RAIRO. RECHERCHE OPÉRATIONNELLE

PHILIPPE FACON

Un théorème de décomposition des questionnaires optimaux identifiant des monômes booléens. Applications

RAIRO. Recherche opérationnelle, tome 13, nº 4 (1979), p. 391-412

<http://www.numdam.org/item?id=RO_1979__13_4_391_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 13, n° 4, novembre 1979, p. 391 à 412)

UN THÉORÈME DE DÉCOMPOSITION DES QUESTIONNAIRES OPTIMAUX IDENTIFIANT DES MONOMES BOOLÉENS. APPLICATIONS (*)

par Philippe FACON (1)

Résumé. — *L'article formalise sous le nom de problèmes d'identification booléens divers problèmes comme celui de la traduction des tables de décision. Le but étant de trouver pour chaque problème un processus d'interrogation de coût minimal, une étude utilisant le formalisme de la théorie des questionnaires est présentée. Un théorème et ses corollaires sont donnés; ils permettent d'accélérer le temps de recherche d'un questionnaire optimal par des règles de décomposition du problème initial et de choix de questions.*

Abstract. — *The paper formalizes with the name of boolean identification problems different problems like that of the translation of decision tables. The aim is to find a testing procedure of minimal cost for each problem; a study using the formalism of the questionnary theory is presented. A theorem and its corollaries are given; they allow to reduce the research time of an optimal questionnary, by some rules of decomposition of the initial problem and of choices of questions.*

1. INTRODUCTION

Le point de départ de cette étude a été la préoccupation de réduire le temps d'exécution de programmes générés à partir d'expressions du type conditions/actions (tables de décision, succession de règles logiques, etc.).

Par exemple :

si CODE $=10$ et NUM <49 faire A_1;

si CODE $\neq 10$ faire A_2;

si NUM $\geqq 49$ et CODE $=10$ faire A_3.

On peut toujours, moyennant certaines précautions, se ramener au cas de conditionnements booléens disjoints et de somme égale à 1.

Un problème identique est la répartition d'éléments en groupes selon différents critères; soit ainsi un responsable d'une colonie de vacances désirant répartir des enfants en cinq groupes :

— randonneurs ne pratiquant pas le football;

— cyclistes n'aimant pas la randonnée;

(*) Reçu février 1979.

(1) Institut d'Informatique d'Entreprise, C.N.A.M., Paris.

R.A.I.R.O. Recherche opérationnelle/Operations Research, 0399-0842/1979/391/$ 4.00

— footballeurs non cyclistes;
— pratiquant les trois activités;
— n'en pratiquant aucune.

L'objectif est de poser les questions permettant de réduire la durée du tri.

Dans les deux cas (le temps d'exécution d'un programme étant couramment considéré comme proportionnel au nombre moyen de tests posés) il faut réduire le coût du processus d'interrogation.

Si on ne s'intéresse qu'au coût en temps d'exécution, on peut modéliser ce processus d'interrogation par une arborescence de questions portant sur les valeurs des variables intervenant dans le problème. Une arborescence résoud un problème si à chacune de ses feuilles il est possible d'associer un des « cas » du problème (c'est-à-dire un des groupes à identifier, une des actions à effectuer, etc.); un même cas peut évidemment être associé à plusieurs feuilles, mais à une feuille ne sera associé qu'un seul cas si les conditionnements sont disjoints.

Supposons maintenant, par exemple, les temps d'exécution des tests égaux; le coût sera alors proportionnel au nombre moyen de questions posées. Le premier problème serait ainsi résolu par les arborescences (*fig.* 1 et 2).

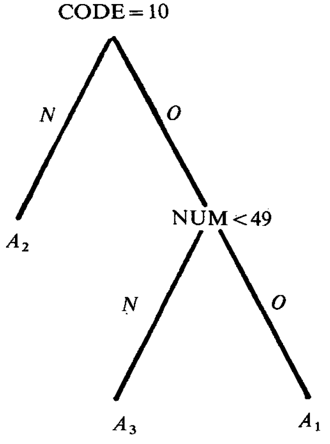

Figure 1

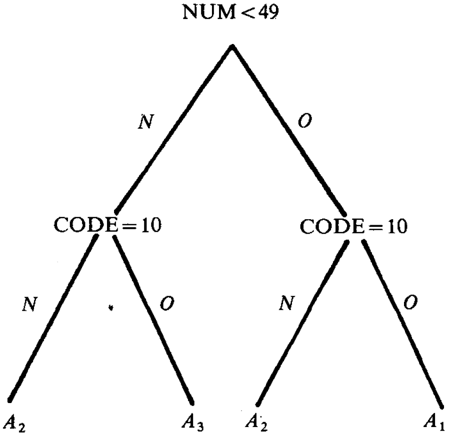

Figure 2

En supposant le temps d'exécution de chaque test égal à 1 et en notant $p(C)$, si C est une condition logique (resp. une fonction booléenne), la probabilité pour que C soit vrai (resp. que $C=1$), le coût de Q_1 est obtenu par :

$$1 \times p(\text{CODE} \neq 10) + 2 \times p(\text{CODE}=10)$$

alors que celui de Q_2 vaut 2.

Q_1 est manifestement meilleur que Q_2, mais il est clair que la connaissance de probabilités peut être nécessaire pour déterminer l'arborescence optimale. Il en est ainsi pour le second problème, qui correspond à l'illustration d'un exemple classique (Kuntzmann [7]) d'ensemble de monômes dont la somme vaut 1 et est pourtant irréductible : $a\bar{b}$, $b\bar{c}$, $c\bar{a}$, abc, $\bar{a}\,\bar{b}\,\bar{c}$. On obtient l'arborescence : Q_3 (*fig.* 10) et les deux autres arborescences déduites de Q_3 par permutation circulaire de a, b, c. Le coût de Q_3, avec les hypothèses précédentes, est donné par :

$$2(p(a\bar{b})+p(c\bar{a}))+3(p(b\bar{c})+p(abc)+p(\bar{a}\,\bar{b}\,\bar{c})).$$

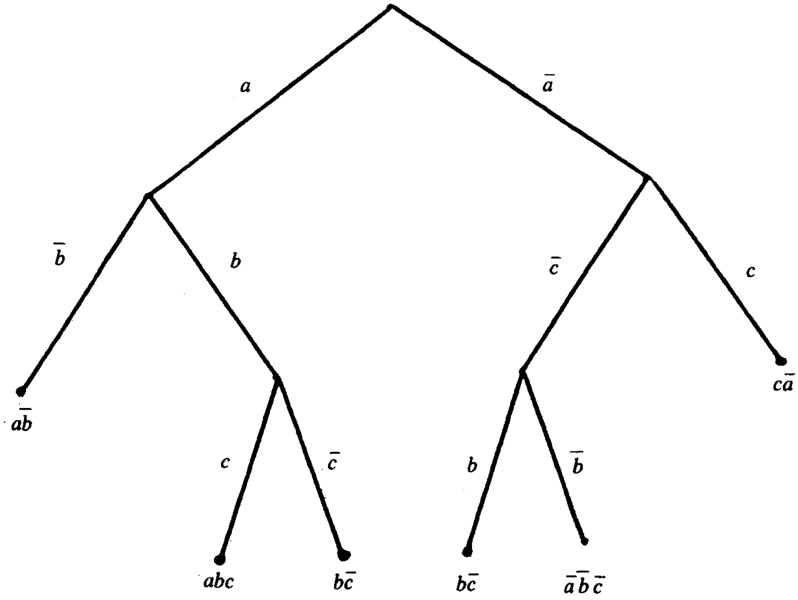

Figure 10

Nous cherchons ainsi de manière générale à résoudre le problème suivant :

— soit un ensemble de monômes booléens disjoints et de somme 1;

— il existe manifestement toujours plusieurs arborescences de questions (chaque question concernant la valeur d'une variable apparaissant dans les monômes) telles qu'à chacune de leur feuille on puisse associer un des monômes, et un seul par les réponses aux différentes questions le long du chemin allant de la racine à cette feuille (on notera que, les monômes étant disjoints et de somme 1, chacun sera alors associé à au moins une feuille); on dispose d'un ensemble de probabilités nous permettant d'évaluer le coût (en particulier le nombre moyen de questions si les coûts des questions sont égaux) de chaque arborescence;

— le problème est alors : quelle est l'arborescence de questions de coût minimal?

