(*Texte public*)

Résumé : Ce texte présente une modélisation d'un paysage politique au moyen de partitions aléatoires. On considère un pays imaginaire constitué d'un grand nombre de citoyens, que l'on suppose infini. Ces citoyens se regroupent progressivement en partis politiques avec un mécanisme aléatoire tenant compte de l'ambition. Le paysage politique correspond à la répartition des citoyens en partis politiques. Le texte a pour but la modélisation et l'étude probabiliste de l'évolution de ce paysage et de son comportement asymptotique, ainsi que l'estimation statistique d'un paramètre lié à l'ambition.

Mots clefs : Partition aléatoire, loi de Bernoulli, estimateur, chaîne de Markov.

1. La modélisation

À l'instant $n \geq 1$, il y a n citoyens membres de partis politiques. Si ces citoyens sont numérotés de 1 à n, les partis politiques qu'ils forment sont modélisés par une partition aléatoire B_n de l'ensemble fini $\{1,\dots,n\}$ en $|B_n|$ partis politiques. À l'instant $n+1$, un nouveau citoyen entreprenant, numéroté $n+1$, décide soit de rejoindre l'un des $|B_n|$ partis déjà existants avec une probabilité proportionnelle à sa taille, soit, s'il est plus ambitieux qu'opportuniste, de créer un nouveau parti. On modélise par $\theta/(n+\theta)$ la probabilité qu'a le citoyen $n+1$ de créer un nouveau parti, où θ est un paramètre réel positif fixé et inconnu, qui sera interprété comme un coefficient d'ambition. On suppose pour simplifier qu'une fois son parti politique choisi, le citoyen engagé y reste fidèle. Dans la culture citoyenne, l'infidélité est très mal vue et durement sanctionnée par tous.

Ce modèle correspond à un processus à temps discret $(B_n)_{n\geq 1}$ où pour tout n, B_n est une partition aléatoire de $\{1,\dots,n\}$ constituée de $|B_n|$ sous-ensembles non vides de $\{1,\dots,n\}$. On convient que $B_1 = \{1\}$. Si B est une partition de $\{1,\dots,n\}$, la notation $b \in B$ signifie que b est un sous-ensemble non vide de $\{1,\dots,n\}$ de cardinal $|b|$ faisant partie de la partition B. On dit que b est un *bloc* de taille $|b|$ de la partition B. Le modèle qui décrit $(B_n)_{n\geq 1}$ est sans mémoire dans la mesure où la loi de B_{n+1} sachant $B_1,\dots,B_n$ ne dépend que de B_n.

Soit B une partition de $\{1,\dots,n\}$ et B' une partition de $\{1,\dots,n+1\}$. On désigne par B_+ la partition de $\{1,\dots,n+1\}$ de taille $|B|+1$ obtenue à partir de B en l'enrichissant avec le singleton $\{n+1\}$. Si $b \in B$, on note $B \to_b B'$ lorsque B' s'obtient à partir de B en remplaçant le bloc b par $b \cup \{n+1\}$. Notons que si $B \to_b B'$ alors $|B| = |B'|$. Par définition de B_{n+1}, on a

$$\mathbb{P}(B_{n+1} = B' \,|\, B_n = B) = \begin{cases} \dfrac{|b|}{\theta+n} & \text{si } B \to_b B', \\[2ex] \dfrac{\theta}{\theta+n} & \text{si } B' = B_+, \\[2ex] 0 & \text{sinon.} \end{cases}$$

Remarque 1 (Interprétation du paramètre θ)**.** *Lorsque $\theta = 0$, la probabilité du cas $B' = B_+$ est nulle, tandis que lorsque $\theta \to \infty$, la probabilité du cas $B \to_b B'$ tend vers 0. Ces probabilités sont de plus monotones en θ. Ainsi, plus θ est grand, plus les citoyens ont tendance à créer de nouveaux partis politiques plutôt que de rejoindre des partis déjà existants. Pour cette raison, le paramètre θ peut être interprété comme un coefficient d'ambition dans le modèle.*

2. Les deux lois des partis politiques

La première loi des partis politiques correspond à la loi de la partition aléatoire B_n pour tout $n \geq 1$, tandis que la seconde loi des partis politiques correspond à la loi des tailles de blocs de B_n pour tout $n \geq 1$.

