La méthode probabiliste

Marie Heyvaert[1], F. Thomas Bruss[2]

1. Introduction

La *méthode probabiliste* désigne l'utilisation de raisonnements probabilistes dans la résolution de problèmes purement déterministes. Nous partons d'un problème qui n'a rien d'aléatoire. Nous lui associons un espace de probabilités judicieusement choisi. C'est en travaillant à l'intérieur de celui-ci que nous aboutirons à une conclusion qui pourra fournir une solution au problème non probabiliste de départ.

Cette méthode s'est révélée très efficace dans l'établissement de nombreux théorèmes tant en théorie des graphes qu'en combinatoire ou en théorie des nombres. On trouve également des résultats obtenus par cette méthode en théorie des jeux, en théorie de la complexité, en géométrie, et en analyse.

Dans la section 2, nous décrivons des techniques souvent utilisées dans les démonstrations via la méthode probabiliste. Ensuite, la section 3 présente une série d'exemples piochés dans plusieurs domaines où les techniques probabilistes ont été appliquées avec succès. La plupart de ces exemples illustrent les arguments de la section 2. Mais il y en a aussi qui font appel à d'autres raisonnements. La démonstration de l'égalité dite *égalité min-max* en utilisant le principe d'inclusion-exclusion, est, à notre connaissance, neuve.

La méthode probabiliste peut également servir à fournir des arguments heuristiques (non rigoureux) pour conforter les mathématiciens dans l'idée qu'une certaine conjecture est vraie. Nous aborderons cet aspect dans la section 4. En fait, nous tenterons d'appliquer la méthode à deux grands problèmes de la théorie des nombres : la conjecture de Goldbach et le théorème de Green-Tao. Nous leur associerons un modèle probabiliste. Cependant, contrairement aux exemples de la section 3, nous ne pourrons pas prouver, ici, que le résultat obtenu dans le contexte probabiliste est encore vrai dans le contexte initial car nous n'avons pas de preuve que le modèle choisi décrit correctement la situation réelle.

Nous nous sommes principalement basés sur le livre de Alon et Spencer [1] et sur celui de Diestel [5] pour rédiger la section 3 de ce travail. Quant à la section 4, elle est inspirée de notes non publiées rédigées par le deuxième auteur.

Le but de ce travail est de donner un bref aperçu de ce qu'il est possible de faire avec la méthode probabiliste, d'exposer ses concepts de base et de montrer la diversité des questions qu'elle peut traiter. Ce qui nous intéresse dans toutes les démonstrations qui suivent, c'est en effet plus la manière de procéder que les

[1] Université Libre de Bruxelles
[2] Université Libre de Bruxelles

résultats obtenus. Nous voulons souligner la puissance et l'élégance qui se cachent derrière les calculs. En exprimant un problème dans le langage probabiliste, nous pouvons gagner en simplicité. Dans l'histoire des mathématiques, la méthode probabiliste est une technique de preuve relativement récente. Sans doute est-elle loin de la fin de ses possibilités.

2. Deux techniques clés

2.1. La technique de base

Le principe de base de la méthode probabiliste peut être décrit de la manière suivante : pour prouver l'existence d'une structure combinatoire jouissant de certaines propriétés, nous construisons un espace de probabilités approprié au problème et nous montrons qu'un élément choisi de façon aléatoire dans cet espace possède les propriétés désirées avec une probabilité strictement positive.

Remarquons que, dans certains ouvrages, c'est ce principe-ci qui porte le nom de *méthode probabiliste*. Sa première apparition remonte à l'année 1943 avec la publication, par Szele, d'un résultat sur l'existence de graphes complets orientés contenant un grand nombre de chemins hamiltoniens.[3] Cependant, c'est Paul Erdös qui est considéré comme le véritable promoteur de cette méthode. En introduisant des probabilités là où personne n'en attendait, il démontra un grand nombre de théorèmes provenant principalement de la théorie des graphes. Depuis, cette approche s'est étendue à d'autre domaines.

2.2. L'argument de l'espérance

Un deuxième argument probabiliste est fréquemment utilisé pour garantir l'existence d'une structure vérifiant une propriété particulière. Il fait intervenir l'espérance d'une variable aléatoire. Le voici :

Soit X une variable aléatoire définie sur un espace de probabilités (Ω, A, P). Si l'espérance $E[X]$ de X est égale à μ, alors il existe un point ω_1 dans Ω pour lequel

$$X(\omega_1) \leqslant \mu$$

et, de même, il existe un point ω_2 dans Ω pour lequel

$$X(\omega_2) \geqslant \mu.$$

La démonstration, par l'absurde, est évidente. Le meilleur moyen d'illustrer ces techniques de preuves est bien entendu de les appliquer à des exemples concrets. Quelques-uns sont présentés ci-dessous. Le lecteur intéressé en trouvera d'autres dans le livre de Alon et Spencer [1].

[3] La démonstration de Szele est donnée dans le livre de Alon et Spencer [1], à la page 15.

3. Diverses applications

3.1. Nombres de Ramsey

Commençons par exposer un résultat de la théorie des graphes. Nous désignons par K_n le graphe complet à n sommets et par V son ensemble de sommets.

Définition 1 (Nombre de Ramsey)**.** *Le* nombre de Ramsey $R(k,l)$ *est le plus petit entier n tel que pour toute coloration des arêtes de K_n en deux couleurs, disons bleu et rouge, il y ait soit un K_k rouge (c'est-à-dire un sous-graphe complet à k sommets dont toutes les arêtes sont coloriées en rouge), soit un K_l bleu.*

Ramsey a montré en 1930 que ces nombres sont tous finis. En 1947, Erdös utilise les probabilités pour obtenir des bornes inférieures pour les nombres de Ramsey diagonaux $R(k,k)$. C'est le premier problème qu'il traite avec sa méthode probabiliste.