Notons que la recherche d'une telle arborescence est un problème polynômial complet [5]. Ce problème a été abordé, sous diverses formes, par maints auteurs. Une très grande partie de la littérature sur les « tables de décision » (on en trouve une synthèse dans U. W. Pooch [9] et E. Humby [4]) lui est consacrée. Les auteurs ont proposé divers algorithmes : la plupart, reposant sur une idée intuitive, recherchent une « bonne arborescence »; ceux qui recherchent une arborescence optimale utilisent en général une méthode PSEP afin d'éviter l'énumération complète des arborescences possibles; tous utilisent des probabilités — qui dans la réalité ne sont pas toujours connues — à chaque étape de l'algorithme.

Nous avons adopté ici une approche de ce problème très différente des précédentes. Nous avons cherché à obtenir des résultats théoriques concernant les arborescences optimales à partir de l'étude de la composition des différents monômes en tant qu'ensembles de variables complémentées ou non. Un théorème a été obtenu; il est basé sur la possibilité de décomposer un problème en sous-problèmes dont les solutions localement optimales font partie d'une solution globalement optimale. On peut en déduire des règles de construction des arborescences optimales; elles sont en particulier indépendantes de toute probabilité.

Certaines conséquences sur le choix ou l'élimination des questions sont directement utilisables par tout algorithme de recherche d'une solution optimale (ou simplement d'une bonne solution). Cependant, les résultats obtenus nécessiteraient une autre étude pour leur application systématique sous forme d'un algorithme; si l'on recherche une arborescence optimale, un tel algorithme fait nécessairement intervenir la connaissance de probabilités à certaines étapes lorsque les décompositions deviennent impossibles. Nous ne précisons pas ici comment on passe de la forme la plus générale du problème (ensemble de fonctions booléennes) à un ensemble de monômes disjoints et de somme 1, ou encore comment les résultats obtenus pour cette dernière forme s'appliquent à la forme la plus générale (sur ces deux points qui ne présentent pas de difficultés théoriques, *voir* Facon [2]) (2).

(2) Notons qu'on peut obtenir, à partir d'un problème quelconque, plusieurs problèmes d'identification booléens.

Concernant le formalisme employé, un problème, dans le sens évoqué plus haut, est formalisé sous le nom de « problème d'identification booléen » et une arborescence de questions sous le nom de « questionnaire booléen ». Cette dernière notion est implicitement utilisée par les auteurs sous des noms divers : arbres de recherche binaires, arbres de décision, etc.

Le vocabulaire de la théorie des questionnaires a été préféré à d'autres pour sa précision et sa rigueur mais la théorie elle-même a été très peu utilisée. Bouchon [1], s'appuyant sur les travaux de Picard [8], introduit la notion de « questionnaire logique » avec l'objectif, différent du nôtre, de représenter des expressions logiques (l'auteur propose cependant un algorithme concernant le problème traité ici; cet algorithme, de même type que certains utilisés dans la traduction des tables de décision, est cité en exemple en 5.5). Les questionnaires booléens sont des types particuliers (entre autres toujours arborescents) de questionnaires logiques. Il n'est pas question ici de problèmes de représentation (qui peuvent nécessiter l'utilisation de « latticiels » ou d'« hémi-questionnaires ») : il s'agit par contre de minimiser le coût en temps d'exécution d'un processus d'interrogation; on peut ainsi ne considérer que des questionnaires arborescents (correspondant aux arborescences des chemins de questionnaires quelconques) et rechercher parmi eux celui, ou ceux de coût minimal.

2. PROBLÈMES D'IDENTIFICATION BOOLÉENS

2.1. Définition

Nous appellerons *problème d'identification booléen* $\mathcal{D}=(\mathcal{A}, \mathcal{P}, \mathcal{M})$ un triplet constitué de :

— $\mathcal{A}$ ensemble fini de variables booléennes :

$$=\{a_1, a_2, \ldots, a_i, \ldots, a_n\};$$

— $\mathcal{P}(\mathcal{A})$, ou plus simplement $\mathcal{P}$, ensemble de probabilités non nulles affectées aux 2^n monômes canoniques $\mu_j=\tilde{a}_1\,\tilde{a}_2, \ldots, \tilde{a}_i, \ldots, \tilde{a}_n$ ($\tilde{a}_i$ désigne a_i ou $\bar{a}_i$);

— $\mathcal{M}$, ensemble de monômes booléens m_k, produits de variables de $\mathcal{A}$, complémentées ou non, tels que :

$$\forall\, i\neq j,\quad m_i.m_j=0 \qquad \text{et} \qquad \sum_k m_k=1.$$

On note $p(m_k)$ la probabilité de m_k : probabilité que m_k soit égal à 1, somme des probabilités des monômes canoniques μ_j, inférieurs ou égaux à m_k. Dans la suite M étant un ensemble de monômes quelconques, on notera $\dot{+}M$ leur somme booléenne. On notera $\tilde{\mathscr{A}}$ l'ensemble des variables de $\mathscr{A}$ et leurs compléments.

Exemple :

$$\mathscr{A}=\{a_1, a_2, a_3\},$$

$$\mathscr{P} : p(\tilde{a}_1\,\tilde{a}_2\,\tilde{a}_3)=1/8 \text{ (les huit probabilités sont identiques)},$$

$$\mathscr{M}=\{a_1\,\bar{a}_2,\ a_2\,\bar{a}_3,\ a_3\,\bar{a}_1,\ a_1\,a_2\,a_3,\ \bar{a}_1\,\bar{a}_2\,\bar{a}_3\}.$$

On obtient :

$$p(a_1\,\bar{a}_2)=p(a_2\,\bar{a}_3)=p(a_3\,\bar{a}_1)=1/4,$$

$$p(a_1\,a_2\,a_3)=p(\bar{a}_1\,\bar{a}_2\,\bar{a}_3)=1/8.$$

2.2. Composition de problèmes d'identification booléens

(*a*) Nous donnons d'abord les deux définitions préliminaires suivantes :

1° deux problèmes $(\mathscr{A}_1, \mathscr{P}_1, \mathscr{M}_1)$ et $(\mathscr{A}_2, \mathscr{P}_2, \mathscr{M}_2)$ seront dits *disjoints* si : $\mathscr{A}_1 \cap \mathscr{A}_2=\emptyset$;

2° un ensemble de probabilités non nulles $\mathscr{P}(\mathscr{A})$ affectées à 2^n monômes canoniques $\mu=\tilde{a}_1\,\tilde{a}_2\,\ldots\,\tilde{a}_n$ sera dit *induit* par deux ensembles de probabilités non nulles $\mathscr{P}_1(\mathscr{A}_1)$ et $\mathscr{P}_2(\mathscr{A}_2)$ si :

– $\mathscr{A}_1$ et $\mathscr{A}_2$ partitionnent $\mathscr{A}$;

– $\forall\mu$: si μ_1 désigne le monôme canonique sur $\mathscr{A}_1 \succ \mu$ et si μ_2 désigne le monôme canonique sur $\mathscr{A}_2 \succ \mu$ $(\mu=\mu_1\,\mu_2)$ alors : $p(\mu)=p_1(\mu_1).p_2(\mu_2)$.

Autrement dit il y a indépendance entre les variables de $\mathscr{A}_1$ et les variables de $\mathscr{A}_2$.

(*b*) Nous définissons maintenant une opération binaire dans l'ensemble des problèmes d'identification booléens :

1° soient deux problèmes $\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M})$ et $\mathscr{D}_1=(\mathscr{A}_1, \mathscr{P}_1, \mathscr{M}_1)$ disjoints. Soit $M \subseteq \mathscr{M}$. On appellera *composition de* $(\mathscr{A}, \mathscr{P}, \mathscr{M})$ *par* $(\mathscr{A}_1, \mathscr{P}_1, \mathscr{M}_1)$ *en* M, noté $\mathscr{D} \underset{M}{\triangle} \mathscr{D}_1$, le problème $(\mathscr{A}', \mathscr{P}', \mathscr{M}')$ avec :

$$\mathscr{A}'=\mathscr{A} \cup \mathscr{A}_1,$$

$$\mathscr{P}' \text{ induit par } \mathscr{P} \cup \mathscr{P}_1,$$

$$\mathscr{M}'=\{m_k \mid m_k \in \mathscr{M}-M\} \cup \{m_k\,m_l^1 \mid m_k \in M,\ m_l^1 \in \mathscr{M}_1\}.$$

Si $M=\mathscr{M}$, on parlera de composition des deux problèmes, qu'on notera $\mathscr{D} \triangle \mathscr{D}_1$;

2° soit $\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M})$ un problème d'identification booléen, M_1, M_2, ..., M_k des sous-ensembles disjoints de $\mathscr{M}$, et $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_k$ des problèmes disjoints de $\mathscr{D}$.

