Sujet d'algèbre
par Philippe Caldero (Université Lyon 1)

Soit $\mathbb{K}$ un corps fini, E un $\mathbb{K}$-espace vectoriel de dimension finie, et u un endomorphisme nilpotent de E. Nous nous intéressons au nombre de sous-espaces stables de u.

Nous notons q le cardinal de $\mathbb{K}$ et n la dimension de E. Dans une base adéquate de E, la matrice de u est diagonale par blocs, avec sur la diagonale des blocs de Jordan de taille λ_1, λ_2, ..., λ_ℓ. (Par construction, $n = \lambda_1 + \lambda_2 + \cdots + \lambda_\ell$.) Nous supposerons que les λ_i sont rangés dans l'ordre décroissant : la partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ est ainsi déterminée par u. Le nombre de sous-espaces stables de u ne dépend que de λ et q : nous le notons $a_\lambda(q)$.

Le cas où u est cyclique (c'est-à-dire $\ell = 1$ et $\lambda = (n)$) est classique : les sous-espaces stables de u sont les sous-espaces de la forme $\ker u^d$, où d est un entier tel que $0 \leq d \leq n$ (notons qu'ici nous avons l'égalité $\ker u^d = \operatorname{im} u^{n-d}$). Par conséquent, $a_{(n)}(q) = n + 1$.

Le cas $u = 0$ (c'est-à-dire $\ell = n$ et $\lambda = (1, 1, \ldots, 1)$) est lui aussi bien connu : si $0 \leq d \leq n$, alors le nombre de sous-espaces vectoriels de dimension d de E est donné par le coefficient binomial de Gauss

$$\binom{n}{d}_q = \frac{(1-q^n)(1-q^{n-1})\cdots(1-q^{n-d+1})}{(1-q^d)(1-q^{d-1})\cdots(1-q)}$$

et ainsi

$$a_{(1,1,\ldots,1)}(q) = \sum_{d=0}^{n} \binom{n}{d}_q .$$

L'exercice consiste à étudier l'énoncé suivant : $a_\lambda(q)$ est un polynôme en q à coefficients entiers positifs, et les valeurs en 0 et 1 de ce polynôme sont $a_\lambda(0) = n + 1$ et

$$a_\lambda(1) = (\lambda_1 + 1)(\lambda_2 + 1)\cdots(\lambda_\ell + 1).$$

Remarque. Vous pouvez commencer par regarder des exemples en petite dimension. À défaut de parvenir à démontrer le cas général, qui est difficile, vous pouvez vous concentrer sur des situations particulières, par exemple le cas où λ n'a que deux parts (c'est-à-dire $\ell = 2$) ou le cas $u^2 = 0$ (c'est-à-dire tous les λ_i sont inférieurs ou égaux à 2). La rédaction d'exemples ou de solutions partielles sera valorisée.

Question subsidiaire. Omettons l'hypothèse que u est nilpotent. Comment alors peut-on exprimer le nombre de sous-espaces stables de u en fonction des facteurs invariants de u ?

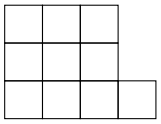
Éléments de solution.

Une vidéo présentant le sujet et sa solution est disponible sur la chaîne YouTube de l'auteur. Les notes qui suivent forment quant à elles un exposé écrit.

La principale difficulté à surmonter pour traiter le sujet est de trouver un procédé capable de gérer la complexité combinatoire et de permettre une rédaction rigoureuse des arguments. Nous proposons deux méthodes pour cela, la seconde étant essentiellement celle présentée dans la vidéo (ces deux méthodes ont été trouvées dans les copies). Nous concluons par une ébauche de réponse à la question subsidiaire.

Première méthode.

Nous représentons une partition λ par son diagramme, avec λ_p cases sur la p-ième ligne du diagramme (en partant du bas). Par exemple, le diagramme de $\lambda = (4, 3, 3)$ est



(Le diagramme comporte ℓ lignes, où ℓ est le nombre de parts non nulles dans λ.) Nous notons D l'ensemble des cases du diagramme qui ne sont pas sur la colonne de gauche, et si $b \in D$, nous notons b^- la case située immédiatement à gauche de b. Avec ces notations, il existe une base (e_b) de E, indexée par les cases du diagramme, telle que $u(e_b) = e_{b^-}$ si $b \in D$ et $u(e_b) = 0$ sinon. Pour simplifier l'écriture dans la suite, nous écrirons parfois abusivement e_{b^-} même si $b \notin D$; il faut alors comprendre que e_{b^-} est le vecteur nul.

Nous numérotons les cases du diagramme de 1 à n, en procédant colonne par colonne de bas en haut, en commençant par la colonne de droite.

10	7	4	
9	6	3	
8	5	2	1

Pour simplifier l'exposé, nous confondrons les cases du diagramme avec leur numéro, notant par exemple e_j plutôt que e_b si la case b a reçu le numéro j. De même, nous notons j^- le numéro de la case b^-. Nous observons que si j et k appartiennent à D et $k > j$, alors $k^- > j^-$.

Rappelons qu'une matrice à m lignes et n colonnes, à coefficients dans $\mathbb{K}$, et de rang m, est dite échelonnée réduite si :

— sur chaque ligne, le premier coefficient non nul, appelé pivot, est un 1 ;

— le pivot d'une ligne est strictement plus à droite que le pivot de la ligne précédente ;

— lorsqu'une colonne contient un pivot, tous les autres coefficients de la colonne sont nuls.

Soit F un sous-espace vectoriel de E, de dimension m. Nous pouvons choisir une base $(f_1, f_2, \ldots, f_m)$ de F. Soit M la matrice à m lignes et n colonnes dont la i-ième ligne est formée des coordonnées de f_i dans la base $(e_1, e_2, \ldots, e_n)$. La matrice M dépend manifestement de la base de F utilisée : un autre choix se traduit par la multiplication de M à gauche par une matrice dans $\mathrm{GL}_m(\mathbb{K})$, c'est-à-dire par l'action d'une suite d'opérations élémentaires sur les lignes de M. Or l'algorithme de Gauss permet, à l'aide de ces mêmes opérations, de mettre

M sous forme échelonnée réduite ; et cette forme échelonnée réduite est unique. Ainsi chaque sous-espace vectoriel F de E de dimension m peut être représenté, de façon unique, par une matrice échelonnée réduite M_F à m lignes et n colonnes de rang m.

Voici dans notre exemple (espace de dimension 10) une forme possible pour la matrice échelonnée réduite M_F d'un sous-espace F de dimension 5.

$$\begin{pmatrix} 0 & 0 & 1 & * & * & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Les colonnes de M_F sont en bijection avec les cases du diagramme. Une façon d'indiquer la position des pivots dans M_F est de griser les cases correspondantes du diagramme.

10	7	4	
9	6	3	
8	5	2	1

Nous allons maintenant regarder comment se traduit sur M_F la condition que F est stable par u. Prenons la i-ième ligne de M_F, dont les coefficients sont les coordonnées du vecteur f_i dans la base $(e_1, e_2, \ldots, e_n)$. Soit j le numéro de la colonne contenant le pivot de cette ligne ; comme M_F est échelonnée, f_i est égal à e_j plus une combinaison linéaire de vecteurs e_k avec $k > j$. Supposons $j \in D$. Nous avons $k^- > j^-$ pour chaque $k > j$; il s'ensuit que $u(f_i)$ est égal à e_{j^-} plus une combinaison linéaire de vecteurs $e_{k'}$ avec $k' > j^-$. Or ce vecteur $u(f_i)$ appartient à F, donc est combinaison linéaire de lignes de M_F. Il est donc nécessaire que j^- soit le pivot d'une ligne de M_F.

Par commodité, nous appellerons clé un ensemble S de cases du diagramme tel que $(j \in S \Rightarrow j^- \in S)$ pour chaque $j \in D$. La discussion précédente montre que si F est un sous-espace stable de u, décrit par une matrice échelonnée réduite M_F, alors l'ensemble des pivots de M_F (les cases grisées du diagramme) est une clé. Nous l'appellerons la clé de F.

Fixons une clé S. Comment décrire l'ensemble $\mathscr{E}_S$ des sous-espaces stables par u de clé S ?

Précisons nos notations : soit m le cardinal de S et soient s_1, s_2, ..., s_m les éléments de S, énumérés dans l'ordre croissant. Pour chaque $p \in [\![1, \ell]\!]$, indiquons par $j(p)$ le numéro de la dernière case de S sur la p-ième ligne du diagramme, notons $i(p)$ l'indice tel que $s_{i(p)} = j(p)$, et appelons $B_S(p)$ l'ensemble des cases du diagramme n'appartenant pas à S de numéro strictement plus grand que $j(p)$. (Si aucune case de la clé n'apparaît sur la p-ième ligne du diagramme, $i(p)$ et $j(p)$ ne sont pas définis et $B_S(p) = \varnothing$.) Dans notre exemple :

p	$i(p)$	$j(p)$	$B_S(p)$
1	3	8	$\varnothing$
2	1	3	$\{4, 5, 7\}$
3	5	10	$\varnothing$

Avec ces notations, si M est une matrice échelonnée réduite de taille $m \times n$ et de rang m et si S est l'ensemble des indices des colonnes de M contenant un pivot, alors :

— M admet un pivot en position $(i(p), j(p))$.
— $B_S(p)$ est l'ensemble des indices k pour lesquels le coefficient matriciel $m_{i(p),k}$ de M_F n'est pas fixé par la condition d'échelonnement.

Nous affirmons à présent la chose suivante : pour chaque $(a_{p,k}) \in \prod_{p=1}^{\ell} \mathbb{K}^{B_S(p)}$, il existe un unique sous-espace stable $F \in \mathscr{E}_S$ contenant les vecteurs

$$f_{i(p)} = e_{j(p)} + \sum_{k \in B_S(p)} a_{p,k} e_k \qquad (*)$$

pour $p \in [\![1, \ell]\!]$. Par conséquent, $\mathscr{E}(S)$ est de cardinal $q^{\mathscr{N}(S)}$, où $\mathscr{N}(S) = \sum_{p=1}^{\ell} |B_S(p)|$.

Démontrons notre affirmation. La donnée $(*)$ fixe les coefficients

$$m_{i(p),k} = a_{p,k} \quad \text{avec } p \in [\![1, \ell]\!] \text{ et } k \in B_S(p)$$

de M_F, c'est-à-dire fixe les lignes de M_F correspondant aux cases les plus à droite de S. Nous voulons voir que les autres lignes de M_F sont alors déterminées par la condition que F soit stable par u. Procédant par récurrence de la droite vers la gauche dans la clé, il nous faut examiner la situation suivante : on choisit une case j de la clé, avec $j \in D$; on appelle i (respectivement i') le numéro de la ligne de M_F dont le pivot est sur la colonne j (respectivement j^-). Nous supposons connue la i-ième ligne de M_F et voulons trouver la i'-ième ligne. Connaître la i-ième ligne signifie avoir déterminé les coefficients de la combinaison linéaire

$$f_i = e_j + \sum_{\substack{k>j \\ k \notin S}} b_k e_k.$$

Nous avons alors

$$u(f_i) = e_{j^-} + \sum_{\substack{k>j \\ k \in D \setminus S}} b_k e_{k^-}.$$

Si k^- appartient à la clé, alors il est en bout de ligne sur la clé (puisque k n'appartient pas à la clé), et est donc de la forme $j(p)$ pour un $p \in [\![1, \ell]\!]$. La donnée $(*)$ nous indique

$$e_{k^-} = f_{i(p)} - \sum_{k' \in B_S(p)} a_{p,k'} e_{k'},$$

égalité que nous pouvons substituer dans l'expression de $u(f_i)$. Ceci fait pour tous les k pertinents, nous obtenons une expression de la forme

$$u(f_i) = e_{j^-} + \sum_{\substack{k'>j^- \\ k' \notin S}} c_{k'} e_{k'} + \sum_{p=1}^{\ell} d_p f_{i(p)},$$

dans laquelle les coefficients $c_{k'}$ et d_p sont déterminés par les b_k et par $(*)$. Ce vecteur doit appartenir à F, donc doit être combinaison linéaire des lignes de M_F. Cela sera le cas si et seulement si la i'-ième ligne de M_F est

$$f_{i'} = e_{j^-} + \sum_{\substack{k'>j^- \\ k' \notin S}} c_{k'} e_{k'}.$$

Nous voyons ainsi que nous pouvons avancer sans ambiguïté ni contrainte dans la récurrence. La matrice M_F est donc complètement déterminée à partir de la donnée $(*)$ et de la condition que F est stable par u.

Ceci justifie notre affirmation et sa conséquence : $\mathscr{E}_S$ possède $q^{\mathscr{N}(S)}$ éléments. Le nombre total de sous-espaces de E stables par u est donc égal à

$$a_\lambda(q) = \sum_{S \text{ clé}} q^{\mathscr{N}(S)}.$$

D'après la formule obtenue dans la question précédente, $a_\lambda(0)$ est égal au nombre de clés S incluses dans le diagramme de λ telles que $\mathscr{N}(S) = 0$. Une telle clé S est de la forme suivante : il existe un entier $j \in [\![0, \dim E]\!]$ tel que S soit l'ensemble des cases de numéro strictement supérieur à j. Par conséquent, $a_\lambda(0) = \dim E + 1$.

La valeur $a_\lambda(1)$ est quant à elle égale au nombre de clés incluses dans le diagramme de λ. Or, choisir une clé, c'est choisir la longueur des « dents » de la clé sur chacune des lignes : $\lambda_1 + 1$ choix pour la première ligne, $\lambda_2 + 1$ sur la deuxième, ..., pour un total de $(\lambda_1 + 1)(\lambda_2 + 1)\cdots(\lambda_\ell + 1)$ clés.

Cette première méthode est empruntée à l'article *A theorem on the fixed point set of a unipotent transformation on the flag manifold* de Naohisa Shimomura, J. Math. Soc. Japan **32** (1980), 55–64.

<u>*Seconde méthode.*</u>

Dans cette seconde méthode, nous démontrons que pour chaque partition μ, le nombre $b_{\lambda,\mu}(q)$ de sous-espaces vectoriels F de E stables par u tels que la restriction $u|_F$ soit de type μ est donné par un polynôme en q à coefficients entiers positifs. Il suffira alors d'écrire

$$a_\lambda(q) = \sum_\mu b_{\lambda,\mu}(q) \qquad (\dagger)$$

pour répondre à la question.

