LECTURE DE TABLEAUX A QUATRE CASES ET TEST D'INDEPENDANCE DU χ^2

Moncef ZAKI

At the end of a statistical course in first year university, we have proposed to students a problem of decision using 2×2 tables. This problem has been presented as a game on micro-computer, in order to make evident the randomness of the situation.
We present here the two types of phenomena wich appared :

- developping decision strategies without using the χ^2 test,
- a systematic under-estimation of the variability of the frequences in the observed tables.

I. Problèmatique et présentation de la situation explorée.

I.1. *Introduction.*

Au terme d'un enseignement de probabilités-statistiques au niveau DEUG scientifique, nous avons mené une observation auprès de quelques étudiants[1] sur leurs critères de décision statistique en situation probabiliste. La situation que nous avons proposée aux étudiants met essentiellement en avant une notion statistique qui leur a été enseignée, à savoir le test d'indépendance du χ^2 appliqué aux tableaux à quatre cases. En fait, ce que nous avons voulu surtout explorer, ce sont des procédures de lecture de tels tableaux pour décider de la validité (ou non) de l'hypothèse d'indépendance stochastique des variables aléatoires qui y sont croisées.

En mettant les étudiants en situation extra-scolaire, la question était de savoir s'ils allaient faire appel à l'outil statistique qui leur a déjà été enseigné, en l'occurence le test du χ^2, ou s'ils n'allaient pas plutôt développer d'autres procédures de traitement pour construire leurs règles de décision. Bien entendu, la pertinence de telles procédures dépend essentiellement de celle de la lecture effectuée des tableaux.

Pour essayer de répondre à cette question, nous avons imaginé une situation de jeu, qui s'inspire d'une situation probabiliste de décision déjà explorée par J. Alarcon (1982) auprès de jeunes enfants de 12-14 ans. Alarcon avait exploré la situation type suivante :

Retrouver parmi deux sacs contenant des boules blanches et noires, celui dans lequel ont été faits un ou plusieurs tirages.

© Annales de Didactique et de Sciences Cognitives
4 (1991) (p.59 - 83) - IREM de Strasbourg

[1] *Il s'agit d'étudiants de l'Université Louis Pasteur à Strasbourg.*

Cette situation devient particulièrement intéressante lorsque la composition de l'un des deux sacs n'est pas donnée. Une telle situation exige un traitement particulier "*pour pondérer l'éventualité d'un tirage fait dans le sac ainsi caché, il faut alors tenir compte des différences entre le résultat du tirage et ce que l'on pouvait attendre si le tirage avait été dans le sac connu*"[1]. Or, les résultats montrent que seule une faible minorité d'élèves pratique systèmatiquement ce type de traitement lorsque le résultat des tirages est compatible avec la composition du sac connu. En effet, les élèves ont tendance à porter leur choix, soit sur le sac connu, soit sur le sac caché.

Nous ne pouvions reprendre telle quelle cette situation explorée par Alarcon avec de jeunes élèves. D'une part, le contenu probabiliste de cette situation de décision se trouve à un niveau assez élémentaire par rapport aux prérequis probabilistes et statistiques des étudiants. D'autre part, le caractère aléatoire des épreuves mises en jeu n'y était pas du tout apparent. C'est pourquoi nous avons retenu dans notre expérimentation une situation de jeu, dans laquelle les étudiants aient à l'aide de leurs critères de décision, à choisir parmi deux lois de probabilité génératrices de tableaux à quatre cases, et dans laquelle l'incrémentation des cases de ces tableaux est simulée aléatoirement par un ordinateur.

L'épreuve que nous avons envisagée aurait pu être réalisée sous la forme d'un travail papier-crayon, sans recours à l'ordinateur. En effet, on aurait pu imaginer le stockage d'un certain nombre de tableaux sur papier, et les présenter aux étudiants un par un. Dans ce cas, la crédibilité de l'expérience aux yeux de l'étudiant aurait pu être mise en doute. A la limite, qu'est-ce qui permet d'être sûr que l'expérimentateur ne construit pas ses verdicts (juste ou faux) au fur et à mesure des réponses données ? Et ceci, même si l'étudiant peut constater que les verdicts sont prêts d'avance. On peut, par exemple, faire la comparaison avec un énoncé où l'on dit "on a jeté 100 fois une pièce de monnaie et l'on a obtenu 65 fois "pile" et 35 fois "face"..". Les cas où de tels énoncés sont des constats d'expériences effectives sont rarissimes. Il était donc important pour nous que le caractère alétoire de l'épreuve soit apparent au cours de l'expérimentation, et que l'étudiant ne soupçone pas les données d'être faites pour créer un problème mathématique qui a "sa" réponse.

1 *Cf. J.Alarcon (1982), page 3.*

I.2. *Situations probabilistes explorées.*

Nous avons élaboré deux progammes informatiques qui simulent respectivement les lois de deux couples de variables aléotoires (X_1 , Y_1) et (X_2 , Y_2) à valeurs dans $\{0,1\}\times\{0,1\}$, tels que :

La loi du couple (X_1 , Y_1) est celle de l'équiprobabilité : elle est communiquée aux étudiants. Le programme correspondant est à rapprocher du sac connu de l'expérience d'Alarcon. Nous avons appelé une épreuve issue de cette loi "*épreuve connue*". Dans ce cas on a :

$$\forall (i,j) \in \{0,1\}\times\{0,1\} \quad , \quad P\{(X_1 , Y_1) = (i,j)\} = 0.5\times0.5 = 0.25 .$$

La loi du couple (X_2 , Y_2) est celle d'une autre loi : elle n'est pas communiquée aux étudiants. Le programme correspondant est à rapprocher du sac caché de l'expérience d'Alarcon. Nous avons appelé une épreuve issue de cette loi "*épreuve mystérieuse*". La loi du couple est construite de la façon suivante :

$$P\{(X_2 , Y_2) = (0,0)\} = P\{(X_2 , Y_2) = (1,1)\} = 0.6\times0.5 = 0.3 .$$
$$P\{(X_2 , Y_2) = (0,1)\} = P\{(X_2 , Y_2) = (1,0)\} = 0.4\times0.5 = 0.2 .$$

La distance distributionnelle du χ^2 entre les deux tableaux théoriques de 100 épreuves issus respectivement des lois connue et mystérieuse est significative, et vaut approximativement $\chi^2 \approx 3.84$.

