



MODÉLISATION EN PROBABILITÉS AU LYCÉE.

GILLES ALDON
Josette FEURLY-REYNAUD

**INSTITUT DE RECHERCHE SUR
L'ENSEIGNEMENT DES MATHÉMATIQUES**

ACADÉMIE DE LYON

**Université Claude Bernard Lyon I - 43 Bd du 11 Novembre 1918
69622 VILLEURBANNE**

Septembre 94 - 40 F

MODÉLISATION EN PROBABILITÉS AU LYCÉE.

GILLES ALDON
Josette FEURLY-REYNAUD

**INSTITUT DE RECHERCHE SUR
L'ENSEIGNEMENT DES MATHÉMATIQUES**

ACADÉMIE DE LYON

**Université Claude Bernard Lyon I - 43 Bd du 11 Novembre 1918
69622 VILLEURBANNE**

Septembre 94 - 40 F

PRÉSENTATION

La mise en place des derniers programmes de probabilités au lycée insiste sur l'introduction de la notion de probabilités à partir des fréquences statistiques. De nombreux travaux ont déjà été réalisés à propos de cette nouvelle approche (cf Bibliographie). L'analyse faite par exemple par A. et M. Henry montre que le passage entre fréquences et probabilités ne va pas de soi, et ne relève pas, en tout cas, de la simple stabilisation des fréquences sur un grand nombre d'expériences. Notre réflexion nous a conduit à repérer la démarche de modélisation comme très importante et incontournable dans cet enseignement : présenter les probabilités comme modélisation mathématique d'une situation physique permet de les distinguer de la réalité statistique.

On peut discuter pour savoir si ce passage de l'expérience au modèle est à charge du mathématicien, lequel préfère souvent travailler en aval, dans le confort du modèle. Mais on doit bien reconnaître qu'un travail de cette nature est demandé aux élèves dans le cadre des problèmes habituels qui leur sont posés : ils doivent à partir de situations "pseudo-réelles", utilisant des objets tels que dés, jetons, boules ou autres, bâtir un modèle en adéquation avec les conditions de l'expérience aléatoire qui leur est décrite. On se doit donc de ne pas escamoter cette démarche, qui est à la fois une étape importante en probabilités, et une exigence de l'institution scolaire.

Le point de vue adopté tout au long de cette brochure est de travailler explicitement sous l'angle de la modélisation, qui constitue donc un fil conducteur tant pour la présentation de situations de probabilités classiques que de probabilités conditionnelles.

Le but de cette publication est d'alimenter une réflexion générale sur l'enseignement des probabilités. Nous nous sommes appuyés sur des exemples divers, parfois classiques, parfois déstabilisants, qui ne sont pas écrits pour être utilisés tel quels avec des élèves, mais qui peuvent être repris et adaptés pour être transposés en situations de classe.

La première partie présente différentes modélisations d'un même problème et les liens entre solution mathématique, dispositif physique de réalisation de l'expérience aléatoire et simulation.

La deuxième partie essaie, à travers un exercice classique de mettre en évidence les implicites liés aux situations de modélisation.

La troisième partie présente, à propos de probabilités conditionnelles, un outil de modélisation qui permet de dégager pour ces problèmes une structure commune.

SOMMAIRE

1 - CHOISIR SON HASARD	Page 5
1 Un problème et des solutions	Page 5
2 Mais alors...	Page 12
3 Est-ce choquant ?	Page 16
 2 - UN PROBLÈME BIEN ORDINAIRE...	 Page 17
1 L'énoncé et "sa" solution	Page 17
2 Deux autres solutions	Page 19
3 Quelques commentaires	Page 21
 3 - ARBRE ET PROBABILITÉS CONDITIONNELLES	 Page 22
1 Un détour par l'outil "arbres"	Page 22
2 Utilisation dans les problèmes	Page 25
3 L'enseignement des probabilités conditionnelles	Page 34
 4 - QUELQUES ELEMENTS DE BIBLIOGRAPHIE	 Page 38

CHOISIR SON HASARD

Enoncé : On trace au hasard une corde dans un cercle. Quelle est la probabilité pour qu'elle soit plus petite que le côté du triangle équilatéral inscrit ?

Consigne : Donnez une solution mathématique à ce problème, décrivez un système physique permettant de simuler l'expérience aléatoire et écrivez un algorithme permettant de simuler cette expérience sur une calculatrice ou un ordinateur.

Ce problème, connu sous le nom de "paradoxe de Bertrand" a donné lieu à de nombreuses controverses ; citons en particulier l'étude que Poincaré en a faite dans "Calcul des probabilités", et celle de Borel, qui écrit dans "Le hasard" :

"Joseph Bertrand, dans son célèbre ouvrage sur le Calcul des probabilités, a critiqué la théorie des probabilités continues...la conclusion de Bertrand paraît être que les problèmes de probabilités continues sont de purs jeux mathématiques, mais ne correspondent à aucune réalité...Cette attitude sceptique n'est pas en rapport avec les faits. On peut s'en rendre compte aisément en empruntant un exemple même de Bertrand..."

Et Borel cite et traite le problème de Bertrand à l'appui de sa thèse.

Il s'agit donc d'un problème qui historiquement a été moteur dans la construction de la théorie des probabilités actuelle.

1 DES SOLUTIONS

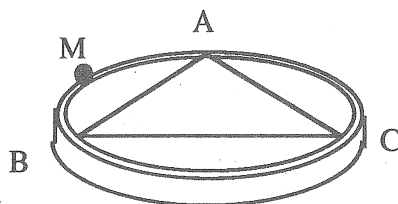
Plusieurs solutions sont envisagées dans la suite, et pour chacune, sont proposés à la fois un dispositif physique de réalisation de l'expérience, une simulation (algorithme et programme sur TI82), et une solution mathématique. L'ordre de ces trois parties dans chaque solution est volontairement différent chaque fois : en effet, la réflexion peut commencer de différentes façons, soit qu'on imagine ou qu'on crée un dispositif pour réaliser l'expérience aléatoire, soit qu'on commence par une simulation, soit enfin qu'on se lance directement dans une démonstration mathématique.

Si l'ordre adopté pour présenter ces trois approches importe peu, la cohérence entre elles est en revanche essentielle du point de vue de la modélisation.

Première solution

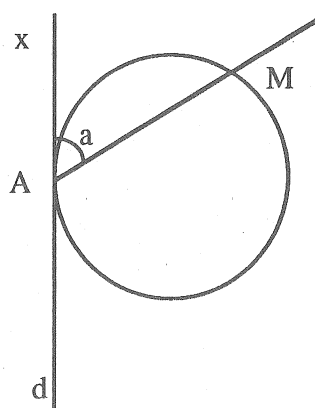
Dispositif physique :

Sur un rail circulaire, est lancée une bille (on suppose que les conditions sont "idéales": pas de frottements, plan parfaitement horizontal...). De l'endroit où cette bille s'arrête, on trace un segment joignant un point A fixé sommet d'un triangle équilatéral ABC dessiné à l'avance. La condition "la corde est plus petite que le côté AB" se réalise si M est à l'extérieur de l'arc BC.



Solution mathématique :

On fait l'hypothèse que l'on peut, pour des raisons de symétrie, fixer arbitrairement l'une des extrémités de la corde en A :

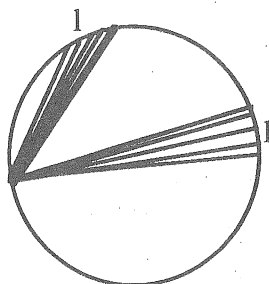


Soit $C(O,R)$ un cercle de centre O et de rayon R . Soit A un point de C . Soit d la tangente à C en A . Pour déterminer la corde AM , on choisit un angle a entre 0 et π . Le point M est le point du cercle tel que l'angle de $[A,x)$ vers $[AM]$ vaille a .

La corde AM est plus petite que le côté du triangle équilatéral lorsque a est compris entre 0 et $\pi/3$ ou entre $2\pi/3$ et π . La probabilité cherchée est alors de $2/3$

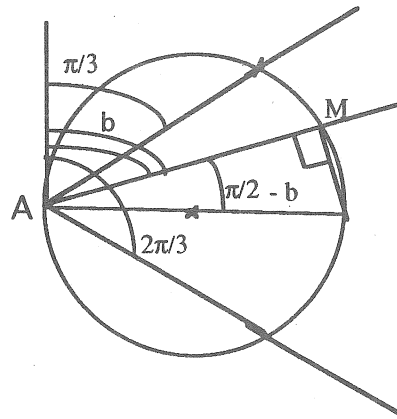
Cette solution suppose que quelle que soit la position d'arcs de cercle de longueur donnée l , il y a la même probabilité pour que la corde choisie au hasard ait sa deuxième extrémité sur l'un ou l'autre de ces arcs.

Cela revient à dire que pour la variable position du deuxième point sur le cercle, la probabilité est uniformément répartie sur la circonférence.



Algorithme pour simuler

On peut faire une simulation de l'expérience aléatoire en choisissant une mesure d'angle entre 0 et π , et en décidant que ce nombre est la mesure de l'angle entre une demi-tangente en A à un cercle de rayon 1, et la corde "choisie au hasard" de cette manière. On détermine ainsi la corde AM :



Initialisation :

s représente le nombre de succès
n comptabilise le nombre d'essais

Tant que n est inférieur à 10000 faire

n: = n+1

choisir un nombre au hasard entre 0 et π : b

calculer la longueur de la corde
a:= 2sinb

Comparer la longueur de la corde et la longueur du côté du triangle équilatéral inscrit

si $a < \sqrt{3}$ alors s:=s+1 fin si

fin tant que

afficher s/10000

Un programme possible sur une calculatrice TI82 , et les résultats qu'il fournit :

```
0->S
0->N
Lbl 0
rand*π->B
2sinB->A
If A<√3
S+1->S
N+1->N
If N<10000
Goto 0
S/N->F
Disp F
```

L'exécution de ce programme sur la machine, lancée trois fois sur 1000 essais seulement, a donné successivement :

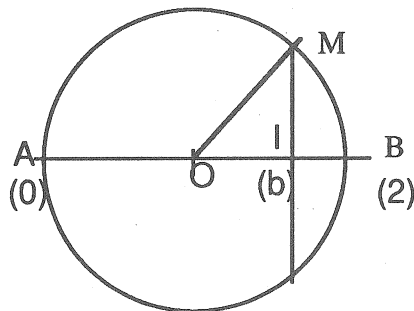
0,669
0,654
0,665

tous résultats proches de 2/3.

Deuxième solution

Algorithme pour simuler :

On peut faire une simulation en choisissant de l'expérience aléatoire en choisissant un nombre entre 0 et 2. On décide que ce nombre détermine la place d'un point I sur un diamètre [AB] d'un cercle de rayon 1. La corde "choisie au hasard" est alors la corde perpendiculaire à [AB] en I.



$$OI^2 = (b - 1)^2 \text{ ou } (1 - b)^2$$

$$IM^2 = 1^2 - (b - 1)^2 \text{ ou } 1^2 - (1 - b)^2 \\ = 2b - b^2$$

initialisation

s représente le nombre de succès

n comptabilise le nombre d'essais

Tant que n est inférieur à 10000 faire

n: = n+1

choisir un nombre au hasard entre 0 et 2: b

Ce nombre détermine un point du diamètre [AB] : I

Calculer la longueur de la corde perpendiculaire au rayon OI

$$a: = 2\sqrt{2b - b^2}$$

Comparer la longueur de la corde et la longueur du côté du triangle équilatéral inscrit

si $a < \sqrt{3}$ alors s : = s+1 fin si

fin tant que

afficher s/10000

Un programme possible sur une calculatrice TI82, et les résultats qu'il fournit :

0->S

0->N

Lbl 0

rand*2->B

$2\sqrt{2B - B^2}$ ->A

If $A < \sqrt{3}$

S+1->S

N+1->N

If $N < 10000$

Goto 0

S/N->F

Disp F

L'exécution de ce programme, lancée trois fois sur 1000 essais seulement, a donné successivement les résultats suivants :

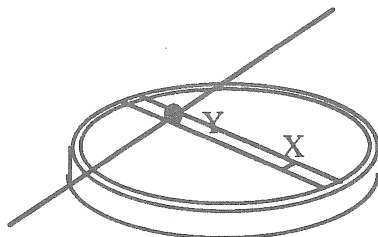
0,499

0,489

0,501

Dispositif physique :

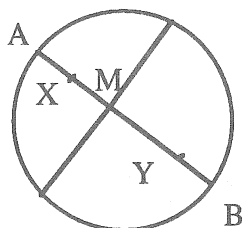
Un rail est placé le long d'un diamètre du cercle. Une bille mobile supporte une tige dont la direction est fixée perpendiculaire au rail. On lance la bille dans la rainure et on suppose là encore que les conditions sont idéales, en particulier qu'il n'y a pas de frottements, que le support est parfaitement horizontal,... La condition "la corde est plus petite que le côté du triangle équilatéral" est réalisée lorsque la bille s'arrête en dehors du segment $[XY]$, où X et Y sont placés au quart et au trois quart du rail.



