

# ANNALES SCIENTIFIQUES DE L'É.N.S.

MAURICE FRÉCHET

**Une hiérarchisation des probabilités nulles déduite d'une  
formule d'Antoine Appert**

*Annales scientifiques de l'É.N.S. 3<sup>e</sup> série*, tome 82, n° 2 (1965), p. 377-385

[http://www.numdam.org/item?id=ASENS\\_1965\\_3\\_82\\_2\\_377\\_0](http://www.numdam.org/item?id=ASENS_1965_3_82_2_377_0)

© Gauthier-Villars (Éditions scientifiques et médicales Elsevier), 1965, tous droits réservés.

L'accès aux archives de la revue « Annales scientifiques de l'É.N.S. » (<http://www.elsevier.com/locate/ansens>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# UNE HIÉRARCHISATION DES PROBABILITÉS NULLES DÉDUITE D'UNE FORMULE D'ANTOINE APPERT

PAR M. MAURICE FRÉCHET.

---

INTRODUCTION. — Dans un Mémoire de 1936 <sup>(1)</sup> Antoine Appert donne, dès la troisième page, la définition d'un nombre  $N(E, r)$  qui est le maximum du nombre  $\nu$  tel qu'il existe  $\nu$  points de l'ensemble  $E$  dont les distances mutuelles sont supérieures à  $r$ . Et il observe que : « il est intuitivement très naturel de considérer la plus ou moins grande rapidité de croissance de cette fonction  $N(E, r)$  quand  $r$  décroît en tendant vers zéro comme mesurant d'une certaine manière l'étendue spatiale ou l'extension de l'ensemble  $E$  ».

Dans les pages suivantes de son Mémoire, Appert part de la définition de  $N(E, r)$  pour en déduire par deux transformations successives, une définition de la « mesure normale de  $E$  ».

C'est ici que nous nous écarterons d'Appert (en même temps que nous nous tiendrons pourtant de plus près aux idées exprimées par lui dans les phrases citées plus haut). Autrement dit, nous n'opérerons aucune transformation de la fonction  $N(E, r)$  pour définir une « probabilité généralisée » de  $E$  au moyen de la limite de  $N(E, r)$  quand  $r \rightarrow 0$ .

Puis nous comparerons les résultats obtenus avec ceux énoncés plus tard (en 1949) par Borel en partant d'une définition (implicite) tout à fait différente. Et nous donnerons quelques exemples où notre définition fournit les mêmes résultats que ceux de Borel.

On notera, toutefois, que l'introduction très importante par Appert de la fonction  $N(E, r)$  permet de définir la probabilité généralisée d'un

---

<sup>(1)</sup> *Mesures normales dans les espaces distanciés* (*Bull. Sc. math.*, Mélanges, t. 28, 1936, p. 329).

ensemble  $E$  appartenant à un espace distancié (dit aussi métrique) quelconque soit  $D$ . Tandis que Borel se restreint au cas des espaces euclidiens.

Il est donc bon de donner des exemples d'application de cette fonction dans un tel espace, ou plutôt de l'allure de la suite des valeurs de cette fonction quand  $r \rightarrow 0$ . Nous observons d'abord, que si  $E$  se réduit à un nombre fini  $\alpha > 1$ , de points de  $\mathcal{D}$ , on a évidemment

$$N(E, r) \leq \alpha.$$

D'autre part, si  $\delta$  est la plus petite des distances mutuelles de ces  $\alpha$  points, on a  $N(E, r) = \alpha$  quand on prend comme  $\alpha$  points de  $E$ , les  $\alpha$  points de  $E$ , pour  $r < \delta$ . Donc  $N(E, r) = \alpha$  pour  $r < \delta$  et par suite :

$$(1) \quad \lim_{r \rightarrow 0} N(E, r) = \alpha.$$

Lorsque  $E$  comporte un nombre infini de points, prenons-en arbitrairement  $\alpha$ , et soit  $E_\alpha$  leur ensemble.