Théorème 1 (Erdös, 1947)**.** *Si les entiers positifs n et k satisfont à l'inégalité* $\binom{n}{k}2^{1-\binom{k}{2}} < 1$*, alors* $R(k,k) \geqslant n$*.*

Démonstration. Tout d'abord, nous remarquons que prouver ce théorème revient à montrer l'existence d'une coloration des arêtes de K_n sans K_k rouge ni K_k bleu.

Supposons que chaque arête est coloriée soit en bleu, soit en rouge avec une probabilité $\frac{1}{2}$ et ceci indépendamment des autres arêtes. Formellement, nous venons de créer un espace de probabilités $(\Omega, \mathcal{A}, P)$ tel que les éléments de Ω sont les $2^{\binom{n}{2}}$ colorations possibles des arêtes de K_n en deux couleurs. Ces éléments sont tous équiprobables.

Pour tout ensemble S de k sommets, soit A_S l'événement que le sous-graphe de K_n induit par S est unicolore. Clairement, pour chaque S,

$$P(A_S) = 2\left(\frac{1}{2}\right)^{\binom{k}{2}} = 2^{1-\binom{k}{2}}.$$

Soit B l'événement qu'il y ait au moins un K_k unicolore dans une coloration de K_n. Alors

$$P(B) = P\left(\bigcup_{\substack{S\subseteq V\\ |S|=k}} A_S\right) \leqslant \sum_{\substack{S\subseteq V\\ |S|=k}} P(A_S) = \binom{n}{k}2^{1-\binom{k}{2}}$$

et le dernier terme est strictement inférieur à 1 par hypothèse. Par conséquent,

$$P(B^c) = 1 - P(B) > 0.$$

Cela assure l'existence d'un point de notre espace de probabilité pour lequel l'événement B^c se réalise. Autrement dit, il existe une coloration de K_n pour laquelle il n'y a aucun K_k unicolore. Ceci termine la preuve. □

Ce premier exemple est une application directe de la méthode probabiliste de base décrite dans la section 2. Notons que cette méthode est *non constructive*. Aucune indication n'est fournie quant à la manière d'obtenir une telle coloration sans K_k unicolore.

Corollaire 2. *$R(k,k) > 2^{\frac{k}{2}}$ pour tout $k \geqslant 3$.*

Démonstration. Le résultat découle directement du Théorème 1 en choisissant $n = \lfloor 2^{k/2} \rfloor$. En effet, dans ce cas, on obtient :

$$\binom{n}{k} 2^{1-\binom{k}{2}} \;<\; \frac{n^k}{k!} 2^{1-\frac{k(k-1)}{2}} \;=\; \frac{2^{1+k/2}}{k!} \frac{n^k}{2^{k^2/2}} \;<\; 1.$$

□

Remarquons encore que $R(2,2) = 2$. Nous allons maintenant montrer que le résultat du Théorème 1 peut être amélioré. Cette fois, c'est la deuxième technique de la section 2 qui nous servira.

Théorème 3. *$R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$ quels que soient les entiers n et k.*

Démonstration. Le début de la preuve est identique à celui de la preuve du Théorème 1. À nouveau, nous considérons une coloration aléatoire des arêtes de K_n obtenue en coloriant indépendamment chaque arête soit en bleu, soit en rouge. Chaque couleur a une probabilité $\frac{1}{2}$ d'être choisie.

Pour tout ensemble S de k sommets, soit X_S la variable aléatoire indicatrice définie par

$$X_S = \begin{cases} 1 & \text{si le sous-graphe de } K_n \text{ induit par } S \text{ est unicolore} \\ 0 & \text{sinon.} \end{cases}$$

Soit X la variable aléatoire qui compte le nombre de K_k unicolores dans une coloration de K_n.

$$X = \sum_{\substack{S \subseteq V \\ |S| = k}} X_S$$

En calculant l'espérance de X, nous obtenons

$$\begin{aligned} E[X] &= \sum_{\substack{S \subseteq V \\ |S| = k}} E[X_S] \\ &= \sum_{\substack{S \subseteq V \\ |S| = k}} P(\text{le sous-graphe de } K_n \text{ induit par } S \text{ est unicolore}) \\ &= \mu \end{aligned}$$

où $\mu = \binom{n}{k} 2^{1-\binom{k}{2}}$. Ceci certifie l'existence d'une coloration des arêtes de K_n pour laquelle $X \leqslant \mu$. Fixons une telle coloration. Ensuite, supprimons de K_n un sommet dans chaque K_k unicolore. Au plus μ sommets sont supprimés. Après leur suppression, il nous reste un graphe complet d'au moins $n - \mu$ sommets. Plus aucun K_k unicolore n'apparaît dans sa coloration.

Nous pouvons alors en conclure que $R(k,k) > n - \mu$. □

Il nous semble intéressant de souligner une particularité du raisonnement qui précède : il utilise le *principe de l'altération* qui existe dans d'autres démonstrations via la méthode probabiliste. Dans un premier temps, un argument probabiliste

prouve l'existence d'une structure (ici, une coloration) qui n'a pas toutes les propriétés souhaitées mais qui possède quelques *défauts*. Ensuite, grâce à de petites modifications, les défauts sont effacés pour laisser apparaître la structure désirée.

