PARTITIONS ALÉATOIRES D'EWENS–PITMAN

mots-clés : partitions aléatoires, sous-martingales et chaînes de Markov rétrogrades.

Ce texte présente un modèle de partitions aléatoires qui sont construites récursivement, et qui ont des propriétés asymptotiques intéressantes. On rappelle qu'une *partition* de l'ensemble $[\![1,n]\!] = \{1,2,3,\ldots,n\}$ est une famille $\pi = (\pi_1,\pi_2,\ldots,\pi_\ell)$ de parties disjointes et non vides de $[\![1,n]\!]$ telles que $[\![1,n]\!] = \pi_1 \sqcup \pi_2 \sqcup \cdots \sqcup \pi_\ell$. Par exemple,

$$\pi = (\pi_1,\pi_2,\pi_3,\pi_4) = (\{1,3,7\},\{2\},\{4,5,8,9\},\{6\}) = \begin{matrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{matrix}$$

est une partition de taille 9. Notons $\mathfrak{P}(n)$ l'ensemble des partitions de taille n. Si $\pi \in \mathfrak{P}(n)$, on ordonnera toujours ses parts $\pi_1,\ldots,\pi_\ell$ en fonction de leurs éléments minimaux : $\min \pi_1 < \min \pi_2 < \cdots < \min \pi_\ell$ (comme dans l'exemple ci-dessus). Ceci permet de définir sans ambiguïté la liste $c(\pi) = (c_1(\pi), c_2(\pi),\ldots,c_\ell(\pi))$ des tailles des parts d'une partition $\pi \in \mathfrak{P}(n)$. Dans l'exemple ci-dessus, $c(\pi) = (3,1,4,1)$. Notons aussi $\ell(\pi)$ le nombre de parts de π, et $\lambda(\pi)$ le réordonnement décroissant des tailles des parts de π ; dans l'exemple précédent, $\ell(\pi) = 4$ et $\lambda(\pi) = (\lambda_1(\pi),\lambda_2(\pi),\lambda_3(\pi),\lambda_4(\pi)) = (4,3,1,1)$. On dira que $\lambda(\pi)$ est la *signature* de π.

Supposons donnée une partition $\pi \in \mathfrak{P}(n)$. Elle représente par exemple la répartition de n personnes en autant de groupes d'opinion. Si une nouvelle personne $n+1$ arrive dans cette population, elle peut :

- soit rejoindre l'un des groupes π_k déjà existant ;
- soit créer un nouveau groupe $\{n+1\}$.

La première section du texte définit des règles de transition qui construisent une partition Π de taille $n+1$ à partir de la partition π de taille n, en prenant en compte deux facteurs qui poussent la personne $n+1$ à créer un nouveau groupe. En itérant cette construction, on obtient ainsi une suite aléatoire

$$(\pi^{(1)},\pi^{(2)},\ldots,\pi^{(n)},\ldots)$$

de partitions, chaque $\pi^{(n)}$ étant de taille n. Cette suite est *cohérente*, au sens suivant : pour tous entiers $n \leq N$, les parts de $\pi^{(n)}$ sont les intersections non vides des parts de $\pi^{(N)}$ avec $[\![1,n]\!]$. La seconde section du texte montre que sous une hypothèse d'*échangeabilité*, les tailles des plus grandes parts d'une suite cohérente de partitions aléatoires ont des *fréquences limites* : pour tout $k \geq 1$,

$$\frac{\lambda_k(\pi^{(n)})}{n} \to_{\binom{n\to\infty}{\text{p.s}}} Y_k$$

où les Y_k sont des variables aléatoires dans $[0,1]$, avec $\sum_{k=1}^{\infty} Y_k \leq 1$. La loi de ces fréquences limites sera évoquée à la toute fin du texte.

1. Les paramètres θ et α

Lorsqu'un nouvel individu $n+1$ arrive dans une population de taille n partitionnée suivant $\pi^{(n)}$, deux facteurs le poussent à créer un nouveau groupe :

- l'ambition *individuelle* θ : l'individu $n+1$ a une propension naturelle à créer son propre groupe.

