LES PROBABILITÉS NULLES

In: *«Bulletin de Sciences Mathématiques»*, Paris, 1936, pp. 275-288

LES PROBABILITÉS NULLES [1];

PAR M. BRUNO DE FINETTI.

Il n'y a pas de doute, ainsi que l'a remarqué M. Borel, et comme cela se trouve très clairement expliqué dans le traité de M. Lévy, que la notion d'événement possible et de probabilité nulle est purement théorique, car il s'agit en général d'événements définis comme des cas limites d'événements pratiquement observables, et leur vérification exigerait par conséquent une *infinité* d'expériences ou une expérience comportant une mensuration absolument *exacte*. On ne peut pas cependant se passer de les considérer et d'étudier les questions qui se posent à ce sujet, car une théorie mathématique ne pourrait être arrêtée au point où la correspondance entre les constructions logiques dont elle s'occupe et les notions pratiques qu'elles doivent représenter perd de sa clarté, de sa netteté : dans le cas contraire, on n'aurait pas le droit, par exemple, d'énoncer que le rapport de la diagonale d'un carré au côté est irrationnel, car aucune mensuration ne peut vérifier cette affirmation. Lorsqu'on considère un nombre aléatoire X, et que l'on dit que $\varphi(x)$ est la densité de probabilité, il est évident que la seule chose qui a un sens pratique c'est que

$$P(a, b) = \int_a^b \varphi(x)\, dx$$

donne à peu près la probabilité que X tombe dans un intervalle (a, b) suffisamment grand pour que les cas où la réponse sera douteuse, parce que X est pratiquement égal à l'un des deux extrêmes a ou b, aient une importance négligeable. Mais comment pourrait-on donner mathématiquement ces probabilités sans le

[1] Conférence faite le 8 mai 1935 à la Société mathématique de France.

FINETTI.

préciser, et sans que, par conséquent, on puisse en déduire l'expression de la probabilité pour des intervalles aussi petits que l'on veut, et conclure qu'à chaque point correspond un cas possible de probabilité nulle, et que la somme de tous ces cas de probabilité nulle est un événement certain ?

Si nous ne pouvons pas nous passer de considérer les probabilités nulles, nous devons aussi chercher à les étudier rigoureusement, et cela est bien difficile, car les principes à appliquer ne peuvent être établis que par une extension plus ou moins arbitraire ou incertaine à des cas idéalisés de raisonnements qui ont une valeur bien connue seulement dans les cas pratiques. La difficulté est plus grande encore car la définition même de probabilité chez les divers auteurs n'est pas la même et souvent est plus ou moins vague. Mais la discussion sur les diverses façons de concevoir la signification de la probabilité et de ses lois [1], nous amènerait trop loin de notre but, plus modeste mais plus déterminé, qui consiste dans l'éclaircissement des relations entre diverses propriétés qu'on pourrait vouloir admettre. Il s'agit de voir si elles sont ou ne sont pas compatibles entre elles : nous verrons qu'il y a des incompatibilités, et il s'ensuit qu'il faut choisir entre deux solutions distinctes.

*
* *

Indiquons avec $P(E)$ la probabilité d'un événement E quelconque, avec $P\left(\frac{E}{E'}\right)$ la probabilité de E subordonnément à E'.

Une notation qui s'explique d'elle-même permet alors de représenter les théorèmes des probabilités totales et composées sous la forme $P(E_1 + E_2) = P(E_1) + P(E_2)$ si E_1 et E_2 sont incompatibles, et $P(E_1 E_2) = P(E_1) . P\left(\frac{E_2}{E_1}\right)$. Les questions que nous examinons sont les suivantes :

1° est-ce qu'on peut penser $P(E)$ et $P\left(\frac{E}{E'}\right)$ définies pour chaque E, ou respectivement pour chaque couple E et E' ?

2° Lorsque $P_1(E)$, $P_2(E)$, ..., $P_n(E)$, ... est une suite de

(1) Ce sujet a été traité dans les conférences que j'ai tenues à la même époque (2-10 mai 1935) à l'Institut H. Poincaré (actuellement sous presse dans les *Annales* de cet Institut)

fonctions de probabilité telle que pour tout événement E la limite $P(E) = \lim_{n \to \infty} P_n(E)$ existe, $P(E)$ est-elle encore une fonction de probabilité ?

3° Le théorème des probalibités totales est-il applicable aussi pour un ensemble d'événements incompatibles en nombre infini ?