Introduisons la partition conjuguée $\lambda' = (\lambda'_1, \lambda'_2, \ldots, \lambda'_{\ell'})$ de λ : par définition,

$$\ell' = \lambda_1 \quad \text{et} \quad \lambda'_j = \mathrm{Card}\{i \in [\![1, \ell]\!] \mid \lambda_i \geq j\}$$

pour $j \in [\![1, \ell']\!]$. Autrement dit, alors que les parts λ_1, λ_2, ..., λ_ℓ de λ sont les longueurs des lignes du diagramme de λ, les parts λ'_1, λ'_2, ..., $\lambda'_{\ell'}$ de λ' sont les hauteurs des colonnes de ce diagramme. Les parts de λ' donnent les dimensions des noyaux itérés de u :

$$\dim \ker u^j = \lambda'_1 + \lambda'_2 + \ldots + \lambda'_j.$$

(Cette égalité a lieu pour tout $j \geq 0$ si l'on définit $\lambda'_j = 0$ pour $j > \ell'$.)

Fixons une seconde partition $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$ et notons $\mu' = (\mu'_1, \mu'_2, \ldots, \mu'_{m'})$ la partition conjuguée de μ. Un sous-espace F stable par u tel que la restriction $u|_F$ est de type μ s'écrit sous la forme

$$F = \mathrm{Vect}\big\{u^k(x_1) \;\big|\; k \in [\![0, \mu_1 - 1]\!]\big\} + \mathrm{Vect}\big\{u^k(x_2) \;\big|\; k \in [\![0, \mu_2 - 1]\!]\big\} + \\ \cdots + \mathrm{Vect}\big\{u^k(x_m) \;\big|\; k \in [\![0, \mu_m - 1]\!]\big\} \qquad (\ddagger)$$

où la famille $(x_1, x_2, \ldots, x_m)$ est telle que la somme (‡) soit directe et que l'on ait $u^{\mu_i}(x_i) = 0$ et $u^{\mu_i - 1}(x_i) \neq 0$ pour chaque $i \in [\![1, m]\!]$.

Dénombrons ces familles $(x_1, x_2, \ldots, x_m)$ en comptant le nombre de possibilités qui s'offrent pour x_i lorsqu'on suppose x_1, x_2, ..., x_{i-1} choisis. Notre hypothèse précise est que pour $j \in [\![1, i-1]\!]$, on a $x_j \in \ker u^{\mu_j} \setminus \ker u^{\mu_j - 1}$, et que la somme

$$V = \mathrm{Vect}\big\{u^k(x_1) \;\big|\; k \geq 0\big\} + \mathrm{Vect}\big\{u^k(x_2) \;\big|\; k \geq 0\big\} + \cdots + \mathrm{Vect}\big\{u^k(x_{i-1}) \;\big|\; k \geq 0\big\} \qquad (\S)$$

est directe.

Proposition. Sous ces conditions :

(a) Soit x_i un élément de $\ker u^{\mu_i}$. Pour qu'il vérifie $u^{\mu_i - 1}(x_i) \neq 0$ et que la somme

$$V + \mathrm{Vect}\big\{u^k(x_i) \;\big|\; k \geq 0\big\}$$

soit directe, il faut et il suffit que x_i n'appartienne pas à $\big(\ker u^{\mu_i - 1}\big) + V$.

(b) L'espace vectoriel $\big(\ker u^{\mu_i}\big) \cap \big(\ker u^{\mu_i - 1} + V\big)$ est de dimension

$$\lambda'_1 + \lambda'_2 + \cdots + \lambda'_{\mu_i - 1} + i - 1.$$

Preuve de la proposition.

(a) Utilisant que μ_i est plus petit que μ_1, μ_2, ..., μ_{i-1}, nous observons que

$$(\ker u) \cap V = \mathrm{Vect}\big\{u^{\mu_j - 1}(x_j) \;\big|\; j \in [\![1, i-1]\!]\big\} \subset u^{\mu_i - 1}(V).$$

Maintenant, si $u^{\mu_i - 1}(x_i) = 0$ ou si l'intersection $V \cap \mathrm{Vect}\big\{u^k(x_i) \;\big|\; k \geq 0\big\}$ n'est pas réduite à $\{0\}$, alors $u^{\mu_i - 1}(x_i)$ appartient à V, donc à $(\ker u) \cap V$, donc à $u^{\mu_i - 1}(V)$, et par conséquent $x_i \in \big(\ker u^{\mu_i - 1}\big) + V$. Réciproquement, si $x_i \in \big(\ker u^{\mu_i - 1}\big) + V$, alors $u^{\mu_i - 1}(x_i) \in V$, et donc soit $u^{\mu_i - 1}(x_i) = 0$, soit $V \cap \mathrm{Vect}\big\{u^k(x_i) \;\big|\; k \geq 0\big\} \neq \{0\}$.

(b) L'espace vectoriel $\big(\ker u^{\mu_i}\big) \cap \big(\ker u^{\mu_i - 1} + V\big) = \big(\ker u^{\mu_i - 1}\big) + \big(\ker u^{\mu_i} \cap V\big)$ est de dimension

$$\dim\big(\ker u^{\mu_i - 1}\big) + \dim\big(\ker u^{\mu_i} \cap V\big) - \dim\big(\ker u^{\mu_i - 1} \cap V\big).$$

Or, utilisant que μ_i est plus petit que μ_1, μ_2, ..., μ_{i-1}, nous vérifions que

$$\dim\big(\ker u^{\mu_i} \cap V\big) - \dim\big(\ker u^{\mu_i - 1} \cap V\big) = i - 1$$

(chacun des $i - 1$ termes de la somme directe (§) contribue d'une unité).

Revenant à notre problème, nous voyons alors qu'il y a

$$q^{\lambda'_1 + \lambda'_2 + \cdots + \lambda'_{\mu_i - 1}} \Big(q^{\lambda'_{\mu_i}} - q^{i-1}\Big)$$

choix possibles pour x_i. Le processus complet nous donne ainsi exactement

$$\prod_{i=1}^{m} q^{\lambda'_1 + \lambda'_2 + \cdots + \lambda'_{\mu_i - 1}} \Big(q^{\lambda'_{\mu_i}} - q^{i-1}\Big)$$

familles de vecteurs $(x_1, x_2, \dots, x_m)$ vérifiant les conditions imposées. En regroupant, pour chaque $j \geq 1$, les facteurs pour lesquels $\mu_i = j$, nous réécrivons ce nombre sous la forme

$$\prod_{j\geq 1} \prod_{i=\mu'_{j+1}+1}^{\mu'_j} \left(q^{\lambda'_1+\lambda'_2+\cdots+\lambda'_{j-1}}\left(q^{\lambda'_j} - q^{i-1}\right)\right),$$

où nous convenons que $\mu'_j = 0$ si $j > m'$.

Cependant l'application $(x_1, x_2, \dots, x_m) \mapsto F$ n'est pas injective : le raisonnement ci-dessus, appliqué au sous-espace F plutôt qu'à l'espace E, prouve que chaque F contient

$$\prod_{j\geq 1} \prod_{i=\mu'_{j+1}+1}^{\mu'_j} \left(q^{\mu'_1+\mu'_2+\cdots+\mu'_{j-1}}\left(q^{\mu'_j} - q^{i-1}\right)\right)$$

familles $(x_1, x_2, \dots, x_m)$. Appliquant le lemme des bergers, nous en déduisons que

$$b_{\lambda,\mu}(q) = \prod_{j\geq 1} \prod_{i=\mu'_{j+1}+1}^{\mu'_j} \frac{q^{\lambda'_1+\lambda'_2+\cdots+\lambda'_{j-1}}(q^{\lambda'_j} - q^{i-1})}{q^{\mu'_1+\mu'_2+\cdots+\mu'_{j-1}}(q^{\mu'_j} - q^{i-1})},$$

ce qui se réécrit

$$b_{\lambda,\mu}(q) = \prod_{j\geq 1} q^{\mu'_{j+1}(\lambda'_j-\mu'_j)} \binom{\lambda'_j - \mu'_{j+1}}{\mu'_j - \mu'_{j+1}}_q.$$

Or les coefficients binomiaux de Gauss sont des polynômes en q à coefficients entiers positifs ; voir par exemple la page Wikipédia qui leur est consacrée. Par suite, $b_{\lambda,\mu}(q)$ est également un polynôme à coefficients entiers positifs.

Les coefficients binomiaux de Gauss, évalués en $q = 1$, sont les coefficients binomiaux ordinaires. L'égalité demandée pour la valeur $a_\lambda(1)$ se présente dès lors comme une identité faisant intervenir une somme de produits de coefficients binomiaux ; la somme vient de l'expression (†) et porte sur toutes les suites décroissantes $(\mu'_1, \mu'_2, \dots, \mu'_{\ell'})$ telles que $\mu'_j \in [\![0, \lambda'_j]\!]$ pour chaque $j \in [\![1, \ell']\!]$. Le calcul semble intimidant à première vue, mais il est en réalité tout simple lorsque l'on le prend par le bon bout, c'est-à-dire si l'on somme d'abord sur μ'_1, puis sur μ'_2, ..., jusqu'à μ'_ℓ. De façon formalisée, en utilisant la formule du binôme de Newton, on vérifie par récurrence sur $i \in [\![1, \ell']\!]$ que pour tout $\mu'_{i+1} \in [\![0, \lambda'_i]\!]$, on a

$$\sum_{\mu'} \prod_{j=1}^{i} \binom{\lambda'_j - \mu'_{j+1}}{\mu'_j - \mu'_{j+1}} = \left(\prod_{j=1}^{i-1} (j+1)^{\lambda'_j-\lambda'_{j+1}}\right) \times (i+1)^{\lambda'_i-\mu'_{i+1}},$$

la somme dans le membre de gauche portant sur l'ensemble des suites $(\mu'_1, \mu'_2, \dots, \mu'_i)$ telles que $\mu'_1 \geq \mu'_2 \geq \cdots \geq \mu'_i \geq \mu'_{i+1}$ et $\mu'_j \in [\![0, \lambda'_j]\!]$ pour chaque $j \in [\![1, i]\!]$. Le cas particulier $i = \ell'$ et $\mu'_{\ell'+1} = 0$ cette identité nous donne

$$a_\lambda(1) = \prod_{j\geq 1} (j+1)^{\lambda'_j-\lambda'_{j+1}},$$

et c'est là l'égalité demandée $a_\lambda(1) = (\lambda_1+1)(\lambda_2+1)\cdots(\lambda_\ell+1)$ puisque $\lambda'_j - \lambda'_{j+1}$ est le nombre de parts de λ égales à j. La vérification concernant la valeur $a_\lambda(0)$ est un petit exercice combinatoire laissé au lecteur.

Le principe de cette seconde méthode apparaît dans l'article *Number of the subgroups of any given abelian group* de G. A. Miller, Proc. Nat. Acad. Sci. **25** (1939), 258–262. La formule explicite pour les nombres que nous avons notés $b_{\lambda,\mu}(q)$ se trouve dans de nombreuses références, dont l'ouvrage *Subgroup Lattices and Symmetric Functions* de Lynne M. Butler, Mem. Amer. Math. Soc. vol. 112, nr. 539 (1994) et l'article *On computing the number of subgroups of a finite abelian group* de Thomas Stehling, Combinatorica **12** (1992), 475–479.

Question subsidiaire.

La décomposition de Frobenius permet d'écrire E comme somme directe $E_1 \oplus E_2 \oplus \cdots \oplus E_\ell$ de sous-espaces cycliques, la matrice de $u|_{E_i}$ étant la matrice compagnon d'un polynôme $P_i \in \mathbb{K}[X]$, avec de surcroît $P_\ell \mid \cdots \mid P_2 \mid P_1$. Ces polynômes P_i sont les facteurs invariants de u.

Soit Π l'ensemble des polynômes irréductibles dans $\mathbb{K}[X]$. Pour chaque $f \in \Pi$, notons $\lambda(f)$ la partition $(\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ où f^{λ_i} est la plus grande puissance de f divisant P_i. On peut alors démontrer que le nombre de sous-espaces stables de u est égal au produit

$$\prod_{f\in\Pi} a_{\lambda(f)}(q^{\deg(f)}).$$

(Lorsque f ne divise pas P_1, la partition $\lambda(f)$ est vide : toutes ses parts sont réduites à zéro. Dans ce cas, $a_{\lambda(f)}$ est le polynôme constant égal à 1. Le produit ci-dessus ne comporte donc qu'un nombre fini de facteurs non triviaux.)

La preuve consiste à décomposer E selon les sous-espaces caractéristiques de u :

$$E = \ker P_1(u) = \bigoplus_{f\in\Pi} E_f \quad \text{où} \quad E_f = \ker f^{\lambda_1(f)}(u).$$

Alors tout sous-espace vectoriel F de E stable par u est nécessairement de la forme $\bigoplus_{f\in\Pi} F_f$, où F_f est un sous-espace de E_f stable par $u|_{E_f}$: de fait, les $F_f = F \cap E_F$ sont les sous-espaces caractéristiques de l'endomorphisme induit $u|_F$.

Il reste encore à comprendre comment ramener le dénombrement de ces sous-espaces F_f à notre étude du cas des endomorphismes nilpotents, de façon à faire apparaître les polynômes $a_{\lambda(f)}$. Trois méthodes (au moins) sont ici possibles. Le film est cependant déjà assez long comme cela et c'est un bon moment pour le clap de fin.

Remarque finale.

Soit p un nombre premier. Pour chaque p-groupe abélien fini G, il existe une unique partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ telle que G soit isomorphe à

$$\mathbb{Z}/p^{\lambda_1}\mathbb{Z} \times \mathbb{Z}/p^{\lambda_2}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{\lambda_\ell}\mathbb{Z}.$$

Alors $a_\lambda(p)$ est le nombre de sous-groupes de G.

Sujet d'analyse
par Dominique Malicet (Université Gustave Eiffel)

On se pose la question de la nature de la série $\sum_{n\geq 1} \frac{1}{n^{1+|\sin(nt)|}}$ pour $t \in \mathbb{R}$. Cette série fait penser à une série de Riemann, mais avec un exposant oscillant entre 1 et 2 et il n'est a priori pas évident de deviner la nature. L'objectif est de répondre à la question dans un contexte plus général où l'on remplace la fonction sinus par une fonction périodique quelconque.

L'étude de la nature de ce type de séries dont le terme général s'exprime à l'aide d'évaluations de fonctions périodiques comme $x \mapsto \sin(xt)$ et $x \mapsto \cos(xt)$ s'avère souvent délicate. Elle est fréquemment liée à des estimées fines sur la dynamique d'une rotation sur le cercle. Parfois, la nature de la série dépend de propriétés arithmétiques de la période (par exemple sur le développement en fraction continue). Quelques exemples illustratifs parmi de nombreux autres :

- C'est un exercice facile et classique que $\sum_{n\geq 1} \frac{\sin(nt)}{n}$ converge quel que soit t.
- Hardy et Littlewood ont déterminé sous quelles conditions arithmétiques précises sur t la série $\sum_{n\geq 1} \frac{\sin(n^2 t)}{n}$ converge (en particulier, la nature de la série dépend de t).
- Hardy et Littlewood ont également démontré que la série $\sum_{n\geq 1} \frac{\sin(n^2 t)}{\sqrt{n}}$ est toujours divergente si $t/2\pi$ est irrationnel.
- Plus récemment, il a été montré que la série $\sum_{n\geq 1} \frac{\sin(n^3 t)}{n}$ est toujours convergente.
- Pour $\alpha > 0$ la série $\sum_{n\geq 1} \frac{(-1)^n}{n^\alpha + \sin(nt)}$ converge pour $t \in [0, 2\pi]$ sauf pour un nombre fini de valeurs de t (exercice assez élémentaire), alors que la série $\sum_{n\geq 1} \frac{(-1)^n}{\ln(n) + \sin(nt)}$ converge par exemple pour $t = 1$ grâce à certaines propriétés arithmétiques de π (degré d'irrationalité), mais il existe un ensemble indénombrable de valeurs de t pour laquelle la série diverge.

