Remarques sur le programme de probabilités dans les sections A, B, C, D et E des lycées (*)

Le Chapitre des Epreuves répétées

par H.L. MORITZ, M. BLANCHARD et C. BLOCH, I.R.E.M. de Poitiers.

On ne peut plus aujourd'hui traiter le moindre exercice de calcul des probabilités sans invoquer un espace probabilisé sous-jacent $(\Omega\ , \mathcal{A})$ où

Ω représente l'ensemble des éventualités (résultats possibles de l'expérience aléatoire envisagée) ;

(*) Cet article précise certains points évoqués dans la brochure A.P.M.E.P. n° 17 (Hasardons-nous) pages 95 sq.

$\mathcal{A}$, support de la probabilité, est une tribu de sous-ensembles de Ω qui constitue l'ensemble des événements à considérer (les questions qu'on se pose relativement à cette expérience).

Notons qu'il y a en général plusieurs façons de définir l'expérience et de choisir $(\Omega, \mathcal{A})$.

Les vieux probabilistes, en somme, faisaient cela sans toujours le dire explicitement. Il faut reconnaître que la nouvelle méthode est sûre et efficace quand on a bien choisi Ω ; elle rassure ceux qui n'ont pas l'instinct combinatoire et elle ne dispense pas totalement d'avoir du bon sens.

En fait, dans l'enseignement secondaire, Ω est toujours un ensemble fini et, pour $\mathcal{A}$, même en Terminale, on prend généralement tout $\mathcal{P}(\Omega)$, si bien qu'on a eu seulement la satisfaction morale (?) d'introduire le mot tribu. Cependant on s'est engagé sur une voie difficile car, si l'on veut se maintenir à cette altitude théorique, comme y invitent les Instructions de la Circulaire n° 71-244 du 26 juillet 1971 (p. 140 de la brochure publiée par le Ministère, *Mathématiques - Classes du Second Cycle. Horaires - Programmes - Instructions*),

(i) on est amené à définir une variable aléatoire réelle comme une application mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ où $\mathcal{B}_{\mathbb{R}}$ est la tribu des boréliens de $\mathbb{R}$ (tribu engendrée par les intervalles) ;

(ii) il faut aussi, dès qu'on envisage une suite d'épreuves *indépendantes*, définir un espace probabilisé-produit (*) Ω (produit cartésien des Ω_i), la tribu-produit $\mathcal{A}$ (engendrée par les pavés : $\prod_{i=1}^{n} A_i$, $A_i \in \mathcal{A}_i$) et la probabilité P portée par cette tribu ; en fait, là encore on prend pour $\mathcal{A}$ tout $\mathcal{P}(\Omega)$ mais la probabilité-produit n'est définie que sur les pavés en posant

$$P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P_i(A_i)$$

(*) A ce propos, les Instructions sont étrangement muettes et se bornent à introduire $\Omega' = \Omega^n$ et la probabilité — sic — de l'événement $(x_1 = a_1, \dots, x_n = a_n)$ négligeant leurs premières recommandations.

et tout sous-ensemble de Ω n'est pas un pavé — heureusement cela ne pose pas de problème pour Ω fini. Il faut quand même remarquer que $\mathcal{T}(E \times F)$ n'est pas $\mathcal{T}(E) \times \mathcal{T}(F)$; c'est d'ailleurs un bon exercice ;

(iii) enfin il faut définir les variables aléatoires de façon cohérente entre les espaces "facteurs" et l'espace-produit (voir ci-dessous b)).

*
* *

(i) Pour le premier point, quand on n'a affaire qu'à des variables discrètes, c'est prendre les choses au rebours de la pratique et du sens commun, alors que rien n'y oblige à ce niveau car, même si tous les élèves sont de futurs utilisateurs de la statistique, ils n'auront généralement pas besoin de ce point de vue axiomatique rigoureux.

En effet, ce qu'on connaît le mieux dans la pratique, et ce dont on a besoin, ce sont les variables aléatoires avec leurs valeurs (ensemble image) (*) et la pondération qu'on y attache (diagramme en bâtons) ; on peut regretter à ce propos que la fonction de répartition soit tombée quasiement en désuétude (surtout dans l'enseignement supérieur) depuis qu'on privilégie la présentation axiomatique et "algébrique" par rapport à l'aspect analytique du calcul des probabilités, qui reste pourtant la base de la statistique mathématique.

