DEUX PARADOXES DE THÉORIE DES PROBABILITÉS

Michel ÉMERY

(C.N.R.S. et Université de Strasbourg)

> Il faut éviter le paradoxe comme une fille publique qu'il est, avec laquelle on couche à l'occasion, pour rire, mais qu'un fou, seul, épouserait.
>
> *G. Courteline*

Le mot *paradoxe* peut revêtir des significations très diverses. Les deux petits paradoxes dont il va être question sont des énoncés mathématiques incontestables, mais qui peuvent, au premier abord, choquer l'intuition. Ils cessent donc d'être paradoxaux dès que l'on s'habitue à eux; ce sont d'ailleurs des résultats classiques (et élémentaires) de théorie des probabilités, que les lecteurs tant soit peu familiers avec celle-ci parcourront d'un œil blasé — de même, toutes proportions gardées, que des découvertes paradoxales en leur temps, comme l'existence d'irrationnels ou l'équicardinalité de $\mathbb{R}$ et $\mathbb{R}^n$ («je le vois, mais je ne le crois pas»), sont devenues aujourd'hui des évidences.

Les exercices en petits caractères sont des remarques ou digressions laissées au lecteur pour éviter d'alourdir le texte. Les deux paradoxes sont indépendants et peuvent être lus dans n'importe quel ordre; leur seul point commun est l'utilisation du petit lemme ci-dessous (dont on pourrait d'ailleurs très aisément se passer) :

LEMME. — *Si N est une variable aléatoire à valeurs dans $\mathbb{N}$, son espérance mathématique est la somme (finie ou infinie) de la série* $\sum_{k\in\mathbb{N}} \mathbb{P}[N > k]$.

Pour le vérifier, il suffit de calculer de deux façons différentes la somme des probabilités figurant dans le tableau triangulaire infini

$$\begin{array}{lllll}
\mathbb{P}[N{=}1] & & & & \\
\mathbb{P}[N{=}2] & \mathbb{P}[N{=}2] & & & \\
\mathbb{P}[N{=}3] & \mathbb{P}[N{=}3] & \mathbb{P}[N{=}3] & & \\
\mathbb{P}[N{=}4] & \mathbb{P}[N{=}4] & \mathbb{P}[N{=}4] & \mathbb{P}[N{=}4] & \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

En sommant ligne par ligne, on trouve $0\,\mathbb{P}[N{=}0] + 1\,\mathbb{P}[N{=}1] + 2\,\mathbb{P}[N{=}2] + \ldots$, c'est-à-dire l'espérance de N; en sommant d'abord chaque colonne, on obtient la série annoncée $\mathbb{P}[N > 0] + \mathbb{P}[N > 1] + \mathbb{P}[N > 2] + \ldots$

Le revers de la loi des séries. — Ce premier paradoxe est un phénomène bien connu dans l'étude des processus aléatoires ponctuels, et en particulier des processus de Poisson.

Dans la longue rue que j'habite, les voitures stationnent le long du trottoir en une file ininterrompue. Mon chien utilise ce trottoir pour sa promenade et ces voitures pour y lever la patte, mais il éprouve une telle aversion pour la couleur bleue qu'il arrête sa promenade, de part et d'autre de chez moi, à la première voiture bleue rencontrée. Son champ d'action est donc le plus grand intervalle de trottoir, autour de ma porte d'entrée, limité par deux voitures bleues. Sachant qu'une voiture sur dix est bleue, quelle est, en moyenne, la longueur de trottoir utilisable par Médor ?

Prenons comme unité de longueur les places de stationnement et supposons que chaque place a, indépendamment des autres, une chance sur dix d'être occupée par une voiture bleue. Pour éviter les effets de bord, supposons aussi la rue infiniment longue des deux côtés. Considérons une portion de rue centrée chez moi et de longueur L voitures, où L est très grand. Elle contient à peu près $L/10$ voitures bleues; les $9L/10$ autres voitures sont donc réparties dans les $L/10$ intervalles entre voitures bleues successives, en comptant les intervalles de longueur nulle. En divisant la somme $9L/10$ des longueurs de tous ces intervalles par le nombre $L/10$ de ces intervalles, on trouve 9 pour leur longueur moyenne.