Solution mathématique :

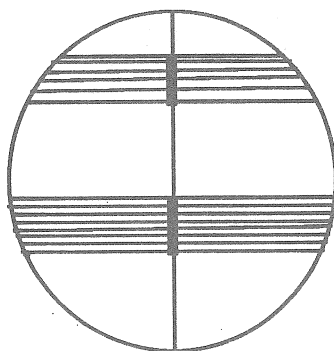
On peut supposer, pour des raisons de symétrie, que l'on s'intéresse aux cordes de directions fixées, perpendiculaires à un diamètre arbitrairement choisi $[AB]$.

Soit $C(O,R)$ un cercle de centre O et de rayon R . Soit $[AB]$ un diamètre du cercle. Soient X et Y les points de $[AB]$ définis par $AX = R/2$ et $YB = R/2$. Pour choisir une corde, on choisit un point M de $[AB]$. La corde est alors le segment perpendiculaire en M à $[AB]$. La condition est réalisée si M appartient à $[AX]$ ou à $[YB]$. La probabilité cherchée est donc $(AX + YB) / AB$, c'est à dire $1/2$.



Cette solution suppose que quelle que soit la place de segments de longueur donnée sur le diamètre, la probabilité d'obtenir une corde dont le milieu est sur l'un de ces segments est la même.

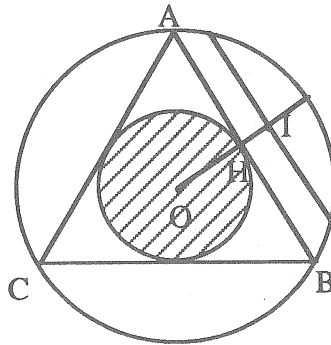
Cela revient à dire que la répartition de probabilité de la variable position du milieu d'une corde est uniforme sur le diamètre $[AB]$.



Troisième solution

Solution mathématique :

La corde [MN] est déterminée par son milieu I. On peut choisir I dans le disque de centre O et de rayon R. L'événement qui nous intéresse est réalisé si et seulement si on a $OI > OH$, c'est à dire $OI > R/2$. La probabilité cherchée est alors le rapport des aires de la couronne autour du disque de rayon R/2 et du disque de rayon R, ou $\frac{\pi R^2 - \pi(R/2)^2}{\pi R^2} = \frac{3}{4}$.



Cette solution suppose que quelle que soit la place et la forme d'un morceau du disque d'aire donnée, la probabilité d'obtenir une corde dont le milieu est dans l'une de ces surfaces est la même. Cela revient à dire que la répartition de probabilité de la variable milieu de la corde est uniforme sur le disque.

Algorithme pour simuler :

On peut réaliser une simulation de cette expérience aléatoire en choisissant un point du disque de rayon 1 par ses coordonnées cartésiennes dans un repère orthonormé dont l'origine est le centre du cercle. Ce point est alors le milieu de la corde cherchée.

```
initialisation
s représente le nombre de succès
n comptabilise le nombre d'essais

Tant que n est inférieur à 10000 faire

    choisir un nombre au hasard entre - 1 et 1 : a
    choisir un nombre au hasard entre - 1 et 1 : b

    si  $a^2 + b^2 \leq 1$  alors le point est dans le disque et faire
         $n := n + 1$ 
    Comparer la longueur de la corde et la longueur du côté du triangle équilatéral inscrit
        si  $2\sqrt{1 - a^2 - b^2} < \sqrt{3}$  alors  $s := s + 1$  fin si
        fin si

    fin tant que

afficher  $s/10000$ 
```

Voici un programme sur TI 82, et le résultat affiché à l'écran lorsqu'on lance le programme plusieurs fois sur 1000 lancers.

```

0->S
0->N
Lbl 0
rand*2-1->A
rand*2-1->B
If A2+B2>1
Goto 0
If  $2\sqrt{1-A^2-B^2} \leq \sqrt{3}$ 
S+1->S
N+1->N
If N<1000
Goto 0
S/N->F
Disp F

```

```

prgmSIM3      .77
               Done
prgmSIM3      .754
               Done

```

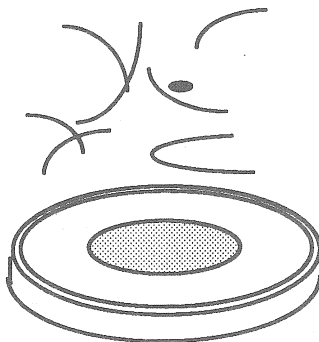
```

prgmSIM3      .732
               Done
prgmSIM3      .73
               Done

```

Dispositif physique :

On choisit au hasard un point dans le disque en lançant un confetti dans une boîte cylindrique où soufflent des courants d'air en tous sens. Le point d'atterrissage du confetti est le milieu de la corde. Les points qui conviennent sont ceux extérieurs au cercle dont le rayon est moitié de celui du fond de la boîte.



2 MAIS ALORS...

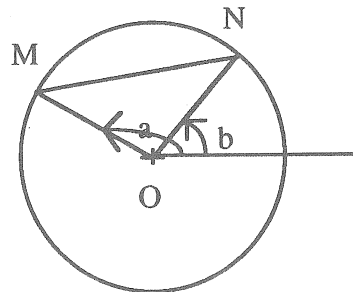
Peut-on trouver d'autres dispositifs physiques, d'autres simulations modélisant cette situation différemment et pouvant amener d'autres résultats à ce problème ?

• On peut par exemple s'étonner que dans la première solution, on fixe l'une des extrémités de la corde a priori et que la corde soit ensuite déterminée par le choix de l'autre extrémité. Le résultat serait-il identique si l'on choisissait deux points sur le cercle ?

Le dispositif physique pourrait dans ce cas se décrire de la manière suivante : deux rails circulaires entourent le disque. Dans chaque rail, on lance une bille (M et N). Quand ces billes s'arrêtent, on considère la corde [MN].

Ce qui nous amène à formaliser de la manière suivante :

Soient M et N deux points du cercle de centre O et de rayon R. On peut déterminer la corde par les coordonnées polaires de ses extrémités $M(R,a)$ et $N(R,b)$. L'univers peut alors être $[-\pi,\pi]^2$, avec a et b entre $-\pi$ et π . L'angle du vecteur OM vers le vecteur ON a une mesure de $b - a$, comprise entre -2π et 2π .



L'angle géométrique MON, saillant ou rentrant, a pour mesure $|b - a|$, qui est donc entre 0 et 2π . Ce nombre est à comparer à l'angle saillant ou rentrant qui intercepte le côté du triangle équilatéral inscrit, qui est de mesure $2\pi/3$ pour l'angle saillant, $2\pi - 2\pi/3$ pour le rentrant.

Pour que les points M et N réalisent l'événement "la corde MN est plus petite que le côté du triangle équilatéral inscrit, on doit avoir

- ou bien l'angle saillant MON inférieur à $2\pi/3$
- ou bien l'angle rentrant MON supérieur à $2\pi - 2\pi/3$

Le problème revient à chercher les couples de réels (a, b) tels que :

$$|b - a| \leq 2\pi/3 \text{ ou } |b - a| \geq 2\pi - 2\pi/3. \text{ Or :}$$

$$|b - a| \leq 2\pi/3 \Leftrightarrow b - a \leq 2\pi/3 \text{ et } b - a \geq -2\pi/3.$$

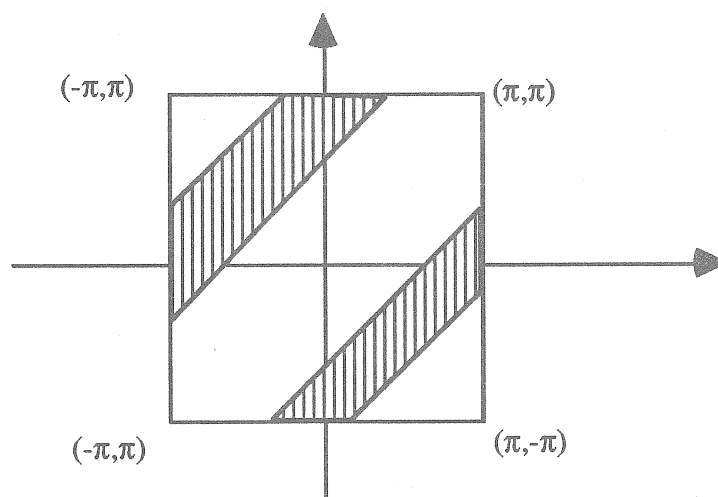
$$|b - a| \geq 2\pi - 2\pi/3 \Leftrightarrow |b - a| \geq 4\pi/3 \Leftrightarrow b - a \geq 4\pi/3 \text{ ou } b - a \leq -4\pi/3$$

Finalement les nombres cherchés sont tels que :

$$\begin{cases} b - a \leq 2\pi/3 \\ \text{et } b - a \geq -2\pi/3 \end{cases} \text{ ou } b - a \geq 4\pi/3 \text{ ou } b - a \leq -4\pi/3$$

On peut modéliser l'ensemble des choix possibles par le carré $[-\pi,\pi]^2$ représenté dans un repère orthonormé, chaque couple (a, b) correspondant à un point dans le plan, et à une corde du cercle.

Pour chercher les points du carré vérifiant les conditions précédentes, on doit tracer à l'intérieur du carré les droites d'équations $y - x = 2\pi/3$, $y - x = -2\pi/3$, $y - x = 4\pi/3$, $y - x = -4\pi/3$.



Une étude élémentaire montre que les points qui fournissent un couple (a, b) qui convient sont ceux intérieurs au carré qui ne sont pas dans les surfaces hachurées. On peut évaluer cette surface non hachurée, le résultat, obtenu par des différences d'aires simples est de $24 \pi^2 / 9$. La probabilité cherchée est alors le rapport de ce nombre à l'aire du carré, soit : $(24 \pi^2 / 9) / 4 \pi^2 = 6 / 9 = 2 / 3$.

Ce résultat est le même que dans la première solution, où l'une des extrémités de la corde avait été fixée arbitrairement.

• On peut remarquer de même que dans la seconde solution, le diamètre considéré est fixé d'emblée, et que toutes les cordes ainsi choisies ont la même direction. Le résultat ne serait-il pas changé si on choisissait au hasard ce premier diamètre ? (en faisant tourner par exemple une tige mobile autour du centre du disque).

La solution mathématique, en utilisant les mêmes notations qu'à la page 6, peut s'obtenir en choisissant le point I par ses coordonnées polaires (r, a) , a représente alors l'inclinaison variable du diamètre $[AB]$. La variable est alors un couple de deux nombres choisis dans l'univers $[0, R] \times [-\pi, \pi]$.

L'événement qui nous intéresse est alors modélisé par le rectangle : $[R/2, R] \times [-\pi, \pi]$, dont la surface est moitié de celle du rectangle univers.

La probabilité cherchée est donc de $1/2$. On obtient le même résultat que dans la deuxième solution, où la direction du diamètre avait été fixée.

• Mais on peut aussi choisir la corde par sa longueur...

C'est un nombre compris entre 0 et $2R$. La corde sera plus petite que le côté du triangle équilatéral inscrit si le nombre choisi est inférieur à $R\sqrt{3}$. La probabilité de cet événement sera alors $\frac{\sqrt{3}}{2}$. La simulation correspondante est simple à mettre en oeuvre. On suppose

dans ce cas que la probabilité d'obtenir une corde de longueur l est la même, quel que soit le nombre l .

• Ou bien on peut, par exemple, lancer deux confettis dans la boîte de la troisième solution, et décider que la corde est celle qui joint les deux points obtenus.

Choisissons le cercle de centre O et de rayon 1 et le repère orthonormé $(O, \vec{i}, \vec{j})$; les deux points K et L ont pour coordonnées (a,b) et (c,d) dans ce repère.
La simulation de cette expérience peut être décrite par l'algorithme suivant :

```

initialisation
s représente le nombre de succès
n comptabilise le nombre d'essais

Tant que n est inférieur à 50000 faire

    choisir un nombre au hasard entre -1 et 1 : a
    choisir un nombre au hasard entre -1 et 1 : b
    choisir un nombre au hasard entre -1 et 1 : c
    choisir un nombre au hasard entre -1 et 1 : d

    si  $(a^2 + b^2 \leq 1)$  et  $(c^2 + d^2 \leq 1)$  alors les points sont dans le disque et faire
        n := n+1
        Calculer la longueur de la corde d
        La comparer à la longueur du côté du triangle équilatéral inscrit
        si  $d < \sqrt{3}$  alors s := s+1 fin si
        fin si

    fin tant que

afficher s/50000

```

Cette simulation a été programmée en Turbo-Pascal et trois réalisations successives ont donné les résultats suivants :

s/50000 = 0,41034
s/50000 = 0,41696
s/50000 = 0,40864

Une question reste en suspens et nous préoccupe : qu'en est-il de la solution mathématique correspondante?

