PROCESSUS DES RESTAURANTS CHINOIS ET LOI D'EWENS

DJALIL CHAFAÏ, YAN DOUMERC, ET FLORENT MALRIEU

Résumé. On étudie une suite aléatoire à valeurs dans les permutations d'ensembles finis, appelée *processus des restaurants chinois*. Ce processus est relié à la loi d'Ewens bien connue en combinatoire élémentaire. Ce processus et cette loi constituent en quelque sorte un analogue pour les permutations du processus de Poisson et de la loi de Poisson, plus classiques en théorie des probabilités.

On s'intéresse à la disposition de clients (numérotés $1, 2$, etc) venant successivement s'asseoir dans un restaurant chinois imaginaire autour de tables circulaires (numérotées $1, 2$, etc) et toutes de capacité infinie. On note $\mathfrak{S}_n$ l'ensemble des permutations de $\{1, \ldots, n\}$ et Π_n l'ensemble des partitions de $\{1, \ldots, n\}$. Avant de décrire précisément comment s'obtient la disposition D_n des n premiers clients autour des tables, on associe à D_n une *permutation* $\sigma_n \in \mathfrak{S}_n$ ainsi qu'une *partition* $p_n \in \Pi_n$, de la façon suivante :
- la décomposition en cycles de σ_n est obtenue par lecture dans le sens horaire des numéros des clients de D_n à chaque table ;
- les blocs de p_n sont les supports des cycles de σ_n i.e. les clients de chaque table de D_n.

Par exemple, si D_6 est comme dans la figure 1 alors

$$\sigma_6 = (1,6,3)(2,4)(5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 1 & 2 & 5 & 3 \end{pmatrix} \quad \text{et} \quad p_6 = \{\{1,6,3\},\{2,4\},\{5\}\}.$$

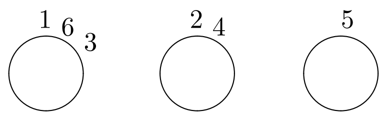
$D_6 =$ 

Figure 1. Exemple de configuration D_6 du restaurant avec 6 clients répartis en 3 tables.

La permutation σ_n encode toute l'information contenue dans D_n. En revanche, la partition p_n s'obtient à partir de la permutation σ_n en considérant le support des cycles, et contient donc moins d'information car la disposition précise des clients sur chaque table est perdue. Décrivons maintenant l'évolution récursive de la suite (D_n). Le premier client s'assoit à la première table. Supposons assis les n premiers clients aux tables $1, \ldots, K_n$, où K_n désigne le nombre de tables occupées par les n premiers clients. Le client $n+1$ a le choix entre s'asseoir entre deux clients quelconques déja attablés ou bien s'asseoir à la nouvelle table de numéro $K_n + 1$. Il pourrait effectuer ce choix avec équiprobabilité. Par souci de généralité, on attache un poids 1 à chaque place entre deux clients, un poids θ à la nouvelle table et le client $n+1$ choisit une place avec une probabilité proportionnelle à son poids. Comme il y a n places disponibles entre tous les clients déjà assis, la probabilité que le client $n+1$ s'asseoit à l'une de ces places est

$$\frac{1}{\theta + n}$$

tandis que la probabilité de s'asseoir à une nouvelle table est

$$\frac{\theta}{\theta + n}.$$

Le choix équiprobable correspond donc au cas particulier $\theta = 1$. Ainsi, nous avons défini une suite (D_n) à laquelle correspondent des suites (σ_n) et (p_n) toutes aléatoires. Dans une première partie, nous examinerons avec précision le mécanisme d'évolution des suites (σ_n), (p_n) et d'autres suites associées. Dans une deuxième partie, nous en expliciterons les lois à n fixé, mettant en évidence la famile des lois

Date: version initiale: 10 mai 2012, révision: 13 septembre 2012.
Manuscrit soumis à RMS – La Revue de la Filière Mathématique.

d'Ewens paramétrées par $\theta \geq 0$ et dont l'équiprobabilité sur $\mathfrak{S}_n$ correspond au cas $\theta = 1$. Enfin, nous regarderons certaines propriétés asymptotiques de la décomposition en cycles : une loi des grands nombres et un théorème central-limite seront donnés pour le nombre de cycles ainsi qu'un remarquable résultat de convergence en loi de la taille des cycles.

Notons que ce mécanisme d'évolution provoque l'apparition de petites tables, et celles-ci peuvent subsister au cours du temps (tout cela dépend du paramètre θ).

1. Évolution markovienne

Une suite $(X_n)_{n\geq 0}$ aléatoire (à valeurs dans un espace au plus dénombrable) est dite markovienne lorsqu'elle vérifie, à partir d'une donnée initiale X_0, une relation de récurrence d'ordre 1 du type

$$X_{n+1} = f(X_n, \epsilon_{n+1})$$

pour tout $n \geq 0$, où ϵ_{n+1} est indépendante de $\epsilon_1, \ldots, \epsilon_n$ et de X_0. En interprétant n comme un temps (discret), ceci signifie que pour décider de son état futur X_{n+1} la suite n'a besoin que de connaître son présent X_n et d'effectuer un choix aléatoire ϵ_{n+1} indépendant de son passé. Ainsi, l'évolution de (X_n) est déterminée par son état initial X_0 et les probabilités de transition $P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x)$ i.e. les probabilités de sauter de l'état x à l'état y à l'instant n. Les suites markoviennes constituent des processus d'évolution stochastiques « sans mémoire ». Il s'agit d'un analogue probabiliste des suites récurrentes ou des équations différentielles.

1.1. Les permutations σ_n.

La suite (D_n) est markovienne, et ϵ_{n+1} correspond au choix de la place par le client $n + 1$. L'application $\phi : D_n \to \sigma_n$ étant bijective, l'évolution de (σ_n) est aussi markovienne (prendre $f'(\sigma, \epsilon) = \phi(f(\phi^{-1}(\sigma), \epsilon))$). Elle se caractérise donc entièrement par la donnée de son état initial $(\sigma_1 = (1))$ et de ses probabilités de transition :

$$\mathbb{P}(\sigma_{n+1} = \sigma' \,|\, \sigma_n = \sigma) = \begin{cases} \dfrac{1}{\theta + n} & \text{si } \sigma' \text{ s'obtient en insérant } n+1 \text{ dans l'un des cycles de } \sigma, \\[2ex] \dfrac{\theta}{\theta + n} & \text{si } \sigma' \text{ s'obtient en ajoutant le cycle } (n+1) \text{ à ceux de } \sigma, \\[2ex] 0 & \text{sinon.} \end{cases}$$

Remarque 1.1 (Remonter le temps)**.** *Voici un exemple de début de trajectoire de la suite aléatoire $(\sigma_n)_{n\geq 1}$, qui mène au σ_6 de l'introduction :*

$$\sigma_1 = (1),\ \sigma_2 = (1)(2),\ \sigma_3 = (1,3)(2),\ \sigma_4 = (1,3)(2,4),\ \sigma_5 = (1,3)(2,4)(5),\ \sigma_6 = (1,6,3)(2,4)(5).$$

On remarque immédiatement qu'il est possible de remonter le temps : σ_n se déduit de σ_{n+1} en enlevant $n + 1$ du cycle où il se trouve quitte à supprimer ce cycle s'il est d'ordre 1. Plus généralement, pour tout $n \geq 1$ et tout $\sigma' \in \mathfrak{S}_{n+1}$, il existe un unique $\sigma \in \mathfrak{S}_n$ tel que $\mathbb{P}(\sigma_{n+1} = \sigma' \,|\, \sigma_n = \sigma) > 0$. Il s'agit là d'une propriété commune à tous les processus dont le présent transporte intégralement le passé (ce n'est évidemment pas le cas de toutes les suite markoviennes). Il est amusant d'observer que le mécanisme d'évolution markovien sans mémoire n'empêche pas la conservation intégrale du passé.