Cette dernière propriété, l'additivité *complète*, est équivalente à cette autre que, si $E_1, E_2, \ldots, E_n, \ldots$ est une suite décroissante d'événements $E_1 \supset E_2 \supset E_3 \supset \ldots \supset E_n \supset \ldots$, et si le produit infini $E_1 E_2 \ldots E_n \ldots$ est impossible, on a toujours $P(E_n) \to 0$.

Il suffit, comme l'avait fait substantiellement M. Cantelli [1], d'observer que

$$E_1 = (E_1 - E_2) + (E_2 - E_3) + (E_3 - E_4) + \ldots + (E_{n-1} - E_n) + E_n;$$

sous cette forme l'énoncé a été considéré par M. Kolmogoroff [2], qui désigne la propriété décrite *axiome de continuité*.

On voit facilement d'abord que l'additivité complète ne peut pas coexister avec la propriété que la limite d'une loi de probabilité soit encore une loi de probabilité.

Si nous avons une infinité dénombrable d'événements incompatibles $E_1, E_2, \ldots, E_h, \ldots$, et si nous posons, par exemple,

$$P_n(E_h) = a_n^{h-1}(1 - a_n),$$

avec $a_n = 1 - \frac{1}{n}$, on a bien $\sum_1^\infty {}_h P_n(E_h) = 1$ pour tout n, mais la loi limite

$$P(E_h) = \lim_{a \to 1} a^{h-1}(1 - a) = 0$$

---

(1) F. P. CANTELLI, *Una teoria astratta del calcolo delle probabilità* (*Giorn. Ist. Ital. Attuari*, A. III, n° 2, 1932). Le même raisonnement, sans modifications essentielles, est reproduit dans la conférence de M. Cantelli : *Sulla estensione del principio delle probabilità totali ad una successione illimitata di eventi incompatibili* (tenue le 29 avril 1935 à l'*Istit. Ital. Attuari* et publiée dans le *Giornale* de cet Institut, A. VI, n° 4, octobre 1935); la « justification » de l'additivité complète repose toujours sur l'acceptation implicite de l' « axiome de continuité », et ne signifie par conséquent rien de plus que l'*équivalence* entre les deux conditions.

(2) A. KOLMOGOROFF, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, *Ergebnisse der Math.*, II, 3, Berlin, Springer, 1933, p. 13.

ne satisfait pas cette relation, car

$$\sum_{1}^{\infty}{}_{h} P(E_h) = \Sigma_h 0 = 0.$$

Le même fait a lieu si l'on pose

$$P_n(E_h) = \frac{1}{n} \quad \text{pour} \quad h \leqq n, \qquad P_n(E_h) = 0 \quad \text{pour} \quad h > n,$$

ce qui correspond au procédé bien connu par lequel on définit ordinairement la probabilité dans le cas d'un *entier choisi au hasard*. S'il y a une infinité dénombrable de cas possibles, l'hypothèse que la limite d'une loi de probabilité variable soit toujours une loi de probabilité, conduirait donc à admettre la possibilité d'attribuer à tous ces cas une probabilité nulle, et cela contredit l'additivité complète car $0 + 0 + 0 + \ldots = 0 \neq 1$.

On parvient à des conclusions analogues mais beaucoup plus générales si l'on admet le concept dont nous allons voir qu'on peut le considérer comme le fondement de la théorie des probabilités nulles, c'est-à-dire que deux probabilités nulles ont toujours un rapport déterminé. C'est une locution générale de dire que l'on considère *également probables* certains cas, par exemple dans la détermination au hasard d'un point ou d'un autre sur un intervalle, mais cela signifie généralement que la *densité* est constante; il s'agit ici de voir si l'on peut définir, non cette limite du rapport entre les probabilités de deux voisinages, mais le rapport entre les probabilités de deux points considérés comme isolés. Pour cela on n'a qu'à appliquer la définition de probabilité subordonnée : si nous supposons que cette définition donne un sens à $p_1 = P\left(\frac{E_1}{E_1 + E_2}\right)$ et à $p_2 = P\left(\frac{E_2}{E_1 + E_2}\right)$, on a évidemment

$$P\left(\frac{E_1}{E_1 + E_2}\right) + P\left(\frac{E_2}{E_1 + E_2}\right) = 1 + P\left(\frac{E_1 E_2}{E_1 + E_2}\right) \geqq 1;$$

les deux probabilités p_1, p_2 ne peuvent donc s'annuler simultanément, et leur rapport, bien défini, peut être appelé le rapport entre les probabilités de E_1 et E_2. Ce rapport est fini si les deux probabilités subordonnées sont positives, et les deux probabilités nulles sont alors *commensurables;* si au contraire une seule des probabilités subordonnées est nulle, l'événement correspondant a une probabilité infiniment plus petite que l'autre. Une telle notion des probabilités nulles considérées comme des infiniment petits actuels a été envisagée par Lomnicki ([1]), qui appelle *poids* des coefficients proportionnels aux probabilités nulles pour des événements qui ont des probabilités commensurables.