Une équation de la droite (KL) est : $x(d-b) + y(c-a) + bc - ad = 0$
La longueur de la corde passant par les points K et L est alors obtenue en cherchant les points d'intersection de (KL) et du cercle. Tous calculs faits, on obtient

$$d = 2\sqrt{-\frac{a^2(d^2 - 1) + 2ac(1 - bd) + b^2(c^2 - 1) + 2bd - c^2 - d^2}{a^2 - 2ac + b^2 - 2bd + c^2 + d^2}}$$

Il s'agit alors de déterminer pour quelles positions des points K et L cette distance est inférieure à $\sqrt{3}$.

Comment continueriez-vous ? Nous avons pour notre part une solution dont nous ne sommes pas sûrs, vos propositions seront les bienvenues !

On pourrait aussi modifier le dispositif de la troisième solution en laissant tomber une bille tenue à la verticale du centre de la base de la boîte, dont le fond serait rempli de sable pour éviter les rebonds.

Dans ce cas, les points proches du centre sont plus souvent obtenus, une répartition uniforme de probabilité des points dans le disque ne convient pas. On peut par exemple choisir une répartition normale en dimension 2, dont les paramètres sont à définir. On peut lire utilement pour mettre ce modèle en oeuvre l'article de M. Henry et H. Lombardi dans Repères n° 13.

3 EST-CE CHOQUANT ?

Le fait qu'il y ait plusieurs résultats possibles pour un problème de formulation simple apparaît d'emblée comme choquant pour le mathématicien. Le "paradoxe" est d'autant plus frappant que l'énoncé ne met en jeu que des objets très élémentaires, reconnus comme mathématiques : cercle, corde, côté du triangle équilatéral inscrit.

Devoir imaginer la réalisation du choix de la corde par un dispositif physique nous montre que plusieurs situations concrètes de choix aléatoires peuvent être imaginées sous le terme "au hasard" employé dans l'énoncé : il y a plusieurs façons de choisir une corde au hasard, le terme "au hasard" est insuffisant pour définir le modèle.

Définir pour chacune des expériences envisagées une probabilité, c'est choisir pour chacune un modèle en adéquation avec cette expérience. Les comparaisons des différentes solutions mathématiques mettent en évidence des choix de distribution de probabilité différents : les variables et donc les populations ne sont pas les mêmes dans ces solutions, non plus que les lois correspondantes : population de nombres, ou de points qui ne jouent pas le même rôle pour la corde (milieu, seconde extrémité, couple de deux points), uniformément répartis ou non, sur des univers différents (intervalle de réels, figure géométrique donnée de dimension 1 ou 2, telle que circonférence, diamètre, disque...)

Le fait que plusieurs modèles soient possibles n'est alors plus paradoxal : le mathématicien est dans une situation analogue à celle du physicien qui modélise une expérience, et en ce sens, il choisit son hasard.

Cette remarque nous renvoie à l'utilisation de l'expression "au hasard", qui figure souvent dans les énoncés de problèmes que nous proposons à nos élèves. Implicitement, nous supposons ainsi qu'il y a répartition uniforme d'une variable dans une certaine population, et nous pensons volontiers que les choix pour la variable et l'univers sont uniques, alors que l'exemple précédent nous montre qu'il n'en est rien.

On pourrait peut-être penser que les moyens actuels de simulation d'expériences aléatoires peuvent nous aider à choisir une modélisation plutôt qu'une autre. Une analyse attentive des propositions faites auparavant pour les simulations montre qu'elles sont très liées à l'expérience choisie : elles en constituent une transcription fidèle. En effet, elles comportent le choix de nombres au hasard qui vont permettre de construire la quantité déterminant la corde, qui correspond au choix de la variable et de l'univers dans le modèle probabiliste. On ne peut pas pour l'expérience version 1, proposer la simulation de l'expérience version 2, ou inversement.

En ce sens, il n'y a pas à attendre de la simulation une "vérification" du modèle probabiliste.

Certains collègues, à qui nous avons proposé le problème précédent lors d'un stage, nous ont objecté que cet apparent paradoxe était lié à la nature continue du problème, et qu'il ne pouvait y avoir pour les activités proposées aux élèves dans le cadre des programmes de lycée des situations similaires, où l'emploi de "au hasard" puisse être interprété par des choix différents de variables et d'univers.

Et vous, êtes-vous du même avis ?

UN PROBLÈME BIEN ORDINAIRE

Pour réfléchir aux éventuels choix implicites présents dans les problèmes que nous proposons à nos élèves, nous avons choisi une situation banale, qui met en scène une classe fictive, où les élèves sont triés suivant deux caractères, par exemple les caractères fille / garçon et externe / demi - pensionnaire.

Cette situation, modulo un habillage différent peut être utilisée :

En première pour l'introduction de la notion de probabilité (passage de la proportion à la fréquence, puis aux probabilités) ou pour illustrer la propriété : $p(A \cup b) = p(A) + p(B) - p(A \cap B)$

En terminale pour introduire la notion de probabilité conditionnelle. Dès que deux caractères sont étudiés dans une population et qu'on se place dans une situation du calcul des probabilités, le cadre des probabilités conditionnelles peut être mobilisé.

Dans un premier temps, nous allons examiner un énoncé et la solution couramment attendue pour en dégager les choix implicites.

Dans un deuxième temps, nous verrons quels autres choix peuvent être faits et leurs conséquences en ce qui concerne la modélisation mathématique et les résultats aux questions posées.

1 UN ÉNONCÉ

Dans une classe, il y a 25 filles, 15 garçons. Parmi eux, 7 filles et 3 garçons sont externes, les autres sont demi-pensionnaires. Je choisis un élève de cette classe au hasard. Quelle est la probabilité que ce soit une fille demi-pensionnaire ?

2 ET LA SOLUTION ATTENDUE...

Il s'agit d'un tirage équiprobable d'un individu parmi 40. Il y a 18 filles demi-pensionnaires donc la probabilité cherchée est de $18/40$ soit $9/20$, ou 0,45.

Pour expliciter davantage pour les élèves, on peut ramener ce problème à celui d'un tirage dans une urne. Ce procédé, souvent utilisé, permet de repérer quelques structures-clés, dans lesquelles les différents problèmes rencontrés habituellement peuvent se classer.

L'énoncé en serait :

Une urne contient 40 jetons indiscernables au toucher, sur 18 d'entre eux est inscrite la mention "Fille demi-pensionnaire". On tire un jeton au hasard, quelle est la probabilité que ce soit un jeton marqué "Fille demi-pensionnaire" ?

Le fait de parler de jetons "indiscernables au toucher" est censé donner à l'élève un renseignement qui n'était pas évident dans l'énoncé initial : le modèle d'équiprobabilité semble plus explicite qu'il ne l'est dans l'énoncé de départ, qui mentionne seulement le choix de l'élève "au hasard".

Remarquons cependant que le décodage de ce type d'expressions reste affaire de conventions entre initiés ! Dans ce cas par exemple, on suppose évident que l'on tire en plongeant la main dans l'urne sans pouvoir voir les jetons (qui ne sont pas indiscernables à la vue, puisque certains ont une inscription, et d'autres pas).

La simulation peut être décrite par :

```
Initialisation
x représente le numéro d'un élève
n représente le nombre d'essais
f représente le nombre de succès

Tant que n est inférieur à 1000 faire
    n:=n+1
    choisir un nombre au hasard entre 1 et 40 : x
    Si x < 19 alors f:=f + 1
    fin tant que
afficher f/1000
```

Rien d'inattendu dans ce problème, cependant...

2 DEUX AUTRES SOLUTIONS

Pour choisir au hasard un élève dans la classe, je décide de faire mettre la classe en rang, les filles d'un côté, les garçons de l'autre, ou bien les externes d'un côté, les demi-pensionnaires de l'autre. Je choisis une file, puis un individu dans la file. Une fois dépassé le côté désuet d'une classe docile attendant le choix du maître, il est intéressant de s'occuper des modèles de probabilités à mobiliser.

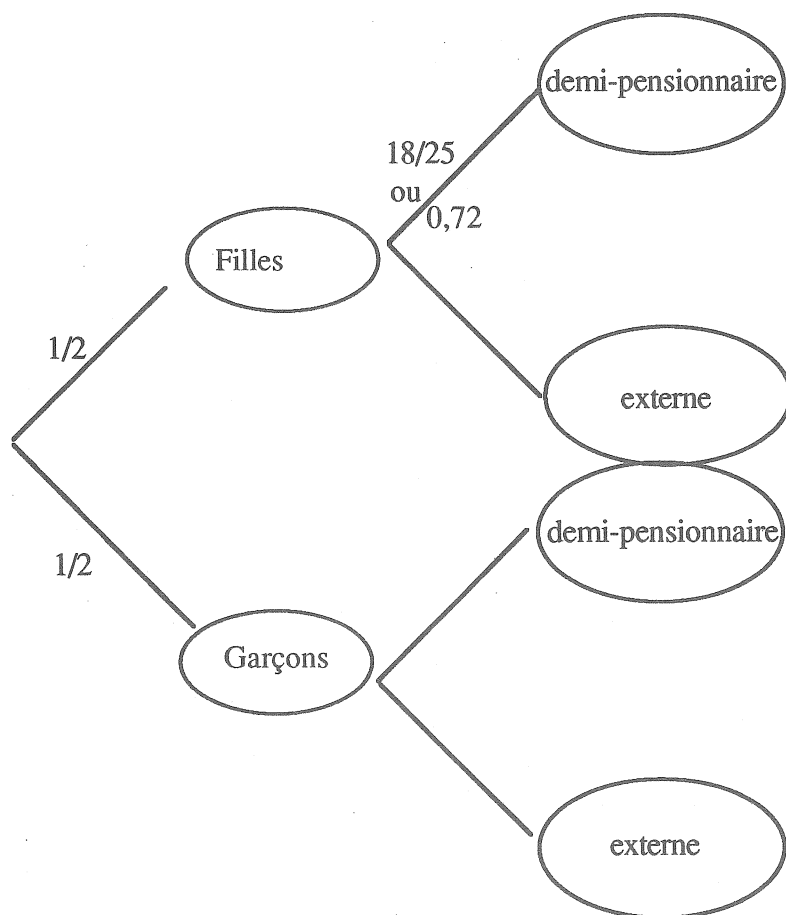
Une file de filles et une file de garçons :

Le problème peut se ramener à un problème d'urne, qui serait :

Deux urnes contiennent respectivement 15 jetons et 25 jetons. Dans la deuxième 18 jetons sont notés Fille demi-pensionnaire. Je tire au hasard une urne. Je prends au hasard un jeton dans l'urne choisie. Quelle est la probabilité que ce soit un jeton marqué Fille demi-pensionnaire ?

Notons que dans cet énoncé, nous supposons qu'on choisit de façon équiprobable entre les deux urnes, ce qui suppose que la rangée de filles et celle de garçons sont équiprobables. Là encore ces hypothèses ne font pas partie de l'énoncé initial, elles sont implicites.

On peut schématiser cette situation de choix à l'aide d'un arbre à deux niveaux, qui peut se présenter ainsi :



Dans cette procédure de choix, la probabilité cherchée s'obtient en multipliant 0,5 par 0,72, ce qui donne 0,36. Le fait qu'on multiplie les probabilités serait bien sûr à expliciter pour les élèves.