Le problème proposé n'est pas aussi complexe que certains des exemples énoncés ci dessus, mais repose tout de même sur certaines subtilités de la répartition des suites $(nt)_{n\in\mathbb{N}}$ modulo 1.

Questions

Etant donné une fonction $f : \mathbb{R} \to \mathbb{R}$ périodique, on considère la série $\sum_{n\geq 1} \frac{1}{n^{1+f(n)}}$.

(a) On suppose que f est lipschitzienne et vérifie $f(0) = 0$. Etudier la nature de la série.

(b) On suppose que f est lipschitzienne, qu'elle s'annule (pas forcément en 0) et que sa période est irrationnelle. Etudier la nature de la série.

(c) Soit $\alpha < 1$. On pose pour $t > 0$ $f(x) = |\sin(2\pi t x)|^\alpha$. Montrer qu'il existe au moins un nombre irrationnel t tel que la série diverge, et au moins un nombre irrationnel t tel que la série converge.

Eléments de solution

Question 1 Nous allons montrer que la série est toujours divergente. Soit T une période de f et L sa constante de Lipschitz. Pour tout réel x, on note $r(x)$ le représentant de x modulo T dans $[-T/2, T/2[$. Il est assez clair que le cas difficile est lorsque T est irrationnel et f est positive (mais la méthode que l'on propose ici ne nécessite pas de distinguer des cas). Il y a plusieurs moyens de démontrer la divergence de la série, et en effet plusieurs approches différentes et correctes ont été proposées par des équipes participantes. Ces approches reposent cependant toujours sur la même idée : justifier, de manière suffisamment fine et quantitative, qu'il existe beaucoup d'entiers n tels que le représentant de $r(n)$ est suffisamment proche de 0, de sorte que l'exposant $1 + f(n)$ soit proche de 1.

Commençons par un lemme :

Lemme 1 *Pour tout n dans $\mathbb{N}$, pour tout ε dans $]0, \frac{1}{2}]$, il existe n entiers positifs distincts $p_1, \ldots, p_n$ vérifiant $p_i \leq \frac{n}{\varepsilon}$ et $|r(p_i)| \leq 2T\varepsilon$.*

Preuve On découpe $[-T/2, T/2[$ en $N = \left[\frac{1}{\varepsilon}\right]$ intervalles de longueur $\frac{T}{N} \leq 2T\varepsilon$. Parmi les $Nn + 1$ nombres $0, 1, \ldots, Nn$, on peut trouver par le principe des tiroirs généralisés $n + 1$ nombres $m_0 < m_1 < \cdots < m_n$ tels que $r(m_0), \ldots, r(m_n)$ soient dans un même intervalle. En posant alors $p_k = m_k - m_0$ pour $k = 1, \ldots, n$, on a bien $|r(p_k)| \leq 2T\varepsilon$ et $p_k \leq nN \leq \frac{n}{\varepsilon}$.

Soit maintenant $(\varepsilon_k)_{k\in\mathbb{N}}$ une suite de réels dans $]0, \frac{1}{2}]$. On définit par récurrence une suite $(n_k)_{k\geq 1}$ comme suit : une fois $n_1, \ldots, n_{k-1}$ définis, on chosit grâce au lemme un entier n_k différent de $n_1, \ldots, n_{k-1}$ vérifiant $n_k \leq \frac{k}{\varepsilon_k}$ et $r(n_k) \leq 2T\varepsilon_k$. Ainsi on obtient une suite $(n_k)_{k\geq 1}$ d'entiers deux à deux distincts vérifiant d'une part

$$\forall k \geq 1, n_k \leq \frac{k}{\varepsilon_k},$$

et d'autre part,

$$\forall k \geq 1, f(n_k) = f(r(n_k)) \leq C\varepsilon_k$$

où $C = 2TL$. Donc on peut minorer la somme à étudier comme suit :

$$\sum_{n=1}^{+\infty} \frac{1}{n^{1+f(n)}} \geq \sum_{k=1}^{+\infty} \frac{1}{n_k^{1+f(n_k)}} \geq \sum_{k=1}^{+\infty} \left(\frac{\varepsilon_k}{k}\right)^{1+C\varepsilon_k}.$$

On choisit $\varepsilon_k = \dfrac{1}{\ln(k)}$ pour $k \geq 8$ (de sorte que $\varepsilon_k \in]0, \frac{1}{2}]$), et par exemple $\varepsilon_k = \dfrac{1}{2}$ pour $k < 8$. Ainsi, lorsque k tend vers $+\infty$,

$$\left(\frac{\varepsilon_k}{k}\right)^{1+C\varepsilon_k} = \frac{1}{(k\ln(k))^{1+\frac{C}{\ln(k)}}} \sim \frac{e^{-C}}{k\ln(k)},$$

Ce dernier terme étant le terme général d'une série divergente, on conclut que $\sum_{k=1}^{+\infty} \left(\frac{\varepsilon_k}{k}\right)^{1+C\varepsilon_k} = +\infty$ et donc que $\sum_{n=1}^{+\infty} \frac{1}{n^{1+f(n)}} = +\infty$.

Question 2 Nous allons montrer de nouveau que la série est toujours divergente. L'idée est la même : si on note x un point où f s'annule, on veut justifier qu'il existe suffisamment d'entiers n tels que $r(n-x)$ est proche de 0, de sorte à ce que l'exposant $1+f(n)$ soit proche de 1. C'est cependant plus difficile à quantifier lorsque x est non nul, car le principe des tiroirs ne suffit plus. Ici l'irrationalité de T aura un rôle important, le résultat étant faux sinon (par exemple si $f(x) = \cos(\frac{2\pi x}{3})$).

En gardant les mêmes notations que précédemment, montrons un nouveau lemme :

Lemme 2 *Il existe un ensemble $E \subset]0, \frac{1}{2}[$ s'accumulant en 0 tel que pour tout $\varepsilon \in E$ et pour tout nombre réel x, il existe un entier positif m tel que $m \leq \frac{1}{\varepsilon}$ et $|r(m-x)| \leq 2T\varepsilon$.*

Preuve Puisque $\frac{1}{T}$ est irrationnel, par le lemme d'approximation diophantienne de Dirichlet, il existe une suite de nombres rationnels tendant vers $\frac{1}{T}$, tel qu'en les écrivant sous la forme $\frac{p}{q}$, $p \wedge q = 1$, on a $|\frac{1}{T} - \frac{p}{q}| \leq \frac{1}{q^2}$. Alors, pour tout entier $m \leq q$, $|m - \frac{mp}{q}T| \leq \frac{T}{q}$. Si x est un réel quelconque, on peut trouver un entier a tel que $|\frac{a}{q}T - x| \leq \frac{T}{q}$, puis trouver $m \leq q$ tel que $mp = a \bmod q$. On obtient alors $|r(m-x)| \leq \frac{2T}{q}$. Ceci donne le résultat voulu pour le nombre particulier $\varepsilon = \frac{1}{q}$. Mais q pouvant être arbitrairement grand, on obtient un ensemble de nombres s'accumulant en 0.

On fixe l'ensemble E donné par le lemme. En cumulant ce lemme avec le lemme 1, on déduit le lemme suivant :

Lemme 3 *Pour tout n dans $\mathbb{N}$, pour tout ε dans E et pour tout nombre réel x, il existe n entiers positifs distincts $p_1, \ldots, p_n$ vérifiant $p_i \leq \frac{n+2}{\varepsilon}$ et $0 < |r(p_i - x)| \leq 4T\varepsilon$.*

Preuve On prend $p_i = m + p'_i$ où m est donné par le lemme 2 et $p'_1, \ldots, p'_{n+1}$ sont donnés par le lemme 1. On a bien $p_i \leq \frac{n+2}{\varepsilon}$ et $|r(p_i - x)| \leq 4T\varepsilon$. De plus, il existe au plus un entier p_i tel que $r(p_i - x) = 0$ (car T est irrationnel), entier qu'on exclut le cas échéant. Quitte à renommer les indices on obtient bien n nombres $p_1, \ldots, p_n$ vérifiant les propriétés souhaitées.

Nous pouvons maintenant prouver la divergence de la série. Soit x un point où f s'annule. Soit ε dans E. Posons

$$A_\varepsilon = \{n \in \mathbb{N}, 0 < |r(n-x)| \leq 4T\varepsilon\}.$$

On construit par récurrence une suite $(n_k)_{k\geq 1}$ comme suit : si $n_1, \ldots, n_{k-1}$ sont définis, on choisit grâce au lemme 3 un entier n_k différent de $n_1, \ldots, n_{k-1}$ vérifiant $n_k \leq \dfrac{k+2}{\varepsilon}$ et $0 < |r(n_k - x)| \leq 4T\varepsilon$. Ceci implique que

$$\forall k \geq 1, f(n_k) = f(n_k) - f(x) \leq L|r(n_k - x)| \leq C\varepsilon$$

où $C = 4TL$. On obtient donc la minoration :

$$\sum_{n\in A_\varepsilon} \frac{1}{n^{1+f(n)}} \geq \sum_{k=1}^{+\infty} \left(\frac{\varepsilon}{k+2}\right)^{1+C\varepsilon} \geq \int_3^{+\infty} \left(\frac{\varepsilon}{t}\right)^{1+C\varepsilon} dt = \frac{1}{C}\left(\frac{\varepsilon}{3}\right)^{C\varepsilon}.$$

Quand ε tend vers 0, le terme de droite tend vers une constante strictement positive. Mais si la somme $\sum_{n=1}^{+\infty} \frac{1}{n^{1+f(n)}}$ est finie, on déduit par convergence dominée que $\sum_{n\in A_\varepsilon} \frac{1}{n^{1+f(n)}} = \sum_{n=1}^{+\infty} \chi_{A_\varepsilon}(n) \frac{1}{n^{1+f(n)}}$ tend vers 0 quand ε tend vers 0. Ceci étant une contradiction, on conclut que la série diverge.

Question 3 Nous proposons ici une approche non constructive pour répondre à la question. Il est cependant possible, bien qu'un peu plus technique, de définir explicitement des nombres réels t répondant à la question, ce que plusieurs équipes ont fait.

Posons $S(t) = \sum_{n=1}^{+\infty} \frac{1}{n^{1+|\sin(2\pi tn)|^\alpha}}$, et $A = \{S = +\infty\}$. Si $t = \frac{p}{q}$ est rationnel, alors $|\sin(2\pi tn)|^\alpha = 0$ quand n est multiple de q, donc on déduit que $S(t) = +\infty$, donc $\mathbb{Q} \subset A$. De plus, pour tout $k \in \mathbb{N}$, l'ensemble $\{S > k\}$ est ouvert, donc A est un G_δ. Mais c'est une conséquence classique du lemme de Baire que $\mathbb{Q}$ n'est pas un G_δ. Donc A contient des nombres irrationnels.

D'autre part, on a

$$\int_0^1 S(t)dt = \sum_{n=1}^{+\infty} \int_0^1 \frac{1}{n^{1+|\sin(2\pi tn)|^\alpha}} dt = \sum_{n=1}^{+\infty} \int_0^1 \frac{1}{n^{1+|\sin(2\pi s)|^\alpha}} ds \leq 1 + \int_0^1 \frac{1}{|\sin(2\pi s)|^\alpha} ds$$

(la dernière inégalité s'obtient par exemple par comparaison série-intégrale). Puisque cette dernière intégrale est finie, on déduit que $S(t) < +\infty$ pour presque tout t dans $[0,1]$. Ainsi A^c est de mesure nulle dans $[0,1]$ donc il contient des irrationnels.

Sujet d'arithmétique
par Jean-Paul Allouche (Sorbonne Université)

Notations :

Soit φ l'indicatrice d'Euler : pour k entier ≥ 1, $\varphi(k)$ est le nombre d'entiers dans l'intervalle $[1,k]$ premiers avec k.

Soit r un entier impair > 1 et soit $\mathfrak{o}_2(r)$ l'ordre multiplicatif de 2 modulo r, c'est-à-dire le plus petit entier $\ell > 0$ tel que $2^\ell \equiv 1 \bmod r$.

Si a et b sont deux entiers ≥ 1, on note $\gcd(a,b)$ leur pgcd. On convient aussi que si b est non nul, $\gcd(0,b) = b$.

On note Card A ou $|A|$ le nombre d'éléments d'un ensemble fini : $\mathrm{Card}\, A = |A| = \sum_{k\in A} 1$.

1. Démontrer que, pour tout entier $n \geq 0$, on a

$$\sum_{\substack{d|2n+1\\ d\neq 1}} \frac{\varphi(d)}{\mathfrak{o}_2(d)} = \left(\frac{1}{\mathfrak{o}_2(2n+1)} \sum_{j=0}^{\mathfrak{o}_2(2n+1)-1} \gcd(2^j-1, 2n+1) \right) - 1.$$

2. On note $(e_1, e_2, \ldots, e_{2n})$ la base canonique de $\mathbb{R}^{2n}$. Soit f l'endomorphisme de $\mathbb{R}^{2n}$ défini par :

$$f(e_j) := \begin{cases} e_{2j} & \text{si } 1 \leq j \leq n; \\ e_{2j-(2n+1)} & \text{si } n+1 \leq j \leq 2n. \end{cases}$$

Montrer que le polynôme caractéristique de f est égal à

$$\prod_{\substack{d|2n+1\\ d\neq 1}} (X^{\mathfrak{o}_2(d)} - 1)^{\varphi(d)/\mathfrak{o}_2(d)}.$$

3. On note log le logarithme népérien. Montrer que

$$\sum_{\substack{d|2n+1\\ d\neq 1}} \frac{\varphi(d)}{\mathfrak{o}_2(d)} = O\left(\frac{n}{\log n}\right).$$

Eléments de solution

1. a) Démonstration « arithmétique ». Définissons, pour $n \geq 1$, la quantité $U(n)$ par :

$$U(n) := \sum_{1 \leq x \leq 2n} \mathrm{Card}\{j \in [0, \mathfrak{o}_2(2n+1)-1] : (2^j - 1)x \equiv 0 \bmod (2n+1)\} = \sum_{1 \leq x \leq 2n} A(x)$$

où $A(x) := \mathrm{Card}\{j \in [0, \mathfrak{o}_2(2n+1)-1] : (2^j - 1)x \equiv 0 \bmod (2n+1)\}$.