Comme, en général,

$$N(F, r) \leq N(G, r) \quad \text{pour } F \subset G.$$

alors

$$\alpha = N(E_\alpha, r) \leq N(E, r)$$

d'où, d'après (1),

$$\lim_{r \rightarrow 0} N(E, r) \geq \lim_{r \rightarrow 0} N(E_\alpha, r) = \alpha$$

donc, ceci ayant lieu pour tout entier  $\alpha$ ,

$$\lim_{r \rightarrow 0} N(E, r) = \infty.$$

Cette conséquence permet d'énoncer une réciproque du résultat de (1).

Considérons le cas où l'on aurait

$$(2) \quad \lim_{r \rightarrow 0} N(E, r) = \beta,$$

$\beta$  étant un nombre fini; il résulte de ce qui précède que  $E$  ne peut être infini. Soit  $\alpha$  le nombre de ses points, on a vu que

$$\lim_{r \rightarrow 0} N(E, r) = \alpha.$$

Donc  $\beta$  est un entier égal à  $\alpha$ . Sous la condition (2),  $E$  n'a qu'un nombre fini de points et ce nombre est égal à  $\beta$ .

*Remarque.* — Quand  $E, F$  sont disjoints et ont chacun un nombre fini de points, on a évidemment

$$(3) \quad \lim_{r \rightarrow 0} N(E + F, r) = \lim_{r \rightarrow 0} N(E, r) + \lim_{r \rightarrow 0} N(F, r).$$

Si l'un des ensembles E, F est infini, par exemple E,  $E + F$ , le sera aussi et l'on aura

$$\lim_{r \rightarrow 0} N(E + F, r) = \infty + \lim_{r \rightarrow 0} N(F, r) = \lim_{r \rightarrow 0} N(E, r) + \lim_{r \rightarrow 0} N(F, r).$$

Ainsi la relation (3) est vérifiée dans tous les cas.

. — Considérons maintenant le cas d'un point aléatoire X dont toutes les positions possibles appartiennent à un ensemble H de points de l'espace distancié  $\mathcal{O}$ . Plaçons-nous dans le cas où la probabilité classique pour que X ne puisse prendre qu'une position  $a$  est nulle.

Intuitivement, puisque cette position n'est pas rigoureusement impossible, sa probabilité ne devrait pas être nulle, mais être considérée comme une probabilité *généralisée*  $\varepsilon$ , où  $\varepsilon$  est un « zéro généralisé » appartenant à un ensemble, posé *a priori*, de « zéros généralisés » formant une hiérarchie et correspondant cependant tous à une probabilité classique qui serait nulle.

Nous pouvons nous placer dans le cas où  $\varepsilon$  est le même pour tous les points de H.

Supposons maintenant que le principe des probabilités totales, s'applique aussi aux probabilités généralisées. Alors la probabilité généralisée de E est, quand E a un nombre fini,  $\alpha$ , de points, égale à la somme  $\varepsilon + \varepsilon + \dots + \varepsilon$  composée de  $\alpha$  termes. Nous la représenterons symboliquement par

$$\alpha \cdot \varepsilon \equiv \varepsilon \lim_{r \rightarrow 0} N(E, r).$$

Dès lors quand E est composé d'un nombre fini  $\alpha$ , de points, la probabilité généralisée que X appartienne à E sera représentée par

$$p_g(E) = \alpha \cdot \varepsilon.$$

Et si E comprend un nombre infini de points, sa probabilité généralisée sera représentée symboliquement par

$$(4) \quad p_g(E) = \infty \cdot \varepsilon.$$

Mais on peut aussi la préciser, *dans le cas* où E est un ensemble *dénombrable infini* de points  $a_1, a_2, \dots$  par la suite des nombres

$$(5) \quad 1 \cdot \varepsilon, 2 \cdot \varepsilon, \dots, n \cdot \varepsilon, \dots \quad (2).$$

On aura la même suite si l'on modifie l'ordre des points de E.

