TD VII : Fragmentation

- Octobre 2025
- Master I ISIFAR
- **Probabilités**

Exercice 1

Exercice 2

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 $\{1, \ldots, n\}$ 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 (politique) 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 $\{1, \ldots, n\}$ constituée de $|B_n|$ sous-ensembles non vides de $\{1, \ldots, n\}$.

On convient que $B_1 = \{1\}$.

Si B est une partition de $\{1, \ldots, n\}$, la notation $b \in B$ signifie que b est un sous-ensemble non vide de $\{1, \ldots, n\}$ de cardinal $|b|$ faisant partie de la partition B (b est une des classes 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 $B_1, \ldots, B_n$ ne dépend que de B_n.

Soit B une partition de $\{1, \ldots, n\}$ et B' une partition de $\{1, \ldots, n+1\}$. On désigne par B_+ la partition de $\{1, \ldots, n+1\}$ de taille $|B|+1$ obtenue à partir de B en l'enrichissant avec le singleton $\{n+1\}$.

Si $b \in B$, on note $B \rightarrow_b B'$, lorsque B' s'obtient à partir de B en inscrivant le citoyen $n+1$ au parti b : on remplace b par $b \cup \{n+1\}$ (on a alors $|B'| = |B|$).

$$\Pr\left\{B_{n+1} = B' \mid B_n = B\right\} = \begin{cases} \frac{|b|}{\theta+n} & \text{si } B \rightarrow_b B' \\ \frac{\theta}{\theta+n} & \text{si } B' = B_+ \\ 0 & \text{sinon.} \end{cases}$$

Remarque 1 (Interprétation du paramètre θ). Lorsque $\theta = 0$, la probabilité du cas $B' = B_+$ est nulle, tandis que lorsque $\theta \rightarrow \infty$, la probabilité du cas $B \rightarrow_b B'$ 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 s'affilier à des partis déjà existants. Pour cette raison, le paramètre θ peut être interprété comme un coefficient d'ambition dans le modèle.

Exercice 3 (Loi de la partition aléatoire)

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 du profil $(C_{n,1}, C_{n,2}, \ldots,)$ des tailles des blocs de B_n (un profil pour chaque $n \geq 1$).

Théorème 1 (Première loi).

Pour tout entier $n \geq 1$ et toute partition B de $\{1, \dots, n\}$,

$$\Pr\left\{B_n = B\right\} = \frac{\theta^n}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{b \in B} (|b|-1)! = \frac{\theta^{|B|}\Gamma(\theta)}{\Gamma(\theta+n)} \prod_{b \in B} (|b|-1)!$$

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 taille k au temps n, de sorte que $n = C_{n,1} + 2C_{n,2} + \dots + nC_{n,n}$. Alors pour tout n-uplet d'entiers $0 \leq a_1, \dots, a_n \leq n$ tels que $a_1 + 2a_2 + \dots + na_n = n$, on a

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

Diversité politique, micropartis, 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

$$\mathbb{E}|B_n| = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k}$$

et

$$\operatorname{var}(|B_n|) = \sum_{k=0}^{n-1} \frac{\theta k}{(\theta+k)^2}$$

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

$$\mathbb{E}C_{n,1} = \frac{n\theta}{n-1+\theta}$$

et

$$\operatorname{var}(C_{n,1}) = \frac{n(n+2\theta-2)(n-1)\theta}{(n+\theta-2)(n+\theta-1)^2}.$$

Si G_n désigne la fonction génératrice de la loi de $C_{n,1}$, on a la relation suivante :

$$G_{n+1}(s) - G_n(s) = \frac{1-s}{n+\theta}\Big(G_n'(s) - \theta G_n(s)\Big)$$

pour $s \in [0, 1]$

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. Lorsque $C_{n,n} = 1$, on a affaire à un parti unique. On a $C_{n,n} = 1$ si et seulement si $|B_n| = 1$. Le théorème suivant montre que le modèle des partis politiques est asymptotiquement pluraliste.

Théorème 4 (Parti unique). Pour tout $n \geq 2$, la probabilité que le paysage politique soit réduit à un parti unique vaut

$$\Pr\left\{C_{n,n} = 1\right\} = \frac{\Gamma(n)\Gamma(\theta+1)}{\Gamma(n+\theta)}$$

Fluctuations

: : :

Théorème 5. La convergence en probabilité suivante a lieu :

$$\lim_n \frac{|B_n|}{\log(n)} = \theta \qquad \text{en probabilité.}$$

: : :

Le théorème suivant précise les fluctuations autour de la moyenne.

