L'exercice de Probabilités au Baccalauréat Série D de Paris-Créteil-Versailles Juin 1988

par Jean Capron et Françoise Boucher
professeurs au Lycée Louis Thuillier
Amiens

En tant que chargé de mission à l'A.P.M.E.P., responsable des sujets d'examen, j'ai été alerté, dès Juillet 88, par notre collègue Michèle SADI, professeur agrégée au Lycée Paul Bert à Paris, de l'existence d'une erreur dans l'exercice de probabilités du Bac D de Paris-Créteil-Versailles.

"Ecrire simultanément : $p(\overline{T_2}) = [p(\overline{T})]^2$ (1)

et $p(\overline{T_2}/V) = [p(\overline{T}/V)]^2$ (2)

et $p(\overline{T_2}/\overline{V}) = [p(\overline{T}/\overline{V})]^2$ (2) est faux".

Effectivement, en appelant $p(V) = a$, $p(\overline{T}/V) = b$ et $p(\overline{T}/\overline{V}) = c$ on a :
$p(\overline{T}) = p(V) \times p(\overline{T}/V) + p(\overline{V}) \times p(\overline{T}/\overline{V})$.

Soit $p(\overline{T}) = ab + (1-a)c = c + ab - ac$, d'où $\left[p(\overline{T})\right]^2 = (c + ab - ac)^2$.

Si l'on suppose que $p(\overline{T}_2/V) = \left[p(\overline{T}/V)\right]^2$ et $p(\overline{T}_2/\overline{V}) = \left[p(\overline{T}/\overline{V})\right]^2$
alors : $p(\overline{T}_2) = p(V) \times p(\overline{T}_2/V) + p(\overline{V}) \times p(\overline{T}_2/\overline{V})$

$$p(\overline{T}_2) = p(V) \times \left[p(\overline{T}/V)\right]^2 + p(\overline{V}) \times \left[p(\overline{T}/\overline{V})\right]^2$$

soit $p(\overline{T}_2) = ab^2 + (1-a)c^2 = c^2 + ab^2 - ac^2$

d'où : $p(\overline{T}_2) - \left[p(\overline{T})\right]^2 = c^2 + ab^2 - ac^2 - c^2 - a^2b^2 - a^2c^2 - 2abc + 2ac^2 + 2a^2bc$

$$= a\left[b^2 + c^2 - 2bc - a\left(b^2 + c^2 - 2bc\right)\right]$$

c'est-à-dire : $= a(1-a)(b-c)^2$

Il ne peut donc y avoir compatibilité entre les deux formules (1) et (2) que dans les deux cas triviaux où $a = 0$ et $a = 1$, ou encore dans le cas où $b = c$, c'est-à-dire $p(\overline{T}/V) = p(\overline{T}/\overline{V})$ ce qui rendrait, bien entendu, le test tout-à-fait inutile ! Dans le cas présent $b \neq c$, et dans la deuxième partie de l'exercice, en suivant l'ordre des questions et en tenant compte du fait que *"on effectue deux tests identiques dans des conditions qui garantissent l'indépendance des résultats"*, on est amené dans le a), à écrire que :

$$p(\overline{T}_2) = \left[p(\overline{T})\right]^2 = (0{,}878)^2 \simeq 0{,}770\,884$$

puisque $p(\overline{T}_2/V) = \left[p(\overline{T}/V)\right]^2 = (0{,}05)^2 = 0{,}002\,5$ d'où on en "déduit" dans le b) que $p(V \cap \overline{T}_2) = p(V) \times p(\overline{T}_2/V) = 0{,}1 \times 0{,}002\,5 = 0{,}000\,25$

$$\text{et} \quad p(V/\overline{T}_2) = \frac{p(V \cap \overline{T}_2)}{p(\overline{T}_2)} = \frac{0{,}000\,25}{0{,}770\,884} \simeq 0{,}000\,324$$

En calculant : $p(\overline{T}_2) = p(V) \times \left[p(\overline{T}/V)\right]^2 + p(\overline{V}) \times \left[p(\overline{T}/\overline{V})\right]^2$
$p(\overline{T}_2) = 0{,}1 \times (0{,}05)^2 + 0{,}9 \times (0{,}97)^2 = 0{,}000\,25 + 0{,}846\,81$ soit
$p(\overline{T}_2) = 0{,}847\,06$. On trouve alors : $p(V/\overline{T}_2) = \dfrac{0{,}000\,25}{0{,}847\,06} \approx 0{,}000\,295$.