L'expression $(\ldots(\ldots((\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1) \underset{M_2}{\triangle} \mathscr{D}_2) \ldots) \underset{M_k}{\triangle} \mathscr{D}_k$ est indépendante de la permutation considérée sur l'ensemble des indices $\{1, 2, \ldots, k\}$. On la notera donc : $\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \ldots \underset{M_k}{\triangle} \mathscr{D}_k$ et on parlera de composition de $\mathscr{D}$ par $\mathscr{D}_1$ en M_1, $\mathscr{D}_2$ en $M_2 \ldots \mathscr{D}_k$ en M_k.

Exemple :

$$\mathscr{D} = (\mathscr{A}, \mathscr{P}, \mathscr{M}),$$

$$\mathscr{M} = \{a\bar{b}, b\bar{c}, c\bar{a}, abc, \bar{a}\bar{b}\bar{c}\},$$

$$M_1 = \{a\bar{b}\},$$

$$M_2 = \{b\bar{c}, c\bar{a}\},$$

$$\mathscr{D}_1 = (\mathscr{A}_1, \mathscr{P}_1, \mathscr{M}_1) \quad \text{avec} \quad \mathscr{M}_1 = \{e, \bar{e}\},$$

$$\mathscr{D}_2 = (\mathscr{A}_2, \mathscr{P}_2, \mathscr{M}_2) \quad \text{avec} \quad \mathscr{M}_2 = \{ef, e\bar{f}, \bar{e}\},$$

alors :

$$\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 = (\mathscr{A}', \mathscr{P}', \mathscr{M}') \quad \text{avec} \quad \mathscr{A}' = \{a, b, c, e, f\},$$

$$\mathscr{P}' \text{ induit par } \mathscr{P}, \mathscr{P}_1 \text{ et } \mathscr{P}_2,$$

$$\mathscr{M}' = \{a\bar{b}e, a\bar{b}\bar{e}, b\bar{c}ef, b\bar{c}e\bar{f}, b\bar{c}\bar{e}, c\bar{a}ef, c\bar{a}e\bar{f}, c\bar{a}\bar{e}, abc, \bar{a}\bar{b}\bar{c}\}.$$

(c) LEMME : *Une condition nécessaire et suffisante pour qu'un problème d'identification booléen $\mathscr{D}'$ soit la composition d'un problème $\mathscr{D}$ par des problèmes $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_K$, soit $\mathscr{D}' = \mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \ldots \underset{M_K}{\triangle} \mathscr{D}_K$ est que :*

– *la suppression dans $\mathscr{M}'$ des variables n'appartenant pas à $\mathscr{D}$ laisse des monômes deux à deux identiques ou disjoints, sans en éliminer;*

– *les variables supprimées et les variables non supprimées sont indépendantes.*

Preuve : La condition est manifestement nécessaire par définition de la composition de problèmes.

Démontrons qu'elle est suffisante. Soient : $\mathscr{D}' = (\mathscr{A}', \mathscr{P}', \mathscr{M}')$ le problème initial; $\mathscr{A}' = \mathscr{A}' \cup \mathscr{A}_1$, où $\mathscr{A}$ est l'ensemble des variables non supprimées et $\mathscr{A}_1$ l'ensemble des variables supprimées; $\mathscr{M} = \{m_k \mid k = 1, 2, \ldots, q\}$ l'ensemble des monômes distincts obtenus à partir de $\mathscr{M}'$ par suppression des variables de $\mathscr{A}_1$.

Donc les m_k sont formés de variables de $\mathscr{A}$; par hypothèse ils sont disjoints :

$$\forall k, k' \in \{1, 2, \ldots, q\}, \qquad k \neq k' \quad \Rightarrow \quad m_k . m_{k'} = 0, \tag{1}$$

$\mathscr{M}'$ est donc de la forme, en désignant par m_l^1 des monômes formés de variables de $\mathscr{A}_1$:

$$\mathscr{M}' = \{m_1 m_l^1 \mid l \in L_1\} \cup \{m_2 m_l^1 \mid l \in L_2\} \cup \ldots$$
$$\cup \{m_k . m_l^1 \mid l \in L_k\} \cup \ldots \cup \{m_{K'} m_l^1 \mid l \in L_{K'}\} \cup \{m_{K'+1}, \ldots, m_q\}.$$

« $\mathscr{D}'$ est un problème d'identification booléen » entraîne

1° $$\dot{+}_{k=1}^{K'} m_k(\dot{+}_{l\in L_k} m_l^1)\dot{+}\dot{+}_{k=K'+1}^{q} m_k=1. \tag{2}$$

$\forall k\in\{1, \ldots, K'\}$, donnons aux variables de $\mathscr{A}$ une valeur telle que $m_k=1$; (1) et (2) entraînent :

$$\dot{+}_{l\in L_k} m_l^1=1. \tag{3}$$

2° $\forall k\in\{1, \ldots, K'\}$, $\forall l, l'\in L_k$, $m_k m_l^1 . m_k m_{l'}^1=0$.

Les m_k étant formés de variables de $\mathscr{A}$, les m_l^1 de variables de $\mathscr{A}_1$, on obtient donc :

$$m_l^1 . m_{l'}^1=0.$$

Donc $\forall k\in\{1, \ldots, K'\}$, $\mathscr{D}_k=(\mathscr{A}_{1k}, \mathscr{P}, \{m_l^1 \mid l\in L_k\})$ est un problème d'identification booléen.

Enfin (2) et (3) entraînent : $\dot{+}_{k=1}^{q} m_k=1$. Donc, avec (1), il résulte que $(\mathscr{A}, \mathscr{P}, \mathscr{M})$ est un problème d'identification booléen. En regroupant les monômes m_k correspondants à des $\mathscr{D}_k$ identiques en les ensembles $M_1, M_2, \ldots, M_K$ de monômes $(K\leqq K')$, on obtient :

$$\mathscr{D}'=\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \underset{M_3}{\triangle} \cdots \underset{M_K}{\triangle} \mathscr{D}_K.$$

2.3. Relation entre problèmes d'identification booléens

Soit deux problèmes

$$\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M}) \qquad \text{et} \qquad \mathscr{D}'=(\mathscr{A}', \mathscr{P}', \mathscr{M}'),$$

avec

$$\mathscr{M}=\{m_1, m_2, \ldots, m_k\},$$

$$\mathscr{P}=\{p(\mu_j)/j=1, \ldots, 2^n\}, \qquad \mathscr{P}'=\{p'(\mu_j')/j'=1, \ldots, 2^{n'}\}.$$

On dira que $\mathscr{D}'$ est *plus fin que* $\mathscr{D}$ si :

- $\mathscr{A}\subseteq\mathscr{A}'$ (donc $\forall j$, $\exists J$ tel que $\mu_j=\dot{+}_{j'\in J} \mu_{j'}'$);
- $\forall j : p(\mu_j)=\sum_{j'\in J} p'(\mu_{j'}')$ si $\mu_j=\dot{+}_{j'\in J} \mu_{j'}'$;
- il existe une partition de $\mathscr{M}'$ en K classes M_k' telles que :

$$\forall k\in\{1, 2, \ldots, K\} : \dot{+}_{m_{k'}'\in M_k'} m_{k'}'=m_k.$$

Il est clair que la relation $\mathscr{R}$: $\mathscr{D}'\mathscr{R}\mathscr{D} \Leftrightarrow \mathscr{D}'$ est plus fin que $\mathscr{D}$ est une relation d'ordre (partiel). On notera $\mathscr{D}' \geqq \mathscr{D}$.