Théorème 1 (Première loi)**.** *Pour tout entier $n \geq 1$ et toute partition B de $\{1, \dots, n\}$,*

$$\mathbb{P}(B_n = B) = \frac{\theta^{|B|}}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{b \in B} (|b|-1)!.$$

Démonstration. Découle d'un raisonnement par récurrence sur n, basé sur l'expression de la loi conditionnelle de B_{n+1} sachant B_n. □

Corollaire 1 (Seconde loi)**.** *Pour tout temps $n \geq 1$ et tout $1 \leq k \leq n$, soit $C_{n,k}$ le nombre de partis politiques de tailles k au temps n, de sorte que $n = C_{n,1} + 2C_{n,2} + \cdots + nC_{n,n}$. Alors pour tout n-uplet d'entiers $0 \leq a_1, \dots, a_n \leq n$ tels que $a_1 + 2a_2 + \cdots + na_n = n$, on a*

$$\mathbb{P}(C_{n,1} = a_1, \dots, C_{n,n} = a_n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{k=1}^{n} \frac{1}{a_k!} \left(\frac{\theta}{k}\right)^{a_k}.$$

Démonstration. Pour tout $1 \leq k \leq n$, l'entier $C_{n,k}$ correspond au nombre de blocs de taille k dans la partition aléatoire B_n. Soit B une partition de $\{1, \dots, n\}$ en blocs $b_1, \dots, b_{|B|}$. Soit $a_1, \dots, a_n$ la suite des nombres de blocs de tailles respectives $1, \dots, n$. On a alors $a_1 + 2a_2 + \cdots + na_n = n$ et $a_1 + \cdots + a_n = |B|$. Le nombre de telles partitions (i.e. à nombres $a_1, \dots, a_n$ prescrits) vaut

$$n! \prod_{k=1}^{n} \frac{1}{(k!)^{a_k} a_k!}.$$

La seconde loi découle alors de la première par un jeu de réécriture. □

3. Diversité politique, micros partis, et parti unique

On suppose à présent que $\theta > 0$. Au temps $n \geq 1$, le paysage politique citoyen se compose de $|B_n|$ partis politiques différents. Le théorème suivant donne les deux premiers moments de l'entier aléatoire $|B_n|$.

Théorème 2 (Diversité politique)**.** *Pour tout $n \geq 1$, la moyenne et la variance de $|B_n|$ sont données par*

$$\mathbb{E}(|B_n|) = \sum_{k=0}^{n-1} \frac{\theta}{\theta + k} \qquad \textit{et} \qquad \mathrm{Var}(|B_n|) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta + k)^2}.$$

Démonstration. La variable aléatoire $|B_n|$ s'écrit $|B_n| = \xi_1 + \cdots + \xi_n$ où $\xi_1, \ldots, \xi_n$ sont des variables aléatoires indépendantes de lois de Bernoulli vérifiant pour tout $1 \leq k \leq n$

$$\mathbb{P}(\xi_k = 1) = 1 - \mathbb{P}(\xi_k = 0) = \frac{\theta}{\theta + k - 1}.$$

Les formules exprimant les deux premiers moments de $|B_n|$ en découlent directement. □

Remarque 2. *Par construction, la suite* $(|B_n|)_{n\geq 1}$ *est presque-sûrement croissante et ne fait que des sauts de* 1*. Elle est cependant de moins en moins croissante en quelque sorte puisque la quantité $p_n = \mathbb{P}(|B_{n+1}| = |B_n| + 1) = \mathbb{P}(\xi_{n+1} = 1) = \theta(\theta + n)^{-1}$ décroît quand n croît. Malgré tout, la moyenne et la variance de $|B_n|$ sont équivalentes à* $\theta \log(n)$ *lorsque n tend vers ∞. Ici, la notation* log *désigne la fonction logarithme naturel (népérien).*