Ce qui amène à la conclusion que les *probabilités généralisées* de tous les ensembles *dénombrables* infinis de points de H *sont les mêmes*. Et qu'en

(2) Quand X n'a qu'une position possible,  $a_1$ , la probabilité de cette position est  $\varepsilon$ . Et  $N(E, r)$  n'ayant plus de sens puisqu'on ne peut parler de la distance des deux positions distinctes de E, on prendra naturellement avec Appert,  $N(a, r) = 1$  pour compléter naturellement la suite (5).

désignant par  $\omega$  le premier nombre transfini de Cantor ou indifféremment la suite  $1, 2, \dots, n, \dots$ , on pourra, pour le cas de chaque ensemble E dénombrable infini, représenter symboliquement sa probabilité généralisée par

$$p_g(E) = \omega \cdot \varepsilon$$

aussi bien que par la suite (5).

Il nous reste à préciser le sens de l'infini dans le cas où E comporte un ensemble *non dénombrable* de points.

Nous le ferons pour deux exemples où E est un ensemble linéaire et, en fin de ce Mémoire, un troisième où  $\mathcal{O}$  est un espace fonctionnel.

EXEMPLE DE L'ENSEMBLE DE CANTOR. — Considérons l'ensemble ternaire C de Cantor et désignons par  $C_s$ , l'ensemble des points

$$X = \frac{X_1}{3} + \dots + \frac{X_s}{3^s},$$

où chaque  $X_i$  est égal à 0 ou 2.

$N(C_s, r)$  où  $r > 0$  désigne le maximum des nombres  $\nu$  tels qu'il existe  $\nu$  points de  $C_s$  dont les distances mutuelles sont  $> r$ .

Pour chaque  $r > 0$  et  $< \frac{2}{3}$ , il existe une valeur de  $s$ , telle que

$$\frac{2}{3^{s+1}} \leq r < \frac{2}{3^s}.$$

En général, si  $r' < r''$ , on a évidemment

$$N(C_s, r'') \leq N(C_s, r').$$

Donc

$$N\left(C_s, \frac{2}{3^s}\right) \leq N(C_s, r) \leq N\left(C_s, \frac{2}{3^{s+1}}\right).$$

Il y a évidemment  $2^s$  points de  $C_s$  et leurs distances mutuelles ont un minimum  $\frac{2}{3^s}$  qui est atteint (quand on les range par ordre de grandeur) par chaque couple  $X'' > X'$  tel que

$$\left. \begin{array}{l} X' \\ X'' \end{array} \right\} = \frac{X_1}{3} + \dots + \frac{X_{s-1}}{3^{s-1}} + \left\{ \begin{array}{l} 0 \\ \frac{2}{3^s} \end{array} \right.$$

D'où

$$X'' - X' = \frac{2}{3^s} > r \left( \geq \frac{2}{3^{s+1}} \right)$$

et par suite :

$$(6) \quad N\left(C_s, \frac{2}{3^s}\right) < 2^s \leq N(C_s, r) \leq N\left(C_s, \frac{2}{3^{s+1}}\right).$$

Reste à évaluer  $N\left(C_s, \frac{2}{3^{s+1}}\right)$ .

Observons que s'il y a  $\nu$  points de  $C_s$  dont les distances mutuelles sont  $> \frac{2}{3^{s+1}}$ ,  $\nu$  peut être égal à  $2^s$  puisque si l'on considère  $2^s$  points de  $C_s$ , c'est-à-dire tous les points de  $C_s$ , le minimum de leurs distances mutuelles est  $\frac{2}{3^s} > \frac{2}{3^{s+1}}$ . Et comme  $\nu \leq 2^s$ , on voit que

$$N\left(C_s, \frac{2}{3^{s+1}}\right) = 2^s.$$

Finalement

$$(7) \quad N(C_s, r) = 2^s,$$

où  $s$  est un entier fonction de  $r$  déterminé par la relation ci-dessus

$$\frac{2}{3^{s+1}} \leq r < \frac{2}{3^s}.$$

Posons  $a = 3^s$ , on déduit de (6) et (7) que

$$N(C_s, r) = a^b, \quad \text{où } b = \frac{\log 2}{\log 3}.$$

La probabilité généralisée de  $C_s$  peut être représentée par la suite des valeurs de  $3^{s^b} \varepsilon$  où  $s = 1, 2, \dots$