Théorème 6. La convergence en loi suivante a lieu : Pour une suite $(a_n)_n$ bien choisie

$$\frac{|B_n| - \mathbb{E}|B_n|}{a_n} \rightsquigarrow \mathcal{N}(0,1)$$

Exercice 3 (Approche du théorème 6)

Pour établir le théorème 6, on peut procéder en deux temps : montrer d'abord que lorsque n est grand le nombre de partis tend à être distribué selon une loi de Poisson de même espérance ; puis montrer qu'après recentrage et renormalisation une variable de Poisson de grande espérance est approximativement gaussienne. On quantifie l'approximation de lois à l'aide de deux notions différentes : dans un premier temps la distance en variation, dans un second temps à l'aide de la notion de convergence faible.

Définition (Distance en variation).

Soient P et Q deux lois sur $(\Omega, \mathcal{F})$. On appelle distance en variation entre P et Q :

$$\mathrm{d}_{\mathrm{TV}}(P,Q) = \sup_{A \in \mathcal{F}} P(A) - Q(A)\,.$$

Théorème : Loi des événements rares (version facile) Soit un entier $n \geq 1$ et $S_n = \xi_1 + \ldots + \xi_n$ 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 d'espérance $p_1 + \ldots + p_n$.

$$\mathrm{d}_{\mathrm{TV}}(\mu_n, \nu_n) \leq \sum_{i=1}^n p_i^2\,.$$

Théorème : Loi des événements rares (version fine) Soit un entier $n \geq 1$ et $S_n = \xi_1 + \ldots + \xi_n$ 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 d'espérance $p_1 + \ldots + p_n$.

$$\mathrm{d}_{\mathrm{TV}}(\mu_n, \nu_n) \leq \frac{\sum_{i=1}^n p_i^2}{\sum_{i=1}^n p_i}\,.$$

Questions

Si on butte sur une question où il est demandé de démontrer un énoncé, on a le droit d'utiliser cet énoncé pour répondre aux questions suivantes.

1. Établir les deux formules du théorème 2.
2. Montrer le théorème 4 (probabilité d'obtenir un parti unique)
3. Montrer que la suite $\mathbb{I}_{C_{n,n}=1}$ tend presque sûrement vers 0.
4. Calculer $\mathbb{E}[C_{n+1,1} \mid \sigma(C_{n,1})]$.

5. Calculer $\mathbb{E}[C_{n,1}]$ pour $n \geq 1$. (Prouver première formule du Théorème 3).
6. Calculer $\operatorname{var}\left(C_{n+1,1} \mid \sigma(C_{n,1})\right)$
7. Calculer $\operatorname{var}\left(C_{n,1}\right)$ pour $n \geq 1$. (Prouver deuxième formule du Théorème 3).
8. Établir l'équation qui relie G_{n+1} G_n et G'_n. (Prouver troisième formule du Théorème 3).
9. Calculer la fonction génératrice G_n de la loi de $C_{n,1}$ lorsque $\theta = 1$
10. Établir une relation entre $\mathbb{E}[C_{n+1,2}]$ et $\mathbb{E}[C_{n,2}]$ $\mathbb{E}[C_{n,1}]$.
11. Montrer que $\lim_{n\to\infty} \mathbb{E}C_{n,2} = \frac{\theta}{2}$.
12. Établir le théorème 5.
13. Montrer que d_{TV} définit une distance sur l'ensemble des lois de probabilités sur $(\Omega, \mathcal{F})$.
14. Soient μ et ν deux lois de probabilité sur $\mathbb{N}$. Montrer que la distance en variation entre μ et ν satisfait
$$\mathrm{d}_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in \mathbb{N}} |\mu(x) - \nu(x)|.$$
15. Montrer que si P est la loi d'une variable de Bernoulli de paramètre $p \in [0,1]$ et Q est une loi de Poisson de paramètre p alors
$$\mathrm{d}_{\mathrm{TV}}(P, Q) \leq p^2\,.$$
16. Montrer que
$$\mathrm{d}_{\mathrm{TV}}(P, Q) \leq \mathbb{P}\{X \neq Y\}$$
si $\mathbb{P}$ est une loi telle que $P = \mathbb{P} \circ X^{-1}$ et $Q = \mathbb{P} \circ Y^{-1}$ (un *couplage* des lois P et Q).
17. Montrer la version facile de la loi des événements rares.
18. En utilisant la version fine de la loi des événements rares, majorer la distance en variation entre la loi de $|B_n|$ et la loi de Poisson de même espérance
19. Si P_X et P_Y sont à distance en variation d, quelle est la distance en variation entre les lois de $\sigma X + \mu$ et de $\sigma Y + \mu$? 1.Montrer que si $X_n \sim \text{Poisson}(n)$ alors $(X_n - n)/\sqrt{n} \rightsquigarrow \mathcal{N}(0,1)$
20. Montrer le théorème 6 (en utilisant la version fine de la loi des événements rares).