(*) C'est magnifiquement illustré par l'exercice donné au bac, C en 1975 à Rennes (Annales p. 47). Visiblement l'auteur est parti d'une très banale variable aléatoire, d'où atomes, d'où tribu, événements dans cette tribu ; puis il a rédigé l'énoncé exactement en sens inverse ; le seul intérêt de cet exercice est sans doute de dresser les élèves à marcher sur la tête. Au moins n'y a-t-il pas d'erreur. Par contre, dans l'exercice donné en D (1975) à Abidjan (Annales p. 101), l'auteur, oubliant qu'il a des *atomes* (oubli compréhensible puisqu'ils ne servent à rien), finit par les couper en deux !...
S'il faut absolument introduire ce vocabulaire, au moins qu'on ne le considère pas comme la quintessence du programme de Probabilités, car alors avec un contenu d'une pauvreté affligeante on arrive à fabriquer un petit casse-tête formel dont seuls viendront à bout les élèves capables de régurgiter le cours à l'endroit, à l'envers et par morceaux.

Il faut pouvoir travailler aussi bien avec la probabilité-image, visualisée par le graphe de la fonction de répartition et celui de la densité quand il y en a une, qu'avec l'espace probabilisé. Ensuite on ne sort pas de la tribu engendrée par la v.a.r. (la mesurabilité est donc assurée) ; et c'est ainsi en définitive qu'on acquiert la connaissance du phénomène représenté par Ω . Si bien qu'au bout du compte les notions essentielles sont introduites, mais cette fois on sait pourquoi et on peut alors, si on y tient, donner la *définition* d'une variable aléatoire réelle dans toutes sa généralité.

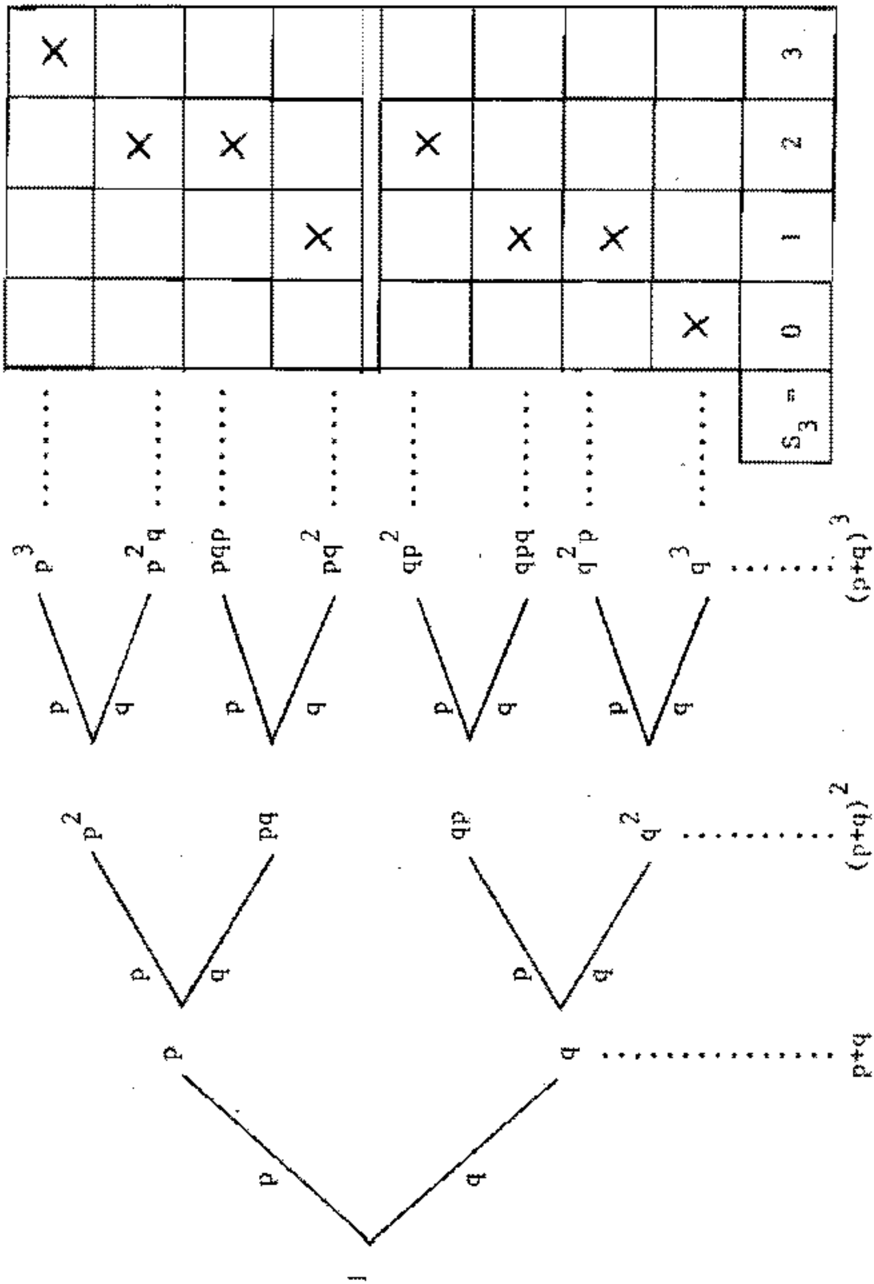
(ii) et (iii) Pour le second et le troisième point c'est plus délicat. Si par exemple on veut définir la variable binomiale comme une somme de variables de Bernoulli, on ne peut se dispenser d'introduire un espace-produit. Or la notion est difficile, aussi plusieurs manuels glissent-ils rapidement — on ne le leur reprocherait pas s'il ne leur arrivait de déraper jusqu'à tomber dans l'erreur.