On peut aussi la représenter symboliquement en remplaçant avec Borel,  $a$ , qui tend vers l'infini par  $\omega$ , le premier nombre transfini de Cantor. On aura alors

$$(8) \quad p_g(C) = \omega^b \varepsilon.$$

UN DES EXEMPLES DE BOREL. — Considérons l'ensemble considéré par Borel consistant dans l'ensemble  $B$  des points  $X$  de l'intervalle  $(0, 1)$  dont la représentation décimale

$$X = \frac{X_1}{10} + \dots + \frac{X_n}{10^n} + \dots \equiv 0, X_1, X_2, \dots, X_n, \dots$$

ne contient pas le chiffre 9<sup>(3)</sup>.

Soit alors  $B_n$  l'ensemble des nombres

$$X^{(n)} = \frac{X_1}{10} + \dots + \frac{X_n}{10^n},$$

où chaque  $X_i$  est l'un des 9 nombres 0, 1, ..., 8.

Nous supposons avec Borel que la répartition des probabilités sur  $(0, 1)$  est uniforme, c'est-à-dire que les  $X_i$  sont indépendants et que pour  $X$

---

<sup>(3)</sup> Borel fait la supposition que les  $X_i$  sont  $\neq 7$ , mais la suite des raisonnements reste parallèle et les résultats identiques.

sur  $(0, 1)$ , chaque  $X_i$  a la probabilité  $\frac{1}{10}$  d'être égal à l'un des nombres  $0, 1, \dots, 9$ .

On a

$$X^{(n)} - X'^{(n)} = \frac{\Delta X_1}{10} + \dots + \frac{\Delta X_n}{10^n}.$$

La plus petite valeur de ces distances des points  $X^{(n)}, X'^{(n)}$  distincts est obtenue pour  $\Delta X_1 = \dots = \Delta X_{n-1} = 0$  et  $\Delta X_n = \pm 1$ . On a par hypothèse

$$X^{(n)} \neq X'^{(n)}.$$

On peut donc prendre  $X^{(n)} > X'^{(n)}$  et dans le cas actuel  $\Delta X_n = 1$ .

La distance  $\frac{1}{10^n}$  est obtenue pour tout couple  $X^{(n)}, X'^{(n)}$  tel que

$$\left. \begin{matrix} X^{(n)} \\ X'^{(n)} \end{matrix} \right\} = \frac{X_1}{10} + \dots + \frac{X_{n-1}}{10^{n-1}} + \frac{1}{10^n} \left\{ \begin{matrix} 0 & \text{ou} & 1 & \text{ou} & 2 & \text{ou} & 3 & \text{ou} & 4 & \text{ou} & 5 & \text{ou} & 6 & \text{ou} & 7 \\ 1 & & 2 & & 3 & & 4 & & 5 & & 6 & & 7 & & 8 \end{matrix} \right\}.$$

D'après sa définition, tout  $N(E, r)$  est au plus égal au nombre de points de  $E$ . Pour  $B_n$  ce nombre est  $9^n$ .

Le nombre  $N(B_n, r)$  est évidemment ici :

$$N(B_n, r) = 9^n \quad \text{pour} \quad r < \frac{1}{10^n}.$$

Posons

$$\alpha = 10^n.$$

On aura

$$\begin{aligned} \log N(B_n, r) &= n \log 9, \\ \log \alpha &= n \log 10. \end{aligned}$$

D'où

$$\log N(B_n, r) = a \log \alpha' \quad \text{avec} \quad a = \frac{\log 9}{\log 10}.$$

D'où

$$(9) \quad N(B_n, r) = \alpha^a.$$

Quand  $n \rightarrow \infty$ ,  $r$  tend vers zéro et  $\alpha$  vers l'infini.

On peut alors représenter la probabilité généralisée de  $B$  par la suite des nombres  $10^n \varepsilon$ , où  $n \rightarrow \infty$ .