- l'ambition *induite* α : chaque groupe $\pi_k^{(n)}$ déjà présent montre à $n+1$ que l'on peut créer sans risque son propre groupe, et ceci le pousse dans cette direction.

Si $\pi^{(n)} = \pi = (\pi_1, \ldots, \pi_\ell)$ est fixée dans $\mathfrak{P}(n)$, il existe $\ell+1$ partitions $\mathit{\Pi} \in \mathfrak{P}(n+1)$ telles que les parts de π sont les intersections non vides des parts de $\mathit{\Pi}$ avec $[\![1,n]\!]$:

- les successeurs $\pi[k] = (\pi_1, \ldots, \pi_k \sqcup \{n+1\}, \ldots, \pi_\ell)$, obtenus en rajoutant $n+1$ à l'une des parts de π ;
- le successeur $\pi[\ell+1] = (\pi_1, \ldots, \pi_\ell, \{n+1\})$, obtenu en créant un nouveau groupe $\{n+1\}$.

Par exemple, si π est la partition de l'introduction, alors $\pi[2] = (\{1,3,7\}, \{2,10\}, \{4,5,8,9\}, \{6\})$, et $\pi[5] = (\{1,3,7\}, \{2\}, \{4,5,8,9\}, \{6\}, \{10\})$.

Si l'on veut prendre en compte les deux paramètres d'ambition θ et α, il est naturel de demander que la probabilité conditionnelle pour obtenir $\pi^{(n+1)} = \pi[\ell+1]$ sachant $\pi^{(n)} = \pi$ soit proportionnelle à $\theta + \alpha\,\ell(\pi)$. Par ailleurs, pour $k \in [\![1,\ell]\!]$, il est naturel de demander que la probabilité conditionnelle pour obtenir $\pi^{(n)} = \pi[k]$ sachant $\pi^{(n)} = \pi$ soit une fonction croissante de la taille c_k de la part π_k. Ainsi, s'il ne crée pas son propre groupe, alors $n+1$ a tendance à rejoindre l'un des plus grands groupes déjà existants. Ceci motive la définition suivante :

Définition 1. *On fixe deux paramètres $\theta > 0$ et $\alpha \in [0,1)$. Soit $\mathfrak{P} = \bigsqcup_{n=1}^{\infty} \mathfrak{P}(n)$. La suite de partitions aléatoires $(\pi^{(n)})_{n\geq 1}$ de paramètres (θ, α) est la chaîne de Markov sur $\mathfrak{P}$:*

- *d'état initial $\pi^{(1)} = (\{1\})$ (l'unique élément de $\mathfrak{P}(1)$) ;*
- *telle que $\pi^{(n)} \in \mathfrak{P}(n)$ pour tout $n \geq 1$;*
- *dont le noyau de transition est défini pour π de taille n et avec ℓ parts par :*

$$P(\pi, \pi[\ell+1]) = \frac{\theta + \alpha\,\ell}{\theta + n} \qquad ; \qquad P(\pi, \pi[k]) = \frac{c_k(\pi) - \alpha}{\theta + n} \quad \text{si } 1 \leq k \leq \ell.$$