Nous pouvons écrire chaque terme de cette somme de la manière suivante :

$$\begin{aligned}
A(x) &= \mathrm{Card}\left\{ j \in [0, \mathfrak{o}_2(2n+1)-1] : (2^j-1)x' \equiv 0 \bmod \frac{2n+1}{\gcd(x, 2n+1)} \right\} \text{ où } x' := \frac{x}{\gcd(x, 2n+1)} \\
&= \mathrm{Card}\left\{ j \in [0, \mathfrak{o}_2(2n+1)-1] : (2^j-1) \equiv 0 \bmod \frac{2n+1}{\gcd(x, 2n+1)} \right\} \text{ car } \gcd\left(x', \frac{(2n+1)}{\gcd(x, 2n+1)}\right) = 1 \\
&= \mathrm{Card}\left\{ j \in [0, \mathfrak{o}_2(2n+1)-1] : j \text{ multiple de } \mathfrak{o}_2\left(\frac{(2n+1)}{\gcd(x, 2n+1)}\right) \right\} \\
&= \frac{\mathfrak{o}_2(2n+1)}{\mathfrak{o}_2\left(\frac{2n+1}{\gcd(x, 2n+1)}\right)}
\end{aligned}$$

(si l'on est pointilleux, il est facile de s'assurer que $\mathfrak{o}_2\left(\frac{2n+1}{\gcd(x, 2n+1)}\right)$ divise $\mathfrak{o}_2(2n+1)$).

Donc, en remarquant que $x \in [1, 2n] \Longrightarrow \gcd(x, 2n+1) < 2n+1$,

$$\begin{aligned}
\frac{U(n)}{\mathfrak{o}_2(2n+1)} &= \sum_{1 \leq x \leq 2n} \frac{1}{\mathfrak{o}_2\left(\frac{2n+1}{\gcd(x, 2n+1)}\right)} = \sum_{\substack{d \mid 2n+1 \\ d \neq 2n+1}} \sum_{\substack{1 \leq x \leq 2n \\ \gcd(x, 2n+1) = d}} \frac{1}{\mathfrak{o}_2\left(\frac{2n+1}{d}\right)} \\
&= \sum_{\substack{d \mid 2n+1 \\ d \neq 2n+1}} \frac{1}{\mathfrak{o}_2\left(\frac{2n+1}{d}\right)} \sum_{\substack{1 \leq x \leq 2n \\ \gcd(x, 2n+1) = d}} 1 = \sum_{\substack{d \mid 2n+1 \\ d \neq 2n+1}} \frac{1}{\mathfrak{o}_2\left(\frac{2n+1}{d}\right)} \sum_{\substack{1 \leq x < 2n+1 \\ \gcd(x, 2n+1) = d}} 1 \\
&= \sum_{\substack{d \mid 2n+1 \\ d \neq 2n+1}} \frac{1}{\mathfrak{o}_2\left(\frac{2n+1}{d}\right)} \sum_{\substack{1 \leq x' < (2n+1)/d \\ \gcd(x', (2n+1)/d) = 1}} 1 \quad \text{(changement d'indice } x' = x/d) \\
&= \sum_{\substack{d \mid 2n+1 \\ d \neq 2n+1}} \frac{\varphi\left(\frac{2n+1}{d}\right)}{\mathfrak{o}_2\left(\frac{2n+1}{d}\right)}
\end{aligned}$$

et donc, en faisant le changement d'indice $e = (2n+1)/d$,

$$\frac{U(n)}{\mathfrak{o}_2(2n+1)} = \sum_{\substack{e \mid 2n+1 \\ e \neq 1}} \frac{\varphi(e)}{\mathfrak{o}_2(e)}. \tag{1}$$

Mais nous pouvons évaluer $U(n)$ d'une autre manière. En effet il est clair que :

$$U(n) = \mathrm{Card}\{(j,x) \in [0, \mathfrak{o}_2(2n+1)-1] \times [1,2n] : (2^j-1)x \equiv 0 \bmod (2n+1)\}$$

$$= \sum_{j=0}^{\mathfrak{o}_2(2n+1)-1} \mathrm{Card}\{x \in [1,2n] : (2^j-1)x \equiv 0 \bmod (2n+1)\}.$$

et donc

$$U(n) = \sum_{j=0}^{\mathfrak{o}_2(2n+1)-1} \mathrm{Card}\{x \in [1,2n] : (2^j-1)x \equiv 0 \bmod (2n+1)\} \tag{2}$$

Pour j fixé dans l'intervalle $[0, \mathfrak{o}_2(2n+1)-1]$, soit $d := \gcd(2^j-1, 2n+1)$.

– La congruence $(2^j-1)x \equiv 0 \bmod 2n+1$, avec $x \in [1,2n]$, implique $((2^j-1)/d)x \equiv 0 \bmod (2n+1)/d$, donc $x \equiv 0 \bmod (2n+1)/d$ car $(2^j-1)/d$ et $(2n+1)/d$ sont premiers entre eux. Le fait que x est multiple de $(2n+1)/d$ et la condition $x \in [1,2n] = [1,2n+1[$ impliquent donc l'existence d'un entier $\lambda \in [1,d[= [1,d-1]$ tel que $x = \lambda(2n+1)/d$.

– Réciproquement, si $x = \lambda(2n+1)/d$, avec $\lambda \in [1,d-1] = [1,d[$, on a $x \in [1,2n+1[= [1,2n]$ et $(2^j-1)x = (2^j-1)\lambda(2n+1)/d = ((2^j-1)/d)\lambda(2n+1) \equiv 0 \bmod (2n+1)$.

Bref :
$\mathrm{Card}\{x \in [1,2n] : (2^j-1)x \equiv 0 \bmod (2n+1)\} = \mathrm{Card}\{\lambda \in [1,d-1]\} = d-1 = \gcd(2^j-1, 2n+1)-1$,
et par conséquent, en utilisant l'égalité (2) :

$$\frac{U(n)}{\mathfrak{o}_2(2n+1)} = \frac{1}{\mathfrak{o}_2(2n+1)} \left(\sum_{j=0}^{\mathfrak{o}_2(2n+1)-1} \gcd(2^j-1, 2n+1) \right) - 1. \tag{3}$$

La comparaison des égalités (1) et (3) donne le résultat annoncé. □

1. b) Démonstration algébrique. Posons $m := 2n+1$. L'égalité à prouver est

$$\sum_{d|m} \varphi(d) \frac{\mathfrak{o}_2(m)}{\mathfrak{o}_2(d)} = \sum_{j=0}^{\mathfrak{o}_2(m)-1} \gcd(2^j-1, m). \tag{4}$$

Comme m est impair, 2 appartient au groupe $(\mathbb{Z}/m\mathbb{Z})^\times$ des éléments inversibles de $\mathbb{Z}/m\mathbb{Z}$. Notons $\langle 2\rangle_m$ le sous-groupe de $(\mathbb{Z}/m\mathbb{Z})^\times$ engendré par 2 : c'est un groupe d'ordre $\mathfrak{o}_2(m)$.

Soit d un diviseur de m. Il existe un unique morphisme d'anneaux de $\mathbb{Z}/m\mathbb{Z}$ vers $\mathbb{Z}/d\mathbb{Z}$. Ce morphisme est surjectif et son noyau est l'idéal $d\mathbb{Z}/m\mathbb{Z}$. Le morphisme induit de $(\mathbb{Z}/m\mathbb{Z})^\times$ vers $(\mathbb{Z}/d\mathbb{Z})^\times$ envoie $\langle 2\rangle_m$ sur $\langle 2\rangle_d$. Par conséquent, $\langle 2\rangle_d$ est le quotient de $\langle 2\rangle_m$ par le sous-groupe

$\{y \in \langle 2\rangle_m : y \equiv 1 \bmod d\}$. Le théorème de Lagrange nous dit alors que le membre de gauche de (4) est

$$\sum_{d|m} \varphi(d) \frac{|\langle 2\rangle_m|}{|\langle 2\rangle_d|} = \sum_{d|m} \varphi(d) \times |\{y \in \langle 2\rangle_m : y \equiv 1 \bmod d\}|.$$

Nous pouvons écrire le membre de droite de cette égalité sous la forme

$$\begin{aligned} \sum_{d|m} \varphi(d) \times |\{j \in [0, \mathfrak{o}_2(m)-1] : d|(2^j-1)\}| &= \sum_{0\le j\le \mathfrak{o}_2(m)-1} \sum_{\substack{d|m \\ d|2^j-1}} \varphi(d) \\ &= \sum_{0\le j\le \mathfrak{o}_2(m)-1} \sum_{d|\gcd(2^j-1,m)} \varphi(d). \end{aligned}$$

La formule d'Euler $\sum_{d|k} \varphi(d) = k$ donne alors le résultat souhaité. □

Remarque. L'égalité démontrée ici est à l'origine l'observation de deux calculs différents du nombre de cycles dans la décomposition de la permutation σ de $[1, 2n]$ définie par

$$\sigma := \begin{pmatrix} 1 & 2 & \dots & n & n+1 & n+2 & \dots & 2n \\ 2 & 4 & \dots & 2n & 1 & 3 & \dots & 2n-1 \end{pmatrix}.$$

On peut aussi voir que ces quantités donnent le nombre d'orbites de l'application "doublement modulo $(2n+1)$", c'est-à-dire de l'application $x \to 2x$, sur $\mathbb{Z}/(2n+1)\mathbb{Z} \setminus \{0\}$. Elles donnent aussi, en ajoutant 1, le nombre de facteurs irréductibles du polynôme $Z^{2n+1} - 1$ sur le corps à deux éléments $\mathbb{F}_2$.

2.

Pour démontrer l'assertion concernant le polynôme caractéristique de f, on considère comme dans la remarque ci-dessus la permutation σ de $[1, 2n]$ définie par

$$\sigma := \begin{pmatrix} 1 & 2 & \dots & n & n+1 & n+2 & \dots & 2n \\ 2 & 4 & \dots & 2n & 1 & 3 & \dots & 2n-1 \end{pmatrix}.$$

Il est immédiat que, pour tout j dans $[1, 2n]$, on a $\sigma(j) \equiv 2j \bmod (2n+1)$ et que $f(e_j) = e_{\sigma(j)}$. Un résultat classique sur le polynôme caractéristique d'une matrice de permutation (ici la matrice associée à f dans la base canonique de $\mathbb{R}^{2n}$) nous donne :

$$\det(Xid - f) = \prod_{1\le k\le n} (X^k - 1)^{c_k},$$

où c_k est le nombre de cycles de longueur k dans la décomposition de σ en produit de cycles (pour redémontrer ce résultat, il suffit de réordonner les (e_j) suivant les cycles de la décomposition de σ : en commençant par $e_1, e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma^t(1)}$, où t est la longueur du cycle contenant 1, puis, après $e_{\sigma(t)}$, on choisit un vecteur e_ℓ non déjà utilisé, etc.). Or l'orbite de j sous σ a pour cardinal $\mathfrak{o}_2((2n+1)/\varphi(j, 2n+1))$, donc, les cycles dans la décomposition de σ ont pour longueurs $\mathfrak{o}_2((2n+1)/d$, où d divise $2n+1$ et $d \neq 2n+1$ (car σ n'a pas de point fixe).

De plus le nombre de cycles de longueur d pour un tel d est $\varphi((2n+1)/d)/(\mathfrak{o}_2((2n+1)/d)$. Ainsi

$$\det(Xid - f) = \prod_{\substack{d|2n+1 \\ d\neq 2n+1}} (X^{\mathfrak{o}_2((2n+1)/d)} - 1)^{\varphi((2n+1)/d)/\mathfrak{o}_2((2n+1)/d)},$$

d'où, avec le changement d'indice de sommation $d' = (2n+1)/d$,

$$\det(Xid - f) = \prod_{\substack{d|2n+1 \\ d\neq 1}} (X^{\mathfrak{o}_2(d)} - 1)^{\varphi(d)/\mathfrak{o}_2(d)}. \qquad \square$$

3.

Pour démontrer le 3., on va couper la somme suivant les « petits » diviseurs et les « grands » diviseurs de $2n+1$. Plus précisément entre les diviseurs $\leq (2n+1)^\alpha$ et les diviseurs $> (2n+1)^\alpha$ pour un $\alpha > 0$ « petit » qui sera choisi plus tard. Écrivons

$$\sum_{\substack{d|2n+1 \\ d\neq 1}} \frac{\varphi(d)}{\mathfrak{o}_2(d)} = S_1(n) + S_2(n),$$

avec

$$S_1(n) := \sum_{\substack{d|2n+1 \\ d\neq 1,\ d\leq (2n+1)^\alpha}} \frac{\varphi(d)}{\mathfrak{o}_2(d)}$$

et

$$S_2(n) := \sum_{\substack{d|2n+1 \\ d> (2n+1)^\alpha}} \frac{\varphi(d)}{\mathfrak{o}_2(d)}.$$

Pour S_1, on minore $\mathfrak{o}_2(d)$ par 1, et l'on observe que, pour $d \leq (2n+1)^\alpha$, on a $\varphi(d) \leq d \leq (2n+1)^\alpha$. D'où

$$S_1(n) \leq (2n+1)^\alpha \sum_{\substack{d|2n+1 \\ d\neq 1,\ d\leq (2n+1)^\alpha}} 1 \leq (2n+1)^\alpha \ \ \mathrm{Card}\{d : d|2n+1\}.$$

Or le nombre de diviseurs d'un entier satisfait à l'inégalité :

$$\mathrm{Card}\{d : d|m\} \leq 2\sqrt{m}$$

(grouper les entiers qui divisent m deux par deux, d et m/d). Bref

$$S_1(n) \leq 2(2n+1)^{\alpha+1/2} = O(n^{\alpha+1/2}) = O\left(\frac{n}{\log n}\right) \text{ pour } \alpha < 1/2.$$

Pour la somme S_2, on note que si $2^k \equiv 1 \bmod d$ avec $k \neq 0$ et $d \neq 1$, on a $2^k \geq d+1$, donc $k \geq \log(d+1)/\log 2$. Donc, pour les d qui apparaissent dans la somme S_2, on a

$$\mathfrak{o}_2(d) \geq \frac{\log(d+1)}{\log 2} \geq \frac{\log(2n+1)}{\alpha \log 2}.$$

On en déduit que :

$$S_2(n) \leq \left(\sum_{\substack{d|2n+1 \\ d>(2n+1)^\alpha}} \varphi(d) \right) O(1/\log n) \leq \left(\sum_{d|2n+1} \varphi(d) \right) O(1/\log n) = (2n+1)\, O(1/\log n) \ \ (*)$$

D'où $S_2(n) = O(n/\log n)$. Il reste juste à démontrer que, pour tout entier $m \geq 1$, on a $\sum_{d|m} \varphi(d) = m$. Pour établir cette égalité, on peut procéder comme suit :

$$m = \sum_{1\leq k\leq m} 1 = \sum_{d|m} \sum_{\substack{1\leq k\leq m \\ \gcd(k,m)=d}} 1 = \sum_{d|m} \sum_{\substack{1\leq k'\leq m/d \\ \gcd(k',m/d)=1}} 1 = \sum_{d|m} \varphi(m/d) = \sum_{d'|m} \varphi(d')$$

où la dernière somme est obtenue à partir de l'avant-dernière en faisant le changement d'indice $d' = n/d$. $\square$

Sujet de combinatoire
par Lingmin Liao (Université Paris-Est Créteil)

On souhaite colorier le réseau $\mathbb{Z}^2$ en bleu et en rouge, suivant la règle suivante : on ne peut pas colorier en même temps les deux points (x, y) et $(2x, 2y)$ en rouge. Pour $n \in \mathbb{N}$, $n \geq 1$, notons X_n le nombre des façons de colorier les points $\mathbb{Z}^2 \cap [-n, n]^2$. Montrer que la suite $\frac{1}{n^2} \log X_n$ converge et calculer la limite.