Remarque 1.2 (Numérotation des tables)**.** *Les cycles sont numérotés par ordre croissant du plus petit élément non encore utilisé dans les cycles précédents.*

1.2. Les partitions p_n.

Si ψ est l'application qui à une permutation associe la partition donnée par sa décomposition en cycles alors $p_n = \psi(\sigma_n)$. Mais ψ n'étant pas bijective, le caractère markovien de (σ_n) ne se transmet pas a priori à (p_n) : pour déterminer p_{n+1}, la connaissance de σ_n est suffisante mais celle de p_n pourrait ne pas l'être car p_n contient moins d'information que σ_n. En l'espèce, il se trouve malgré tout que $p_{n+1} = \psi(\sigma_{n+1}) = \psi(f(\sigma_n, \epsilon_{n+1})) = g(p_n, \epsilon'_{n+1})$ pour une fonction g bien choisie. Ainsi, la suite

(p_n) est bien markovienne [1], et ses probabilités de transition sont données par

$$\mathbb{P}(p_{n+1} = p' \,|\, p_n = p) = \begin{cases} \dfrac{|b|}{\theta+n} & \text{si } p' \text{ s'obtient en ajoutant } n+1 \text{ au bloc } b \text{ de } p, \\[2ex] \dfrac{\theta}{\theta+n} & \text{si } p' \text{ s'obtient en ajoutant le singleton } \{n+1\} \text{ à } \ p, \\[2ex] 0 & \text{sinon.} \end{cases}$$

Remarque 1.3 (Monotonie en θ)**.** *Plus θ est grand, plus les clients ont tendance à s'asseoir à une table vide plutôt que de rejoindre une table occupée. À l'extrême, si $\theta = 0$, on a $p_n = \{\{1, \dots, n\}\}$ tandis que si $\theta = \infty$ alors $p_n = \{\{1\}, \dots, \{n\}\}$. Bien sûr, cette monotonie en θ est seulement probabiliste : un p_n obtenu avec θ peut avoir plus de blocs qu'un p'_n obtenu avec θ' même si $\theta' > \theta$ de la même manière que je peux obtenir face et mon voisin pile même si nos pièces offrent respectivement des probabilités 0,9 et 0,1 de pile ! On parle dans ce cas d'*ordre stochastique*, cf.* [SS].

1.3. **Les effectifs A_n.** Définissons le vecteur $A_n = (A_{n,1}, \dots, \dots, A_{n,n})$ des effectifs de p_n par

$$A_{n,i} = \text{nombre de tables (ou de cycles, de blocs) de cardinal } i \text{ de } D_n \text{ (resp. de } \sigma_n\text{, de } p_n\text{)}$$

de telle sorte que

$$\sum_{i=1}^{n} i A_{n,i} = n.$$

Comme pour (p_n), on a $A_n = \psi(p_n)$ où cette fois $\psi : p \mapsto$ « vecteur des effectifs de p ». Bien que ψ ne soit pas bijective, (A_n) est à nouveau markovienne. De plus, si $a \in \mathbb{N}^n$, $a' \in \mathbb{N}^{n+1}$ et (e_i) est la base canonique de $\mathbb{R}^{n+1}$, les transitions de (A_n) sont données par :

$$\mathbb{P}(A_{n+1} = a' \,|\, A_n = a) = \begin{cases} \dfrac{i a_i}{\theta+n} & \text{si } a' = (a,0) + e_{i+1} - e_i \text{ avec } 1 \le i \le n, \\[2ex] \dfrac{\theta}{\theta+n} & \text{si } a' = (a,0) + e_1, \\[2ex] 0 & \text{sinon.} \end{cases}$$

1.4. **Le nombre de tables K_n.** Notons K_n le nombre de tables occupées de D_n si bien que

$$K_n = \text{ nombre de cycles de } \sigma_n \ = \text{ nombre de blocs de } p_n \ = \sum_{i=1}^{n} A_{i,n}.$$

La suite (K_n) est aléatoire et $K_{n+1} - K_n$ prend ses valeurs dans $\{0, 1\}$. Comme au paragraphe précédent, $K_n = \psi(p_n)$ où $\psi : p \mapsto |p|$ n'est pas bijective mais ici encore, la suite (K_n) est tout de même markovienne de transitions données par

$$\mathbb{P}(K_{n+1} = K_n + 1) = \frac{\theta}{\theta+n}, \quad \mathbb{P}(K_{n+1} = K_n) = 1 - \frac{\theta}{\theta+n}.$$

Une autre façon de décrire cette évolution est d'introduire

$$\xi_n = \begin{cases} 1 & \text{si le client } n+1 \text{ s'asseoit à une nouvelle table,} \\ 0 & \text{sinon.} \end{cases}$$

Alors on a $K_n = \xi_1 + \dots + \xi_n$ et les (ξ_n) sont indépendantes et de lois données par

$$\mathbb{P}(\xi_n = 1) = \frac{\theta}{\theta+n-1}, \quad \mathbb{P}(\xi_n = 0) = 1 - \frac{\theta}{\theta+n-1}. \tag{1.1}$$

1. Il s'agit là d'une application du critère de Dynkin, qui assure que certaines images de suites markoviennes sont encore markoviennes, cf. [BMP, p. 121].

1.5. **Les solitaires.** La variable aléatoire $S_n = A_{n,1}$ est le nombre de clients solitaires de D_n, i.e. le nombre de points fixes de σ_n. La suite $(A_{n,1})$ est markovienne de transition

$$\mathbb{P}(S_{n+1} = s' \,|\, S_n = s) = \begin{cases} \dfrac{\theta}{\theta+n} & \text{si } s' = s+1 \text{ (s'attabler seul à une table vide),} \\[2ex] \dfrac{s}{\theta+n} & \text{si } s' = s-1 \text{ (rejoindre la table d'un solitaire),} \\[2ex] \dfrac{n-s}{\theta+n} & \text{si } s' = s \text{ (rejoindre une table comptant 2 clients ou plus)} \\[2ex] 0 & \text{sinon.} \end{cases}$$

1.6. **Table unique.** La variable aléatoire $A_{n,n}$ vaut 1 ou 0 selon qu'il y a une seule ou plusieurs tables. La suite $(A_{n,n})$ est décroissante et markovienne de transitions

$$\mathbb{P}(A_{n+1,n+1} = 0 \,|\, A_{n,n} = 0) = 1, \quad \mathbb{P}(A_{n+1,n+1} = 0 \,|\, A_{n,n} = 1) = \frac{\theta}{\theta+n}.$$

2. Lois à n fixé

Dans toute la suite, nous utiliserons la notation

$$x_{(n)} = x(x+1)(x+2)\cdots(x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)}, \quad x_{(0)} = 1.$$

2.1. **Les permutations σ_n.** On dispose du théorème suivant concernant les permutations σ_n.