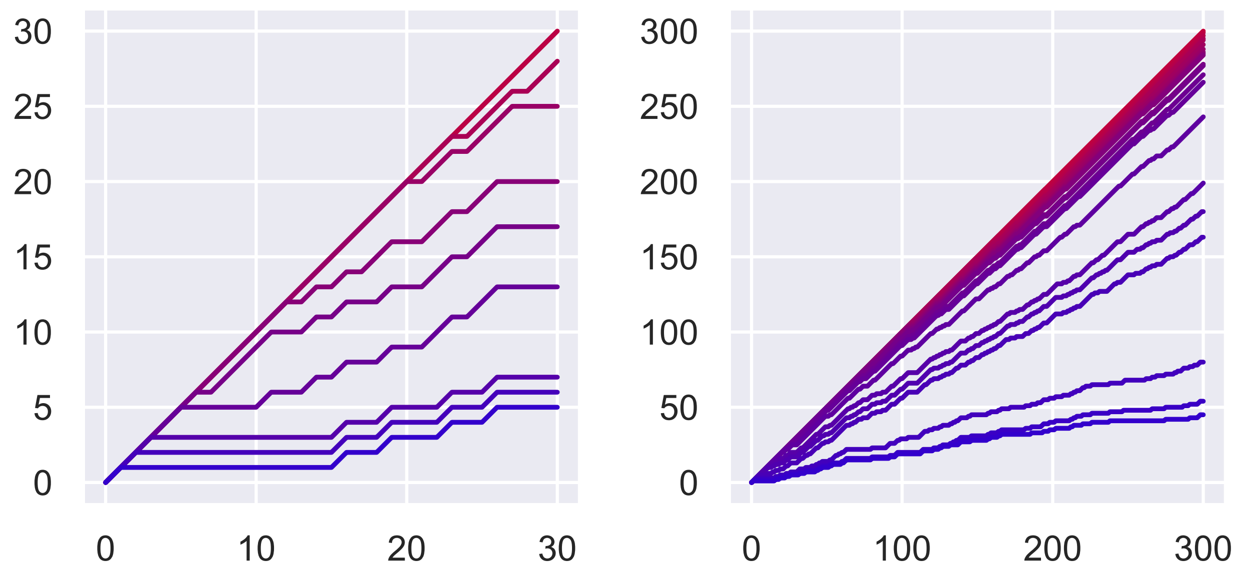

FIG. 1. Représentation graphique de deux suites cohérentes de partitions.

On peut représenter graphiquement une suite cohérente $(\pi^{(n)})_{n\geq 1}$ comme suit. Pour tout $k \geq 1$, la quantité

$$c_k(\pi^{(n)}) = \begin{cases} \text{taille de la } k\text{-ième part de } \pi^{(n)} & \text{si } \ell(\pi^{(n)}) \geq k, \\ 0 & \text{si } \ell(\pi^{(n)}) < k \end{cases}$$

est croissante avec n, et pour tout $n \geq 1$, une unique quantité $c_k(\pi^{(n+1)})$ est plus grande d'une unité que $c_k(\pi^{(n)})$ (celle telle que $n+1$ appartient à la k-ième part de $\pi^{(n+1)}$). Par conséquent, si l'on connaît toutes les suites $(c_k(\pi^{(n)}))_{n \in [\![1,N]\!]}$, alors on connaît entièrement $(\pi^{(n)})_{n \in [\![1,N]\!]}$. Une représentation assez intuitive de cette suite cohérente de partitions est donc donnée par le tracé de toutes les fonctions

$$n \in [\![1,N]\!] \mapsto c_1(\pi^{(n)}) + c_2(\pi^{(n)}) + \cdots + c_k(\pi^{(n)})$$

pour $k \leq \ell(\pi^{(n)})$. L'espace entre deux courbes consécutives représente alors la taille $c_k(\pi^{(n)})$ de la k-ième part, qui croît avec n.

On a dessiné sur la Figure 1 deux tirages de la chaîne de paramètres $\theta = 1$ et $\alpha = 0.3$: l'un jusqu'à $N = 30$, et l'autre jusqu'à $N = 300$. Sur le second graphe, il apparaît assez clairement que pour tout $k \geq 1$, la proportion $\frac{c_k(\pi^{(n)})}{n}$ d'individus dans le k-ième groupe créé tend vers une limite. L'objectif de la suite de ce texte est de comprendre ce résultat asymptotique.

Remarque 2. *Dans la suite, on ne s'intéressera pas au nombre $\ell(\pi^{(n)})$ de groupes créé au temps n. On peut néanmoins montrer que, au sens de la convergence presque sûre :*

$$\ell(\pi^{(n)}) \simeq_{n\to\infty} \begin{cases} \theta \log n & \text{si } \alpha = 0, \\ S_{\theta,\alpha}\, n^{\alpha} & \text{si } \alpha > 0, \end{cases}$$

où $S_{\theta,\alpha}$ est une certaine variable aléatoire.