Pouvons-nous supposer, lorsque les cas possibles constituent une infinité non dénombrable, que leurs probabilités soient commensurables? Évidemment non ([2]), si l'on admet l'additivité complète et la comparabilité entre les probabilités nulles : les *cas possibles* ayant des probabilités commensurables avec un cas donné constituent alors au plus une infinité dénombrable, et la somme de leurs probabilités (par rapport à une *unité* quelconque commensurable) est convergente.

([1]) A. Lomnicki, *Nouveaux fondements de la théorie des probabilités* (*Fund. Math.*, t. IV, 1923, p. 34-71). M. Paul Lévy n'accepte pas par contre la notion qui est à la base de cette conception. « S'il y a une notion, dit-il en effet, qui n'a pas par elle-même de sens précis, c'est celle de probabilité conditionnelle, dans le cas où la condition C qu'elle suppose vérifiée a elle-même une probabilité nulle. Si la probabilité $\nu = P(C)$ n'est pas nulle, la probabilité conditionnelle de A est le nombre qui, multiplié par ν, donne la probabilité de C et A. Mais si $\nu = 0$, la probabilité de C et A est nulle aussi, et la probabilité conditionnelle de A n'a par elle-même aucun sens » [*Sur la notion de probabilité conditionnelle* (*Bull. Sciences math.*, 1936); *voir* aussi *Sur l'application du théorèm ede Fubini au calcul des probabilités* (*Enseign. Math.*, 1935)]. Cette observation serait juste seulement si la notion de « probabilité conditionnelle » (ou subordonnée) était considérée comme notion dérivée, *définie* par la relation mentionnée $\nu x = p_{CA}$. Quiconque considère cette relation comme le « théorème des probabilités composées », et non pas comme « définition déguisée de la probabilité subordonnée », doit admettre que la notion de la probabilité subordonnée ait un sens intrinsèque connu, indépendamment de la relation $\nu x = p_{CA}$, et indépendamment donc de la circonstance que $\nu \neq 0$ ou $\nu = 0$, cette circonstance n'ayant pas un rôle dans la nature intrinsèque du concept, mais seulement par rapport à la relation $\nu x = p_{CA}$. On dira seulement que la connaissance des nombres ν et p_{CA} permet de calculer $x = p_{CA} : \nu$ si $\nu \neq 0$, tandis qu'il ne suffit pas de savoir que $\nu = 0$ (et donc $p_{CA} = 0$) pour déterminer la valeur de x dans le cas contraire.

([2]) Dans cette hypothèse, il y a évidemment une infinité d'événements également commensurables, c'est-à-dire de probabilités comprises entre p_1 et p_2, p_1 et p_2 étant commensurables, et cette infinité sera dénombrable ou contiendra une sous-classe dénombrable $E_1, E_2, \ldots, E_n, \ldots$. Subordonnément à

$$E = E_1 + E_2 + E_3 + \ldots + E_n + \ldots$$

les cas possibles sont les E_i et l'additivité complète n'est pas satisfaite, car leurs probabilités, étant également commensurables, sont évidemment nulles.

Si l'on admet encore que la probabilité doit pouvoir être définie pour tous les événements, on peut préciser le résultat en appliquant le théorème donné par Kuratowski-Banach [1], et généralisé ensuite par S. Ulam [2], qui a démontré la non-existence de fonctions d'ensemble complètement additives [3] dans un domaine infini non dénombrable de puissance $\mathfrak{m}$, s'il n'y a, parmi les nombres cardinaux $\leqq \mathfrak{m}$, aucun nombre cardinal inaccessible (*unerreichbar*) selon *Sierpinski-Tarski*. Ce théorème est donc démontré pour $\mathfrak{m} = \aleph_1$, $\mathfrak{m} = \aleph_2$, ..., $\mathfrak{m} = \aleph_\omega$, etc.; il l'est par conséquent pour $\mathfrak{m} = \mathfrak{c}$ (puissance du continu), si l'hypothèse du continu, ou même une hypothèse bien plus faible, subsiste; il a enfin une validité tout à fait générale si les nombres cardinaux inaccessibles, dont l'existence n'est pas prouvée, n'existent pas [4].

