Le trésor et les Sophies

Les solutions paradoxales des plus simples problèmes de probabilité engendrent de passionnantes controverses que les simulations permettent de trancher. Elles ont des prolongements en théorie de l'information quantique.

Notre raison est parfois attirée dans un piège dont nous ne nous libérerons qu'au prix d'un violent effort de concentration et de réflexion. Les deux exemples que nous présentons illustrent notre difficulté à raisonner juste, même lorsque nous sommes informés de la présence d'une difficulté. Il se peut d'ailleurs que vous ne soyez pas en accord avec les solutions que j'exposerai ; pire, vous pourriez les trouver stupides. Avant de m'envoyer des messages d'insultes, réfléchissez bien, car la très grande majorité des gens s'accordent sur les solutions que je présente, et je suis intimement persuadé qu'elles sont exactes.

Prendre la porte

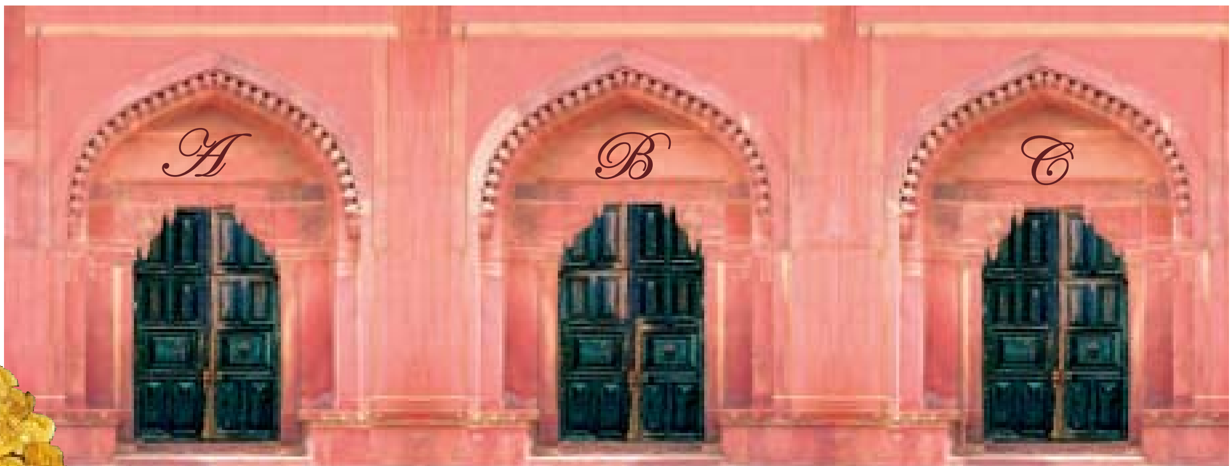
Le premier de nos deux problèmes est maintenant assez classique, mais il est tellement frappant qu'il serait dommage de l'ignorer. Sa première version, présentée en 1959 par Martin Gardner dans sa rubrique de *Scientific American,* invoquait des condamnés à mort, mais sa forme la plus connue a été portée à l'attention du public anglo-saxon par la journaliste Marilyn vos Savant qui, en 1990, le décrivit dans *Parade Magazine.* Cette nouvelle présentation suscita une série de terribles controverses.

Marilyn vos Savant le mentionne sous le nom du problème du *Monty Hall* en référence à une émission de télévision américaine (c'est sous ce nom que vous devez mener vos recherches sur Internet). Voici le problème.

Un vizir montre trois portes *A, B, C*, à Sindbad, son éventuel futur gendre. Il lui explique que derrière l'une d'elles se trouve un trésor et que derrière chacune des deux autres il n'y a rien. Sindbad doit choisir l'une des portes et gagne le trésor s'il choisit la bonne porte. Supposons que Sindbad désigne la porte *A*. Ce choix fait, il y a une autre étape. Le vizir explique :

« Je vais maintenant ouvrir une des deux portes que vous n'avez pas choisie (ici *B* ou *C*) et derrière laquelle il n'y a rien ; si vous avez choisi la porte du trésor, je tire au hasard (secrètement) l'une des deux portes restantes, et si vous avez choisi une porte derrière laquelle il n'y a rien, j'ouvre la seconde où il n'y a rien non plus. »

