1
ETUDES

Théorie de l'information et Master-Mind

par Gilles DUBOIS, Lycée de l'Isle Adam, 95290

1. Introduction

L'amateur de Master-Mind se sera inévitablement posé les questions suivantes :

— Y a-t-il des combinaisons plus difficiles à trouver que d'autres ?

— Combien d'essais infructueux peut faire en moyenne un bon joueur avant de trouver la combinaison cachée ?

— Doit-on se réjouir d'obtenir telle réponse plutôt que telle autre à la première demande ? Par exemple, trois fiches noires et deux blanches ainsi que quatre fiches noires sont généralement considérées comme de bonnes réponses, c'est-à-dire des réponses indiquant que l'on est près du but. Laquelle des deux est la meilleure ? Un joueur débutant peut se lamenter de recevoir la réponse 0 fiche noire et 0 fiche blanche. A-t-il raison ?

— S'il est possible de prouver que toute combinaison peut être découverte après un certain nombre, disons n, d'essais infructueux, est-il possible de mettre sur pied une stratégie visant à découvrir effectivement toute combinaison avec au plus n demandes ?

La théorie élémentaire de l'information, que l'on présente actuellement comme un appendice de la théorie du calcul des probabilités, permet de répondre à ces questions en restant à un niveau théorique accessible (officiellement, celui du bac C). En effet, toutes les questions posées précédemment sont liées ; il s'agit d'évaluer quantitativement l'information fournie par chaque réponse aux demandes successives.

Le jeu de Master-Mind a des ancêtres célèbres, par exemple le jeu du mot caché que l'on découvre en nommant d'autres mots ayant même nombre de lettres et en notant chaque fois les coïncidences. En outre, le jeu de Master-Mind oblige à une démarche de l'esprit que l'on a l'occasion de faire en de nombreuses situations.

— Une enquête policière — façon Agatha Christie — est une forme de Master-Mind. Il s'agit d'identifier un coupable parmi plusieurs suspects et les petites "cellules grises" d'Hercule Poirot jouent au Master-Mind en procédant par recoupements.

— Un médecin établissant son diagnostic à partir d'une liste de symptômes "joue" au Master-Mind.

2. Mettons-nous d'accord !

Le jeu de Master-Mind auquel il est fait référence est celui distribué en France par CAPIEPA S.A. sous l'appellation commerciale SUPER MASTER-MIND.

Le lecteur est supposé connaître la règle du jeu rappelée dans la notice du constructeur, dont nous reprendrons la terminologie.

Qu'il soit d'abord bien clair que nous n'envisageons ici que la première variante (32768 combinaisons), le codificateur s'interdisant de laisser libre aucune des 5 cases, celles-ci devant donc toutes être remplies avec une des huit couleurs prévues.

La règle du jeu, fournie avec la boîte, et qui est à ma disposition, comporte une petite ambiguïté qu'il importe de lever avant toute chose.

Le décodeur cherche à découvrir la combinaison choisie par le codificateur parmi les 32768 possibles ; pour cela, il fait des "demandes" successives, une "demande" consistant à poser une combinaison de son choix et à recueillir la "réponse" du décodeur, laquelle réponse est elle-même codée. Cette réponse consiste en un certain nombre de fiches noires et un certain nombre de fiches blanches, les fiches noires correspondant aux coïncidences et les fiches blanches aux pions de couleur mal placés.

Il n'y a pas plusieurs interprétations possibles pour utiliser les fiches noires ; par contre, en ce qui concerne le nombre de fiches blanches, il est utile de préciser l'interprétation de la règle. Voici comment peut se comprendre la notice :

— pour le décompte des fiches blanches, il ne faut plus tenir compte des couleurs bien placées dont le nombre est indiqué par les fiches noires ;

— si le codificateur a dissimulé :

Rouge Bleu Bleu Noir Noir

et si le décodeur fait la demande pour

Rouge Jaune Rouge Blanc Blanc

le codificateur répondra simplement : "une fiche noire" et non "une fiche noire et une fiche blanche" sous prétexte que la couleur rouge est une fois bien placée, et une fois mal placée ;

— de plus, si le codificateur a dissimulé

Bleu Bleu Vert Vert Vert

et si le décodeur demande

Jaune Jaune Bleu Bleu Bleu

selon la façon dont on interprète la règle, on peut répondre :

a) 1 fiche blanche correspondant à la couleur bleue, présente, mais mal placée.

b) 2 fiches blanches signifiant qu'en changeant simultanément deux pions de place dans la demande, il est possible d'obtenir 2 fiches noires de plus (on ne précise pas quels pions doivent bouger).

c) 3 fiches blanches signifiant qu'il y a 3 pions qui, individuellement, s'ils étaient placés ailleurs, donneraient une fiche noire.

