Entropie et Information

F. Didier-IREM Campus Luminy
didier@irem.univ-mrs.fr

Juin 2011

1 L'entropie, mesure du degré d'incertitude

L'étude des évènements aléatoires a pour propriété principale l'absence d'une complète certitude quant à leur réalisation, ce qui entraine une certaine indétermination dans la réalisation des expériences liées à ces évènements. C'est pourquoi nous allons chercher un moyen de mesurer le degré d'incertitude des expériences liées à des évènements aléatoires afin de pouvoir les comparer sous cet angle.

Nous donnons ici une approche intuitive de la fonction H qui mesure le degré d'incertitude d'une expérience, mais on peut définir rigoureusement cette fonction par ses propriétés mathématiques.

Considérons une expérience possédant k issues également probables, il est naturel de penser que lorsque k augmente, le degré d'incertitude doit augmenter. Donc, H doit être une fonction croissante de k. Cette fonction doit aussi s'annuler pour $k = 1$, car dans ce cas il n'y a plus d'incertitude.

De plus, considérons deux expériences indépendantes α et β. Supposons que l'expérience α possède k issues également probables et que l'expérience β possède j issues équiprobables. L'expérience composée $\alpha\beta$ qui consiste à effectuer successivement les deux expériences a $k \times j$ issues équiprobables. Il est naturel de penser que le degré d'incertitude de $\alpha\beta$ est plus grand que celui de l'expérience α, car à l'incertitude de l'expérience α s'ajoute ici l'incertitude concernant l'expérience β.
On doit avoir

$$f(k \times j) = f(k) + f(j).$$

Cette dernière condition nous suggère de prendre pour f une fonction logarithme. En fait, la fonction logarithme est l'unique fonction vérifiant les propriétés énoncées ci-dessus. Notons que le choix de la base des logarithmes a peu d'importance ici car cela revient à multiplier la fonction $f(k) = Log(k)$ par un facteur constant ($log_b(k) = log_b(a) \times log_a(k)$) et est donc équivalent à un changement d'unité de mesure. Nous utiliserons partout dans la suite le logarithme néperien $ln(k)$.
Nous avons raisonné jusqu'à maitenant en considérant une expérience α dont les k issues $A_1, A_2, \ldots, A_k$ étaient équiprobables, et nous avons dit que le degré d'incertitude était égal à $ln(k)$. On peut donc considérer que chaque issue A_i apporte une incertidude égale à

$$\frac{ln(k)}{k} = -(\frac{1}{k}) \times ln(\frac{1}{k})$$

Ainsi dans le cas plus général, où l'expérience α admet k issues $A_1, A_2, \ldots, A_k$ ayant respectivement pour probabilité de se produire $p_1, p_2, \ldots, p_k$, il est naturel de poser pour mesure du degré d'incertitude de α :

$$H(\alpha) = -p_1 \times ln(p_1) - p_2 \times ln(p_2) \ldots - p_k \times ln(p_k)$$

2 Propriétés de l'entropie

Le lecteur trouvera ci-dessous quelques propriétés de l'entropie concernant une expérience α.

- **proposition 1 :**
  $H(\alpha) \geq 0$, l'entropie d'une expérience α est toujours positive ou nulle.
- **proposition 2 :**
  $H(\alpha) = 0 \Leftrightarrow \exists i_0$ tel que $P(A_{i_0}) = 1$ et $\forall i,\, i \neq i_0\ P(A_i) = 0$
- **proposition 3 :**
  $H(\alpha)$ maximum pour $P(A_1) = P(A_2) = \ldots = P(A_k) = \frac{1}{k}$
  Pour démontrer ce résultat nous allons utiliser le fait que $xln(x)$ est une fonction convexe et l'inégalité de Jensen.
  Inégalité de Jensen :
  Soit$f$ une fonction convexe sur un intervalle $[a, b]$
  $\forall x_1, x_2, \ldots, x_k \in [a, b]$ et $\forall \lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ tels que $\sum_{i=1}^{k} \lambda_i = 1$ on a :
  $\lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_k f(x_k) \geq f(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k)$
  Appliquons l'inégalité ci-dessus à la fonction $f(x) = xln(x)$
  $\forall P_1, P_2, \ldots, P_k \in [0, 1]$ on aura :
  $\sum_{i=1}^{k} \lambda_i \times P_i \times ln(P_i) \geq (\sum_{i=1}^{k} \lambda_i \times P_i) \times ln(\sum_{i=1}^{k} \lambda_i \times P_i)$
  ou encore
  $(-\sum_{i=1}^{k} \lambda_i \times P_i) \times ln(\sum_{i=1}^{k} \lambda_i \times P_i) \geq -\sum_{i=1}^{k} \lambda_i \times P_i \times ln(P_i)$
  Cette inégalité est vraie $\forall \lambda_1, \lambda_2, \ldots, \lambda_k$ avec $\sum_{i=1}^{k} \lambda_i = 1$ donc en particulier pour $\lambda_i = \frac{1}{k}\ \forall i$, l'inégalité ci-dessus devient
  $((-\frac{1}{k}) \times \sum_{i=1}^{k} P_i) \times ln((\frac{1}{k}) \times \sum_{i=1}^{k} P_i) \geq (-\frac{1}{k}) \times \sum_{i=1}^{k} P_i \times ln(P_i)$
  or $\sum_{i=1}^{k} P_i = 1$
  on obtient
  $(-\frac{1}{k}) \times ln(\frac{1}{k}) \geq (\frac{1}{k}) \times H(\alpha)$
  $-ln(\frac{1}{k}) \geq H(\alpha)$ ou encore $H(\alpha) \leq ln(k)$
- **proposition 4 :**
  Soit α et β deux expériences indépendantes on a $H(\alpha\beta) = H(\alpha) + H(\beta)$
- **proposition 5 :**
  Soit α et β deux expériences quelconques α ayant k issues $A_1, A_2, \ldots, A_k$ et β ayant l issues $B_1, B_2, \ldots, B_l$ alors
  $$H(\alpha\beta) = H(\alpha) + H_\alpha(\beta)$$

