Agrégation externe de mathématiques
Épreuve de modélisation, option A : Probabilités et Statistiques

(*Texte public*)

Résumé : On présente quatre algorithmes de génération de permutations aléatoires. Les trois premiers construisent des permutations quelconques et le quatrième produit des permutations à cycles de longueur paire. L'analyse de ces algorithmes permet d'étudier les lois de certaines quantités liées à ces permutations : rangs relatifs, nombres de points fixes, nombres de cycles, longueurs de ces cycles.

Mots clefs : Cycle, permutation, loi de Poisson, fonction caractéristique, fonction génératrice, convergence en loi

► *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. Vous êtes laissé(e) libre d'organiser votre discussion comme vous l'entendez. Des suggestions de développement, largement indépendantes les unes des autres, vous sont proposées en fin de texte. Vous n'êtes pas tenu(e) de les suivre. Il vous est conseillé de mettre en lumière vos connaissances à partir du fil conducteur constitué par le texte. Le jury appréciera que la discussion soit accompagnée d'exemples traités sur ordinateur.*

Nous étudions quelques techniques de réalisation d'une permutation aléatoire de loi uniforme sur le groupe symétrique S_n.

1. Rangs relatifs ; points fixes

Fixons un entier $n \geq 2$ et considérons des variables aléatoires indépendantes X_1, …, X_n telles que chaque X_k soit de loi uniforme sur $\{1,\dots,k\}$. Voici tout d'abord deux façons de construire une permutation σ de $\{1,\cdots,n\}$ de loi uniforme sur S_n à partir de $X_1,\cdots,X_n$.

La première consiste à poser $\sigma(n) = X_n$, puis par récurrence descendante sur $1 \leq i \leq n-1$ et en écrivant $\{1,\cdots,n\}\backslash\{\sigma(n),\cdots,\sigma(i+1)\} = \{j_1,\cdots,j_i\}$ à définir $\sigma(i) = j_{X_i}$. On vérifie que σ est bien de loi uniforme. Cela fournit par exemple une information sur les rangs relatifs :

Proposition 1. *Pour tout $1 \leq k \leq n$, la position relative de $\sigma(k)$ parmi $\{\sigma(1),\cdots,\sigma(k)\}$ est de loi uniforme sur $\{1,\dots,k\}$.*

Une autre méthode est de partir de σ_0 la permutation identité. Pour tout $1 \leqslant k \leqslant n$, on définit alors la permutation σ_k de la manière suivante : en notant τ_k la transposition échangeant k et X_k (τ_k est l'identité si $X_k = k$), on pose $\sigma_k = \sigma_{k-1} \circ \tau_k$. Par récurrence sur n et en examinant l'image de n, on vérifie que toute permutation de $\{1,\dots,n\}$ s'écrit de façon unique sous la forme $(1\ x_1) \circ \dots \circ (n\ x_n)$ avec, pour tout $k \in \{1,\dots,n\}$, $x_k \in \{1,\dots,k\}$. Ainsi, la permutation finale $\sigma = \sigma_n$ qu'on obtient par l'algorithme ci-dessus peut être n'importe

quelle permutation, et on ne peut obtenir une permutation τ donnée que d'une seule façon. Ainsi pour toute permutation donnée τ,

$$\mathbb{P}(\sigma = \tau) = \frac{1}{n!}.$$

Considérons à présent la loi du nombre de nombre de points fixes, c'est-à-dire la loi du nombre d'entiers $1 \leqslant k \leqslant n$ tels que $\sigma(k) = k$. Remarquons tout d'abord que si Q est une partie de cardinal m, il y a exactement $(n-m)!$ permutations σ fixant les éléments de Q. Ainsi

$$\mathbb{P}(\forall i \in Q, \sigma(i) = i) = \frac{1}{n(n-1)\cdots(n-(m-1))}. \tag{1}$$

Notons A_n le nombre de points fixes. On a le résultat suivant :

Théorème 1. *(i) Soit f_n la fonction génératrice de A_n : pour z complexe,*

$$f_n(z) = \mathbb{E}\left[z^{A_n}\right] = \sum_{k=0}^{n} \mathbb{P}(A_n = k) z^k.$$

Alors

$$f_n(z) = \sum_{j=0}^{n} \frac{(z-1)^j}{j!}.$$

(ii) Quand $n \to \infty$, A_n converge en loi vers une loi de Poisson de paramètre 1 :

$$\forall i \geq 0,\ \mathbb{P}(A_n = i) \to_{n\to\infty} \frac{e^{-1}}{i!}.$$

Esquisse de la démonstration.
Pour un entier j, $0 \leqslant j \leqslant n-1$, calculons

$$\mathbb{E}[A_n(A_n-1)\dots(A_n-j)].$$

Soit $\mathbf{F}_n$ l'ensemble des points fixes de σ et rappelons donc que A_n est le cardinal de $\mathbf{F}_n$. Soit $F_n(j)$ le nombre de sous-ensembles de $\mathbf{F}_n$ de cardinal $j+1$. Alors