L'interprétation c) est rejetée, car, non symétrique, elle ne donne pas le même nombre de fiches blanches lorsqu'on inverse les rôles du codificateur et du décodeur, et se prête mal à une mathématisation.

Entre les interprétations a) et b), il semble que b) soit plus courante et c'est donc celle qui sera adoptée pour l'étude qui suit ; la même étude a été faite pour a) et l'auteur la tient à la disposition de qui voudra la consulter.

3. Description de l'espace probabilisé fini qui servira à notre étude mathématique et introduction des notations qui seront utilisées par la suite

Nous numérotons les cases de un à cinq, chaque case recevant une couleur choisie parmi les huit possibles. Un événement élémentaire (une disposition des couleurs) s'identifie avec une application d'un ensemble à 5 éléments dans un ensemble à 8 éléments. Si W désigne l'espace des événements élémentaires, W est donc un ensemble à $8^5 = 32768$ éléments, un élément $w \in W$ sera noté $w = (w_1, w_2, w_3, w_4, w_5)$, w_i désignant la couleur ayant été placée dans la case n° i. Nous admettrons que chaque combinaison a la même probabilité d'être choisie, c'est-à-dire que tous les événements élémentaires sont équiprobables, chacun d'eux ayant alors la probabilité $1/32768$ d'être choisi. Il est aussi commode de considérer l'espace produit $W \times W$, lequel comporte $(32768)^2$ couples (w, w') et la variable aléatoire $f : W \times W \to N \times N$ telle que $f(w, w') = (n, m)$ où n est le nombre de coïncidences entre w et w' (fiches noires) et m le nombre de pions de couleur mal placés selon l'interprétation b) (fiches blanches). En pratique, dans un couple (w, w'), w représente la combinaison dissimulée (laissée au libre choix du codificateur) et w' la "question" posée par le décodeur ; $f(w, w')$ est la "réponse" du codificateur au décodeur.

On remarquera que si w est fixée :

$w' \to f(w, w')$ est une variable aléatoire sur W ,

et si w' est fixé

$w \to f(w, w')$ est une autre variable aléatoire sur W .

4. Un peu de classement

On peut, en première approximation, départager les dispositions selon le nombre de couleurs différentes. Ainsi on trouve les monocolores, (au nombre de 8), les bicolores (840), les tricolores (8400), les quadricolores (16800) et les pentacolores (6720). Par ailleurs, on s'aperçoit que les bicolores sont de deux types : les bicolores 4-1 (280) et les bicolores 3-2 (560). De même, les tricolores sont de 2 types : les tricolores 3-1-1 (3360) et les tricolores 2-2-1 (5040).

Voici les probabilités des différents groupes :

Monocolores	0,000244
Bicolores 4-1	0,008545
Bicolores 3-2	0,017090
Tricolores 3-1-1	0,102539
Tricolores 2-2-1	0,153809
Quadricolores	0,512695
Pentacolores	0,205078

On remarquera que la probabilité d'avoir le type (2-2-1) sachant qu'on a un tricolore est 0,6 , tandis que la probabilité d'avoir le type 3-1-1 est 0,4 . De même, pour les bicolores, le type 3-2 est plus fréquent que le type 4-1 ; probabilités respectives 2/3 et 1/3.

5. Suivons Shannon !

La théorie de l'information, bien qu'ayant maintenant plus de 30 ans d'âge, reste peu connue du grand public, même chez les mathématiciens. Les linguistes, les biologistes, les physiciens et les informaticiens ont vu depuis longtemps ce que cette théorie apportait à leur science, et ils font référence à ce modèle mathématique, même dans des ouvrages de vulgarisation ; certains ont même vu une étude possible de la création artistique par la théorie de l'information. Un des buts de cet article est d'attirer l'attention des enseignants en mathématiques sur la théorie de l'information.