2. Fréquences limites des parts de partitions échangeables

Supposons donnée une suite cohérente de partitions aléatoires $(\pi^{(n)})_{n\geq 1}$, définie sur un espace de probabilités $(\Omega, \mathscr{F}, \mathbb{P})$. On notera $\mathbb{P}_n$ la loi de $\pi^{(n)}$:

$$\forall \pi \in \mathfrak{P}(n),\ \ \mathbb{P}_n[\pi] = \mathbb{P}[\pi^{(n)} = \pi].$$

Si $(\pi_1, \pi_2, \ldots, \pi_\ell) = \pi \in \mathfrak{P}(n)$ et $\sigma \in \mathfrak{S}(n)$ est une permutation de taille n, notons $\sigma(\pi)$ la partition de taille n dont les parts sont $(\sigma(\pi_1), \sigma(\pi_2), \ldots, \sigma(\pi_\ell))$.

Définition 3. *La suite cohérente $(\pi^{(n)})_{n\geq 1}$ est dite* échangeable *si, pour tout $n \geq 1$ et toute permutation $\sigma \in \mathfrak{S}(n)$, les deux partitions $\pi^{(n)}$ et $\sigma(\pi^{(n)})$ ont la même loi.*

Théorème 4. *La chaîne de Markov de paramètres (θ, α) donne une suite $(\pi^{(n)})_{n\geq 1}$ qui est échangeable.*

Démonstration. Si $\pi, \pi' \in \mathfrak{P}(n)$, alors il existe une permutation $\sigma \in \mathfrak{S}(n)$ telle que $\sigma(\pi) = \pi'$ si et seulement si les parts de π et de π' ont les mêmes longueurs, à réordonnement près :

$$\lambda(\pi) = \lambda(\pi').$$

Une suite cohérente est donc échangeable si et seulement si $\mathbb{P}_n[\pi]$ est une fonction de la signature $\lambda(\pi)$, pour tout $n \geq 1$. Établissons ceci pour la chaîne de Markov $(\pi^{(n)})_{n\geq 1}$ de paramètres (θ, α). Il suffit de voir que si $\pi \in \mathfrak{P}(n)$ a pour liste de tailles des parts $c(\pi) = (c_1, \ldots, c_\ell)$, alors

$$\mathbb{P}_n[\pi] = \frac{\alpha^\ell}{\theta^{\uparrow n}} \left(\frac{\theta}{\alpha}\right)^{\uparrow \ell} \prod_{k=1}^{\ell} (1-\alpha)^{\uparrow c_k - 1}, \tag{1}$$

où $x^{\uparrow m} = x(x+1)\cdots(x+m-1)$ (avec par convention $x^{\uparrow 0} = 1$). En effet, la fonction ci-dessus est symétrique en $(c_1, \ldots, c_\ell)$, donc elle ne dépend en fait que du réordonnement décroissant $(\lambda_1, \ldots, \lambda_\ell) = \lambda(\pi)$. □

Dans ce qui suit, on se concentre sur la suite $(\lambda^{(n)})_{n\geq 1}$ des signatures d'une suite échangeable $(\pi^{(n)})_{n\geq 1}$: $\lambda^{(n)} = \lambda(\pi^{(n)})_{n\geq 1}$, et chaque $\lambda^{(n)}$ est un élément de l'ensemble $\mathfrak{Y}(n)$ des suites finies décroissantes d'entiers dont la somme est égale à n. On note

$$\mathscr{F}_{-n} = \sigma(\lambda^{(n)}, \lambda^{(n+1)}, \lambda^{(n+2)}, \ldots)$$