On peut donc s'attendre à ce que la réponse soit 9, ou en tout cas de l'ordre de 9. Mais on négligerait alors une donnée essentielle (au moins à mes yeux) : le trottoir devant MA maison n'est pas n'importe quel trottoir ! Nous allons voir que cela fait une grosse différence : en réalité, l'intervalle entre les deux voitures bleues de part et d'autre de chez MOI est en moyenne de 18 voitures.

Chaque voiture est de couleur soit bleue (notée b), soit non-bleue. Appelons C_0, C_1, C_2, etc. les couleurs des voitures successivement rencontrées en partant de chez moi vers la gauche. Le numéro N de la première d'entre elles à être bleue, c'est-à-dire $N = \inf\{n \in \mathbb{N} : C_n = b\}$, est aussi la longueur disponible pour Médor à gauche de ma porte. La longueur totale disponible est $N + N'$, où N', longueur disponible à droite, a par symétrie même loi que N. La quantité à laquelle nous nous intéressons, l'espérance $\mathbb{E}[N+N']$, vaut $\mathbb{E}[N] + \mathbb{E}[N'] = 2\,\mathbb{E}[N]$; il nous reste à trouver l'espérance de N.

Commençons par calculer la probabilité pour que N excède une valeur donnée $k \geqslant 0$. Ceci a lieu si et seulement si aucune des couleurs C_0, C_1, ..., C_k n'est bleue, c'est-à-dire si et seulement si $C_0 \neq b$ et $C_1 \neq b$ et ... et $C_k \neq b$. En utilisant l'indépendance des C_i, la probabilité de cet événement est

$$\begin{aligned}\mathbb{P}[N > k] &= \mathbb{P}[C_0 \neq b \quad \text{et} \quad C_1 \neq b \quad \text{et} \quad \dots \quad \text{et} \quad C_k \neq b] \\ &= \mathbb{P}[C_0 \neq b] \times \mathbb{P}[C_1 \neq b] \times \dots \times \mathbb{P}[C_k \neq b] = \left(\frac{9}{10}\right)^{k+1} ,\end{aligned}$$

et, d'après le lemme, l'espérance vaut

$$\mathbb{E}[N] = \sum_{k=0}^{\infty} \mathbb{P}[N > k] = \sum_{k=0}^{\infty} \left(\frac{9}{10}\right)^{k+1} = \frac{9/10}{(1-9/10)} = 9 .$$

La longueur moyenne disponible à gauche de ma porte est donc 9 voitures, de même à droite, et la longueur totale est en moyenne 18.

EXERCICE. — Le calcul ci-dessus suppose qu'aucune voiture n'est garée juste devant ma porte : la première voiture vers la gauche et la première vers la droite ne sont pas la même. Vérifier que si une place de stationnement se trouve face à la porte, l'espérance n'est plus 18 mais 17,1.

D'où vient le désaccord entre ces deux résultats, 9 et 18, double l'un de l'autre? Les calculs sont bien entendu tous deux corrects, mais ils ne parlent pas de la même chose.

Pour le premier, nous avons pris $L/10$ voitures bleues et $9L/10$ non-bleues, nous les avons réparties au hasard dans les L places, et nous nous sommes intéressés à la moyenne des longueurs des $L/10$ intervalles entre deux bleues successives. En appelant $\mathcal{I}$ l'ensemble de ces $L/10$ intervalles et $\ell(i)$ la longueur de l'intervalle i, nous avons donc calculé la moyenne arithmétique

$$\frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \ell(i) .$$

Nous aurions d'ailleurs obtenu le même résultat en faisant la moyenne sur un ensemble beaucoup plus gros, prenant en compte toutes les répartitions possibles des $L/10 + 9L/10$ voitures dans les L places, plutôt que de tirer d'abord au sort l'une de ces répartitions.