Y \ X	0	1
0	25	25
1	25	25

Epreuve connue

Y \ X	0	1
0	30	20
1	20	30

Epreuve mystérieuse

Ainsi, au seuil habituel 5%, ces deux tableaux "s'écartent" juste assez l'un de l'autre au sens du χ^2, ce qui constitue un bon terrain d'évaluation de l'appréciation des étudiants pour les fluctuations éventuelles des valeurs observées autour des valeurs espérées.

I.3. *Présentation de la situation de jeu proposée aux étudiants.*

Le jeu est le suivant :

On choisit au hasard l'une des deux lois précédentes. Ensuite, on présente aux étudiants sur l'écran d'un ordinateur un tableau à double entrées (X et Y) qui s'incrémente de 100 épreuves. La simulation de l'évolution des quatre cases du tableau se fait suivant la loi choisie préalablement au hasard.

A l'issue des 100 épreuves, on demande aux étudiants de parier sur l'une ou l'autre des deux lois :

- connue, correspondant au couple (X_1 , Y_1),
- mystérieuse, correspondant au couple (X_2 , Y_2).

Au départ, l'étudiant dispose d'une fortune de 16, et doit parier sur l'issue d'un tableau, en misant l'une des sommes entières 1, 8, 15 ou 22, avec la contrainte de ne pas miser plus d'argent qu'il n'en possède. Les gains se répartissent selon le tableau suivant :

Mises	1	8	15	22
Gains	0	1	2	3

Notons au passage que le choix de cette répartition de mises et de gains a été fait dans le but d'amener les étudiants autant que possible à "bien" pondérer leur décision à chaque étape du jeu. Nous reviendrons d'ailleurs sur ce point avec plus de détails dans le prochain paragraphe.

Enfin, le jeu s'arrête dès que l'étudiant atteint une somme supérieure ou égale à 23 ou inférieure ou égale à 7.

Au terme du jeu, nous avons prévu une sortie sur imprimante des tableaux observés. Cela, pour permettre à chaque étudiant de faire des commentaires sur sa stratégie de décision au cours du jeu.

Par ailleurs, avant d'entamer le jeu, chaque étudiant a suivi une séance de démonstrations durant laquelle il pouvait simuler autant de tableaux de 100 épreuves qu'il le voulait, en connaissant préalablement à chaque fois la loi génératrice (connue ou mystérieuse) du tableau observé. Cette séance permettait donc d'observer les évolutions et les écarts propres à chaque loi génératrice.

I.4. *Quelques remarques sur les stratégies de décision relatives au jeu.*

Le choix de la loi de l'épreuve au cours du jeu se fait au hasard, à l'insu de l'étudiant et de l'expérimentateur. Cependant, les positions respectives de l'étudiant et de l'expérimentateur restent différentes, puisque le premier ne connait la loi du couple (X,Y) que pour l'épreuve connue alors que le second la connait pour les deux épreuves. Par conséquent, du point de vue de l'étudiant, l'outil statistique qui s'adapte au jeu sera le test du χ^2, alors que du point de vue de l'expérimentateur, la fonction de décision sera celle d'un test d'hypothèses simples.

Pour analyser les stratégies utilisées par les étudiants, il est donc nécessaire de tenir compte de cette différence. Ainsi, nous utiliserons et comparerons dans notre analyse deux outils statistiques, à savoir le test du χ^2 et le test (d'hypothèses simples) de Neyman-Pearson[1].

Par ailleurs, dans la situation de jeu que nous explorons, nous avons choisi de faire intervenir un système particulier de mises et de gains. Le but de ce choix ne se limite pas, comme nous l'avons déjà dit, à éviter que les étudiants ne procèdent à un choix au hasard lors de leurs décisions, mais aussi à voir s'il y a une relation entre les critères de décision des étudiants et leurs pondérations des choix à l'issue des tableaux. En d'autres termes, nous voulions faire intervenir dans la situation de décision présentée ici aux étudiants une pondération des risques de première et de seconde espèce : l'acceptation d'une hypothèse fausse et le refus d'une hypothèse juste.

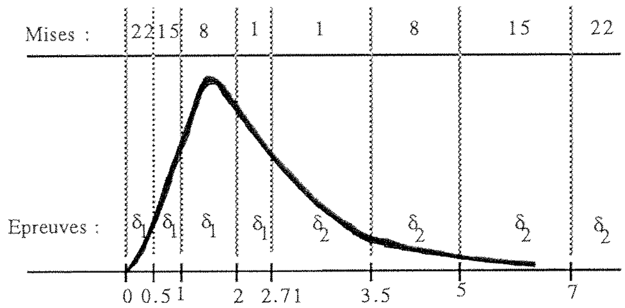
Or on peut estimer que, s'il faut tenir compte des mises pour la construction d'une fonction de décision dans la situation explorée, il faudrait en principe faire appel à la théorie de jeux ; mais ce n'est là ni le propos de notre étude, ni celui de notre investigation.

Maintenant, bien que le seul point de vue statistique ne permette pas de construire la fonction de décision optimale de ce jeu, on peut quand même arriver à construire une stratégie qui tient compte des mises. Par exemple, on peut penser, aussi bien pour le test du χ^2 que pour le test de Neyman-Pearson, à une fonction de décision δ qui prendrait ses deux valeurs δ_1 et δ_2 (δ_1 = connue, δ_2 = mystérieuse) selon une partition des deux régions de rejet et d'acceptation de l'hypothèse nulle (indépendance des variables X et Y) :

---

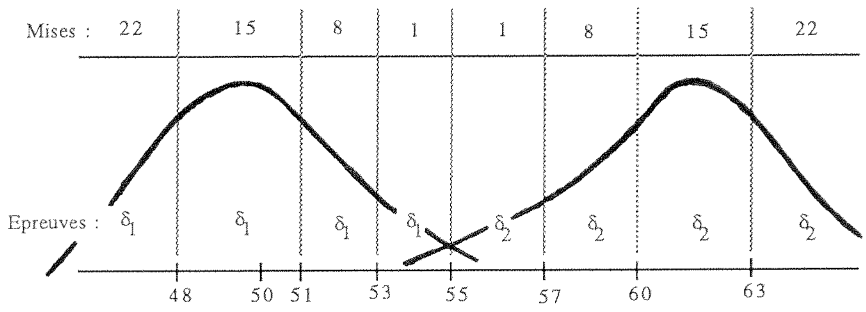
1 *Le détail de ces deux tests se trouve en annexe 1.*

- Du point de vue de l'étudiant (test de χ^2), nous proposons la stratégie suivante :



Un essai sur machine a permis, à l'aide de cette stratégie, de réussir à 4 jeux[1] au cours de 10 jeux consécutifs.

