Ecart à l'indépendance d'événements : encadrement et comportement limite

Jean-Baptiste Hiriart-Urruty
Départment de mathématiques
Université Paul Sabatier de Toulouse

Cette note pédagogique est rédigée en hommage à Paul-Louis Hennequin récemment disparu. PLH, comme nous l'appelions familièrement, m'acueillit, avec d'autres de ses collègues, quand je commençais ma carrière d'enseignant-chercheur au département de mathématiques appliquées de l'université de Clermont-Ferrand. Nommé assistant agrégé, je venais de l'enseignement secondaire.

Bien que ne travaillant que partiellement dans des domaines "relevant du stochastique", PLH nous a toujours montré son soutien et son intérêt en assistant à tous les exposés de séminaires que je donnais ; d'ailleurs, c'était aussi le cas à l'endroit des autres conférenciers du département. J'ai eu l'honneur et le plaisir de l'avoir comme membre examinateur de ma Thèse de Doctorat ès Sciences Mathématiques.

Plus tard, lorsque j'étais Professeur à l'université Paul Sabatier de Toulouse, PLH et moi avons continué à avoir des échanges épisodiques, parfois sur un livre que j'avais pu écrire, parfois sur des actions de popularisation mathématique, qu'il appréciait particulièrement. C'est d'ailleurs dans cet esprit que j'ai écrit ce texte.

L'écart à l'indépendance... les débuts

"$P(A \text{ et } B) = P(A)P(B)$ lorsque les événements A et B sont indépendants" est l'une des premières choses que l'élève ou l'étudiant débutant apprend en cours de Probabilités..., c'est même la définition de "l'indépendance de 2 événements A et B". Lorsque A et B ne sont pas indépendants, la différence $e(A,B) := P(A \text{ et } B) - P(A)P(B)$ n'est pas nulle..., mais évidemment comprise entre -1 et 1, puisque c'est la différence de 2 nombres réels compris entre 0 et 1. On pourrait s'en tenir là, et c'est d'ailleurs ce qu'on fait usuellement dans un cours de Probabilités...

On pourrait donc penser que "l'écart à l'indépendance" $e(A,B)$ de 2 événe-ments A et B peut prendre n'importe quelle valeur entre -1 et 1... Or, il n'en est rien : cet écart est toujours compris entre $-1/4$ et $1/4$! C'est en consultant un récent recueil d'exercices d'oraux de concours ([1, Exercice 3.7]) que j'ai été arrêté par ce résultat... Peut-être l'ai-je appris dans une vie antérieure, en tout cas je l'avais complètement oublié. Aussitôt surgissent les questions : comment le démontrer (si possible de plusieurs façons très différentes) ? Qui a publié le permier ce résultat ? Comment généraliser au cas de n événements ? C'est à cette tâche que nous allons nous atteler.

Cas de 2 événements

Comme souvent en mathématiques, mais pas toujours, une fois qu'on sait ce qu'il faut démontrer, on peut s'attaquer à la démonstration... La démonstration dépend du niveau de connaissances préalables dans lequel on se place, c'est bien ce qui sera le cas ici.

Après quelques recherches bibliographiques et la consultation de collègues, probabilistes ou pas, j'ai pu détecter la première occurrence de ce résultat : l'encadrement et une paire de démonstrations sont dûs à Mme Edith Kosmanek en 1996 ([2]) ; ils sont repris comme exercice dans un livre d'enseignement des Probabilités en universités la même année ([3, Exercice 18 de la page 66]. Ensuite c'est dans les exercices d'oraux de concours que j'ai vu apparaître ce résultat.

Voici donc le résultat et quelques-unes des démonstrations possibles.

Théorème 1 (E. Kosmanek). *Soit A et B deux événements d'un espace probabilisé $(\Omega, \mathcal{F}, P)$. Alors, l'écart à l'indépendance de A et B, à savoir $e(A,B) := P(A \text{ et } B) - P(A)P(B)$, est encadré comme suit :*

$$-1/4 \leqslant e(A,B) \leqslant 1/4. \tag{1}$$

Notons que l'on a bien réduit l'encadrement trivial par -1 et 1 signalé au début. De plus, on ne pourra pas faire mieux que (1) : en effet, pour un événement A dont la probabilité est $1/2$, nous avons $e(A,A) = 1/4$ tandis que $e(A,A^c) = -1/4$ (A^c désigne ici et dans la suite l'événement contraire de A).