Le lecteur intéressé par l'histoire de cette théorie est renvoyé à la lecture de la préface d'A. Moles de la traduction française du livre de W. Weaver et C.E. Shannon [1].

Il va sans dire que la théorie de l'information, en France comme ailleurs, peut être pour de nombreuses années encore un champ de recherches fructueux, et l'inscription de cette spécialité dans les enseignements de second cycle à l'Université pourrait favoriser des vocations.

Nous présentons maintenant, en raccourci, la base de la théorie élémentaire de l'information, dans laquelle on développe peu à peu le concept de quantité d'information. Cette partie, ne nécessitant pas un bagage mathématique important, a l'avantage de pouvoir être exposée devant un large public. Tout ce qui suit peut être compris d'un lecteur possédant le niveau du Bac C. Nous donnerons, ci-après, cinq extensions successives de la notion de quantité d'information :

a) Quantité d'information liée au choix d'un élément dans un ensemble fini dont le nombre d'éléments est connu.
b) Quantité d'information liée à la réalisation d'un événement de probabilité donnée.
c) Quantité d'information attachée à une variable aléatoire ne prenant qu'un nombre fini de valeurs.
d) Quantité d'information attachée à une variable aléatoire, sachant qu'un événement donné, de probabilité non nulle, est réalisé.
e) Quantité d'information liée à une variable aléatoire connaissant la loi d'une autre variable aléatoire définie sur le même espace probabilisé que la première.

Nous essaierons, à chaque étape, de mettre en évidence le lien entre le nouveau concept et ceux qui précèdent.

a) Considérons un ensemble E ayant 2^n éléments. Chaque élément peut être numéroté en système binaire par un naturel à n chiffres. Ainsi, pour déterminer avec précision un élément d'un tel ensemble, il faut faire n choix successifs entre 0 et 1. Définissant un tel choix comme l'unité d'information (*bit*), on voit qu'il faut $n = \log_2(2^n)$ unités d'information pour préciser un élément particulier de E. Partant de là, si E a un nombre d'éléments m qui n'est plus nécessairement une puissance de 2, nous définirons la quantité d'information liée au choix d'un élément de E par la formule $I = \log_2(m)$; c'est la formule de Hartley, la quantité d'information apparaissant maintenant, non plus comme une grandeur exclusivement entière, mais réelle.

b) Voyons tout de suite une première extension de cette notion. Soit $(\Omega, \mathcal{B}, P)$ un espace probabilisé, et $A \in \mathcal{B}$ un événement de probabilité p. On veut mesurer quantitativement l'information I(A) : "l'élément w de Ω est situé dans A", de sorte que si Ω est un ensemble fini, muni de la tribu $\mathcal{T}(\Omega)$ de toutes ses parties, tous les événements élémentaires étant équiprobables et de probabilité $1/\mathrm{Card}(\Omega)$, on retrouve la quantité $\log_2(m)$ dans le cas particulier où A est un événement élémentaire (un

singleton), avec de plus $I(A)=0$ dans le cas particulier où $A=\Omega$ (le fait d'affirmer que $w\in\Omega$ sachant que $w\in\Omega$ n'apporte aucune information supplémentaire). Une possibilité consiste à prendre $I(A)=\log_2(1/p)$ où $p=P(A)$. Ce choix a beaucoup d'avantages ; nous voyons en particulier que $I(A)\geqslant 0$, et que l'application $A \to I(A)$ est une fonction décroissante de p , ce qu'elle doit être car, si $p=P(A)$ est petit, savoir qu'un élément w de Ω appartient à A apporte beaucoup d'information ; si p est grand, cela en apporte peu. En ce sens, la quantité d'information doit être comprise comme une mesure statistique de l'incertitude.

La base des logarithmes peut être choisie quelconque ; choisir la base revient à fixer l'unité d'information. Quand la base des logarithmes est *deux*, l'unité d'information est le *bit* (abréviation de "binary digit") ; si la base est *dix*, c'est le *dit* ("decimal digit"), qui vaut 3,32 bits ; en base e , l'unité vaut 1,44 bit.

c) Passons immédiatement à une deuxième généralisation du concept de quantité d'information. A partir de maintenant et pour toute la suite, $(\Omega, \mathcal{B}, P)$ désigne un espace probabilisé et X une variable aléatoire ne prenant qu'un nombre fini de valeurs $x_1, x_2, \ldots, x_n$. Pour $1\leqslant i\leqslant n$, on pose

$$A_i = \{w\in\Omega \mid X(w)=x_i\} = X^{-1}(x_i) .$$

Les A_i forment alors un système complet d'événements avec des probabilités $p_i=P(A_i)=P(X=x_i)$ vérifiant $\sum_i p_i=1$.