- Du point de vue de l'expérimentateur (test de Neyman-Pearson), nous proposons la stratégie suivante :



Un essai sur machine a permis, à l'aide de cette stratégie, de réussir à 5 jeux[1] au cours de 10 jeux consécutifs. Ceci peut “prouver” que malgré l'apparence de la répartition retenue pour les gains par rapport aux mises, ce jeu reste assez équitable.

II. Analyses et conclusions.

II.1. *Analyses des résultats de l'expérimentation.*

L'expérimentation a été conduite auprès de 10 étudiants en fin de première ou de deuxième année de DEUG scientifique, que nous désignons par les initiales : DEB, PFE, FUR, BUR RIC, SMI, STO, ROT, PAU et CER.

---

1 *Il est sous-entendu que l'on a atteint au moins la somme 23.*

Tous ces étudiants ont reçu le même enseignement de probabilités-statistiques :
En première année, on leur a dispensé un enseignement de probabilités (variables aléatoires, espérance mathématique, variance, lois usuelles) en vue d'une introduction à des notions statistiques telles que l'estimation, les test d'hypothèses, la corrélation et la régression. En fin de ce cycle, l'enseignement des statistiques a alors été essentiellement orienté vers l'analyse des données, où les étudiants sont amenés à appliquer à des cas concrets des modèles statistiques multivariés (modèle linéaire général et analyses factorielles).
Avant de procéder à l'analyse des stratégies utilisées par les étudiants, nous commencerons par faire une description générale des tableaux pour lesquels les étudiants ont fait un mauvais choix[1]. Par ailleurs, nous élaborerons deux codages : dans le premier, nous ferons appel aux tests de χ^2 et de Neyman-Pearson, et dans le second, aux mises introduites dans le jeu. Ces deux codages serviront respectivement à faire une classification des choix faits par les étudiants et une pondération de ces choix. Ensuite, à partir de ces deux codages, nous ferons une comparaison entre les stratégies déclarées par les étudiants et celles utilisées effectivement au cours du jeu, et nous analyserons les "profils" des étudiants, c'est-à-dire leurs critères de décision. Enfin, nous terminerons l'analyse par une étude sur la conformité des choix des étudiants avec l'information qui leur a été communiquée, à savoir la loi du couple (X,Y) pour l'épreuve connue.

II.1.1. *Description des tableaux d'échec*[1]*par rapport au jeu programmé.*
En passant en revue l'ensemble de tous ces tableaux, on peut faire deux constatations globales:
Sur 14 tableaux issus de l'épreuve mystérieuse, il y a 11 tableaux où l'on peut considérer que le choix de l'épreuve connue semble être raisonnable (voir les valeurs du χ^2 correspondantes), et 3 tableaux[2] où le choix de l'épreuve connue irait à l'encontre aussi bien de la distribution espérée des effectifs de l'épreuve connue que de la valeur statistique du χ^2, pour accepter cette épreuve.
Sur 17 tableaux issus de l'épreuve connue, il n'y a que trois tableaux[3] où l'étudiant, qui n'a pas remarqué la particularité de la distribution des effectifs de l'épreuve mystérieuse, peut parier presque sûrement sur l'épreuve mystérieuse : en fait, il faut aussi noter que deux de ces trois tableaux portent le plus grand poids d'effectif sur la deuxième diagonale par opposition à un tableau de type épreuve mystérieuse.

[1]*L'ensemble de ces tableaux se trouve en annexe 2.*
[2]*Voir ceux qui sont accompagnés du signe *.*
[3]*Voir ceux qui sont accompagnés du signe double *.*

Il se dégage de ces remarques une forte exigence de la part des étudiants en ce qui concerne la distribution des effectifs des tableaux qui sortent au cours du jeu, par rapport aux constats qu'ils ont pu faire au cours des démonstrations qui ont précédé le jeu. Des écarts entre des effectifs obtenus et les effectifs théoriques pour l'épreuve connue produisent vite une gêne pour la décision des étudiants. Ces écarts conduisent au rejet de l'épreuve connue et donc à un choix, que l'on peut qualifier de refuge, de l'épreuve mystérieuse.
Nous reviendrons d'ailleurs au cours de notre analyse sur ce point. En attendant, nous nous contenterons de faire remarquer que J.Alarcon a observé le même type de stratégie chez de jeunes élèves pour la situation de décision qu'il avait retenue (cf. § I.1.).

II.1.2. *Codages des choix et des mises des étudiants.*

Codage des choix :

Pour chacune des deux épreuves, connue et mystérieuse, nous avons établi un tableau d'analyse que nous avons partagé en quatre modalités par rapport au jeu programmé :

- **Réussite** :	on déclare un étudiant en situation de réussite lorsque son choix est correct.
- **Echec** :	on déclare un étudiant en échec, lorsqu'il rejette une issue correcte alors qu'elle est acceptée par le test du χ^2.
- **Non échec confirmé** :	on déclare un étudiant en non échec confirmé (N. E. C.), lorsqu'il rejette une issue correcte, rejetée par le test du χ^2 et le test de Neyman-Pearson.
- **Non échec non confirmé** :	on déclare un étudiant en non échec non confirmé (N. E. N. C.), lorsqu'il rejette une issue correcte, rejetée par le test du χ^2 mais acceptée par le test de Neyman-Pearson.

Codage des mises par rapport à une perte éventuelle :

Afin de pondérer le choix des étudiants, nous avons adopté un codage concernant leurs mises par rapport à une perte éventuelle.

Pour cela nous avons partagé l'avoir du joueur (étudiant), que nous désignons par A, en quatre catégories :

$A \leq 7$	Partie perdue
$8 \leq A \leq 14$	Mise maximum 8
$15 \leq A \leq 21$	Mise maximum 15
$A \geq 22$	Mise maximum 22

Ainsi, nous avons obtenu le codage suivant :

0	Non changement de catégorie
1	Diminution de catégorie, mais non perte de la partie.
2	Perte de la partie

Dans l'élaboration des tableaux d'analyse, nous utiliserons ce codage, en faisant la distinction entre la réussite et les autres modalités :

- Pour la réussite, on retiendra le mode des mises ainsi que le nombre de mises gagnantes inférieures aux mises maximales.
- Pour les trois autres modalités, on retiendra également le mode des mises avec, cette fois-ci, le nombre de mises "perdantes" supérieures aux mises minimales.