Eléments de solution

1. Partition de $\mathbb{Z}^2 \cap [-n,n]^2$. D'après la règle de coloriage, on considère une partition de $\mathbb{Z}^2 \cap [-n,n]^2$. Posons

$$\mathcal{B}_n := \big\{(i,j) \in \mathbb{Z}^2 \cap [-n,n]^2 : 2 \nmid i, \text{ ou } 2 \nmid j\big\}.$$

Alors,

$$\mathbb{Z}^2 \cap [-n,n]^2 = \{(0,0)\} \cup \bigsqcup_{(i,j)\in\mathcal{B}_n} \mathcal{N}(i,j),$$

avec

$$\mathcal{N}(i,j) := \Big\{(2^k i, 2^k j) : k \geq 0 \text{ et } -n \leq 2^k i \leq n, -n \leq 2^k j \leq n\Big\}.$$

Remarquons que pour différents $(i,j),(i',j') \in \mathcal{B}_n$, la méthode de coloriage pour $\mathcal{N}(i,j)$ ne concerne pas cela pour $\mathcal{N}(i',j')$. Ainsi, si on peut calculer le nombre des façons $N(i,j)$ pour colorier $\mathcal{N}(i,j)$, alors,

$$X_n = 2 \cdot \prod_{(i,j)\in\mathcal{B}_n} N(i,j),$$

où le facteur 2 sorte du fait qu'il y a deux façons à colorier le point $(0,0)$.

Pour chaque $(i,j) \in \mathcal{B}_n$, notons

$$\ell(i,j) = \max\big\{k \geq 1 : -n \leq 2^{k-1} i \leq n, -n \leq 2^{k-1} j \leq n\big\}.$$

Nous trouverons, dans l'étape suivante, que $N(i,j)$ ne dépend que $\ell(i,j)$. Précisément, si $\ell(i,j) = m \geq 1$, alors, $N(i,j) = F_m$, où (F_m) est la suite de Fibonacci définie par

$$F_1 = 2, F_2 = 3, \text{ et } F_{m+2} = F_{m+1} + F_m, \ (\forall m \geq 1).$$

Nous regroupons donc les points (i,j) qui ont la même valeur de ℓ :

$$\mathbb{Z}^2 \cap [-n,n]^2 = \{(0,0)\} \cup \bigsqcup_{m=1}^{\lfloor \log_2 n \rfloor + 1} \ \bigsqcup_{(i,j)\in\mathcal{B}_n,\ \ell(i,j)=m} \mathcal{N}(i,j).$$

Notons $K_{n,m}$ le cardinal de $\{(i,j) \in \mathcal{B}_n, \ \ell(i,j) = m\}$. Alors,

$$X_n = 2 \cdot \prod_{m=1}^{\lfloor \log_2 n \rfloor + 1} F_m^{K_{n,m}}.$$

2. Comptage sur $N(i,j)$. Nous montrerons que si $\ell(i,j) = m \geq 1$, alors $N(i,j) = F_m$. En fait, nous comptons le nombre de façons pour colorier les points

$$(i,j), \ (2i,2j), \ (2^2 i, 2^2 j), \ldots, (2^{m-1} i, 2^{m-1} j),$$

avec la règle que nous ne pouvons pas colorier en rouge deux points consécutifs dans la liste ci-dessus.

Nous montrons par récurrence en m. Si $m = 1$, évidemment, nous avons deux façons pour colorier le seul point (i,j), d'où $N(i,j) = F_1$. Si $m = 2$, nous avons trois façons pour colorier

les deux points (i,j) et $(2i,2j)$ car nous devons éliminer le cas du coloriage : rouge, rouge pour (i,j), $(2i,2j)$.

Supposons les conclusions pour m et $m+1$ sont correctes. Alors, pour $\ell(i,j)=m+2$, considérons le coloriage du dernier point $(2^{m-1}i, 2^{m-1}j)$. Si on colorie ce point en bleu, alors il n'y a pas de contrainte pour le coloriage des $m+1$ premiers points et donc nous avons F_{m+1} façons pour colorier ces points par l'hypothèse de récurrence. Si on colorie le dernier point en rouge, alors, on doit colorier le point précédant $(2^{m-2}i, 2^{m-2}j)$ en bleu et puis il n'y a pas de contrainte pour les m premiers points. D'où, nous avons F_m façons de colorier. Au total, nous avons donc $F_{m+1}+F_m$ façons pour colorier nos points. Ceci complète la démonstration.

3. Comptage sur $K_{n,m}$. Nous comptons le cardinal de l'ensemble $\{(i,j)\in\mathcal{B}_n,\ \ell(i,j)=m\}$. Pour ceci, remarquons que

$$\begin{aligned}
&\{(i,j)\in\mathbb{Z}^2\cap[-n,n]^2 : \ell(i,j)=m\}\\
=&\left\{(i,j) : \frac{n}{2^m}<|i|\le\frac{n}{2^{m-1}},\ |j|\le\frac{n}{2^m}\right\}\bigsqcup\left\{(i,j) : |i|\le\frac{n}{2^m},\ \frac{n}{2^m}<|j|\le\frac{n}{2^{m-1}}\right\}\\
&\bigsqcup\left\{(i,j) : \frac{n}{2^m}<|i|\le\frac{n}{2^{m-1}},\ \frac{n}{2^m}<|j|\le\frac{n}{2^{m-1}},\right\}.
\end{aligned}$$

Donc,

$$\begin{aligned}
&\sharp\{(i,j)\in\mathbb{Z}^2\cap[-n,n]^2 : \ell(i,j)=m\}\\
=&2\cdot\left(2\cdot(\frac{n}{2^{m-1}}-\frac{n}{2^m})+\epsilon_1\right)\cdot\left(2\cdot\frac{n}{2^m}+\epsilon_2\right)+\left(2\cdot(\frac{n}{2^{m-1}}-\frac{n}{2^m})+\epsilon_1\right)^2,
\end{aligned}$$

où ϵ_1,ϵ_2 sont certains nombres réels tels que $|\epsilon_1|,|\epsilon_2|<2$ (pour rendre les nombres dans la formule soient les entiers que nous souhaitons). On en déduit

$$\sharp\{(i,j)\in\mathbb{Z}^2\cap[-n,n]^2 : \ell(i,j)=m\}=\frac{12n^2}{2^{2m}}+\epsilon\frac{n}{2^m},\quad\text{avec } |\epsilon|<48.$$

De plus,

$$\begin{aligned}
&\{(i,j)\in\mathbb{Z}^2\cap[-n,n]^2\setminus\mathcal{B}_n,\ \ell(i,j)=m\}\\
=&\left\{(i,j) : 2|i, \frac{n}{2^m}<|i|\le\frac{n}{2^{m-1}},\ 2|j, |j|\le\frac{n}{2^m}\right\}\bigsqcup\left\{(i,j) : 2|i, |i|\le\frac{n}{2^m},\ 2|j, \frac{n}{2^m}<|j|\le\frac{n}{2^{m-1}}\right\}\\
&\bigsqcup\left\{(i,j) : 2|i, \frac{n}{2^m}<|i|\le\frac{n}{2^{m-1}},\ 2|j,\ \frac{n}{2^m}<|j|\le\frac{n}{2^{m-1}}\right\}.
\end{aligned}$$

En faisant des calculs similaires, nous avons

$$\sharp\{(i,j)\in\mathbb{Z}^2\cap[-n,n]^2\setminus\mathcal{B}_n, \ell(i,j)=m\}=\frac{3n^2}{2^{2m}}+\epsilon'\frac{n}{2^m},\quad\text{avec } |\epsilon'|<12.$$

Donc,

$$K_{n,m}=\frac{12n^2}{2^{2m}}+\epsilon\frac{n}{2^m}-\frac{3n^2}{2^{2m}}-\epsilon'\frac{n}{2^m}=\frac{9n^2}{2^{2m}}+\epsilon''\frac{n}{2^m},\quad\text{avec } |\epsilon''|<60.$$

4. Calculs de la limite. Nous avons

$$\frac{1}{n^2}\log X_n=\frac{1}{n^2}\log\left(2\cdot\prod_{m=1}^{\lfloor\log_2 n\rfloor+1}F_m^{K_m}\right)=\frac{1}{n^2}\left(\log 2+\sum_{m=1}^{\lfloor\log_2 n\rfloor+1}K_m\log F_m\right).$$

En remarquant que $F_m \leq C\beta^m$ avec $\beta = \frac{\sqrt{5}+1}{2}$ et C une constante positive, on déduit que la série

$$\sum_{m=1}^{\infty} K_{n,m} \mathbf{1}_{\{m \leq \lfloor \log_2 n \rfloor + 1\}} \log F_m$$

uniformément converge. Donc,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n^2} \log X_n &= \lim_{n\to\infty} \frac{1}{n^2} \sum_{m=1}^{\lfloor \log_2 n \rfloor + 1} K_{n,m} \log F_m = \sum_{m=1}^{\infty} \lim_{n\to\infty} \frac{K_{n,m} \mathbf{1}_{\{m \leq \lfloor \log_2 n \rfloor + 1\}}}{n^2} \log F_m \\ &= \sum_{m=1}^{\infty} \frac{9}{2^{2m}} \log F_m.\end{aligned}$$

Commentaires sur les solutions proposées : Parmi les solutions proposées, celle de l'équipe Mathadors est excellente. L'équipe a même trouvé une solution pour la dimension quelconque de la question. L'utilisation de la notation de valuation simplifie les arguments. L'idée réduisant les comptages au calcul du nombre de vecteurs dans une couronne ayant la valuation 2-adique supérieure à un entier fixé est très intelligente. En fait, l'équipe donne une formule plus simple de X_n :

$$X_n = \prod_{k=0}^{\infty} \left(\frac{F_{k+1}}{F_k} \right)^{\sharp\{x \in \Gamma_n :\ v_2(x) \geq k\}},$$

où $\Gamma_n = \mathbb{Z}^d \cap ([-n, n]^d \setminus [-n/2, n/2]^d)$, et $v_2(\cdot)$ est la valuation 2-adique. Alors, le nombre

$$\sharp\{x \in \Gamma_n :\ v_2(x) \geq k\}$$

est plus facile à compter.

Sujet de modélisation
par Filippo Santambrogio (Université Lyon 1)

Pour fêter son anniversaire, comme d'habitude en cette circonstance, un enfant apporte en classe une tarte au chocolat et une tarte à la praline de même taille.

En début d'année, dans cette situation, la maîtresse demandait à chaque enfant ce qu'il préfèrait. Par exemple, 20 élèves choisissaient le chocolat et 5 la praline. Puis la maîtresse annonçait que la tarte au chocolat serait donc divisée en 20 parts égales et la tarte à la praline seulement en 5. Mais un 1/20e de tarte ce n'est vraiment pas beaucoup, et la maîtresse donnait toujours à chaque enfant un peu de temps pour changer d'avis, éventuellement plusieurs fois, avant de procéder aux coupes.

Maintenant, la maitresse connaît bien ses élèves et elle aimerait procéder au partage des deux tartes de sorte qu'il ne soit plus nécessaire de donner du temps aux enfants pour changer d'avis. Il faudrait trouver d'emblée un partage que chacun accepterait.

Or, chaque enfant a ses propres préférences sous la forme d'un coefficient de plaisir à manger une unité de tarte au chocolat (on appelle $c_i > 0$ le coefficient de l'enfant i) et d'un coefficient de plaisir à manger une unité de tarte à la praline (on appelle $p_i > 0$ le coefficient de l'enfant i) : entre une part de chocolat de taille t_c et une part de praline de taille t_p, l'enfant i préfère le chocolat si $c_i t_c > p_i t_p$ et la praline si $c_i t_c < p_i t_p$.

Les enfants n'ayant pas la possibilité de se réunir et discuter ensemble de leurs choix, chacun déclare son choix (chocolat ou praline) et chacun décide de changer d'avis selon si cela lui permet ou pas d'augmenter sa satisfaction, en supposant que les choix des autres ne changent pas. On cherche une configuration de choix qui soit stable, c'est-à-dire dans laquelle aucun enfant ne souhaite changer d'avis. Ce que la maîtresse va faire est d'attribuer elle-même un choix (chocolat ou praline) à chaque enfant ; elle demandera ensuite à chacun indépendamment si, au vu de la repartition proposée, il préfère changer, et elle compte bien faire en sorte que personne ne change, parce qu'elle aura trouvé une configuration stable (qu'on appelle également un équilibre).

(a) Démontrer qu'une telle configuration stable existe toujours.

(b) Construire un exemple de configuration stable où une des tartes n'est choisie par aucun enfant.

(c) Construire un exemple de non-unicité de la configuration stable.

(d) Peut-on trouver, dans une configuration stable, un couple d'enfants qui auraient (chacun individuellement) intérêt à échanger leurs parts mais qui n'ont pas pu le faire à cause de l'interdiction de s'accorder ?

(e) On imagine pendant un instant que les enfants ne se connaissent pas encore très bien et qu'on n'arrive pas forcément tout de suite à une configuration stable. Supposons qu'au début chaque enfant choisisse sa tarte préférée (selon $p_i > c_i$ ou $p_i < c_i$, sans se soucier de la taille des parts) et qu'ensuite, à chaque fois que la maîtresse propose, en même temps à tous, de changer d'avis, il recalcule ce qu'il préfère en imaginant que les autres ne changeraient pas d'avis : cela donne lieu à une suite de configurations de

choix ; peut-on dire que ça converge forcément (et, si oui, en un nombre fini d'itérations ou asymptotiquement ?) vers une configuration stable ?