Pour plus de commodité nous développerons les considérations suivantes en les énonçant comme s'il était prouvé que le résultat d'Ulam subsiste pour $\mathfrak{m}$ nombre cardinal quelconque; il est sousentendu que, si la validité de ce résultat est plus restreinte, celle de nos considérations serait à restreindre conformément.

En précisant, nous partons de cette admission : si $\mathcal{E}$ est un ensemble non dénombrable, la seule façon d'y définir une fonction d'ensemble complètement additive, c'est de la définir sur un sousensemble $\mathcal{E}_0$ dénombrable, et de lui attribuer la valeur o en dehors de $\mathcal{E}_0$.

La probabilité sur un ensemble $\mathcal{E}$ non dénombrable de ces cas possibles (ou *points*) est alors distribuée de la façon suivante : les points qui ont une probabilité positive constituent un ensemble $\mathcal{E}_0$ fini ou dénombrable, et la somme (respectivement la série) des probabilités de ces points est 1; la probabilité ρ_1 de $\mathcal{E} - \mathcal{E}_0$ est donc nulle. Mais il existe des points de $\mathcal{E} - \mathcal{E}_0$ de probabilité commensurable avec ρ_1, et ils constituent un ensemble $\mathcal{E}_1$ fini ou dénombrable, et la somme (ou série) des probabilités de ces points,

---

(1) *Fund. Math.*, t. XIV.

(2) S. Ulam, *Zur Masstheorie in der allgemeinen Mengenlehre* (*Fund. Math.*, t. XVI, p. 140-150).

(3) Sauf la fonction nulle, et celles définies effectivement seulement dans un domaine partiel fini ou dénombrable (et nulles en dehors).

(4) J'ai remanié ce point d'après une observation dont je dois remercier M. Aronszajn et M. Kurepa, qui m'ont indiqué le mémoire d'Ulam que je ne connaissais pas.

par rapport à ρ_1 pris comme unité, est 1 ; la probabilité ρ_2 de $\mathcal{E} - \mathcal{E}_0 - \mathcal{E}_1$ est infiniment plus petite que ρ_1. L'itération transfinie du procédé conduit à décomposer $\mathcal{E}$ dans une suite bien ordonnée d'ensembles finis ou dénombrables

$$\mathcal{E}_0 \mathcal{E}_1 \mathcal{E}_2 \ldots \mathcal{E}_\omega \mathcal{E}_{\omega+1} \ldots \mathcal{E}_{\omega+\omega} \ldots$$

avec des probabilités $\rho_0 = 1$, ρ_1, ρ_2, $\ldots$, ρ_ω, $\rho_{\omega+1}$, $\ldots$ $\rho_{\omega+\omega}$, $\ldots$, chacune infiniment petite par rapport à la précédente; chaque ensemble $\mathcal{E}_\alpha$ est constitué d'un nombre fini ou dénombrable de points

$$E_\alpha^1 E_\alpha^2 \ldots E_\alpha^n \ldots$$

tels que les rapports p_α^i de la probabilité de E_α^i à ρ_α donnent

$$p_\alpha^1 + p_\alpha^2 + \ldots + p_\alpha^n + \ldots = 1.$$

En admettant ces hypothèses, le théorème des probabilités totales pourrait être étendu non seulement au cas d'une infinité dénombrable d'événements incompatibles, mais aussi au cas d'une infinité même non dénombrable : si E est un événement quelconque (si l'on préfère : un sous-ensemble quelconque de $\mathcal{E}$), la probabilité de E est la somme des probabilités de tous les points. Soit en effet α le premier nombre ordinal pour lequel E et $\mathcal{E}_\alpha$ ne sont pas incompatibles : E contient donc un ensemble fini ou dénombrable de points de $\mathcal{E}_\alpha$, qui a une probabilité $c\rho_\alpha$ ($0 < c \leqq 1$; l'égalité $c = 1$ subsiste seulement si E contient entièrement $\mathcal{E}_\alpha$); l'ensemble complémentaire $E - \mathcal{E}_\alpha$, étant contenu dans

$$\mathcal{E} - (\mathcal{E}_0 + \mathcal{E}_1 + \ldots + \mathcal{E}_\alpha),$$

a une probabilité $\leqq \rho_{\alpha+1}$ infiniment petite par rapport à ρ_α et donc $P(E) = c\rho_\alpha$. La probabilité est donc la somme des probabilités de tous les cas, la somme de celles qui sont infiniment petites d'ordre supérieur n'ayant d'ailleurs aucune influence.