II.1.3. *Tableaux d'analyse et représentations des profils des étudiants.*

EPREUVE CONNUE

Ind \ Mod	Réussite	Echec	N.E.C.	N.E.N.C.	Mode des mises gagnantes	Mode des mises perdantes	Mode des mises N.E.C.	Mode des mises N.E.N.C.	Mises gagn. < mises max.	Mises perd. > mises min.	Mises N.E.C. > mises min.	Mises N.E.N.C. > mises min.
DEB	8	1	0	1	2),2,2,2,2,2,2,2,2	0 (0)	0 (0)		1	0	0	
PFE	5	2	0	0	2 (2,2,2,2)	1 (1,1)			1	2		
FUR	1	1	0	0	0 (0)	2 (2)			1	1		
BUR	1	3	0	0	1 (1)	(0,1,2)			1	2		
CER	1	0	0	0	1 (1)				1			
ROT	4	1	0	1	2 (2,2,0,2)	1 (1)		0 (0)	1	1		0
PAU	1	2	0	0	0 (0)	(0,2)			1	1		
SMI	2	3	0	0	(1,2)	(0,1,2)			1	2		
RIC	1	2	1	0	2 (2)	0 (0,0)		1 (1)	0			1
STO	2	2	0	0	0 (0,0)				2	1		

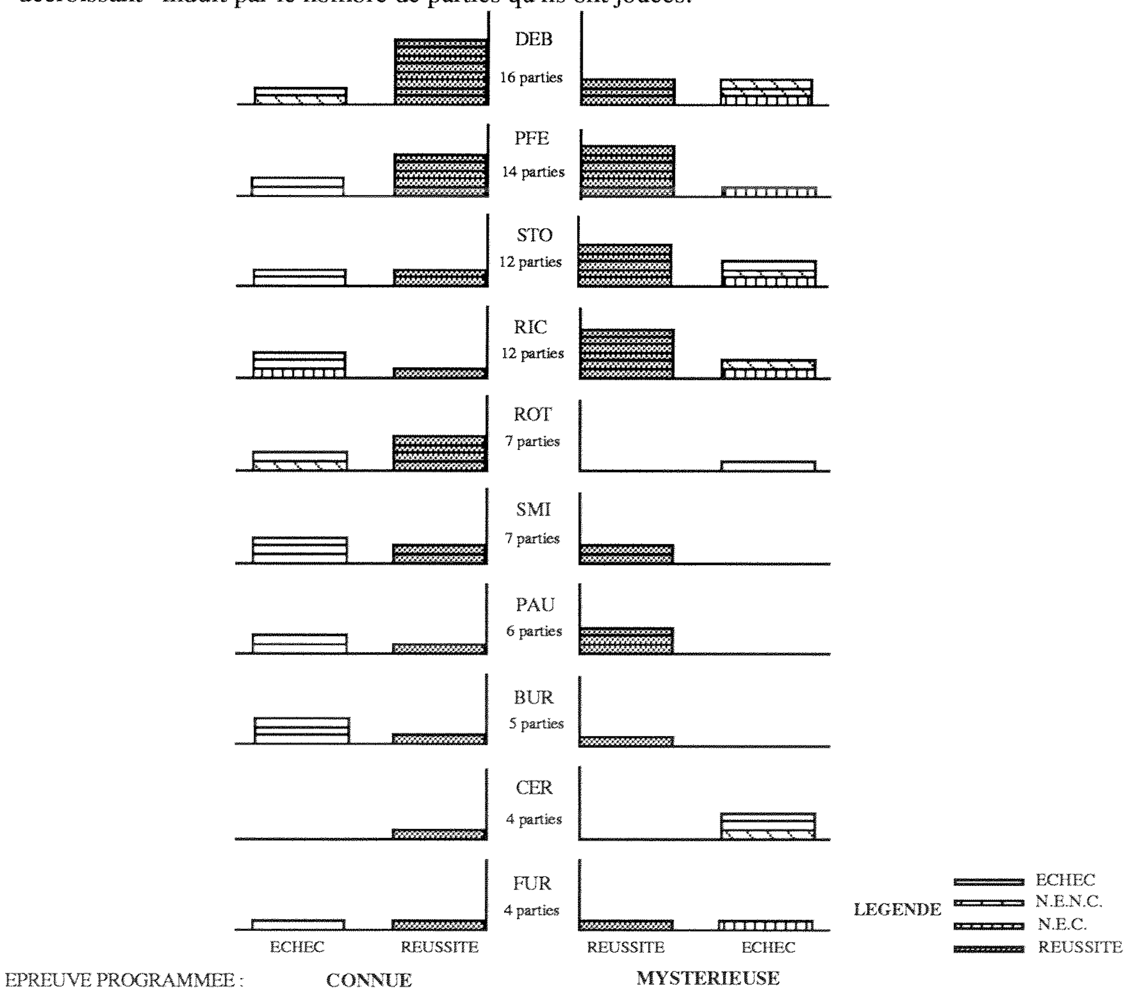
EPREUVE MYSTERIEUSE

Ind \ Mod	Réussite	Echec	N.E.C.	N.E.N.C.	Mode des mises gagnantes	Mode des mises perdantes	Mode des mises N.E.C.	Mode des mises N.E.N.C.	Mises gagn. < mises max.	Mises perd. > mises min.	Mises N.E.C. > mises min.	Mises N.E.N.C. > mises min.
DEB	3	0	1	2	2 (2,2,2)		1 (1)	(0,2)	0		1	1
PFE	6	0	1	0	2 (2,2,2,2,2,2)		2 (2)		1		1	
FUR	1	0	1	0	1 (1)		0 (0)		1		0	
BUR	1	0	0	0	1 (1)				1			
CER	0	2	0	1		(0,2)		1 (1)		1		1
ROT	0	1	0	0		2 (2)				1		
PAU	3	0	0	0	2 (2,2,2)				0			
SMI	2	0	0	0	(1,2)				1			
RIC	6	0	1	1	2 (1,2,2,2,2,2)		2 (2)	0 (0)	1		1	0
STO	5	1	1	1	0 (0,1,0,0,0)	0 (0)	0 (0)	0 (0)	5	0	0	0

Pour dégager le profil de chaque individu (étudiant), nous avons utilisé deux représentations en briques (l'une se déduisant de l'autre) : la première se réfère à la réussite par rapport au jeu programmé et la seconde à la réussite par rapport au choix de l'étudiant.