Théorème 2.1 (Loi d'Ewens sur $\mathfrak{S}_n$)**.** *Pour tout $\sigma \in \mathfrak{S}_n$, on a*

$$\mathbb{P}(\sigma_n = \sigma) = \frac{\theta^{K(\sigma)}}{\theta_{(n)}} \quad \text{où } K(\sigma) \text{ est le nombre de cycles de } \sigma.$$

Démonstration. La formule est vraie pour $n = 1$. Supposons-la vraie pour n. Soit $\sigma' \in \mathfrak{S}_{n+1}$. Il existe un unique $\sigma \in \mathfrak{S}_n$ tel que $\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) > 0$. Si $K(\sigma') = K(\sigma) + 1$ alors $\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) = \theta/(\theta+n)$ et si $K(\sigma') = K(\sigma)$ alors $\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) = 1/(\theta+n)$. Dans les deux cas, on a bien

$$\mathbb{P}(\sigma_{n+1} = \sigma') = \mathbb{P}(\sigma_n = \sigma)\,\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) = \frac{\theta^{K(\sigma)}}{\theta_{(n)}},$$

ce qui conclut la récurrence. □

Remarque 2.2 (Simulation de la loi uniforme par algorithme de Fisher-Yates-Knuth [Kh1])**.** *Le cas particulier $\theta = 1$ correspond à la loi uniforme sur $\mathfrak{S}_n$, c'est-à-dire la loi accordant une probabilité $1/n!$ à chaque permutation. Ce processus fournit donc un algorithme pour simuler la loi uniforme sur $\mathfrak{S}_n$. Cet algorithme revient en fait à multiplier à droite par des transpositions. En effet, rappelons que si $\sigma \in \mathfrak{S}_n$, si $\tau = (i,j)$ désigne la transposition (i,j) avec $1 \leq i,j \leq n$, et si c_i et c_j sont les cycles de σ contenant i et j respectivement, alors la décomposition en cycles de la permutation $\sigma\tau$ s'obtient à partir de celle de σ en fusionnant c_i et c_j en i et j si $c_i \neq c_j$ ou bien en scindant en deux c_i si $c_i = c_j$. L'algorithme de simulation de la loi uniforme sur $\mathfrak{S}_n$ consiste à présent à calculer le produit de transpositions*

$$(1U_1)\cdots(nU_n)$$

où $U_1, \ldots, U_n$ sont indépendantes, avec U_k de loi uniforme sur $\{1, \ldots, k\}$ pour tout $1 \leq k \leq n$ (notons que $(1U_1)$ peut être omis car $\mathbb{P}(U_1 = 1) = 1$). Voici l'implémentation de l'algorithme en pseudo-code :

```
for k from 1 to n; do v[k] := k; endfor
for k from n downto 2; do i := rand(1,k); swap(v[i],v[k]); endfor
```

Cet algorithme remarquablement simple est connu sous le nom de Fisher-Yates shuffle *ou* Knuth shuffle*. Sa complexité est linéaire en n. L'algorithme alternatif consistant à trier n nombres réels indépendants et équidistribués a une complexité de l'ordre de $n\ln(n)$, cf.* [MR]*. D'autre part, l'algorithme alternatif consistant à effectuer n tirages sans remise dans $\{1, \ldots, n\}$ est moins commode à implémenter efficacement car il faut tenir compte des éléments déjà tirés.*

2.2. **Les partitions p_n.** Cette section est consacrée à l'étude des partitions p_n.

Théorème 2.3 (Loi d'Ewens sur Π_n)**.** *Soit $p \in \Pi_n$. En notant $|p|$ le nombre de blocs de p et $b \in p$ le fait que b soit un bloc de p, on a*

$$\mathbb{P}(p_n = p) = \frac{\theta^{|p|}}{\theta_{(n)}} \prod_{b \in p} (|b| - 1)!.$$

Démonstration. Soit E l'ensemble des éléments $\sigma \in \mathfrak{S}_n$ dont la décomposition en cycles correspond à p. Il y a $(k-1)!$ cycles d'ordre k d'un ensemble à k éléments donc $|E| = \prod_{b \in p}(|b|-1)!$. Ensuite

$$\mathbb{P}(\sigma_n = \sigma) = \frac{\theta^{|p|}}{\theta_{(n)}}$$

pour tout $\sigma \in E$ donc

$$\mathbb{P}(p_n = p) = \sum_{\sigma \in E} \mathbb{P}(\sigma_n = \sigma) = \frac{\theta^{|p|}}{\theta_{(n)}} |E|,$$

ce qui est la réponse. □

Remarque 2.4 (Loi d'Ewens et loi uniforme sur Π_n)**.** *Quelle que soit la valeur de θ, la loi d'Ewens sur Π_n n'est jamais la loi uniforme pour $n \geq 3$. En effet, notons*

$$p = \{\{1\}, \{2\}, \ldots, \{n\}\}, \quad q = \{\{1,2\}, \{3\}, \ldots, \{n\}\}, \quad \textit{et} \quad r = \{\{1, 2, \ldots, n\}\}.$$

Alors

$$\mathbb{P}(p_n = p) = \frac{\theta^n}{\theta_{(n)}}, \quad \mathbb{P}(p_n = q) = \frac{\theta^{n-1}}{\theta_{(n)}}, \quad \textit{et} \quad \mathbb{P}(p_n = r) = \frac{\theta\,(n-1)!}{\theta_{(n)}}.$$

Ainsi, $\mathbb{P}(p_n = p)$ ne vaut $\mathbb{P}(p_n = q)$ que pour $\theta = 1$ et dans ce cas $\mathbb{P}(p_n = p) \neq \mathbb{P}(p_n = r)$.

Remarque 2.5 (Conjugaison)**.** *Le théorème 2.1 montre que la loi de σ_n ne dépend que du nombre de cycles de σ_n donc qu'elle est invariante par conjugaison. Ceci assure que, conditionnellement à $p_n = p$, la loi de σ_n est uniforme sur l'ensemble des permutations dont la décomposition en cycles correspond à p, ce qui correspond à l'évidence intuitive d'invariance par permutation des clients à chaque table.*

2.3. **Loi uniforme sur les partitions.** La remarque 2.4 nous incite à étudier la loi uniforme sur Π_n, qui accorde à chaque partition une probabilité $1/B_n$ où $B_n = |\Pi_n|$ est le n-ième nombre de Bell. On a $B_0 = 1$ par convention, $B_1 = 1$ et $B_2 = 2$, et

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k, \tag{2.1}$$

formule qui se démontre en examinant les k éléments qui n'appartiennent pas au même bloc que $n+1$. En introduisant la série formelle $G(X) := \sum_{n=0}^{\infty} \frac{B_n}{n!} X^n$, un produit de Cauchy montre que $G'(X) = \exp(X)G(X)$, ce qui donne $G(X) = \exp(\exp(X) - 1)$. On reconnaît la transformée de Laplace de la loi de Poisson de paramètre 1. Ainsi, les nombres de Bell sont les moments de cette loi de Poisson, ce qui donne la formule de Dobinski :

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}. \tag{2.2}$$

Notons par ailleurs que

$$B_n = \sum_{k=1}^{n} \left\{ n \atop k \right\}$$

où la notation entre accolades désigne le nombre de partitions à k blocs d'un ensemble à n éléments (notation de Knuth pour le nombre de Stirling de seconde espèce). On dispose de la formule de récurrence

$$\left\{ n \atop k \right\} = \left\{ n-1 \atop k-1 \right\} + k \left\{ n-1 \atop k \right\} \quad \text{avec conditions au bord} \quad \left\{ n \atop 1 \right\} = 1 \quad \text{et} \quad \left\{ n \atop n \right\} = 1$$

car pour choisir une partition de $\{1, \ldots, n+1\}$ ayant k blocs, il faut et il suffit soit de choisir une partition de $\{1, \ldots, n\}$ ayant $k-1$ blocs et de la compléter avec le bloc singleton $\{n+1\}$, soit d'ajouter l'élément $n+1$ à l'un des k blocs d'une partition de $\{1, \ldots, n\}$ ayant k blocs. Si X est une variable aléatoire de loi de Poisson de paramètre λ alors

$$\mathbb{E}(X^n) = \sum_{k=1}^{n} \left\{ n \atop k \right\} \lambda^k \quad \text{en particulier} \quad \mathbb{E}(X^n) = B_n \quad \text{si } \lambda = 1.$$