Il faut remarquer qu'il n'en résulte pas cependant une échelle de probabilités infiniment petites définies par elles-mêmes : seulement les *rapports* sont bien définis. Si, dans deux problèmes distincts on a ρ_α et ρ'_α, aucune relation nécessaire n'a lieu entre les deux échelles.

Nous venons de voir que, si on admet l'additivité complète pour la somme d'une infinité dénombrable d'événements, et si l'on suppose la probabilité définie pour tout événement, les autres hypothèses que nous avons faites étendent l'additivité complète aux sommes d'une puissance quelconque : cela pourrait sembler commode d'un point de vue abstrait, mais pour les applications on aurait des difficultés frappantes. Un nombre aléatoire ne pourrait avoir qu'une distribution de probabilité discontinue, concentrée dans un ensemble fini ou dénombrable de points de probabilité positive, la probabilité nulle résiduelle serait distribuée de la même façon discontinue, et ainsi de suite pour les résidus successifs. Il serait bien étrange d'admettre que telle forme de distribution soit la plus générale possible, plus encore, si l'on observe que dans ce cas on ne pourrait considérer une suite d'épreuves indépendantes à pile ou face, car toutes les suites possibles ont, dans le cas de l'indépendance, une probabilité nulle !

Pour ce qui est du comportement de la fonction de répartition, et des conclusions pratiquement absurdes qui en ont été déduites, cela ne dépend pas de l'admission sur la comparabilité, mais seulement de celle relative à l'existence de la probabilité pour tout ensemble. L'additivité complète est donc pratiquement en contradiction avec toutes les autres propriétés considérées (dans le sens que leur union conduit, sinon à des absurdités, du moins à des conclusions pratiquement peu acceptables) ; pour qu'elle soit applicable sans difficulté on devrait se borner à un *corps borélien* d'événements suffisamment simples, peut-être, ce corps ne devrait être déterminé que par une infinité dénombrable d'événements, c'est-à-dire, être au plus semblable au corps des ensembles linéaires boréliens.

∴

Si par contre on abandonne l'additivité complète, et l'on considère l'additivité simple comme condition nécessaire et suffisante pour qu'une fonction $P(E)$ soit une loi de probabilité, la limite d'une loi variable est encore une loi de probabilité, car si $P_n(E_1 + E_2) = P_n(E_1) + P_n(E_2)$,

$$\lim_{n \to \infty} P_n(E_1 + E_2) = \lim_{n \to \infty} P_n(E_1) + \lim_{n \to \infty} P_n(E_2),$$

et aucun obstacle n'empêche que l'on puisse, du moins idéalement, en principe, considérer les probabilités de tous les événements comme bien déterminées et comparables entre elles, sans arriver, je crois, à aucune conclusion étrange. Rien n'empêche, en particulier, qu'un nombre infini d'événements incompatibles aient des probabilités également comparables, ou même des probabilités égales.

La somme d'un nombre infini de probabilités égales est évidemment infiniment plus grande que chaque terme; il faut observer cependant que si ρ' est la probabilité de la somme d'une infinité d'événements incompatibles de probabilité ρ il y a toujours une autre infinité de ces mêmes événements qui a une probabilité ρ'' infiniment plus petite que ρ'. Considérons en fait une suite d'événements $E_1 E_2 \dots E_n \dots$ parmi ceux dont la somme avait la probabilité ρ'; si la somme de cette suite a une probabilité infiniment plus petite que ρ', la démonstration est achevée, dans le cas contraire considérons la suite décomposée en deux, formée des éléments avec indice pair ou respectivement impair. Désignons par $E_i^{(1)}$ celle de deux suites de probabilité plus petite (pour éliminer toute ambiguïté, la suite paire si les deux probabilités sont égales); on aura donc

$$E_i^{(1)} = E_{2i} \quad \text{respectivement} \quad E_i^{(1)} = E_{2i-1}$$

selon que

$$P(\Sigma_i E_{2i}) \leqq P(\Sigma_i E_{2i-1}) \qquad \text{ou} \qquad P(\Sigma_i E_{2i}) > P(\Sigma_i E_{2i-1}).$$