Le vizir, comme promis, ouvre la porte *C* dévoilant un espace vide et ajoute :



1. Changer de choix ? Le vizir propose à Sindbad d'ouvrir l'une des trois portes *A*, *B* et *C*. Il précise que derrière une seule des portes, on trouvera un trésor (dont l'emplacement a été choisi au hasard). Sindbad gagnera le trésor s'il ouvre (finalement) la bonne porte. Sindbad choisit la porte *A*, mais n'ouvre pas immédiatement cette porte. Le vizir ouvre la porte *C* et montre

« Je vous offre la possibilité de changer votre choix de la porte *A* et d'ouvrir plutôt la porte *B*. Que faites-vous ? » Notons bien qu'à chaque partie, les choses se déroulent de la même façon : Sindbad a la possibilité de changer son premier choix après la vision de la porte ne donnant sur rien.

La très grande majorité des gens pensent, de prime abord, qu'il n'y a aucun intérêt à changer de choix, c'est-à-dire à ouvrir la porte *B* plutôt que la porte *A*. En effet, dès que le vizir a ouvert la porte *C*, Sindbad sait que le trésor est derrière l'une des deux portes *A* ou *B* et il ne sait rien de plus. Le trésor est donc soit derrière la porte *A* avec une probabilité de 50 pour cent, soit derrière la porte *B* avec une probabilité de 50 pour cent : il n'y a aucun avantage à changer de choix.

Ce raisonnement est faux : vous devez modifier votre choix car, en changeant de porte, vous doublerez vos chances de repartir avec le trésor. Pour vous en convaincre, avançons plusieurs raisonnements en gardant à l'esprit qu'il faut toujours profiter d'une information supplémentaire.

La méthode la plus simple consiste à imaginer qu'après que Sindbad a choisi la porte *A*, le vizir lui propose : « Soit vous ouvrez *A*, soit vous ouvrez *B et C*. » Il ne fait aucun doute que Sindbad a intérêt à ouvrir *B* et *C*, car cela double ses chances de gagner le trésor. Cette variante est équivalente à celle réellement proposée par le vizir, qui, en réalité, ne fait qu'aider Sindbad à ouvrir les deux portes en s'occupant d'une des portes lui-même. Donc, en changeant son choix, Sindbad double ses chances.

Contrairement à ce que suppose le raisonnement spontané, les deux portes *A* et *B* ne sont pas en position symétrique. L'une, *B*, est le résultat d'un processus de sélection qui a concentré sur elle les probabilités de gain de deux portes *B* et C, alors que l'autre, *A*, avait une chance sur trois de cacher le trésor avant que le vizir n'ouvre une porte muette et a toujours une chance sur trois de cacher un trésor après l'intervention du vizir. Le vizir n'a évidemment rien changé à la probabilité que le trésor soit derrière la porte *A*.

Si ces considérations ne vous persuadent pas, voici une analyse plus minutieuse. Les cas possibles sont :

Cas 1 : porte *A* trésor ; porte *B* rien ; porte *C* rien.
Cas 2 : porte *A* rien ; porte *B* trésor ; porte *C* rien.
Cas 3 : porte *A* rien ; porte *B* rien ; porte *C* trésor.

Chacun de ces cas a une probabilité 1/3 de se produire.