la tribu engendrée par les signatures des partitions de taille plus grande que n dans la suite échangeable. Par construction, chaque $\mathscr{F}_{-n}$ est une sous-tribu de $\mathscr{F}$, et si $l \leq m$ sont deux entiers négatifs, alors $\mathscr{F}_l \subset \mathscr{F}_m$. Autrement dit, $(\mathscr{F}_l)_{l\leq -1}$ est une *filtration rétrograde* : c'est la même définition que pour une filtration d'un espace de probabilité, à ceci près que les indices des sous-tribus sont des entiers négatifs. Dans ce qui suit, on considérera également des *chaînes de Markov rétrogrades* indicées par les entiers négatifs : ce sont des suites de variables aléatoires $(Y_l)_{l\leq -1}$ à valeurs dans un ensemble dénombrable $\mathfrak{Y}$, telles qu'il existe une matrice stochastique Q sur $\mathfrak{Y}$ avec

$$\mathbb{P}[Y_{l+1} = y \,|\, Y_l, Y_{l-1}, \ldots] = Q(Y_l, y)$$

pour tout $l \geq -2$. L'équation ci-dessus est la même formule que pour une chaîne de Markov standard, mais avec des indices négatifs pour les variables.

Proposition 5. *Fixons $k \geq 1$, et notons*

$$V_{-n,k} = \frac{\lambda_1^{(n)} + \lambda_2^{(n)} + \cdots + \lambda_k^{(n)}}{n},$$

avec par convention $\pi_j^{(n)} = \emptyset$ et $\lambda_j^{(n)} = 0$ si $j > \ell(\pi^{(n)})$. La variable $V_{-n,k}$ est la proportion d'entiers de $[\![1,n]\!]$ qui appartiennent aux k plus grandes parts de $\pi^{(n)}$. La suite $(V_{l,k})_{l\leq -1}$ est une sous-martingale rétrograde pour la filtration $(\mathscr{F}_l)_{l\leq -1}$:

$$\forall l \leq -1,\ \ \mathbb{E}[V_{l,k} \,|\, \mathscr{F}_{l-1}] \geq V_{l-1,k}.$$

Pour démontrer ceci, introduisons la notion d'*effacement aléatoire*. Si Π est une partition de taille $n+1$, son effacement aléatoire est la partition $\pi \in \mathfrak{P}(n)$ obtenue :

- en choisissant aléatoirement $X_{n+1} \in [\![1, n+1]\!]$ de loi uniforme, indépendamment de Π ;
- en retirant X_{n+1} de la part de Π correspondante (supprimant la part si c'était $\{X_{n+1}\}$) ;
- et en utilisant l'unique bijection croissante $[\![1, n+1]\!] \setminus \{X_{n+1}\} \to [\![1,n]\!]$ pour transformer la liste de parts restantes en une partition de $[\![1,n]\!]$.

Notons cette construction $\pi = E(\Pi)$. Par exemple, si $\Pi = (\{1,3,7\}, \{2,10\}, \{4,5,8,9\}, \{6\})$ et si l'on tire $X_{10} = 5$, alors $\pi = (\{1,3,6\}, \{2,9\}, \{4,7,8\}, \{5\})$.

Lemme 6. *Si $(\pi^{(n)})_{n\geq 1}$ est une suite échangeable de partitions, alors $(\lambda^{(-l)})_{l\leq -1}$ est une chaîne de Markov rétrograde sur $\mathfrak{Y} = \bigsqcup_{n=1}^{\infty} \mathfrak{Y}(n)$. Son noyau de transition est*

$$Q(\Lambda, \mu) = \mathbb{P}[\lambda(E(\Pi)) = \mu] \tag{2}$$

où Π est une partition arbitraire de signature Λ.

Esquisse de preuve. On montre d'abord que conditionnellement à $\mathscr{F}_{-n}$, la loi de $\pi^{(n)}$ est uniforme sur l'ensemble des partitions de $[\![1,n]\!]$ dont la signature est $\lambda^{(n)}$. Autrement dit, si π a signature Λ, alors

$$\mathbb{P}[\pi^{(n)} = \pi \,|\, \mathscr{F}_{-n}] = \frac{1_{(\lambda^{(n)}=\Lambda)}}{B_{n,\Lambda}}, \tag{3}$$

où $B_{n,\Lambda}$ est le nombre de partitions de $[\![1,n]\!]$ avec signature Λ. Ce nombre est donné par la formule :