On suppose maintenant que le choix des enfants n'est pas binaire : au lieu de choisir entre chocolat et praline ils peuvent indiquer deux proportions σ_i et π_i, avec $\sigma_i, \pi \geq 0$ et $\sigma_i + \pi_i = 1$, qui représentent un choix mixte. L'enfant qui choisir (σ_i, π_i) recevra une part de tarte au chocolat de taille $\frac{\sigma_i}{\sum_j \sigma_j}$ et une de tarte à la praline de taille $\frac{\pi_i}{\sum_j \pi_j}$. On cherche à nouveau une configuration stable.

(f) Prouver que dans une configuration stable chaque tarte est bien mangée par au moins deux enfants différents.

(g) Démontrer qu'une configuration stable existe toujours.

(h) Prouver que dans une configuration stable on ne peut pas avoir un enfant qui ne mange que du chocolat, un autre qui ne mange que de la praline, mais les deux auraient préféré échanger leurs parts.

Discuter enfin le cas où le nombre d'enfants est infini, chacun ayant un poids négligeable [1]et les liens de ce cas avec les deux cas présentés ici (choix binaires ou mixtes) dans la limite où le nombre d'enfant de la classe tend vers l'infini.

1. Les jeux où chaque joueur est négligeable devant les autres sont appelés *jeux non-atomiques* et sont utilisés comme une approximation très raisonnable de beaucoup de situations réelles comme le problème de l'équilibre en trafic routier, où le temps de parcours d'une route dépend du nombre de véhicules sur la même route. Les jeux où l'utilité de chaque joueur dépend du nombre d'individus faisant le même choix que lui (en pénalisant le cas où ce nombre est trop grand) sont appelés *jeux de congestion* et les problèmes de trafic, tout comme le problème de partage de gâteaux qu'on vient d'analyser, en sont un exemple. Les équilibres dans les problèmes de trafic sont appelés *équilibres de Wardrop* et ont été étudiés dans des cas discrets - sur une réseau - ou continu - dans un domaine, toute courbe étant admise, en lien avec des EDP satisfaites par l'intensité de trafic réalisé par les choix des joueurs. Une théorie mathématique récente qui reprend beaucoup de ces aspects est celle des *jeux à champ moyen*. Dans tous ces jeux, les équilibres sont plus faciles à décrire et à trouver lorsqu'il existe une fonctionnelle de congestion globale à minimiser, typiquement convexe, ce qui est souvent le cas dans les jeux de congestion continus. Les jeux admettant une telle fonctionnelle s'appellent *jeux potentiels* parce que la fontion qu'on cherche à minimiser joue le rôle d'un potentiel comme lorsqu'on dit qu'une force est potentielle quand elle a une forme gradient. Dans le cas atomique il est souvent plus difficile de trouver un potentiel.

Eléments de solution

Dans le cas binaire, où les enfants doivent choisir entre chocolat et praline, on considère une configuration et on se pose la question si elle est un équilibre. On définit

$$\mathcal{P} := \{i : \text{ l'enfant } i \text{ choisit praline}\},\ \mathcal{C} := \{i : \text{ l'enfant } i \text{ choisit chocolat}\}$$

ainsi que

$$P := \#\mathcal{P},\ C := \#\mathcal{C}.$$

Alors, cette configuration est d'équilibre si pour tout $i \in \mathcal{P}$ on a $\frac{p_i}{P} \geq \frac{c_i}{C+1}$ (le plaisir à prendre une part de praline quand on est P à en prendre une n'est pas dépassé par le plaisir à prendre une part de chocolat si le même enfant se rajoute aux C qui l'ont déjà choisi) et pour tout $i \in \mathcal{C}$ on a $\frac{p_i}{P+1} \leq \frac{c_i}{C}$.

Pour répondre à la question **(a)** (existence d'un équilibre), on construira une configuration comme suit : étant donné $t \in \mathbb{R}$, on attribue la praline à tous les enfants tels que $\frac{p_i}{c_i} > t$, le chocolat à tous ceux avec $\frac{p_i}{c_i} < t$, et on partage selon une proportion à choisir les éventuels enfants tels que $\frac{p_i}{c_i} = t$ entre les deux options. Pour que cela soit un équilibre, il suffit que P, C et t satisfassent $t \in [\frac{P}{C+1}, \frac{P+1}{C}]$ (si le dénominateur C s'annulle, on considère que cet intervalle est une demi-droite). Si on indique $P(t) = \#\{i : \frac{p_i}{c_i} > t\}$, $C(t) = \#\{i : \frac{p_i}{c_i} < t\}$, et $E(t) = \#\{i : \frac{p_i}{c_i} = t\}$, on peut choisir P et C de la manière suivante : on prend un entier $j \in [P(t), P(t) + E(t)]$ et on définit $P = j$ et $C = N - j$, où N est le nombre total d'enfants (on a donc $P(t) + E(t) = N - C(t)$). La condition pour que t donne lieu à un équilibre est donc $t \in \Gamma(t)$ où

$$\Gamma(t) = \bigcup_{j=P(t),\dots,P(t)+E(t)} \left[\frac{j}{N+1-j}, \frac{j+1}{N-j}\right] = \left[\frac{P(t)}{N+1-P(t)}, \frac{N+1-C(t)}{C(t)}\right].$$

Les fonctions $t \mapsto P(t)$ et $t \mapsto C(t)$ sont semi-continues inférieurement, ce qui implique que $t \mapsto \frac{P(t)}{N+1-P(t)}$ est auss s.c.i. alors que $t \mapsto \frac{N+1-C(t)}{C(t)}$ est s.c.s. Comme fonctions à valeurs ensembles (mutlifonction), Γ est donc semicontinue supérieurement (parce que $\Gamma(t)$ contient les limites de $\Gamma(t_n)$ lorsque $t_n \to t$). De plus, elle est à valeurs convexes. On peut fixer un T qui majore tous les ratios $\frac{p_i}{c_i}$ et considérer $[0, T] \ni t \mapsto \Gamma(t) \cap [0, T]$. Cette multifonction à valeurs convexes compacts non-vides satisfait les hypothèses du théorème de point fixe de Kakutani, et il existe donc t tel que $t \in \Gamma(t)$, et donc un équilibre.

On remarque que si la solution proposée ci-dessus est assez logique puisqu'elle ne fait que détailler la condition pour que t donne lieu à un équilibre, remarquer qu'elle correspond à une condition de point fixe multivalué, et en prouver l'existence par un théorème général, elle n'est sans doute pas la plus intuitive, et aucune des solutions proposées par les participants ne se basait à ce stade sur le théorème de Kakutani. Une construction plus explicite peut en effet être utilisée. On classe les enfants selon leur ratio $\frac{p_i}{c_i}$, en commençant par celui qui a plus de goût pour la praline ($\frac{p_i}{c_i}$ maximal) et en descendant. On attribue d'abord le chocolat à tous les enfants et on demande au premier s'il est satisfait ou s'il préfère changer. S'il est satisfait, a fortiori tous les autres, qui aiment un peu plus le chocolat que lui, le sont, et on a donc un équilibre. Sinon on change la tarte attribuée à cet enfant en lui attribuant la praline, et on pose la même question au deuxième, et ainsi de suite. On s'arrête quand on tombe sur un enfant qui ne souhaite pas changer, et on peut facilement prouver que cela donne un équilibre.

Plus précisément, cette construction donne un équilibre pour tout jeu où la satisfaction des enfants choisissant la praline serait de la forme $p_i f(P,C)$ et celle des enfants choisissant le chocolat serait $c_i g(P,C)$, sans forcément prendre $f(P,C) = 1/P$ et $g(P,C) = 1/C$ (cela n'a pas été observé par les participants, mais par un collègue).

(b) Pour construire un exemple où la tarte au chocolat n'est choisie par aucun enfant, il suffit de prendre des p_i, c_i tels que $\frac{p_i}{c_i} > N$ pour tout i : de cette manière chaque enfant prefère partager la tarte à la praline avec tout le monde plutôt qu'avoir la tarte au chocolat à lui seul.

Considérons, ensuite, le cas $N = 2$ avec $\frac{p_i}{c_i} \in]\frac{1}{2}, 2[$ pour tout i : de cette manière chaque enfant préfère toujours une tarte entière à la moitié de l'autre. Cela signifie qu'on peut leur attribuer les tartes comme on veut, une chacun, et ils ne changeront pas d'avis. Cela donne en même temps un exemple de non-unicité mais également, si $\frac{1}{2} < \frac{p_1}{c_1} < 1 < \frac{p_2}{c_2} < 2$, un exemple où l'on peut attribuer la tarte à la praline à l'enfant qui préfère le chocolat et viceversa. Cela répond à **(c)** et **(d)**.

Pour la question **(e)** sur l'évolution des choix des enfants on peut considérer l'exemple suivant :

$$N = 4,\ \frac{p_1}{c_1} = 5,\ \frac{p_2}{c_2} = 1.2,\ \frac{p_3}{c_3} = 1.1,\ \frac{p_4}{c_4} = 0.2.$$

De cette manière, l'enfant $i = 1$ choisit toujours la praline indépendamment de ce qui font les autres, et l'enfant $i = 4$ toujours le chocolat. Au début les deux autres choisissent la praline aussi, mais se retrouvent avec des parts de taille $1/3$ et ils préfèrent donc, chacun, une moitié de la tarte au chocolat. Ils changent donc d'avis mais, comme ils sont deux à le faire, ils n'obtiennent pas la moitié de cette tarte mais seulement un tiers, ce qui leur convient moins bien de la moitié de la tarte à la praline. Ils changent à nouveau d'avis mais...
On ne peut donc pas affirmer que cette évolution converge toujours vers une configuration stable.

Nous passons maintenant au cas des choix mixtes. La réponse à la question **(f)** est assez simple. On commence par prouver que chaque tarte est mangée par au moins un enfant. Si jamais personne ne mange, par exemple, du chocolat, donc $\sigma_i = 0, \pi_i = 1$ pour tout i, n'importe quel enfant aurait intérêt à changer en $\sigma_i = \varepsilon, \pi_i = 1 - \varepsilon$. De cette manière, il pourrait s'attendre à manger la tarte au chocolat en entier, en gagnant donc c_i dans son plaisir, et il ne perdrait qu'une quantité $O(\varepsilon)$ en ce qui concerne son plaisir à la praline. Montrons ensuite que chaque tarte est mangée par au moins deux enfants. Si jamais un seul enfant mangeait, par exemple, du chocolat, cet enfant aurait $\sigma_i > 0$ alors que pour $i \neq j$ on aurait $\sigma_j = 0$. Or, l'enfant mangeant la tarte au chocolat pourrait librement réduire son σ_i sans s'attendre à perdre du chocolat, mais cela lui permettrait d'augmenter π_i et donc sa part de praline. Cela est absurde.

Pour répondre aux autres questions il est nécessaire d'écrire précisément la condition pour qu'une configuration $(\pi_i, \sigma_i)_i$ soit stable. On définit $P = \sum_i \pi_i$, $C = \sum_i \sigma_i$, et, selon une notation typique de la théorie des jeux, $P_{-i} = P - \pi_i = \sum_{j \neq i} \pi_j$ et $C_{-i} = C - \sigma_i = \sum_{j \neq i} \sigma_j$. Chaque enfant, si les choix des autres sont connus, résout le problème d'optimisation suivant

$$\max_{\sigma,\pi \in [0,1] \,:\, \sigma+\pi=1} p_i \frac{\pi}{P_{-i} + \pi} + c_i \frac{\sigma}{C_{-i} + \sigma}.$$

On remarque que ce problème d'optimisation n'est pas bien posé si P_{-i} ou C_{-i} sont nuls, parce que dans ce cas il faudrait définir la valeur de cette fonction pour $\pi = 0$ ou $\sigma = 0$, respectivement. Or, il serait raisonnable de dire que choisir $\pi = 0$ apporte une utilité nulle en ce qui concerne la praline (et, de même, $\sigma = 0$ en ce qui concerne le chocolat) mais si on définit de cette manière la fonction à optimiser elle serait discontinue et il n'y aurait pas de minimiseur. On suppose pour l'instant $P_{-i}, C_{-i} > 0$.

Comme cela peut être réécrit (en retirant des constantes et en changeant de signe) sous la forme

$$\min_{\sigma,\pi\in[0,1]\,:\,\sigma+\pi=1} p_i \frac{P_{-i}}{P_{-i}+\pi} + c_i \frac{C_{-i}}{C_{-i}+\sigma}$$

on voit qu'il s'agit de la minimisation d'une fonction strictement convexe, qui admet donc un unique minimiseur.

En dérivant (et en utilisant $\sigma = 1 - \pi$) on trouve la solution du problème d'optimisation ci-dessous, qui est donnée par un point (σ_i^*, π_i^*) caractérisé par

$$p_i \frac{P_{-i}}{(P_{-i}+\pi_i^*)^2} = c_i \frac{C_{-i}}{(C_{-i}+\sigma_i^*)^2} \qquad \text{si } \pi_i^*, \sigma_i^* > 0, \tag{5}$$

$$p_i \frac{P_{-i}}{(P_{-i}+\pi_i^*)^2} \leq c_i \frac{C_{-i}}{(C_{-i}+\sigma_i^*)^2} \qquad \text{si } \pi_i^* = 0, \tag{6}$$

$$p_i \frac{P_{-i}}{(P_{-i}+\pi_i^*)^2} \geq c_i \frac{C_{-i}}{(C_{-i}+\sigma_i^*)^2} \qquad \text{si } \sigma_i^* = 0. \tag{7}$$

En pratique, on peut trouver σ_i^* (et donc π_i^* en utilisant $\pi_i^* = 1 - \sigma_i^*$) en prenant le point de minimum de la fonction $\sigma \mapsto p_i \frac{P_{-i}}{P_{-i}+1-\sigma} + c_i \frac{C_{-i}}{C_{-i}+\sigma}$ sur l'intervalle ouvert $]-C_{-i}, P_{-i}+1[$, qui est caractérisé donc par (5) et en prenant sa partie positive s'il est négatif, ou en tronquant à 1 s'il dépasse 1. De ce fait, on voit que l'optimiseur σ_i^* dépend de manière continue de P_{-i} (le paramètre C_{-i} étant une fonction continue de P_{-i}). Tout cela marche sous l'hypothèse $P_{-i}, C_{-i} > 0$.

Si $C_{-i} = 0$ on peut définir $\sigma_i^* = 0$ et, de même, si $P_{-i} = 0$ on prend $\sigma_i^* = 1$ (ce qui équivaut à $\pi_i^* = 0$). Cette définition préserve la continuité de l'application qui a $\sigma = (\sigma_1, \ldots, \sigma_N)$ associe $\sigma^* = (\sigma_1^*, \ldots, \sigma_N^*)$. Pour le voir, il suffit de prendre une suite $\sigma(n)$ telle que $C_{-i}(\sigma(n)) \to 0$ et prouver $\sigma_i^*(n) \to 0$. Or, si on avait $\sigma_i^*(n) \geq \varepsilon > 0$, on pourrait appliquer les relations (5) et (6) et voir que le membre de droite tend vers 0, alors que le membre de gauche est borné inférieurement par $p_i P_{-i}/(P_{-i}+1)^2$ qui ne tend pas vers 0 parce que $P_{-i} + C_{-i} = N - 1$. La preuve pour le cas $P_{-i} = 0$ est analogue.