3.2. Ensemble dominant de sommets dans un graphe

Voici encore un résultat concernant la théorie des graphes.

Définition 2 (Degré minimum)**.** *Soit $G = (V, E)$ un graphe simple (non dirigé). Nous désignons l'ensemble des voisins d'un sommet v de G par $N(v)$. Le nombre $d(v) := |N(v)|$ est le* degré *du sommet v et $\delta(G) := min\{d(v) \mid v \in V\}$ est le* degré minimum *de G.*

Définition 3 (Ensemble dominant)**.** *Dans un graphe $G = (V, E)$, un ensemble de sommets $U \subseteq V$ est appelé dominant si tout sommet $v \in V \backslash U$ possède au moins un voisin dans U.*

Théorème 4. *Si $G = (V, E)$ est un graphe de n sommets dont le degré minimum est $\delta > 1$, alors G a un ensemble dominant d'au plus $n\left(\frac{1+ln(\delta+1)}{\delta+1}\right)$ sommets.*

Démonstration. Nous commençons par définir un espace de probabilités à l'intérieur duquel nous travaillerons. L'idée est la suivante : posons $p = \frac{\ln(\delta+1)}{\delta+1}$, choisissons aléatoirement et indépendamment chaque sommet de V avec une probabilité p et considérons X l'ensemble aléatoire formé de tous les sommets sélectionnés.

Formellement, ceci signifie que nous travaillons dans un espace $(\Omega, \mathcal{A}, P)$ où Ω est l'ensemble des sous-ensembles X de V, $\mathcal{A}$ est l'ensemble des parties de Ω et où la mesure de probabilité P est définie de façon à ce que les événements $[v \in X] \equiv \{X \in \Omega \mid v \in X\}$ soient indépendants avec $P(v \in X) = p$ pour tout $v \in V$. L'ensemble X considéré est un élément quelconque de cet espace.

Soit Y_X l'ensemble des sommets de $V \backslash X$ qui n'ont aucun voisin dans X. L'espérance de $|X|$ est clairement égale à np. La variable aléatoire $|Y_X|$ peut s'écrire comme une somme de n variables indicatrices χ_v $(v \in V)$ où

$$\chi_v = \begin{cases} 1 & \text{si } v \in Y_X \\ 0 & \text{sinon.} \end{cases}$$

Pour chaque sommet $v \in V$,

$$\begin{aligned} P(v \in Y_X) &= P(v \text{ et ses voisins ne sont pas dans } X) \\ &= (1-p)^{1+d(v)} \\ &\leqslant (1-p)^{1+\delta}. \end{aligned}$$

Nous en déduisons que

$$\begin{aligned} E[|X| + |Y_X|] &= np + \sum_{v \in V} E[\chi_v] \\ &= np + \sum_{v \in V} P(v \in Y_X) \\ &\leqslant np + n(1-p)^{1+\delta}. \end{aligned}$$

En utilisant l'inégalité $(1-x) \leqslant e^{-x}\ \forall x \in \mathbb{R}$, et en remplaçant p par sa valeur $\frac{\ln(\delta+1)}{\delta+1}$, nous obtenons

$$\begin{aligned} E[|X| + |Y_X|] &\leqslant np + ne^{-p(\delta+1)} \\ &= n\left(\frac{\ln(\delta+1)}{\delta+1}\right) + ne^{-\ln(\delta+1)} \\ &= n\left(\frac{1+\ln(\delta+1)}{\delta+1}\right). \end{aligned}$$

Donc, il existe au moins un choix de $X \subseteq V$ pour lequel

$$|X| + |Y_X| \leqslant n\left(\frac{1+\ln(\delta+1)}{\delta+1}\right).$$

Pour un tel X, l'ensemble $U = X \cup Y_X$ est un ensemble dominant d'au plus $n\left(\frac{1+\ln(\delta+1)}{\delta+1}\right)$ sommets. □

Deux ingrédients classiques de la méthode probabiliste sont incorporés à cette preuve. Il s'agit de l'argument de l'espérance et du principe d'altération, car l'ensemble aléatoire X ne fournit pas immédiatement l'ensemble dominant recherché. Nous devons le modifier en lui ajoutant Y_X.

3.3. Graphes avec nombre chromatique et tour de taille arbitrairement grands

Voici une dernière application en théorie des graphes.

Le *tour de taille* $g(G)$ d'un graphe G est la longueur de son plus petit cycle. $\alpha(G)$ est la taille du plus grand stable dans G et $\chi(G)$ est le nombre chromatique de G.

Théorème 5 (Erdös, 1959)**.** *Pour tous $k, l \in \mathbb{N}$, il existe un graphe H avec $\chi(H) > k$ et $g(H) > l$.*

En 1959, Erdös a démontré ceci grâce à la méthode probabiliste. Pour beaucoup, c'est l'application la plus surprenante de la méthode car personne ne s'attend à une preuve non constructive de ce genre de théorème. La démonstration d'Erdös combine la technique de base (cf. section 2.1), le principe d'altération (cf. fin section 3.1) et l'inégalité de Markov. Nous décrivons ci-dessous le raisonnement suivi et les arguments probabilistes utilisés. Nous renvoyons le lecteur au livre de Diestel [5] pour le détail des calculs.