A noter que dans le corrigé qui figurait au Minitel après les épreuves du Baccalauréat on pouvait lire : " $p(\overline{T_2}) = [p(\overline{T})]^2 \approx 0{,}771$ " (on aurait pu envisager de traiter la question plus subtilement et de ne pas considérer les événements comme indépendants, dans ce cas, on trouve $p(\overline{T_2}) \approx 0{,}847$ " l'auteur se contentant de cette remarque sans prendre position, mais en utilisant la première valeur 0,771 pour terminer la question b) et trouver $p(V / \overline{T_2}) \approx 32.10^{-5}$.

Françoise BOUCHER, professeur au Lycée Louis Thuillier d'Amiens, a effectué une simulation sur ordinateur qui donne des valeurs de $p(\overline{T_2})$ voisines de 0,847 06 ; elle présente ici le programme qu'elle a utilisé.

Simulation sur ordinateur :

Le programme qui suit ne traite que du seul point contesté de l'exercice : le calcul de la probabilité de $\overline{T_2}$ (2a). Ce programme est écrit en Turbo-Pascal pour compatible PC.

La simulation se fait sur une population d'effectif variable (n). A condition de choisir n assez grand, une bonne approximation de la probabilité sera donnée par la fréquence de réalisation de l'événement "les deux tests sont négatifs" : $f = n / r$, r étant le nombre d'individus pour lesquels les tests sont négatifs.

Tout le problème réside dans le choix de la méthode de simulation; elle dépend elle aussi de l'interprétation du texte et, par conséquent, l'informatique ne permet pas de prouver de façon absolue que la solution de Jean CAPRON est correcte !

Cependant, comme l'approche du problème est totalement différente et qu'elle conduit à un résultat numérique très voisin, on peut raisonnablement penser que l'événement "la solution de J.Capron est la bonne" est un événement presque certain...

Techniques de simulation :

Chaque individu I de la population est codé par un nombre de trois chiffres écrits en binaire : a b c.

- Si I est porteur de virus	$a = 1$,	sinon :	$a = 0$
- Si le 1er test est positif	$b = 1$,	sinon :	$b = 0$
- Si le 2d test est positif	$c = 1$,	sinon :	$c = 0$.

Par exemple : si I a le code 110, I est porteur du virus, mais seul le premier test est positif..

Le codage est effectué dans une procédure "codage", appelée dans une boucle où i varie de 1 à n.

Le code x est mis à 000 à l'entrée de la procédure et les hypothèses sont alors traduites de la manière suivante :

1) Le virus : traduction de $P(V) = 0.1$

On tire au hasard un entier entre 0 et 9 (random (10)).
- Si random(10) = 0 (1 chance sur 10), l'individu est considéré comme porteur du virus : x prend la valeur 100
- sinon (9 chances sur 10) x reste nul: 000.

2) Premier test : traduction de $P(T/V) = 0.95$ et $P(T/\overline{V}) = 0.03$.

2 cas se présentent :

◊ si I est porteur du virus (sachant V, donc à ce stade x = 100), on tire au hasard un entier entre 0 et 99 (random(100)).
- si random(100)<95 (95 chances sur 100), le premier test est considéré comme positif x prend la valeur x + 10 = 110
- sinon (5 chances sur 100) x garde sa valeur 100

◊ si I n'est pas porteur du virus (sachant $\overline{V}$ à ce stade : x = 000), on tire au hasard un entier entre 0 et 99.
- si random(100)<3 (3 chances sur 100), x prend la valeur x+10=010
- sinon (97 chances sur 100) x garde sa valeur 000

3) Second test (passage délicat !).

On ignore les résultats du premier test : c'est l'hypothèse du 2) "conditions qui garantissent l'indépendance des résultats". On calcule c exactement de la même manière que b, en tenant compte seulement de la valeur de a que l'on suppose inchangée, car on peut penser que le malade est resté malade et n'a pas été guéri par le premier test !

◊ si x >= 100 (I est porteur du virus: à ce stade : x = 110 ou x = 100)
- si random(100)<95 x prend la valeur x + 1 (111 ou 101)
sinon x garde sa valeur: (110 ou 100)
◊ si x<100 (I n'a pas le virus : à ce stade : x = 010 ou x = 000)
- si random(100)<3 x prend la valeur x + 1 (011 ou 001)
sinon x garde sa valeur (010 ou 000)

4) A la fin du codage, les deux tests sont négatifs lorsque x se termine par 00, soit encore lorsque x est divisible par 100. Si c'est le cas, r (nombre d'individus pour lesquels les deux tests sont négatifs) prend la valeur r + 1.

A la sortie de la boucle codant et examinant les n individus, il suffit alors de calculer $f = r / n$, fréquence de réalisation de l'événement étudié (fréquence qui est calculée six fois de suite sur la même population).