Si $S_{n,k}$ désigne le nombre de surjections de $\{1,\dots,n\}$ dans $\{1,\dots,k\}$, alors on dispose également de la formule explicite suivante, qui peut s'obtenir grâce au principe d'inclusion-exclusion :

$$\left\{ n \atop k \right\} = \frac{S_{n,k}}{k!} = \frac{1}{k!}\sum_{j=1}^{k}(-1)^{k-j}\binom{k}{j}j^n.$$

2.4. **Simulation de la loi uniforme sur Π_n.** Par analogie avec les permutations, il est naturel de chercher à construire une suite markovienne (q_n) où q_n est uniforme sur Π_n, obtenue en insérant aléatoirement (selon une loi à déterminer) n dans q_{n-1}. Démontrons par l'absurde que cela est impossible. Pour tout $q \in \Pi_n$, notons $s_n(q) \in \Pi_{n-1}$ la partition obtenue en supprimant n de q. On aurait alors $s_n(q_n) = q_{n-1}$, et donc

$$\mathbb{P}(q_n = q) = \mathbb{P}(q_{n-1} = s_n(q))\mathbb{P}(q_n = q | q_{n-1} = s_n(q)).$$

Pour un $q' \in \Pi_{n-1}$ fixé, on somme sur tous les $q \in \Pi_n$ tels que $s_n(q) = q'$ pour obtenir

$$\sum_{q\in\Pi_n : s_n(q)=q'} \mathbb{P}(q_n = q) = \mathbb{P}(q_{n-1} = q').$$

Comme q_{n-1} et q_n sont uniformes sur Π_{n-1} et Π_n respectivement, on obtiendrait, pour tout $q' \in \Pi_{n-1}$,

$$\frac{|q'|+1}{B_n} = \frac{1}{B_{n-1}},$$

ce qui est faux. Ainsi, il n'existe pas d'algorithme récursif de simulation de la loi uniforme sur Π_n consistant en une insertion aléatoire de n à chaque étape. En revanche, il est possible d'obtenir, en utilisant les formules (2.1) et (2.2), deux algorithmes de simulation de la loi uniforme sur Π_n.

Algorithme tiré de la formule 2.1. On simule d'abord le cardinal du complémentaire du bloc contenant n, puis on choisit les éléments de ce complémentaire, puis on partitionne ce complémentaire. Plus précisément, soit K une variable aléatoire prenant la valeur $k \in \{0,\dots,n-1\}$ avec probabilité $\binom{n-1}{k}B_k/B_n$. Conditionnellement à K, on choisit une partie S de $\{1,\dots,n-1\}$ à K éléments de façon équiprobable. Conditionnellement à S, on choisit récursivement une partition p' de S de loi uniforme (sur l'ensemble des partitions de S). Alors $p = \{p', \{1,\dots,n\}\backslash S\}$ est une partition de $\{1,\dots,n\}$ suivant la loi uniforme. Comme $|S| < n$, ceci fournit l'algorithme annoncé, qui peut s'écrire de manière récursive ou itérative. Évidemment, l'inconvénient de cet algorithme est le calcul des B_k par récurrence.

Algorithme (de Stam) tiré de la formule (2.2). Cet algorithme est décrit dans [Kh2]. On choisit un entier M aléatoire puis on attribue à chaque élément de $\{1,\dots,n\}$ une couleur aléatoire (allant de 1 à M) et on considère la partition dont les blocs sont les éléments de même couleur. Plus précisément, on décide que M est aléatoire et prend la valeur $m \in \mathbb{N}$ avec probabilité $m^n/(m!eB_n)$ (la somme de ces probabilités vaut 1 grâce à (2.2)). Conditionnellement à M, soient $C_1,\dots,C_n$ les couleurs respectives de $1,\dots,n$, choisies de manière i.i.d. selon la loi uniforme sur $\{1,\dots,M\}$. On construit à présent la partition aléatoire $p \in \Pi_n$ en décidant que i et j sont dans le même bloc si et seulement si $C_i = C_j$. Alors p suit la loi uniforme sur Π_n car pour tout $p_0 \in \Pi_n$ à k blocs, on a

$$\mathbb{P}(p = p_0) = \sum_{m=k}^{\infty}\mathbb{P}(p = p_0 | M = m)\mathbb{P}(M = m) = \sum_{m=k}^{\infty}\frac{m(m-1)\cdots(m-k+1)}{m^n}\frac{m^n}{m!eB_n} = \frac{1}{B_n}.$$

2.5. **Les effectifs A_n.** Cette section est consacrée à l'étude des effectifs A_n.

Théorème 2.6 (Loi d'Ewens sur les effectifs)**.** *Pour tout $n \in \mathbb{N}^*$ et tout $(a_1,\dots,a_n) \in \mathbb{N}^n$ vérifiant $a_1 + 2a_2 + \cdots + na_n = n$, on a*

$$\mathbb{P}(A_{n,1} = a_1,\dots,A_{n,n} = a_n) = \frac{n!}{\theta_{(n)}}\prod_{j=1}^{n}\frac{1}{a_j!}\left(\frac{\theta}{j}\right)^{a_j}. \quad (2.3)$$

Démonstration. Soit $(a_1,\dots,a_n) \in \mathbb{N}^n$ fixé vérifiant $a_1 + 2a_2 + \cdots + na_n = n$. Notons F l'ensemble des $p \in \Pi_n$ dont les effectifs associés sont $(a_1,\dots,a_n)$. On a

$$|F| = n!\prod_{j=1}^{n}\frac{1}{(j!)^{a_j}a_j!}.$$

Ensuite, pour tout $p \in F$, on a

$$\prod_{b\in p}(|b|-1)! = \prod_{j=1}^{n}((j-1)!)^{a_j} \quad \text{et} \quad |b| = a_1 + a_2 + \cdots + a_n.$$

Le résultat découle alors du lemme 2.3 puisqu'il donne

$$\mathbb{P}(A_{n,1} = a_1, \dots, A_{n,n} = a_n) = \sum_{p \in F} \mathbb{P}(p_n = p) = |F| \, \frac{\theta^{a_1 + \cdots + a_n}}{\theta_{(n)}} \prod_{j=1}^{n} ((j-1)!)^{a_j},$$

ce qui est le résultat après simplification des factorielles. □

Remarque 2.7 (Contrainte et dépendance)**.** *La forme produit de la formule* (2.3) *pourrait faire croire que les composantes $A_{n,1}, \dots, A_{n,n}$ sont indépendantes mais il n'en n'est rien à cause de la contrainte $A_{n,1} + \cdots + nA_{n,n} = n$. En revanche, cette contrainte est en quelque sorte le seul obstacle à l'indépendance des $A_{n,i}$ comme en témoigne le théorème suivant.*

Théorème 2.8 (Description poissonnienne de la loi d'Ewens)**.** *Soit (Z_i) une suite de variables aléatoires indépendantes telle que Z_i suit une loi de Poisson de paramètre θ/i. Alors la loi de $(A_{n,1}, \dots, A_{n,n})$ est la loi conditionnelle de $(Z_1, \dots, Z_n)$ sachant que $Z_1 + 2Z_2 + \cdots + nZ_n = n$:*

$$\mathbb{P}(A_{n,1} = a_1, \dots, A_{n,n} = a_n) = \mathbb{P}(Z_1 = a_1, \dots, Z_n = a_n \,|\, Z_1 + 2Z_2 + \cdots + nZ_n = n).$$

