCU examinée:

k=1: une PCU de longueur un affirme l'inexistence d'une réponse donnée à une question donnée.

Ainsi la PCU " a' " affirme que la réponse négative à la première question est entièrement contenue dans  $R_\beta$  et n'apparaît donc jamais dans l'ensemble des réponses non négligeables. La première question recevra donc toujours une réponse positive.

k=2. une PCU de longueur deux affirme une implication directe entre les réponses à deux questions.

Ainsi la PCU "bc'" affirme que le couple de réponses "oui à la seconde question" et "non à la troisième" est entièrement contenu dans  $R_{\beta}$  et n'apparaît jamais dans l'ensemble des réponses non-négligeables.

Il en découle les implications logiques :

$b \rightarrow c$  ("oui à la seconde" implique "oui à la troisième")  
 et  $c' \rightarrow b'$  ("non à la troisième" implique "non à la seconde")

D'une manière analogue, plus brièvement (\*) :

k=3. La PCU "cd'f" détermine les six implications :

$c \wedge d' \rightarrow f'$  ("oui à la troisième" et "non à la quatrième" impliquent "non à la sixième")

$c \wedge f \rightarrow d \dots$

$d' \wedge f \rightarrow c' \dots$

$c \rightarrow d \vee f'$  ("oui à la troisième" implique "oui à la quatrième" ou "non à la sixième")

$d' \rightarrow c' \vee f' \dots$

$f \rightarrow c' \vee d \dots$

k=4. La PCU "abc'd" détermine les 14 implications :

$a \wedge b \wedge c' \rightarrow d'$

$d \rightarrow a' \vee b' \vee c$

$a \wedge b \wedge d \rightarrow c$

$c' \rightarrow a' \vee b' \vee d'$

$a \wedge c' \wedge d \rightarrow b'$

$b \rightarrow a' \vee c \vee d'$

$b \wedge c' \wedge d \rightarrow a'$

$a \rightarrow b' \vee c \vee d'$

$a \wedge b \rightarrow c \vee d'$

$c' \wedge d \rightarrow a' \vee b'$

$a \wedge c' \rightarrow b' \vee d'$

$b \wedge d \rightarrow a' \vee c$

$a \wedge d \rightarrow b' \vee c$

$b \wedge c' \rightarrow a' \vee d'$

(\*) Les opérateurs booléens seront notés  $\wedge$  et  $\vee$  (notés par ailleurs aussi . et +) et correspondent aux relations logiques "et" et "ou" ("ou" non exclusif = "et/ou")

Par souci de clarté, mais d'une manière redondante, nous écrirons généralement toutes les implications résultant d'une PCU ( les unes étant les implications duales des autres).

Il va de soi, d'un point de vue purement pratique, que les implications découlant d'une PCU perdent la plus grande partie de leur intérêt dès que la longueur de la PCU atteint 3 ou 4.

L'analyse des implications socio-psychologiques d'un questionnaire par la recherche des PCU de longueur  $k \leq 3$  (ou même  $\leq 2$ ) semble être l'élément pratique le plus immédiatement utilisable de cette méthode.

### 3. SEUILS DE DICHOTOMISATION ET COHESION D'UN PROTOCOLE.

En examinant les effectifs des différents patrons de réponses à un questionnaire et en les ordonnant, on peut obtenir une suite strictement croissante d'effectifs:

$$0 = S_0 < S_1 < S_2 < S_3 < S_4 < \dots < S_m$$

Chacun de ces effectifs  $S_i$  pouvant servir de seuil de dichotomisation, nous aboutissons à  $m$  analyses différentes du protocole examiné.

A chaque seuil de dichotomisation  $S_i$  nous associerons le nombre  $N_i$  total du nombre de réponses appartenant aux patrons non-négligés et la proportion  $P_i$  de réponses non-négligées par la dichotomisation:

$$P_i = N_i / N$$

$$1 = P_0 > P_1 > P_2 > P_3 > P_4 > \dots > P_m = 0$$

Le passage d'un seuil  $S_i$  à un seuil supérieur  $S_j$  ( $S_i < S_j$ ) augmentera le nombre de patrons négligés par l'analyse et peut ainsi diminuer le nombre ou la longueur des PCU.

Le problème du choix optimal d'un seuil de dichotomisation repose sur deux critères contradictoires:

-respecter au maximum la totalité du protocole et donc négliger le moins possible de patrons à effectif non-nuls ( choisir  $S_i$  le plus petit possible).

-obtenir des implications socio-psychologiques intéressantes et donc obtenir des PCU de longueur minimale ( choisir  $S_i$  le plus grand possible).

Il n'est pas possible de fournir une réponse mathématiquement satisfaisante à ce problème mais il nous a semblé intéressant d'associer à tout protocole une mesure du degré de cohésion de celui-ci:

soit  $S_i$  le plus petit seuil de dichotomisation pour lequel tous les PCU sont de longueur  $k \leq 2$ , nous appellerons degré de cohésion du protocole la proportion  $P_i$  de réponses retenues par cette dichotomisation.