$$B_{n,\Lambda} = \frac{n!}{\prod_{i=1}^{s} (i!)^{m_i}(m_i!)} \tag{4}$$

si Λ est une signature avec m_1 entrées égales à 1, m_2 entrées égales à 2, etc. de sorte que $n = \sum_{i=1}^{s} i\,m_i$. Ces nombres sont également en jeu dans le noyau de transition $Q(\lambda,\mu)$: si $\Lambda = (s^{m_s},(s-1)^{m_{s-1}},\dots,2^{m_2},1^{m_1})$ est la signature d'une partition Π et si

$$\mu = (s^{m_s},\dots,r^{m_r-1},(r-1)^{m_{r-1}+1},\dots,1^{m_1})$$

est la signature de $\pi = E(\Pi)$, alors

$$Q(\Lambda,\mu) = \frac{(m_{r-1}+1)\,B_{n-1,\mu}}{B_{n,\Lambda}}. \tag{5}$$

Le lemme s'en déduit, en calculant d'abord la loi conditionnelle sachant $\mathscr{F}_{-n}$ de $\pi^{(n-1)}$, puis celle de $\lambda^{(n-1)}$. □

Preuve de la Proposition 5. Notons X_{n+1} une variable uniforme sur $[\![1,n+1]\!]$ et indépendante de $\mathscr{G} = \sigma(\pi^{(n+1)},\mathscr{F}_{-(n+1)})$. Pour tous indices $\rho(1)\neq\rho(2)\neq\cdots\neq\rho(k)$,

$$\left(|\pi^{(n+1)}_{\rho(1)}|+\cdots+|\pi^{(n+1)}_{\rho(k)}|\right) - 1_{\left(X_{n+1}\in\pi^{(n+1)}_{\rho(1)}\sqcup\cdots\sqcup\pi^{(n+1)}_{\rho(k)}\right)} = |E(\pi^{(n+1)})_{\rho(1)}|+\cdots+|E(\pi^{(n+1)})_{\rho(k)}|. \tag{6}$$

On choisit la permutation ρ de sorte que les parts $\pi^{(n+1)}_{\rho(1)},\dots,\pi^{(n+1)}_{\rho(k)}$ soient les k plus grandes :

$$|\pi^{(n+1)}_{\rho(1)}|+\cdots+|\pi^{(n+1)}_{\rho(k)}| = \lambda^{(n+1)}_1+\cdots+\lambda^{(n+1)}_k = (n+1)\,V_{-(n+1),k}.$$

Il suffit de connaître $\pi^{(n+1)}$ pour choisir ρ, donc on peut supposer que ρ est $\mathscr{G}$-mesurable, et en particulier indépendante de X_{n+1}. Par conséquent,

$$X_{n+1}\quad\text{et}\quad(\pi^{(n+1)}_{\rho(1)},\dots,\pi^{(n+1)}_{\rho(k)})\quad\text{sont indépendants.}$$

En prenant l'espérance conditionnelle de (6) sachant $\mathscr{F}_{-(n+1)}$, on obtient donc :

$$n\,V_{-(n+1),k} = \mathbb{E}[(|E(\pi^{(n+1)})_{\rho(1)}|+\cdots+|E(\pi^{(n+1)})_{\rho(k)}|)\,|\,\mathscr{F}_{-(n+1)}] \tag{7}$$

$$\leq \mathbb{E}[\lambda_1(E(\pi^{(n+1)}))+\cdots+\lambda_k(E(\pi^{(n+1)}))\,|\,\mathscr{F}_{-(n+1)}] = \mathbb{E}[n\,V_{-n,k}\,|\,\mathscr{F}_{-(n+1)}]. \tag{8}$$

en utilisant le Lemme 6 pour établir l'égalité sur la seconde ligne. □

L'inégalité du nombre de montées est valable pour les sous-martingales rétrogrades, avec la même preuve que dans le cas non rétrograde. Comme $(V_{l,k})_{l\leq -1}$ est une suite bornée entre 0 et 1, elle n'a donc presque sûrement qu'un nombre fini de montées entre deux niveaux a et b, pour tous niveaux $a<b$. Ceci implique comme dans le cas standard la convergence presque sûre :

$$\lim_{n\to\infty}\frac{V_{-n,k}}{n} = V_k$$

pour une certaine variable aléatoire $V_k\in[0,1]$. En posant $Y_k = V_k - V_{k-1}$, on obtient donc le résultat annoncé dans l'introduction :