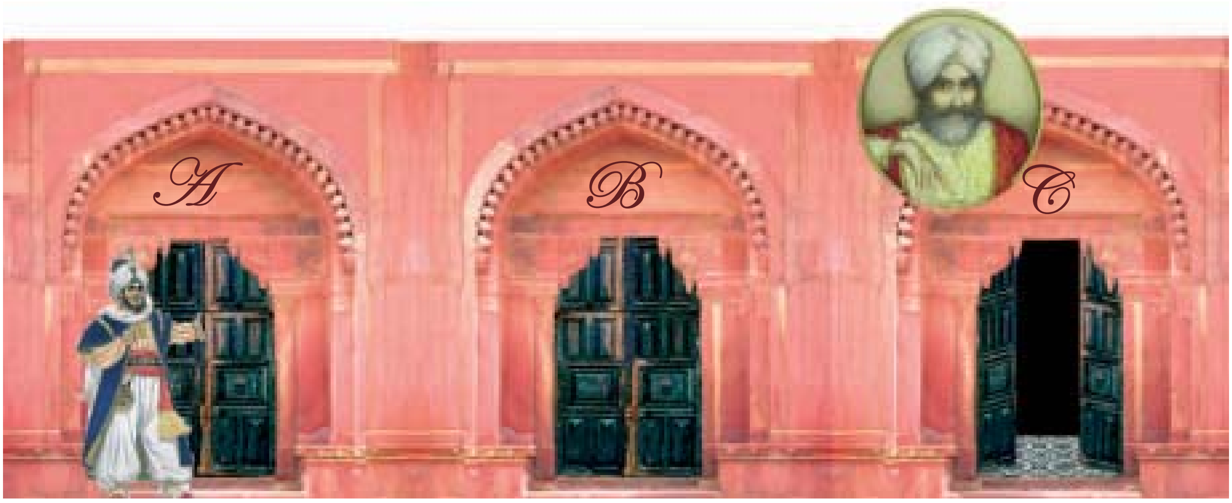
Dans le cas 1, quelle que soit la porte que le vizir ouvre (*B* ou *C*) si Sindbad change, il perd.

Dans les cas 2 et 3, en revanche, Sindbad, en changeant, gagne à chaque fois. Donc, au total, en changeant de porte, Sindbad gagne 2 fois sur 3.

Si ce raisonnement ne vous persuade toujours pas, imaginez la variante suivante du jeu. Il y a maintenant 100 portes, 99 sont perdantes et une masque un trésor. Après le premier choix de Sindbad, le vizir ouvre 98 des 99 portes qui restent, montrant 98 vides ! Sindbad a la possibilité de maintenir son choix initial ou de le changer pour la porte qui n'a pas été ouverte parmi les 99 autres. Sans aucun calcul ni raisonnement, ne vous semble-t-il pas certain qu'elle a plus de chances de cacher le trésor ?

Un sujet de débats

Si vous n'êtes toujours pas convaincu, sachez que même le grand mathématicien Paul Erdös n'a accepté l'idée qu'il fallait changer son choix initial qu'après une longue discussion avec son collègue Ron Graham et que certains experts en statistiques sont restés longtemps convaincus que la solution défendue par Marilyn vos Savant était idiote. Ils lui ont écrit des lettres très longues et parfois peu aimables pour lui expliquer qu'elle était dans l'erreur.



qu'elle ne cache pas le trésor espéré (lorsque le trésor est derrière la porte choisie par Sindbad, le vizir choisit au hasard entre les deux autres, sinon il choisit la seule des deux autres portes derrière lesquelles il n'y a rien). Le vizir propose à Sindbad de changer son choix (donc d'ouvrir *B* plutôt que *A* choisi en premier). Question : Sindbad a-t-il intérêt à changer de choix ? Que feriez-vous à sa place ?

2. Les variantes du trésor caché

Dans toutes ces variantes du problème de *Monty Hall*, la question est la même : Sindbad a-t-il intérêt à maintenir son choix initial ?

1. Après le choix initial de Sindbad, le vizir ouvre au hasard une des deux portes non choisies. S'il découvre le trésor, la partie est annulée et la partie est recommencée. Est-il encore vrai que Sindbad a intérêt à changer de choix lors des parties qui ne sont pas annulées ?
2. Après le choix initial de Sindbad, le vizir ouvre au hasard une des deux portes ne donnant sur rien. S'il ouvre la porte choisie par le joueur, la partie est annulée et la partie est recommencée. Même question : Sindbad doit-il changer ?
3. Lorsque le trésor se trouve derrière la porte choisie par Sindbad, le vizir, qui normalement doit choisir au hasard entre les deux autres, choisit systématiquement la porte la plus à gauche parmi les portes restantes (si le joueur a choisi *A* et qu'elle cache le trésor, le vizir ouvre donc la porte *B*). Même question : Sindbad doit-il changer ?