Pour le deuxième calcul, nous avons fixé à l'avance un point, ma porte, et la longueur moyenne de l'intervalle enjambant ce point a été calculée sur toutes les répartitions possibles des $L/10 + 9L/10$ voitures. Il est clair que choisir un autre point aurait conduit au même résultat; on aurait d'ailleurs également pu tirer d'abord au sort une répartition, puis, cette répartition étant fixée, faire la moyenne, lorsqu'un point décrit toute la portion L, de la longueur de l'intervalle enjambant ce point. Par rapport à la moyenne arithmétique, dans laquelle tous les intervalles de $\mathcal{I}$ jouaient le même rôle, ce deuxième calcul avantage les intervalles plus longs, parce qu'ils contiennent plus de points, et fournit en fait la moyenne pondérée

$$\sum_{i \in \mathcal{I}} p(i)\, \ell(i) ,$$

où les coefficients de pondération p, de somme 1, sont eux-mêmes proportionnels aux longueurs des intervalles :

$$p(i) = \frac{\ell(i)}{\sum_{j \in \mathcal{I}} \ell(j)} .$$

De façon tout à fait analogue, le nombre moyen n de personnes vivant au foyer d'un Français pris au hasard est largement supérieur à l'effectif moyen m des foyers français, parce que m est la moyenne arithmétique des effectifs des foyers français, alors que n est la moyenne pondérée par les effectifs eux-mêmes des différents foyers (lors du calcul de n, un foyer de 5 personnes sera compté 5 fois).

EXERCICE. — Plus précisément, en désignant par v la variance des effectifs des foyers considérés, $n = m + v/m$.

Remarquez que si la répartition aléatoire des $L/10$ voitures bleues parmi L avait fourni le résultat périodique (excessivement improbable) où deux voitures bleues successives sont toujours séparées par exactement 9 non-bleues, l'ensemble $\mathcal{I}$ serait uniquement constitué d'intervalles de longueur 9, et les deux formules ci-dessus donneraient toutes deux 9. Mais contrairement à ce que peut dicter l'intuition, la répartition aléatoire fournit, outre des voitures bleues isolées et des intervalles de l'ordre de 9, bon nombre d'agrégats de voitures bleues, et quelques intervalles bien plus longs que 15; les valeurs proches de la moyenne, 9, ne sont pas du tout typiques, et il y a d'ailleurs environ 2,5 fois plus d'intervalles de longueur 0 que de longueur 9 (la valeur la plus probable est 0, puis 1, puis 2, etc.).
Vous avez reconnu la loi des séries. Supposons que l'on observe en moyenne une fois par siècle une explosion de supernova. En considérant ces phénomènes comme instantanés et indépendants, le modèle probabiliste employé pour décrire leur structure temporelle est ce qu'on appelle un processus de Poisson : si s est le nombre de secondes dans un siècle, on admet qu'à chaque seconde, indépendamment de ce qui se passe avant et après, on a une chance sur s d'assister à une supernova.[1] La situation est tout à fait semblable à celle vue plus haut, les secondes jouant le rôle des places de stationnement et la paramètre $1/10$ étant remplacé par $1/s$. Certains siècles observeront une supernova, mais d'autres jusqu'à trois ou quatre; et en compensation certains intervalles entre deux supernovæ dureront plusieurs siècles. À tout instant, que l'on soit en train d'en observer une ou que l'on n'en ait pas vu depuis 350 ans, l'espérance mathématique du temps à attendre jusqu'à la suivante est un siècle. Mais l'espérance de la durée qui séparera la dernière supernova avant l'an 4000 de la première à partir de l'an 4000 est, elle, de deux siècles, les deux morceaux d'intervalle avant et après 4000 y contribuant chacun pour un siècle. D'après ce que nous venons de voir, ce doublement de la durée moyenne de l'intervalle enjambant un instant donné s'explique qualitativement par la variabilité des durées des intervalles entre supernovæ. Venant en contrepartie de l'existence de petits intervalles, ce doublement n'est finalement que l'autre versant de la loi des séries.

Le paradoxe de Feller. — Il se trouve, sous le nom de « persistence of bad luck », dans le traité de W. Feller *An Introduction to Probability Theory and Its Applications* (Wiley, 1966), volume 2, chapitre I, § 5.
Rappelons qu'une variable aléatoire X est dite *diffuse* si, pour tout nombre x, la probabilité $\mathbb{P}[X{=}x]$ est nulle. Tel est le cas des variables gaussiennes, ou exponentielles, ou uniformes sur un intervalle, etc. ; au contraire, une variable aléatoire discrète, à valeurs dans $\mathbb{N}$ ou $\mathbb{Z}$, n'est jamais diffuse.
Vous effectuez une expérience aléatoire dont le résultat est une certaine variable aléatoire diffuse. Par exemple, au bureau de poste, vous mesurez la durée passée à attendre d'être servi (elle n'est diffuse qu'à condition d'admettre qu'en entrant, la