Exemple :

$$\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M}) \qquad \text{avec} \quad \mathscr{M}=\{ab, a\bar{b}, \bar{a}\},$$

$$\mathscr{D}'=(\mathscr{A}', \mathscr{P}', \mathscr{M}') \qquad \text{avec} \quad \mathscr{M}'=\{abe, ab\bar{e}, a\bar{b}, \bar{a}fg, \bar{a}f\bar{g}, \bar{a}\bar{f}\}.$$

Si :

$$p(ab)=p'(abe)+p'(ab\bar{e}),$$

$$p(a\bar{b})=p'(a\bar{b}),$$

$$p(\bar{a})=p'(\bar{a}fg)+p'(\bar{a}f\bar{g})+p'(\bar{a}\bar{f}),$$

alors $\mathscr{D}'$ est plus fin que $\mathscr{D}$.

On montre facilement :

LEMME : Soient

$$\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M}), \qquad \mathscr{D}'=(\mathscr{A}', \mathscr{P}', \mathscr{M}'),$$

$\mathscr{D}' \geqq \mathscr{D} \Leftrightarrow \exists M_1, M_2, \ldots, M_k$ sous-ensembles disjoints de $\mathscr{M}$, $\exists \mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_k$ problèmes disjoints de $\mathscr{D}$, tels que :

$$\mathscr{D}'=\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \ldots \underset{M_k}{\triangle} \mathscr{D}_k.$$

Dans l'exemple précédent

$$M_1=\{ab\}, \qquad M_2=\{\bar{a}\},$$

$$\mathscr{D}_1=(\mathscr{A}_1, \mathscr{P}_1, \mathscr{M}_1) \qquad \text{avec} \quad \mathscr{M}_1=\{e, \bar{e}\},$$

$$\mathscr{D}_2=(\mathscr{A}_2, \mathscr{P}_2, \mathscr{M}_2) \qquad \text{avec} \quad \mathscr{M}_2=\{fg, f\bar{g}, \bar{f}\}.$$

3. QUESTIONNAIRES BOOLÉENS

3.1. Rappel

La théorie des questionnaires, développée par C. F. Picard, définit un questionnaire comme un graphe valué quasi-fortement connexe inférieurement $Q=(S, G, P_G)$ tel que l'ensemble des sommets admet la partition $S=E \cup F$ où E est formé des sommets terminaux (réponses) et F des sommets non terminaux (questions). Un certain nombre d'axiomes précisent cette définition [8]. En particulier, en notant $p(i, j)$ la valuation de l'arc (i, j) Γ_i (resp. Γ_i^{-1}) l'ensemble des successeurs (resp. prédécesseurs) de i :

— $\sum_{j \in \Gamma_i} p(i, j) = \sum_{h \in \Gamma_i^{-1}} p(h, i)$ pour tout $i, i \in F$, tel que Γ_i et Γ_i^{-1} sont non vides;

— $\sum_{i \in E} \sum_{h \in \Gamma_i^{-1}} p(h, i) = 1$.

3.2. Définitions

Soient une arborescence dichotomique (S, G), S étant l'ensemble des sommets et G l'ensemble des arcs, $(\mathscr{A}, \mathscr{P})$ le couple d'un ensemble $\mathscr{A}$ de variables booléennes et d'un ensemble $\mathscr{P}$ de probabilités qui sont affectées aux différents monômes canoniques construits à partir de $\mathscr{A}$. Soit λ une application de G dans $\tilde{\mathscr{A}}$, $\lambda : (i, j) \to \lambda(i, j)$ vérifiant :

— pour tout i non terminal : $\lambda(i, j_1) = \overline{\lambda(i, j_2)}$, j_1 et j_2 étant les deux successeurs de i;

— une variable booléenne n'apparaît jamais deux fois sur le même chemin.

Affectons à tout sommet j de S différent de la racine r le monôme booléen $m(j)$ produit des variables apparaissant sur le chemin allant de r à j; posons

$$\forall j \in S - \{r\}, \qquad p(j) = p[m(j)] \quad \text{[probabilité pour que } m(j) = 1\text{]}.$$

On montre alors que le graphe (S, G) valué sur les arcs par $p : (i, j) \to p(j)$, est un questionnaire. Nous appellerons $Q = (S, G, \mathscr{A}, \mathscr{P}, \lambda)$ un *questionnaire booléen*. Notons qu'il s'agit là d'un type particulier de « questionnaire logique » (défini dans Bouchon [1]).

Exemple :

$$\mathscr{A} = \{a_1, a_2\};$$

$$\mathscr{P} = \{p(a_1 a_2) = p(a_1 \bar{a}_2) = 1/4,\ p(\bar{a}_1 a_2) = 1/12,\ p(\bar{a}_1 \bar{a}_2) = 5/12\};$$

(S, G, λ) (*fig.* 3). On obtient alors le questionnaire (*fig.* 4).

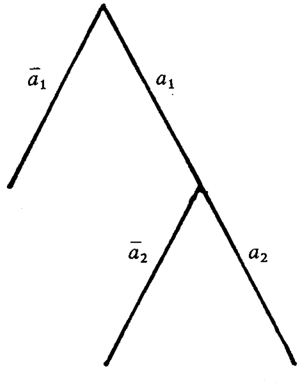

Figure 3

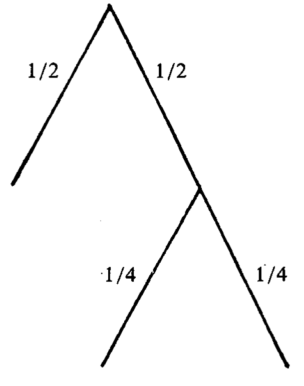

Figure 4

Nous dirons que deux questionnaires booléens $(S_1, G_1, \mathscr{A}_1, \mathscr{P}_1, \lambda_1)$ et $(S_2, G_2, \mathscr{A}_2, \mathscr{P}_2, \lambda_2)$ sont *disjoints* si $\mathscr{A}_1 \cap \mathscr{A}_2 = \emptyset$.

La longueur de cheminement L d'un questionnaire Q [8] est donnée par : $L = \sum_{e \in E} p(e) r(e)$, E étant l'ensemble des réponses de Q, $r(e)$ le rang de e ou le nombre de lettres du monôme $m(e)$. Elle correspond au nombre de questions posées en moyenne.

Une question dont les deux successeurs sont des réponses sera appelée *question terminale*.

Une question origine de deux arcs — values par j et $\bar{j}$ — sera appelée *question* $\tilde{j}$.

3.3. Produits de questionnaires booléens

(*a*) *Rappel :* produit d'arborescences.

Soient deux arborescences A_1 et A_2, e un sommet terminal de A_1. Le produit de A_1 par A_2 en e est l'arborescence obtenue en concaténant A_1 et A_2, e devenant la racine de la sous-arborescence isomorphe à A_2.

(*b*) *Produit de questionnaires booléens :* Soient deux questionnaires booléens disjoints $Q = (S, G, \mathscr{A}, \mathscr{P}, \lambda)$ et $Q_1 = (S_1, G_1, \mathscr{A}_1, \mathscr{P}_1, \lambda_1)$; E étant l'ensemble des réponses de Q, soit $E_1 \subseteq E$. On appellera *produit de Q par Q_1 en E_1*, noté $Q \underset{E_1}{\diamond} Q_1$ le questionnaire booléen : $Q' = (S', G', \mathscr{A}', \mathscr{P}', \lambda')$ avec : (S', G') produit de (S, G) par (S_1, G_1) en E_1, c'est-à-dire en chacun des sommets de E_1,

$$\mathscr{A}' = \mathscr{A} \cup \mathscr{A}_1,$$

$$\mathscr{P}' \text{ induit par } \mathscr{P} \text{ et } \mathscr{P}_1,$$

λ' défini par :

$$\lambda'(i, j) = \lambda(i, j) \qquad \text{pour} \quad (i, j) \in G,$$

$$\lambda'(i, j) = \lambda_1(i, j) \qquad \text{pour} \quad (i, j) \in G_1^*,$$

sous-arborescence isomorphe à G_1.

Si $E_1 = E$ on notera $Q \diamond Q_1$ pour $Q \underset{E_1}{\diamond} Q_1$.

Remarquons qu'il s'agit là d'une forme particulière du produit de questionnaires défini dans Picard [8].