On définit $I(X)$ comme étant la moyenne des $I(A_i)$ pondérés par les p_i , c'est-à-dire

$$I(X) = \sum_{1\leqslant i\leqslant n} p_i \log_2(1/p_i)$$

avec la convention

$$p_i \log_2(1/p_i) = 0 \qquad \text{si} \quad p_i=0$$

laquelle est d'ailleurs conforme à :

$$\lim_{x\to 0} x\log(1/x) = 0 .$$

$I(X)$ ainsi définie s'appelle la quantité d'information attachée à la variable aléatoire X . On remarquera que $I(X)\geqslant 0$ et que la quantité d'information associée au choix d'un élément dans l'ensemble Ω coïncide avec la quantité d'information de la variable aléatoire application identique de Ω dans Ω .

On montre facilement par des considérations de convexité que la quantité d'information est maximale quand $p_i = \frac{1}{n}$, $1\leqslant i\leqslant n$. L'incertitude sur la valeur de X est alors en effet maximale puisqu'aucune valeur n'est plus probable qu'une autre.

Observons maintenant l'effet du conditionnement :

d) Soit maintenant B un événement de Ω, de probabilité non nulle. La restriction de X à B est une nouvelle variable aléatoire sur l'espace probabilisé B, muni de la tribu trace de $\mathcal{A}$ sur B et de la mesure $P|_B$ où

$$P|_B(C) = P(C \cap B)/P(B) = P(C|B)$$

La quantité $I(X|_B)$ sera notée $I(X|B)$ et appelée "quantité d'information liée à la variable X sachant l'événement B réalisé".

$$I(X|B) = \sum_{1 \leqslant i \leqslant n} P(A_i|B) \log_2(1/P(A_i|B))$$

Naturellement, $I(X|B)=0$ si X est constante sur B, et de plus $I(X|B)$ est égal à $I(X)$ si X est indépendante de B. On remarquera en outre que $I(X|B) \leqslant I(X)$, la connaissance du fait que B est réalisé ne pouvant que diminuer l'incertitude sur X. Nous voyons ainsi que la quantité d'information "conditionnelle" se comporte comme l'intuition le suggère.

e) Nous en venons maintenant à l'ultime généralisation de la théorie élémentaire, et nous introduisons la notion de quantité d'information attachée à une variable aléatoire, connaissant une autre variable aléatoire, définie sur le même espace. Soit donc Y une nouvelle variable sur Ω, prenant des valeurs $y_1, \ldots, y_p$ et soient $B_j = Y^{-1}(y_j)$ et $q_j = P(B_j)$.

La moyenne des $I(X|B_j)$ pondérée par les q_j, soit

$$\sum_{1 \leqslant j \leqslant p} q_j I(X|B_j),$$

est notée $I(X|Y)$, et est appelée "l'information liée à la variable aléatoire X, connaissant la variable aléatoire Y".

On peut facilement voir que $0 \leqslant I(X|Y) \leqslant I(X)$, que $I(X|Y)=0$ s'il existe une relation fonctionnelle entre X et Y, c'est-à-dire si X se factorise via Y, autrement dit si la valeur de X est parfaitement déterminée par la connaissance de la valeur de Y ; à l'autre extrême, $I(X|Y)=I(X)$ chaque fois que X et Y sont des variables indépendantes.

Cela dit, on s'intéresse à la variable aléatoire "couple" (X,Y), laquelle est encore définie sur Ω, mais prend ses valeurs dans un produit cartésien fini. Cette variable aléatoire réalise une partition d'éléments

$$C_{ij} = (X,Y)^{-1}(x_i, y_j) = X^{-1}(x_i) \cap Y^{-1}(y_j)$$

de probabilités

$$r_{ij} = P(X = x_i \text{ et } Y = y_j) \quad \text{où} \quad 1 \leqslant i \leqslant n \text{ et } 1 \leqslant j \leqslant p.$$

Conformément aux définitions précédentes :

$$I(X,Y) = \sum_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant p}} r_{ij} \log_2(1/r_{ij}).$$

Un calcul simple montre que :

$$I(X,Y) = I(X) + I(Y|X) = I(Y) + I(X|Y)$$

ce qui prouve que $I(X,Y) = I(X) + I(Y)$ lorsque X et Y sont deux variables aléatoires indépendantes.