Solutions

1. Le problème est de toute évidence équivalent au problème initial. Sindbad a donc intérêt à changer de choix.
2. Problème encore équivalent. Sindbad a donc intérêt à changer.
3. Supposons que Sindbad ait choisi *A*. Il y a 3 cas équiprobables :

Cas 1 : *A* trésor ; *B* rien ; *C* rien.
Cas 2 : *A* rien ; *B* trésor ; *C* rien.
Cas 3 : *A* rien ; *B* rien ; *C* trésor.

Dans le cas 1 et 3, le vizir indique *B*. Dans le cas 2, il indique *C*.

Quand il indique *C*, Sindbad est certain que le trésor est en *B*, et donc il change son choix pour être certain de gagner. En revanche, quand le vizir indique *B*, une fois sur deux, le trésor est en *A*, et une fois sur deux, il est en *C*, Sindbad n'a donc pas d'intérêt particulier à changer son choix. Au total, la bonne stratégie consiste encore à changer de choix. Selon que le vizir indique la porte *C* ou la porte *B*, Sindbad gagnera avec certitude ou avec une probabilité de 1/2. En moyenne, il gagnera avec une probabilité de : $1/3+1/2\times 2/3 = 2/3$, comme dans le jeu initial !

Il existe heureusement une méthode ultime pour se persuader de la justesse de l'idée qu'il faut changer de porte : la simulation. Jouez au jeu une dizaine de fois en adoptant la stratégie « ne pas changer » et ensuite une dizaine de fois en adoptant la stratégie « changer » et faite le bilan. En cherchant sur Internet à partir des mots *Monty Hall* et *Applet* vous trouverez des programmes informatiques qui faciliteront ces simulations et qui normalement devraient vous convaincre qu'on double ses chances en changeant de porte

En résumé, si vous ne changez pas de porte vous n'avez qu'une chance sur trois de trouver le trésor, et si vous décidez d'en changer, vous gagnerez lorsque votre premier choix sera une porte derrière laquelle il n'y a rien, c'est-à-dire deux fois sur trois.

Mais dans le monde quantique ?

Une théorie nouvelle prend en compte les caractéristiques particulières de l'information dans notre monde physique, qui est quantique. Cette théorie quantique de l'information éclaire certains jeux sous une lumière nouvelle et le problème de *Monty Hall* a ainsi été étudié sous l'angle quantique. Dans la version quantique, le trésor est mis derrière une porte, mais les états superposés de « le trésor est en *A* », « le trésor est en *B* » et « le trésor est en C » sont possibles, ce qui veut dire que le trésor est derrière *A* en même temps qu'il est derrière *B* et derrière *C*. Le vizir pour indiquer une porte ne comportant pas le trésor doit faire une opération de mesure, et le joueur est autorisé à utiliser des stratégies mixtes du type une fois sur deux je ne change pas mon choix, une fois sur deux je le change.

La conclusion des calculs assez compliqués (le monde quantique ne se laisse pas appréhender facilement) se résume de la manière suivante : si le vizir peut utiliser des moyens quantiques, mais pas le joueur, alors le vizir peut s'arranger pour ne donner qu'une chance sur deux au joueur de gagner le trésor. Si c'est le joueur qui peut pleinement utiliser les manipulations quantiques et que le vizir n'y a pas accès, alors le joueur gagnera toujours.

Dans le cas où les deux joueurs peuvent utiliser les moyens quantiques, alors assez étrangement, les résultats obtenus sont les mêmes que dans le cas classique : le joueur gagne deux fois sur trois. Vous trouverez des détails et les calculs dans les articles cités à la fin de l'article.

L'énigme des Sophies

Le second exemple de piège troublant est *L'énigme des Sophies*, proposé par Jacques Patarin – mathématicien spécialiste de cryptologie – qui me l'a communiqué il y a quelques mois (d'autres exemples proches, mais moins frappants ont été mentionnés dans des textes de probabilités).

