MP*1

# Probabilités sur le groupe symétrique

## Corrigé du problème

### I. Quelques calculs élémentaires

1. a) Un élément de $\mathcal{S}_n$ qui stabilise $A$ est déterminé :
   - par sa restriction à $A$, qui est une permutation de $A$, pour laquelle il y a $k!$ choix ;
   - par sa restriction à $\{1, \ldots, n\} \setminus A$, pour laquelle il y a $(n-k)!$ choix.

   Il s'ensuit que
   $$P(\sigma_n(A) = A) = \frac{1}{\binom{n}{k}}.$$

   b) Si $A$ est une partie de $\{1, \ldots, n\}$, soit $\chi_A$ l'indicatrice de l'événement
   $$(\sigma_n(A) = A).$$
   Si $\mathcal{P}_{k,n}$ est l'ensemble des parties de cardinal $k$ de $\{1, \ldots, n\}$, on a
   $$N_n^k = \sum_{A \in \mathcal{P}_{k,n}} X_A,$$
   d'où, par linéarité de l'espérance
   $$E(N_n^k) = \sum_{A \in \mathcal{P}_{k,n}} E(X_A) = 1.$$

   c) Le nombre de partie de $\{1, \ldots, n\}$ stables par $\sigma_n$ est
   $$\sum_{k=0}^{n} N_n^k.$$
   Par linéarité de l'espérance, son espérance est $n+1$.

2. a) Soit $m$ dans $\{1, \ldots, n\}$. Un $m$-cycle est déterminé :
   - par son support, pour lequel il y a $\binom{n}{m}$ choix ;
   - par sa restriction à ce support, pour laquelle il y a $(m-1)!$ choix.

   Il s'ensuit que
   $$P(A_m^n) = \frac{(m-1)!}{m!(n-m)!} = \frac{1}{m(n-m)!}.$$

Si $m$ est fixé, la formule de Stirling montre que :

$$P(A_{m,n}) \underset{n\to+\infty}{\sim} \frac{e^{n-m}}{m\sqrt{2\pi n}(n-m)^{n-m}} \underset{n\to+\infty}{\sim} \frac{e^n n^{m-\frac{1}{2}}}{m\sqrt{2\pi}n^n},$$

où la seconde équivalence provient de

$$(n-m)^{n-m} = n^m \left(1 - \frac{m}{n}\right)^n \underset{n\to+\infty}{\sim} \frac{n^m}{e}.$$

b) On a, pour $n \geq \ell$ :

$$P(A^n_{n-\ell}) = \frac{1}{(n-\ell)\ell!} \underset{n\to+\infty}{\sim} \frac{1}{\ell! n}.$$

c) Les questions précédentes suggèrent que ce sont les indices $m$ proches de $n$ qui « pèsent le plus lourd » dans la somme

$$\sum_{m=1}^{n} P(A^n_m) = \sum_{\ell=0}^{n-1} \frac{1}{(n-\ell)\ell!},$$

donc que cette somme équivaut à $\frac{1}{ne}$. Prouvons-le. On écrit

$$n \sum_{m=1}^{n} P(A^n_m) = \sum_{\ell=0}^{n-1} \frac{n}{(n-\ell)\ell!}.$$

Posons

$$\alpha_{n,\ell} = \frac{n}{(n-\ell)\ell!}.$$

Pour $\ell$ fixé, la suite $(\alpha_{n,\ell})_{n\geq\ell}$ converge vers $\dfrac{1}{\ell!}$. Si $\ell \leq \dfrac{n}{2}$, alors

$$0 \leq \alpha_{n,\ell} \leq \frac{2}{\ell!}.$$

En appliquant le théorème de convergence dominée pour les séries, on a donc

$$\sum_{1\leq\ell\leq n/2} \alpha_{n,\ell} \underset{n\to+\infty}{\longrightarrow} e.$$

Reste à voir que

$$\sum_{n/2<\ell\leq n} \alpha_{n,\ell} \underset{n\to+\infty}{\longrightarrow} 0.$$

Mais, si $\ell > n/2$ :

$$\alpha_{n,\ell} \leq \frac{n}{\ell!} \leq \frac{n}{\lfloor n/2 \rfloor !}.$$

Par conséquent

$$\sum_{n/2<\ell\leq n} \alpha_{n,\ell} \leq \frac{n^2}{2\lfloor n/2 \rfloor !}.$$

Le résultat est alors clair par croissance comparée.