Mais la loi binomiale est au programme ainsi que la loi des grands nombres ; il faut donc parler d'épreuves répétées. Alors comment faire ?

a) Remarquons d'abord qu'on peut définir globalement l'épreuve binomiale (comme n coups à pile ou face) et la variable binomiale (nombre des succès) en représentant Ω comme un ensemble de cheminements sur un arbre ; à l'étage i on a tous les résultats possibles en i coups. Donc :

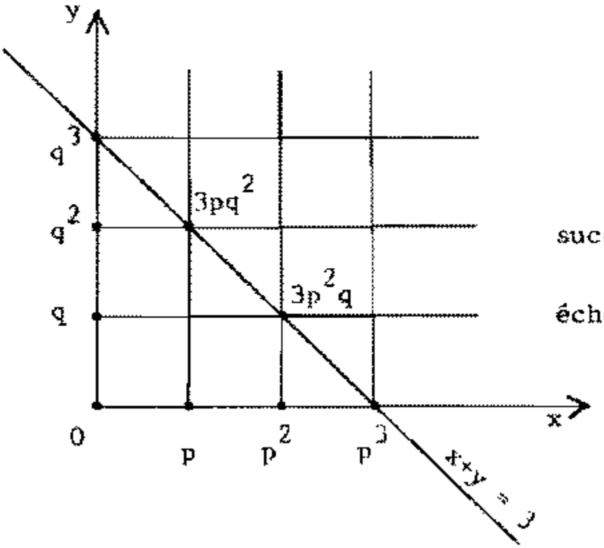
— S_n nombre de succès en n coups
— succès (p) : rameau vers le haut de la page
  échec ($1-p = q$) : rameau vers le bas de la page.
— p constant
— à chaque noeud la probabilité se partage en proportion
  $\left.\begin{array}{l} p \text{ pour succès} \\ q \text{ pour échec} \end{array}\right\} (p + q = 1)$,

(On aura d'abord opéré sur les résultats d'une enquête avec effectifs fractionnés en pourcentages, bonne occasion de vérifier que les élèves (ne) savent (pas) faire les calculs). On obtient au n-ième étage de l'arbre, tous les termes du développement de $(p + q)^n$; par conséquent l'ensemble des chemins correspondant à k succès a pour probabilité la somme des termes en p^k , soit $C_n^k \; p^k \; q^{n-k}$.