Il faut traiter simultanément deux problèmes :

Problème 1 : Une famille de deux enfants a au moins une fille. Quelle est la probabilité P_1, pour que cette famille ait deux filles ?

Problème 2 : Une famille de deux enfants a au moins une fille qui s'appelle Sophie. Quelle est la probabilité P_2 pour que cette famille ait deux filles ?

Les deux énoncés semblent appeler des réponses identiques : qu'importe que la fille dont on affirme l'existence se nomme Sophie ou Gertrude ! L'information supplémentaire donnée dans le second énoncé ne sert à rien, semble-t-il, et donc certainement P_1 est égal à P_2.

Menons soigneusement l'analyse des différents énoncés.

Pour le problème 1, considérons une famille de deux enfants prise au hasard composée d'un aîné E_1 et d'un cadet E_2. Il y a 4 cas :

E_1 garçon, E_2 garçon ;
E_1 garçon, E_2 fille ;
E_1 fille, E_2 garçon ;
E_1 fille, E_2 fille.

Chacun des cas a la même probabilité, 1/4, de se produire, car on néglige la petite différence de proportion qu'il y a entre garçons et filles dans les populations humaines (elle dépend d'ailleurs de l'âge), et les difficultés créées par les jumeaux : prendre en compte ces complications ne changerait les résultats que de manière infime.

L'énoncé du problème 1 nous indique que nous sommes dans l'un des trois derniers cas. Parmi les trois derniers, un seul correspond à une famille avec deux filles, la réponse est donc $P_1 = 1/3$. C'est un peu choquant, et différent de la probabilité 1/2 qu'on donnerait si, sans faire de calcul, on devait répondre instantanément ; cependant quiconque maîtrise un tant soit peu le calcul élémentaire des probabilités acquiescera, je crois, à la réponse $P_1 = 1/3$.

Considérons maintenant le second problème. On supposera que, dans une famille, jamais deux filles n'ont le même prénom et, donc, l'une d'elles seulement se prénomme Sophie. Soit une famille tirée au hasard ayant une fille prénommée Sophie, notons E_1 l'enfant qui s'appelle Sophie (on peut introduire cette notation puisque cet enfant est unique) et E_2 l'enfant qui ne s'appelle pas Sophie. Deux cas sont possibles et de même probabilité :

E_2 est un garçon ;
E_2 est une fille.

La probabilité pour que le second enfant soit une fille est donc 1/2 : $P_2 = 1/2$. Contrairement à ce qui semblait évident à première vue, P_1 et P_2 sont différents. Notons que si la formulation du problème 2 était qu'un enfant s'appelle Dominique, nous ne serions pas plus avancés, dans la mesure où Dominique peut être une fille ou un garçon et ne nous donne aucune information supplémentaire.

Mieux comprendre ?

Une sorte d'explication de la différence entre P_1 et P_2 consiste à dire que dans le problème 2 une famille ayant 2 filles a deux fois plus de chances d'avoir une fille qui s'appelle Sophie qu'une famille ayant une seule fille. Donc que le poids des familles ayant deux filles est en quelque sorte doublé dans le second problème alors qu'il ne l'est pas dans le premier. Je crains cependant que cette explication ne soit pas jugée suffisante par ceux qui doutent du résultat $P_1 \neq P_2$.

Le meilleur argument, comme l'indique Jacques Patarin, provient de l'introduction d'un paramètre continu. On suppose que la probabilité pour qu'une fille se prénomme Sophie est p (on n'exclut pas ici qu'une famille ait deux filles portant le prénom Sophie). Lorsque p sera égal à 1, on retombera sur le problème 1, et lorsque p sera très petit on s'approchera du problème 2.

Considérons une famille de deux enfants, en prenant en compte le prénom des filles. Neuf cas sont possibles ayant chacun une probabilité de se produire qui dépend de p. Notons E_1 l'aîné des enfants et E_2 le cadet.