1. Pour être tout à fait rigoureux, il faudrait travailler non pas avec une durée fixe comme la seconde, mais avec une durée qui tend vers zéro. Le nom de processus de Poisson vient de ce que, pour tout intervalle de temps A, le nombre (aléatoire) de supernovæ observées pendant A suit la loi de Poisson ayant pour paramètre la durée de A exprimée en siècles.

probabilité est nulle de trouver tout de suite un guichet libre tenu par un employé disponible).
Vous faites refaire indépendamment la même expérience par un ami (vous l'envoyez donc dans le même bureau, à la même heure, un jour d'affluence comparable), puis par un autre, et ainsi de suite jusqu'à ce que l'un d'eux obtienne un résultat supérieur au vôtre (attende plus longtemps que vous). Cet ami est le N-ième; ceci définit une nombre aléatoire N, que l'on peut considérer comme un indicateur de votre malchance : plus N est grand, plus il a été difficile de trouver quelqu'un de plus malchanceux que vous. Quelle est l'espérance mathématique de la variable aléatoire N ? (Avant de lire la suite, essayez de deviner, puis de faire vous-même le calcul, qui se révèlera bien plus facile qu'il n'y paraît.)

Cette variable aléatoire est à valeurs dans $\mathbb{N}^* \cup \{\infty\}$, le cas $N = \infty$ signifiant qu'aucun de vos amis n'a obtenu un résultat dépassant le vôtre. Pour calculer la loi de probabilité de N, cherchons d'abord la probabilité pour que N excède un entier $k \geqslant 0$ donné. Dire que $N > k$ revient à dire que les k premiers à avoir tenté l'expérience après vous ont eu des résultats en-deçà du vôtre; ou encore que, parmi les $k+1$ premiers à tenter l'expérience (vous compris), c'est vous qui avez obtenu le résultat le plus grand. (C'est pour éviter les éventuels ex æquo que l'on a exigé une loi diffuse.) Mais ces $k+1$ personnes ont effectué indépendamment la même expérience aléatoire; par symétrie, chacune d'elle a donc précisément une chance sur $k+1$ d'avoir le plus grand résultat; et ceci établit la formule

$$\mathbb{P}[N > k] = \frac{1}{k+1} \,.$$

Exercice. — Cette expression ne dépend pas de la loi diffuse dont nous sommes partis. Comment aurait-on pu s'en douter a priori ? (Indication : si une variable aléatoire X est diffuse, la fonction $f(x) = \mathbb{P}[X < x]$ est croissante et la variable aléatoire $f(X)$ suit la loi uniforme sur $[0, 1]$.)

Cette formule entraîne que la probabilité $\mathbb{P}[N = \infty]$, majorée par $1/(k+1)$ pour tout k fini, est nulle, et N est en fait finie. Mais son espérance, donnée, grâce au lemme, par

$$\mathbb{E}[N] = \sum_{k=0}^{\infty} \mathbb{P}[N > k] = \sum_{k=0}^{\infty} \frac{1}{k+1} = +\infty \,,$$

est infinie! Ainsi, le nombre d'amis qu'il vous faudra tester pour en trouver un plus mal loti que vous est en moyenne infini, et vous pouvez à bon droit vous plaindre d'être le plus malchanceux des hommes. Naturellement, chacun de vos amis peut suivre le même raisonnement, sans même devoir recommencer l'expérience : il lui suffit d'utiliser les valeurs déjà obtenues et d'intervertir sa place et la vôtre dans la succession des tests; la valeur de N que lui obtient ainsi — et qui peut être différente de la vôtre — a, inévitablement, la même loi que la vôtre et la même espérance infinie.

À l'aide d'une machine disposant d'un programme qui choisit aléatoirement des nombres entre 0 et 1 (fonction « random »), ou à l'aide d'une table de nombres au hasard, il est très facile de simuler cette situation : un premier appel de cette fonction ayant fourni un résultat R, compter le nombre de nouveaux appels nécessaires pour obtenir une valeur dépassant R. On obtient ainsi une réalisation de la variable aléatoire N. C'est tout à fait faisable en pratique, et l'on y parvient raisonnablement vite, bien que le temps nécessaire, évidemment proportionnel à N, soit d'espérance infinie.