## II. Nombre de points fixes

3. a) Une permutation qui fixe $i_1, \ldots, i_k$ est déterminé par sa restriction au complémentaire de $\{i_1, \ldots, i_k\}$ dans $\{1, \ldots, n\}$, qui est une permutation quelconque de ce complémentaire. Par suite :

$$P\left(\bigcap_{j=1}^{k} (\sigma_n(i_j) = i_j)\right) = \frac{(n-k)!}{n!}.$$

b) Si $1 \leq i_1 < i_2 < \cdots < i_k$, notons $E_{i_1,\ldots,i_k}$ l'événement

$$\forall j \in \{1, \ldots, k\}, \qquad \sigma_n(i_k) = i_k.$$

On a

$$(C_n^1 \neq 0) = \bigcap_{i=1}^{n} E_i.$$

La formule du crible donne donc :

$$P(C_n^1 \neq 0) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\left(\bigcap_{j=1}^{k} E_{i_j}\right).$$

La question précédente montre alors que :

$$P(C_n^1 \neq 0) = \sum_{k=1}^{n} (-1)^{k-1} \frac{\binom{n}{k}(n-k)!}{n!} = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}.$$

Il reste à passer au complémentaire pour obtenir la formule désirée.

Le développement en série entière de l'exponentielle permet enfin de conclure que

$$P(C_n^1 = 0) \underset{n \to +\infty}{\longrightarrow} \frac{1}{e}.$$

c) Pour $k$ dans $\mathbb{N}$, notons $D_k$ le nombre de bijections sans point fixe d'un ensemble de cardinal $k$ sur lui-même, en convenant que $D_0 = 1$. On a

$$D_k = k!\, P(C_k^1 = 0) = k! \sum_{i=0}^{k} \frac{(-1)^i}{i!}.$$

Soient $j$ dans $\mathbb{N}^*$, $n \geq j$. Un élément de $\mathcal{S}_n$ est déterminé par l'ensemble de ses points fixes et sa restriction au complémentaire de cet ensemble, sur lequel elle induit une permutation sans point fixe. Par conséquent

$$P(C_n^1 = j) = \frac{\binom{n}{j} D_{n-j}}{n!} = \frac{P(C_{n-j}^1 = 0)}{j!}.$$

Il s'ensuit que

$$P(C_n^1 = j) \underset{n \to +\infty}{\longrightarrow} \frac{1}{ej!}.$$

Ainsi, la suite de variables aléatoires $(C_n^j)_{n \geq j}$ converge en loi vers une variable de Poisson de paramètre 1.

4. Notons $X_i$ l'indicatrice de $E_i$. Alors

$$C_n^1 = \sum_{i=1}^{n} X_i.$$

On a

$$P(X_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Par linéarité de l'espérance

$$E(C_n^1) = 1.$$

D'autre part, si $1 \leq i < j \leq n$,

$$P(X_i = X_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

Par conséquent

$$\operatorname{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}.$$

La formule qui donne la variance d'une somme fournit alors

$$V(C_n^1) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{\leq i<j\leq n} \operatorname{Cov}(X_i, X_j) = n\frac{1}{n}\left(1 - \frac{1}{n}\right) + 2\frac{n(n-1)}{2}\frac{1}{n^2(n-1)} = 1.$$

## III. Loi faible pour le nombre de cycles

5. a) On démontre l'égalité de polynômes

$$\sum_{\sigma \in \mathcal{S}_n} X^{k_n(\sigma)} = \prod_{k=0}^{n-1} (X + k)$$

par récurrence sur $n$. Le cas $n = 1$ est clair, car $\mathcal{S}_1$ ne contient que l'identité.