$H_\alpha(\beta)$, entropie de l'expérience β sous condition de l'expérience α est égale à la valeur moyenne des entropies conditionnelles $H_{A_i}(\beta)$
ie $H_\alpha(\beta)$=$\sum_{i=1}^{k} P(A_i) \times H_{A_i}(\beta)$ où $H_{A_i}(\beta)$ est l'entropie de l'expérience β sachant que α a eu l'issue A_i :

$$H_{A_i}(\beta) = \sum_{j=1}^{l} -P_{A_i}(B_j) \times ln(P_{A_i}(B_j)) \text{ avec } P_{A_i}(B_j) = \frac{P(A_i \cap B_j)}{P(A_i)}$$

- **proposition 6 :**
  $\forall \alpha, \beta$ on a $H_\alpha(\beta) \leq H(\beta)$
- **proposition 7 :**
  Si $A_k = \alpha_1\alpha_2 \ldots \alpha_k$ nous avons
  $H(A_k) = H(\alpha_1) + H_{\alpha_1}(\alpha_2) + \ldots + H_{\alpha_1\alpha_2,\ldots,\alpha_{k-1}}(\alpha_k)$
  $H(A_k) \leq H(\alpha_1) + H(\alpha_2) + \ldots + H(\alpha_k)$

3 Définition de la notion d'information

Une mesure ou une observation α quelconque précédant β peut restreindre la quantité des issues possibles de l'expérience β et en diminue par là même le degré d'incertitude. Ainsi le degré d'incertitude de l'expérience β qui consiste à trouver quel est le plus lourd de trois poids se trouve diminué après comparaison de deux d'entre eux. Afin que le résultat de la mesure (de l'observation) α puisse se faire sentir sur l'expérience β qui la suit, il est bien entendu nécessaire que ce résultat ne soit pas connu d'avance ; on peut par conséquent considérer α comme une expérience auxiliaire qui peut également admettre plusieurs issues. Le fait que la réalisation de α diminue le degré d'incertitude de β se reflète dans le fait que l'entropie conditionnelle $H_\alpha(\beta)$ de l'expérience β sous la condition de la réalisation de α s'avère inférieure ou égale à l'entropie initiale $H(\beta)$ de cette même expérience.
De plus si β est indépendante de α, la réalisation de α ne diminue pas l'entropie de β : $H_\alpha(\beta) = H(\beta)$; par contre si le résultat de α prédétermine l'issue de de β, l'entropie conditionnelle de β devient nulle $H_\alpha(\beta) = 0$
De la sorte la différence $I(\alpha, \beta) = H(\beta) - H_\alpha(\beta)$ indique dans quelle mesure la réalisation de l'expérience α diminue l'incertitude de β, c'est à dire ce que nous apprenons de nouveau sur l'issue de l'expérience β en effectuant la mesure (l'observation) α ; cette différence est appelée quantité d'information relative à l'expérience β contenue dans l'expérience α, ou plus brièvement, information relative à β contenue dans α. L'intérêt de cette notion provient du fait qu'elle permet d'évaluer l'influence, sur une expérience déterminée β, d'une autre expérience α, par la "différence d'entropie" :$I(\alpha, \beta) = H(\beta) - H_\alpha(\beta)$
On peut définir l'entropie $H(\beta)$ de l'expérience β comme l'information concernant β et contenue dans l'expérience β elle même (car $H_\beta(\beta) = 0$)

ou comme la plus grande information possible concernant β (l'information totale concernant β). Autrement dit l'entropie $H(\beta)$ de l'expérience β est égale à l'information que nous obtenons après avoir effectué cette expérience, c'est à dire à l'information moyenne contenue dans une issue de l'expérience β.