On a donc définit une application $Z = [0,1]^N \to [0,1]^N$ qui est continue et associe à chaque σ un vecteur σ^* correspondant à la meilleur réponse aux données $(\sigma_1, \ldots, \sigma_N)$. On en cherche un point fixe tel que pour tout i on ait $P_{-i}, C_{-i} > 0$. La fonction Z, en tant que fonctinue d'un compacte convexe dans lui-même, admet surement au moins un point fixe, mais comme $\sigma = (0, \ldots, 0)$ et $\sigma = (1, \ldots, 1)$ sont des points fixes non-acceptables il faut en produire d'autres.

Pour ce faire, on fixe $\varepsilon > 0$ et on considère $K := \{\sigma \in [0,1]^N : \sum_i \sigma_i \in [\varepsilon, N - \varepsilon]\}$. Cet ensemble est aussi un convexe, mais on ne sait pas si $Z(K) \subset K$. On considère donc $\tilde{Z} = \Pi_K \circ Z$, où Π_K est la projection sur le convexe fermé K, qui est continue. Il existe donc

un point fixe σ de $\tilde{Z}$ et on veut démonter que, pour $\varepsilon > 0$ bien choisi, ce point est aussi un point fixe de Z et qu'il satisfait $P_{-i}, C_{-i} > 0$ pour tout i.

Supposons maintenant que le point fixe σ est tel qu'une valeur C_{-i} est nulle (le cas où $P_{-i} = 0$ se traite de la même manière). On a donc $\sigma_j = 0$ pour tout $j \neq i$ et donc $\sigma_i \geq \varepsilon$. Or, un point de ce type (avec une seule composante non nulle) n'est la projection sur K que de lui-même (il n'y a pas d'autres points de $[0,1]^N$ qui se projetterait sur ce point). Donc σ serait un point fixe de Z mais $C_{-i} = 0$ impliquerait alors $\sigma_i^* = 0$, ce qui contredit $\sigma_i \geq \varepsilon$ et $\sigma = \sigma^*$.

On sait maintenant que le point fixe σ de $\tilde{Z}$ est tel que $P_{-i}, C_{-i} > 0$ pour tout i, ce qui nous permet d'utiliser les relations (5),(6) et (7). Si on suppose $\sigma \neq Z(\sigma)$, la condition $\sigma = \Pi_K(Z(\sigma))$ implique $\sigma \in \partial K$ et $\sigma^* = Z(\sigma) \notin K$. On peut supposer $\sum_i \sigma_i^* < \varepsilon$ (le cas où la somme serait plus grande que $N - \varepsilon$ se traite de la même manière) et $\sum_i \sigma_i = \varepsilon$ (parce que la projection envoie les points avec somme des composantes inférieure à ε sur les points où cette somme vaut ε, et pas $N - \varepsilon$). En particulier on a $\sigma_i^* < \varepsilon$ pour tout i. Aussi, on peut choisir i tel que $C_{-i} = O(\varepsilon)$ (plus précisément, on a toujours $C_{-i} \leq \varepsilon$ mais on a $C_{-i} \geq \frac{N-1}{N}\varepsilon$ pour au moins un index i) et $P_{-i} \geq 1 - \varepsilon$. On applique (5) ou (7) et on a, dans tous les cas

$$\frac{p_i}{P_{-i}} \geq p_i \frac{P_{-i}}{(P_{-i} + \pi_i^*)^2} \geq c_i \frac{C_{-i}}{(C_{-i} + \sigma_i^*)^2}.$$

Le membre de droite est $O(\varepsilon^{-1})$ alors que celui de gauche est borné, ce qui montre que cela est impossible pour ε petit. Par conséquent, si ε est suffisamment petit le point fixe de $\tilde{Z}$ dans K est aussi un point fixe de Z, et est finalement un équilibre. Cela répond à la question **(g)**.

Même si une des solutions proposées passait en effet par une approximation avec un paramètre $\varepsilon \to 0$ la plupart des solutions correctes proposées par les participants se basent sur une idée différente. Cette idée peut se résumer comme suit : pour toute valeur de P_{-i} (en posant donc $C_{-i} := N - 1 - P_{-i}$) il existe unique un choix optimal π_i pour l'enfant i et on peut l'indiquer comme $\pi_i^*(P_{-i})$: on se donne une valeur de $P > 0$ et, pour tout i, on cherche une valeur $\pi_i \in (0, P)$ telle que $\pi_i^*(P - \pi_i) = \pi_i$; on prouve que cette valeur existe et est unique et on l'appelle $\pi^{**}(P)$; on cherche ensuite une valeur de P telle que $P = \sum_i \pi^{**}(P)$ et $P \in (0, N)$; pour ce faire, on prouve que la fonction $P \mapsto \sum_i \pi^{**}(P)$ est au dessus de l'identié pour $P > 0$ proche de 0 et en dessous de l'identité pour $P < N$ proche de N, ce qui implique qu'il y a un P point fixe. Dès qu'une telle valeur de P est trouvée, chaque enfant choisira $\pi_i := \pi_i^{**}(P), \sigma_i := 1 - \pi_i$. Ce sera un équilibre parce que, avec ce choix, la condition $\pi_i^*(P - \pi_i) = \pi_i$ indique exactement que ce choix de π_i est optimal si $P_{-i} = P - \pi_i$.

En ce qui concerne la question **(h)**, on suppose d'avoir un point fixe σ pour l'application Z avec $\sigma_i = 1$ et $\pi_j = 1$, pour un couple (i, j), $i \neq j$. On a donc, en utilisant (6) pour i et (7) pour j,

$$\frac{p_i}{P_{-i}} \leq c_i \frac{C_{-i}}{(C_{-i} + 1)^2} \quad \text{et} \quad p_j \frac{P_{-j}}{(P_{-j} + 1)^2} \geq \frac{c_j}{C_{-j}}. \tag{8}$$

Or, la condition "les deux enfants préfèrent échanger leurs parts" signifie

$$\frac{c_i}{C} < \frac{p_i}{P} \quad \text{et} \quad \frac{c_j}{C} \geq \frac{p_j}{P}. \tag{9}$$

On a aussi $C = C_{-i} + \sigma_i = C_{-i} + 1$ ainsi que $P = P_{-j} + \pi_j = P_{-j} + 1$. On a donc, en remplaçant cela dans (8)

$$\frac{p_i}{P} \leq \frac{p_i}{P_{-i}} \leq c_i \frac{C_{-i}}{C^2} \leq \frac{c_i}{C} \quad \text{et} \quad \frac{p_j}{P} \geq p_j \frac{P_{-j}}{P^2} \geq \frac{c_j}{C_{-j}} \geq \frac{c_j}{C}.$$

Ces conditions sont en contradiction avec (9).

Enfin, plusieurs considérations sont possibles dans le cas où on considère des enfants négligeables. La différence principale consiste en le fait que, lorsqu'ils changent d'avis, ils n'influencent pas la valeur de P ou C, ce qui rend la condition d'équilibre plus simple, puisque (5) est remplacée par $\frac{p_i}{P} = \frac{c_i}{C}$. On peut définir formellement le problème comme la recherche d'une mesure sur un ensemble $R \times \{\text{praline,chocolat}\}$ où R est l'ensemble des types d'enfant (caractérisés par leur ratio p_i/c_i) et la marginale sur R est prescrite. Ce type de jeux, où les agents ont des types différents et leur utilité est composée de la somme d'une fonction du type et du choix et d'un terme de congestion dépendant seulement du choix et de l'autre marginale de la mesure sont connus sous le nom d'équilibres de Cournot-Nash et admettent même un potentiel. On pourrait obtenir quelque chose de similaire en considérant M copies des N enfants de départ et en faisant tendre M vers l'infini. cela ressemblerait beaucoup au cas avec choix mixtes, à une différence importante près : ce n'est pas la même chose que d'avoir un joueur qui contribue au calcul du nombre de part où M joueurs indépendants (chacun optimisant son propre intérêt et ne se concertant pas avec les autres du même type) contribuant chacun pour un M-ème.

Une considération intéressante sur le cas des enfants négligeables (mais qui n'a pas été reperée par les participants) est que on peut trouver l'équilibre de manière variationnelle. Une mesure sur $R \times \{\text{praline,chocolat}\}$ avec marginale sur R prescrite et égale à ρ (où l'on fixe par simplicité $\int_R d\rho = 1$) peut s'identifier à une fonction $\pi : R \to [0,1]$ et on cherche une telle fonction π de manière à avoir, en considérant les fonctions $p, c : R \to \mathbb{R}_+$ de préférences des enfants

$$\begin{aligned}
\frac{p}{\int_R \pi d\rho} &= \frac{c}{1 - \int_R \pi d\rho} \quad \text{p.p. sur } \{\pi \in (0,1)\}, \\
\frac{p}{\int_R \pi d\rho} &\geq \frac{c}{1 - \int_R \pi d\rho} \quad \text{p.p. sur } \{\pi = 1\}, \\
\frac{p}{\int_R \pi d\rho} &\leq \frac{c}{1 - \int_R \pi d\rho} \quad \text{p.p. sur } \{\pi = 0\}.
\end{aligned}$$

Une telle fonction π peut être trouvée comme le minimiseur de la fonction convexe

$$\pi \mapsto -\int_R [\pi \log p + (1-\pi)\log c] d\rho + (\int_R \pi d\rho) \log(\int_R \pi d\rho) + (\int_R (1-\pi) d\rho) \log(\int_R (1-\pi) d\rho).$$

Sujet de probabilité
par Thomas Budzinski (Ecole normale supérieure de Lyon)

Le *processus de Bienaymé–Galton–Watson* est un des processus classiques utilisés pour modéliser l'évolution d'une population, qui se reproduit de manière asexuée, homogène et stationnaire dans le temps. Il admet une *transition de phase* remarquable : si m désigne le nombre moyen d'enfants d'un individu, alors la population s'éteint presque sûrement si $m \leq 1$ (en excluant le cas particulier où un individu a toujours un seul enfant), et survit avec probabilité strictement positive si $m > 1$.

Il s'agit ici d'étudier une "tropicale" de ce processus, où la somme dans la définition du processus de Bienaymé–Galton–Watson est remplacée par un max, et de se demander comment ceci influence la transition de phase.

Plus précisément, on fixe une mesure de probabilité μ sur l'ensemble $\mathbb{N}$ des entiers naturels, qui vérifie $\mu(0) > 0$. Soient $(Z_{n,i})_{n\geq 1, i\geq 1}$ des variables aléatoires indépendantes et identiquement distribuées de loi μ. On définit récursivement la suite de variables aléatoires $(X_n)_{n\geq 0}$ par $X_0 = 1$ et, pour $n \geq 1$:

$$X_n = \max\left\{Z_{n,i} | 1 \leq i \leq X_{n-1}\right\},$$

avec la convention $\max \emptyset = 0$. On se demande ainsi, en fonction de μ, si la probabilité de survie $\mathbb{P}(\forall n, X_n > 0)$ vaut 0 ou non.

(a) Montrer que si μ est à support fini, alors $\mathbb{P}(\forall n, X_n > 0) = 0$.

(b) Montrer que si $\sum_{i=1}^{\infty} i\mu(i) < 1$, alors $\mathbb{P}(\forall n, X_n > 0) = 0$.

(c) Montrer que si $\sum_{i=1}^{\infty} i\mu(i) < +\infty$, alors $\mathbb{P}(\forall n, X_n > 0) = 0$.

(d) Montrer que si $\mu(i) \sim \frac{c}{i^\alpha}$ quand $i \to +\infty$ avec $c > 0$ et $1 < \alpha < 2$, alors $\mathbb{P}(\forall n, X_n > 0) > 0$.

(e) Étudier le cas $\mu(i) \sim \frac{c}{i^2}$ quand $i \to +\infty$ avec $c > 0$.

Eléments de solution

(a) La manière la plus rapide de répondre à cette question est de faire appel à la classification des états d'une chaîne de Markov. Comme X_n s'écrit en fonction de X_{n-1} et de variables aléatoires indépendantes de tout se qui précède, (X_n) est une chaîne de Markov sur l'espace d'états $\mathrm{supp}(\mu)$, et on peut noter Q sur matrice de transition. L'espace d'état est fini donc la chaîne admet au moins un point récurrent x. Alors 0 est accessible depuis x (car $Q(x,0) = \mu(0)^x > 0$), donc 0 est aussi récurrent, donc X_n atteint 0 presque sûrement.

Si on n'est pas familier avec la théorie des chaînes de Markov, on peut résoudre la question de manière plus calculatoire : notons m le maximum du support de μ. Pour tout $n \geq 1$, si $X_0, \ldots, X_{n-1} > 0$, alors

$$\mathbb{P}\left(X_n = 0 | X_0, \ldots, X_{n-1}\right) = \mu(0)^{X_{n-1}} \geq \mu(0)^m.$$

On en déduit, par récurrence sur n :

$$\mathbb{P}\left(X_0, \ldots, X_n > 0\right) \leq (1 - \mu(0)^m)^n \xrightarrow[n\to+\infty]{} 0,$$

donc $\mathbb{P}\left(\forall n \geq 0, X_n > 0\right) = 0$.

(b) Dans ce cas, le plus rapide est de "dominer" (X_n) par un processus de Galton–Watson classique. En utilisant les mêmes variables $Z_{n,i}$, on définit les variables Y_n par récurrence par $Y_0 = 1$ et, pour $n \geq 1$:

$$Y_n = \sum_{i=1}^{Y_{n-1}} Z_{n,i}.$$

Alors Y_n est un processus de Galton–Watson de loi de reproduction μ. Comme μ est d'espérance au plus 1 on sait que Y s'éteint presque sûrement, i.e. $Y_n = 0$ pour n assez grand. D'autre part, on vérifie par récurrence sur n que $X_n \leq Y_n$ pour tout n, d'où le résultat.

Si on n'est pas familier avec les processus de Galton–Watson, on peut aussi résoudre le problème en considérant l'espérance de X_n. Ici aussi, l'idée est de borner le max des $Z_{n,i}$ par leur somme :

$$\mathbb{E}\left[X_n | X_0, \ldots, X_{n-1}\right] \leq \sum_{i=1}^{X_{n-1}} \mathbb{E}[Z_{n,i}] = X_{n-1} \times \sum_i i\mu(i).$$

En prenant l'espérance des deux côtés, on obtient $\mathbb{E}[X_n] \leq \mathbb{E}[X_{n-1}] \times \sum_i i\mu(i)$, d'où par récurrence

$$\mathbb{E}[X_n] \leq \left(\sum_i i\mu(i)\right)^n \xrightarrow[n\to+\infty]{} 0.$$

En utilisant l'inégalité de Markov, on a $\mathbb{P}\left(X_n \geq 1\right) \leq \mathbb{E}[X_n] \to 0$, donc $\mathbb{P}\left(X_n > 0\right) \to 0$, ce qui conclut.