Nous en resterons là pour cette présentation rapide de la théorie de l'information.

6. Un peu de calcul sur le Master-Mind

Le décodeur choisit une disposition, disons w', et attend la valeur de la variable aléatoire $w \to f(w,w')$ qui lui est fournie en réponse par le codificateur. Il importe donc de savoir :

a) la quantité d'information liée à chaque réponse du codificateur ;

b) la quantité d'information liée à la variable aléatoire $w \to f(w,w')$.

Si pr_2 désigne la seconde projection $(w,w') \to w'$ de $\Omega \times \Omega$ dans Ω, il faut donc calculer, conformément aux notations que nous avons adoptées, $I(f|pr_2 = w')$. Naturellement, cette quantité ne dépend que du type de w' ; c'est pourquoi on trouvera ci-après 7 tableaux correspondant aux 7 types différents recensés. Les réponses possibles ne sont pas les mêmes pour tous les types ; le nombre de réponses possibles est fonction croissante du nombre de couleurs différentes de w'. Dans chaque case correspondant à une réponse possible est indiqué :

— N : le nombre de combinaisons donnant cette réponse
— P : la probabilité de cette réponse $P = N/32768$
— L : la quantité d'information liée à la réponse en question $L = \log_2(1/P)$.

Ensuite, il faut faire le calcul de $I(f|pr_2)$, qui a une valeur statistique : c'est la quantité d'information moyenne reçue par le décodeur après une réponse. Cette quantité s'obtient comme moyenne des $I(f|pr_2 = w')$ pondérées par les probabilités des différents types.

Il faut aussi calculer, et c'est fondamental, la quantité d'information à rassembler pour déterminer la combinaison cachée.

Enfin, il peut être important de connaître les quantités d'information liées à la connaissance de certains renseignements sur w, notamment le nombre de couleurs distinctes, car c'est un renseignement que l'on peut obtenir facilement et rapidement. Les résultats sont les suivants :

Quantité d'information à rassembler pour déterminer la combinaison cachée :

$$\log_2(32768) = 15$$

Quantité d'information reçue à chaque réponse, selon le type posé :

Si w' est monocolore	$I(f \mid pr_2 = w') = 1{,}467274$
Si w' est bicolore 4-1	$I(f \mid pr_2 = w') = 2{,}642040$
Si w' est bicolore 3-2	$I(f \mid pr_2 = w') = 2{,}876883$
Si w' est tricolore 3-1-1	$I(f \mid pr_2 = w') = 3{,}059295$
Si w' est tricolore 2-2-1	$I(f \mid pr_2 = w') = 3{,}180208$
Si w' est quadricolore	$I(f \mid pr_2 = w') = 3{,}228778$
Si w' est pentacolore	$I(f \mid pr_2 = w') = 3{,}231553$

Quantité d'information moyenne reçue après une réponse :

$$I(f \mid pr_2) = 3{,}193041$$

Quantité d'information résultant de la connaissance du nombre de couleurs différentes :

Monocolore	I = 11,999998	Quadricolore	I = 0,963826
Bicolore	I = 5,285754	Pentacolore	I = 2,285754
Tricolore	I = 1,963826		

7. Analyse des résultats

Certaines des questions posées dans l'introduction sont mal posées. Elles sont néanmoins reproduites telles quelles, car c'est bien sous cette forme qu'on se les pose pour la première fois.

Ainsi, on comprendra tout de suite que cela n'a aucun sens de comparer deux réponses quant aux quantités d'information qu'elles apportent si l'une est obtenue en posant un monocolore et l'autre en posant un pentacolore. La réponse "0N – 0B" (0 fiche noire et 0 fiche blanche) vaut 0,96 unités d'information si l'on a posé un monocolore, c'est-à-dire très peu relativement à la moyenne de 3,19 que l'on doit recevoir en une réponse, alors que cette même réponse vaut 7,07 unités d'information si l'on a posé un pentacolore. Signalons par contre que, dans tous les cas, "3N – 2B" est nettement meilleure que "4N – 0B".