La seconde loi des partis politiques est une loi sur $\mathbb{N}^n$ portée par l'ensemble

$$\{(a_1, \ldots, a_n) \in \mathbb{N}^n ; a_1 + 2a_2 + \cdots + na_n = n\}.$$

Théorème 3 (Loi des micro partis)**.** *Pour tout temps $n \geq 1$, le nombre $C_{n,1}$ de partis politiques réduits à leur créateur vérifie*

$$\mathbb{E}(C_{n,1}) = \frac{n\theta}{n + \theta - 1} \qquad \textit{et} \qquad \mathrm{Var}(C_{n,1}) = \frac{n(n-1)(n-2+2\theta)\theta}{(n+\theta-2)(n+\theta-1)^2}.$$

Démonstration. Il s'agit de décrire la loi de la première composante $C_{n,1}$ d'un vecteur aléatoire C_n de $\mathbb{N}^n$ qui suit la seconde loi des partis politiques. Il est cependant plus commode de voir $C_{n,1}$ comme le nombre de blocs de taille 1 dans B_n. Pour tous entiers $0 \leq a \leq n$ et $0 \leq a' \leq n + 1$, la probabilité conditionnelle $\mathbb{P}(C_{n+1,1} = a' \,|\, C_{n,1} = a)$ se calcule facilement en remarquant que seuls les cas où $|a - a'| \leq 1$ comptent. Cela spécifie totalement la loi conditionnelle $\mathscr{L}(C_{n+1,1} \,|\, C_{n,1} = a)$. En particulier, on obtient

$$\mathbb{E}(C_{n+1,1} \,|\, C_{n,1} = a) = \frac{a(\theta + n - 1) + \theta}{\theta + n}.$$

Cela donne $(n+\theta)m_{n+1} = (\theta + n - 1)m_n + \theta$ où m_n désigne l'espérance de $C_{n,1}$. La formule annoncée pour m_n s'en déduit. Pour la variance, on procède de la même manière avec la formule

$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X \,|\, Y)) + \mathrm{Var}(\mathbb{E}(X \,|\, Y)).$$

Les calculs sont cependant un peu plus lourds que pour l'espérance. □

Remarque 3. *Les formules pour les deux premiers moments de $C_{n,1}$ entraînent*

$$\lim_{n\to\infty} \mathbb{E}(C_{n,1}) = \theta \qquad et \qquad \lim_{n\to\infty} \mathrm{Var}(C_{n,1}) = \theta.$$

Ainsi, en temps grand, la moyenne et la variance du nombre de partis ne comprenant qu'un seul membre (le créateur) sont proches du coefficient d'ambition θ.

L'entier $C_{n,n}$ représente le nombre de partis de taille n, autrement dit le nombre de blocs de taille n dans B_n. Il s'agit du nombre de partis uniques. On a $C_{n,n} = 1$ si et seulement si $|B_n| = 1$. Le théorème suivant précise les choses, et montre en particulier que le modèle des partis politiques est asymptotiquement démocratique en quelque sorte.

Théorème 4 (Parti unique)**.** *On a $C_{1,1} = 1$ et $C_{n,n} \in \{0,1\}$ pour tout $n \geq 1$. De plus, la suite $(C_{n,n})_{n\geq 1}$ est presque-sûrement décroissante, et pour tout $n \geq 2$, la probabilité que le paysage politique soit réduit à un parti unique vaut*

$$\mathbb{P}(C_{n,n} = 1) = \prod_{k=1}^{n-1} \frac{k}{\theta + k}.$$

Enfin, la suite $(C_{n,n})_{n\geq 1}$ converge presque-sûrement vers 0.