Supposons $n \geq 2$, le résultat vrai à l'ordre $n-1$. Partitionnons $\mathcal{S}_n$ en l'ensemble $F_n$ des permutations qui fixent $n$ et en son complémentaire $G_n$. Alors

$$\sum_{\sigma \in \mathcal{S}_n} X^{k_n(\sigma)} = \sum_{\sigma \in F_n} X^{k_n(\sigma)} + \sum_{\sigma \in G_n} X^{k_n(\sigma)}.$$

On dispose d'une bijection naturelle entre $\mathcal{S}_{n-1}$ et $F_n$, qui à $\sigma \in \mathcal{S}_{n-1}$ associe l'unique élément $\sigma^*$ de $D_n$ dont la restriction à $\{1, \ldots, n-1\}$ est $\sigma$. On a

$$\forall \sigma \in \mathcal{S}_{n-1}, \qquad k_n(\sigma^*) = k_{n-1}(\sigma) + 1.$$

Par conséquent

$$(1) \qquad \sum_{\sigma \in F_n} X^{k_n(\sigma)} = X \sum_{\sigma \in \mathcal{S}_{n-1}} X^{k_{n-1}(\sigma)} = X \prod_{k=0}^{n-2} (X + k).$$

Par ailleurs, $G_n$ est réunion disjointe des ensembles

$$G_{n,j} = \{\sigma \in \mathcal{S}_n \ ; \ \sigma^{-1}(n) = j\} \quad 1 \leq j \leq n-1.$$

Pour $j$ dans $\{1, \dots, n-1\}$, on construit une bijection de $G_{n,j}$ sur $\mathcal{S}_n$ en associant à $\sigma$ l'unique élément $\sigma_j^*$ de $\mathcal{S}_{n-1}$ obtenu en éliminant $n$ du cycle de $\sigma$ où il apparaît. Comme $n$ n'est pas un point fixe, $\sigma_j^*$ a le même nombre total de cycles que $\sigma$. Par conséquent

$$(2) \qquad \sum_{\sigma \in G_{n,j}} X^{k_n(\sigma)} = \sum_{\sigma \in \mathcal{S}_{n-1}} X^{k_{n-1}(\sigma)} = \prod_{k=0}^{n-2} (X+k).$$

En sommant les relations (2), on arrive à

$$(3) \qquad \sum_{\sigma \in G_n} X^{k_n(\sigma)} = (n-1) \prod_{k=0}^{n-2} (X+k).$$

Enfin, en sommant (1) et (3), il vient

$$\sum_{\sigma \in \mathcal{S}_n} X^{k_n(\sigma)} = \prod_{k=0}^{n-1} (X+k)$$

b) Pour $x$ dans $[-1, 1]$, on a donc

$$G_{K_n}(x) = E(x^{K_n}) = \frac{1}{n!} \sum_{j=1}^{n} P(K_n = j) x^j = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} x^{k_n(\sigma)}.$$

Par suite

$$\frac{G'_{K,n}}{G_{K_n}} = \sum_{k=0}^{n-1} \frac{1}{X+k}, \qquad \frac{G''_{K,n}}{G_{K_n}} - \left(\frac{G'_{K_n}}{G_{K_n}}\right)^2 = -\sum_{k=0}^{(n-1} \frac{1}{(X+k)^2}.$$

En particulier

$$G'_{K_n}(1) = H_n, \qquad G''_{K_n}(1) = {H_n}^2 - \sum_{k=1}^{n} \frac{1}{k^2}.$$

La variable aléatoire $K_n$ étant à valeurs dans un ensemble fini, elle admet des moments de tous ordres. On sait alors que

$$E(K_n) = G'_{K_n}(1) = H_n, V(K_n) = G''_{K_n}(1) + G'_{K_n}(1) - G'_{K_n}(1)^2 = H_n - \sum_{k=1}^{n} \frac{1}{k^2}.$$

6. a) Soit $\varepsilon$ dans $\mathbb{R}^{+*}$. Il suffit d'appliquer l'inégalité de Tchebychev à $K_n$ pour obtenir :

$$P\left(|K_n - E(K_n)| \geq \varepsilon E(K_n)\right) \leq \frac{V(K_n)}{\varepsilon^2 E(K_n)^2}.$$