Démonstration. Soient $k, l \in \mathbb{N}$. Prenons un réel $\varepsilon \in]0, 1/l[$ et posons $p = n^{\varepsilon-1} \in [0,1]$ où $n \in \mathbb{N}\backslash\{0\}$ est un paramètre qui sera fixé plus tard. Définissons un espace de probabilités $(\Omega, \mathcal{A}, P)$ où Ω est l'ensemble des graphes sur les sommets $\{1, 2, \ldots, n\}$, où $\mathcal{A}$ contient toutes les parties de Ω et où la mesure de probabilité P est définie de façon à ce que les événements

$$A_{i,j} \equiv \{G \in \Omega \mid \{i,j\} \text{ est une arête de } G\}$$

soient indépendants avec $P(A_{i,j}) = p$ pour toute paire $\{i,j\}$ de sommets.

Tirons au hasard un graphe G dans Ω. Soit X la variable aléatoire qui compte le nombre de cycles de longueur au plus l dans G. Nous pouvons alors montrer (cf. Diestel [5]) que, pour n suffisamment grand, nous avons

(i) $P(\alpha(G) \geqslant \frac{n}{2k}) < 1/2$
(ii) $E[X] < n/4$.

Par l'inégalité de Markov, (ii) implique que $P(X \geqslant \frac{n}{2}) < 1/2$. Donc, en fixant n suffisamment grand, nous obtenons

$$\begin{aligned} P\left(\alpha(G) < \frac{n}{2k} \text{ et } X < \frac{n}{2}\right) &= 1 - P\left(\alpha(G) \geqslant \frac{n}{2k} \text{ ou } X \geqslant \frac{n}{2}\right) \\ &\geqslant 1 - P\left(\alpha(G) \geqslant \frac{n}{2k}\right) - P\left(X \geqslant \frac{n}{2}\right) \\ &> 1 - 1/2 - 1/2 = 0 \end{aligned}$$

En utilisant la technique de base de la méthode probabiliste, nous venons ainsi de prouver l'existence d'un graphe G^* dans Ω avec $\alpha(G^*) < \frac{n}{2k}$ et ayant moins de $n/2$ cycles de longueur au plus l.

Il reste une dernière étape d'altération à effectuer. En supprimant de G^* un sommet par cycle de longueur inférieur ou égale à l, nous faisons apparaître un nouveau graphe H. Ce H n'a plus de cycle de longueur inférieur ou égale à l. D'où $g(H) > l$.

De plus,

$$\chi(H) \geqslant \frac{\text{nombre de sommets de } H}{\alpha(H)} \geqslant \frac{n/2}{\alpha(H)} \geqslant \frac{n/2}{\alpha(G^*)} > \frac{n/2}{n/2k} = k.$$

□

3.4. Ensemble sans somme

La méthode probabiliste a aussi été utilisée pour prouver un théorème de théorie combinatoire des nombres.

Définition 4 (Ensemble sans somme)**.** *Un sous-ensemble A d'un groupe abélien $(G, +)$ est appelé* sans somme *si $(A + A) \cap A = \varnothing$, c'est-à-dire si l'équation $a_1 + a_2 = a_3$ n'a pas de solution avec a_1, a_2, $a_3 \in A$.*

Théorème 6 (Erdös, 1965)**.** *Tout ensemble $B = \{b_1, \ldots, b_n\}$ de n entiers non nuls contient un sous-ensemble sans somme A de taille $|A| > \frac{n}{3}$.*

Démonstration. Soit p un nombre premier de la forme $p = 3k+2$ avec p strictement supérieur à $2 \max_{1 \leqslant i \leqslant n} |b_i|$. L'existence de p est garantie par le théorème de la progression arithmétique de Dirichlet. [4]

Posons $C = \{k + 1, k + 2, \ldots, 2k + 1\}$. Clairement C est un sous-ensemble sans somme du groupe cyclique $\mathbb{Z}_p$.

Munissons $\{1, 2, \ldots, p - 1\}$ d'une distribution uniforme et choisissons au hasard un entier x dans cet ensemble. Pour tout $1 \leqslant i \leqslant n$, définissons d_i par $d_i \equiv xb_i \pmod p$, $0 \leqslant d_i < p$. Puisque x peut prendre toutes les valeurs entre 1 et $p - 1$

[4] Théorème de la progression arithmétique de Dirichlet : pour tous naturels non nuls a et b premiers entre eux, il existe une infinité de nombres premiers de la forme $a + nb$, où $n > 0$.

et puisque b_i est non nul (par hypothèse), alors d_i, lui, peut être n'importe quel élément non nul de $\mathbb{Z}_p$. D'où, pour $i = 1, \ldots, n$,

$$P(d_i \in C) = \frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.$$

Appelons A l'ensemble (aléatoire) de tous les éléments b_i $(1 \leqslant i \leqslant n)$ pour lesquels $d_i \in C$. Alors

$$\begin{aligned} E[|A|] &= \text{le nombre moyen de } d_i \text{ qui sont dans } C \\ &= \sum_{i=1}^{n} P(d_i \in C) \\ &> n \cdot \frac{1}{3}. \end{aligned}$$

Par l'argument de l'espérance (cf. section 2.2), nous sommes certains qu'il existe un $x \in \{1, 2, \ldots, p-1\}$ tel que $|A| > \frac{n}{3}$. Fixons un tel x. Ainsi, on obtient un sous-ensemble A de B de taille $|A| > \frac{n}{3}$ dont tous les éléments $a \in A$ vérifient $xa \pmod p \in C$.