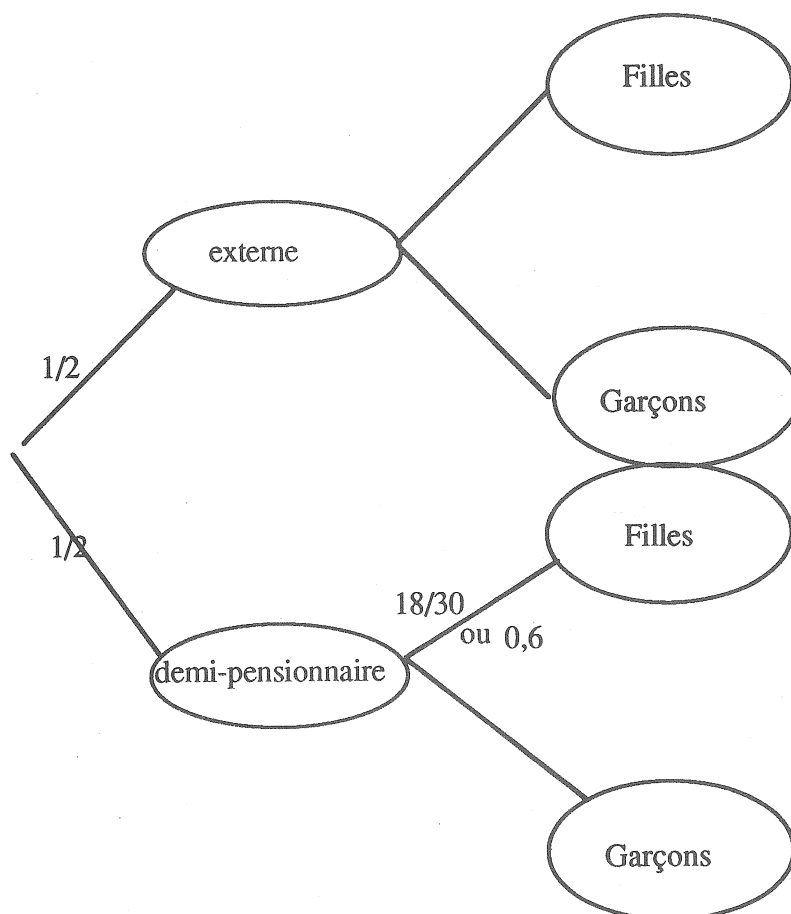
La simulation correspondante peut être réalisée de la façon suivante :

Initialisations
n représente le nombre d'essais
x représente le numéro de la file
y représente le numéro d'un élève
f représente le nombre de succès

tant que n est inférieur à 1000 faire
 n:=n+1
Choisir un entier x au hasard dans {1,2}
 Si x = 1 alors
 choisir un entier y entre 1 et 25 et si y<19 alors f:=f+1
 fin tant que
afficher f/1000

Une file d'externes, une file de demi-pensionnaires :

En termes d'urnes, le problème pourrait se formuler ainsi :
Deux urnes contiennent respectivement 10 jetons et 30 jetons. Dans la deuxième 18 jetons sont notés Fille demi-pensionnaire. Je tire au hasard une urne. Je prends au hasard un jeton dans l'urne choisie. Quelle est la probabilité que ce soit un jeton marqué Fille demi-pensionnaire ?



La probabilité ainsi trouvée est alors de $0,5 \times 0,6$, ou de $0,3$.
On obtiendrait une simulation de la même façon que dans le cas précédent.

3 QUELQUES COMMENTAIRES

Ces deux dernières façons de choisir les élèves peuvent paraître arbitraires, mais il est important de noter que dans chacune des trois solutions proposées, le modèle utilisé résulte d'un choix qui n'est pas explicitement décrit par l'énoncé. Certes, la première solution peut paraître plus naturelle, car faire mettre les élèves en rang, séparer les garçons des filles est éloigné des coutumes actuelles et peut faire rejeter ces deux solutions. Mais doit-on attendre des probabilités qu'elles modélisent les coutumes d'une époque donnée ?

Plus sérieusement, il semble que nous introduisons dans les problèmes, à l'aide de locutions comme "au hasard", une hypothèse d'équiprobabilité, sans que soit clairement désignée la population sur laquelle elle porte : faute d'une définition suffisamment détaillée des circonstances de l'expérience aléatoire, ces hypothèses ne sont pas spécifiées de façon explicite dans la plupart des énoncés proposés. L'usage courant, le bon sens, le contrat didactique font que sont rétablies presque toujours les hypothèses imaginées par l'auteur. Mais n'est-ce pas là une pratique surprenante en mathématiques, et peut-être une des causes de difficulté d'enseignement et d'apprentissage des probabilités ?

C'est en effet une des rares occasions où nous avons à nous occuper du passage entre le réel et les mathématiques, autrement dit où nous avons à construire une démarche de modélisation. Il faut admettre qu'une partie du travail mis en oeuvre pour résoudre un problème usuel de probabilités est en dehors du domaine mathématique appelé "Calcul des probabilités". Il serait donc vain d'attendre que les mathématiques donnent une réponse aux questions soulevées par cette démarche de modélisation, qui représente un saut entre deux mondes qu'il convient de ne pas confondre.

Le problème étudié auparavant n'est pas seul à soulever la question des divers choix possibles de modélisation pour un problème donné : on peut trouver un autre exemple dans "Enseigner les probabilités en classe de première" (IREM de Strasbourg, page 79).

Au fait, comment choisissez-vous, en début d'année le premier "porteur" de cahier de textes ? En "plantant" au hasard un stylo sur la liste, les noms placés aux extrémités ont-ils autant de chance d'être choisis que ceux du milieu de liste ? Et si c'est le premier redoublant par ordre alphabétique qui est choisi, est-il choisi au hasard ?

ARBRES ET PROBABILITÉS CONDITIONNELLES

Le problème précédent peut donner lieu à des questions telles que :

"On tire au hasard un élève, et il est extenué. Quelle est la probabilité pour que ce soit un garçon ?"

Une telle question est classiquement repérée comme une question de probabilités conditionnelles. Plus généralement, toute expérience aléatoire dont le résultat est un "individu" repéré suivant deux caractères est un problème qui peut rentrer dans le cadre des probabilités conditionnelles. Dans un tel problème, un individu peut posséder ou non la propriété A, il peut posséder ou non la propriété B, et la structure d'arbre à deux niveaux de choix est un outil commode, dont l'utilisation a déjà fait l'objet de propositions et d'analyses (voir la bibliographie).

La première partie présente une étude très technique de l'outil arbre, dont la maîtrise permet, nous semble-t-il, de dégager une structure commune aux problèmes de probabilités conditionnelles et d'unifier ainsi leur traitement.

La seconde partie propose des situations utilisables en classe, où on peut mettre en oeuvre des arbres pour la résolution.

La troisième partie propose quelques éléments de réflexion sur la question de savoir s'il faut enseigner cet outil, et comment, question qui mérite d'être examinée soigneusement, et mise en débat.

1 UN DÉTOUR PAR L'OUTIL "ARBRE"

Nous avons choisi dans cette partie une présentation purement formelle de l'outil arbre, sans faire référence à des exemples d'utilisations.

Nous tenons à insister sur le fait que ce qui suit n'est donc pas directement à l'usage de l'élève, et n'a pas à être enseigné tel quel. Il y aurait à le faire un grave risque de perte de sens pour l'élève, et de glissement métacognitif, qui consisterait à remplacer l'enseignement des probabilités conditionnelles par l'enseignement d'algorithmes de résolution, ce qui serait en contradiction complète avec notre propos.

En reprenant les notations introduites plus haut, on peut considérer d'abord la réalisation de A, puis celle de B, ou l'inverse.

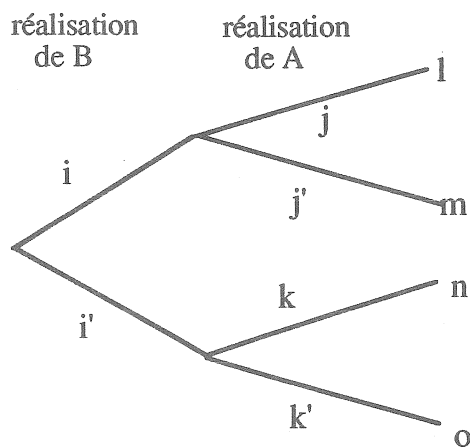
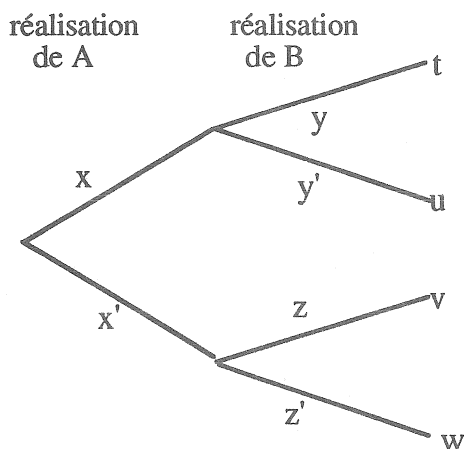
Convenons que les branches du haut signifient la réalisation de l'événement indiqué en tête de colonne.

On peut schématiser alors la situation à l'aide de l'un ou l'autre des deux arbres ci-dessous, et retrouver les nombres qui apparaissent habituellement dans les problèmes de probabilités conditionnelles.

Nous avons au départ :

$$x = p(A), \quad x' = p(\overline{A}),$$

$$y = p_A(B), \quad y' = p_A(\overline{B}), \quad z = p_{\overline{A}}(B), \quad z' = p_{\overline{A}}(\overline{B})$$



• Les nombres qu'on peut lire directement et les relations immédiates qui les relient.

Pour le premier arbre, nous avons au départ :

$$x = p(A), x' = p(\bar{A}),$$

$$y = p_A(B), y' = p_A(\bar{B}), z = p_{\bar{A}}(B), z' = p_{\bar{A}}(\bar{B})$$

$$t = p(A \cap B), u = p(A \cap \bar{B}), v = p(\bar{A} \cap B), w = p(\bar{A} \cap \bar{B}).$$

On peut immédiatement écrire : $x' = 1 - x, y' = 1 - y, z' = 1 - z$.

Les règles de multiplication des probabilités sur une même branche donnent alors :

$$xy = t, xy' = u, x'z = v, x'z' = w,$$

$$\text{On en déduit : } t + u = x, v + w = x', t + u + v + w = 1.$$

On a des relations analogues entre les nombres du deuxième arbre.

• A l'intérieur de chaque arbre, les renseignements qu'on peut obtenir par un calcul simple.

Pour le premier arbre toujours :

- On lit directement la probabilité de A, mais on peut avoir très facilement celle de B : il suffit d'ajouter les probabilités de $A \cap B$ et de $\bar{A} \cap B$, c'est à dire les nombres t et v.

$$\text{Donc } p(B) = t + v = xy + (1-x)z$$

$$\text{De même, } p(\bar{B}) = u + w$$

- On peut résoudre facilement la question de l'indépendance des événements A et B, en partant de la définition $p_A(B) = p(B)$, on aura :

$$A \text{ et } B \text{ sont indépendants} \Leftrightarrow p_A(B) = p(B)$$

$$\Leftrightarrow y = xy + (1-x)z$$

$$\Leftrightarrow (1-x)y = (1-x)z$$

$$\Leftrightarrow y = z, \text{ (on peut en effet supposer que } x \text{ est différent de } 1,$$

sinon A étant certain, on ne se trouve plus dans le cadre du tri suivant deux caractères)

Le résultat obtenu est : "A et B sont indépendants lorsque la probabilité d'avoir le caractère B est la même que l'on considère les individus présentant A ou $\bar{A}$ ".

C'est d'ailleurs une définition possible de l'indépendance en probabilité.

On a bien sûr les mêmes propriétés pour les nombres du deuxième arbre.

• Dans un arbre, quels nombres suffisent à déterminer tous les autres ?

Plaçons nous toujours dans le premier arbre. En prenant toujours le parti de négliger le sens de la représentation utilisée, l'examen très technique des relations algébriques précédentes nous permet les observations suivantes :

- Le fait de connaître x, y, z , permet de déterminer x', y', z' , comme leurs compléments à 1, et permet de calculer les nombres t, u, v, w , comme produits des nombres situés sur les parcours y aboutissant. La situation est alors entièrement connue.

- Si l'on connaît x, t , et z , la relation $xy = t$ permettra de calculer y , et on est alors ramené au cas précédent. De même si x, y , et v sont connus, ou si x, t , et v sont connus, ou x, u , et v , le sont,...etc..

- Dans tous les choix précédents, on peut remplacer la connaissance d'un nombre quelconque par la connaissance de son complémentaire à 1, par exemple connaître x', t et v suffit à connaître tous les autres nombres.

- La connaissance de trois des quatre nombres t, u, v, w , permet également de retrouver tous les autres. Par exemple, si l'on connaît t, u, v , on aura :
 $w = 1 - (t + u + v)$ puis $t + u = xy + xy' = x(y+y') = x$, puis $x' = 1 - x$.
On pourra alors calculer y et y' ($y = t / x, y' = u / x$), puis z et z' ($z = v / x' et z' = w / x'$).

Il en est de même pour les nombres du deuxième arbre.

• Les relations entre les deux arbres

- Le fait que A et B soient indépendants (ou non) peut se lire sur la premier arbre tout comme sur le second, donc : $y = z \Leftrightarrow j = k$.