EXERCICE. — Pour simuler N plus rapidement, en n'appelant qu'une seule fois la fonction random, on peut prendre simplement la partie entière de $1/R$.

En répétant cette simulation, on peut fabriquer une suite de copies indépendantes N_1, N_2, ... de N. La loi des grands nombres affirme que la moyenne

$$\frac{N_1 + N_2 + ... + N_k}{k}$$

des valeurs observées tend vers l'infini quand k tend vers l'infini.

EXERCICE. — La moyenne arithmétique de N_1, ..., N_k tend vers l'infini, mais leur moyenne harmonique tend vers $6/(\pi^2-6)$.

EXERCICE. — Pour k grand, la variable aléatoire $k/\sup(N_1, ..., N_k)$ suit à peu près la loi exponentielle de densité $\exp(-x)$ pour $x > 0$ (cette loi est indépendante de k).

L'exercice précédent montre que, pour k grand, le plus grand des nombres N_1, ..., N_k est déjà de l'ordre de k ; la somme $N_1 + ... + N_k$ est, bien sûr, encore plus grande, mais pas tellement. De fait, sa croissance est à peu près en $k \ln k$, et la moyenne $(N_1+...+N_k)/k$ tend vers l'infini avec seulement une vitesse logarithmique. Cette moyenne évolue d'ailleurs de façon assez erratique, les très grandes valeurs des N_i se trouvant suffisamment grandes (ou suffisamment peu rares) pour contribuer au comportement, contrairement à ce qui se passe pour les lois des grands nombres à espérance finie. C'est par ses grandes valeurs, sporadiques mais pas très rares, que N a un comportement inhabituel, indigne des honnêtes variables aléatoires habituelles d'espérance finie.

Une illustration de ce phénomène est fournie par les records météorologiques, tels que : « Ce mois d'avril est le plus froid jamais enregistré en Basse-Normandie depuis que les archives météorologiques existent. » Très friands d'observations de ce type, les médias peuvent nous donner l'impression de vivre une époque exceptionnellement riche en catastrophes de tous genres. Et pourtant, sans avoir à prendre en compte une éventuelle évolution à long terme du climat, le paradoxe de Feller affirme qu'*il est normal, même si c'est peu intuitif, que de tels records soient souvent annoncés* : Considérons la température moyenne d'avril en Basse-Normandie cette année, l'année dernière, il y a deux ans, etc. En admettant que l'on puisse considérer ces données comme des variables aléatoires indépendantes, diffuses et de même loi, le nombre d'années qu'il faut remonter pour trouver un mois d'avril plus froid que celui de cette année est un nombre aléatoire N dont la loi et l'espérance sont celles calculées ci-dessus; il est donc beaucoup plus enclin à prendre de grandes valeurs que l'on ne s'y attend intuitivement. Ajoutez à cela la variété des situations que l'on peut considérer (la grêle la plus violente jamais observée à Bordeaux, la plus forte crue du Rhin, la température la plus haute pour un 12 juin, la plus longue période de gel ininterrompu à Lyon, le mois d'octobre le plus arrosé depuis 1934, la plus importante chute de neige à Strasbourg un jour de Noël, etc., etc., etc.), et vous serez peut-être à l'avenir un peu moins surpris par les sécheresses exceptionnelles et autres tornades du siècle. Je ne suis d'ailleurs pas loin de croire que la Météorologie Nationale utilise des programmes informatiques pour rechercher systématiquement les records battus par le jour, le mois ou l'année en cours, afin de les communiquer à la presse s'ils semblent suffisamment sensationnels!

Bien sûr, nul ne prétend que les records de chômage ou de pollution, par exemple, ne sont que des épiphénomènes probabilistes sans signification ni importance particulière. Mais le paradoxe de Feller nous enseigne qu'inversement, même en l'absence de tendance à long terme ou d'évolution sous-jacente, le jeu normal des fluctuations statistiques peut suffire à donner l'impression que nous traversons une période anormalement riche en records de toutes sortes, de chaleur et de froid, de pluie et de sécheresse.

ERRATA :

Dans l'article de C. MERCAT : "*Théorie des nœuds et enluminure celte*" ('*L'Ouvert*' n° 84 de septembre 1996, page 1)
il fallait lire :
Kelles (**VIIème siècle**) et non pas (XIIème siècle).