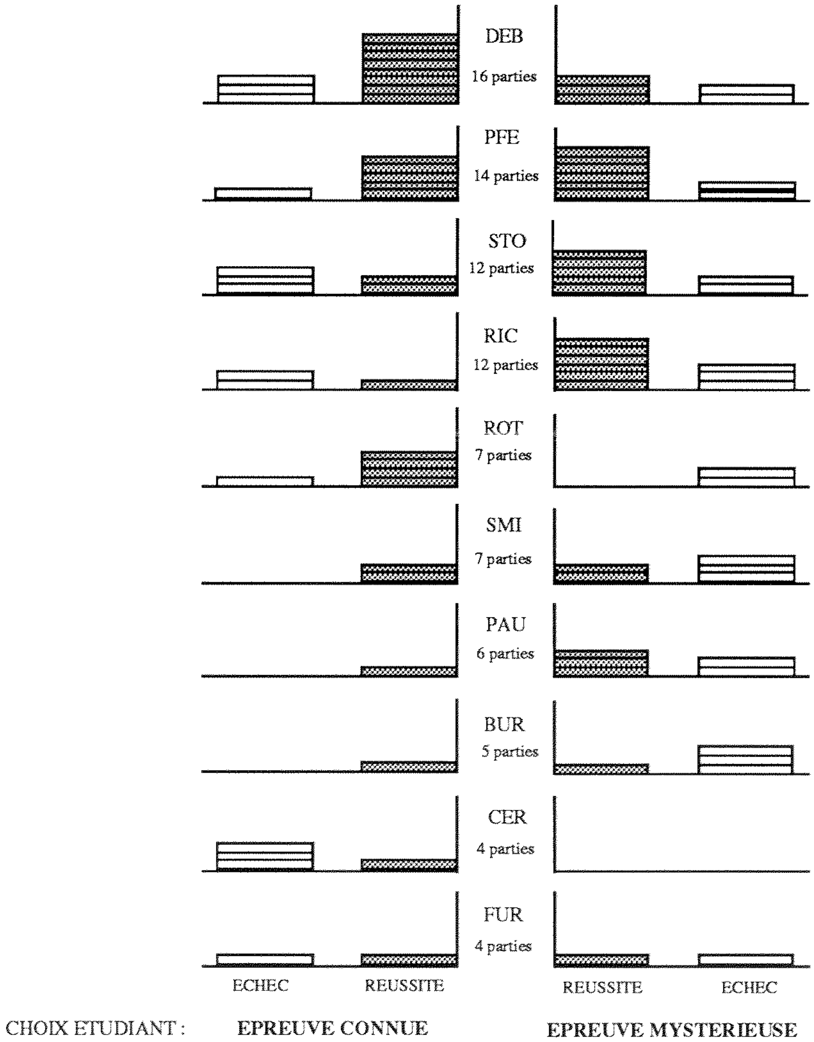
Première représentation :

Nous y avons distingué les quatre modalités, réussite, échec, N. E. C. et N. E. N. C., par rapport au jeu programmé. Chaque brique représente une issue de tableau au cours du jeu pour un étudiant donné. Les individus ont été classés dans cette représentation dans l'ordre "décroissant" induit par le nombre de parties qu'ils ont jouées.



Deuxième représentation :

Nous y avons distingué les deux modalités, réussite et échec, par rapport au choix de l'étudiant.



II.1.4. *Comparaison entre les stratégies déclarées et les résultats de tableaux d'analyse.*

Comme nous l'avions déjà dit précédement, à la fin de chaque jeu, une récapitulation des tableaux issus au cours du jeu a été prévue. Le but en était que chaque étudiant devait commenter sa stratégie concernant son choix d'épreuve (connue ou mystérieuse) à l'issue d'un tableau, ainsi que la mise qu'il avait parié sur ce choix. La comparaison des commentaires faits

par les étudiants aux résultats de tableaux d'analyse va nous permettre d'étudier la comformité de leurs stratégies déclarées.
De ces entretiens, nous avons relevé en gros deux types de populations :

première population : Elle se base uniquement sur les lois marginales. Autrement dit, cette population ne se réfère qu'à la distribution des effectifs du tableau final : elle parie sur la loi connue si les quatre valeurs des deux marges du tableau se rapprochent de la valeur 50, sinon, si l'une au moins des quatre valeurs marginales s'écarte de 5 de la valeur 50, elle parie sur la loi mystérieuse.

deuxième population : Elle se base sur le déroulement des quatre valeurs "intérieures" du tableau. L'évolution des quatre cases intérieures du tableau lui permet de se faire "a priori" une idée sur la nature de la loi du couple. Une fois la partie terminée, cette population compare le minimum et le maximum des quatre valeurs intérieures du tableau. Si leur différence est inférieure à 10, elle parie sur la loi connue, sinon sur la loi mystérieuse. Dans le cas où le traitement précédent ne confirme pas l'hypothèse "a priori", cette population se livre à une décision au hasard.

1) La première population est constituée de BUR, CER et FUR. Voici d'ailleurs une partie de leurs commentaires :
BUR : " Je ne sais pas sur quoi me baser, le déroulememt des chiffres ne me dit rien pour faire mon choix. Ma stratégie n'est pas toujours bonne".
CER : "Je ne fais pas attention aux mises ... le fait de regarder les valeurs des 4 cases revient en même que de regarder les 4 valeurs sur les côtés".
FUR : "Je ne comprends pas pourquoi ça ne marche pas, je crois que c'est un jeu de hasard. Je ne sais pas comment m'orienter".
Si l'on se réfère maintenant aux résultats de jeux de l'ensemble de ces trois étudiants, on constate que, par rapport au jeu programmé, on a :
- 43% de réussite, 57% d'échec, 0% de N. E. C. et 0% de N. E. N. C. pour les tableaux issus

de la loi connue. (Ces pourcentages sont relatifs à un ensemble de 7 tableaux, cf. Tableaux d'analyse.)

- 33.33% de réussite, 33.33% d'échec, 16.66% de N. E. C. et 16.66% de N. E. N. C. pour les tableaux issus de la loi mytérieuse. (Ces pourcentages sont relatifs à un ensemble de 6 tableaux, cf. Tableaux d'analyse.).

Comme le N. E. C. (resp. le N. E. N. C.) se rapproche d'autant mieux de la réussite (resp. de l'échec), en regroupant respectivement deux à deux chacune des deux modalités, on voit que sur l'ensemble il y a plus d'"échec" que de "réussite", ce qui tend à confirmer plus ou moins la stratégie déclarée par ces étudiants.