Démonstration. Soit $(a_1, \dots, a_n) \in \mathbb{N}^n$. Si $a_1 + 2a_2 + \cdots + na_n \neq n$ alors les deux membres de l'égalité à prouver sont nuls. Sinon, en notant $W = \mathbb{P}(Z_1 + 2Z_2 + \cdots + nZ_n = n)$, on a

$$\mathbb{P}(Z_1 = a_1, \dots, Z_n = a_n | Z_1 + 2Z_2 + \cdots + nZ_n = n)$$

$$\begin{aligned}
&= W^{-1}\mathbb{P}(Z_1 = a_1, \dots, Z_n = a_n, Z_1 + 2Z_2 + \cdots + nZ_n = n) = W^{-1}\mathbb{P}(Z_1 = a_1, \dots, Z_n = a_n) \\
&= W^{-1} \prod_{j=1}^{n} \mathbb{P}(Z_j = a_j) = Z^{-1} \prod_{j=1}^{n} e^{-1/j} \, \frac{1}{a_j!} \left(\frac{\theta}{j}\right)^{a_j} \\
&= V \, \mathbb{P}(A_{n,1} = a_1, \dots, A_{n,n} = a_n) \quad \text{où} \quad V = W^{-1} \exp\left(-\sum_{j=1}^{n} \frac{1}{j}\right) \frac{\theta_{(n)}}{n!}.
\end{aligned}$$

Or deux lois de probabilité différant d'une constante multiplicative sont égales, d'où le résultat. □

Remarque 2.9 (Formule accessoire)**.** *Au passage, la preuve précédente fournit $V = 1$ i.e.*

$$\mathbb{P}(Z_1 + 2Z_2 + \cdots + nZ_n = n) = \exp\left(-\sum_{j=1}^{n} \frac{1}{j}\right) \frac{\theta_{(n)}}{n!},$$

ce qui n'est pas facile à prouver directement.

2.6. **Le nombre de tables K_n.** La loi de K_n s'obtient de façon élémentaire. En effet,

$$\sum_{\sigma \in \mathfrak{S}_n} \mathbb{P}(\sigma_n = \sigma) = \sum_{\sigma \in \mathfrak{S}_n} \frac{\theta^{K(\sigma)}}{\theta_{(n)}} = 1.$$

En notant $c_{n,k}$ le nombre de $\sigma \in \mathfrak{S}_n$ ayant exactement k cycles, on a

$$\sum_{k=0}^{n} c_{n,k} \theta^k = \theta_{(n)} = \theta(\theta + 1) \cdots (\theta + n - 1). \tag{2.4}$$

Ceci est la fonction génératrice bien connue des $(c_{n,k})$ que nous avons retrouvée simplement en arguant qu'une somme de probabilités vaut 1 ! Maintenant, en utilisant

$$\mathbb{P}(K_n = k) = c_{n,k} \frac{\theta^k}{\theta_{(n)}}$$

et en remplaçant θ par $t\theta$ dans (2.4), il vient

Théorème 2.10 (Fonction génératrice de K_n)**.**

$$\mathbb{E}(t^{K_n}) = \sum_{k=1}^{n} \mathbb{P}(K_n = k) t^k = \frac{(t\theta)_{(n)}}{\theta_{(n)}}. \tag{2.5}$$

Par identification dans (2.4) et dans (2.5), il vient

$$c_{n,k} = \text{coefficient de } t^k \text{ dans } t(t+1)\cdots(t+n-1) = \mathrm{sym}_{n-k}(1,2,\ldots,n-1)$$

où sym_{n-k} est la $(n-k)$-ème fonction symétrique élémentaire, et

$$\mathbb{P}(K_n = k) = \text{coefficent de } t^k \text{ dans le polynôme } \frac{(t\theta)_{(n)}}{\theta_{(n)}} = \frac{\theta^k}{(\theta)_{(n)}}\mathrm{sym}_{n-k}(1,2,\ldots,n-1).$$

La formule $K_n = \xi_1 + \cdots + \xi_n$ permet facilement d'accéder à l'espérance et à la variance :

$$\mathbb{E}(K_n) = \sum_{k=0}^{n-1}\frac{\theta}{\theta+k} \qquad \text{et} \qquad \mathrm{Var}(K_n) = \sum_{k=1}^{n-1}\frac{\theta k}{(\theta+k)^2}. \tag{2.6}$$

Remarque 2.11 (Autre méthode)**.** *Le théorème 2.10 aurait aussi pu être prouvé en utilisant la formule $K_n = \xi_1 + \cdots + \xi_n$ et l'indépendance :*

$$\mathbb{E}(t^{K_n}) = \mathbb{E}(\prod_{i=1}^n t^{\xi_i}) = \prod_{i=1}^n \mathbb{E}(t^{\xi_i}) = \prod_{i=1}^n \left(t^0\mathbb{P}(\xi_i=0) + t\mathbb{P}(\xi_i=1)\right) = \prod_{i=1}^n \frac{t\theta+i-1}{\theta+i-1} = \frac{(t\theta)_{(n)}}{\theta_{(n)}}.$$

2.7. **Solitaires.** Posons $G_n(t) = \mathbb{E}(t^{S_n})$. La règle d'évolution de (S_n) permet d'écrire

$$\begin{aligned}
G_{n+1}(t) &= \sum_{s\in\mathbb{N}} \mathbb{P}(S_n = s)\mathbb{E}(t^{S_{n+1}} \mid S_n = s) \\
&= \sum_{s\in\mathbb{N}} \mathbb{P}(S_n = s)\left(\frac{\theta}{\theta+n}t^{s+1} + \frac{s}{\theta+n}t^{s-1} + \frac{n-s}{\theta+n}t^s\right) \\
&= \frac{t\theta+n}{\theta+n}G_n(t) + \frac{1-t}{\theta+n}G_n'(t) \\
&= G_n(t) + \frac{t-1}{\theta+n}(\theta G_n(t) - G_n'(t))
\end{aligned}$$

Dans le cas $\theta = 1$, on voit par récurrence que

$$G_n(t) = \sum_{k=0}^n \frac{(t-1)^k}{k!}.$$

Dans ce cas, S_n est le nombre de points fixes d'une permutation de $\mathfrak{S}_n$ équiprobable. Comme sa fonction génératrice est la troncature d'ordre n de la fonction génératrice d'une loi de Poisson $\mathcal{P}(1)$, ses moments factoriels jusqu'à l'ordre n sont ceux de la loi de Poisson $\mathcal{P}(1)$

$$\forall\ 0 \le k \le n, \quad \mathbb{E}(S_n(S_n-1)\cdots(S_n-k+1)) = 1.$$

Par changement de base dans $\mathbb{R}_n[t]$, les moments usuels jusqu'à l'ordre n sont aussi ceux d'une loi de Poisson $\mathcal{P}(1)$.

Remarque 2.12 (Formule classique)**.** *En développant la fonction génératrice en puissance de t et en identifiant, on retrouve la formule classique*

$$\mathbb{P}(S_n = k) = \binom{n}{k}\sum_{i=0}^{n-k}\frac{(-1)^i}{i!},$$

qui s'obtient habituellement par la formule du crible de Poincaré (principe d'inclusion-exclusion).