– **proposition 8 :**
  $I(\alpha,\beta) = I(\beta,\alpha)$
  En effet
  $H(\alpha\beta) = H(\beta\alpha)$
  $H(\alpha) + H_\alpha(\beta) = H(\beta) + H_\beta(\alpha)$
  $H(\alpha) - H_\beta(\alpha) = H(\beta) - H_\alpha(\beta)$
  $I(\beta,\alpha) = I(\alpha,\beta)$

– **proposition 9 :**
  Soit $A_k = \alpha_1\alpha_2\ldots\alpha_k$ une expérience composée directement orientée vers l'expérience β ($H_\beta(A_k) = 0$) alors
  A_k détermine entièrement β (ie $H_{A_k}(\beta) = 0$) si et seulement si $H(A_k) = H(\beta)$

4 Résolution de quelques problèmes de logique

4.1 Jeux de logique

Pour illustrer les notions et propositions ci-dessus, nous examinerons ici quelques problèmes de logique bien connus.
a) On sait que les habitants d'une ville A disent toujours la vérité, tandis que les habitants d'une ville voisine B mentent toujours. Un observateur X se trouve dans une des deux villes, mais ignore laquelle.
En interrogeant un passant (les habitants de A peuvent aller en B et vice versa) il doit déterminer :

– dans quelle ville il se trouve.
– dans quelle ville vit son interlocuteur.
– dans quelle ville il se trouve et dans quelle ville vit son son interlocuteur.

On demande quel est le nombre minimal de questions à poser, et quelles sont ces questions. Le passant ne répondant que par OUI ou par NON. Supposons que X ait à déterminer dans quelle ville il se trouve. L'expérience β, dont le résultat nous intéresse ici, peut avoir deux issues équiprobables, $H(\beta) = ln(2)$. Par ailleurs l'expérience α, consistant en une question posée par X à un passant, a également deux issues (OUI ou NON) $H(\alpha) \leq ln(2)$. Dans le problème on demande s'il est possible de poser une question α de telle sorte que l'information $I(\alpha,\beta)$ contenue

dans la question α sur l'expérience β soit égale à l'entropie $H(\beta) = ln(2)$ de l'expérience β, c'est à dire de sorte que l'issue de α détermine entièrement l'issue de β. On a vu (proposition 9) qu'il faut que $H(\alpha) = H(\beta)$.
Il suffit pour cela que la question α soit telle qu'une réponse affirmative et une réponse négative y soit également probable et que l'issue de l'expérience β détermine entièrement l'issue de α (c'est sous cette condition seulement qu'a lieu l'égalité $H_\beta(\alpha) = 0$, qui indique que la question α est "entièrement orientée" vers la détermination de l'issue de β). Dans ces conditions on aura :

$$I(\alpha, \beta) = H(\alpha) = H(\beta) = ln(2)$$

Ces conditions sont remplies par la question "Vivez-vous dans cette ville ?". Une réponse affirmative ne peut être donnée que dans la ville A et une réponse négative que dans la ville B. On voit encore plus simplement comment X peut à l'aide d'une seule question, établir dans quelle ville vit son interlocuteur. Il lui suffit pour cela de poser une question quelconque dont il connaît d'avance la réponse, par exemple "Est-ce que deux fois deux font quatre ?"
Si X doit déterminer à la fois dans quelle ville il se trouve et dans quelle ville vit son interlocuteur, une seule question ne peut suffire car l'expérience β qui nous intéresse peut avoir 4 issues équiprobables, $H(\beta) = ln(4)$. Il lui faudra poser les deux questions précédentes.
b) Soit trois villes A, B et C telles que les habitants de A disent toujours la vérité, les habitants de B mentent toujours, tandis que les habitants de C donnent alternativement une réponse vraie et une réponse fausse. Un observateur X veut déterminer dans quelle ville il se trouve et dans quelle ville vit son interlocuteur. Combien de questions doit-il poser à celui-ci si son interlocuteur ne répond que par OUI ou NON à toutes les questions ?
L'expérience β a ici 9 issues équiprobables, $H(\beta) = ln(9)$. Soit $A_k = \alpha_1, \alpha_2, \ldots, \alpha_k$ l'expérience composée de k questions posée par X.
L'entropie des questions $\alpha_1, \alpha_2, \ldots, \alpha_k$ est au plus égale à $ln(2)$. Or

$$H(A_k) = H(\alpha_1 \alpha_2 \ldots \alpha_k) \leq H(\alpha_1) + H(\alpha_2) + \ldots + H(\alpha_k) \leq k \times ln(2)$$
(proposition 7)

Si $k = 3$ l'information obtenue peut tout au plus être égale $3 \times ln(2) = ln(8) < ln(9)$. Ainsi trois questions ne peuvent donc suffire à élucider à la fois la situation de X et le lieu de résidence de son interlocuteur. Il est par contre possible que quatre questions judicieusement posées permettent de résoudre le problème ($4 \times ln(2) = ln(16) > ln(9)$).
Le lecteur vérifiera que les quatre questions suivantes permettent à l'observateur X d'élucider ce qu'il souhaite.

- 1) Suis-je dans une des villes A et B ?

– 2) Suis-je dans la ville C ?
– 3) Habitez-vous la ville C ?
– 4) Suis-je dans la ville A ?

Des réponses identiques aux questions 1) et 2) montrent que l'interlocuteur de X habite C ...