1. E_1 Sophie, E_2 garçon ; probabilité : 1/2 x p x 1/2.
2. E_1 Sophie, E_2 fille non Sophie ; probabilité :1/2 x px1/2x(1– p).
3. E_1 Sophie, E_2 Sophie ; probabilité : 1/2 x p x 1/2 x p.
4. E_1 fille non Sophie, E_2 garçon ; probabilité : 1/2 x (1 – p) x 1/2.
5. E_1 fille non Sophie, E_2 fille non Sophie ; probabilité : 1/2x(1 – p) x 1/2 x (1 – p).
6. E_1 fille non Sophie, E_2 Sophie ; probabilité : 1/2x(1–p)x 1/2 x p.
7. E_1 garçon, E_2 garçon ; probabilité : 1/2 x 1/2
8. E_1 garçon, E_2 fille non Sophie ; probabilité : 1/2 x 1/2 x (1 – p)
9. E_1 garçon, E_2 Sophie ; probabilité : 1/2 x 1/2 x p (on n'envisage pas qu'un garçon puisse se prénommer Sophie !)

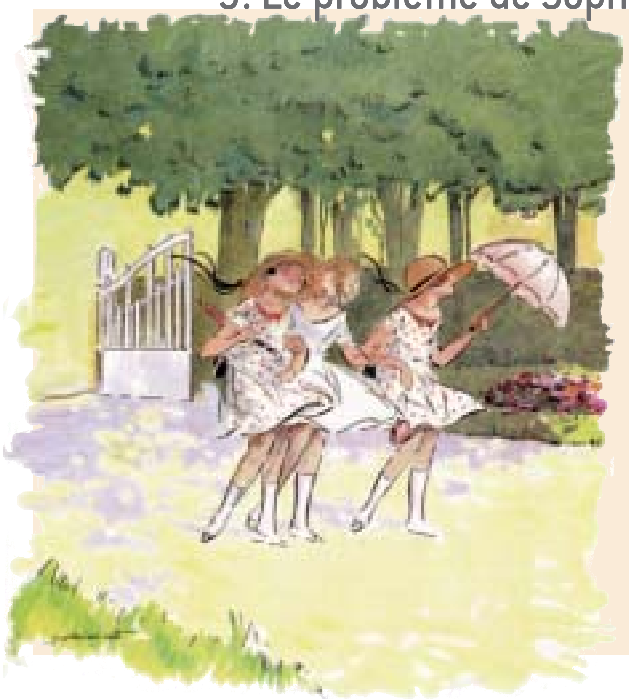
Nous allons calculer la probabilité p pour qu'une famille de deux enfants ayant une fille qui s'appelle Sophie ait deux filles. Cette probabilité est donnée par le rapport P :

P = (probabilité pour qu'il y ait 2 filles dont une au moins s'appelle Sophie)/(probabilité pour qu'il y ait au moins une fille qui s'appelle Sophie).

Notons A le numérateur et B le dénominateur. Nous obtenons A en additionnant les probabilités des cas 2, 3 et 6. Similairement, nous calculons B en additionnant les probabilités des cas 1, 2, 3, 6, 9. En définitive, après des calculs assez simples, nous obtenons :

$A = p/4 + p(1 - p)/4 = p(2 - p)/4$
$B = p/2 + p(2 - p)/4 = (4p - p^2)/4$
$P = [2p - p^2]/[4p - p^2] = (2 - p)/(4 - p)$

3. Le problème de Sophie dans une famille plus nombreuse



Sophie, dans la Comtesse de Ségur, n'avait pas de frère et sœur ! Mais elle côtoyait de grandes familles, et ses problèmes ou ses *malheurs* étaient nombreux. Nous symboliserons les problèmes de Sophie par l'illustration d'André Pécoud dans l'édition de 1930 des *Malheurs de Sophie*.

Problème 1 : Une famille de trois enfants a au moins une fille. Quelle est la probabilité, R_1, pour que cette famille ait trois filles ?
Problème 2 : Une famille de trois enfants a au moins une fille qui s'appelle Sophie. Quelle est la probabilité, R_2, pour que cette famille ait trois filles ?