- On peut calculer les nombres du second arbre à l'aide de ceux du premier : remarquons d'abord que les événements au bout de chacune des quatre branches des deux arbres sont les mêmes, mais dans un ordre différent, leurs probabilités se retrouvent, donc :

$$l = t, m = v, n = u, o = w.$$

Comme on l'a déjà vu, la connaissance de trois des quatre nombres au bout des branches permet d'obtenir tous les autres. Pour mémoire, on aura :

$$i = l + m, i' = 1 - i, j = l / i, j' = m / i, k = n / i', k' = o / i'.$$

On peut donc très facilement passer du premier arbre au second, et inversement, convenons d'appeler cette opération "inversion d'arbre".

2 UTILISATION DANS LES PROBLÈMES

Voici plusieurs problèmes, pour lesquels nous proposons l'utilisation de l'outil étudié auparavant. Pour chacun, nous avons choisi d'effectuer le traitement de la même manière, que nous essayons de décrire d'abord :

- On choisit de schématiser la situation probabiliste simultanément par les deux arbres pouvant la décrire, en y indiquant directement les nombres donnés dans l'énoncé, et en gardant pour les nombres non connus d'emblée les lettres employées dans la partie précédente.

- Les événements concernés sont indiqués en haut de chaque colonne des arbres, et les branches du haut correspondent toujours à la réalisation de l'événement indiqué en tête de colonne.

- On montre que les questions du problème peuvent être traduites par la recherche de certains nombres des arbres.

- En utilisant les méthodes exposées dans la première partie, on fait le calcul de ces nombres en indiquant les diverses étapes et le remplissage progressif des arbres.

- On traduit ensuite les résultats précédents dans les termes du problème.

Comme le lecteur l'aura sans doute remarqué, certaines de ces règles ne sont que des conventions commodes, d'autres au contraire relèvent de choix délibérés : par exemple le fait d'appliquer les techniques exposées dans la première partie nous oblige à certains moments à ne plus nous préoccuper du sens et à privilégier l'aspect progression dans les calculs.

Comme pour la première partie, les résolutions proposées ici sont donc destinées à des lecteurs enseignants, et si les problèmes proposés peuvent être utilisés en classe, leur résolution par contre ne saurait être présentée de cette manière pour des élèves.

PROBLÈME 1

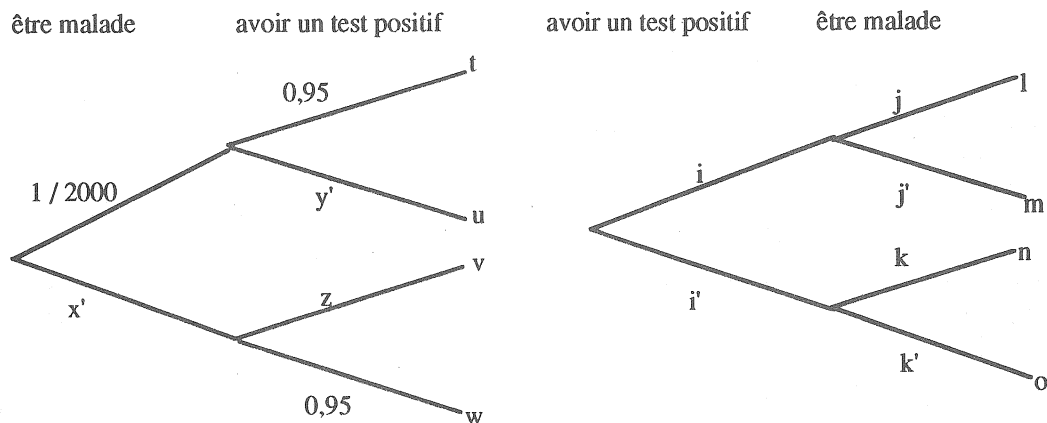
Dans une population, une personne sur 2000 est atteinte de la maladie X. Un test donne un résultat correct de dépistage - c'est-à-dire positif si la personne est malade, négatif si elle ne l'est pas - avec une probabilité de 0,95 dans chacun des deux cas évoqués.

- *On choisit une personne au hasard, et elle subit le test. Le résultat est positif. Quelle est la probabilité pour que cette personne soit atteinte de la maladie X ?*
- *On choisit une personne au hasard, et elle subit le test. Le résultat est négatif. Quelle est la probabilité pour que cette personne n'ait pas la maladie X ?*

Pour une personne choisie au hasard, deux événements sont en jeu :

- Être malade
- Présenter un test positif

Les deux arbres correspondants nous permettent de placer les trois nombres donnés dans l'énoncé



Les questions posées :

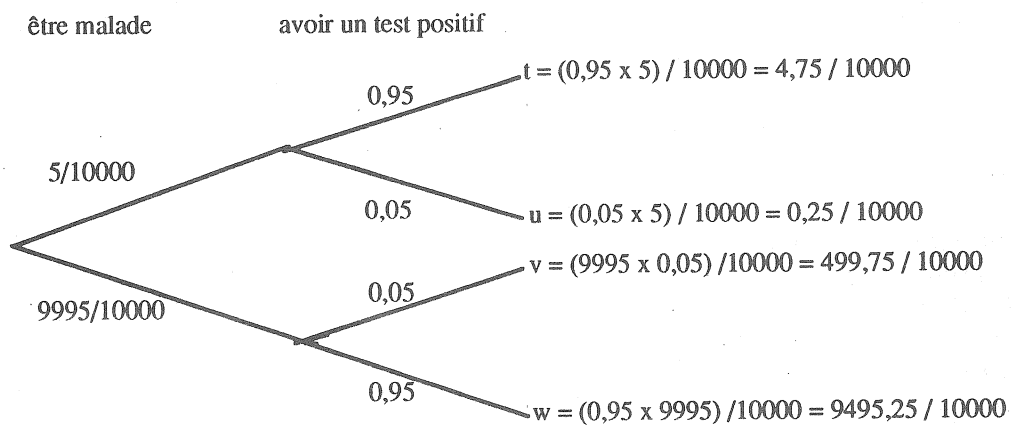
Il faut déterminer la probabilité, pour une personne choisie au hasard :

- d'être malade sachant que le test est positif, c'est à dire il faut déterminer j .
- d'être non malade sachant que le test est négatif, c'est à dire il faut déterminer k' .

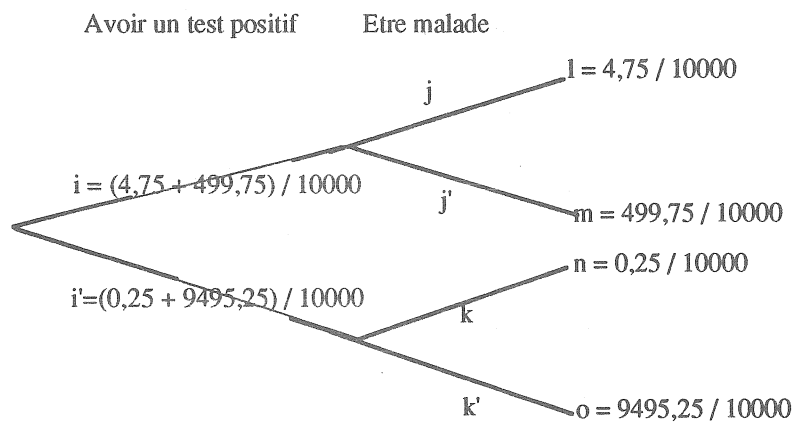
Calculs :

Les trois nombres connus dans le premier arbre nous assurent qu'on peut calculer tous les autres. Les nombres à trouver sont dans le second arbre, il s'agira donc d'inverser le premier arbre.

En suivant la stratégie indiquée dans la partie précédente, et en effectuant les calculs en dix millièmes et de façon exacte, on a d'abord :



On connaît alors les nombres qui sont au bout de l'arbre inversé, et on peut obtenir les nombres de la première colonne par addition :



On obtient alors : $j = l/i = 4,75 / (4,75 + 499,75) = 0,009415...$
 Ensuite, on obtient $k' = o/i' = 9495,25 / (9495,25 + 0,25) = 0,999973...$

Solution

On voit donc que la probabilité d'être malade sachant que le test est positif est inférieure à 1%, ce qui est très surprenant.

Alors que la probabilité d'être non malade sachant que le test est négatif est supérieure à 99,99%, ce qui est bien plus rassurant !

PROBLÈME 2

Un appareil fabriqué en très grande série peut être défectueux à cause de deux défauts seulement, désignés par A et B.

Dans un lot de 10 000 appareils prélevés on a constaté que :

- 1000 appareils présentaient le défaut A
- 800 appareils présentaient le défaut B
- 400 appareils présentaient à la fois les défauts A et B.

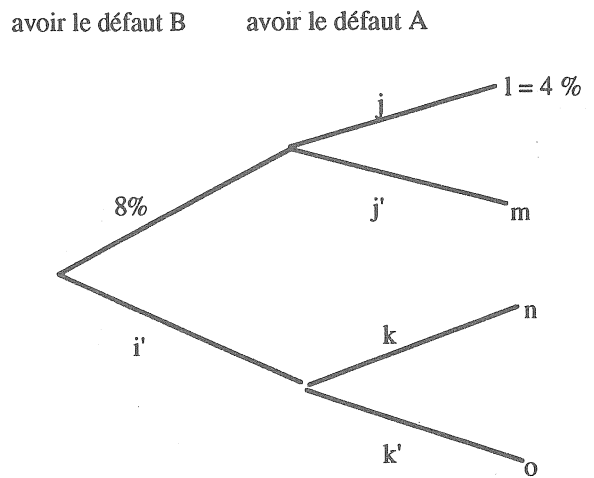
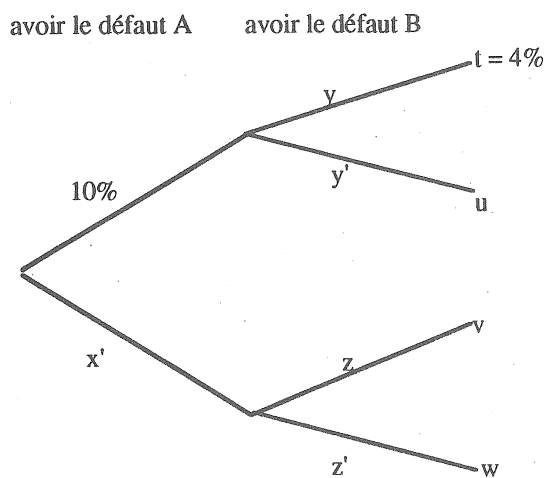
Un client achète l'un des appareils.

Quelle est la probabilité pour qu'il ne présente aucun défaut ? Le défaut A seulement ? le défaut B seulement ? Les événements "présenter le défaut A" et "présenter le défaut B" sont-ils indépendants ?

Les événements qui sont en jeu pour le choix de l'appareil sont "présenter le défaut A" et "présenter le défaut B".

On peut traduire les données en termes de probabilités, que l'on place dans les deux arbres :

- la probabilité pour l'appareil choisi de l'évènement "avoir le défaut A" est de $1000 / 10000$, ou 10 %,
- celle de "avoir le défaut B" est de $800 / 10000$, ou 8 %,
- celle d'avoir à la fois A et B est de $400 / 10000$, ou 4 %.

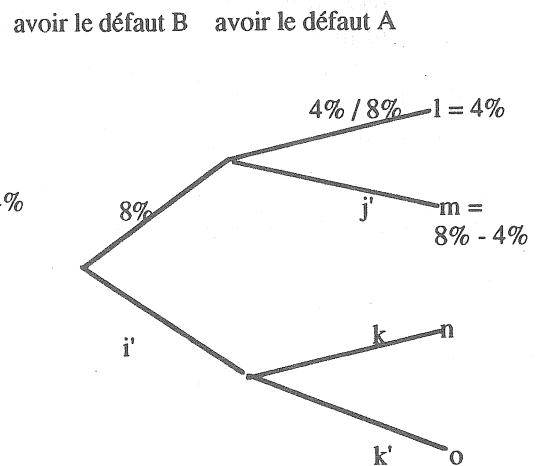
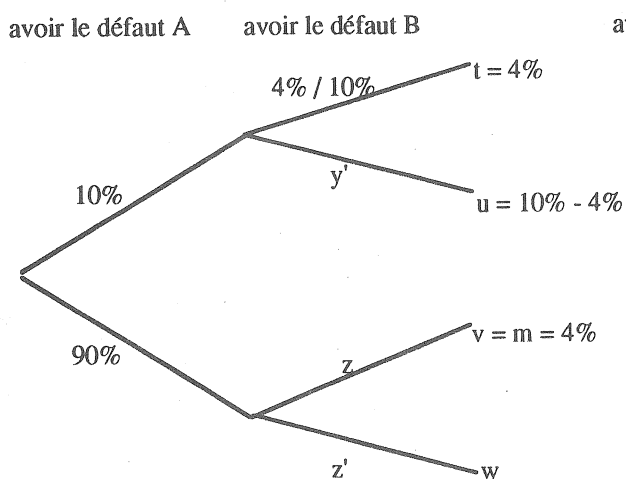


Les questions posées :

Pour l'appareil choisi, chercher la probabilité de n'avoir aucun défaut, c'est chercher w, ou o, qui sont les mêmes ;
chercher la probabilité de n'avoir que le défaut A, c'est chercher u, ou n, qui sont égaux ;
chercher la probabilité de n'avoir que le défaut B, c'est chercher v ou m, qui sont égaux ;
chercher si les événements "présenter le défaut A" et "présenter le défaut B" sont indépendants, c'est chercher si y et z sont égaux. (ou si j et k le sont.)