$$F_n(j) = \frac{A_n(A_n-1)\dots(A_n-j)}{(j+1)!}.$$

Soit $P_j(n)$ l'ensemble des parties à $j+1$ éléments de $\{1,2,\dots n\}$. On peut aussi écrire :

$$F_n(j) = \sum_{Q\in P_j(n)} \mathbf{1}_{\{\forall i\in Q,\ \sigma(i)=i\}}.$$

Il en résulte que

$$\mathbb{E}[A_n(A_n-1)\dots(A_n-j)] = (j+1)! \sum_{Q\in P_j(n)} \mathbb{P}(\forall i \in Q,\ \sigma(i) = i).$$

En utilisant (1), on en déduit

$$\mathbb{E}[A_n(A_n-1)\dots(A_n-j)] = 1$$

pour j vérifiant $0 \leqslant j \leqslant n$. Il s'ensuit que f_n vérifie $f_n^{(j)}(1) = 1$ pour $0 \leqslant j \leqslant n$ et $f_n^{(j)}(1) = 0$ pour $j > n$, d'où (i).

On déduit de (i) que pour tout complexe z, $f_n(z) \to e^{z-1}$ quand $n \to \infty$. Par conséquent, la fonction caractéristique de A_n converge vers la fonction caractéristique de la loi de Poisson quand $n \to \infty$, ce qui prouve (ii). □

2. Nombre de cycles

On introduit ici un nouvel algorithme qui va permettre d'étudier le nombre de cycles de la permutation et la longueur du cycle contenant 1. On construit la permutation aléatoire σ de la manière suivante. On choisit d'abord $\sigma(1)$, puis $\sigma(\sigma(1))$, etc, jusqu'à ce que le cycle de 1 soit bouclé. Puis on continue en construisant un autre cycle, etc. Formellement, on pose

$$m_1 = r_1 = 1,$$
$$I_1 = \{1, 2, \dots n\}.$$

À la k-ième étape, m_k désigne le point pour lequel on va choisir une image, I_k l'ensemble des images possibles et r_k le point d'origine du cycle en construction. On itère alors n fois la procédure suivante :

À la k-ième itération, on choisit $\sigma(m_k)$ au hasard, uniformément dans l'ensemble I_k, indépendamment du passé. Puis on pose :

— $I_{k+1} = I_k - \{\sigma(m_k)\}$,
— si $\sigma(m_k) \neq r_k$, alors $m_{k+1} = \sigma(m_k)$ et $r_{k+1} = r_k$,
— si $\sigma(m_k) = r_k$, alors $m_{k+1} = r_{k+1} = \min I_{k+1}$.

Puis on passe à la $(k+1)$-ième itération.

On a alors :

Proposition 2. *La permutation σ obtenue après n itérations est une permutation aléatoire uniforme sur l'ensemble des permutations de* $\{1, 2, \dots, n\}$.

Une manière de démontrer cette proposition consiste à considérer une permutation quelconque ρ de $\{1, 2, \dots, n\}$ et à vérifier que

$$\mathbb{P}(\sigma = \rho) = \frac{1}{n!},$$

ce qu'on fait en observant que $\mathbb{P}(\sigma(1) = \rho(1)) = \frac{1}{n}$ et que pour tout $k \in \{2, \dots, n\}$,

$$\mathbb{P}(\sigma(k) = \rho(k) | \sigma(1) = \rho(1), \dots, \sigma(k-1) = \rho(k-1)) = \frac{1}{\#I_k} = \frac{1}{n-k+1}.$$

Soit U_n la taille du cycle contenant 1 et C_n le nombre de cycles dans la permutation σ.

Théorème 2. *(i) La variable aléatoire U_n est uniformément répartie sur* $\{1, 2, \dots n\}$. *Ainsi quand $n \to \infty$, la loi de U_n / n converge vers la loi uniforme sur* $[0, 1]$.

(ii) Quand $n \to \infty$, $\mathbb{E}(C_n) \sim \log n$.

(iii) Quand $n \to \infty$, $Var(C_n) \sim \log n$.

Esquisse de la démonstration.

On dit que $k \in \{1,2,\dots,n\}$ est un temps de boucle si $\sigma(m_k) = r_k$. On voit que k est un temps de boucle si un cycle est complété à la k-ième itération de l'algorithme. On note E_k l'événement que k est un temps de boucle.

Alors on montre que $\mathbb{P}(E_k) = 1/(n-k+1)$ et que la famille d'événements $(E_k, 1 \leqslant k \leqslant n)$ est une famille indépendante d'événements. Or on a $U_n = k$ si et seulement si $1,2,\dots,k-1$ ne sont pas des temps de boucle et k est un temps de boucle, d'où (i).