1
p
q
p+q
p²
pq
qp
q²
(p+q)²
p³
p²q
pqp
pq²
qp²
qpq
q²p
q³
(p+q)³
S₃ =
0
1
2
3

Voici une autre représentation de l'épreuve binomiale : *un* chemin partant de 0 et aboutissant (en n coups) sur la droite $x + y = n$ au point $M(k, n-k)$ a pour probabilité $p^k q^{n-k}$ (indépendance) ; il y a C_n^k chemins OM.

La probabilité est lue immédiatement mais on voit moins bien se former le développement du binome.



succès $\longrightarrow$ (p)

échec $\uparrow$ (q)

Ensuite il faut calculer, directement à partir de la définition, l'espérance et la variance de S_n et démontrer directement le théorème de Bernoulli :

$$E(S_n) = \sum_{k=0}^{n} k\, C_n^k\, p^k\, q^{n-k} = np \sum_{k=1}^{n} C_{n-1}^{k-1}\, p^{k-1}\, q^{(n-1)-(k-1)} = np\,.$$

$$\mathrm{var}(S_n) = E[S_n\,(S_n - 1)] + E(S_n) - [E(S_n)]^2 = \ldots = npq.$$

En posant $I = \; k\,;\, 0 \leqslant k \leqslant n,\ |k - np| > n\epsilon \; \subseteq \; 0, 1, \ldots, n$,

$$\mathrm{Prob}\ \ |\frac{S_n}{n} - p| > \epsilon \ = \sum_{k \in I} C_n^k\, p^k\, q^{n-k}\,,$$

$$< \frac{1}{n^2\,\epsilon^2} \sum_{k \in I} (k - np)^2\ C_n^k\, p^k\, q^{n-k}$$

$$\leqslant \frac{1}{n^2\,\epsilon^2}\ npq \leqslant \frac{1}{4n\epsilon^2}\,.$$

De toutes façons cet exercice en vaut bien un autre.

b) Mais naturellement il est intéressant de pouvoir considérer S_n comme *la somme* de n variables de Bernoulli indépendantes et de même loi. *Il faut donc que toutes ces variables soient définies sur un même espace de probabilité*, or on considère généralement n variables X_i, chacune définie sur un espace

$$(\Omega_i, \mathcal{P}(\Omega_i), P_i) \text{ avec } \forall i\ P_i(X_i = 1) = p,\ P_i(X_i = 0) = 1 - p,$$

i étant le rang du tirage ($1 \leqslant i \leqslant n$).

Remarque : les v.a. X_i peuvent avoir même loi sans que les Ω_i soient identiques.

. Si alors on pose $\Omega = \prod_{i=1}^{n} \Omega_i$, un élément de Ω est un n—uple $(\omega^1, ..., \omega^i, ..., \omega^n)$. Soit f_i l'application i-ième coordonnée : $\Omega \to \Omega_i$ et soit $Y_i = X_i \circ f_i$. Il faut montrer que Y_i est une v.a. définie sur Ω qui prend la valeur 1 avec la probabilité p et la valeur 0 avec la probabilité $1 - p$; or, si

$$A_i = X^{-1}(1) = \omega^i \in \Omega_i \ / \ X_i(\omega^i) = 1 ,$$

on a

$$B_i = Y^{-1}(1) = \Omega_1 \times \Omega_2 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n$$

$B_i \in \mathcal{A} = \mathcal{P}(\Omega)$ c'est une section de Ω et, par définition de la probabilité-produit P, l'indépendance des n tirages de Bernoulli se traduit par :

$$P(B_i) = P_1(\Omega_1) \dots P_{i-1}(\Omega_{i-1})\, P_i(A_i) . P_{i+1}(\Omega_{i+1}) \dots P_n(\Omega_n)$$
$$= p.$$

Sur $(\Omega, \mathcal{A}, P)$ on définit la v.a.r.

$$S_n = \sum_{i=1}^{n} Y_i \text{ d'où } E(S_n) = nE(Y_i) = np,\ \operatorname{var}(S_n) = n \operatorname{var}(Y_i) = npq,$$

et comme il en résulte

$$E\ \frac{S_n}{n} = p \text{ et } \operatorname{var}\left(\frac{S_n}{n}\right) = \frac{pq}{n}$$

on en déduit que $\frac{S_n}{n}$ tend vers p en probabilité quand $n \to \infty$, en appliquant l'inégalité de Bienaymé-Tchebychev.