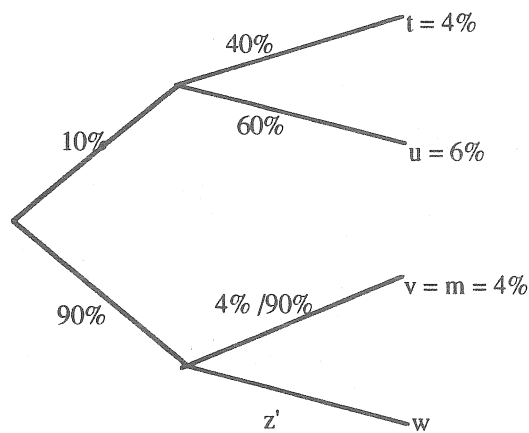
Calculs :

Aucun des deux arbres ne peut suffire à lui seul à trouver les nombres manquants, il faut traiter les deux simultanément :
x' et i' sont obtenus de façon immédiate, ainsi que u et m. Puis n et v, qui leur sont respectivement égaux ; on a alors trois des quatre nombres du bout des quatre branches de chaque arbre, on peut trouver tous les manquants.
Voici les principales étapes de ces calculs :

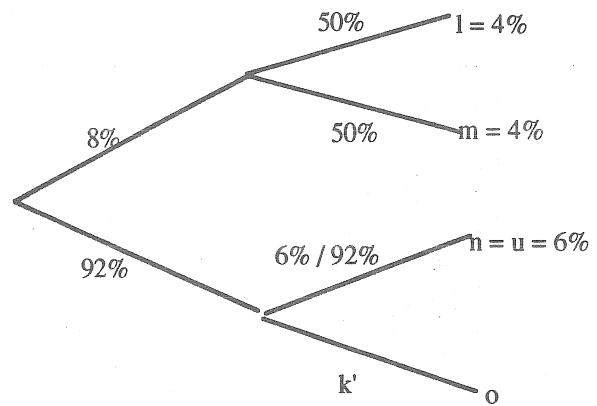


Puis :

avoir le défaut A avoir le défaut B



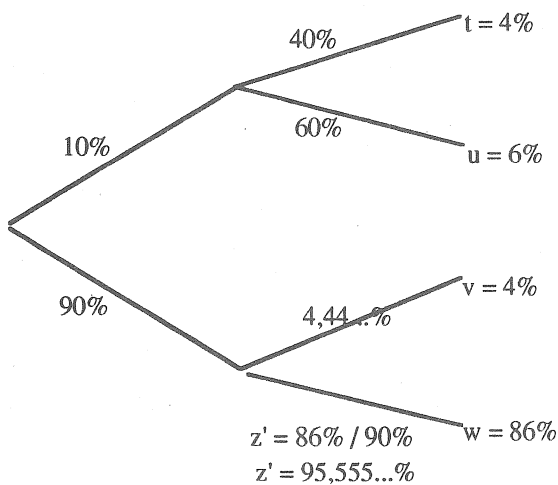
avoir le défaut B avoir le défaut A



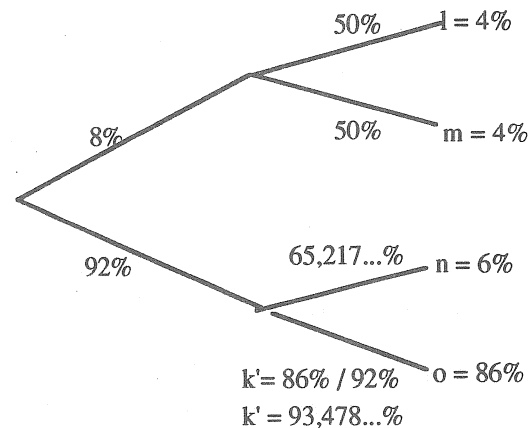
Puis w et o s'obtiennent en complétant à 1 la somme des trois nombres du bout de chaque arbre : $w = o = 1 - (4\% + 6\% + 4\%) = 86\%$.

On obtient z' et k' en dernier lieu :

avoir le défaut A avoir le défaut B



avoir le défaut B avoir le défaut A



Solutions :

La probabilité pour que l'appareil ne présente aucun défaut est donc de 86%.
La probabilité pour que l'appareil ne présente que le défaut A est donc de 6%.
La probabilité pour que l'appareil ne présente que le défaut B est donc de 4%.

Comme y et z ne sont pas égaux, les événements "avoir le défaut A" et "avoir le défaut B" ne sont pas indépendants : on constate que les appareils qui ont déjà le défaut A ont une probabilité d'avoir aussi le défaut B bien plus forte que les appareils n'ayant pas le défaut A.

PROBLÈME 3

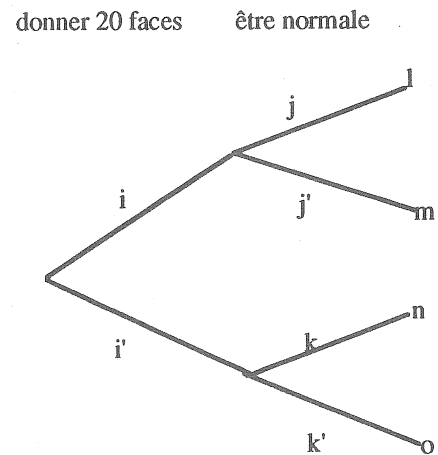
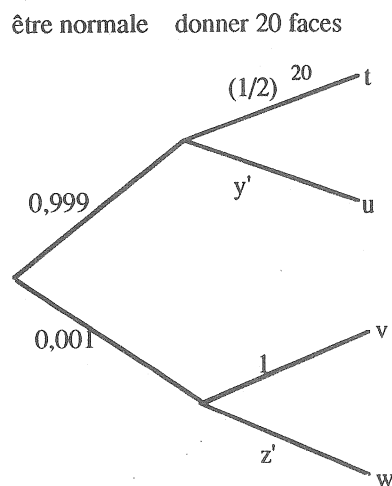
Une pièce sur 1000 présente deux "face" au lieu de "pile et face". On choisit une pièce au hasard, et on la lance 20 fois. On obtient 20 "face". Quelle est la probabilité pour que la pièce choisie soit normale ?

Pour la pièce qu'on choisit, deux événements sont en jeu : "être normale" et "produire une suite de 20 faces lors de 20 lancers successifs".

La probabilité pour la pièce de n'être pas normale est de $1/1000$, et celle d'être normale est de $999/1000$.

Si la pièce est normale, la probabilité d'obtenir 20 faces est de $(1/2)^{20}$.

Mais si la pièce est truquée et possède deux faces, la probabilité d'obtenir 20 faces lors de 20 lancers est de 1. Schématisons la situation par les deux arbres :



La question posée :

On cherche la probabilité pour que la pièce soit normale, sachant qu'elle a produit 20 faces lors des 20 lancers, c'est à dire qu'on cherche j.

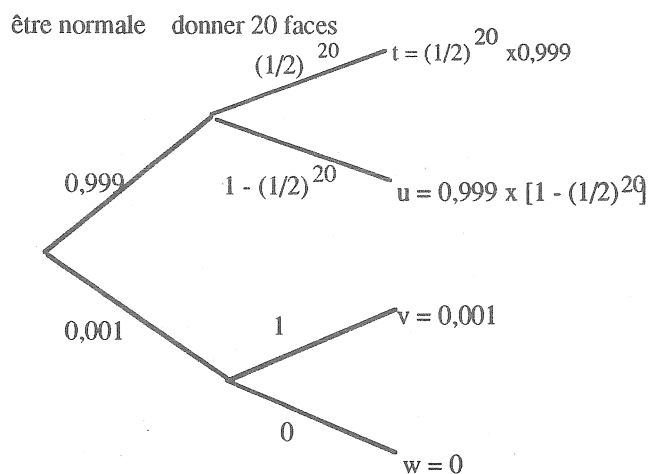
Calculs:

On peut d'abord remarquer qu'on connaît suffisamment de nombres dans le premier arbre pour tous les déterminer, le problème que nous aurons à résoudre sera d'inverser cet arbre pour avoir j.

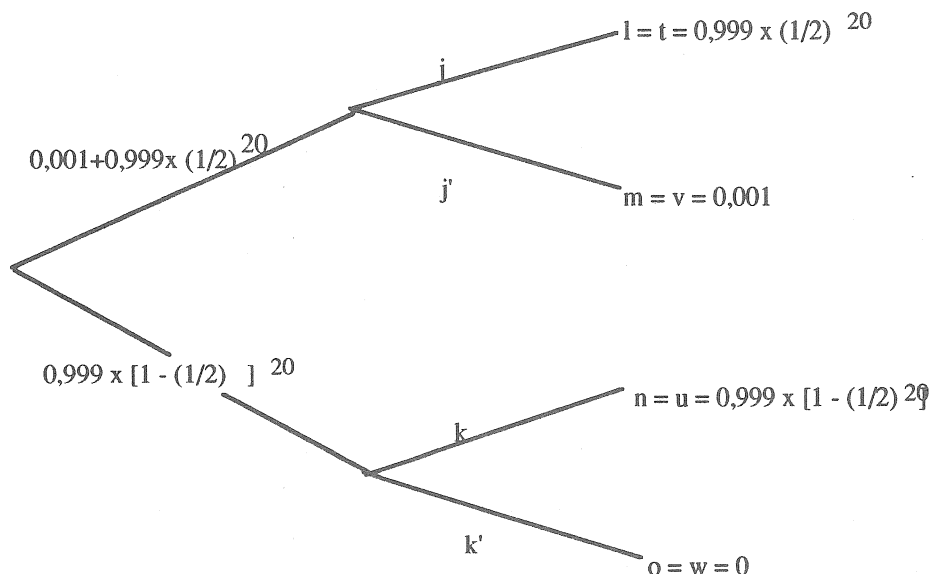
Dans le premier arbre, on peut trouver immédiatement y' et z', on obtient :

$y' = 1 - (1/2)^{20}$, et $z' = 0$. Pour préserver l'uniformité de traitement des problèmes de ce paragraphe, on peut choisir de conserver la branche sur laquelle est 0 dans la suite, (mais c'est une branche morte, correspondant à l'événement impossible, on pourrait ne pas la dessiner.)

Les nombres du bout des branches du premier arbre sont alors obtenus par les produits des nombres des branches y conduisant :



D'où en reportant dans un ordre adéquat les nombres situés au bout des branches, on a :



On aura alors : $j = l / i = [0,999 \times 0,5^{20}] / [0,999 \times 0,5^{20} + 0,001] = 0,0009518...$

Solution :

Avec 20 faces obtenus sur 20 lancers, la probabilité pour que la pièce soit normale est très faible, de moins d'un millièrme.

Cependant, avec la même proportion de pièces truquées, si l'on obtient n faces sur n

lancers, une étude analogue à la précédente donne $j = \frac{0,999 \times 0,5^n}{0,999 \times 0,5^n + 0,001}$.

Si $n = 5$, on a obtenu 5 faces sur 5 lancers et la probabilité pour qu'alors la pièce soit normale est de 97% environ.

Si $n = 10$, on a obtenu 10 faces sur 10 lancers et la probabilité pour qu'alors la pièce soit normale est de 49% environ.

Si $n = 15$, on a obtenu 15 faces sur 15 lancers et la probabilité pour qu'alors la pièce soit normale est de 3% environ.

Il faut dépasser 17 lancers pour que la probabilité d'avoir une pièce normale dans ces conditions chute à moins de 1%.