On peut améliorer l'identité sur les moments de $S_n = A_{n,1}$: si les variables aléatoires Z_i sont indépendantes et suivent des lois de Poisson $\mathcal{P}(1/i)$, alors les moments factoriels joints de $(A_{n,1},\ldots,A_{n,n})$ sont exactement ceux de $(Z_1,\ldots,Z_n)$ jusqu'à un certain ordre :

$$\mathbb{E}\left(\prod_{i=1}^n A_{n,i}^{[m_i]}\right) = \mathbb{E}\left(\prod_{i=1}^n Z_i^{[m_i]}\right)\mathbf{1}_{\{m_1+\cdots+nm_n\le n\}},$$

où $x^{[m]} = x(x-1)\cdots(x-m+1)$. La démonstration de ce résultat dû à Watterson se trouve par exemple dans le livre [ABT]. L'étude de $A_{n,1}$ montre que cette formule poissonnienne ne subsiste pas quand $\theta \neq 1$.

2.8. **Table unique.** Ici, peu de choses à dire tellement la loi de $A_{n,n}$ est simple : $A_{n,n} = (1-\xi_2)\cdots(1-\xi_n)$ et $A_{n,n}$ suit une loi de Bernoulli de paramètre

$$\mathbb{P}(A_{n,n} = 1) = \mathbb{P}(\xi_2 = \ldots = \xi_n = 0) = \prod_{i=2}^n\left(1 - \frac{\theta}{\theta+i-1}\right).$$

3. Comportement asymptotique

3.1. **Le nombre de tables K_n.** Les formules (2.6) ainsi qu'une comparaison série-intégrale permettent de montrer le lemme suivant.

Lemme 3.1 (Espérance et variance asymptotiques du nombre de tables)**.** *On a*

$$\mathbb{E}(K_n) \sim \mathrm{Var}(K_n) \sim \theta \ln(n) \quad \textit{quand } n \to +\infty.$$

On peut préciser ceci par une convergence en probabilités.

Théorème 3.2 (Comportement asymptotique du nombre de tables en probabilité)**.** *La suite* $(K_n/\ln(n))$ *converge en probabilités vers* θ*, ce qui signifie que*

$$\forall \varepsilon > 0, \qquad \lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{K_n}{\ln(n)}\right| > \varepsilon\right) = 0.$$

Démonstration. Par l'inégalité de Markov et le lemme 3.1, pour tous $\varepsilon > 0$ et $n \geq 2$,

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

□

Remarque 3.3 (Convergence presque sûre)**.** *On peut raffiner ce résultat en établissant une convergence presque sûre, soit par une méthode de martingale, soit en montrant que la sous-suite* $(K_{2^n}/\ln(2^n))$ *converge presque sûrement, puis en procédant par encadrement.*

Théorème 3.4 (Comportement asymptotique en loi du nombre de tables)**.** *On a*

$$K_n^* = \frac{K_n - \theta\ln(n)}{\sqrt{\theta \ln(n)}} \xrightarrow[n\to+\infty]{\text{loi}} \mathcal{N}(0,1)$$

où $\mathcal{N}(0,1)$ *désigne la loi normale centrée réduite (de moyenne* 0 *et de variance* 1*).*

Démonstration. Le critère de convergence en loi basé sur les fonctions génératrices est approprié. En effet, la fonction génératrice G_n de K_n étant donnée par le théorème 2.10, celle de K_n^* se calcule facilement :

$$G_n^*(t) = \mathbb{E}(t^{K_n^*}) = t^{-\sqrt{\theta\ln(n)}} G_n(t^{u_n}) = t^{-1/u_n} \frac{\Gamma(n + t^{u_n})\Gamma(\theta)}{\Gamma(\theta+n)\Gamma(t^{u_n}\theta)} \quad \text{où} \quad u_n = \frac{1}{\sqrt{\theta\ln(n)}}.$$

En utilisant $u_n \to 0$, $t^{u_n} \to 1$, $(1+\frac{a}{x})^x \to e^a$ et la formule de Stirling $\Gamma(x) \sim \sqrt{2\pi} x^{x-1/2} e^{-x}$ quand $x \to +\infty$, il vient facilement

$$G_n^*(t) \sim \exp\left(-\frac{\ln(t)}{u_n} + (t^{u_n} - 1)u_n^2\right).$$

Un développement limité d'ordre 2 prouve que $-\frac{\ln(t)}{u_n} + (t^{u_n}-1)u_n^2 \to (\ln(t))^2/2$, ce qui assure que

$$G_n^*(t) \to \exp\left(\frac{(\ln(t))^2}{2}\right) = \mathbb{E}(t^N),$$

où N suit une loi $\mathcal{N}(0,1)$. Enfin, la convergence des fonctions génératrices assure celle des lois. □

3.2. **Effectifs des petites tables.** On dispose du théorème suivant concernant les petites tables.

Théorème 3.5 (Convergence en loi de $(A_{n,1},\ldots,A_{n,k})$)**.** *Fixons* $k \in \mathbb{N}^*$*. Lorsque* $n \to +\infty$*, le vecteur aléatoire* $(A_{n,1},\ldots,A_{n,k})$ *converge en loi vers la loi du vecteur aléatoire* $(Z_1,\ldots,Z_k)$ *où* $Z_1,\ldots,Z_k$ *sont indépendantes et* Z_i *suit une loi de Poisson de paramètre* θ/i*. C'est-à-dire que pour tout* $(a_1,\ldots,a_k) \in \mathbb{N}^k$*,*

$$\lim_{n\to+\infty} \mathbb{P}(A_{n,1} = a_1, \ldots, A_{n,k} = a_k) = \mathbb{P}(Z_1 = a_1, \ldots, Z_k = a_k).$$

Démonstration. Pour $0 \leq l \leq m$, posons

$$T_{l,m} = (l+1)Z_{l+1} + \cdots + mZ_m$$

avec $T_{m,m} = 0$. Soit $(a_1,\ldots,a_k) \in \mathbb{N}^k$, $a = a_1 + \cdots + ka_k$ et $n \geq a$. On a

$$\begin{aligned}
\mathbb{P}(A_{n,1} = a_1, \ldots, A_{n,k} = a_k) &= \mathbb{P}(Z_1 = a_1, \ldots, Z_k = a_k \,|\, T_{0,n} = n) \\
&= \frac{\mathbb{P}(Z_1 = a_1, \ldots, Z_k = a_k, T_{0,n} = n)}{\mathbb{P}(T_{0,n} = n)} \\
&= \frac{\mathbb{P}(Z_1 = a_1, \ldots, Z_k = a_k)\mathbb{P}(T_{k,n} = n - a)}{\mathbb{P}(T_{0,n} = n)}
\end{aligned}$$

La fonction génératrice de $T_{k,n}$ est

$$\phi(x) = \mathbb{E}(x^{T_{k,n}}) = \prod_{i=k+1}^{n} \mathbb{E}(x^{iZ_i}) = \prod_{i=k+1}^{n} \exp\left(\frac{\theta(x^i - 1)}{i}\right) = C_{k,n}^{-1} \exp\left(\theta \sum_{i=k+1}^{n} \frac{x^i}{j}\right)$$

où

$$C_{k,n} = \exp\left(\theta \sum_{i=k+1}^{n} \frac{1}{i}\right).$$

En notant $[x^l]g(x)$ le coefficient de x^l dans le développement en série entière en 0 de $g(x)$, on a

$$\mathbb{P}(T_{k,n} = n-a) = [x^{n-a}]\phi(x).$$

Ainsi, en posant $u_{n-a} = C_{k,n}\,\mathbb{P}(T_{k,n} = n-a)$, on a :

$$\begin{aligned}
u_{n-a} &= [x^{n-a}] \exp\left(\theta \sum_{i=k+1}^{n} \frac{x^i}{i}\right) \\
&= [x^{n-a}] \exp\left(\theta \sum_{i=k+1}^{+\infty} \frac{x^i}{i}\right) \\
&= [x^{n-a}] \exp\left(\theta \sum_{i=1}^{+\infty} \frac{x^i}{i}\right) \exp\left(-\theta \sum_{i=1}^{k} \frac{x^i}{i}\right) \\
&= [x^{n-a}]\,(1-x)^{-\theta}\,\psi(x) \quad \text{où} \quad \psi(x) = \exp\left(-\theta \sum_{i=1}^{k} \frac{x^i}{i}\right).
\end{aligned}$$

On a alors recours au lemme suivant.