Démonstration. Les premières assertions du théorème découlent directement de la définition de B_n. L'expression de la probabilité s'obtient en considérant des variables aléatoires de Bernoulli ξ indépendantes mais non équidistribuées. Pour la convergence presque-sûre, on remarque tout d'abord que, puisque $\theta > 0$, le produit infini $\prod_{k=1}^{\infty}(1 + \theta k^{-1})$ diverge car la série harmonique diverge, et par conséquent $\lim_{n\to\infty} \mathbb{P}(C_{n,n} = 1) = 0$. D'autre part, la suite d'événements $(A_n)_{n\geq 1}$ définie par $A_n = \{C_{n,n} = 0\}$ est croissante. □

4. Estimation du paramètre

Le théorème suivant permet d'estimer θ avec une seule réalisation de la suite $(|B_n|)$, c'est-à-dire avec une seule réalisation de la vie politique.

Théorème 5. *La convergence en probabilité suivante a lieu :*

$$\frac{|B_n|}{\log(n)} \xrightarrow[n\to\infty]{\mathbb{P}} \theta.$$

Démonstration. En vertu de l'inégalité de Bienaymé-Tchebychev et du théorème de diversité politique, on a pour tout réel $\varepsilon > 0$ fixé, et pour tout entier $n \geq 2$,

$$\mathbb{P}\left(\left|\frac{|B_n|}{\log(n)} - \theta\right| > \varepsilon\right) \leq \frac{\mathrm{Var}(|B_n|) + (\mathbb{E}(|B_n|) - \theta \log(n))^2}{(\varepsilon \log(n))^2} = O\left(\frac{1}{\log(n)}\right) = o(1).$$

□

Le théorème suivant précise les fluctuations autour de la moyenne.

Théorème 6. *La convergence en loi suivante a lieu :*

$$\frac{|B_n| - \mathbb{E}(|B_n|)}{\sqrt{\mathrm{Var}(|B_n|)}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0,1).$$

De plus, $\mathbb{E}(|B_n|)$ *et/ou* $\mathrm{Var}(|B_n|)$ *peuvent être remplacés par* $\theta\log(n)$.

La preuve du théorème 6 fait appel à la définition et au lemme suivants. On admettra le lemme, très subtil bien que d'apparence élémentaire.

Définition 1 (Distance en variation)**.** *Soient* μ *et* ν *deux lois de probabilité sur* $\mathbb{N}$. *On appelle* distance en variation *entre* μ *et* ν *le réel* $d(\mu,\nu) \in [0,1]$ *défini par*

$$d_V(\mu,\nu) = \frac{1}{2}\sum_{x\in\mathbb{N}} |\mu(\{x\}) - \nu(\{x\})|.$$

Lemme 1 (Approximation poissonnienne)**.** *Soit un entier* $n \geq 1$ *et* $S_n = \xi_1 + \cdots + \xi_n$ *la somme de* n *variables aléatoires indépendantes* $\xi_1,\ldots,\xi_n$ *de lois de Bernoulli de paramètres respectifs* $p_1,\ldots,p_n$ *dans* $]0,1]$. *On note* μ_n *la loi de* S_n *et* ν_n *la loi de Poisson de moyenne* $p_1+\cdots+p_n$. *Alors la majoration suivante a lieu*

$$d_V(\mu_n,\nu_n) \leq \frac{p_1^2+\cdots+p_n^2}{p_1+\cdots+p_n}.$$

Démonstration du théorème 6. Pour tout entier $n \geq 1$, on note μ_n la loi de $|B_n|$ et ν_n la loi de Poisson de moyenne $\lambda_n = \mathbb{E}(|B_n|)$. On pose également $\sigma_n = \sqrt{\mathrm{Var}(|B_n|)}$. Soit enfin P_n une variable aléatoire de Poisson de moyenne λ_n (et donc de loi ν_n et de variance λ_n). Le lemme 1 combiné au théorème de diversité politique fournit $d_V(\mu_n,\nu_n) = o(1)$. Par conséquent, pour toute fonction continue et bornée $f:\mathbb{R}\to\mathbb{R}$, on a

$$\left|\mathbb{E}\left(f\left(\frac{|B_n|-\lambda_n}{\sigma_n}\right)\right) - \mathbb{E}\left(f\left(\frac{P_n-\lambda_n}{\sigma_n}\right)\right)\right| \leq 2\|f\|_\infty\, d_V(\mu_n,\nu_n) = o(1).$$