Mais, d'après la question précédente, l'équivalent $H_n \sim \ln(n)$ et la convergence de $\sum \frac{1}{k^2}$, on a

$$E(K_n) \sim \ln(n) \quad \text{et} \quad V(K_n) \sim \ln(n).$$

Par suite

$$P\left(|K_n - E(K_n)| \geq \varepsilon E(K_n)\right) \underset{n\to+\infty}{=} O\left(\frac{1}{\ln(n)}\right).$$

b) Comme $\ln(n)$ et $H_n$ sont équivalents, on dispose de $N$ dans $\mathbb{N}^*$ tel que, pour $n \geq N$ :

$$\left(1 - \frac{\varepsilon}{2}\right) H_n \geq (1-\varepsilon)\ln(n) \quad \text{et} \quad (1+\varepsilon)\ln(n) \geq \left(1 + \frac{\varepsilon}{2}\right) H_n.$$

On a donc, pour $n \geq N$

$$\left(\left|\frac{K_n}{\ln(n)} - 1\right| \geq \varepsilon\right) \subset \left(\left|\frac{K_n}{E(K_n)} - 1\right| \geq \frac{\varepsilon}{2}\right).$$

La question b) permet de conclure. On a donc un résultat type « loi faible des grands nombres » pour la suite de variables aléatoires $(K_n)_{n\geq 1}$.

## IV. Loi asymptotique des cycles courts

### A. Préliminaire analytique

7. La fonction

$$x \mapsto \frac{1}{1-x}$$

est développable en série entière sur $]-1,1[$, avec

$$\forall x \in ]-1,1[, \qquad \frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n.$$

Le théorème sur le produit de Cauchy de deux séries entières assure que $f$ est développable en série entière sur $]-1,1[$ et que, si on pose

$$\forall x \in ]-1,1[, \qquad f(x) = \sum_{n=0}^{+\infty} a_n x^n,$$

alors

$$\forall n \in \mathbb{N}, \qquad a_n = \sum_{k=0}^{n} b_k.$$

Or, par définition :

$$\sum_{k=0}^{+\infty} b_k = g(1).$$

Il s'ensuit que $(a_n)_{n\geq 0}$ converge vers $g(1)$.

8. On a
$$V(x)-U(x)=e^{u(x)}\left(e^{v(x)-u(x)}-1\right)\underset{x\to 0}{\sim}e^{u(0)}\left(v(x)-u(x)\right).$$
Or
$$e^{v-u}(x)-1\underset{x\to 0}{\sim}v(x)-u(x)\quad\text{donc}\quad e^{v-u}(x)-1\underset{x\to 0}{\sim}(v-u)(x).$$
Par conséquent
$$V(x)-U(x)\underset{x\to 0}{=}O(x^{p+1}).$$
La formule de Taylor-Young et l'unicité du développement limité assurent que :
$$\forall i\in\{0,\dots,p\},\qquad V^{(i)}(0)=U^{(i)}(0).$$

### B. Description poissonienne de la distribution des types cycliques

9. Se donner une permutation $\sigma$ de type cyclique $(c_1,\dots,c_n)$, c'est faire les choix suivants.

- Choix des supports des cycles : il a
$$\frac{n!}{(n-c_1)!}\times\frac{(n-c_1)!}{2!^{c_2}(n-c_1-2c_2)!}\times\frac{(n-c_1-\dots-(n-1)c_{n-1})!}{n!^{c_n}}$$
façon de choisir succesivement $c_1$ parties de cardinal 1, $c_2$ parties de cardinal 2,..., $c_n$ parties de cardinal $n$ de $c_n$ formant une partition de $\{1,\dots,n\}$. Ce résultat vaut en supposant les supports des cycles de longueur $i$ rangés pour tout $i$. Il faut donc le diviser par $c_1!\times\dots\times c_n!$ pour obtenir le choix des supports.