Lemme 3.6. *Si ψ est une fonction entière*

$$[x^l]\,(1-x)^{-\theta}\,\psi(x) = \frac{1}{l!} \sum_{i=0}^{+\infty} (-1)^i \frac{(\theta - i)_{(l)}}{i!}\,\psi^{(i)}(1).$$

Preuve du lemme. Posons $x = 1-h$. Alors, pour $|x| < 1$,

$$\begin{aligned}
(1-x)^{-\theta}\,\psi(x) = h^{-\theta}\,\psi(1-h) &= \sum_{i=0}^{+\infty} \frac{\psi^{(i)}(1)(-1)^i h^{i-\theta}}{i!} = \sum_{i=0}^{+\infty} \frac{\psi^{(i)}(1)(-1)^i (1-x)^{i-\theta}}{i!} \\
&= \sum_{i=0}^{+\infty} \frac{\psi^{(i)}(1)(-1)^i}{i!} \sum_{l=0}^{+\infty} \binom{i-\theta}{l} (-1)^l x^l = \sum_{l=0}^{+\infty} x^l \sum_{i=0}^{+\infty} \frac{\psi^{(i)}(1)(-1)^{i+l}}{i!} \binom{i-\theta}{l} \\
&= \sum_{l=0}^{+\infty} x^l \sum_{i=0}^{+\infty} \psi^{(i)}(1)(-1)^i \frac{(\theta-i)_{(l)}}{i!\,l!},
\end{aligned}$$

l'interversion des sommes étant justifiée par la convergence absolue de la série double précédente. □

Ainsi, en mettant $\theta_{(n-a)}$ en facteur, on a

$$u_{n-a} = \frac{\theta_{(n-a)}}{(n-a)!}\,v_{n-a} \quad \text{où} \quad v_{n-a} = \sum_{i=0}^{+\infty} (-1)^i \frac{(\theta-i)_{(n-a)}}{\theta_{(n-a)}} \frac{\psi^{(i)}(1)}{i!}.$$

Ensuite, à $i \geq 1$ fixé et quand $l \to +\infty$,

$$\frac{(\theta-i)_{(l)}}{\theta_{(l)}} = (\theta-i)_{(i)} \frac{\theta_{(l-i)}}{\theta_l} \sim (\theta-i)_{(i)} l^{-i} \to 0.$$

On peut aussi montrer (cf. [W, Théorème 5.3.1]) que

$$v_{n-a} = \psi(1) + o(1) \quad \text{quand } n \to +\infty.$$

Comme $\psi(1) = \exp\left(-\theta \sum_{i=1}^{k} \frac{1}{i}\right)$, il vient

$$\mathbb{P}(T_{k,n} = n-a) = \frac{\theta_{(n-a)}}{(n-a)!} \exp\left(-\theta \sum_{i=1}^{k} \frac{1}{i}\right)(1+o(1)).$$

Ceci vaut en particulier pour $k = 0$ et $a = 0$ (en fait, le reste $o(1)$ est nul dans ce cas comme le montre la remarque 2.9). On obtient donc

$$\frac{\mathbb{P}(T_{k,n} = n - a)}{\mathbb{P}(T_{0,n} = n)} = \frac{\theta_{(n-a)}\, n!}{\theta_{(n)}\, (n-a)!}(1 + o(1)) \to 1$$

par une nouvelle application de la formule de Stirling. □

En particulier, le nombre de points fixes $S_n = A_{n,1}$ converge en loi vers Z_1 de loi de Poisson de paramètre θ, le nombre de transpositions $A_{n,2}$ converge en loi vers Z_2 de loi de Poisson de paramètre $\theta/2$, etc. Il est remarquable qu'asymptotiquement les $(A_{n,i})_{1\leq i\leq k}$ deviennent indépendants entre eux. Ceci n'a été démontré qu'à k fixé et resterait valable pour un k tendant vers l'infini avec n tant que $k = o(n)$, cf. le livre [ABT]. Le nombre de cycles de grande taille ne saurait être poissonnien comme le montre l'exemple qui suit de $A_{n,n} \in \{0, 1\}$ qui tend presque-sûrement vers 0.

3.3. **Table unique.** La suite $(A_{n,n})$ est décroissante et à valeurs dans $\{0, 1\}$. Considérons le premier instant (éventuellement $+\infty$) où $A_{n,n} = 0$, soit

$$T = \inf\{n \geq 1 : A_{n,n} = 0\}.$$

On a alors, pour tout $n \geq 1$,

$$\mathbb{P}(T > n) = \mathbb{P}(A_{n,n} = 1) = \prod_{i=2}^{n}\left(1 - \frac{\theta}{\theta + i - 1}\right).$$

Comme la série $\sum \frac{\theta}{\theta+i-1}$ diverge, on a $\lim_{n\to+\infty} \mathbb{P}(T > n) = 0$ i.e. $\mathbb{P}(T = +\infty) = 0$. Donc $(A_{n,n})$ est nulle à partir d'un certain rang. Par comparaison avec une intégrale, on peut facilement montrer que,

$$\mathbb{P}(T > n) \sim \frac{\text{constante}}{n^{\theta}} \quad \text{quand } n \to +\infty.$$

4. Restaurant de Feller

Pour tout $n \geq 1$, le restaurant chinois fournit le tableau triangulaire aléatoire $(A_{i,k})_{1\leq k\leq i\leq n}$. D'après le théorème 2.6, pour tout $1 \leq i \leq n$, la i^{e} ligne $(A_{i,1}, \ldots, A_{i,i})$ de ce tableau suit la loi d'Ewens de taille i. Pour tout $k \geq 1$, le théorème 3.5 assure la convergence en loi de $(A_{n,1}, \ldots, A_{n,k})$ quand $n \to \infty$. Feller a proposé une méthode de construction d'un autre tableau triangulaire aléatoire, pour lequel la i^{e} ligne suit également la loi d'Ewens, pour tout $1 \leq i \leq n$, et pour lequel, pour tout $k \geq 1$, on a cette fois-ci convergence presque sûre de $(A_{n,1}, \ldots, A_{n,k})$ quand $n \to \infty$.

Remarque 4.1 (Couplage)**.** *Soient μ_1 et μ_2 deux lois de probabilités. On appelle* couplage *de μ_1 et μ_2 tout couple (X_1, X_2) de variables aléatoires définies sur un même espace de probabilité tel que X_1 est de loi μ_1 et X_2 de loi μ_2. Plus généralement, et par abus de langage, on parle de couplage d'une famille finie ou dénombrable de lois de probabilités.*

Le restaurant chinois ainsi que la méthode de Feller constituent deux couplages de la famille des lois d'Ewens. Pour cette raison, la méthode de Feller est souvent appelée *Feller coupling*, qu'on appelle ici *restaurant de Feller* pour rester dans la gastronomie. Le restaurant de Feller est en quelque sorte dual du restaurant chinois car les rôles des clients et des tables sont presque inversés.