Notons que l'on peut faire usage des tables au cours du jeu pour estimer la quantité d'information reçue et apprécier la quantité restant à découvrir. Ainsi, si le décodeur commence à poser un pentacolore, et reçoit la réponse "2N – 3B", il aura en une seule fois réuni 10,68 unités d'information, et il lui restera 4,32 unités à rassembler pour déterminer la combinaison cachée. Cependant, il est bon de préciser tout de suite les limites de cette méthode. Raisonnons sur un exemple simple :

Supposons qu'à la première demande, le décodeur reçoive la réponse "1N – 3B" après avoir posé un quadricolore, cette réponse lui rapportant 5,65 unités d'information, et supposons qu'à la question suivante il pose les mêmes quatre couleurs dans un ordre différent et qu'il recueille la réponse "2N – 2B" qui vaut 6,37 unités. Il ne faut pas croire qu'après les deux réponses il aura obtenu globalement la quantité

$5,65 + 6,36 = 12,02$ unités d'information, et cela parce que si w_1' et w_2' sont les 2 quadricolores (différant d'une permutation) qu'il a successivement posés, les 2 variables aléatoires

$$f_1 : w \mapsto f(w, w_1') \quad \text{et} \quad f_2 : w \mapsto f(w, w_2')$$

ne sont pas indépendantes, de sorte que la quantité d'information liée au couple (f_1, f_2) n'est pas la somme des quantités $I(f_1)$ et $I(f_2)$. De même, à la troisième demande, si w_3' est la troisième combinaison posée et si $f_3 : w \mapsto f(w, w_3')$ est la variable aléatoire correspondante, f_3 n'est sans doute pas indépendante de f_1 et de f_2, et encore moins du couple (f_1, f_2), de sorte que, non seulement

$$I(f_1, f_2, f_3) < I(f_1) + I(f_2) + I(f_3)$$

mais

$$I(f_1, f_2, f_3) < I(f_1, f_2) + I(f_3)$$

à supposer qu'on connaisse $I(f_1, f_2)$. Le lecteur voit donc ici la nécessité de tabuler les $I((f_1, f_2) | (w_1', w_2'))$. On signale au passage qu'il y a 29 cas à considérer compte tenu de $I(f_1, f_2) = I(f_2, f_1)$, et que le nombre de couples de réponses possibles ne dépend pas seulement du type de w_1' et du type de w_2', mais encore des coïncidences entre eux. Ce travail, s'il présente un intérêt dans certains cas particuliers, peut difficilement être généralisé ; il ne présente d'ailleurs aucun intérêt sur le plan de la pratique du jeu.

Venons-en maintenant à une autre question : Qu'est-ce qu'un bon score ?

L'idée pourrait venir de diviser la quantité d'information à recueillir, soit 15, par la quantité d'information moyenne reçue après chaque réponse, soit 3,19, puis d'affirmer que l'on doit trouver la bonne réponse après 5 essais infructueux au plus. C'est ce que font en pratique, et sans autre justification, les auteurs traitant de ce sujet (cf. bibliographie) lorsqu'ils étudient des jeux consistant à déterminer un élément caché dans un ensemble de cardinal connu, en utilisant des *tests indépendants* (par exemple, dichotomie). Cet argument ne peut être repris ici, car, comme nous venons de le voir plus haut, les variables aléatoires "réponses du codificateur" ne sont pas indépendantes. En outre, il convient de s'arrêter un peu sur le cas des jeux où les tests sont indépendants. Dans ce cas, on peut dire que la quantité d'information apportée par n tests est n fois supérieure à la quantité d'information apportée par un seul test. Les auteurs dont il est question utilisent alors implicitement le résultat suivant : "Quand l'information est réunie, et à partir de ce moment seulement, on peut déterminer l'élément caché à coup sûr", étant bien entendu que l'on peut trouver avant par chance, intuition, etc.

Voici comment on peut comprendre ce résultat : A tout instant du jeu, la quantité d'information recueillie est en rapport direct avec le nombre des "possibles" par la formule $I = \log_2(1/P(A))$ où A désigne précisément l'ensemble des possibles. Il arrive un moment où I atteint la

valeur $\log_2(m)$, m étant le cardinal de l'ensemble ; à partir de ce moment, toute information supplémentaire est superflue et l'élément peut être trouvé par le raisonnement.