PROBLÈME 4

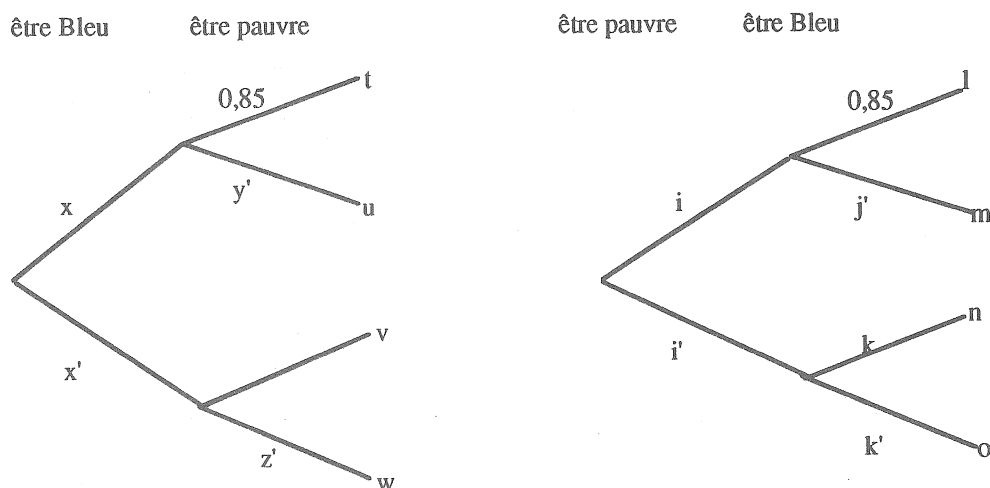
Dans une lointaine planète, vivent des Verts et des Bleus : 85 % des Bleus sont pauvres, et 85 % des pauvres sont Bleus.

On choisit un individu au hasard. Les événements "obtenir un individu bleu" et "obtenir un individu pauvre" sont-ils indépendants ? Y a-t-il sur cette planète une inégalité sociale due à la couleur ?

85% des Bleus sont pauvres, on sait donc que la probabilité pour un individu d'être pauvre sachant qu'il est bleu est de 85%, et $y = 0,85$.

85% des pauvres sont Bleus, donc la probabilité pour un individu d'être bleu sachant qu'il est pauvre est de 85%, et $j = 0,85$.

Les deux arbres correspondants à cette situation sont :



Les questions posées :

La question de savoir si les événements considérés sont ou non indépendants revient à savoir si y et z sont égaux (ou si j et k le sont).

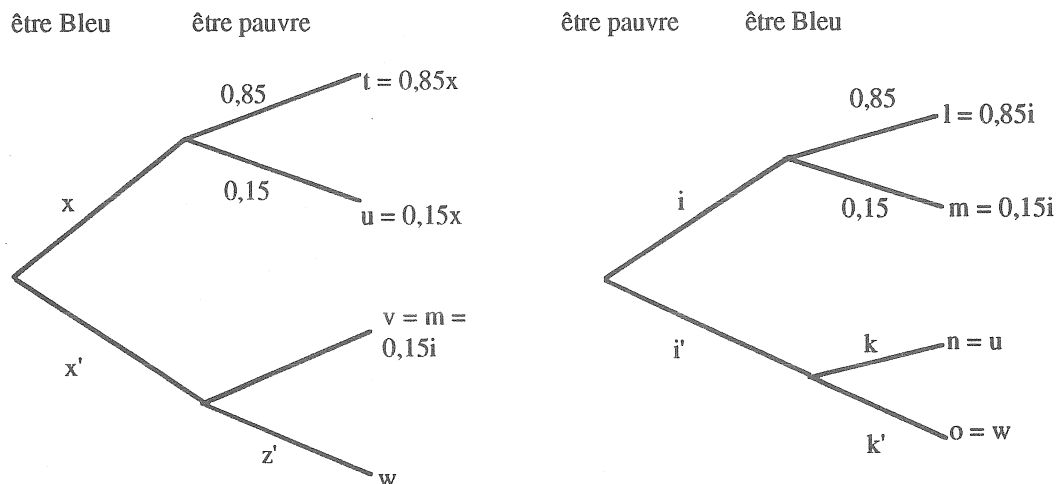
Pour répondre à la question concernant l'inégalité sociale éventuelle entre les Bleus et les Verts, il faut définir un critère : le mieux adapté semble être de comparer la probabilité d'être pauvre chez les Bleus et chez les Verts, ce qui revient à comparer y et z . La réponse est donc la même que pour la question précédente.

Calculs :

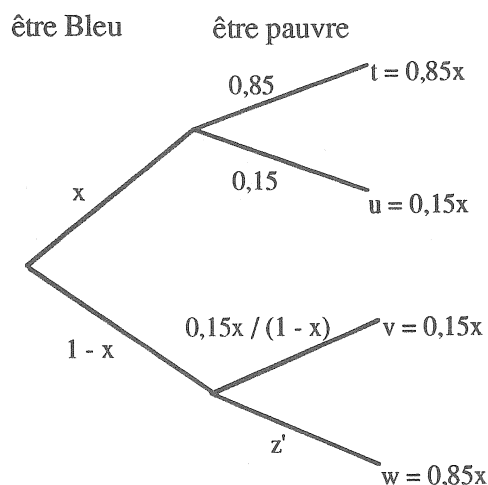
Seuls peuvent être complétés y' et j' dans les deux arbres précédents. On ne peut obtenir aucun autre des nombres qui sont sur les autres branches, ce qui nous montre que l'énoncé est incomplet.

On ne peut donc pas répondre aux questions posées, les données sont insuffisantes.

Ce qu'il faudrait rajouter, c'est par exemple la répartition des Bleus et des Verts dans cette lointaine planète. Suivant cette répartition, on peut avoir des résultats différents aux questions posées : on peut, pour aller un peu plus loin, conduire les calculs en fonction de x , probabilité d'être Bleu pour un individu choisi au hasard. On obtient :



Comme $t = l$, on aura $0,85 x = 0,85 i$, donc $i = x$.
On peut alors compléter le premier arbre par exemple, on aura :



Il faut alors comparer $0,85$ à $0,15x / (1 - x)$.

La probabilité d'être pauvre pour un Bleu est supérieure à celle d'être pauvre pour un Vert dès que $0,85$ dépasse $0,15x / (1 - x)$. On a :

$$0,85 \geq \frac{0,15x}{1-x} \Leftrightarrow 0,85(1-x) \geq 0,15x \Leftrightarrow 0,85 \geq x \Leftrightarrow x \leq 0,85$$

Si par exemple on a 85% de Bleus, la probabilité d'être pauvre pour un Vert est de 0,85 et il y a égalité sociale. Mais si il y a 86% de Bleus, la probabilité pour un Vert d'être pauvre est de 92,1...%, et dans ce cas, il y a une inégalité sociale, mais pas dans le sens où l'énoncé pouvait nous le laisser penser : mieux vaut alors être Bleu que Vert !

3 L'ENSEIGNEMENT DES PROBABILITÉS CONDITIONNELLES

Dans les problèmes examinés auparavant, l'utilisation conjointe des deux arbres de probabilités apparaît comme un outil qui peut permettre un traitement quasi-automatique de ce type de problèmes. Il s'agit à partir de quelques nombres de déterminer tous les autres, et le plus souvent, la démarche employée est de compléter l'un des arbres pour l'inverser et obtenir l'autre.

Comme pour tout outil de type "instrumental", ou toute technique, il y a des dangers à l'utiliser de manière systématique dans l'enseignement : les automatismes risquent de se substituer au sens, et l'enseignement de l'outil risque de remplacer l'enseignement du concept. C'est pourquoi il ne faudrait pas considérer que ce qui précède est destiné à être communiqué aux élèves tel quel, ce que nous avons déjà dit plusieurs fois.

On peut tenter d'analyser les difficultés du concept de probabilités conditionnelles, pour en tirer ensuite quelques conséquences pour l'enseignement. Des travaux en ce sens existent déjà (voir bibliographie).

Rappelons simplement que Laplace écrit dans le sixième principe à propos de probabilités conditionnelles : "C'est le principe fondamental de cette branche de l'Analyse des hasards, qui consiste à remonter des événements aux causes."

Historiquement, la démarche fondatrice du concept de probabilités conditionnelles a été en effet construite pour résoudre un champ de problèmes où il s'agissait d'inférer les probabilités des causes, ayant observé celles de leurs conséquences, ce qui oblige à revenir en pensée en arrière dans le temps.

Nous voyons ici apparaître deux obstacles qui sont d'une part la causalité, d'autre part la temporalité, et qui sont constitutifs du concept lui-même. En ce sens, nous pouvons parler d'obstacles épistémologiques, au sens où A. Sierpiska les définit (Revue recherche en didactique des mathématiques, 6 - 1) :

"Pour notre part, nous retiendrons deux aspects de la notion d'obstacle épistémologique selon G. Bachelard :

- l'apparition des obstacles a un caractère inévitable...
- la répétition de leur apparition dans la philogénèse et l'ontogénèse des concepts..

Dans la conception qui va être utilisée ici de la notion d'obstacle, c'est à ces deux aspects que nous faisons référence en disant que, pour un concept mathématique déterminé, il est possible de mettre en évidence des causes de lenteurs et de troubles dans l'acquisition du concept qui :

- 1) sont spécifiques de ce concept et de lui seul
- 2) sont telles que leur prise de conscience est indispensable pour le développement de ce concept."

Analysons de ce point de vue le problème 1 :

Les renseignements donnés au départ par l'énoncé pourraient l'être par un responsable de la santé publique, ce qui est décrit l'est dans l'ordre où une telle personne voit les choses advenir : une nouvelle maladie survient, se développe dans des proportions que nos sociétés ne peuvent considérer comme négligeables ; la recherche médicale tente de trouver des traitements, et met au point des tests de dépistage qui ne sont jamais fiables à 100%, mais dont on connaît statistiquement l'efficacité. De ce point de vue, l'ordre maladie - test de dépistage suit la logique du déroulement du temps.

Mais la question posée concerne un individu, qui ne sait pas encore s'il est malade, et qui va d'abord passer le test pour en attendre ensuite la verdict. Pour résoudre le problème, l'élève doit donc parvenir à comprendre qu'il est possible d'inverser l'ordre des

renseignements donnés par l'énoncé, c'est cette procédure qui va lui donner la solution. Notons par ailleurs que ce problème n'est pas neutre affectivement pour l'élève, qui, en tant qu'individu risquant d'être confronté à cette situation, s'identifie plutôt à la personne qui passe le test. Les résultats numériques paraissent alors à peine crédibles, seul le point de vue de ce qui se passe globalement pour la population permet de comprendre le phénomène. La première fois où j'ai proposé ce travail à des élèves, je n'avais posé que la première question : la faible probabilité d'être malade sachant que le test est positif a amené plusieurs de ces élèves à se poser par eux-mêmes la seconde question, pour laquelle ils soupçonnaient la possibilité d'une réponse tout aussi déconcertante.

C'est la comparaison de ces deux résultats qui amène à comprendre le phénomène : les taux de fiabilité des tests sont les mêmes dans les deux cas, mais les conséquences sont différentes à cause de la taille des populations sur lesquelles ils agissent, les personnes non malades étant beaucoup plus nombreuses que les personnes malades.

La compréhension au-delà des calculs ne peut se produire que si l'on dépasse le point de vue de l'individu, alors que c'est justement du point de vue de l'individu que la question est posée. L'individu qu'est l'élève se met en situation d'avoir un test positif, et tente d'évaluer une probabilité d'un point de vue subjectif, qui n'est évidemment pas celle du probabiliste.

De façon plus générale, le passage pour l'élève de la probabilité de "A sachant B" à celle de "B sachant A" demande que B ne soit pas considéré d'emblée comme ayant déjà été réalisé. Il y a donc à accepter que la locution "sachant que..." ne signifie pas que l'événement cité est réalisé, ce qui est en rupture avec l'emploi habituel de la langue. Une représentation erronée des situations de ce type est donc à craindre dans l'enseignement. La situation peut constituer un obstacle encore plus fort pour le concept de probabilités conditionnelles s'il y a entre A et B une relation sémantique du type "A a pour conséquence B" ou "la réalisation de A précède celle de B". Dans ce cas, la réalisation de "B sachant A" est confondue avec celle de B et celle de $A \cap B$, d'où les confusions fréquentes entre les nombres $p_A(B)$ et $p(A \cap B)$. Dans l'exemple 1, pour l'élève en tant qu'individu, si l'on se réfère à notre environnement sociologique, le test de dépistage précède la révélation de la maladie, qui peut, même si chacun sait qu'elle préexiste, apparaître comme postérieure au test positif.

Nous pensons que le côté technique de l'outil arbre permet de dépasser ces contradictions.

En effet, les événements qui interviennent sont traités indépendamment de leur sémantique. La possibilité d'inverser un arbre quels que soient les événements en jeu est mise en évidence dès lors qu'on passe un peu de temps en classe pour repérer sur les branches les nombres $p_A(B)$ et $p(A \cap B)$, et pour analyser les différentes étapes du traitement de cet outil.

