Indépendance et indépendance conditionnelle

L'article de Jean CAPRON et Françoise BOUCHER sur l'exercice de Probabilités du Bac D de Paris (Juin 1988) paru dans le *Bulletin* n° 372 (Février 1990), page 89, a suscité un abondant courrier.

Pierre CARRIQUIRY (Ecole Nationale de Commerce de Paris) suggère d'utiliser un modèle d'urne et met en garde l'appel à l'intuition par l'exercice "paradoxal" suivant :

n virus $V_1, V_2, \ldots V_n$ sont répartis dans n éprouvettes, chaque éprouvette ne contenant qu'un seul type de virus. Considérons alors un "savant menteur" (il y en a ... mais pas en France, bien sûr) qui analyse le contenu d'une éprouvette choisie au hasard et qui annonce le résultat de la manière suivante : si le virus contenu dans l'éprouvette est V_i, la probabilité pour qu'il dise que c'est V_i est 0,1 et la probabilité pour qu'il dise que c'est V_j est $\frac{0,9}{n-1}$ (pour tout $j \neq i$). Autrement dit, le savant menteur dit la vérité une fois sur 10 et ment 9 fois sur 10.

Supposons alors que 2 savants menteurs analysent le contenu de la même éprouvette et disent tous les deux que le virus contenu dans cette éprouvette est V_1. Calculer la probabilité P_n pour que le virus V_1 soit effectivement contenu dans l'éprouvette.

Soient les événements :
A_1 : "l'éprouvette contient le virus V_i ". $i = 1,2,\ldots n$. $P(A_i) = 1/n$.

S_1 : "le premier savant menteur dit que l'éprouvette contient le virus V_1".
S_2 : "le 2ème savant menteur dit qu'elle contient le virus V_1".

On doit calculer $P(A_1 / S_1 \cap S_2) = \dfrac{P(A_1) \times P(S_1 \cap S_2 / A_1)}{P(S_1 \cap S_2)}$

Comme dans l'exercice précédent, l'indépendance des savants menteurs ne permet pas de dire que les événements S_1 et S_2 sont indépendants.
$P(S_1 \cap S_2 / A_1) = P(S_1 / A_1) \times P(S_2 / A_1)$ en supposant que les savants menteurs soient indépendants.

$$P(S_1 \cap S_2) = P(A_1) \times P(S_1 \cap S_2 / A_1) + \dots P(A_n) \times P(S_1 \cap S_2 / A_n) =$$
$$= \frac{1}{n}(0{,}1)^2 + \frac{1}{n}\left(\frac{0{,}9}{n-1}\right)^2 (n-1)\ .$$

On obtient, après simplification : $P(A_1 / S_1 \cap S_2) = \dfrac{n-1}{n+80}$.

Le paradoxe est que $\lim\limits_{n \to +\infty} \dfrac{n-1}{n+80} = 1$ Autrement dit, si les deux savant menteurs disent la même chose, et s'il y a beaucoup de virus, il faut les croire.

Alain PICHEREAU (Lycée Marguerite de Valois , Angoulème) donne une situation où la formule $P(\overline{T}_2) = P(\overline{T})^2$ s'applique : elle consiste à faire les deux tests successifs non pas sur le même individu, mais sur deux individus tirés indépendamment au hasard dans la population.

Bernard PARZYSZ indique que l'indépendance conditionnelle par rapport à V de T_1 et T_2 : $P(T_1 \cap T_2) / V) = P(T_1 / V)\, P(T_2 / V)$ n'implique pas celle de T_1 et T_2 par rapport à $\overline{V}$:

$$P(T_1 \cap T_2 / \overline{V}) = P(T_1 / V)\, P(T_2 / \overline{V})$$

et que supposer les deux revient à imposer l'indépendance conditionnelle à la partition $(V,\overline{V})$. Il ajoute la remarque suivante sur l'"usage des probabilités conditionnelles :

Je relève depuis quelques temps, dans les sujets de Bac, un type particulier de formulations "vicieuses", qui me semblent tout à fait fâcheuses quant à la confusion qu'elles risquent d'introduire dans l'esprit des élèves. Ainsi :
Bac D Paris Juin 90 : "on pourra noter A/B l'événement "A sachant que B".

Bac D Paris Juin 89 : "les deux événements "D sachant que S" et "G sachant que S" sont-ils indépendants ?"
Bac A1 Polynésie Française Juin 89 : "Calculer les probabilités des événements (...) H : "Catherine fait partie du jury sachant que Michel en fait partie".
Bac D Paris Juin 88 : "On admet que, en désignant pat p(E) la probabilité de l'événement E, (...) p(T sachant que V) = 0,05 ...".
Bac D Lyon Juin 87 : "En notant P(A/B) la probabilité de l'événement : "A sachant que B"...".

Le danger réel que j'y vois (les auteurs des textes ci-dessus en sont-ils conscients ?) est que la notation P(A/B) induit, par assimilation avec la notation p(A), l'idée *fausse* que A/B est un événement au même titre que A et B, d'où les formulations ci-dessus. C'est pourquoi il me semble de beaucoup préférable - au moins dans les débuts de l'enseignement - d'utiliser la notation indicielle $P_B(A)$, qui évite la confusion, et en outre met en évidence le fait que P_B est :
- une probabilité
- différente de P.

Peut-être quelqu'un pourrait-il faire une mise au point là-dessus dans le *Bulletin*, dans le but de mettre fin à cette pratique ? Cela me semble urgent.

Enfin, Bernard LAFARGE (Pontoise) fait référence à l'article de Bernard PARZYSZ sur *"l'arbre probabiliste"* paru dans le *Bulletin* n°372 (Février 1990) page 47 et indique :

J'ai personnellement testé cet exercice sur ma classe de T.D. J'ai beaucoup employé dans mon cours sur les Probabilités conditionnelles les arbres pondérés. Tous les élèves ayant utilisé cette technique sont arrivésd au résultat, les autres (malheureusement influencés par *"succession de DEUX tests identiques réalisés INDÉPENDAMMENT l'un de l'autre"* et *"des conditions qui garantissent l'indépendance des résultats"*) n'y sont pas arrivés.