Théorème 7. *Si $(\lambda^{(n)})_{n\geq 1}$ est la suite des signatures d'une suite échangeable de partitions, alors pour tout $k\geq 1$,*

$$\frac{\lambda^{(n)}_k}{n}\to_{\text{p.s.}} Y_k$$

pour certaines variables aléatoires Y_k avec $\sum_{k=1}^{\infty} Y_k\leq 1$.

En particulier, le théorème s'applique aux chaînes de Markov introduites par la Définition 1. Si l'on revient à la Figure 1, notons qu'on avait conjecturé la convergence presque sûre de $\frac{c_k(\pi^n)}{n}$, ce qui diffère de la convergence après réordonnement décroissant des fréquences $\frac{\lambda_k(\pi^{(n)})}{n}$. Le résultat suivant, dont la démonstration est simple mais assez longue, décrit les limites des fréquences non réordonnées :

Théorème 8. *Notons $(W_k)_{k\geq 1}$ une suite de variables aléatoires indépendantes, avec chaque W_k qui suit une loi beta de paramètres $(1-\alpha, \theta+k\alpha)$:*

$$W_k \sim 1_{(x\in[0,1])}\,\frac{\Gamma(1+\theta+(k-1)\alpha)}{\Gamma(1-\alpha)\,\Gamma(\theta+k\alpha)}\,x^{-\alpha}\,(1-x)^{\theta+k\alpha-1}\,\mathrm{d}x\,.$$

On pose $X_k = (1-W_1)(1-W_2)\cdots(1-W_{k-1})W_k$, et on introduit également une suite $(U^{(n)})_{n\geq 1}$ de variables uniformes sur $[0,1]$ et indépendantes entre elles et de $(W_k)_{k\geq 1}$.

(1) Définissons une suite cohérente $(\pi^{(n)})_{n\geq 1}$ de partitions aléatoires par la récurrence :

$$\pi^{(n+1)} = \begin{cases} \pi^{(n)}[k] & \text{si } X_1+\cdots+X_{k-1} \leq U^{(n)} < X_1+\cdots+X_k, \\ \pi^{(n)}[\ell(\pi^{(n)})+1] & \text{si } U^{(n)} \geq X_1+\cdots+X_{\ell(\pi^{(n)})}. \end{cases}$$

Au sens de la Définition 1, $(\pi^{(n)})_{n\geq 1}$ est une chaîne de Markov de paramètres (θ,α).

(2) Cette chaîne vérifie $\frac{c_k(\pi^{(n)})}{n} \to_{\text{p.s.}} X_k$.

(3) La suite $(Y_k)_{k\geq 1}$ est donc obtenue par réordonnement décroissant de la suite $(X_k)_{k\geq 1}$.

Questions

(1) Vérifier que le noyau de transition de la Définition 1 est bien une matrice stochastique : pour toute partition $\pi \in \mathfrak{P}(n)$, si ℓ est le nombre de parts de π, alors $\sum_{k=1}^{\ell+1} P(\pi,\pi[k]) = 1$.

(2) Écrire un programme `markov_partition(N, theta, alpha)` qui simule le N-ième état $\pi^{(N)}$ de la chaîne de Markov de paramètres (θ,α). En Python, on pourra représenter une partition $\pi \in \mathfrak{P}(N)$ par une liste de listes π_k dont la somme des tailles vaut N. Expliquer pourquoi $\pi^{(N)}$ détermine entièrement la suite $(\pi^{(n)})_{n\in[1,N]}$.