Exemple :

$$Q\ (\textit{fig. } 5) \quad E_1 = \{e_1, e_3\}, \qquad Q_1\ (\textit{fig. } 6);$$

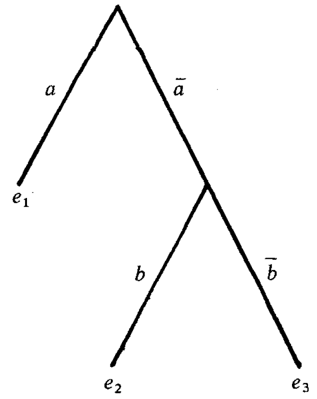

Figure 5

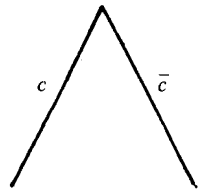

Figure 6

alors :

$$Q \underset{E_1}{\diamond} Q_1 \text{ (\textit{fig.} 7).}$$

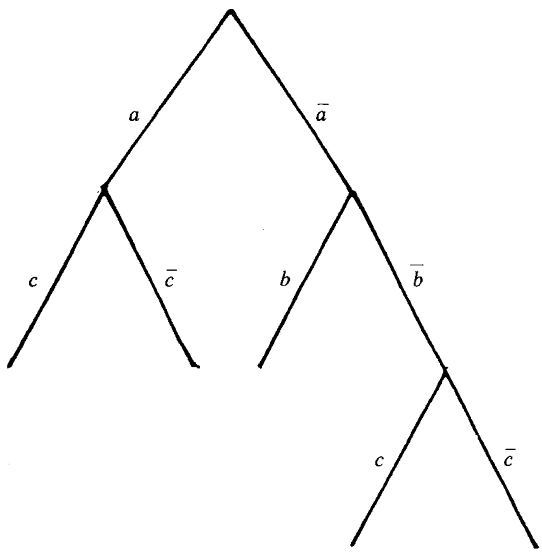

Figure 7

(*c*) *Produit d'un par plusieurs questionnaires booléens :* Soit un questionnaire binaire Q, E l'ensemble de ses terminaux, $E_1, E_2, \ldots, E_k$ des sous-ensembles disjoints de E.

L'expression $(\ldots(\ldots((Q \underset{E_1}{\diamond} Q_1) \underset{E_2}{\diamond} Q_2)\ldots)\ldots) \underset{E_k}{\diamond} Q_k$ est indépendante de la permutation considérée sur l'ensemble des indices $\{1, 2, \ldots, k\}$. On la notera donc

$$Q \underset{E_1}{\diamond} Q_1 \underset{E_2}{\diamond} Q_2 \ldots \underset{E_k}{\diamond} Q_k.$$

4. PROBLÈMES D'IDENTIFICATION BOOLÉENS ET QUESTIONNAIRES BOOLÉENS

4.1. Problème d'identification booléen issu d'un questionnaire booléen

Soient $Q=(S, G, \mathscr{A}, \mathscr{P}, \lambda)$ un questionnaire booléen et E l'ensemble de ses réponses. Alors $(\mathscr{A}, \mathscr{P}, \mathscr{M})$ avec $\mathscr{M}=\{m(e), e \in E\}$ forme un problème d'identification booléen que nous dirons *issu* de Q, et que nous noterons $\mathscr{D}(Q)$.

Exemple : Reprenons le questionnaire booléen donné en exemple : Q (*fig*. 3). On obtient :

$$\mathscr{D}(Q)=(\mathscr{A}, \mathscr{P}, \{\bar{a}_1, a_1 \bar{a}_2, a_1 a_2\}).$$

Nous dirons que deux questionnaires booléens Q_1 et Q_2 sont *équivalents* si

$$\mathscr{D}(Q_1)=\mathscr{D}(Q_2).$$

4.2. Questionnaire booléen résolvant un problème d'identification booléen

(*a*) Nous dirons qu'un questionnaire booléen Q résoud un problème $\mathscr{D}$ si $\mathscr{D}(Q) \geqq \mathscr{D}$. Pratiquement cela signifie qu'à chaque réponse de Q on pourra associer un et un seul monôme de $\mathscr{D}$ et que les probabilités intervenant dans Q sont compatibles avec celles de $\mathscr{D}$.

(*b*) *Dédoublement par une question*. Soient un problème $\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M})$, Q un questionnaire résolvant $\mathscr{D}$. Nous dirons qu'il y a dans Q dédoublement par une question $\tilde{e}$ si il existe des monômes associés à des réponses de Q contenant $\tilde{e}$ et inférieurs à des monômes de $\mathscr{M}$ ne contenant pas $\tilde{e}$ (la question $\tilde{e}$ a été posée « inutilement » pour ces éventualités).

(*c*) *Questionnaire booléen optimal résolvant un problème d'identification booléen*. On dira qu'un questionnaire booléen est optimal pour résoudre un problème $\mathscr{D}$ (ou encore pour $\mathscr{D}$), et on le notera $Q^*(\mathscr{D})$, si :

– Q est un questionnaire résolvant $\mathscr{D}$;

– il n'existe pas de questionnaire résolvant $\mathscr{D}$ de longueur de cheminement inférieure à celle de Q.

Remarquons que cette définition d'un questionnaire optimal par sa longueur de cheminement sous-entend des coûts égaux pour les différentes questions. Cette hypothèse ne semble pas trop restrictive dans la pratique.

Nous laisserons au lecteur le soin de vérifier que la longueur $L^*(\mathscr{D})$ d'un questionnaire booléen optimal pour un problème $\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M})$ admet les bornes suivantes :

$$\sum_{m \in \mathscr{M}} l(m)\, p(m) \leqq L^*(\mathscr{D}) \leqq \operatorname{card} \mathscr{A}.$$

La borne inférieure n'est atteinte que si les monômes se présentent directement sous la forme d'une écriture lexicographique. Notons que si aucune lettre de $\mathscr{A}$ n'apparaît dans tous les monômes de $\mathscr{M}$, en notant $\mathscr{M}_x^-$ l'ensemble des monômes ne contenant pas $\tilde{x}$, et x_0 la lettre telle que $p(\dot{+}\mathscr{M}_x^-)$ soit minimal, cette borne peut être immédiatement améliorée par :

$$L^*(\mathscr{D}) \geqq \sum_{m \in \mathscr{M}} l(m)\, p(m) + p(\dot{+}\mathscr{M}_{x_0}^-)$$

(il y a dédoublement par la première question).

4.3. Restrictions de problèmes d'identification booléens à un monôme et sous-questionnaires booléens

(*a*) *Restriction d'un problème à un monôme.* Soient un problème $\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M})$ et a une variable de $\mathscr{A}$. Nous considérons

$$\mathscr{A}' = \mathscr{A} - \{a\},$$
$$\mathscr{P}' = \{p(\mu'_j = 1/a = 1)\},$$

c'est-à-dire l'ensemble des probabilités conditionnées par $a=1$ des monômes canoniques μ'_j des lettres de $\mathscr{A}'$;

$$\mathscr{M}' = \bigcup_{i \in I} \{m_i\} \cup \bigcup_{k \in K} \{m_k\} \qquad \text{si} \quad \mathscr{M} = \bigcup_{i \in I} \{a m_i\} \cup \bigcup_{j \in J} \{\bar{a} m_j\} \cup \bigcup_{k \in K} \{m_k\}$$

et laisserons une nouvelle fois au lecteur le soin de vérifier que $(\mathscr{A}', \mathscr{P}', \mathscr{M}')$ est un problème d'identification booléen que nous appellerons *restriction du problème* $\mathscr{D}$ *à* a et noterons $\mathscr{D}' = \mathscr{R}_a \mathscr{D}$. On définirait semblablement $\mathscr{R}_{\bar{a}} \mathscr{D}$.

Exemple :

$$\mathscr{D} = (\mathscr{A}, \mathscr{P}, \mathscr{M}); \qquad \mathscr{M} = \{abc, ab\bar{c}, a\bar{b}f, a\bar{b}\bar{f}, \bar{a}\}.$$

Alors $\mathscr{R}_b \mathscr{D}$ admet pour ensemble de monômes $\{ac, a\bar{c}, \bar{a}\}$ et $\mathscr{R}_{\bar{b}} \mathscr{D}$: $\{af, a\bar{f}, \bar{a}\}$.