Il ne reste plus qu'à montrer que ce A est sans somme. Par l'absurde, supposons qu'il existe a_1, a_2, $a_3 \in A$ tels que $a_1 + a_2 = a_3$. Alors nous aurions $xa_1 + xa_2 \equiv xa_3 \pmod p$. Mais ceci contredit le fait que C est un sous-ensemble sans somme de $\mathbb{Z}_p$. □

3.5. Théorème d'approximation de Weierstrass

Poursuivons avec une très belle application de la méthode probabiliste dans le domaine de l'analyse. C'est le célèbre théorème d'approximation uniforme de Weierstrass. Ce théorème affirme que l'ensemble des polynômes réels sur $[0, 1]$ est dense dans l'ensemble des fonctions continues réelles sur $[0, 1]$. Formellement, il s'énonce comme suit :

Théorème 7 (Weierstrass)**.** *Si $f : [0, 1] \to \mathbb{R}$ est une fonction continue, alors pour tout $\varepsilon > 0$, il existe un polynôme $p(x)$ tel que $|p(x) - f(x)| \leqslant \varepsilon$ pour tout $x \in [0, 1]$.*

Parmi toutes les preuves de ce théorème, celle de Bernstein [2] est sûrement la plus élégante. (Voir aussi [7].) C'est une preuve constructive. L'utilisation des probabilités la rend particulièrement intéressante. Elle repose sur les propriétés de la distribution binomiale et sur l'inégalité de Chebyshev que nous rappelons : si X est une variable aléatoire réelle de moyenne μ et de variance σ^2, alors

$$\forall \lambda > 0 : P(|X - \mu| \geqslant \lambda) \leqslant \frac{\sigma^2}{\lambda^2}.$$

Voici la preuve probabiliste du Théorème 7 donnée par Bernstein :

Démonstration. Soit $f : [0, 1] \to \mathbb{R}$ une fonction continue, et soit $\varepsilon > 0$. Puisque f est continue sur un domaine compact, nous avons

(i) f est uniformément continue ;
(ii) f est bornée.

Par (i), il exite un $\delta > 0$ tel que si $x, x' \in [0,1]$ et si $|x - x'| \leqslant \delta$, alors $|f(x) - f(x')| \leqslant \varepsilon/2$. De plus, (ii) implique l'existence d'un $M > 0$ tel que pour tout $x \in [0,1]$ on a $|f(x)| \leqslant M$.

Pour chaque $n \in \mathbb{N}$ et chaque $x \in]0,1[$, nous définissons $B_{n,x}$ comme étant une variable aléatoire de loi binomiale de paramètres n et x. Donc, $B_{n,x}$ compte le nombre de succès obtenus en répétant de façon indépendante n épreuves de paramètre de succès x. Par conséquent

$$\forall j = 0, 1, \ldots, n : P(B_{n,x} = j) = \binom{n}{j} x^j (1-x)^{n-j}.$$

L'espérance de $B_{n,x}$ est nx et sa variance est $nx(1-x)$. Alors, pour tout $n \in \mathbb{N}$ et tout $x \in [0,1]$, nous avons

$$P\left(|B_{n,x} - nx| \geqslant n^{2/3}\right) \leqslant \frac{nx(1-x)}{n^{4/3}} \leqslant \frac{1}{n^{1/3}}$$

où la première inégalité résulte de l'inégalité de Chebyshev et la deuxième vient du fait que $x(1-x) \leqslant 1$ lorsque $x \in]0,1[$. Dès lors, nous pouvons choisir un entier n suffisamment grand de façon à ce que les deux inégalités suivantes soient satisfaites :

$$P\left(|B_{n,x} - nx| \geqslant n^{2/3}\right) < \frac{\varepsilon}{4M} \tag{1}$$

et

$$\frac{1}{n^{1/3}} < \delta. \tag{2}$$

Nous définissons le polynôme

$$P_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right)$$

et nous vérifions que $|P_n(x) - f(x)| \leqslant \varepsilon$ pour tout $x \in [0,1]$. Si $x = 0$ ou $x = 1$, $P_n(x) = f(x)$. Soit donc $x \in]0,1[$.

D'abord, remarquons que

$$\sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} = (x + (1-x))^n = 1.$$

Par conséquent, nous pouvons écrire

$$|P_n(x) - f(x)| = \left| \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} \left(f\left(\frac{i}{n}\right) - f(x) \right) \right|.$$

Puis, en appliquant l'inégalité triangulaire à deux reprises, nous obtenons

$$\begin{aligned} |P_n(x) - f(x)| \quad \leqslant \quad & \sum_{i;|i-nx|<n^{2/3}} \binom{n}{i} x^i (1-x)^{n-i} \left| f\left(\frac{i}{n}\right) - f(x) \right| \\ & + \sum_{i;|i-nx|\geqslant n^{2/3}} \underbrace{\binom{n}{i} x^i (1-x)^{n-i}}_{P\left(B_{n,x}=i\right)} \underbrace{\left(\left| f\left(\frac{i}{n}\right) \right| + |f(x)| \right)}_{\leqslant 2M} \end{aligned}$$

Dans la première somme, nous avons

$$\left|\frac{i}{n}-x\right| = \frac{|i-nx|}{n} < \frac{n^{2/3}}{n} = \frac{1}{n^{1/3}} < \delta \qquad \text{par (2)}$$

et donc

$$\left|f\left(\frac{i}{n}\right)-f(x)\right| \leqslant \varepsilon/2.$$

Nous pouvons majorer cette première somme par

$$\varepsilon/2 \cdot \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} = \varepsilon/2 \cdot 1 = \varepsilon/2.$$

La deuxième somme quant à elle est majorée par

$$\begin{aligned} \sum_{i;|i-nx|\geqslant n^{2/3}} P\left(B_{n,x}=i\right) 2M &= P\left(|B_{n,x}-nx| \geqslant n^{2/3}\right) \cdot 2M \\ &\leqslant \frac{\varepsilon}{4M} 2M \qquad \text{par (1)} \\ &= \varepsilon/2. \end{aligned}$$

Nous pouvons enfin conclure :

$$|P_n(x)-f(x)| \leqslant \varepsilon/2 + \varepsilon/2 \leqslant \varepsilon.$$

□

Pour une autre application de la méthode probabiliste en analyse, nous citons le travail de [3] sur une approche probabiliste du polynôme de Taylor.