Le restaurant de Feller fournit en fait des permutations aléatoires, tout comme le restaurant chinois. Plus précisément, soit $(\xi_n)_{n\geq1}$ une suite de variables aléatoires indépendantes de loi de Bernoulli avec

$$\mathbb{P}(\xi_n = 1) = 1 - \mathbb{P}(\xi_n = 0) = \theta/(\theta + n - 1)$$

(cet ingrédient est identique à celui des restaurants chinois (1.1)). Pour tout $n \geq$ fixé, on considère la suite $\xi_n, \ldots, \xi_1, 1$. On commence par remplir la table 1 avec le client 1. Si $\xi_n = 1$, on clôt cette table et on commence à asseoir le client 2 à la table 2. Si $\xi_n = 0$, on choisit un client au hasard parmi $\{2, \ldots, n\}$ que l'on ajoute à la table 1. Puis on examine $\xi_{n-1}, \xi_{n-2}, \ldots$ en ajoutant des clients choisis au hasard parmi ceux encore en attente tant qu'on voit des 0 et en fermant la table en cours dès qu'on voit un 1, pour commencer la suivante avec le client de plus petit numéro non encore vu. Après lecture de ξ_1, la permutation construite σ_n suit la loi d'Ewens.

L'avantage de ce processus est que la taille de chaque cycle est constante à partir d'un certain rang car dès qu'une table est close, elle n'accueille plus jamais personne. Dans ce cas, la suite des effectifs complétée par des 0 $(A_{n,1}, \ldots, A_{n,n}, 0, 0, \ldots)$ converge presque-sûrement vers une suite $(A_{\infty,1}, A_{\infty,2}, \ldots)$ de variables de Poisson indépendantes et de paramètres $\theta, \theta/2, \ldots$

Ce couplage permet de préciser la distance en variance totale et la distance de Wasserstein entre les lois de $(A_{n,1}, \ldots, A_{n,k})$ et de $(A_{\infty,1}, \ldots, A_{\infty,k})$ et d'en donner de bonnes majoration et minoration. Le fait que $(A_{\infty,1}, A_{\infty,2}, \ldots)$ soient des variables de Poisson indépendantes et de paramètres $\theta, \theta/2, \ldots$ provient alors du théorème précédent mais le prouver directement n'est pas évident, et nous renvoyons à [ABT].

Dans la suite infinie de variables aléatoires de Bernoulli hétérogènes (ξ_n), la variable aléatoire $A_{\infty,i}$ est le nombre d'espacements de longueur $i-1$ entre deux 1 consécutifs et alors les $A_{\infty,i}$ sont indépendantes de lois de Poisson de paramètres θ/i. Pour comparaison, si (ξ_n) était une suite de variables aléatoires de Bernoulli homogènes, alors les espacements seraient de même loi géométrique, et on aurait $A'_{\infty,i} = +\infty$ pour tout $i \in \mathbb{N}^*$. Dans notre cas, à i fixé, la probabilité p_n d'un espacement de longueur $i-1$ entre n et $n+i$ est inférieure à $\theta^2/((\theta+n-1)(\theta+n+i-1))$ à cause des 1 de début et de fin. Ceci prouve que $\sum p_n < \infty$, et donc qu'à partir d'un certain rang, il n'y a plus d'espacement de taille $i-1$ i.e. $A_{\infty,i} < \infty$.

5. Épilogue

Warren Ewens est un professeur de biologie mathématique né en 1937 en Australie. C'est à la fin des années 1960 qu'il découvre la loi qui porte aujourd'hui son nom, en étudiant un problème d'échantillonnage en génétique des populations, lié à un célèbre modèle de Fisher et Wright. Le paramètre θ apparaît comme un taux de mutation des allèles. Une synthèse sur le sujet se trouve dans son livre [E], ainsi que dans celui de Kingman [Kn1]. De nombreux aspects statistiques sont abordés dans le cours de Tavaré [T]. Le travail d'Ewens a engendré un nombre considérable de travaux en biologie quantitative et en probabilités. La loi d'Ewens apparaît dans une large gamme de structures aléatoires discrètes dites logarithmiques, allant de la combinatoire à la théorie des nombres. On pourra consulter à ce sujet le livre de Arratia, Barbour, et Tavaré [ABT]. Il semble que le processus des restaurants chinois doive son nom à Pitman et Dubin. Il apparaît sous ce nom dans un cours d'Aldous [A]. On peut le relier aux urnes de Pólya ainsi qu'aux processus de Dirichlet. De nos jours, le processus des restaurants chinois et la loi d'Ewens font désormais partie du folklore d'une théorie plus générale de la (fragmentation et de la) coalescence. On pourra à ce sujet consulter les livres de Kingman [Kn2], de Bertoin [Bn], de Pitman [P], ainsi que de Berestycki [Bi].

Références

[A] D. J. Aldous. *Exchangeability and related topics*, volume 1117 de *Lecture Notes in Math.*, pages 1–198, Springer, 1985. Notes de cours de la XIIIème École d'été de Probabilités de Saint-Flour, été 1983.

[ABT] R. Arratia, A. D. Barbour, et S. Tavaré. *Logarithmic combinatorial structures : a probabilistic approach.* EMS Monographs in Mathematics. European Mathematical Society (EMS), 2003.

[BMP] P. Baldi, L. Mazliak, et P. Priouret. *Martingales et chaînes de Markov*, Hermann, 1998.

[Bi] N. Berestycki. *Recent progress in coalescent theory*, volume 16 de *Ensaios Matemáticos*. Sociedade Brasileira de Matemática, 2009.

[Bn] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2006.

[E] W. J. Ewens. *Mathematical population genetics. I*, volume 27 de *Interdisciplinary Applied Mathematics*. Springer, seconde édition, 2004. Theoretical introduction.

[Kh1] D. E. Knuth. *The Art of Computer Programming*, volume 2, Addison and Wesley, 1997.

[Kh2] D. E. Knuth. *The Art of Computer Programming*, volume 4, Addison and Wesley, 2001.

[Kn1] J. F. C. Kingman. *Mathematics of genetic diversity*, volume 34 de *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), 1980.

[Kn2] J. F. C. Kingman. *Poisson processes*, volume 3 de *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, 1993. Oxford Science Publications.

[MR] R. Motwani et P. Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995.

[P] J. Pitman. *Combinatorial stochastic processes*, volume 1875 de *Lecture Notes in Mathematics*. Springer, 2006. Notes de cours de la XXXIIème École d'été de Probabilités de Saint-Flour, juillet 7–24, 2002.

[SS] M. Shaked et J. G. Shanthikumar. *Stochastic orders*, Springer Series in Statistics. Springer, 2007. xvi+473 pp.

[T] S. Tavaré. *Ancestral inference in population genetics*, volume 1837 de *Lecture Notes in Math.*, pages 1–188. Springer, 2004. Notes de cours de la XXXIème École d'été de Probabilités de Saint-Flour, juillet 8–25 2001.

[W] H. S. Wilf. *generatingfunctionology*, Academic Press, 1994.

(D. Chafaï, auteur correspondant) Université Paris-Est Marne-la-Vallée, UMR CNRS 8050, France.
URL: http://djalil.chafai.net/
E-mail address: djalil@chafai.net

(Y. Doumerc) Classes Préparatoires aux Grandes Écoles, Lycée Gaston Berger, Lille.
E-mail address: y.doumerc@yahoo.fr

(F. Malrieu) Université Rennes 1, UMR CNRS 6627, France.
URL: http://perso.univ-rennes1.fr/florent.malrieu/