Soit $\rho'_1 = P(\Sigma E_i^{(1)})$; si $\frac{\rho'_1}{\rho'} = 0$ la démonstration est achevée; dans le cas contraire $\left(0 < \frac{\rho'_1}{\rho'} \leqq \frac{1}{2}\right)$ définissons par le même procédé $E_i^{(2)}$, et ensuite $E_i^{(3)}, \dots, E_i^{(n)}, \dots$. Deux cas peuvent se vérifier : si après n répétitions du procédé on a $\frac{\rho'_n}{\rho'} = 0$ la démonstration est achevée; si $\frac{\rho'_n}{\rho'} > 0$ pour tout n, considérons la suite $E_i^{(i)}$, qui a une probabilité ρ'' plus petite que tout ρ'_n. Puisque $\frac{\rho'_n}{\rho'} \leqq \frac{1}{2^n}$, on a

$$\frac{\rho''}{\rho'} = 0 \qquad\qquad \text{C. Q. F. D.}$$

Si nous indiquons $E_i^{(i)} = E_i^{(\mathrm{I})}$ et $\rho'' = \rho^{(\mathrm{I})}$, puis définissons par le même procédé $E_i^{(\mathrm{II})}$ et $\rho^{(\mathrm{II})}$, $E_i^{(\mathrm{III})}$ et $\rho^{(\mathrm{III})}$, ... nous avons une suite $\rho^{(I)}$, $\rho^{(II)}$, $\rho^{(III)}$,... de probabilités dont chacune infiniment petite par rapport à la précédente; il existe cependant des suites, par exemple E_1, $E_2^{(I)}$, $E_3^{(II)}$, $E_4^{(III)}$, ... dont la probabilité est infiniment plus petite que toutes les $\rho^{(I)}$, $\rho^{(II)}$, $\rho^{(III)}$, ...

La probabilité d'un point, isolément, n'a évidemment rien à faire avec la densité; on peut donner (¹) des exemples tels que les suivants (les cas possibles étant les points de l'intervalle (0,1) : tous les points de (0,1) sont également probables, mais sur (0,1/2) la densité est 2, et nulle sur (1/2,1), ou vice versa, la densité est constante et la probabilité ρ' de chaque point de (0,1/2) est infiniment plus petite que la probabilité ρ'' de chaque point de (1/2,1) ou même, la densité est infiniment plus petite dans la moitié où la probabilité de chaque point est infiniment plus grande.

Une autre propriété qui cesse d'être vraie lorsqu'on abandonne l'additivité complète est celle qu'on peut appeler, comme je l'ai fait, propriété conglomérative (²), et qui consiste en ce que, si l'on a

$$P\left(\frac{E}{A}\right) = p \quad (\text{ou } \leqq p, \text{ ou } \geqq p)$$

lorsque A est un cas quelconque d'une certaine classe complète de cas possibles, alors la probabilité P(E) est encore p (resp. $\leqq p$, ou $\geqq p$) (³). Si par exemple dans un problème tous les nombres entiers sont également probables, la probabilité des nombres pairs est $\frac{1}{2}$ subordonnément à chacune des hypothèses (1, 2), (3, 4), (5, 6), ..., mais égale à $\frac{2}{3}$ subordonnément à chacune des hypothèses (1, 2, 4), (3, 6, 8), (5, 10, 12), ...; la propriété conglomérative nous conduirait donc à admettre à la fois que les nombres pairs ont une probabilité égale à $\frac{1}{2}$ et à $\frac{2}{3}$ (et évidemment à toute fraction que l'on veut), ce qui est absurde. Un paradoxe analogue qui contredit la même propriété a été donné par M. Borel : la probabilité que la latitude d'un point choisi au hasard sur une sphère

(¹) *Voir* ma note *Sulle probabilità numerabili e geometriche* (*Rend. R. Ist. Lombardo*, vol. LXI, 1928).

(²) *Voir* ma note *Sulla proprietà conglomerativa delle probabilità subordinate* (*Rend. R. Ist. Lombardo*, vol. LXIII, 1930).

(³) Cette propriété admise d'abord par M. Paul Lévy dans son traité, a été par la suite abandonnée dans un mémoire ultérieur.

soit comprise entre deux limites données n'a pas la même valeur qu'elle a subordonnément à l'hypothèse que le méridien soit donné.

Il y a beaucoup d'énoncés pratiquement importants qui devraient être modifiés si l'additivité complète était abandonnée [1]. Les phrases qui contiennent un passage à la limite, comme, dans l'énoncé de la loi forte, lorsqu'on parle de *tous les événements*, doivent être affaiblis, en spécifiant qu'il s'agit de *tous les événements d'un ensemble fini aussi grand que l'on veut*.