D'une manière analogue, le seuil minimal de dichotomisation  $S_j$  déterminant des PCU de longueur  $k \leq 3$  permettrait de définir un degré complémentaire de cohésion  $P_j$  ( $P_j \geq P_i$ ) ...

En d'autres termes, il sera possible d'établir, pour une proportion  $P_i$  ( resp.  $P_j$  ) de la population interrogée, un schéma d'implications logiques entre les questions ( resp. entre questions et paires de questions ). (\*)

L'association de ces seuils à la méthode de recherche des PCU permet d'établir des schémas d'implications au sein d'un groupe cohérent de réponses et remplace avantageusement la méthode classique du "croisement" des questions deux-à-deux (les différents groupes de population retenus par divers croisements étant pratiquement incomparables alors que les PCU concernent un même groupe de population).

---

(\*) L'existence et la valeur intrinsèque des degrés de cohésion d'un protocole n'ont pas été abordées mais mériteraient une étude mathématique plus approfondie.

#### 4. ANALYSE DE DEUX EXEMPLES: LE PAPE ET LA RELIGION (198<sup>r</sup>).

Pour illustrer l'apport de la méthode dans l'analyse de questionnaires, nous avons tiré sept questions équivalentes de deux enquêtes effectuées à l'occasion de la visite du Pape en Belgique en mars-avril 1985 (enquêtes Le Vif- Di Marso (3) et Panorama- Edit (4). Il est utile de signaler que la seconde enquête se limite à la partie néerlandophone pays ).

Chacune de ces enquêtes reprenait, à la formulation près, les questions suivantes:

- a**: la visite du Pape influencera-t-elle la vie politique belge ?
- b**: faut-il suivre les messages du Pape ou de l'Eglise sur la participation des femmes dans les fonctions ecclésiastiques ?
- c**: faut-il suivre les messages du Pape ou de l'Eglise sur la contraception ?
- d**: faut-il suivre les messages du Pape ou de l'Eglise sur le mariage?
- e**: faut-il suivre les messages du Pape ou de l'Eglise en matière d'avortement ?
- f**: faut-il suivre les messages du Pape ou de l'Eglise sur la sexualité ?
- g**: selon vous le Pape est-il progressiste ( ou conservateur ) ?

Parmi les réponses exprimées ( respectivement environ 1500 et 1000 personnes interrogées ) respectivement 874 et 403 personnes ont formulées une opinion sur chacune de ces questions.

La première enquête détermine 23 seuils de dichotomisation possibles (ou effectifs de patrons):

1,2,3,4,5,6,7,8,9,10,11,14,15,16,18,19,24,28,36,72,97,153,156

tandis que la seconde détermine 18 seuils:

1,2,3,4,5,6,7,8,10,11,13,14,18,19,28,30,34,43

Une recherche systématique des PCU pour chaque seuil de dichotomisation  $S_i$  (suivant par exemple les valeurs décroissantes de  $S_i$ ) détermine très rapidement (en quelques minutes sur un micro-ordinateur 16 bits) les seuils correspondants aux PCU de longueur au plus égale à deux et trois. Les proportions de réponses retenues correspondantes et les schémas d'implications en découlent immédiatement.

Ainsi pour la première enquête :

seuil de cohésion principal : ( PCU de longueur  $\leq 2$  ) :  $S_i = 24$

degré de cohésion principal : 68.6 % (600 réponses retenues sur 874)

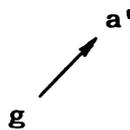
seuil secondaire : ( PCU de longueur  $\leq 3$  )  $S_j = 7$

degré de cohésion secondaire : 82.4 %

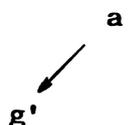
Au premier seuil correspondent les PCU :

ag    b'c    bc'    b'd    bd'    b'e    be'    b'f    bf'

et par conséquent les implications logiques (\*) :


  
 $b = c = d = e = f$

$B = 24$  ( 2.7 % )  
population  
retendue : 68.6 %


  
 $b' = c' = d' = e' = f'$

dont l'interprétation socio-psychologique est assez immédiate.

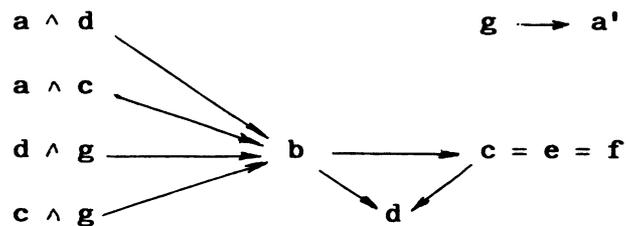
(\*) nous noterons la double implication  $b \rightarrow c$  et  $c \rightarrow b$  (équivalence de  $b$  et de  $c$ ) par une égalité :  $b = c$

Près de 70 % de la population interrogée répond d'une manière identique aux cinq questions b,c,d,e,f et sans rapport significatif avec les deux autres questions. Pour la même population, ces dernières questions s'opposent (la réponse "oui" à l'une impliquant "non" à l'autre).