Solutions

Problème 1. Cas possibles équiprobables : FFF FFG FGF FGG GFF GFG GGF GGG. Cas où il y a au moins une fille : FFF FFG FGF FGG GFF GFG GGF. Donc $R_1 = 1/7$.
Problème 2. Soit E_2 et E_3 les deux enfants qui ne se prénomment pas Sophie. Il y a 4 cas équiprobables :
(1) E_1 = F ; E_2 = F. (2) E_1 = F ; E_2 = G (3) E_1 = G ; E_2 = F (4) E_1 = G E_2 = G.
Donc $R_2 = 1/4$.
(Remarque : dans le cas de N enfants, on trouverait $R_1 = 1/[2^N - 1]$ et $R_2 = 1/2^{N-1}$).

Maintenant, nous pouvons vérifier et confirmer nos évaluations de P_1 et P_2.

Si $p = 1$ alors $P = 1/3$. Cela correspond au cas où toutes les filles se prénomment Sophie, et donc cela correspond au problème 1. Nous retrouvons la valeur $P_1 = 1/3$ de tout à l'heure.

Le problème 2 correspond au cas où très peu de filles se prénomment Sophie, c'est-à-dire au cas où p tend vers 0. Nous trouvons alors que, P tend vers 1/2, ce qui est bien ce que nous avions calculé $P_2 = 1/2$.

Là encore ceux qui doutent peuvent avoir recours à la simulation : on se donne par exemple 100 cartons avec un carton marqué Sophie, 49 marqués fille et 50 marqués garçons. On évalue P_1 en tirant deux cartons au hasard, en rejetant les deux cartons si aucun n'est une fille, puis en comptant parmi les tirages retenus la proportion de ceux qui sont de type fille-fille. On évalue P_2 en tirant deux cartons au hasard, en les rejetant si aucun ne porte le prénom Sophie, et en comptant parmi les tirages retenus la proportion de ceux qui sont de type Sophie-fille ou fille-Sophie.

Notons que dans ces expériences, on n'obtiendra pas exactement 1/3 et 1/2, mais des valeurs approchées. Cela est dû, d'une part, à l'aspect statistique de l'expérience, mais aussi au fait qu'on effectue des tirages sans remise, ce qui fausse légèrement les probabilités limites. Les résultats obtenus expérimentalement permettent cependant de bien mettre en évidence les valeurs théoriques de P_1 et P_2 et donc de lever vos derniers doutes… si vous en aviez.

Les deux problèmes pièges montrent que des exercices de probabilité, même élémentaires, cachent parfois des difficultés que notre intuition est impuissante à maîtriser totalement. Heureusement les techniques et les instruments de calcul – voire de simulation – dont nous disposons maintenant permettent de résoudre ces énigmes et de se mettre d'accord… je l'espère. N'oublions pas qu'il n'en fut pas toujours ainsi, et que par exemple Jean le Rond d'Alembert pensait que jeter une pièce de monnaie trois fois de suite n'est pas équivalent à jeter trois pièces à la fois, et croyait (comme aujourd'hui certains joueurs de casino mal informés) qu'après une longue série de tirages pile, le tirage face est plus probable...

Auteur & Bibliographie

Jean-Paul DELAHAYE est professeur d'informatique à l'Univ. de Lille.

Martin GARDNER, *The Colossal Book of Mathematics*, W.W. Norton Company, New York, 2001 (voir le chapitre 21 qui contient une version mise à jour de l'article d'octobre 1959 de *Scientific American*).

Marilyn VOS SAVANT, *Ask Marilyn,* in *Parade Magazine*, p. 12, 17 février 1990.

G.M. D'ARIANO, R.D. GILL, M. KEYL, R.F. WERNER, B. KÜMERRER et H. MAASSEN, *The Quantum Monty Hall Problem,* 2001 : http://www.imaph.tu-bs.de/qi/monty/

A.P. FLITNEY, et D. ABBOTT, *Quantum version of the Monty Hall Problem,* quant-ph/0109035 (2001) : http://arxiv.org/abs/quant-ph/0109035

Encyclopédie Wikipedia : *Monty Hall Problem* 2005 http://en.wikipedia.org/wiki/Monty_Hall_problem

Ian STEWART, *Persée, Pégase et Andromède*, *Pour la science,* n°167, pp. 92-95, septembre 1991.