Néanmoins, en se référant aux résultats obtenus par CER (cf. Tableaux d'analyse), on peut avancer que la stratégie qu'il a déclarée ne concorde pas tout à fait avec les choix qu'il a effectués au cours du jeu. En effet, pour les deux épreuves, connue et mystérieuse, on peut relever les deux points suivants :

- le mode des mises gagnantes est 1, c'est-à-dire que les mises dans ce cas là sont toujours inférieures aux mises maximales ;
- les modes des mises perdantes, N. E. C. et N. E. N. C., prennent comme valeurs 0 ou 2, c'est-à-dire que les mises dans ce cas là correspondent, soit à un non changement de catégorie (des mises), soit à la perte de la partie de jeu.

Bien que cela ne soit pas aussi nette chez les deux autres étudiants, la répartition des mises de cette population, permet de conclure que la mise ne constitue pas pour elle un élément de stratégie, puisqu'il n'y a pas de concordance entre les mises et la pondération des choix faits par ces étudiants.

2) <u>La deuxième population</u> est constituée de DEB, PFE, STO, RIC, ROT, SMI et PAU. Voici quelques uns de leurs commentaires :

PFE : "J'ai remarqué que (0,0) et (1,1) sortaient plus souvent dans l'épreuve mystérieuse que dans l'épreuve connue. Au début je me base sur le déroulement et à la fin je regarde le tableau final pour prendre une décision".

PAU : "Je regardais le compteur : (0,1) et (1,0) sortaient moins fréquement que les autres dans l'épreuve mystérieuse. Dans ce jeu, il s'agit soit de gagner soit de perdre vite".

Pour cette pouplation, les pourcentages des quatre modalités par rapport au jeu programmé sont les suivants :

- 60% de réussite, 34% d'échec, 2% de N. E. C. et 4% de N. E. N. C. pour les tableaux issus de la loi connue. (Ces pourcentages sont relatifs à un ensemble de 39 tableaux, cf. Tableaux d'analyse.)
- 71% de réussite, 7% d'échec, 11% de N. E. C. et 11% de N. E. N. C. pour les tableaux issus de la loi mystérieuse. (Ces pourcentages sont relatifs à un ensemble de 35 tableaux, cf. Tableaux d'analyse.).

Dans l'épreuve connue, et particulièrement dans l'épreuve mystérieuse, la réussite est beaucoup plus importante que l'échec, le N. E. C. et le N. E. N. C.

Ceci tend, de manière encore plus nette que dans le cas de la première population, à confirmer la stratégie déclarée par les étudiants de cette population.

Dans les deux épreuves, connue et mystérieuse, on peut constater que :

- le mode des mises gagnantes est 2, et dans la majorité des cas les mises sont égales aux mises maximales ;
- les modes des mises perdantes, N. E. C. et N. E. N. C., prennent la valeur 0. Cette fois-ci, dans la plupart des cas les mises "perdantes" sont égales aux mises minimales.

Il se dégage de cette répartition des mises une bonne concordance entre les mises et les pondérations des choix, et ce, aussi bien pour les choix positifs que pour les choix incertains.

Jusqu'à maintenant, nous nous sommes contentés de faire une comparaison entre les stratégies déclarées et les résultats de tableaux d'analyse. Une deuxième étape de notre analyse serait d'étudier les profils des étudiants pour mieux comprendre leurs critères de décision. Pour cela, nous ne tiendrons pas compte des mises, mais nous nous intéresserons uniquement aux modalités de réussite et d'"échecs", pris sous deux points de vue :

- Réussite et échec par rapport au choix de l'étudiant ;
- Réussite, échec, N. E. C. et N. E. N. C., par rapport au jeu programmé.

II.1.5. *Analyse des profils des étudiants.*

Les deux représentations en briques suggèrent une partition de l'ensemble des étudiants en trois sous-populations :

SMI, PAU, BUR :

Dans la représentation "Choix étudiant", ces trois étudiants ne présentent aucun échec à l'épreuve connue. Les échecs sont plutôt du côté de l'épreuve mystérieuse. Ceci nous a donc incités à regarder les tableaux de jeu de chacun de ces étudiants, pour voir quelle était la nature des tableaux pour lesquels ils avaient choisi l'épreuve connue.

Nous avons pu constater que tous ces tableaux étaient très proches du tableau espéré de l'épreuve connue (équidistribution sur les quatre cases). Ceci traduit donc une forte **exigence** sur la distribution des effectifs des tableaux de la part de ces étudiants pour un choix positif concernant l'épreuve connue.
Maintenant, si l'on se réfère à la représentation "Jeu programmé", on constate que les échecs à l'épreuve connue sont de "vrais" échecs, confirmés aussi bien par le test du χ^2 que par le test de Neyman-Pearson : des **fluctuations** autour des valeurs espérées de l'épreuve connue traduisent chez ces étudiants une insatisfaction qui les mène à considérer l'épreuve mystérieuse comme un **refuge**, plutôt qu'un élément de stratégie.

DEB, PFE, STO, RIC, ROT :
Pour cette population, la situation est beaucoup moins nette que ce qu'elle ne l'est pour la précédente. En effet, dans la représentation "Choix de l'étudiant", il y a 11 échecs contre 20 réussites à l'épreuve mystérieuse. Cependant, il faut se rappeler que cette population fait partie de celle qui avait déclaré avoir constaté une plus grande fréquence d'apparition des couples (0,0) et (1,1) dans l'épreuve mystérieuse que dans l'épreuve connue. En outre, à la base de cette constation, elle se construisait une hypothèse "a priori" sur la nature de l'épreuve en cours, pour prendre sa décision. Les résultats d'expériences confirment tout à fait cette stratégie, puisqu'à l'issue d'un tableau, si la distribution se "rapproche" de celle que ces étudiants se sont faits "a priori" pour l'épreuve connue ou mystérieuse, leur choix a toujours été positif. (D'ailleurs, les mises dans ce cas là sont très fortes, à part pour STO qui avait déclaré jouer la prudence.) Sinon, si le hasard conduit à des fluctuations qui produisent des "écarts" par rapport à leur hypothèse "a priori", ils décident l'épreuve mystérieuse, en minimisant le risque d'échec, puisque dans ce cas là leurs mises sont faibles.
Par ailleurs, dans la représentation "Jeu programmé", il y a beaucoup plus de vrais échecs à l'épreuve connue qu'à l'épreuve mystérieuse, à laquelle correspondent surtout des N. E. C. et des N. E. N. C. ; ce qui confirme encore une fois la stratégie utilisée par cette population.

CER, FUR :
En ce qui concerne l'étudiant FUR, nous ne pouvons conclure à un élément de stratégie. Au vu des (quatre) tableaux de son jeu et des choix associés à leur issue (cf. représentations en briques), on ne peut avancer aucune interprétation.