- Choix de l'action de $\sigma$ sur le support de chaque cycle. Comme il y a $(k-1)!$ permutations circulaires d'un ensemble de cardinal $k$, le nombre de choix est
$$\prod_{i=1}^{n}(i-1)!^{c_i}$$
Au total, il y a bien
$$n!\prod_{i=1}^{n}\frac{1}{i^{c_i}\,c_i!}$$
éléments de $\mathcal{S}_n$ de type cyclique $(c_1,\dots,c_n)$.

10. Si $c_1+2c_2+\dots+nc_n\neq n$, les deux membres de l'égalité sont nuls.

Supposons désormais $c_1+2c_2+\dots+nc_n=n$. Notons d'abord que
$$P\left(Z_1=c_1,\dots,Z_n=c_n\right)=\prod_{i=1}^{n}\frac{e^{-1/i}}{c_i!\,i^{c_i}}.$$
Par suite
$$P\left(Z_1=c_1,\dots,Z_n=c_n\,\middle|\,\sum_{j=1}^{n}jZ_j=n\right)=\frac{1}{P\left(\sum_{j=1}^{n}jZ_j=n\right)}\prod_{i=1}^{n}\frac{e^{-1/i}}{c_i!\,i^{c_i}}.$$

Cette relation se réécrit

$$P\left(Z_1 = c_1, \ldots, Z_n = c_n \,\middle|\, \sum_{j=1}^{n} jZ_j = n\right) = \frac{1}{e^{H_n}\ P\left(\sum_{j=1}^{n} jZ_j = n\right)} \prod_{i=1}^{n} \frac{1}{c_i!\ i^{c_i}}.$$

Ou encore

$$P\left(Z_1 = c_1, \ldots, Z_n = c_n \,\middle|\, \sum_{j=1}^{n} jZ_j = n\right) = \frac{1}{e^{H_n}\ P\left(\sum_{j=1}^{n} jZ_j = n\right)} P(C_n^1 = c_1, \ldots, C_n^n = c_n).$$

Or deux lois de probabilités qui diffèrent d'une constante multiplicative sont égales, d'où le résultat.[1]

**C. Convergence en loi du vecteur** $(C_n^1, \ldots, C_n^k)$

11. a) Soit $Z$ une variable de Poisson de paramètre $\lambda$. Alors, pour $x$ dans $[-1,1]$ (et même dans $\mathbb{R}$) :

$$G_Z(x) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k x^k}{k!} = e^{\lambda(x-1)}.$$

b) Soit $x$ dans $[-1,1]$. Alors, puisque, pour tout $j$, $x^{jZ_j}$ est fonction déterministe de $Z_j$, les variables aléatoires $x^{jZ_j}$ sont mutuellement indépendantes. Par suite, on a, en appliquant le théorème du produit des espérances

$$G_{T_{k,n}}(x) = E(x^{\sum_{j=k+1}^{n} jZ_j}) = E(\prod_{j=k+1}^{n} x^{jZ_j}) = \prod_{j=k+1}^{n} E(x^{jZ_j}),$$

c'est-à-dire

$$G_{T_{k,n}}(x) = \prod_{j=k+1}^{n} \exp\left(\frac{1}{j}(x^j - 1)\right) = \exp\left(H_k - H_n\right)\ \exp\left(\sum_{j=k+1}^{n} \frac{x^j}{j}\right).$$

12. Comme $n \geq k$, la loi de $(C_1^n, \ldots, C_n^k)$ est la loi de $(Z_1, \ldots, Z_k)$ conditionnellement à $\left(\sum_{j=1}^{n} jZ_j = n\right)$. Ainsi :

$$P(C_n^1 = c_1, \ldots, C_n^k = c_k) = \frac{P(Z_1 = c_1, \ldots, Z_k = c_k, T_{0,n} = n)}{P(T_{0,n} = n)}.$$

Or

$$(Z_1 = c_1, \ldots, Z_k = c_k, T_{0,n} = n) = (Z_1 = c_1, \ldots, Z_k = c_k, T_{k,n} = n - c).$$