Dans le cas du Master-Mind, la situation est plus complexe car, d'une part, les tests ne sont pas indépendants, d'autre part ils apportent des quantités d'information différentes. Néanmoins, voici un raisonnement qu'on peut tenir : le test qui apporte le plus d'information est celui qui consiste à poser un quadricolore ; ce test apporte en moyenne 3,231 unités d'information. Supposons qu'il soit possible, ce qui est douteux, de répéter 4 tels tests indépendants entre eux. Comme le calcul le prouve, le maximum d'information disponible par ce procédé est en moyenne inférieur à 13 bits ; il manque donc encore deux unités pour déterminer à coup sûr la combinaison cachée. Ainsi, si l'on accepte ce qui vient d'être dit plus haut, nous disposons du résultat négatif suivant :

— Il est impossible de mettre sur pied une stratégie donnant dans tous les cas la réponse exacte à la cinquième demande, après 4 essais infructueux. Ainsi "réussir à tout coup, en 6 coups" apparaît comme le maximum de ce que l'on peut faire au "5 trous, huit couleurs".

Pour finir, parlons de stratégie :

De nombreux joueurs déterminent d'abord, en posant en deux demandes successives, et sans intersection, les 8 couleurs distinctes. Cela peut se faire de deux manières, ou bien 5 couleurs d'abord puis les 3 restantes ensuite, ou bien 4 et puis les 4 autres. Cette méthode a l'avantage de donner rapidement un renseignement précis : le nombre de couleurs distinctes de la combinaison cachée, ainsi que la répartition de ces couleurs dans une partition 5-3 ou 4-4. La méthode 4-4 permet de découvrir rapidement les monocolores et les bicolores dès qu'ils sont détectés. Sans savoir s'il y a mieux à faire, je n'envisagerai pas de stratégie débutant autrement.

Du point de vue du codificateur, a priori, toutes les combinaisons sont aussi difficiles à trouver si on ne dispose d'aucune information complémentaire sur la stratégie qu'utilisera le décodeur ; mais dès qu'on connaît suffisamment le décodeur pour savoir qu'il débutera comme il est dit plus haut, il devient maladroit de dissimuler un monocolore ou un bicolore, par la quantité d'information importante liée à la connaissance de leur nature (un monocolore est obligatoirement découvert après un maximum de 4 essais infructueux). Dans cette optique, et pour répondre à la question "Y a-t-il des combinaisons plus difficiles à trouver que d'autres ?", on ne peut que conseiller le choix des quadricolores, qui sont de loin les plus nombreux.

Du point de vue du décodeur, il n'y a qu'un principe pour une bonne stratégie et c'est le suivant :

Si le décodeur a fait p essais infructueux, en posant successivement w'_1, w'_2, ..., w'_p et en recevant, dans l'ordre, les réponses R_1, R_2, ..., R_p

et si nous notons $(R_i|w_i')$ l'événement : "Le décodeur a reçu la réponse R_i après avoir posé la combinaison w_i' ", il faut déterminer w_{p+1}' de sorte que :

$$I(f|w'=w_{p+1}' \quad \text{et les} \quad (R_i|w_i') \quad \text{pour} \quad 1 \leqslant i \leqslant p)$$

soit le plus près possible de la quantité $I(f|w'=w_{p+1}')$, autrement dit que la variable aléatoire $w \mapsto f(w,w_{p+1}')$ soit "la plus indépendante possible" de l'événement conjonction des $(R_i|w_i')$, au sens qu'elle doit être celle qui laisse le maximum d'incertitude quant à la réponse qu'elle recevra en retour.

Toute la difficulté est là, et ce n'est pas un problème facile que de trouver des critères simples, applicables instantanément pour la détermination du meilleur w_{p+1}' sachant les $(R_i|w_i')$.