3) Combien faut-il de questions pour deviner un nombre entier positif au plus égal à n, choisi en pensée par un interlocuteur, si celui-ci ne répond que par OUI ou NON aux questions qui lui sont posées ?

La réponse bien connue est le plus petit entier$k$ tel que $k \geq \frac{ln(n)}{ln(2)} = log_2(n)$. La recherche dichotomique est un exemple parmi tant d'autres où les notions d'information et d'entropie s'appliquent.

4.2 Problème de détermination de monnaies fausses à l'aide de pesées

On dispose de 12 pièces de monnaies, 11 d'entre elles ont le même poids tandis que la dernière qui est fausse a un poids différent des autres. Quel est le nombre minimum de pesées sur une balance à plateaux qui permettra, sans utiliser de poids de découvrir la pièce fausse et de déterminer si elle plus lourde ou plus légère que les autres.

L'expérience β possède ici 24 issues possibles (chacune des 12 pièces peut être plus lourde ou plus légère). Ainsi $H(\beta) = ln(24) = 3.178$

Chaque pesée a trois issues possibles : **E**quilibre, penche à **G**auche, penche à **D**roite. Soit $A_k = \alpha_1\alpha_2 \ldots \alpha_k$ où α_i représente une pesée. Pour répondre au problème il est nécessaire que $H(A_k) \geq ln(24)$. Or $H(A_k) \leq k \times ln(3) \Rightarrow k \geq 3$ $(ln(3) = 1.098)$

A priori trois pesées peuvent suffire.

a) Choix de la première pesée α_1 :

On place i pièces sur chaque plateau $P(E) = \frac{(24-4i)}{24}; P(G) = \frac{2i}{24}; P(D) = \frac{2i}{24}$
$H(\alpha_1)$ sera maximum si et seulement si $P(E) = P(G) = P(D) \Rightarrow i = 4$ dans ce cas on aura $H(\alpha_1) = ln(3) = 1.098$.

b) choix de la deuxième pesée

b-1) Lors de la première pesée avec 4 pièces dans chaque plateau les plateaux restent en équilibre, donc il reste 4 pièces suspectes. Pour la seconde pesée, il est sans intérêt de placer des pièces notoirement vraies sur chaque plateau de la balance. La seconde pesée consiste à placer sur le plateau droit de la balance i de nos pièces suspectes et sur le plateau gauche $j \leq i$ de ces pièces et $i-j$ des pièces vraies. Pour chaque couple (i, j) on va calculer les trois probabilités P(E), P(G) et P(D) et retenir comme pesée celle où l'entropie $H(\alpha_2)$ est la plus grande.

i	j	P(E)	P(D)	P(G)	H(α_2)
1	1	1/2	1/4	1/4	1.039...
1	0	3/4	1/8	1/8	0.735...
2	2	0	1/2	1/2	0.693
2	1	1/4	3/8	3/8	1.082... *
2	0	1/2	1/4	1/4	1.039...
3	1	0	1/2	1/2	0.693...
3	0	1/4	3/8	3/8	1.082... *
4	0	0	1/2	1/2	0.693...

On voit sur ce tableau que deux pesées nous donnent la plus grande entropie. Choisissons la pesée qui consiste à placer 2 pièces suspectes sur le plateau droit, une pièce suspecte et une pièce vraie sur le plateau gauche. On voit facilement qu'on peut déterminer ensuite à l'aide d'une seule pesée l'issue de β. En effet si la pesée α_2 possède l'issue E, la pièce fausse sera l'unique pièce douteuse qui ne participe pas à la seconde pesée, pour savoir si elle plus lourde ou plus légère il suffira de la comparer à une pièce notoirement vraie à l'aide d'une troisième pesée.
Si la pesée α_2 possède l'issue D (l'issue G est symétrique), la pièce fausse est soit l'une des deux pièces de droite (elle est alors plus lourde que les autres) ; soit l'unique pièce douteuse du plateau gauche (elle est alors plus légère que les autres).
En comparant les deux pièces de droite à l'aide d'une troisième pesée nous connaîtrons l'issue de β. Si on obtient l'équilibre la pièce fausse est la troisième pièce douteuse qui est plus légère ; dans le cas contraire la pièce fausse est la plus lourde des 2 pièces comparées.
b-2) Lors de la première pesée le plateau de droite l'emporte (l'autre cas est symétrique).
Dans ce cas, la pièce fausse est l'une des quatre pièces de droite et elle est alors plus lourde que les autres, ou une des quatre pièces de gauche et elle est alors plus légère. Pour la seconde pesée, nous pouvons mettre sur le plateau droit : i_1 pièces de "droite" et i_2 pièces de "gauche" et sur le plateau gauche, j_1 pièces de "droite" et j_2 pièces de "gauche", et en plus $(i_1 + i_2) - (j_1 + j_2)$ pièces vraies si l'on suppose que $(i_1 + i_2) \geq (j_1 + j_2)$. On peut ici encore dresser un tableau analogue au précédent.
Comme le nombre de variantes possibles est assez élevé, il est plus astucieux d'exclure dès le début certaines d'entre elles. Remarquons que comme l'information que peut donner la troisième pesée sur β ne peut excéder $ln(3)$, il ne doit nous rester après deux pesées qu'au plus trois issues possibles de l'expérience β car sinon la troisième pesée α_3 ne pourra suffire à déterminer entièrement l'issue de β. Il en résulte que le nombre de pièces suspectes ne participant pas à la pesée ne doit pas