Il suffit donc d'établir que la suite $((P_n-\lambda_n)/\sigma_n)$ converge en loi vers $\mathcal{N}(0,1)$. Pour cela, on remarque tout d'abord que

$$\frac{P_n-\lambda_n}{\sigma_n} = \frac{P_n-\lambda_n}{\sqrt{\lambda_n}}\frac{\sqrt{\lambda_n}}{\sigma_n}.$$

Maintenant, la suite $((P_n-\lambda_n)/\sqrt{\lambda_n})$ converge en loi vers $\mathcal{N}(0,1)$ tandis que la suite déterministe $(\sqrt{\lambda_n}/\sigma_n)$ converge vers 1. □

Suggestions pour le développement

- *Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.*

— *Aspects mathématiques.*

Soit $\mathscr{B}_n$ l'ensemble des partitions de $\{1,\ldots,n\}$. On pourra vérifier que $\sum_{B'\in\mathscr{B}_{n+1}} \mathbb{P}(B_{n+1}=B' \mid B_n=B)=1$ pour tout $B\in\mathscr{B}_n$. Les suites $(B_n)_{n\geq 1}$, $(|B_n|)_{n\geq 1}$, $(C_{n,1})_{n\geq 1}$, $(C_{n,n})_{n\geq 1}$, $(C_n)_{n\geq 1}$ constituent-elles des chaînes de Markov ? On pourra préciser le cas échéant l'espace d'état et l'homogénéité. On pourra détailler les preuves des théorèmes. À quels théorèmes classiques vous font penser les théorèmes 5 et 6 ? Quels sont le biais et la vitesse de l'estimateur $|B_n|/\log(n)$ de θ ? Cette vitesse vous semble-t-elle satisfaisante ? Vous pourrez préciser une méthode de construction d'intervalles de confiance pour θ. La preuve du théorème 5 peut-elle conduire à une convergence p.-s. au moyen du lemme de Borel-Cantelli ? Les suites $(|B_n|)_n$ et $(|B_n|-\mathbb{E}(|B_n|)_n$ sont-elles des martingales ?

— *Programmes et simulations.*

On pourra écrire un programme qui simule des réalisations de la suite aléatoire $(B_n)_{n\geq 1}$. On pourra également illustrer le comportement de $(|B_n|)_{n\geq 1}$, $(C_{n,1})_{n\geq 1}$ et $(C_{n,n})_{n\geq 1}$ par exemple pour différentes valeurs du paramètre θ. Comment simuler $(|B_n|)_{n\geq 1}$ sans passer par $(B_n)_{n\geq 1}$? On pourra illustrer l'estimation du paramètre θ. Seriez-vous en mesure de fournir un intervalle de confiance associé à l'estimation de θ ?

— *Modélisation.*

Le modèle des partis politiques favorise-t-il l'apparition de petits partis ? le maintien de petits partis ? Critiquer le modèle. Proposer un modèle dans lequel les citoyens déjà membres d'un parti peuvent changer de parti. Expliquer quelques aspects qualitatifs de ce nouveau modèle par rapport à celui proposé dans le texte. Comment modéliser de manière plus souple le fait que l'infidélité est durement sanctionnée ? Expliquer pourquoi les suites $(|B_n|)_{n\geq 1}$ et $(C_{n,n})_{n\geq 1}$ sont reliées à un jeu de pile ou face dans lequel la probabilité de gagner change à chaque lancer. Seriez-vous en mesure d'imaginer d'autres situations justiciables d'une modélisation par un tel jeu de pile ou face ?