(c) On cherche à réutiliser l'idée de la question précédente, mais cette fois borner le max par la somme devient trop brutal. La raison est que de petites valeurs présentes un grand nombre de fois ne contribue pas au max, mais un grand nombre de fois à la

somme. L'idée va donc être de ne compter qu'une seule fois les valeurs plus petites que a, pour une grande constante a.

Plus précisément, pour $x \in \mathbb{Z}$, on note $x^+ = \max(x, 0)$. On fixe une constante $a > 0$ assez grande pour que $\sum_{i>a} i\mu(i) < 1$. On va s'intéresser à $\mathbb{E}\left[(X_n - a)^+\right]$. Pour tout $x \geq 1$, on a

$$\mathbb{P}\left(X_n \geq x + a | X_{n-1}\right) = \mathbb{P}\left(\bigcup_{i=1}^{X_{n-1}} \{Z_{n,i} \geq x + a\}\right) \leq X_{n-1} \times \mathbb{P}\left(Z \geq x + a\right),$$

où Z est une variable de loi μ. En prenant l'espérance, on obtient

$$\mathbb{P}\left(X_n \geq x + a\right) \leq \mathbb{E}[X_{n-1}]\mathbb{P}\left(Z \geq x + a\right).$$

Or, pour toute variable entière X, on a

$$\mathbb{E}\left[(X - a)^+\right] = \sum_{x \geq 1} \mathbb{P}\left((X - a)^+ \geq x\right) = \sum_{x \geq 1} \mathbb{P}\left(X \geq x + a\right).$$

En sommant l'équation précédente sur $x \geq 1$, on obtient donc

$$\mathbb{E}\left[(X_n - a)^+\right] \leq \mathbb{E}[X_{n-1}]\mathbb{E}[(Z - a)^+],$$

donc

$$\mathbb{E}\left[(X_n - a)^+\right] \leq \left(a + \mathbb{E}\left[(X_{n-1} - a)^+\right]\right) \times \mathbb{E}[(Z - a)^+],$$

avec $\mathbb{E}[(Z-a)^+] = \sum_{i>a}(i-a)\mu(i) < 1$ par choix de a. On a donc borné $\mathbb{E}\left[(X_n - a)^+\right]$ par une suite vérifiant une relation de récurrence arithmético-géométrique de raison strictement inférieur à 1, ce qui montre que $\mathbb{E}\left[(X_n - a)^+\right]$ est bornée.

Plus précisément, si u_n vérifie $u_0 = 1$ et $u_n = (a + u_{n-1}) \times \mathbb{E}[(Z - a)^+]$ pour $n \geq 1$, alors on a $\mathbb{E}\left[(X_n - a)^+\right] \leq u_n$ par récurrence sur n. D'autre part, on peut calculer explicitement u_n et obtenir

$$u_n \xrightarrow[n \to +\infty]{} \frac{a\mathbb{E}[(Z - a)^+]}{1 - \mathbb{E}[(Z - a)^+]},$$

donc $\mathbb{E}[(Z_n - a)^+]$ est bornée.

On en conclut que $\mathbb{P}\left(X_n \to +\infty\right) = 0$, donc la chaîne de Markov (X_n) a au moins un état récurrent x (i.e. visité une infinité de fois). De même que dans la question 1, l'état 0 est accessible depuis x, donc X_n atteint presque sûrement 0.

(d) On pose $f(x) = \mathbb{P}\left(X_n \leq X_{n-1} | X_{n-1} = x\right)$ (notons que cette quantité ne dépend que de x et pas de n car (X_n) est une chaîne de Markov). On va montrer que $f(x)$ tend vers 0 très vite quand $x \to +\infty$.

En effet, on a

$$f(x) = \mathbb{P}\left(Z_{n,1}, \ldots, Z_{n,x} \leq x\right) = \mu\left([0, x]\right)^x.$$

On sait que $\mu\left([0, x]\right) = 1 - \frac{c}{\alpha - 1}\frac{1}{x^{\alpha-1}} + o\left(\frac{1}{x^{\alpha-1}}\right)$ quand $x \to +\infty$, donc

$$f(x) = \exp\left(-\frac{c}{\alpha - 1}x^{2-\alpha} + o\left(x^{2-\alpha}\right)\right).$$

En particulier $\sum_{x\geq 1} f(x) < +\infty$.
L'idée est maintenant la suivante : avec probabilité strictement positive X_1 prend une grande valeur x. Une fois que c'est le cas, avec très grande probabilité, on a $X_2 > X_1$, puis X_2 est aussi grand donc avec très grande probabilité $X_3 > X_2$ et ainsi de suite. On veut majorer la probabilité que $X_1 = x$ soit grand mais que (X_n) ne soit pas strictement croissante. Pour cela, on décompose selon le premier pas où la suite ne croît pas :

$$\begin{aligned}&\mathbb{P}\left(X_1 = x \text{ mais } (X_n) \text{ n'est pas croissante}\right)\\ &= \sum_{n\geq 2} \mathbb{P}\left(X_1 = x \text{ et } X_1 < X_2 < \cdots < X_{n-1} \text{ mais } X_n \leq X_{n-1}\right).\end{aligned}$$

Conditionnellement à $X_1, \ldots, X_{n-1}$, la probabilité que $X_n \leq X_{n-1}$ vaut $f(X_{n-1})$. De plus, si $X_1 = x$ et $X_1 < X_2 < \cdots < X_{n-1}$, alors $X_{n-1} \geq x + n - 2$. Par conséquent, on a

$$\mathbb{P}\left(X_1 = x \text{ et } X_1 < X_2 < \cdots < X_{n-1} \text{ mais } X_n \leq X_{n-1}\right) \leq \mu(x) \times \max_{y\geq x+n-2} f(y),$$

donc en sommant

$$\mathbb{P}\left(X_1 = x \text{ mais } (X_n) \text{ n'est pas croissante}\right) \leq \mu(x) \sum_{n'\geq 0} \max_{y\geq x+n'} f(y).$$

D'après notre borne sur $f(x)$, la somme tend vers 0 quand $x \to +\infty$. En particulier il existe x tel que cette somme est strictement inférieure à 1. Dans ce cas, on a

$$\mathbb{P}\left(x = X_1 < X_2 < \ldots\right) > 0,$$

et en particulier $\mathbb{P}\left(\forall n \geq 0, X_n > 0\right)$.
Remarque Pour la dernière question, une analyse plus précise montre que X_{n+1} est de l'ordre de $X_n^{1/(\alpha-1)}$. La croissance de X_n est donc en fait très rapide.

(e) Dans ce cas, pour X_n grand $\frac{X_{n+1}}{X_n}$ est proche d'une variable R dont la loi a pour densité

$$\frac{\mathrm{d}}{\mathrm{d}a} e^{-c/a} = \frac{c}{a^2} e^{-c/a}$$

par rapport à la mesure de Lebesgue sur $\mathbb{R}^+$. Il faut alors étudier le signe de $\mathbb{E}[\log R]$. En utilisant la fonction exponentielle intégrale, on calcule $\mathbb{E}[\log R] = \log(c) + \gamma$. On trouve donc extinction p.s. si $c < e^{-\gamma}$ et survie si $c > e^{-\gamma}$, où $\gamma = \lim_{k\to+\infty} \log k - \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)$.
Seule l'équipe Mathadors a résolu complètement cette question. Leur solution utilise des arguments de couplage très élégants pour rendre rigoureuse l'idée décrite ci-dessous. Cette équipe a même obtenu des résultats plus précis dans la fenêtre critique : si $\mu(i)$ est de la forme $\frac{e^{-\gamma}}{i^2} + \frac{b}{i^2 \log i} + o\left(\frac{1}{i^2 \log i}\right)$, alors il y a extinction presque sûre si $b < b_c$ et survie avec probabilité strictement positive si $b > b_c$, où

$$b_c = e^{\gamma}\left(\gamma^2 + \frac{\pi^2}{12} + 1\right).$$

Sujet de topologie
par Ramanujan Santharoubane (Université Paris-Saclay)

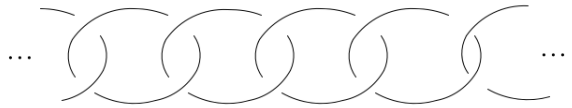
On considère dans $\mathbb{R}^3$ une droite D et $C = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 = 1\,,\ z = 0\}$ un cercle. On note X le complémentaire de D dans $\mathbb{R}^3$ et Y le complémentaire de C dans $\mathbb{R}^3$.

(a) Calculer les groupes fondamentaux de X et Y.

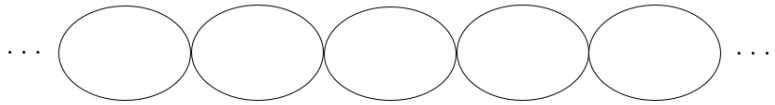
(b) Trouver une suite croissante d'ouverts d'adhérences compactes dont l'union recouvre X. Même question pour Y.

(c) Les espaces topologiques X et Y sont-ils homéomorphes ?

(d) On considère L_1 l'entrelac infini suivant



et L_2 le recollement infini de cercles suivant

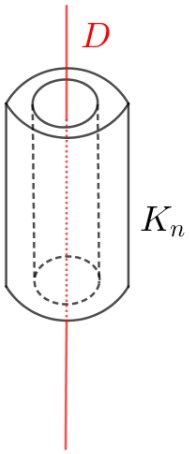


A-t-on $\mathbb{R}^3 - L_1$ homéomorphe à $\mathbb{R}^3 - L_2$?

Eléments de solution

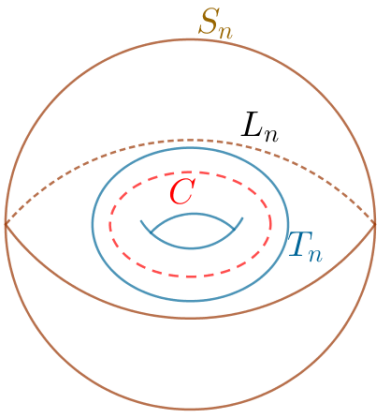
Question 1 : $\mathbb{Z}$ est le groupe fondamental de X et de Y, on peut utiliser par exemple le théorème de Van Kampen pour le démontrer.

Question 2 : En faisant varier la hauteur et le rayon du compact suivant :



on peut construire une suite croissante de compacts (K_n) telle que l'union des intérieurs des K_n est égale à X. On note que les complémentaires des K_n sont connexes.

De même on peut construire une suite de compacts (L_n) telle que l'union des intérieurs des L_n est égale à Y.



Le compact L_n est à l'extérieur du tore T_n et un l'intérieur de la sphère S_n. On note que les complémentaires des L_n ont deux composantes connexes. Pour chaque n, on note $(Y - L_n)^0$ la composante connexe dans la boule et $(Y - L_n)^1$ la composante connexe à l'extérieur de la boule.

Question 3 : Supposons qu'il existe un homéomorphisme $f : X \to Y$. Fixons un entier n. Comme f est un homéomorphisme, $f(\text{Int}(K_n))$ est ouvert et on a

$$L_n \subset \bigcup_{k=0}^{\infty} f(\text{Int}(K_k))$$

Par compacité de L_n, on peut trouver k_n tel que $L_n \subseteq f(K_{k_n})$, de plus on peut supposer que la suite (k_n) est strictement croissante. Pour simplifier, renommons la suite $(f(K_{k_n}))$ par (K'_n). On voit que la suite (K'_n) est toujours une suite croissante de compacts dont les intérieurs recouvrent Y et dont les complémentaires sont connexes. Nous avons démontré que $L_n \subseteq K'_n$ ce qui implique $Y - K'_n \subseteq Y - L_n$. Comme $Y - K'_n$ est connexe, ceci implique $Y - K'_n \subseteq (Y - L_n)^0$ ou $Y - K'_n \subseteq (Y - L_n)^1$. La suite $Y - K'_n$ étant décroissante, on peut vérifier que

$$(\forall n \; Y - K'_n \subseteq (Y - L_n)^0) \text{ ou } (\forall n \; Y - K'_n \subseteq (Y - L_n)^1)$$

Supposons que $Y - K'_n \subseteq (Y - L_n)^0$ pour tout n (l'autre cas est similaire). Maintenant, en refaisant le raisonnement plus haut, on voit que pour tout n il existe l_n tel que $(Y - L_{l_n}) \subseteq (Y - K'_n)$ (K'_n est un compact recouvert par les intérieurs de L_k). De plus on peut choisir la suite (l_n) strictement croissante. Ceci aboutit à $(Y - L_{l_n}) \subseteq (Y - K'_n) \subseteq (Y - L_n)^0$ et en particulier à

$$(Y - L_{l_n})^1 \subseteq (Y - L_n)^0$$

La contradiction vient du fait que $(Y - L_a)^1 \cap (Y - L_b)^0 = \emptyset$ pour tout a, b. En conclusion X et Y ne sont pas homéomorphes.

Certains candidats ont utilisé l'homologie pour démontrer que X n'est pas homéomorphe à Y, en effet $H_2(X, \mathbb{Z}) = 0$ et $H_2(Y, \mathbb{Z}) = \mathbb{Z}$. Cette solution était tout à fait valide.

Question 4 : Soit D_n une boule fermée de rayon $n \geq 1$ dans $\mathbb{R}^3$. La suite de boules (D_n) permet de construire une suite croissante de compacts A_n dans $\mathbb{R}^3 - L_1$ telle que

$$\bigcup_{k=0}^{\infty} \text{Int}(A_k) = \mathbb{R}^3 - L_1$$

De même on peut construire une suite croissante de compacts B_n dans $\mathbb{R}^3 - L_1$ telle que

$$\bigcup_{k=0}^{\infty} \text{Int}(B_k) = \mathbb{R}^3 - L_2$$

On remarque alors que pour tout n, le complémentaire de B_n est connexe alors que le nombre de composantes connexes de A_n tend vers l'infini quand n tend vers l'infini. Un raisonnement similaire à celui fait à la quesiton 3 permet de conclure que $R^3 - L_1$ n'est pas homéomorphe à $R^3 - L_2$. La notion cachée ici est la notion de bouts à l'infini d'un espace topologique. Si X est un espace topologique et (K_n) est une suite croissante de compacts dont les intérieurs recouvrent X alors un bout à l'infini de X est une suite croissante (U_n) d'ouverts telle que pour tout n, U_n est une composante connexe de $X - K_n$. Le point important est que le cardinal de l'ensemble des bouts de X est indépendant de la suite (K_n). Le cardinal de l'ensemble des bouts d'un espace topologique constitue donc un invariant topologique.