Passons aux démonstrations de (1).

Nous commençons par la démonstration qui nous paraît la plus simple, de niveau Lycée (mais, bien sûr, c'est un avis subjectif).

- *Démonstration faisant intervenir des probabilités conditionnelles* (niveau Lycées). Avec l'apport de P. Lassère.

Nous avons

$$\begin{aligned} e(A,B) &:= P(A \text{ et } B) - P(A)P(B) = P(A|B)P(B) - P(A)P(B) \\ &= P(B)[P(A|B) - P(A)]. \end{aligned} \tag{2}$$

Or, $P(A) = P(A \text{ et } B) + P(A \text{ et } B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$. Injectons ceci dans (2), de manière à obtenir

$$e(A,B) = P(B)[P(A|B) - P(A|B)P(B) - P(A|B^c)P(B^c)]$$

$$= P(B)[P(A|B)(1 - P(B)) - P(A|B^c)P(B^c)]$$
$$= P(B)[P(A|B)P(B^c) - P(A|B^c)P(B^c)].$$

Ainsi,
$$e(A, B) = P(B)P(B^c)[P(A|B) - P(A|B^c)]. \qquad (3)$$

Alors : $P(A|B)$ et $P(A|B^c)$ sont 2 nombres réels compris entre 0 et 1, donc $|P(A|B) - P(A|B^c)| \leqslant 1$; $P(B)P(B^c)$ est de la forme $x(1-x)$ avec x compris entre 0 et 1, donc majoré par $\max_{x\in[0,1]} x(1-x) = 1/4$. L'inégalité $|e(A, B)| \leqslant 1/4$ est bien démontrée.

Remarque. L'expression (3) est, bien sûr, "symétrisable" de manière à aboutir à

$$e(A, B) = e(B, A) = P(A)P(A^c)[P(B|A) - P(B|A^c)]. \qquad (4)$$

Par ailleurs, il vient immédiatement de (3) que $e(A, B^c) = -\, e(A, B)$. Cela milite en faveur du fait que la borne supérieure et la borne inférieure dans (1) doivent être opposées ; ce ne sera pas le cas pour plus de 3 événements (voir plus loin). Toujours en tirant sur le même fil : A et B sont indépendants (c'est-à-dire $e(A, B) = 0$) si et seulement si A et B^c sont indépendants (puisque $e(A, B^c) = -\, e(A, B)$)

- *Démonstration par l'optimisation d'une fonction de plusieurs variables* (l'une des preuves données dans [2]), (niveau Bac + 2).

En se servant de A et B, on crée une partition de Ω de la manière suivante :

$$C_1 = A \text{ et } B\,;\ C_2 = A \text{ et } B^c\ ;\ C_3 = A^c \text{ et } B\,;\ C_4 = A^c \text{ et } B^c.$$

Notons x, y, z, t leurs probabilités respectives. On a alors :

$$e(A, B) = f(x, y, z, t) = x - (x + y)(x + z)$$
$$= x(1 - x - y - z) - yz$$
$$= xt - yz \text{ (puisque } 1 = x + y + z + t). \qquad (5)$$

L'optimisation, c'est-à-dire la maximisation et la minimisation, de la fonction (de 4 variables) $(x, y, z, t) \mapsto f(x, y, z, t) = xt - yz$ sous les contraintes $x \geqslant 0, y \geqslant 0, z \geqslant 0, t \geqslant 0$ et $1 = x + y + z + t$ est facile. C'est encore la règle $\max_{x\in[0,1]} x(1 - x) = 1/4$ qui va servir. En effet :

$$\begin{aligned} f(x, y, z, t) &= xt - yz \leqslant xt \leqslant x(1 - x) \leqslant 1/4 \\ (\text{puisque } x &\geqslant 0 \text{ et } t \leqslant 1 - x)\,; \\ -f(x, y, z, t) &= yz - xt \leqslant yz \leqslant y(1 - y) \leqslant 1/4. \end{aligned}$$

Donc, l'inégalité $|f(x, y, z, t)| \leqslant 1/4$ pour les probabilités x, y, z, t est démontrée.

Remarque. L'évaluation (5) de $e(A, B)$ est intéressante pour avoir, par exemple, ceci : A et B sont indépendants si et seulement si $xt = yz$, soit

$$P(A \text{ et } B)P(A^c \text{ et } B^c) = P(A \text{ et } B^c)P(A^c \text{ et } B). \tag{6}$$

- *Démonstration par application ad hoc de l'inégalité de Cauchy-Schwarz* (preuves données dans [2, 1]), (niveau Bac + 3).