3.6. Égalité Min-Max

Voici une autre facette d'un raisonnement probabiliste pour un résultat purement déterministe, à savoir, pour ce qu'on appelle l'Égalité Min-Max. Notre démonstration semble être neuve.

Théorème 8. *Tout n-uplet de réels* $(x_1,\ldots,x_n) \in \mathbb{R}^n$ *satisfait à l'égalité*

$$\begin{aligned} \max\{x_1,\ldots,x_n\} = &\sum_{1\leqslant k\leqslant n} x_k - \sum_{1\leqslant k_1<k_2\leqslant n} \min\{x_{k_1},x_{k_2}\} \\ &+ \sum_{1\leqslant k_1<k_2<k_3\leqslant n} \min\{x_{k_1},x_{k_2},x_{k_3}\} \\ &- \ldots \\ &+ (-1)^{n-1} \min\{x_1,\ldots,x_n\}. \end{aligned}$$

Comment prouver une telle égalité ? Une approche standard serait celle par induction sur n, mais les calculs sont laborieux. Voici notre approche.

Démonstration. Soient $x_1, \ldots, x_n$ des réels fixés. Posons $m := \min\{x_1, \ldots, x_n\}$ et $M := \max\{x_1, \ldots, x_n\}$. Nous introduisons une variable aléatoire continue Y de loi uniforme sur $[m, M]$.

$$\Longrightarrow P(Y \leqslant y) = \frac{y-m}{M-m} \qquad \text{pour tout } y \in [m, M].$$

Pour $1 \leqslant i \leqslant n$, appelons A_i l'événement $[Y \leqslant x_i]$ et calculons de deux manières la probabilité de l'événement $\left[\bigcup_{i=1}^n A_i\right]$.
D'une part,

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(\ Y \leqslant x_1 \text{ ou } Y \leqslant x_2 \text{ ou } \ldots \text{ou } Y \leqslant x_n) \\ &= P(\ Y \leqslant \max\{x_1, \ldots, x_n\}) \\ &= P(\ Y \leqslant M) \\ &= 1. \end{aligned} \tag{3}$$

D'autre part, la formule d'inclusion-exclusion fournit

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_i P(A_i) \ - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \ - \ \ldots \\ &\quad + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n) \\ &= \sum_i P(Y \leqslant x_i) \ - \sum_{i<j} P(Y \leqslant \min\{x_i, x_j\}) \\ &\quad + \sum_{i<j<k} P(Y \leqslant \min\{x_i, x_j, x_k\}) \ - \ \ldots \\ &\quad + (-1)^{n-1} P(Y \leqslant \min\{x_1, \ldots, x_n\}) \\ &= \sum_i \left(\frac{x_i - m}{M-m}\right) \ - \sum_{i<j} \left(\frac{\min\{x_i, x_j\} - m}{M-m}\right) \\ &\quad + \sum_{i<j<k} \left(\frac{\min\{x_i, x_j, x_k\} - m}{M-m}\right) \ - \ \ldots \\ &\quad + (-1)^{n-1} \left(\frac{\min\{x_1, \ldots, x_n\} - m}{M-m}\right). \end{aligned} \tag{4}$$

En multipliant (3) et (4) par $M-m$, nous obtenons

$$\begin{aligned} M-m &= \sum_i (x_i - m) - \sum_{i<j} (\min\{x_i, x_j\} - m) + \sum_{i<j<k} (\min\{x_i, x_j, x_k\} - m) \\ &\quad - \ \ldots \ + \ (-1)^{n-1} (\min\{x_1, \ldots, x_n\} - m) \\ &= \sum_i x_i - \sum_{i<j} \min\{x_i, x_j\} + \sum_{i<j<k} \min\{x_i, x_j, x_k\} \\ &\quad - \ \ldots \ + \ (-1)^{n-1} \min\{x_1, \ldots, x_n\} \\ &\quad - \underbrace{\left[\binom{n}{1} - \binom{n}{2} + \binom{n}{3} - \ldots + (-1)^{n-1}\binom{n}{n}\right]}_{=1} m. \end{aligned}$$

Il ne reste plus qu'à additionner m aux deux membres de l'égalité pour faire apparaître le résultat souhaité. □

4. Quelques arguments probabilistes heuristiques

4.1. Conjecture de Goldbach

Nous allons à présent nous intéresser à l'une des plus célèbres et des plus anciennes conjectures de la théorie des nombres. La conjecture de Goldbach, qui date de 1743 et qui aujourd'hui encore reste non résolue, s'énonce comme suit :

Conjecture 9 (Conjecture de Goldbach)**.** *Tout entier pair strictement supérieur à 2 est la somme de deux nombres premiers.*

Nous présentons ici deux arguments probabilistes qui montrent que nous avons de bonnes raisons de penser que la conjecture est vraie. Tous deux reposent sur le théorème des nombres premiers démontré simultanément et indépendamment par Jacques Hadamard (France) et Charles de la Vallée Poussin [4] (Belgique) en 1896.