En revanche, pour l'étudiant CER, nous pouvons prétendre qu'il nourrit un **refus** catégorique en faveur de l'épreuve mystérieuse. Pour les quatre tableaux de son jeu, il a toujours parié sur l'épreuve connue, en faisant trois échecs sur ses choix. D'ailleurs, sur la représentation "Jeu programmé", ces choix se traduisent par deux vrais échecs et un N. E. N. C. (qui, rappelons-le, est très proche de l'échec).
On peut croire que le fait de ne rien savoir sur l'épreuve mystérieuse a poussé cet étudiant à ignorer cette épreuve, malgré la très grande proximité de la distribution des tableaux présentés au cours du jeu du tableau théorique de l'épreuve mystérieuse (cf. Annexe 2).

II.1.6. *Types de tableaux observés et pondérations des choix.*
Pour vérifier si les étudiants présentaient lors de leurs choix une conformité avec l'épreuve connue (épreuve pour laquelle ils connaissent a priori la distribution du tableau théorique), nous procéderons à une analyse descriptive, d'une part de l'échec et de la réussite des étudiants à l'épreuve mystérieuse par rapport aux choix des étudiants, d'autre part de la pondération de leurs choix (qui se se traduit ici par l'importance de leurs mises), en référence au type tableau présenté au cours du jeu.
Les tableaux observés conduisent à trois types de tableaux[1] :

Premier type de tableaux :

Y \ X	0	1
0	+	—
1	—	+

Mises	0	1	2
R / choix étudiants	1	4	14
E / choix étudiants	0	1	0
Total	1	5	14

L'interprétation des données ci-dessus rejoint celle de l'analyse des profils des étudiants. Effectivement, il apparait clairement pour ce type de tableau que le choix des étudiants est pratiquement toujours positif, et ceci est d'ailleurs confirmé par le poids important des mises avec un risque de perte de la partie (il y en a 14 contre 6).

[1]*Dans ce qui suit, les signes "+" et "-" désignent respectivement une surcharge et une décharge en* effectif.

Deuxième type de tableaux :

Y \ X	0	1
0	+	+
1	—	+

Y \ X	0	1
0	+	—
1	+	+

Y \ X	0	1
0	+	—
1	—	—

Y \ X	0	1
0	—	—
1	—	+

Mises	0	1	2
R / choix étudiants	4	0	4
E / choix étudiants	0	1	2
Total	4	1	6

Les tableaux présentés ici, malgré leur écart (non statistique) dû à une surchage ou au contraire à une décharge en effectif par rapport au tableau espéré de l'épreuve mystérieuse, sont les moins génants pour les étudiants lors de leur décision de choix d'épreuve.

Ceci se traduit par le fait, d'une part qu'il y a 8 réussites contre 3 échecs, d'autre part qu'il y a 6 mises avec un risque de perte de la partie et une seule mise avec risque de changement de catégorie d'avoir. Ceci signifie que pour ce type de tableau la décision se trouve à un stade intermédiaire entre un choix positif et un choix où l'on réduit le risque de perte en cas d'échec.

Troisième type de tableaux : Autres tableaux.

Mises	0	1	2
R / choix étudiants	1	0	0
E / choix étudiants	8	4	3
Total	9	4	3

Ici, sont regroupés les tableaux les plus génants pour le choix de l'épreuve mystérieuse. La sous-estimation des étudiants vis à vis des fluctuations dues au générateur (pseudo) aléatoire, et leur exigence envers la conformité de la distribution des effectifs des tableaux qui sortent en cours de jeu par rapport à la distibution de l'épreuve connue, les conduisent à l'issue de ce type de tableaux à rejeter l'épreuve connue et à choisir l'épreuve mystérieuse en pariant le plus souvent la mise minimale. En effet, mise à part une réussite dont la mise correspondante est d'ailleurs codée par 0, il y a 16 échecs dont 9 étaient sans risque de changement de catégorie d'avoir et 4 avec changement de catégorie d'avoir mais sans risque de perte de la partie. Par conséquent, comme c'était le cas pour l'analyse des profils, le choix de l'épreuve mystérieuse pour ce type de tableaux (après rejet de l'épreuve connue) se fait encore une fois avec une minimisation du risque de perte en cas d'échec.

II.2. *Conclusions.*

Au terme de cette analyse descriptive, nous voyons se dégager de manière très importante la sous-estimation des étudiants vis à vis des fluctuations, inhérentes au hasard, des effectifs des tableaux observés par rapport à l'équidistribution de l'épreuve connue, dont ils étaient informés. Et ceci malgré les séances de démonstrations, précédant les parties de jeux, durant lesquelles ils étaient informés au préalable sur la loi génératrice (connue ou mystérieuse) de chaque tableau observé.

En effet, mis à part les cas où les tableaux tendaient vers une grande conformité aux épreuves, connue ou mystérieuse, et où la décision du choix était correcte, on remarque que l'épreuve mystérieuse, qui possède un caractère "inconnu" pour l'étudiant, semble jouer le rôle d'un refuge d'insatisfaction, sans prendre pour autant le caractère d'un élément de stratégie au cours du jeu. Nous retrouvons là pratiquement le même type de comportement qu'Alarcon avait rencontré durant son expérimentation auprès de jeunes élèves. En ce sens, nous pouvons considérer notre travail comme un prolongement de celui de J.Alarcon.

Pour un complément d'observation, il serait intéressant de refaire l'expérience en présentant cette fois-ci la distribution de l'épreuve mystérieuse et en rendant "inconnue" celle de l'épreuve connue. Mais nous pensons que ce n'est pas le critère d'équidistribution de l'épreuve connue qui rend important ce phénomène de sous-estimation des fluctuations autour des valeurs espérées, c'est plutôt l'exigence, de la part des étudiants, de "petits écarts" entre les tableaux observés et la distribution du tableau sur laquelle ils sont informés.

Pour un type d'observation de nature probabiliste, on pourrait aussi penser à utiliser une variante expérimentale de la présente situation qui consisterait à rendre les deux épreuves, connue et mystérieuse, "connues". A ce moment là, la décision du choix ne serait plus de nature statistique, mais de nature probabiliste puisqu'il s'agirait de discriminer entre deux lois par une quantification des probabilités. Cela pourrait avoir son intérêt, à partir du moment où il s'agirait de voir comment les étudiants vont traiter les données issues de tableaux observés pour traduire cette quantification.