excéder 3, puisque dans le cas où la pesée α_2 admet l'issue E, ce sont ces pièces qui resteront suspectes et une troisième pesée ne suffira pas pour répondre. Nous aurons donc $i_1 + i_2 + j_1 + j_2 \geq 5$. Par ailleurs, si l'expérience α_2 admet l'issue D, alors la pièce fausse est soit l'une des i_1 pièces qui se trouve sur le plateau de droite et est plus lourde ; soit l'une des j_2 pièces placées sur le plateau de gauche et est plus légère. Le raisonnement est analogue dans le cas où α_2 admet l'issue G. Dans les deux cas pour conclure avec une seule pesée il est nécessaire d'avoir $i_1+j_2 \leq 3$ et $i_2+j_1 \leq 3$, en conclusion $i_1+i_2+j_1+j_2$ =5 ou 6. Il est clair d'autre part que doivent être également satisfaites les inégalités $i_1 + j_1 \leq 4, i_2 + j_2 \leq 4$ et $(i_1 + i_2) - j_1 + j_2) \leq 4$.

i_1	i_2	j_1	j_2	P(E)	P(D)	P(G)	H(α_2)
2	1	2	1	1/4	3/8	3/8	1.082...
2	1	2	0	3/8	3/8	1/4	1.082...
2	1	1	1	3/8	3/8	1/4	1.082...
1	2	1	2	1/4	3/8	3/8	1.082...
1	2	0	2	3/8	3/8	1/4	1.082...
2	2	1	1	3/8	1/4	3/8	1.082...
3	1	1	0	3/8	3/8	1/4	1.082...
1	3	0	1	3/8	1/4	3/8	1.082...
2	2	1	1	1/4	3/8	3/8	1.082...
2	2	1	0	3/8	1/4	3/8	1.082...
2	2	0	1	3/8	3/8	1/4	1.082...
3	2	1	0	1/4	3/8	3/8	1.082...
2	3	0	1	1/4	3/8	3/8	1.082...

Pour toutes les valeurs possibles de i_1, i_2, j_1, j_2 on a $H(\alpha_2) = 1.082\ldots$
Choisissons par exemple $i_1 = 2, i_2 = 1, j_1 = 2, j_2 = 1$.

- α_2 admet l'issue E
  La pièce fausse est une des deux pièces suspectées être plus légères ne participant pas à la pesée $\alpha 2$. Il suffit pour la déterminer, de comparer ces deux pièces entre elles ou de comparer l'une d'entre elles avec une pièce certainement vraie.
- α_2 admet l'issue D :
  La pièce fausse est soit l'une des deux pièces de droite, elle est alors plus lourde, soit l'unique pièce de gauche, elle est alors plus légère. Pour découvrir la pièce fausse, il nous suffit de comparer entre elles les deux pièces suspectes du plateau de droite.
- α_2 admet l'issue G :
  Avec un raisonnement analogue à celui du cas précédent, il nous suffit de

comparer entre elles les deux pièces suspectées être plus lourdes du plateau de gauche.

c) Calcul des entropies :
On sait que $H(\beta) = 3,0178\ldots$, d'autre part
$H(\alpha_1\alpha_2\alpha_3) = H(\alpha_1) + H_{\alpha_1}(\alpha_2) + H_{\alpha_1\alpha_2}(\alpha_3)$

- $H(\alpha_1) = ln(3) = 1.098\ldots$
- $H_{\alpha_1}(\alpha_2) = \sum_{i=1}^{3} P(A_i) \times H_{A_i}(\alpha_2) = (\frac{1}{3}) \times H_{A_i}(\alpha_2)$
  On a constaté que pour tout i
  $H_{A_i}(\alpha_2) = -(\frac{1}{4}) \times ln(\frac{1}{4}) - (\frac{3}{8}) \times ln(\frac{3}{8}) - (\frac{3}{8}) \times ln(\frac{3}{8}) = 1.082\ldots$
- Pour la troisième pesée deux cas sont à considérer :
  - α_2 a l'issue de probabilité $\frac{1}{4}$, soit B_1 cette issue.
    Dans ce cas on comparera deux pièces de poids différents, α_3 n'a que deux issues possibles $H_{B_1}(\alpha_3) = ln(2)$.
  - α_2 a l'une des deux issues de probabilité $\frac{3}{8}$, soit B_2 et B_3 ces deux issues.
    Dans chacun de ces deux cas α_3 a trois issues équiprobables.
    $H_{B_2}(\alpha_3) = ln(3) = 1.098\ldots$ et $H_{B_3}(\alpha_3) = ln(3) = 1.098\ldots$
    Ainsi $H_{\alpha_1\alpha_2}(\alpha_3) = (\frac{1}{4}) \times H_{B_1}(\alpha_3) + (\frac{3}{8}) \times H_{B_2}(\alpha_3) + (\frac{3}{8}) \times H_{B_3}(\alpha_3) = (\frac{1}{4}) \times ln(2) + (\frac{3}{4}) \times ln(3) = 0.997\ldots$

  Finalement $H(\alpha_1\alpha_2\alpha_3) = 1.098\ldots + 1.082\ldots + 0.997\ldots = H(\beta)$.