*

* *

Que disent les manuels de Première et de Terminale CE ? Certains sont parfaitement corrects. D'autres adoptent en somme notre démarche a) sans la représenter et en appelant la variable S_n "compteur" ou "totalisateur" des succès ; ils ajoutent en "Remarque" que S_n peut *aussi* être définie sur U^n (il s'agit d'épreuves répétées, Ω_i est toujours le même U). Mais il y a plus grave : certains disent que les v.a. X_i sont *égales*. Or des v.a. *de même loi de probabilité* ne sont pas égales en tant qu'applications, a fortiori si elles ne sont pas définies sur le même ensemble ! Il y a donc intérêt à individualiser les n univers même s'ils sont égaux pour éviter une telle erreur. Bien entendu si on définit les v.a. Y_i sur $\Omega = \prod_{i=1}^{n} \Omega_i$, comme ci-dessus en b), on voit bien que $\forall \omega \in \Omega$ et $i \neq j$

$$P(\ \omega \in \Omega / Y_i(\omega) = Y_j(\omega)\ ) = p^2 + q^2\ ,$$

$$P(\ \omega \in \Omega / Y_i(\omega) \neq Y_j(\omega)\ ) = 2pq\ ,$$

donc ces variables ne sont pas égales (elles le sont seulement presque sûrement si $p = 0$ ou 1).

D'autre part si chaque X_i est définie sur Ω_i on ne peut écrire $S_n = X_1 + X_2 + ... + X_n$ car cela n'a pas de sens. Pourtant on trouve même, dans un manuel couramment utilisé, textuellement, pour S_n :

... "*$X + X + ... + X$, cette somme comportant n termes, puisqu'il s'agit de v.a. de même loi et* indépendantes" [?] *(souligné par nous)*,

mais (plus loin, en "Remarque") :

"*il ne faudrait pas écrire nX, ce qui signifierait que la valeur prise par X, 0 ou 1, est multipliée par n et conduirait à [S_n =] 0 ou n, ce qui est faux*".

Les probabilistes avaient été surpris, il y a quelques années par l'apparition au niveau du secondaire du vocable "aléa numérique" (est-ce d'après la remarque des Instructions ?). Cette fois ce sont les élèves qui pourraient s'étonner, malheureusement ils se soumettent docilement au dressage, ou bien ils se démettent — et nul ne peut mesurer le dommage qui en résulte.

Il a fallu déjà reconnaître à la v.a. le statut de fonction. Voilà maintenant que f + f + ... + f (n fois) n'est plus égal à nf. Cela montre qu'il faudrait s'astreindre à écrire assez souvent, non pas, par exemple, $P(X = x)$ mais $P(\ \omega/X(\omega) = x\ )$, c'est indispensable à ce niveau, cela reste souvent utile à des niveaux plus élevés. On verrait bien alors que dans

$$S_n(\omega) = X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega),$$

c'est l'argument ω qui reste le même, alors que les fonctions X_i ne sont pas la même fonction X, mais des fonctions différentes définies sur *un même ensemble*.

On arrive alors à l'espace vectoriel $\mathcal{E}$ des v.a.r. étagées (Ω fini ou $\mathcal{A}$ finie) définies sur un même espace probabilisé, et à l'espérance considérée comme opérateur linéaire sur $\mathcal{E}$. Est-ce trop difficile ? N'a-t-on pas le temps de traiter correctement cette question ? Alors il ne faut pas s'y lancer. L'essentiel n'est pas d'éblouir les élèves (?) par un langage sophistiqué, ce devrait être de les amener à résoudre des problèmes pratiques et à ne pas dire de bêtises ni s'en laisser conter.