Ou, comme Borel, la représente symboliquement par la notation inspirée de (8) :

$$(10) \quad p_g(B) = \omega^a \varepsilon.$$

## CONSÉQUENCES DE LA SECONDE DÉFINITION DE BOREL.

*Définition de Borel* <sup>(4)</sup>. — Plusieurs années après M. Appert, Borel établit aussi une hiérarchisation des probabilités classiques nulles, suivant quatre définitions *non* équivalentes. Nous allons ici rappeler les conséquences de la *seconde*. Celle-ci, en effet, comme nous allons le voir, conduit à des formules semblables à celles que nous avons obtenues précédemment en modifiant les définitions d'Appert. Et pourtant les points de départ sont très différents. En particulier, celui d'Appert a l'*avantage* de s'appliquer, *sans aucune complication*, à un élément aléatoire appartenant à *n'importe quel espace distancié* (dit aussi métrique).

Au contraire, Borel considère seulement un élément aléatoire  $X$  qui est un nombre compris entre zéro et un, avec distribution uniforme des probabilités.

*Exemples de Borel.* — On peut écrire dans le système décimal

$$X = 0, x_1, \dots, x_n, \dots$$

I. Si l'ensemble  $E$  est réduit à un point  $C = 0, c_1, \dots, c_n, \dots$ , la probabilité que  $x_i = c_i$  est  $\frac{1}{10}$  et la probabilité  $P_n$  que  $0, x_1, \dots, x_n = 0, c_1, \dots, c_n$  est  $\frac{1}{10^n}$  comme le précise Borel qui représente symboliquement par  $\frac{1}{\omega}$  la probabilité que  $X = C$ .

II. Quand  $E$  se réduit à  $\alpha$  points, la probabilité  $P$  pour que  $0, x_1, \dots, x_n$  soit égal à l'un des  $\alpha$  nombres  $0, c_1^1, \dots, c_n^1, \dots; 0, c_1^\alpha, \dots, c_n^\alpha$  sera  $\frac{\alpha}{10^n}$ . Et Borel représenterait symboliquement  $P$  par  $\frac{\alpha}{\omega}$ .

III. Quand  $E$  est l'ensemble des points pour lesquels tous les  $x_i$  sont  $\neq 9$ , la probabilité  $P_n$  pour que  $x_1, \dots, x_n$  soient  $\neq 9$  sera

$$\frac{9^n}{10^n} = \varepsilon_n^{1-a},$$

avec

$$\varepsilon_n = \frac{1}{10^n}, \quad a = \frac{\log 9}{\log 10}.$$

La probabilité généralisée de  $E$  sera représentée par la suite des

$$P_n = \frac{\varepsilon_n}{\varepsilon_n^a}$$

<sup>(4)</sup> On trouvera plus de détails sur cette définition dans notre Note : *C. R. Acad. Sc.*, t. 252, 1962, p. 2033 et dans notre article : *Sur une des classifications des probabilités nulles dues à Émile Borel (Bull. Sc. math., t. 88, 1965, p. 113-127)*.



ou symboliquement par

$$P \sim \frac{\omega^a}{\omega}$$

et la raréfaction de E par

$$R = \frac{\omega}{\omega^a}$$

Quand E est l'ensemble ternaire de Cantor, on verrait de même que

$$P \sim \frac{\omega^b}{\omega}, \quad R = \frac{\omega}{\omega^b}, \quad \text{où } b = \frac{\log 2}{\log 3}.$$

On retrouve ainsi les mêmes formules que les formules (1), (2 bis), (8), (9), toutefois en remplaçant dans nos formules  $\varepsilon$  par  $\frac{1}{\omega}$ . Nous préférons d'ailleurs séparer ces notations, car dans la nôtre,  $\varepsilon$  a un sens précis, qui disparaît dans  $\frac{1}{\omega}$ .

*Remarque générale.* — Les formules de Borel ont été obtenues en choisissant une variable convenable  $\frac{1}{10^n}$ , (par exemple, à la page 383). Comme nous l'avons déjà fait observer ailleurs, ses formules changeraient si l'on prenait, par exemple  $\frac{1}{n}$ , comme variable.