(3) Écrire un autre programme `dessin_partition(pi)` qui prend en argument une partition $\pi = \pi^{(N)}$ de taille N, et qui dessine la représentation graphique de la suite cohérente $(\pi^{(n)})_{n\in[1,N]}$ correspondante, comme sur la Figure 1. Tester ce programme avec différentes valeurs pour les paramètres (θ,α).

(4) Illustrer par des programmes le résultat annoncé dans la Remarque 2.

(5) On souhaite donner une preuve de la convergence $\frac{\ell(\pi^{(n)})}{\log n} \to_{\text{p.s.}} \theta$ lorsque $\alpha = 0$.

- Montrer que si $\alpha = 0$, alors $\ell(\pi^{(n)})$ a la loi d'une somme de variables de Bernoulli indépendantes $\sum_{k=0}^{n-1} B_k$, avec $\mathbb{P}[B_k = 1] = 1 - \mathbb{P}[B_k = 0] = \frac{\theta}{\theta+k}$.
- On considère des variables de Poisson indépendantes P_k, avec $P_k \sim \text{Poisson}(\frac{\theta}{\theta+k})$. Posons $B_k = 1_{(P_k\geq 1)}$. En utilisant le lemme de Borel–Cantelli, montrer que la suite

$$\left(\sum_{k=0}^{n-1} P_k - \sum_{k=0}^{n-1} B_k\right)_{n\geq 1}$$

reste bornée presque sûrement.
- En déduire le résultat souhaité, en considérant un processus de Poisson $(\mathcal{P}_t)_{t\geq 0}$ d'intensité 1, qu'on regarde aux temps $t_n = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k}$.

(6) Montrer que deux partitions π et π' de taille n sont reliées par une permutation $\sigma \in \mathfrak{S}(n)$ si et seulement si $\lambda(\pi) = \lambda(\pi')$.

(7) Établir la Formule (1) pour la loi du n-ième état $\pi^{(n)}$ d'une chaîne de paramètres (θ, α). On pourra procéder par récurrence sur n, en séparant le cas où $n+1$ crée son propre groupe dans $\pi^{(n+1)}$, et le cas où $n+1$ rejoint une part de $\pi^{(n)}$.

(8) On souhaite compléter la preuve du Lemme 6.

- Remarquons que la tribu $\mathscr{F}_{-n}$ est engendrée par les événements $A = A(\mu^{(n)}, \dots, \mu^{(n+k)})$ du type :
$$\{\lambda(\pi^{(n)}) = \mu^{(n)}, \lambda(\pi^{(n+1)}) = \mu^{(n+1)}, \dots, \lambda(\pi^{(n+k)}) = \mu^{(n+k)}\},$$
où $\mu^{(n)}, \dots, \mu^{(n+k)}$ sont des signatures fixées de tailles $n, \dots, n+k$. En revenant alors à la définition de l'espérance conditionnelle, montrer que pour tout $n \geq 1$ et toute partition π de signature Λ, l'Équation (3) est vérifiée.
- Établir la formule (4). On pourra construire une surjection de l'ensemble des permutations $\mathfrak{S}(n)$ vers l'ensemble des partitions de $[\![1, n]\!]$ dont la signature est Λ.
- Montrer que le noyau $Q(\Lambda, \mu)$ défini par l'Équation (2) est donné par la formule explicite (5).
- Déterminer les lois conditionnelles de $\pi^{(n-1)}$ et de $\lambda^{(n-1)}$ sachant $\mathscr{F}_{-n}$, et conclure.

(9) Expliquer pourquoi l'espérance conditionnelle du terme de gauche de l'Équation (7) vaut $n\, V_{-(n+1),k}$, et pourquoi le Lemme 6 implique l'égalité dans l'Équation (8).

(10) Avec les paramètres $\theta = 1$ et $\alpha = 0.5$, vérifier expérimentalement que la loi de la limite Y_1 de la proportion d'individus dans le plus grand groupe est la loi du maximum de la suite $(X_k)_{k \geq 1}$ définie par le Théorème 8. On pourra utiliser `scipy.stats.beta(a, b).rvs()` pour engendrer une variable de loi beta de paramètres (a, b).