Théorème 10 (Hadamard et de la Vallée Poussin, 1896)**.**

$$\frac{\#\{p \leqslant n \mid p \text{ est un nombre premier}\}}{n} \sim \frac{1}{\ln(n)}$$

Premier argument Soit $2n$ un entier pair strictement supérieur à 2. Il existe n façons de le décomposer en une somme de la forme $m + (2n - m)$ avec $m \in \{1, 2, \ldots, n\}$. Soit $N(2n)$ le nombre de telles décompositions pour lesquelles m et $2n - m$ sont des nombres premiers. Choisissons un entier m au hasard dans l'ensemble $\{1, 2, \ldots, n\}$. La probabilité que ce m et $2n - m$ soient tous les deux des nombres premiers est

$$\frac{\#\text{ choix de } m \text{ pour lesquels } m \text{ et } 2n-m \text{ sont premiers}}{\#\text{ choix possibles pour } m} = \frac{N(2n)}{n}.$$

De plus, si on fait l'hypothèse que les événements [m est premier] et [$2n - m$ est premier] sont indépendants, cette même probabilité est asymptotiquement

$$\frac{1}{(\ln(n))^2}.$$

En effet, $P(m \text{ et } 2n-m \text{ premiers})$

$$\begin{aligned} &= P(m \text{ premier}) \cdot P(2n-m \text{ premier}) \\ &= \left(\frac{\#\{p \leqslant n \mid \text{p premier}\}}{n}\right) \cdot \left(\frac{\#\{n+1 \leqslant p \leqslant 2n \mid \text{p premier}\}}{n}\right) \\ &= \left(\frac{\#\{p \leqslant n \mid \text{p premier}\}}{n}\right) \cdot \left(\frac{\#\{p \leqslant 2n \mid \text{p premier}\}}{n} - \frac{\#\{p \leqslant n \mid \text{p premier}\}}{n}\right) \end{aligned}$$

et, par le théorème des nombres premiers, ceci se comporte asymptotiquement comme

$$\frac{1}{\ln(n)}\left(\frac{2}{\ln(2n)} - \frac{1}{\ln(n)}\right) \sim \frac{1}{\ln(n)}\left(\frac{2}{\ln(n)} - \frac{1}{\ln(n)}\right) \sim \frac{1}{(\ln(n))^2}.$$

Donc, nous en déduisons que

$$N(2n) \sim \frac{n}{(\ln(n))^2}$$

et la fonction à droite tend vers ∞ lorsque n tend vers ∞.

Autrement dit, plus n est grand, plus nous devons nous attendre à ce qu'il existe un grand nombre de décompositions de n en somme de deux nombres premiers. Donc, plus n est grand, moins nous aurons de chances de trouver un contre-exemple à la conjecture de Goldbach (et nous savons déjà qu'elle a été vérifiée par ordinateur pour tous les entiers pairs jusqu'à $1,1 \times 10^{18} \ldots$) !

Deuxième argument Pour un entier pair $2n$ strictement supérieur à 2, considérons le modèle probabiliste suivant : tout nombre de l'ensemble $\{1, 2, \ldots, 2n\}$ est un nombre premier avec probabilité (asymptotique) $\frac{1}{\ln(2n)}$ indépendamment des autres nombres de l'ensemble. Nous pouvons estimer la probabilité que, dans notre modèle, $2n$ ne soit pas la somme de deux nombres premiers.

Appelons E_n l'événement qui correspond à un tel *échec*, c'est-à-dire E_n est l'événement $[N(2n) = 0]$. En utilisant les hypothèses faites dans notre modèle, nous calculons la probabilité (asymptotique) d'un *échec*.

$$\begin{aligned} P(E_n) &= P(\ \forall 1 \leqslant m \leqslant n,\ m \text{ ou } 2n-m \text{ n'est pas premier}) \\ &= \prod_{m=1}^{n} P(m \text{ ou } 2n-m \text{ n'est pas premier}) \\ &\sim \left(1 - \frac{1}{(\ln(2n))^2}\right)^n \\ &= \left[\left(1 - \frac{1}{(\ln(2n))^2}\right)^{(\ln(2n))^2}\right]^{n/(\ln(2n))^2} \\ &\sim e^{-n/(\ln(2n))^2} \end{aligned}$$

Puisque $\sum_{n=2}^{\infty} e^{-n/(\ln(2n))^2} < \infty$, nous avons aussi que $\sum_{n=2}^{\infty} P(E_n) < \infty$ et le lemme de Borel-Cantelli implique alors que

$$P(\limsup_{n\to\infty} E_n) = 0$$

c'est-à-dire

$$0 \;=\; P(\cap_n \cup_{i\geqslant n} E_i) \quad = \quad P(\ \forall n\ \exists i \geqslant n \text{ tq } N(2i) = 0)$$

ou encore

$$P(\text{il n'y a qu'un nombre fini d'}\textit{échecs}) = 1.$$

Autrement dit, dans notre modèle, pour n suffisamment grand, la conjecture de Goldbach est vraie avec probabilité 1.

Les deux raisonnements ci-dessus sont de nature heuristique, car nous n'avons pas la preuve que notre modèle décrit correctement la situation réelle. En fait, nous savons même que le modèle n'est pas exact puisqu'il ignore l'existence de corrélations entre, par exemple, les événements [m premier] et [$2n - m$ premier]. Il faut donc bien faire une distinction entre, d'une part, ce que le modèle probabiliste montre et, d'autre part, ce qui pourrait être la situation réelle.