Généralisons à un monôme quelconque la notion de restriction de problème. Soit un problème $\mathscr{D}=(\mathscr{A}, \mathscr{P}, \mathscr{M})$ et m un monôme formé de lettres de $\mathscr{A}$. Soient maintenant : $\mathscr{R}_m \mathscr{M}$, l'ensemble des monômes de $\mathscr{M}$ de produits avec m non nul et après enlèvement des lettres apparaissant dans m; $\mathscr{A}'$, ensemble des lettres de $\mathscr{R}_m \mathscr{M}$; $\mathscr{P}'$, ensemble des probabilités de $\mathscr{P}$ conditionnées par $m=1$ non nulles.

En remarquant que $\mathscr{R}_{m\bar{a}} \mathscr{M} = \mathscr{R}_{\bar{a}} \mathscr{R}_m \mathscr{M}$, on démontre par récurrence sur la longueur de m que $(\mathscr{A}', \mathscr{P}', \mathscr{M}')$ vérifie bien les hypothèses d'un problème d'identification booléen.

Nous appellerons $(\mathscr{A}', \mathscr{P}', \mathscr{M}')$ restriction de $\mathscr{D}$ à m et le noterons $\mathscr{R}_m \mathscr{D}$.

Exemple :

$$\mathscr{M} = \{ a\bar{b},\ b\bar{c},\ c\bar{a},\ abce,\ abc\bar{e},\ \bar{a}\bar{b}\bar{c} \},$$

Alors

$$\mathscr{R}_{ac} \mathscr{M} = \{ \bar{b},\ be,\ b\bar{e} \},$$

$$\mathscr{A}' = \{ b, e \},$$

$$\mathscr{P}' = \{ p(be/ac),\ p(b\bar{e},\ ac),\ p(\bar{b}e/ac),\ p(\bar{b}\,\bar{e}/ac) \}.$$

(*b*) *Problèmes associés aux sommets d'un questionnaire résolvant un problème.* A tout sommet j d'un questionnaire booléen $Q(\mathscr{D})$ résolvant $\mathscr{D}$, on associe le problème noté $\mathscr{D}(j)$ défini par : $\mathscr{D}(r) = \mathscr{D}$ si r est la racine de $Q(\mathscr{D})$, $\mathscr{D}(j) = \mathscr{R}_{m(j)} \mathscr{D}$ si $j \neq r$, $m(j)$ désignant le monôme booléen produit des variables étiquetant le chemin d'origine r et d'extrémité j. $\mathscr{D}(j)$ sera appelé le *problème associé au sommet de* $Q(\mathscr{D})$. Soit j un sommet non terminal d'un questionnaire booléen Q. Rappelons que le *sous-questionnaire* de Q de sommet j [8] est le questionnaire dont l'arborescence est la sous-arborescence de sommet j de l'arborescence de Q, les probabilités affectées aux arcs étant divisées par $p(j)$. On le notera $\mathscr{R}_j Q$. On montre facilement que si Q est un questionnaire résolvant $\mathscr{D}$, alors $\mathscr{R}_j Q$ résoud $\mathscr{R}_{m(j)} \mathscr{D}$.

Puisque tout sous-questionnaire d'un questionnaire optimal est optimal [8], on peut même affirmer que $\mathscr{R}_j Q^*(\mathscr{D})$ est optimal pour résoudre $\mathscr{R}_{m(j)} \mathscr{D}$.

4.4. Lemme

Soient un problème d'identification booléen $\mathscr{D} = (\mathscr{A}, \mathscr{P}, \mathscr{M})$, $Q^*(\mathscr{D})$ *un questionnaire optimal pour* $\mathscr{D}_1$ *et* $M_1, M_2, \ldots, M_K$ *des sous-ensembles disjoints de* $\mathscr{M}$. *Soient* $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_K$ *des problèmes tous disjoints de* $\mathscr{D}$ *et* $Q_1^*(\mathscr{D}_1), Q_2^*(\mathscr{D}_2), \ldots, Q_K^*(\mathscr{D}_K)$ *des questionnaires booléens optimaux pour résoudre respectivement* $\mathscr{D}_1, \mathscr{D}_2, \ldots, \mathscr{D}_K$. *Alors le problème composé*

$$\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \underset{M_3}{\triangle} \cdots \underset{M_K}{\triangle} \mathscr{D}_K$$

admet pour questionnaire booléen optimal

$$Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q_1^*(\mathscr{D}_1) \underset{M_2}{\diamond} Q_2^*(\mathscr{D}_2) \underset{M_3}{\diamond} \cdots \underset{M_K}{\diamond} Q_K^*(\mathscr{D}_K).$$

Preuve : (*a*) Démontrons d'abord le lemme pour $K=1$. Soit $\mathscr{D}'=\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1$. Nous procéderons par récurrence sur le nombre de variables booléennes de $\mathscr{A}'$. Le résultat est facilement vérifiable pour card $\mathscr{A}'=2$. Soit donc $Q^*(\mathscr{D}')$ un questionnaire booléen optimal pour résoudre $\mathscr{D}'$.

(a_1) Si $Q^*(\mathscr{D}')$ a pour première question $\tilde{b}$, $\tilde{b} \in \mathscr{A}_1$, alors $Q^*(\mathscr{D}')$ est, d'après le dernier résultat de 4.3 (*b*), de la forme

$$Q^*(\mathscr{D}')=B \underset{b}{\diamond} Q^*(\mathscr{R}_b \mathscr{D}') \underset{\bar{b}}{\diamond} Q^*(\mathscr{R}_{\bar{b}} \mathscr{D}'), \tag{1}$$

B étant le questionnaire booléen d'unique question $\tilde{b}$.

Or on remarque que

$$\mathscr{R}_b \mathscr{D}'=\mathscr{D} \underset{M_1}{\triangle} \mathscr{R}_b \mathscr{D}_1,$$
$$\mathscr{R}_{\bar{b}} \mathscr{D}'=\mathscr{D} \underset{M_1}{\triangle} \mathscr{R}_{\bar{b}} \mathscr{D}_1$$

(les probabilités des variables de $\mathscr{D}$ ne changent pas puisqu'elles sont indépendantes de b). Donc, par récurrence, $\mathscr{R}_b \mathscr{D}'$ et $\mathscr{R}_{\bar{b}} \mathscr{D}'$ ayant au moins une variable de moins que $\mathscr{D}'$, on peut, sans augmenter la longueur de cheminement de Q, remplacer

$$Q^*[\mathscr{R}_b \mathscr{D}'] \text{ par } Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{R}_b \mathscr{D}_1),$$
$$Q^*[\mathscr{R}_{\bar{b}} \mathscr{D}'] \text{ par } Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{R}_{\bar{b}} \mathscr{D}_1).$$

On obtient alors un questionnaire booléen optimal pour $\mathscr{D}'$ de la forme

$$B \underset{b}{\diamond} [Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{R}_b \mathscr{D}_1)] \underset{\bar{b}}{\diamond} [Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{R}_{\bar{b}} \mathscr{D}_1)].$$

Or ce questionnaire peut aussi s'écrire :

$$[B \diamond Q^*(\mathscr{D})] \underset{bM_1}{\diamond} Q^*(\mathscr{R}_b \mathscr{D}_1) \underset{\bar{b}M_1}{\diamond} Q^*(\mathscr{R}_{\bar{b}} \mathscr{D}_1).$$

Il est équivalent au questionnaire

$$[Q^*(\mathscr{D}) \diamond B] \underset{bM_1}{\diamond} Q^*(\mathscr{R}_b \mathscr{D}_1) \underset{\bar{b}M_1}{\diamond} Q^*(\mathscr{R}_{\bar{b}} \mathscr{D}_1) \tag{2}$$

(1) et (2) ont même longueur de cheminement.