L'erreur la plus fréquemment commise, sur le plan tactique, consiste à recherche une certitude, plutôt qu'une quantité d'information ; les certitudes coûtent cher au Master-Mind, comme le montre l'exemple suivant :

Supposons qu'à la première demande le décodeur ait posé

Vert Vert Vert Vert Rouge

et reçu la réponse "Une fiche noire". Cette fiche noire peut correspondre à la couleur rouge bien placée, ou bien à l'un des quatre pions verts bien placés. Voyons ce qu'il en coûte de vouloir lever tout de suite cette ambiguïté. Le décodeur peut poser en deuxième question :

Vert Vert Vert Vert Vert

Il recevra alors (0N − 0B) ou (1N − 0B) en réponse suivant l'une ou l'autre des deux éventualités avec des probabilités 1/6 et 5/6, ce qui donnera une quantité d'information égale à

$$(1/6)\log_2(6) + (5/6)\log_2(6) - (5/6)\log_2(5),$$

soit environ 0,65 unité d'information, c'est-à-dire très peu comparativement à la moyenne de 3,19.

La quête permanente de certitudes est sécurisante et donne l'impression d'avancer, mais, comme nous venons de le voir, cette tactique est coûteuse et contraire à l'esprit du jeu.

Note historique

Nous signalons seulement, en grandes lignes, les faits les plus importants.

Parmi les précurseurs, il faut citer Nyquist et Hartley (1928), des Laboratoires Bell, ainsi qu'A. Einstein à Princeton et V.A. Kotelnikov (Material for the first All-Union Conf. on questions of communications, 1933), mais l'acte de naissance est vraiment l'article *A mathematical*

theory of communication de C.E. Shannon paru en juillet 1948 dans le "Bell Systems Technical Journal", conjointement avec le livre de N. Wiener *Cybernetics* paru en novembre 1947.

L'article de Shannon présentait bien des imperfections, mais il avait au moins l'avantage de jeter les bases de la recherche dans toutes les directions. Par la suite, des mathématiciens professionnels, tels Mac Millan, Feinstein, et finalement Khinchin, s'efforcèrent de donner à la théorie des fondements plus rigoureux. Aujourd'hui encore, l'article de Khinchin (voir [4]) *On the fundamental theorems of Information Theory*, paru en 1956, fait autorité, mais il ne couvre qu'une partie de la théorie de Shannon (celle des sources discrètes). Des chapitres entiers de l'article de Shannon, et de nombreux autres articles du même auteur, manquent encore de rigueur mathématique. La théorie de Khinchin est elle-même susceptible de nombreuses et fructueuses généralisations.

L'étape suivante consiste à définir l'entropie d'une source Markovienne (ou ergodique) stationnaire, mais ces notions ne sont pas utilisées dans cet article. Une bonne connaissance de la théorie générale des processus stochastiques, en général et des chaînes de Markov en particulier est indispensable pour une bonne compréhension du sujet. Le développement de la théorie dans ce sens est parfaitement adapté à l'étude statistique des textes écrits, des codes génétiques et de toutes les chaînes de caractères, mais ne peut pas rendre compte quantitativement de processus continus de transmission d'information comme l'émission de paroles ou d'images. Un deuxième volet de l'article de Shannon ouvre des perspectives dans cette voie, le traitement de ce sujet relève de l'analyse de Fourier. La théorie de l'information est aujourd'hui un sujet très vaste utilisant les outils mathématiques les plus modernes.

Bibliographie

1. *Théorie mathématique de la communication*, par W. Weaver et C.E. Shannon. Traduction française des articles de Shannon cités ci-dessus avec des notes historiques intéressantes dans la préface d'A. Moles. Editeur Retz, 114 avenue des Champs Elysées, 75008 Paris. Collection "Les Classiques des Sciences Humaines". 1975.
2. *Cybernetics* or Control and communication in the animal and the machine, par N. Wiener. Editeur Hermann, Paris. 1958.
3. *Mathematical foundations of information theory*, par A.I. Khinchin. Traduction en anglais par Silverman & Friedman, Dover Books. 1957.
4. *Key papers in the development of information theory*. Edited by David Slepian, I.E.E.E. Press.
5. *Probabilité et Information*, par A.M. Yaglom et I.M. Yaglom. Collection Sigma n° 17. Dunod. 1969.

6. *La théorie mathématique de l'information*, par S. Guiasu et A. Theodorescu. Collection Sigma n° 16. Dunod (épuisé). 1969.

7. *Incertitude et information*, par les mêmes auteurs. Presses de l'Université Laval (Québec). Vuibert. 1971.

8. *Calcul des probabilités*, avec un appendice sur la théorie de l'information, par Renyi. Dunod.

9. Cours Polycopié de G. Battail à l'E.N.S.T. 1977.
Cours de théorie des communications. (Notions de théorie de l'information).