La notion de risque que nous avons voulu faire intervenir pour notre observation ne semblait pas prendre un caractère important pour les étudiants comme nous l'avions cru au départ. Il serait peut-être plus adéquat de remplacer le système de mises que nous avons retenu par un dispositif plus sensibilisant, par exemple, un segment gradué où une particule avancerait (resp. reculerait) de p unités (resp. q unités) selon que le choix est positif ou négatif, avant d'être absorbée par les extrémités du segment . Un tel dispositif aurait au moins l'avantage d'illustrer la pondération du choix au cours du jeu.

Par ailleurs, il faut aussi faire remarquer qu'aucun étudiant n'a eu recours à l'outil statistique du χ^2 pour élaborer sa stratégie de décision. Bien que la situation explorée s'adapte parfaitement à cet outil statistique, on voit que le fait que les étudiants se trouvent en situation extra-scolaire les a tous conduits à "oublier" l'outil qui leur a été enseigné, et à ne faire qu'une lecture empirique des tableaux observés pour prendre leurs décisions. Nous pensons qu'une variabilité au niveau des situations présentées aux étudiants s'impose, si l'on veut que l'enseignement des statistiques atteigne complètement son objectif.

ANNEXE 1

Pour présenter ces deux tests, nous allons nous référer au tableau général suivant :

Y \ X	0	1	
0	N_{11}	N_{12}	$N_{1.}$
1	N_{21}	N_{22}	$N_{2.}$
	$N_{.1}$	$N_{.2}$	N

Le test du χ^2 :

Il utilise la statistique $T = \dfrac{N_{11}N_{22} - N_{12}N_{21}}{N_{1.}N_{2.}N_{.1}N_{.2}}$.

La région de rejet est de la forme $D = \{T > c\}$. Au seuil 10%, $P_{\chi 2}\{T > c\} = 0.1$ permet de calculer pour $N = 100$ la région de rejet :

$$D = \{100\,\frac{N_{11}N_{22} - N_{12}N_{21}}{N_{1.}N_{2.}N_{.1}N_{.2}} > 2.71\}.$$

Le test de Neyman-Pearson :

Pour utiliser ce test, nous allons tout d'abord nous ramener à un modèle statistique exponentiel où les variances des lois mises en jeu sont pratiquement égales.

Dans l'épreuve mystérieuse, nous avons construit la loi du couple (X,Y) selon les relations suivantes :

$$P\{X = 0\} = 0.5 \quad \text{et } P\{Y = 0/ X = 0\} = P\{Y = 1/ X = 1\} = 0.6\ .$$

Les variables aléatoires $1 - Y$ et Y suivent, respectivement sous les hypothèses $X = 0$ et $X = 1$, une loi de *Bernoulli* de paramètre 0.6. Par conséquent, pour 100 épreuves, la variable $N_{11} + N_{22}$ suit une loi *Binomiale* $\mathcal{B}(100,0.6)$, que l'on peut approximer par la loi normale $\mathcal{N}(60,24)$.

Dans l'épreuve connue, les variables X et Y sont indépendantes. Par conséquent, pour 100 épreuves, la variable $N_{11} + N_{22}$ suit une loi *Binomiale* $\mathcal{B}(100,0.5)$, que l'on peut approximer par la loi normale $\mathcal{N}(50,25)$.

Comme les variances de ces deux lois normales sont très proches l'une de l'autre, le rapport de leurs vraisemblances permet d'affirmer (à deux approximations près) que $N_{11} + N_{22}$ est une statistique exhaustive.

Ainsi, on est amené à tester les deux hypothèses simples :

$$H_0)\ \mathcal{N}(50,25) \quad \text{contre} \quad H_1)\ \mathcal{N}(60,24).$$

Dans ce cas là, le test de Neyman-Pearson[1] est uniformément le plus puissant parmi tous les tests de seuil α.

La région de rejet de ce test est de la forme $D = \{N_{11} + N_{22} > c\}$. Au seuil 10%, $P_{H0}(D) = 0.1$ permet de calculer cette région : $D = \{N_{11} + N_{22} > 56.4\}$. Par ailleurs, la puissance de ce test vaut $P_{H1}(D) = 0.77$.

1 *Pour plus de détails sur ce test, consulter M. Fisz.*

ANNEXE 2

L'ensemble des tableaux auxquels les étudiants ont échoué sont les suivants : (T représente la valeur du χ^2 du tableau observé et “Bonne réponse” représente l'épreuve programmée.)

DEB

25	26
22	27

Bonne réponse : myst

T = 0.17

24	18
24	34

Bonne réponse : myst

T = 2.42

22	34
24	20

Bonne réponse : connue

T = 2.30

*

15	32
31	22

Bonne réponse : connue

T = 7.08

26	25
17	32

Bonne réponse : myst

T = 2.70

PFE

28	22
28	22

Bonne réponse : connue

T ≈ 0

34	20
29	17

Bonne réponse : myst

T ≈ 0

30	18
30	22

Bonne réponse : connue

T = 0.24

FUR

32	22
28	18

Bonne réponse : myst

T = 0.02

24	23
34	19

Bonne réponse : connue

T = 1.75

RIC

			*							
	23	30	34	22	28	29	25	25	25	22
	25	22	16	28	26	17	18	32	30	23
Bonne réponse :	connue		connue		connue		myst		myst	
	T = 0.95		T = 5.84		T = 1.26		T = 1.99		T = 0.11	

STO

			**							
	32	18	36	14	27	26	22	24	30	27
	25	25	18	32	31	16	26	28	24	19
Bonne réponse :	myst		myst		connue		myst		connue	
	T = 1.99		T = 13.04		T = 2.30		T ≈ 0		T = 0.09	

ROT

			*		**	
	26	25	23	29	29	17
	22	27	32	16	21	33
Bonne réponse :	connue		connue		myst	
	T = 0.37		T = 5.07		T = 5.79	

PAU

	28	22	34	21
	31	19	26	19
Bonne réponse :	connue		connue	
	T = 0.37		T = 0.16	

SMI

23	27
20	30

Bonne : connue

T = 0.36

35	17
26	22

connue

T = 1.81

26	24
21	29

connue

T = 1.00

CER

32	17
25	26

Bonne réponse : myst

T = 2.70

**

32	25
15	28

myst

T = 4.44

26	19
22	33

myst

T = 3.13

BIBLIOGRAPHIE

Alarcon, J. (1982) : *L'appréhension des situations probabilistes chez des élèves de 12-14 ans : résultats d'une enquête proposée à des élèves de 4ème et de 5ème.* Thèse de 3ème cycle, Université Louis Pasteur, Strasbourg.

Fisz, M. (1965) : *Probability theory and mathematical statistics.* Third edition, John Wiley.