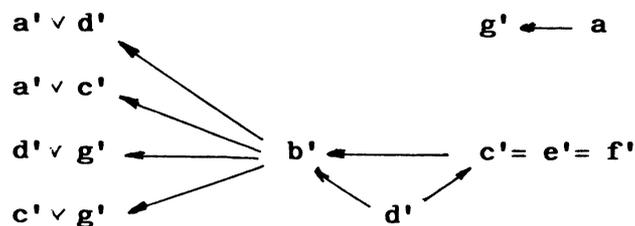
Le second seuil de cohésion détermine, par son nombre élevé de PCU de longueur 3, un schéma d'implications assez complexe et peu interprétable.

Le choix, relativement arbitraire, d'un seuil de dichotomisation compris entre les deux valeurs de cohésion ( $7 < S_i < 24$ ) permet de construire des schémas d'implications plus complexes que le précédent ci-dessus mais valable pour une partie plus importante de la population.

Par exemple pour le seuil immédiatement inférieur,  $B = 19$  :  
(population retenue : 70.8 %) :



et les implications duales en résultant.



De même pour la seconde enquête :

seuil de cohésion principal : ( PCU de longueur  $\leq 2$  ) :  $S_i = 10$

degré de cohésion principal : 54.8 % (221 réponses retenues sur 403)

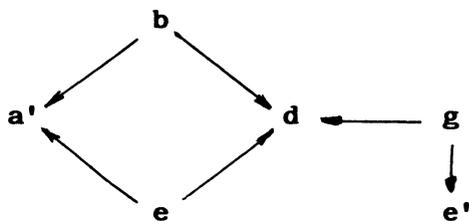
seuil secondaire : ( PCU de longueur  $\leq 3$  )  $S_j = 6$

degré de cohésion secondaire : 65.2 %

Au premier seuil correspondent les PCU :

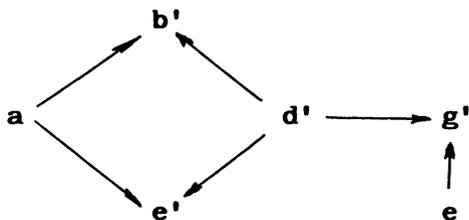
c f ab ae bd' d'e d'g eg

et les implications logiques :



c'      f'

B = 10 ( 2.5 % )  
population  
retenue : 54.8 %



dont l'interprétation socio-psychologique est assez facile mais fort différente de celle de la première enquête.

Pour une bonne moitié de la population interrogée et répondant négativement aux questions c et f, une réponse négative à la question d implique des réponses négatives aux questions b, e et g.

Nous ne nous attacherons pas à rechercher les causes de ces différences importantes entre les deux enquêtes, le passage du protocole

aux schémas d'implications étant ici notre seul objectif.

Nous nous bornerons à remarquer qu'en dehors des différences entre les graphes d'implications, la valeur et l'écart entre les deux indices de cohésion permettent de comparer utilement plusieurs enquêtes du point de vue de l'homogénéité des échantillons ou dans les cas extrêmes du point de vue de la fiabilité des enquêtes (écart entre les degrés de cohésion d'enquêtes ,ou de sous-enquêtes, parallèles et comparaison avec la cohésion nulle correspondant à une distribution uniforme des réponses).

#### BIBLIOGRAPHIE

- (1) FLAMENT C., L'analyse Booléenne de questionnaire, Paris, Mouton, 1976.
- (2) GUENOCHÉ A., "Fonctions booléennes sur un tableau en 0/1", Colloque Analyse des données et informatique Versailles 1985.
- (3) LE VIF - DI MARSO Enquête " La visite du Pape en Belgique" Bruxelles 1985.
- (4) PANORAMA - EDIT Opiniepeiling "Het bezoek van de Paus in België " Brussel 1985.
- (5) VAN BUGGENHAUT J., "Software voor booleaanse analyse" CMP - VUB Brussel 1986.

Ce programme, en cours de développement au Centre for Manpower Planning de la Vrije Universiteit Brussel, est rédigé en (GW)BASIC sur micro-ordinateur Olivetti M24 (IBM- compatible , sous MS-DOS). Provisoirement seulement disponible en néerlandais, il comporte également une version compilée permettant un traitement plus rapide des données.

Les effectifs des  $2^n$  patrons de réponses ( $n = 11$ ) sont introduits manuellement ou via un fichier intermédiaire. Le programme détermine tous les seuils de dichotomisation possibles et calcule les PCU pour tous les seuils supérieurs ou égaux au second seuil de cohésion défini plus haut.

Il compare également les diverses dichotomies associées à la méthode de la case vide et à la méthode des seuils de cohésion.