Les sauts d'une fonction de répartition ne représentent plus nécessairement la probabilité des points de discontinuité : si

$$\Phi(x - 0) = a < \Phi(x + 0) = b$$

la probabilité c des valeurs $< x$ et la probabilité d des valeurs $\leqq x$, peuvent être des valeurs quelconques $a \leqq c \leqq d \leqq b$, et non plus nécessairement $a = c$, $d = b$. En effet, supposons par exemple que les valeurs possibles soient

$$x - 1, \quad x - \frac{1}{2}, \quad x - \frac{1}{3}, \quad \ldots, \quad x - \frac{1}{n}, \quad \ldots,$$

et qu'elles soient également probables; il y a un saut pour $\xi = x$, car $\Phi(\xi) = 0$ si $\xi < x$, $\Phi(\xi) = 1$ si $\xi > x$, mais la probabilité de la valeur x est cependant nulle (elle est même impossible) : il n'y a donc pas une probabilité finie pour $\xi = x$, mais pour que ξ soit très voisin de x à gauche. Si le saut $\Phi(x + 0) - \Phi(x - 0)$ est, comme nous l'avons supposé, $b - a$, et s'il est décomposé arbitrairement en $b - a = (c - a) + (d - c) + (b - d)$, nous pourrons toujours construire un exemple analogue où $(c - a)$ et $(b - d)$ est la probabilité des valeurs très voisines de x à gauche ou respectivement à droite, et $(d - c)$ la probabilité de la valeur x exacte.

On ne peut même pas dire en général que pour une fonction de répartition on a nécessairement

$$\lim_{\xi \to -\infty} \Phi(\xi) = 0, \qquad \lim_{\xi \to +\infty} (\xi) = 1;$$

---

[1] *Voir* mes notes : *Sui passaggi al limite nel calcolo delle probabilità; A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili; Ancora sull' estensione alle classi numerabili del teorema delle probabilità totali* et deux réponses de M. Fréchet dans *Rend. R. Ist. Lombardo*, vol. LXIII, 1930.

il suffit que les deux limites soient $\geqq 0$ respectivement $\leqq 1$. Par exemple, pour un *nombre entier choisi au hasard,* la fonction de répartition est $\Phi(\xi) = 0$; avec les *infiniment petits actuels* on pourrait écrire $\Phi(\xi) = \rho E(\xi)$, avec $\rho =$ probabilité de chaque entier déterminé, $E(\xi) =$ partie entière de ξ pour $\xi > 0$, $E = 0$ pour $\xi < 0$, mais une telle forme n'est possible que lorsque les probabilités de tous les intervalles sont commensurables, ou, du moins, s'il existe un intervalle tel qu'aucun autre intervalle ne soit infiniment plus probable.

Plus en général, on ne peut plus affirmer que du fait que $\varphi(x)$ est la densité de probabilité, il en découle que la probabilité pour un ensemble soit donnée par l'intégrale de Lebesgue, dans tous les cas où elle existe; la seule conclusion légitime serait que la probabilité peut être une valeur quelconque comprise entre l'intégrale inférieure et l'intégrale supérieure dans le sens de Riemann (¹).

Un grand nombre de résultats devrait donc être spécifié et précisé d'une façon différente de l'ordinaire, si l'on abandonne l'additivité complète; dans certains cas il ne s'agirait d'ailleurs que d'exprimer avec plus de précision les hypothèses. Par exemple les paradoxes de la non-conglomérabilité s'évanouissent naturellement, comme l'observe M. Kolmogoroff, en remplaçant la subordination à une hypothèse isolée, par celle relative à la limite d'une décomposition finie; dans l'exemple cité de Borel, en considérant non plus la probabilité subordonnée à *un* méridien, mais la limite de la probabilité subordonnée à un fuseau qui comprend ce méridien et se rétrécit de plus en plus. Mais cela ne signifie pas que la conception des probabilités subordonnées aux cas de probabilité nulle doive être nécessairement considérée comme illégitime, et que le paradoxe ne soit pas effectif.

*
* *

Nous n'avons fait que présenter une liste d'*inconvénients* auxquels nous conduisent les diverses solutions du problème des

(¹) *Voir* B. de Finetti et M. Jacob, *Sull'integrale di Stieltjes-Riemann* (*Giorn. Ist. It. Attuari,* A. VI, nº 4, oct. 1935).

probabilités nulles. Du point de vue formel, on devrait *choisir*. Mais, lorsqu'on définit la probabilité, d'après un point de vue quelconque, mais bien déterminé, la réponse n'est nullement arbitraire : elle est une conséquence de la conception admise.

On ne saurait pas justifier, en effet, le procédé consistant dans l'admission, par convention, de certaines propriétés, relatives à des notions déjà définies ; du moins, on n'aurait pas alors le droit d'interpréter les conclusions comme contenant la notion en question dans son sens propre. Si l'on conclut, par exemple, en appliquant une propriété admise *par convention* que la probabilité d'un certain événement est nulle, nous n'avons pas le droit de penser que cet événement soit *presque impossible*.