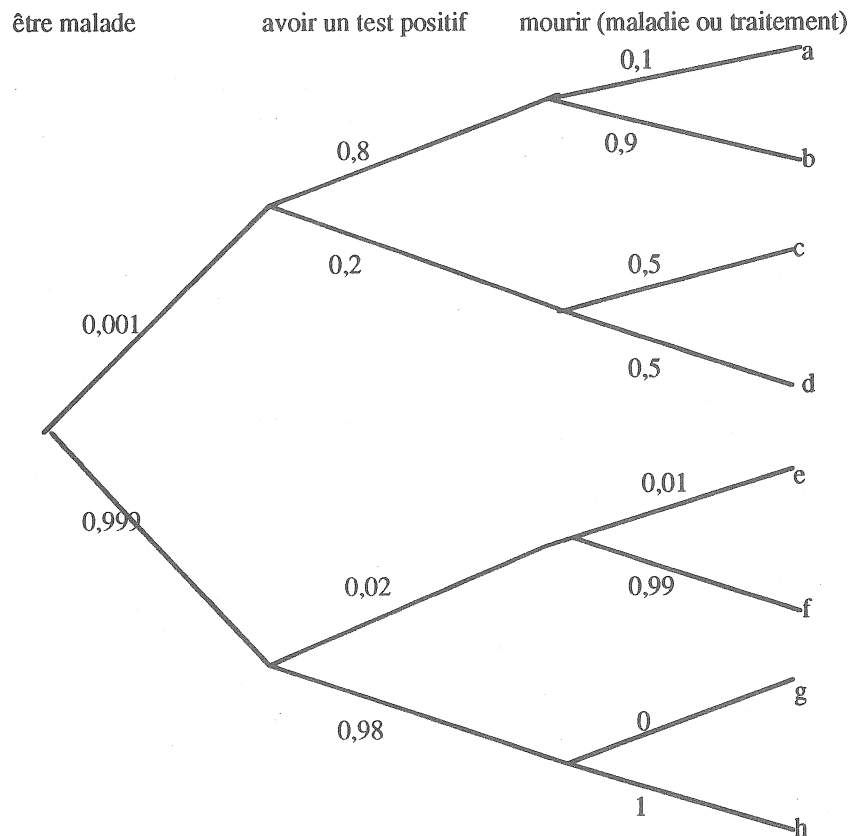
Prenons un dernier exemple, où trois événements peuvent se produire ou non pour un individu. Voici cet énoncé, dont le traitement va faire intervenir un arbre à trois niveaux, que nous envisagerons d'inverser :

Dans la population de Papirémie, sévit une mystérieuse maladie, appelée toxoprobamathose. La probabilité, pour un papirémien, d'en être atteint est de 0,001. La toxoprobamathose est mortelle, avec une probabilité de 0,5 pour tout individu atteint et non traité, et avec une probabilité de 0,1 pour tout individu atteint, dépisté et traité. Le seul test de dépistage disponible permet de détecter 80 % des personnes atteintes, mais désigne aussi à tort comme malades 2 % des personnes non atteintes. Le traitement utilisé est mortel pour 1 % des personnes traitées à tort.

Si on décide un dépistage et un traitement systématiques des personnes ainsi désignées, quelle est pour un papirémien la probabilité de mourir de toxoprobamathose ou de son traitement ?

Les événements en jeu pour un individu sont au nombre de trois, dont la réalisation a lieu ou non dans l'ordre suivant :
être malade, avoir un test positif (donc être traité), et mourir de la maladie ou de son traitement.

La représentation en arbre conserve cet ordre, on a en gardant la convention que les branches du haut correspondent à la réalisation de l'évènement indiqué en tête de colonne:



Pour répondre à la question posée, il faut chercher dans les nombres du bout de l'arbre tous ceux qui correspondent à la réalisation de l'évènement "mourir de la maladie ou de son traitement", c'est à dire les nombres a, c, e, et g, et les ajouter. Comme ces nombres sont les produits de ceux qui sont sur les branches qui y aboutissent, on obtient :

$$a + c + e + g = 0,001 \times 0,8 \times 0,1 + 0,001 \times 0,2 \times 0,5 + 0,999 \times 0,02 \times 0,01 + 0,999 \times 0,98 \times 0 = 0,00038...$$

La probabilité cherchée est donc d'environ 3,8 pour dix mille.

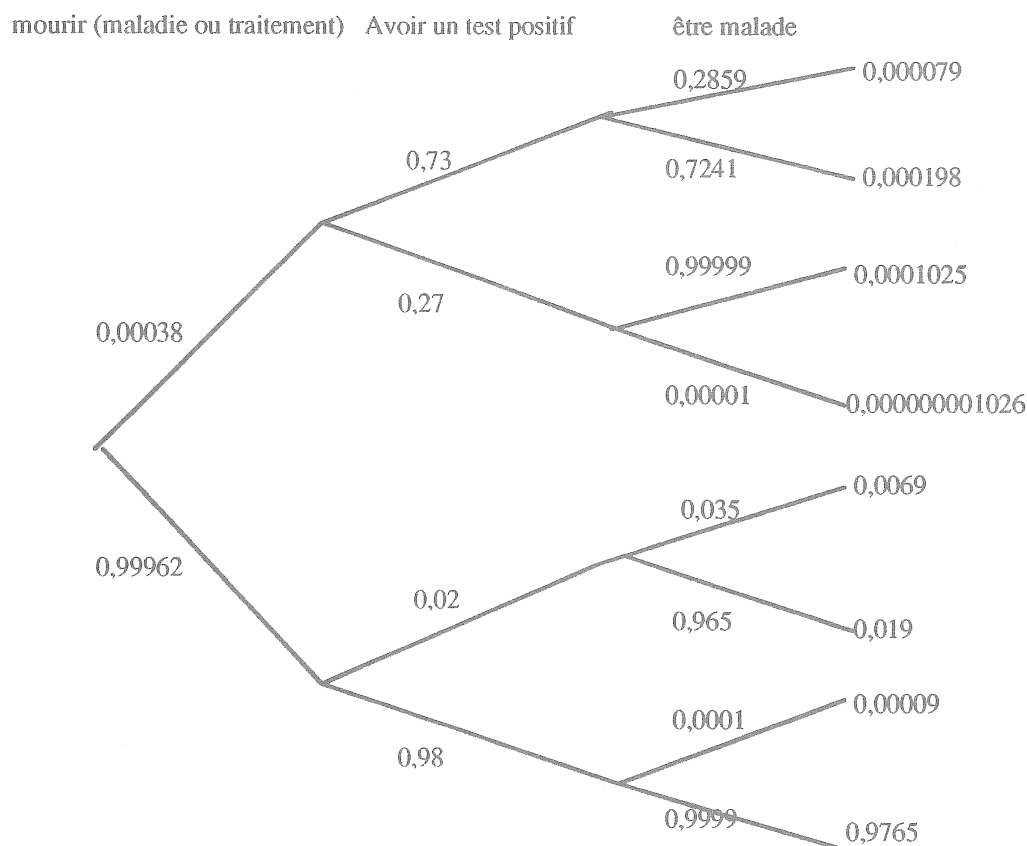
La représentation en arbre est ici un outil performant pour traiter des renseignements qui peuvent paraître nombreux et confus d'emblée.

On peut remarquer que le calcul de la probabilité trouvée précédemment constitue la première étape de l'inversion de cet arbre.

Nous avons vu qu'il est toujours possible d'inverser un arbre à deux niveaux, on peut en déduire que, de proche en proche, l'inversion complète est possible pour un arbre à n niveaux, ce n'est qu'une question de persévérance ! On peut même penser qu'avec quelques compétences en programmation, il serait facile d'automatiser un tel traitement.

A titre de curiosité, voici, tous calculs faits, ce que donne l'arbre précédent après inversion :

(les calculs sont approchés)



On peut y lire par exemple la probabilité d'avoir un test positif sachant qu'on est mort, ou celle d'être malade sachant qu'on est mort et qu'on a un test négatif...
SHOCKING, ISN'T IT ?

En guise de conclusion très provisoire...

L'objet de ce chapitre n'était pas de proposer une méthode d'enseignement des probabilités conditionnelles par les arbres. L'utilisation de cet outil nécessiterait au préalable une familiarisation. De nombreux problèmes peuvent être traités avec cet outil tout au long de la scolarité : depuis l'introduction de la multiplication aux comptages de cas, aux tris de population ou à l'organisation de données. Ces notions peuvent être enseignés avec cet outil bien avant une quelconque référence aux probabilités. Le principe de multiplication des cas pourrait alors faciliter le passage à l'utilisation des arbres dans la domaine des probabilités : un parcours possible serait d'utiliser ce principe sur des effectifs puis sur des proportions et/ou des pourcentages (passage d'une branche pour chaque individu à l'utilisation de branches pondérées) dans des problèmes de tris sur une population.

Nous espérons seulement avoir contribué à amener quelques éléments de réflexion sur les rapports entre l'outil (les arbres) et l'objet d'enseignement (les probabilités conditionnelles), l'analyse de ces rapports nous paraissant un élément important dans la nécessaire réflexion sur la démarche de modélisation.

QUELQUES ELEMENTS DE BIBLIOGRAPHIE

STATISTIQUE ET PROBABILITÉ (Bigot-Verlant) Editions Foucher

ELEMENTS DE CALCUL DES PROBABILITES (J. Bass) Masson et cie.

STATISQUES ET PROBABILITES (H. Egon) Hachette technique.

CONTREMANUEL DE STATISTIQUES ET DE PROBABILITÉS. (Peltier- Rouche-Manderick) Editions Vie ouvrière. Bruxelles.

LES CERTITUDES DU HASARD (Arthur Engel) Aleas editeur.

L'ENSEIGNEMENT DES PROBABILITÉS ET DE LA STATISTIQUE. (Arthur Engel).Cedic, 1975, deux volumes.

LES PROBABILITÉS À L'ÉCOLE (Glaymann-Varga) Cedic, 1975.

LES STRUCTURES DU HASARD (J.L. Boursin) Collection Points Sciences, 1986.

LES DÉES ET LES URNES. (J.L. Boursin) Seuil.

ARS CONJECTANDI.(J. Bernouilli) publication de l'IREM de Rouen, 1987.

ELEMENTS DU CALCUL DES PROBABILITÉS. (Condorcet) , publication de l'IREM de Paris VII, 1986.

ESSAI PHILOSOPHIQUE SUR LES PROBABILITÉS.(Laplace), Christian Bourgois éditeur.

MATHÉMATIQUES AU FIL DES AGES.(IREM groupe épistémologie et histoire) Gauthier-Villars.

LE HASARD (E.Borel) PUF.

HASARD ET CHAOS (D.Ruelle) Editions Odile Jacob.

LA PHYSIQUE DU HASARD (Ruhla) Hachette.

MATHÉMATIQUES : APPROCHE PAR DES TEXTES HISTORIQUES (LE PROBLÈME DES PARTIS) : publication de l'IREM de Paris VII, janvier 86

Articles récents

L'enseignement des probabilités en classe de Première : article de Annie et Michel Henry, paru dans Repères n° 6, janvier 92

L'enseignement des probabilités en classe de Terminale : article de Michel Henry, paru dans Repères n°

Paradoxes et lois de probabilités : article de Michel Henry et Henri Lombardi, paru dans Repères n° 13, octobre 93

Des statistiques aux probabilités, exploitons les arbres : article de Bernard Parzysz, paru dans Repères n° 10, janvier 93

Introduction du concept de probabilités conditionnelles : avantages et inconvénients de l'arborescence : article de Totohasina, paru dans Repères n°15, avril 94

Notation et sens, une étude de cas : article de Bernard Parzysz sur les notations en probabilités conditionnelles, paru dans le Bulletin Vert de l'APMEP n° 392, février mars 94

Persée, Piégase et Andromède : article de Ian Stewart paru dans Pour la science n° 167, septembre 91

Trois portes, deux chèvres et une auto : article de Bouziat et Pagès paru dans Quadrature n° 13 Septembre octobre 92

Une erreur de raisonnement à éviter dans l'interprétation de certains tests médicaux : article de M.D. Indjoudjian, paru dans Quadrature n° 8 mai juin 91

Publications IREM récentes pour l'enseignement

ENSEIGNER LES PROBABILITÉS EN CLASSE DE PREMIÈRE : publication de l'IREM de Strasbourg, octobre 92

ENSEIGNER LES PROBABILITÉS EN CLASSE DE TERMINALE : publication de l'IREM de Strasbourg, juin 94

A PROPOS DE L'ENSEIGNEMENT DU CALCUL DES PROBABILITES : par J C Thiénard, publication de l'IREM de Poitiers, octobre 93