Considérons maintenant le questionnaire

$$Q^*(\mathscr{D}) \underset{M_1}{\diamond} [B \underset{b}{\diamond} Q^*(\mathscr{R}_b \mathscr{D}_1) \underset{\bar{b}}{\diamond} Q^*(\mathscr{R}_{\bar{b}} \mathscr{D}_1)]. \tag{3}$$

Il est clair que l'ensemble des monômes du problème issu de (3) est obtenu à partir de l'ensemble des monômes du problème issu de (2) en remplaçant les couples $\{ bm, \bar{b}m / m \in \mathscr{M} - M_1 \}$ de ce dernier par $\{ m / m \in \mathscr{M} - M_1 \}$. (2) étant un questionnaire résolvant $\mathscr{D}'$ et $\tilde{b}$ n'apparaissant pas dans les monômes de $\mathscr{M} - M_1$, (3) est donc aussi un questionnaire résolvant $\mathscr{D}'$. En outre ce questionnaire a une longueur de cheminement inférieure de $p(\dot{+}(\mathscr{M} - M_1))$ à celle du questionnaire initial, puisque les terminaux associés à des monômes de $\mathscr{M} - M_1$ voient leur rang diminué de 1 (la question $\tilde{b}$ n'est plus posée), le rang des autres terminaux demeurant identique. Le questionnaire obtenu est de la forme $Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q_1$. Or Q_1 est un questionnaire résolvant $\mathscr{D}_1$ puisque à chaque réponse e de $Q^*(\mathscr{D})$ de monôme associé inférieur à $\dot{+} M_1$, on a $\mathscr{R}_e \mathscr{D} = \mathscr{D}_1$ Q_1 est nécessairement optimal pour $\mathscr{D}_1$ sinon on pourrait réduire encore la longueur de Q. Il existe donc bien un questionnaire booléen optimal pour $\mathscr{D}'$ de la forme

$$Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{D}_1).$$

Remarque : Si $M_1 \neq \mathscr{M}$, alors $p(\dot{+}(\mathscr{M} - M_1)) \neq 0$ et la démonstration précédente prouve qu'il n'existe aucun questionnaire optimal commençant par une lettre de $\mathscr{A}_1$.

(a_2) Si Q a pour première question $\tilde{a}$, $a \in \mathscr{A}$, alors $Q^*(\mathscr{D}')$ est de la forme

$$Q^*(\mathscr{D}') = A \underset{a}{\diamond} Q^*(\mathscr{R}_a \mathscr{D}') \underset{\bar{a}}{\diamond} Q^*(\mathscr{R}_{\bar{a}} \mathscr{D}'),$$

A étant le questionnaire d'unique question $\tilde{a}$. Or on remarque que :

$$\mathscr{R}_a \mathscr{D}' = \mathscr{R}_a \mathscr{D} \underset{M_1}{\diamond} \mathscr{D}_1,$$

$$\mathscr{R}_{\bar{a}} \mathscr{D}' = \mathscr{R}_{\bar{a}} \mathscr{D} \underset{M_1}{\diamond} \mathscr{D}_1.$$

En appliquant la récurrence, on obtient alors un questionnaire optimal de forme

$$A \underset{a}{\diamond} [Q^*(\mathscr{R}_a \mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{D}_1)] \underset{\bar{a}}{\diamond} [Q^*(\mathscr{R}_{\bar{a}} \mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{D}_1)].$$

Or ce questionnaire est de la forme

$$Q \underset{M_1}{\diamond} Q^*(\mathscr{D}_1).$$

Q résoud $\mathscr{D}$ sinon $Q \underset{M_1}{\diamond} Q^*(\mathscr{D}_1)$ ne pourrait résoudre $\mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1$. Q est optimal pour $\mathscr{D}$ sinon on pourrait réduire encore la longueur de $Q \underset{M_1}{\diamond} Q^*(\mathscr{D}_1)$.

On obtient donc bien un questionnaire optimal sous la forme

$$Q^*(\mathscr{D}) \underset{M_1}{\diamond} Q^*(\mathscr{D}_1).$$

(*b*) La même démonstration peut être faite pour

$$\mathscr{D}' = \mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \dots \underset{M_K}{\triangle} \mathscr{D}_K.$$

REMARQUE : Dans ce cas, si $\bigcup_{i=1}^{K} M_i \neq \mathscr{M}$ tout questionnaire optimal commence nécessairement par une lettre de $\mathscr{A}$; si $\bigcup_{i=1}^{K} M_i = \mathscr{M}$ tout questionnaire commence nécessairement par une lettre de $\mathscr{A}$ ou par une lettre b telle que : $\forall \mathscr{A}_i$, $b \in \mathscr{A}_i$. En effet, à partir d'un questionnaire commençant par une autre lettre, on obtient par les transformations précédentes un questionnaire de longueur de cheminement strictement inférieure.

5. QUESTIONNAIRE OPTIMAL POUR UN PROBLÈME D'IDENTIFICATION BOOLÉEN DÉCOMPOSABLE

5.1. Théorème

Si, dans un problème d'identification booléen $\mathscr{D}'$, la suppression de certaines variables, indépendantes des variables non supprimées, laisse des monômes deux à deux identiques ou disjoints, sans en éliminer, alors il existe un questionnaire optimal pour $\mathscr{D}'$, produit d'un questionnaire optimal où n'apparaissent que les variables non supprimées par des questionnaires optimaux où n'apparaissent que les variables supprimées.

Démonstration : Le lemme 2.2(*c*) entraîne que $\mathscr{D}'$ est de la forme

$$\mathscr{D}' = \mathscr{D} \underset{M_1}{\triangle} \mathscr{D}_1 \underset{M_2}{\triangle} \mathscr{D}_2 \dots \underset{M_k}{\triangle} \mathscr{D}_k.$$

On peut donc appliquer le lemme 4.4 et obtenir le questionnaire optimal pour $\mathscr{D}'$:

$$Q^*(\mathscr{D}')=Q^*(\mathscr{D})\underset{M_1}{\diamond}Q^*(\mathscr{D}_1)\underset{M_2}{\diamond}Q^*(\mathscr{D}_2)\ldots\underset{M_k}{\diamond}Q^*(\mathscr{D}_k).$$

Dans $Q^*(\mathscr{D})$ n'apparaissent que des variables non supprimées et dans $Q^*(\mathscr{D}_k)$ n'apparaissent que des variables supprimées.

REMARQUE : Les remarques faites dans le lemme précédent permettent d'affirmer qu'il n'existera aucun questionnaire optimal commençant par une question $\tilde{b}$ telle que $b \notin \mathscr{A}$; $\exists\, k : b \notin \mathscr{A}_k$.

5.2. Corollaire 1

Si une variable $\tilde{a}$ apparaît dans tous les monômes d'un problème d'identification booléen $\mathscr{D}$, alors il existe un questionnaire optimal pour $\mathscr{D}$ commençant par la question $\tilde{a}$.

En effet :

– si a est indépendante des autres variables, il s'agit du cas particulier du théorème où, après suppression de toutes les variables excepté a, il reste dans $\mathscr{M}$, après réduction, $\{a, \bar{a}\}$;

– sinon, une démonstration directe, par récurrence sur le nombre de lettres, montre que la proposition demeure valable.

5.3. Corollaire 2

Si un problème d'identification booléen $\mathscr{D}$ demeure inchangé par permutation de e et $\bar{e}$, e étant indépendante des autres variables, alors il existe un questionnaire optimal pour $\mathscr{D}$ dont toutes les questions $\tilde{e}$ sont des questions terminales.

Il s'agit là d'un second cas particulier du théorème précédent : seule la lettre e est supprimée. Notons que si $\tilde{e}$ n'apparaît pas dans tous les monômes, il n'existe aucun questionnaire optimal ayant $\tilde{e}$ comme première question. Cette remarque étant appliquée aux différents sous-questionnaires d'un questionnaire optimal pour $\mathscr{D}$, on peut même affirmer que dans aucun questionnaire optimal pour $\mathscr{D}$ il n'existe de dédoublement par $\tilde{e}$.

5.4. Remarque

Tout problème à deux variables est décomposable.

Les problèmes indécomposables à trois variables sont, à une permutation près, de la forme $a\bar{b}$, $b\bar{c}$, $c\bar{a}$, abc, $\bar{a}\,\bar{b}\,\bar{c}$.

Les problèmes courants étant rarement à plus de six ou sept variables, il suffira la plupart du temps d'appliquer les corollaires 1 et 2 et de résoudre les sous-problèmes à trois variables.