5 Le jeu du Master Mind

5.1 Règle du jeu

Le jeu se joue à deux joueurs A et B. Le joueur A dispose de pions de couleur et cache une combinaison de p pions. Soit n le nombre de couleurs différentes. Le joueur B doit chercher à deviner la combinaison cachée. Pour cela il propose des combinaisons de p pions au joueur A. Pour chaque combinaison proposée par le joueur B, A indique :

- Le nombre de pions biens placés
- Le nombre de pions mals placés

On raisonnera dans la suite de ce document avec $n = 6$ (six couleurs) et $p = 4$ (quatre positions). On associera aux six couleurs les entiers 1,2,3,4,5,6.
Avec ce codage si le joueur A a caché par exemple la combinaison 1 5 2 5 et que le joueur B propose la combinaison 3 5 1 2, La réponse de A sera 1 bien placé (le 5 en position 2) et 2 mals placés (le 1 et le 2).
Le joueur B doit s'efforcer de deviner la combinaison cachée en posant le moins de questions possibles ; chaque combinaison proposée est une question.
Notations : x et y signifieront respectivement "bien placés" et "mal placés", ainsi

la réponse $1x - 2y$ indiquera que la combinaison proposée comporte 1 pion bien placé et 2 mal placés par rapport à la combinaison cachée.

5.2 Construction d'une suite de questions optimales

L'expérience β dont le résultat nous intéresse ici a 6^4 issues possibles et il est naturel de considérer ces issues comme équiprobables de sorte que $H(\beta) = ln(6^4) = 4 \times ln(6)$. L'expérience α_1 consistant en une première question peut avoir 14 issues, 14 réponses : $0x-0y, 0x-1y, 0x-2y, 0x-3y, 0x-4y, 1x-0y, 1x-1y, 1x-2y, 1x-3y, 2x-0y, 2x-1y, 2x-2y, 3x-0y, 4x-0y$, par conséquent $H(\alpha_1) \leq ln(14)$ (proposition 3).
Considérons l'expérience composée $A_k = \alpha_1\alpha_2\ldots\alpha_k$ qui consiste en k questions posées. L'entropie d'une question α_i est au plus égale à $ln(14)$, ainsi $H_{A_k} \leq k \times ln(14)$ (proposition 7).
On veut que l'issue de notre expérience composée A_k détermine entièrement dans tous les cas l'issue β (i.e $H_{A_k}(\beta) = 0$). Pour que l'expérience composée A_k (formée de k questions α_i) permette de trouver sans ambiguïté l'issue de β on doit avoir $H(A_k) = H(\beta)$ et $H_\beta(A_k) = 0$ (proposition 9).
Comme on l'a vu $H(\beta) = 4 \times ln(6) = 7.167\ldots$ et $H(A_k) \leq k \times ln(14)$. Or pour que les k questions $\alpha_1\alpha_2\ldots\alpha_k$ permettent de trouver l'issue de β nous devons avoir $H(\beta) = H(A_k)$ c'est à dire $4 \times ln(6) \leq k \times ln(14)$ ou encore $k \geq \frac{4 \times ln(6)}{ln(14)} \approx 2.7$
Ce qui veut dire qu'il est impossible de trouver à coup sûr la combinaison cachée en moins de trois questions. En fait, il en faudra plus car l'entropie d'une question α_i (conditionnée par $\alpha_1 \ldots \alpha_{i-1}$ sera strictement inférieure à $ln(14)$).
Ainsi l'algorithme de construction de la suite optimale de questions commencera par déterminer quelle est la question auxiliaire α_1 qui contient la plus grande quantité d'information $I(\alpha_1, \beta)$ sur l'expérience β ; on choisit ensuite la seconde question α_2 (qui dépend de l'issue de α_1) de telle sorte que l'information $I(\alpha_1\alpha_2, \beta)$ soit la plus grande possible etc... ; cette façon de procéder nous permet de construire une suite de questions qui, en moyenne, sera optimale pour découvrir la combinaison cachée.