Comme $T_{k,n}$ est fonction déterministe de $(Z_{k+1}, \ldots, Z_n)$, les événements $(Z_1 = c_1, \ldots, Z_k = c_k)$ et $(T_{k,n} = n - c)$ sont indépendants. On en déduit bien

$$P(C_n^1 = c_1, \ldots, C_n^k = c_k) = P(Z_1 = c_1, \ldots, Z_k = c_k)\ \frac{P(T_{k,n} = n - c)}{P(T_{0,n} = n)}.$$

[1] Au passage, on obtient la formule non évidente

$$P(Z_1 + 2Z_2 + \cdots + nZ_n = n) = e^{-H_n}.$$

13. a) On va appliquer la question 8 aux fonctions $U = G_{T_{k,n}}$ et

$$V = H_{k,n} \ : x \longmapsto \frac{\exp\left(H_k - H_n\right)}{1-x} \exp\left(-\sum_{j=1}^{k} \frac{x^j}{j}\right).$$

On note que, pour $x$ dans $]-1,1[$

$$-\ln(1-x) = \sum_{j=1}^{+\infty} \frac{x^j}{j}, \quad H_{k,n}(x) = \exp\left(H_k - H_n\right)\ \exp\left(\sum_{j=k+1}^{+\infty} \frac{x^j}{j}\right).$$

Les fonctions

$$x \longmapsto \sum_{j=k+1}^{\infty} \frac{x^j}{j} \quad \text{et} \quad x \longmapsto \sum_{j=k+1}^{n} \frac{x^j}{j}$$

vérifient

$$u(x) - v(x) \underset{x \to 0}{=} O(x^{n+1}).$$

Il s'ensuit que $U = e^{H_k - H_n} e^u$ et $V = e^{H_k - H_n} e^v$ vérifient

$$\forall j \in \{0, \ldots, n\}, \qquad U^{(j)}(0) = V^{(j)}(0).$$

C'est dire que le coefficient de $x^i$ dans le développement en série entière de $G_{T_{k,n}}$ est celui de $x^i$ dans le développement en série entière de $H_{k,n}$.

b) La fonction

$$W \ : x \longmapsto \exp\left(-\sum_{j=1}^{k} \frac{x^j}{j}\right)$$

est développable en série entière de rayon infini, comme produit des fonctions

$$x \longmapsto \exp\left(-\frac{x^j}{j}\right) = \sum_{k=0}^{+\infty} \frac{(-1)^k x^{kj}}{j^k k!}.$$

On peut lui appliquer la question 7 : si $c_p$ est le $p$-ième coefficient du développement en série entière de

$$x \longmapsto \frac{1}{1-x} \exp\left(-\sum_{j=1}^{k} \frac{x^j}{j}\right),$$

alors $(c_p)$ converge vers $e^{-H_k}$ lorque $p$ tend vers $+\infty$.

Si $\ell$ est un élément de $\mathbb{N}$ et $n \geq \ell$, la question a) entraîne que $P(T_{k,n} = n - \ell)$ est le coefficient d'indice $n - \ell$ de $H_{k,n}$. Ainsi

$$P(T_{k,n} = n - \ell) = e^{H_k - H_n} c_{n-\ell}.$$

On a donc

$$P(T_{k,n} = n - \ell) \underset{n \to +\infty}{\sim} e^{-H_n} \underset{n \to +\infty}{\sim} \frac{e^{\gamma}}{n}.$$

14. En combinant les questions 12 et 13.b), on voit que, pour $c_1, \ldots, c_k$ dans $\mathbb{N}$ :

$$P\left(C_n^1 = c_1, \ldots, C_n^k = c_k\right) \underset{n \to +\infty}{\longrightarrow} P(Z_1 = c_1, \ldots, Z_k = c_k).$$

C'est dire que $(C_n^1, \ldots, C_k^n)$ converge en loi vers $(Z_1, \ldots, Z_k)$. Asymptotiquement, $(C_n^1, \ldots, C_k^n)$ se comporte donc comme un $k$-uplet de variables de Poisson indépendantes de paramètres respectifs $1, 1/2, \ldots, 1/k$.

La première marge de $(C_n^1, \ldots, C_n^k)$ est $C_n^1$. On retrouve bien le résultat de la question 3.