Listing et exécution du programme :

Ci-dessous figurent le listing du programme (virus) ainsi que les résultats obtenus pendant l'exécution pour des valeurs de *n* égales à 1 000 000. On peut constater que les valeurs obtenues sont plus proches de 0.847 que de 0.771.

```
program virus;
var
f,i,n,r:real;
j:integer;

procedure codage(var r:real);

   var x:integer;
   begin
   x:=0;
   (*virus : traduction de P(V)=0.1*)
   if random(10)<1 then x:=x+100;

   (*1er test : traduction de P(T/V)=0.95 et P(T/V̅)=0.03*)

   if x=100 then if random(100)<95 then x:=x+10;
   if x=0 then if random(100)<3 then x:=x+10;

   (*2nd test : traduction de P(T/V)=0.95 et P(T/V̅) = 0.03 une
   seconde fois*)
```

```
if x>=100 then if random(100)<95 then x:=x+1;
if x<100 then if random(100)<3 then x:=x+1;

(*comptage des codes divisibles par 100: les 2 tests sont négatifs*)

if x mod 100 =0 then r:=r+1;
end;
begin
writeln('les fréquences obtenues sont successivement:');
for j:=1 to 6 do
begin
r:=0;i:=1;n:=1E+006;
randomize;
while i<=n do
            begin
            codage(r);i:=i+1;
            end;
f:=r/n;
writeln(f:6:4);
end;
end.
```

```
Running
les fréquences sont successi-
vement:
0.8467 0.8469 0.8474
0.8478 0.8469 0.8473
```

Comment obtenir quand même 0.771, mais en interprétant (à notre avis,!) les hypothèses d'indépendance de manière erronée ...

Il suffit simplement de redistribuer au hasard la maladie, après le premier test, en ignorant complètement tout ce qui a dû se passer avant, *les malades n'étant plus nécessairement les mêmes individus:* ce qui ne semble pas avoir beaucoup de sens, l'état des malades ne pouvant avoir changé.

Dans le programme (virus-bis) qui suit, après le premier test, un nombre h est tiré au hasard entre 0 et 9.

Si $h = 0$ (I porteur du virus), c prend la valeur 1 avec la probabilité souhaitée 95%, sinon (I non porteur du virus) c prend la valeur 1 avec une probabilité égale à 3%.

Le listing du programme est suivi d'une exécution répétée 10 fois de suite sur une population de 100 000 individus ... et l'on peut constater que la fréquence est effectivement proche de 0.771 !

```
program virus_bis;
var
f,i,n,r:real;
j:integer;

procedure codage(var r:real);
var   x,h:integer;
begin
x:=0;
(*virus : traduction de P(V)=0.1 *)
if random(10)<1 then x:=x+100;

(* 1er test : traduction de P(T/V)=0.95 et P(T/V̄)=0.03 *)
if x=100 then if random(100)<95 then x:=x+10;
if x=0 then if random(100)<3 then x:=x+10;

(* 2nd test : traduction de P(T/V)=0.95 et P(T/V̄) =0.03  une seconde*)
(* fois en redistribuant au hasard la maladie *)
h:=random(10);
if h=0 then if random(100)<95 then x:=x+1;
if h>0 if random(100)<3 then x:=x+1;

(* comptage des codes divisibles par 100: les 2 tests sont négatifs *)
if x mod 100 =0 then r:=r+1;
end;

begin
writeln(' les fréquences obtenues sont successivement:');
for j:=1 to 10 do
begin
r:=0;i:=1;n:=1E+006;
randomize

while i<=n do
          begin
          codage(r);i:=i+1;
          end;
f:=r/n;

writeln(f/6/');
end;
end.
```

```
Running
les fréquences obtenues sont
successivement :
0.7716  0.7713  0.7714
0.7687  0.7721  0.7715
0.7713  0.7719  0.7706
0.7690
```

En guise de conclusion :

Nous citerons une phrase de la lettre de Michèle SADI :"*Vue la manière dont l'ensemble des questions du II 2° étaient posées et s'enchaînaient et ont été comprises par la plupart des professeurs et des élèves, je pense que même celui qui a produit ce texte s'est planté*" et nous rapporterons cette phrase de P.L.HENNEQUIN : "*Il est absurde de supposer que les deux tests successifs sont indépendants. Cela revient à dire que le test n'apporte aucune information sur la maladie, c'est la négation de la médecine !*". Il semble donc raisonnable de faire l'hypothèse d'indépendance conditionnelle (relation (2)) et dans ce cas, la "bonne réponse" au 2° b) serait donc $p\left(V / \overline{T_2}\right) = 0{,}000\ 295$.

Il est souhaitable qu'à l'avenir, les textes concernant les exercices de probabilités au baccalauréat soient non ambigus et de difficulté modérée.