5.3 Algorithme de construction des questions

Nous avons le choix pour α_1 entre l'une des 1296 possibilités mais nous choisirons une des questions qui possède plus forte entropie (il y en a plusieurs qui rendent maximum l'entropie). Pour calculer l'entropie, il faut calculer les 14 probabilités associées à chacune des 14 réponses possibles. Pour cela on dénombrera chacun des 14 sous ensembles qui forment une partition des 1296 possibilités.
En fait pour choisir la première question il n'y a fondamentalement que 5 cas à

examiner (aux permutations des couleurs près et aux permutations des pions d'une question près).
Ces cinq cas sont :

- Question avec 4 pions de la même couleur
- Question avec 3 pions de la même couleur et un pion d'une autre couleur
- Question avec 2 pions d'une couleur et 2 pions d'une autre même couleur
- Question avec deux pions d'une même couleur et les deux autres pions d'une couleur différentes
- Question avec quatre pions de couleurs différentes

Le lecteur trouvera dans le tableau ci-dessous des exemples de questions recouvrant ces cinq cas. Pour chacune de ces questions nous avons fait figurer dans le tableau le nombres d'éléments des 14 sous-ensembles qui lui sont associés ainsi que son entropie.
Par exemple si nous posons la question "5521" il y a 276 quadruplets parmi les 1296 possibles où la réponse sera $0x1y$. Donc la probabilité qu'on nous réponde $0x1y$ à la question "5521" est égale à $\frac{276}{1296}$.

	$0x0y$	$0x1y$	$0x2y$	$0x3y$	$0x4y$	$1x0y$	$1x1y$	$1x2y$	$1x3y$	$2x0y$	$2x1y$	$2x2y$	$3x0y$	$4x0y$	entropie
"5555"	625	0	0	0	0	500	0	0	0	150	0	0	20	1	1.038
"5111"	256	308	61	0	0	317	156	27	0	123	24	3	20	1	1.866
"5511"	256	256	96	16	1	256	208	36	0	114	32	4	20	1	1.999
"5521"	81	276	222	44	2	182	230	84	4	105	40	5	20	1	2.109
"1234"	16	152	312	136	9	108	252	132	8	96	48	6	20	1	2.118

Ainsi nous constatons que la question α_1 qui possède la plus grande entropie doit être composée de 4 couleurs différentes. C'est ce type de question qui partitionne "le plus équitablement" l'ensemble des 1296 possibilités.

5.4 Quelques propriétés utiles à la compréhension du programme

5.4.1 Calcul du nombre de réponses possibles

Soit nbc et nbp les constantes qui indiquent respectivement le nombre de couleurs et le nombre de positions possibles. Calculons nbr le nombres de réponses possibles. Pour x (nombre de biens placés) donné ($0 \leq x \leq nbp$), y (le nombre de mals placés) peut prendre pour valeur $0, 1, 2, \ldots, nbp - x$. Sauf dans le cas où x vaut $nbp - 1$ qui implique que y ne peut prendre que la seule valeur 0. Ainsi, pour $x = nbp$ (tous biens placés) on a une seule valeur pour y ($y = 0$), pour $x = nbp - 1$ on a encore qu'une seule valeur pour y. Pour x, $0 \leq x \leq nbp - 1$, on a $nbp - x + 1$ réponses possibles ($y = 0, 1, \ldots, nbp - x$).
Le nombre nbr de réponses possibles est donc $2 + \sum_{x=0}^{nbp-2}(nbp - x + 1) = 2 + \sum_{i=3}^{nbp+1} i = \frac{(nbp+1)(nbp+2)}{2} - 1$.

5.4.2 Construction d'une injection de l'ensemble des couples (x, y) correspondants à une réponse possible dans l'ensemble $0, 1, \ldots, nbr$

Pour calculer l'entropie d'une question q, il va falloir connaitre combien d'éléments possède chacun des nbr membres de la partition des solutions possibles. Pour cela on devra comparer chacun des nbp-uplets candidats à la question q pour déterminer la réponse en calculant le couple (x, y) et incrémenter ainsi le compteur correspondant. Cette injection va nous permettre connaissant le couple (x, y) d'incrémenter directement le "bon" compteur en evitant de faire toute une serie de test en cascade…
Soulignons que l'on va construire une injection et non pas une bijection. En effet la valeur $nbr - 1$ qui correspondrait au couple $(nbr - 1, 1)$ qui n'est pas une réponse possible ne sera donc jamais atteinte.
Pour une valeur de x donnée, dénombrons le nombre de réponses possibles où le nombre de bien placés est strictement plus petit que x. Pour chacune de ces x valeurs possibles $(0, 1, \ldots, x - 1)$ y pourrait théoriquement prendre $nbp + 1$ valeurs $(0, 1, \ldots, nbp)$ mais seules les valeurs $0, 1, \ldots, nbr - x$ sont admissibles. Ainsi le nombre de réponses possibles où le nombre de bien placés est strictement inférieur à x est égal à $x \times (nbp + 1) - \frac{x \times (x-1)}{2}$. On choisira donc d'associer à une réponse possible (x, y) l'entier $ind \in 0, 1, \ldots, nbr$

$$ind \leftarrow x \times (nbp + 1) - \tfrac{x \times (x-1)}{2} + y$$

5.4.3 Détermination du nombre x de biens placés et du nombre y de mals placés

Soit q_1 et q_2 deux nbp-uplets.