Mais, comme nous l'avons déjà observé, il est difficile de donner des conclusions à ce sujet, qui soient bien démontrées, car la probabilité même n'est pas généralement définie d'une façon assez précise. Le point de vue subjectif, que je crois préférable, nous amènerait à *nier* l'additivité complète, car l'additivité sur les ensembles finis est suffisante à assurer la *cohérence* (1), qui est la seule condition à laquelle la probabilité soit assujettie. Mais aussi selon les théories qui conçoivent la probabilité comme la limite de la fréquence, on devrait admettre la même conclusion [comme il est observé explicitement par M. Kamke (2)]. En choisissant au hasard un entier N, la probabilité $P(N \varepsilon E_0)$ ou, plus simplement, $P(E_0)$, qu'il appartienne à la classe E_0 des nombres carrés $\left[\text{limite de la fréquence } \frac{1}{n} E(\sqrt{n}) \leqq \frac{1}{\sqrt{n}};\ E(x) = \text{partie entière de } x\right]$ est évidemment nulle, et l'est donc aussi la probabilité qu'il appartienne à la classe E_1 des nombres de la forme n^2+1, à la classe E_2 des nombres de la forme $n^2+2, \ldots$, à la classe E_k des nombres de la forme n^2+k (et pas n^2+h, avec $h<k$). On a donc

$$P(E_0)+P(E_1)+P(E_2)+\ldots+P(E_k)+\ldots = 0+0+\ldots+0+\ldots = 0,$$

tandis que

$$P(E_0+E_1+E_2+\ldots+E_k+\ldots) = 1,$$

(1) *Voir* mes conférences à l'Inst. H. Poincaré [citées en note (1), p. 276].

(2) E. Kamke, *Einführung in die Wahrscheinlichkeitstheorie*, Hirzel, Leipzig, 1932.

la somme des classes E_i comprenant évidemment tout nombre entier.

Je ne connais, par contre, aucune définition substantielle [1] qui ait été proposée pour la probabilité, et qui amènerait à la conclusion opposée : à la démonstration de l'additivité complète.

---

[1] M. Fréchet m'a fait l'objection que deux cas sont possibles : on peut admettre que la probabilité doit être définissable pour tous les événements, ou bien seulement pour des événements particuliers, que l'on pourrait appeler *probabilisables*, analogues aux ensembles *mesurables;* dans ce deuxième cas, l'additivité complète pourrait subsister.

Or, il va sans dire que l'additivité complète peut subsister dans le domaine des *événements probabilisables*, lorsqu'on appelle, par définition, *probabilisables*, les événements auxquels une fonction de probabilité peut être étendue grâce à l'hypothèse de l'additivité complète. Mais je ne saurais considérer que ce fait soit en contradiction avec mon affirmation, et qu'il constitue une démonstration de l'additivité complète. Il faudrait pour cela avoir comme point de départ une *conception* de la probabilité, justifier la distinction d'événements probabilisables et non probabilisables d'après cette conception et en faire découler la validité du théorème complet des probabilités totales dans le domaine des événements probabilisables. Si par contre on admet ladite propriété arbitrairement, en introduisant arbitrairement les restrictions nécessaires pour qu'elle ne soit pas absurde, on peut constater évidemment qu'elle est vraie pour toutes les lois de probabilité qu'on a considérées précisément parce qu'elles la vérifient, mais on n'a aucune raison de penser que les lois que l'on a voulu écarter n'étaient pas également admissibles.

Enfin la définition des événements probabilisables construite sur l'analogie avec les ensembles mesurables repose sur les notions d'ordre qui transforment l'*ensemble* des *cas possibles* dans un *espace*. Si, par exemple, les *cas possibles* ont la puissance du continu, et nous les représentons sur un segment de deux façons différentes, le même événement pourra résulter *probabilisable* selon la première image et *non probabilisable* selon l'autre (si la correspondance biunivoque entre les deux images ne transforme pas toujours ensembles mesurables en ensembles mesurables). La définition n'est donc pas intrinsèque, mais dépend du choix d'une image, à laquelle je ne vois pas comment on pourrait attacher une signification quelconque dans la théorie des probabilités.

(Extrait du *Bulletin des Sciences mathématiques*
2e série, t. LX, septembre 1936.)

105623 Paris. Imprimerie Gauthier-Villars, quai des Grands-Augustins 55.