On considère les variables aléatoires particulières que sont les indicatrices 1_A et 1_B. Alors,

$$\begin{aligned} \mathrm{cov}(1_A, 1_B) &= \mathrm{E}(1_A . 1_B) - \mathrm{E}(1_A)\mathrm{E}(1_B) \\ &= \mathrm{E}(1_{(A \text{ et } B)}) - \mathrm{E}(1_A)\mathrm{E}(1_B) \\ &= P(A \text{ et } B) - P(A)P(B) = e(A, B). \end{aligned}$$

Par l'inégalité de Cauchy-Schwarz,

$$[\mathrm{cov}(1_A, 1_B)]^2 \leqslant \mathrm{var}(1_A)\mathrm{var}(1_B) = x(1-x)y(1-y), \tag{7}$$

$$\text{où } x = P(A) \text{ et } y = P(B).$$

Encore et toujours l'inégalité $u(1-u) \leqslant 1/4$ pour tout u entre 0 et 1 permet de conclure.

Remarque. Observons l'inégalité (7) écrite avec les probabilités, qui a une certaine esthétique :

$$|P(A \text{ et } B) - P(A)P(B)| \leqslant \sqrt{P(A)P(A^c)}\sqrt{P(B)P(B^c)}. \tag{8}$$

Et pour n événements ?

Les mathématiciens sont incorrigibles pour cela... Inévitablement ils poseront la question :

Quid de l'écart à l'indépendance $P(A_1 \text{ et } A_2 \text{ ... et } A_n) - P(A_1)P(A_2) \text{ ... } P(A_n)$?

J'avoue avoir commencé par le cas $n = 3$ et une démonstration parallèle à la deuxième du Théorème 1 au-dessus : avec la partition de Ω engendrée par 3 évenements A, B et C, on arrive à une fonction f de 8 variables... difficile à optimiser. L'approche ci-dessous est générale, valable pour tout n. Là, comme d'habitude, une fois qu'on sait ce qu'il faut démontrer, les choses deviennent plus faciles.

Soit donc n événements $A_1, A_2, ..., A_n$ et $e(A_1, A_2, ..., A_n) := P(A_1$ et A_2 ... et $A_n) - P(A_1)P(A_2)$ ... $P(A_n)$. L'encadrement de $e(A_1, A_2, ..., A_n)$, général et optimal, est comme suit.

Théorème 2. *Pour tout $n \geqslant 2$, nous avons*

$$-\left(\frac{n-1}{n}\right)^n \leqslant e(A_1, A_2, ..., A_n) \leqslant (n-1)\left(\frac{1}{n}\right)^{\frac{n}{n-1}}. \tag{9}$$

Deux observations avant de faire une démonstration.

- L'encadrement (9) est optimal, au sens où on peut se trouver dans des situations où il y a égalité dans l'une des deux inégalités proposées. Considérons par exemple une partition de Ω en n cellules C_j d'égales probabilités $1/n$, puis les n événements A_i définis par $A_i = \cup_{j \neq i} C_j$. Alors, $P(A_i) = \frac{n-1}{n}$ pour tout i, tandis que $\cap_i A_i = \emptyset$; ainsi $-\left(\frac{n-1}{n}\right)^n = e(A_1, A_2, ..., A_n)$. Considérons un événement A de probabilité $\left(\frac{1}{n}\right)^{\frac{1}{n-1}}$, puis les n événements A_i tous égaux à A. Alors, $e(A_1, A_2, ..., A_n) = P(A) - P(A)^n = \left(\frac{1}{n}\right)^{\frac{1}{n-1}} - \left(\frac{1}{n}\right)^{\frac{n}{n-1}}$. En utilisant la décomposition $\frac{n}{n-1} = 1 + \frac{1}{n-1}$, on constate que cette dernière expression n'est autre que $(n-1)\left(\frac{1}{n}\right)^{\frac{n}{n-1}}$.

- Le comportement limite quand $n \to +\infty$ des deux bornes dans (9) est facile à obtenir :

$$-\left(\frac{n-1}{n}\right)^n \searrow -\frac{1}{e} \text{ quand } n \to +\infty;$$
$$(n-1)\left(\frac{1}{n}\right)^{\frac{n}{n-1}} \nearrow 1 \text{ quand } n \to +\infty.$$

L'apparition de la constante e dans la première limite peut surprendre, mais ce n'est pas la première fois qu'elle apparaît dans une limite de calcul des probabilités (dans le "problème des rencontres" par exemple [3, page 39]).