// Calcul de x
$x \leftarrow 0$
pour i **variant de** 0 **jusqu'à** $nbp - 1$ **faire**
  si $q_1[i] = q_2[i]$ **alors** $x \leftarrow x + 1$

Pour calculer y on va pour chaque couleur c_i ($0 \leq i \leq nbc$) compter son nombre d'apparitions $c1_i$ dans le nb-uplet q_1 et $c2_i$ son nombre d'apparitions dans le nb-uplet q_2. On calcule ensuite la somme $s = \sum_{i=0}^{nbc-1} min(c1_i, c2_i)$. Il est facile de voir que le nombre de mals placés est égal à $s - x$.

// Calcul de y
$s \leftarrow 0$
pour j **variant de** 0 **jusqu'à** $nbc - 1$ **faire**
    $c1 \leftarrow 0$
    $c2 \leftarrow 0$
    pour i **variant de** 0 **jusqu'à** $nbp - 1$ **faire**
        si $q_1[i] = j$ **alors** $c1 \leftarrow c1 + 1$
        si $q_2[i] = j$ **alors** $c2 \leftarrow c2 + 1$
    $s \leftarrow s + min(c1, c2)$
$y \leftarrow s - x$

5.4.4 Enumération de tous les nbp-uplets possibles

Il y en a nbc^{nbp}. Pour énumérer tous les nbp-uplets possibles, il faut "simuler" nbp boucles imbriquées, contrôlées par les variables $i_0, i_1, \ldots, i_{nbp-1}$. Pour chacune des boucles la variable de contrôle variera de 0 à $nbc - 1$. On notera sol le tableau qui contient l'ensemble de tous les nbp-uplets possibles. On obtient ainsi dans la boucle la plus interne un nbp-uplet qu'il suffit alors de ranger dans une ligne du tableau sol.
Il n'est pas question d'écrire "physiquement" ces boucles imbriquées. En effet le nombre d'instructions du programme dépendrait du nombre nbp de positions possibles. On devrait reécrire un nouveau programme pour chaque nouvelle valeur de nbp !
La récursivité va nous permettre d'écrire un programme général qui sera capable d'engendrer tous les nbp-uplets possibles. En effet, on va écrire une procédure récursive qui va s'imbriquer, au niveau des appels, nbp fois. On simulera ainsi les nbp boucles imbriquées. La variable locale i, qui décrit à chaque imbrication les nbc couleurs possibles, aura autant "d'incarnations" qu'il y aura eu d'appels imbriqués. Plus précisément la variable i du premier appel représentera i_0, celle du second appel la variable i_1, etc ...
Par ailleurs cette procedure utilisera et agira sur trois variables globales, le tableau sol qui doit contenir au départ tous les nbp-uplets possibles, la variable entière d qui indique le nombre de nbp-uplets déjà placés, la variable t qui dénote le nbp-uplet en cours de construction.

procedure $creer - sol$ (**entier** j)
entier i
si $j = nbp$ **alors**
    // On a généré les nbp composantes d'un nbp-uplet
    // On les range dans la ligne d du tableau sol
    pour i **variant de** 0 **jusqu'à** $nbp - 1$ **faire**
        $sol[d][i] \leftarrow t[j]$
    $d \leftarrow d + 1$
sinon
    pour i **variant de** 0 **jusqu'à** $nbc - 1$ **faire**
        // On choisit la couleur i pour la composante j du nbp-uplet t
        $t[j] \leftarrow i$
        // Appel récursif de $creer - sol$ pour remplir la composante suivante
        $creer - sol(j + 1)$

5.4.5 Génération d'une permutation aléatoire de l'ensemble des nbp-uplets possibles

Pour que le programme ait des comportements différents à chaque exécution, il est nécessaire que les nbc^{nbp} nbp-uplets possibles soient rangés dans un ordre différent. Pour cela on va simuler $nbc^{nbp} - 1$ tirages sans remise afin d'obtenir une permutation aléatoire des lignes du tableau sol. Après avoir créé dans un premier temps tous les nbp-uplets possibles dans le tableau sol, on va dans un second temps générer une permutation aléatoire des lignes de ce même tableau.
L'idée est de tirer au sort un premier nbp-uplet. En fait on tire au sort un nombre entre 0 et $nbp-1$. Sortir ce nbp-uplet du tableau sol, revient à le placer en dernière position en effectuant un échange avec le dernier. Puis un tire au sort un nombre entre 0 et $nbp - 2$ (il ne reste plus que $nbp - 1$ nbp-uplets dans "l'urne" sol) et on effectue l'échange entre le nbp-uplet qui porte ce numéro et le nbp-uplet qui se trouve en avant dernière position, etc ...

pour i **variant de** $nbc^{nbp} - 1$ **jusqu'à** 1 **faire**

// On tire au sort un nombre k entre 0 et i)

$k \leftarrow rand()$ **mod** $(i + 1)$

// On échange les lignes k et i du tableau sol

pour j **variant de** 0 **jusqu'à** $nbp - 1$ **faire**

// aux est une variable de type nbp-uplet qui sert à réaliser l'échange des lignes i et k

$aux[j] \leftarrow sol[i][j]$

$sol[i][j] \leftarrow sol[k][j]$

$sol[k][j] \leftarrow sol[j]$

Références

[1] A.M Yaglom, I.M Yaglom, *Probabilité et Information*, Collection Sigma.