5.5. Exemple

Supposons que dans le deuxième exemple décrit en introduction, chaque groupe, à l'exception du groupe ne pratiquant aucune des trois activités, soit subdivisé en deux sous-groupes, l'un pratiquant le tennis, l'autre pas. Supposons d'autre part les probabilités pour chaque enfant de pratiquer un sport toutes indépendantes et égales à $1/2$. On aura donc $4\times 2+1=9$ groupes. Avec des notations évidentes (par exemple $R\overline{F}T$ représentera les randonneurs ne jouant pas au football mais jouant au tennis), ces groupes sont caractérisés par les monômes booléens :

$$R\overline{F}T,\ R\overline{F}\,\overline{T},\ C\overline{R}T,\ C\overline{R}\,\overline{T},\ F\overline{C}T,\ F\overline{C}\,\overline{T},\ RFCT,\ RFC\overline{T},\ \overline{R}\,\overline{F}\,\overline{C}.$$

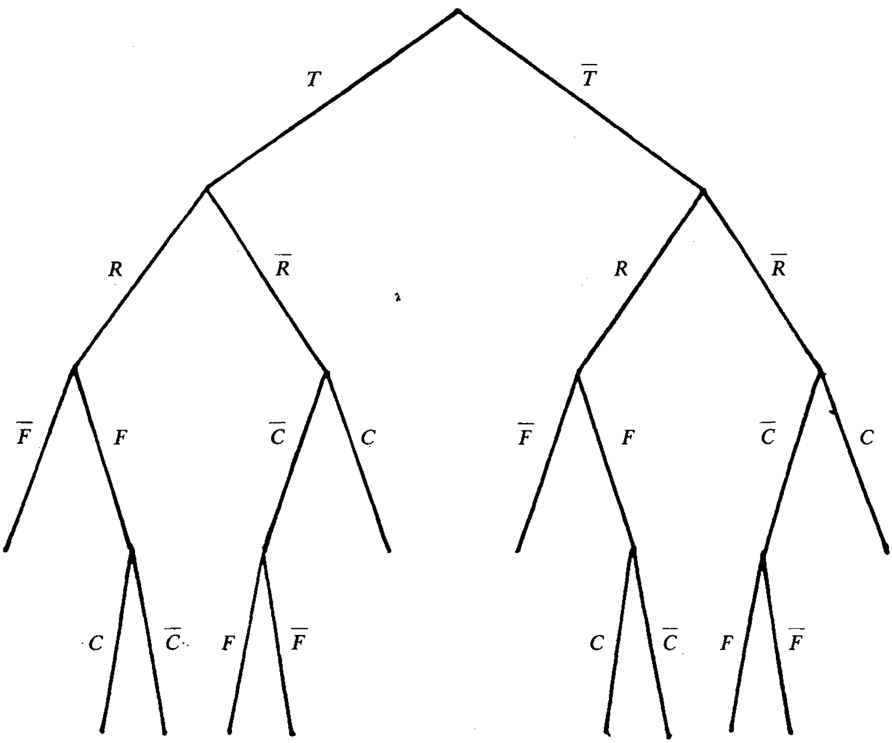

Figure 8

En appliquant l'algorithme le plus fréquemment utilisé pour ce genre de problème (choisir comme première question une variable telle que la probabilité des monômes où apparaît cette variable soit maximale, et continuer ainsi dans les sous-problèmes obtenus) on obtient, en appelant $\mathcal{M}_{\tilde{X}}$ l'ensemble des monômes où apparaît $\tilde{X}$:

$$p(\dot{+}\mathcal{M}_{\tilde{R}})=p(\dot{+}\mathcal{M}_{\tilde{F}})=p(\dot{+}\mathcal{M}_{\tilde{C}})=3/4; \qquad p(\dot{+}\mathcal{M}_{\tilde{T}})=7/8$$

et le questionnaire (*fig.* 8) (les autres questionnaires optimaux s'obtenant par permutation de R, F, C) de longueur

$$L=8\times 4/2^4+4\times 3/2^3=3{,}5.$$

Or, en remarquant que la variable T satisfait aux conditions du corollaire 2, ou encore que le problème est la composition d'un problème où n'apparaissent que R, F, C par un problème où n'apparaît que la lettre T, on obtient le questionnaire optimal (*fig.* 9) de longueur

$$L=6\times 4/2^4+5\times 3/2^3=3{,}375.$$

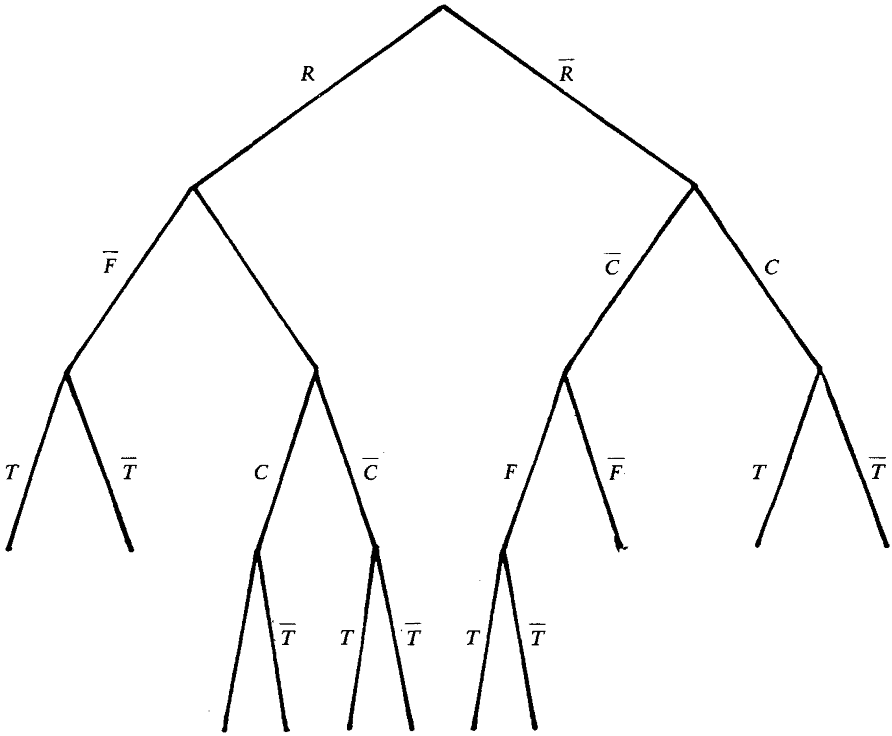

Figure 9

Le gain est de $1/8$, ce qui correspond à $p(\overline{R}\,\overline{C}\,\overline{F})$.

5.6. Conclusion

Ce théorème et ses corollaires peuvent être utilisés :

– soit dans la recherche par une méthode PSES ou PSEP d'un questionnaire optimal; le temps de recherche pourra alors être réduit par décompositions successives du problème initial jusqu'à des sous-problèmes indécomposables dont on cherchera directement des solutions optimales;

– soit intégrés à une méthode heuristique de recherche d'un « bon » questionnaire; ce seront alors les performances des questionnaires trouvés qui pourront être améliorées, par l'élimination de certains choix, et l'optimisation de sous-questionnaires associés à des probabilités élevées.

REMERCIEMENTS

L'auteur tient à remercier le Professeur Étienne Pichat pour ses conseils et le soin qu'il a apporté à la lecture du manuscrit de cet article.

BIBLIOGRAPHIE

1. B. BOUCHON, *Réalisations de questionnaires et propositions logiques*, Thèse de 3e cycle, Paris-VI, 1972.
2. P. FACON, *Conception et optimisation des programmes en informatique de gestion*, Thèse de 3e cycle, Paris-VI, 1978.
3. M. R. GAREY et R. L. GRAHAM, *Performance Bounds on the Splitting Algorithms for Binary Testing*, Acta Informatica, vol. 3, n° 4, 1974, p. 347-355.
4. E. HUMBY, *Programs from Decision Tables*, MacDonald/American Elsevier, 1973.
5. L. HYAFIL et R. L. RIVEST, *Rapport de recherche*, n° 33, I.R.I.A., octobre 1973.
6. A. KAUFMANN et E. PICHAT, *Méthodes mathématiques non numériques et leurs algorithmes*, tome 1, Dunod, Paris, 1977.
7. J. KUNTZMANN, *Algèbre de Boole*, Dunod, Paris, 1969.
8. C. F. PICARD, *Graphes et questionnaires*, tomes 1 et 2, Gauthier-Villars, Paris, 1972.
9. U. W. POOCH, *Translation of Decision Tables*, Computing surveys, vol. 6, juin 1974.