Pour la bonne bouche, voici quelques valeurs numériques approchées de l'encadrement (9) pour n allant de 3 à 10 :

$$\begin{aligned}
n &= 3 : -0,2962 \leqslant \; ... \; \leqslant 0,7313 \\
n &= 4 : -0,3164 \leqslant \; ... \; \leqslant 0,8238 \\
n &= 5 : -0,3276 \leqslant \; ... \; \leqslant 0,8662 \\
n &= 6 : -0,3348 \leqslant \; ... \; \leqslant 0,8911 \\
n &= 7 : -0,3399 \leqslant \; ... \; \leqslant 0,9076 \\
n &= 8 : -0,3436 \leqslant \; ... \; \leqslant 0,9195 \\
n &= 9 : -0,3464 \leqslant \; ... \; \leqslant 0,9284 \\
n &= 10 : -0,3486 \leqslant \; ... \; \leqslant 0,9355.
\end{aligned}$$

Démonstration du Théorème 2.

Commençons par la partie facile, la majoration dans (9). Soit $x = P(A_1$ et A_2... et $A_n)$. Puisque $P(A_i) \geqslant x$ pour tout $i = 1, ..., n$, nous avons

$$e(A_1, A_2, ..., A_n) \leqslant x - x^n \leqslant \max_{x\in[0,1]} (x - x^n) .$$

Ce dernier maximum étant atteint en $\overline{x} = \left(\frac{1}{n}\right)^{\frac{1}{n-1}}$, on a

$$\max_{x\in[0,1]} (x - x^n) = \left(\frac{1}{n}\right)^{\frac{1}{n-1}} - \left(\frac{1}{n}\right)^{\frac{n}{n-1}} = (n-1)\left(\frac{1}{n}\right)^{\frac{n}{n-1}} .$$

Passons maintenant à la minoration dans (9). Voici la démonstration obtenue via H. Gianella (communication personnelle).

On a $e(A_1, A_2, ..., A_n) = x - [1 - P(A_1^c)] \times ... \times [1 - P(A_n^c)]$. L'astuce ici consiste à utiliser l'inégalité $\frac{u_1 + \ ... \ + u_n}{n} \geqslant (u_1 \times \ ... \ \times u_n)^{\frac{1}{n}}$ avec $u_i = 1 - P(A_i^c)$. Par ce biais,

$$e(A_1, A_2, ..., A_n) \geqslant x - \left[1 - \frac{1}{n}(P(A_1^c) + \ ... \ + P(A_n^c))\right]^n . \tag{10}$$

Mais $P(A_1^c) + \ ... \ + P(A_n^c) \geqslant P(\cup_{i=1}^n A_i^c) = P((\cap_{i=1}^n A_i)^c) = 1 - x$. Par conséquent, il vient de (10) :

$$e(A_1, A_2, ..., A_n) \geqslant x - \left(1 - \frac{1-x}{n}\right)^n . \tag{11}$$

La fonction de x apparaissant dans le membre de droite de (11) est une fonction croissante de x (facile à vérifier avec sa dérivée), sa valeur minimale est donc atteinte en $\overline{x} = 0$, elle vaut $-\left(1 - \frac{1}{n}\right)^n$. D'où l'inégalité annoncée $e(A_1, A_2..., A_n) \geqslant -\left(1 - \frac{1}{n}\right)^n = -\left(\frac{n-1}{n}\right)^n$.

Remerciements. Je remercie H. Gianella, P. Lassère, F. Barthe, ainsi que d'autres collègues de mon université qui ont réagi à mes questionnements sur ce sujet.

Références

1. S. Francinou, H. Gianella et S. Nicolas, *Oraux X-ENS Mathématiques*, vol. 6 (2022).

2. E. Kosmanek, *Mini-contribution à la l'étude de la dépendance probabiliste.* Magazine L'ouvert 83 (1996), $16 - 18$.

3. D. Foata et A. Fuchs, *Calcul des probabilités ; Cours et exercices corrigés.* Editions Masson (1996).

La présente note, dédiée à PLH, a été écrite dans les objectifs et valeurs que nous avons toujours partagés : celui de garder constamment un lien avec l'enseignement secondaire, et celui de diffuser et populariser les mathématiques.