4.2. Théorème de Green-Tao

En 2004, Ben Green et Terence Tao (lauréat de la médaille Fields en 2006) démontrent un résultat qui fait progresser remarquablement la recherche dans le domaine des nombres premiers. Voici leur théorème :

Théorème 11 (Green-Tao, 2004)**.** *La suite des nombres premiers contient des progressions arithmétiques arbitrairement longues, c'est-à-dire que pour tout entier $k \geqslant 3$, il existe des nombres premiers p_1, p_2, ..., p_k tels que*

$$p_2 - p_1 = p_3 - p_2 = p_4 - p_3 = \ldots = p_k - p_{k-1}.$$

Plus précisément, Green et Tao ont démontré le résultat ci-dessous.

Pour tout entier $k \geqslant 3$, il existe une constante $\delta_k > 0$ telle que

$$\#\left\{(n,d) \in [1,N]^2 \mid n, n+d, \ldots, n+(k-1)d \text{ sont premiers}\right\} \geqslant \delta_k \frac{N^2}{(\ln(N))^k}.$$

Pour vérifier que ce résultat implique le théorème de Green-Tao, il suffit de remarquer que pour tout $\delta_k > 0$, le terme $\delta_k \frac{N^2}{(\ln(N))^k}$ est supérieur à 1 pour N suffisamment grand.

Un argument probabiliste semblable à ceux présentés dans le cadre de la conjecture de Goldbach suggère que le théorème de Green-Tao est vrai. Fixons un entier $k \geqslant 3$ et N un grand nombre naturel. Soit

$$X_N := \#\left\{(n,d) \in [1,N]^2 \mid n, n+d, \ldots, n+(k-1)d \text{ sont premiers}\right\}.$$

Comme précédemment, nous définissons notre modèle probabiliste en supposant que chaque nombre de l'ensemble $\{1, 2, \ldots, kN\}$ est un nombre premier avec probabilité asymptotique $\frac{1}{\ln(kN)}$ indépendamment des autres nombres de l'ensemble.

Choisissons au hasard deux nombres n et d dans $\{1, 2, \ldots, N\}$. La probabilité que (n, d) donne une progression arithmétique de nombres premiers (dont le premier nombre est n, la raison d et la longueur k) est

$$P(n, n+d, n+2d, \ldots, n+(k-1)d \text{ sont tous premiers})$$
$$= \frac{\#\text{ choix de } (n,d) \text{ pour lesquels ces } k \text{ nombres sont premiers}}{\#\text{ choix possibles pour } (n,d)} = \frac{X_N}{N^2}.$$

D'autre part, puisque $n+(k-1)d \geqslant kN$, notre hypothèse d'indépendance implique que cette probabilité est égale au produit

$$\prod_{i=1}^{k} P(\ n+(i-1)d \ \text{ est premier})$$

qui est asymptotiquement

$$\frac{1}{(\ln(kN))^k} \sim \frac{1}{(\ln(N))^k}.$$

Donc,

$$X_N \sim \frac{N^2}{(\ln(N))^k}$$

et cette fonction tend vers ∞ lorsque N tend vers ∞.

Ceci signifie que plus N est grand, plus nous devons nous attendre à ce que beaucoup de couples $(n, d) \in [1, N]^2$ fournissent des progressions arithmétiques de nombres premiers de longueur k. A nouveau, ce raisonnement n'est en rien une démonstration du théorème de Green-Tao car, dans la réalité, les événements [n premier], [$n+d$ premier], ..., [$n+(k-1)d$ premier] ne sont pas indépendants.

La preuve de Green et Tao repose sur ce raisonnement probabiliste. Les auteurs ont réussi l'admirable tour de force de tourner tous les arguments heuristiques en des arguments parfaitement rigoureux. Ils se sont intéressés à des observations du type : si p est pair alors p, $p+d$, $p+2d$ ne peuvent pas être tous premiers car au moins deux d'entre eux sont pairs ; par contre, si p est premier et d est impair alors la probabilité que p, $p+d$, ..., $p+(k-1)d$ soit une progression de nombres premiers est clairement plus élevée. Ils vont mettre en place des techniques sophistiqués pour formaliser tout cela et obtiendront finalement la constante $\delta_k > 0$ de leur résultat. Pour l'entièreté de leur preuve voir [6].

5. Références

[1] Alon, N. and Spencer, J.H. (1992). *The Probabilistic Method*. Wiley-Interscience Series in Discrete Mathematics and Optimization, New-York.

[2] Bernstein, S. N. (1912). *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*. Communications de la Société mathématique de Kharkow, 13, n° 2, pp. 1-2.

[3] Bruss, F. T. (1982). *A Probabilistic Approach to an Approximation Problem*. Annales de la Société Scientifique de Bruxelles, 96, Vol. 2, pp. 91-97.

[4] de la Vallée Poussin, C. (1896). *Recherches analytiques sur la théorie des nombres premiers*. Annales de la Société Scientifique de Bruxelles, 20, part II, pp. 183-256, 281-397.

[5] Diestel, R. (2005) *Graph Theory*. 3e édition, Graduate Texts in Mathematics, Vol. 173, Springer-Verlag, New York, pp. 299-301.

[6] Green, B. and Tao, T. (2008). *The primes contain arbitrarily long arithmetic progressions*. Annals of Mathematics, Vol. 167, N° 2, pp. 481-547. (http ://www.citebase.org/abstract ?id=oai :arXiv.org :math/0404188.)

[7] Sheynin, O. (2004). *A demonstration of the Weierstrass theorem based on the theory of probability by S.N. Bernstein.* The Math. Scientist, Vol. 29, pp. 127-128.