C'est pour mettre plus en évidence la similitude de nos résultats et de ceux de Borel que nous avons opéré de même.

#### CAS D'UN ESPACE FONCTIONNEL.

*Exemple.* — Sans prétendre donner l'exemple le plus simple où E est non dénombrable, dans un espace non linéaire, nous avons songé à l'exemple qui va suivre.

Soit un rectangle R :

$$a \leq x \leq b, \quad c \leq y \leq d.$$

Prenons d'abord pour espace  $\mathcal{O}$ , l'ensemble H des fonctions  $f(x)$  définies et continues pour  $a \leq x \leq b$  et dont les valeurs sont comprises entre  $c$  et  $d$ .

On pourra indifféremment désigner par H l'ensemble correspondant des courbes continues comprises dans le même rectangle, R, et dont chacune n'est coupée qu'en un seul point par une parallèle à Oy, d'abscisse comprise entre  $a$  et  $b$ .

Et appelons distance de deux fonctions  $f(x)$ ,  $g(x)$  de H, le maximum ( $f$ ,  $g$ ) de  $|f(x) - g(x)|$  sur  $(a, b)$ .

Comme dans le cas général examiné plus haut, nous considérons un choix aléatoire de  $X = F(x)$  et nous supposons que la probabilité clas-

sique que  $X$  soit une fonction déterminée  $f_0(x)$  de  $H$  soit nulle mais que la probabilité généralisée correspondante soit un « zéro généralisé »,  $\varepsilon$ , indépendant de  $f_0$ . Enfin que le principe des probabilités totales soit applicable aux probabilités généralisées.

Ceci étant, cherchons la probabilité généralisée que  $X$  appartienne à un certain ensemble  $E$  que nous allons définir.

Et, pour cela, nous étudierons d'abord  $N(E, r)$ . Commençons par la définition de  $E$ . A cet effet, considérons  $n$  nombres  $x_i$ , avec

$$a \leq x_1 < x_2 < \dots < x_n \leq b.$$

Puis, portons sur l'ordonnée  $x = x_1$  un ensemble non dénombrable,  $e_1$  de mesure nulle de points d'ordonnées comprises entre  $c$  et  $d$ . Déplaçons  $e_1$  par une translation arbitraire éventuellement oblique, et portons-le ainsi sur un ensemble égal  $e_i$  porté sur l'ordonnée  $x = x_i$  et encore compris entre  $y = c$  et  $y = d$ . Chaque point  $b_i$  de  $e_1$  viendra en un point  $b_i$  de  $e_i$ .

Il y a une ligne brisée  $f_b$ , joignant successivement les points de coordonnées  $(x_i, b_i)$ . Elle n'est coupée qu'en un seul point par toute ordonnée et est comprise dans le rectangle  $R$ . Elle est donc représentée par une des fonctions, soit  $f_b$ , de  $H$ .

L'ensemble indéfini,  $E$  sera constitué par toutes les fonctions  $f_b$ .

Il est clair que la distance  $(f_b, f_c)$  définie plus haut est égale ici à la distance des points  $b_i, c_i$  de  $e_1$ .

Dès lors  $N(E, r)$  est égal au nombre maximal des points de l'ensemble linéaire  $e_1$  situé sur l'ordonnée  $x = x_1$  et dont les distances mutuelles sont  $> r$ .

On se trouve ainsi ramené pour étudier

$$\lim_{r \rightarrow 0} N(E, r),$$

où  $E$  est un ensemble de *fonctions*, à l'étude de

$$\lim_{r \rightarrow 0} N(e_1, r),$$

où  $e_1$  est un ensemble linéaire, étude qui a été entreprise plus haut.

On pourrait d'ailleurs généraliser de bien des manières l'ensemble  $E$ , par exemple en supprimant la condition que les  $e_i$  soient égaux moyennant son remplacement par des conditions moins restrictives convenablement choisies.