Pour prouver (ii) et (iii), on définit, pour $1 \leqslant i \leqslant n$, les variables aléatoires $X_i = \mathbf{1}_{E_i}$: les X_i sont des variables de Bernoulli indépendantes avec

$$\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = \frac{1}{n-i+1}.$$

Or on remarque que

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

3. Permutations à cycles de longueur paire

On veut construire maintenant une permutation aléatoire τ de $\{1,2,\dots,2n\}$ n'ayant que des cycles de longueur paire. L'idée est la suivante. D'abord on choisit $\tau(1)$, qui ne peut être égal à 1, donc on a $2n-1$ choix possibles. Puis on choisit $\tau(\tau(1))$, qui peut être égal à 1 mais ne peut pas être égal à $\tau(1)$, donc de nouveau on a $2n-1$ choix possibles. On continue ainsi jusqu'à ce qu'on ait "bouclé" le cycle contenant 1. Puis on passe au cycle suivant, etc.

Formellement, on pose

$$m_1 = r_1 = 1,$$

$$I_1 = \{1,2,\dots,2n\}.$$

On itère alors n fois la procédure suivante :

À la k-ième itération, on choisit $\tau(m_k)$ au hasard, uniformément dans l'ensemble $I_k - \{r_k\}$, indépendamment du passé. Puis on choisit $\tau(\tau(m_k))$ au hasard, uniformément dans l'ensemble $I_k - \{\tau(m_k)\}$, indépendamment du passé. On pose alors :

— $I_{k+1} = I_k - \{\tau(m_k), \tau(\tau(m_k))\}$,
— si $\tau(\tau(m_k)) \neq r_k$, alors $m_{k+1} = \tau(\tau(m_k))$ et $r_{k+1} = r_k$,
— si $\tau(\tau(m_k)) = r_k$, alors $m_{k+1} = r_{k+1} = \min I_{k+1}$.

Puis on passe à la $(k+1)$-ième itération. On a alors :

Proposition 3. *(i) La permutation σ obtenue après n itérations est une permutation aléatoire uniforme sur l'ensemble des permutations de $\{1,2,\dots,2n\}$ n'ayant que des cycles de longueur paire.*

(ii) Le nombre de permutations de $\{1,2,\dots,2n\}$ n'ayant que des cycles de longueur paire est

$$1^2 \times 3^2 \times \dots \times (2n-1)^2.$$

(iii) Si on choisit au hasard, uniformément, une permutation de $\{1,2,\ldots,2n\}$*, la probabilité que cette permutation n'ait que des cycles de longueur paire est équivalente à* $1/\sqrt{\pi n}$ *quand* $n \to \infty$*.*

Comme pour la Proposition 1, une manière de démontrer (i) dans la Proposition 2 consiste à considérer une permutation quelconque ρ de $\{1,2,\ldots,2n\}$ n'ayant que des cycles de longueur paire et à vérifier que

$$\mathbb{P}(\tau = \rho) = \frac{1}{1^2 \times 3^2 \times \ldots \times (2n-1)^2}.$$

On déduit aussi (ii) de l'égalité ci-dessus. Enfin, (iii) résulte de (ii) et de la formule de Stirling : $n! \sim (n/e)^n \sqrt{2\pi n}$.

On peut étudier la taille V_n du cycle contenant 1 et le nombre D_n de cycles dans la permutation τ.

Théorème 3. *(i)* $\mathbb{P}(V_n = 2) = 1/(2n-1)$ *et pour* $2 \leqslant k \leqslant n$*,*

$$P(V_n = 2k) = \frac{1}{2n-1} \times \frac{2n-2}{2n-3} \times \ldots \times \frac{2n-2k+2}{2n-2k+1}.$$

On en déduit que quand $n \to \infty$*,* $V_n/2n$ *converge en loi vers la loi de densité* $1/(2\sqrt{1-x})$ *à support* $[0,1]$*. Autrement dit, pour tout* $y \in [0,1]$*, quand* $n \to \infty$*,* $\mathbb{P}(V_n/2n \geq 1-y) \to \sqrt{y}$

(ii) Quand $n \to \infty$*,* $\mathbb{E}(D_n) \sim \frac{1}{2}\log n$*.*

(iii) Quand $n \to \infty$*,* $Var(D_n) \sim \frac{1}{2}\log n$*.*

La démonstration est analogue à celle du Théorème 1. On dit que l'événement E_k a lieu si et seulement si $\tau(\tau(m_k)) = r_k$. On montre que $\mathbb{P}(E_k) = 1/(2n-2k+1)$ et que $V_n = 2k$ si et seulement si E_k a lieu mais que E_i n'a pas lieu pour $i < k$, d'où (i). Enfin on montre que les E_k sont des événements indépendants pour (ii) et (iii).

Suggestions pour le développement

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

— *Modélisation.*

- Disposer d'une permutation aléatoire peut s'avérer utile, par exemple pour déterminer l'ordre de passage des candidats lors d'un concours. On pourra envisager d'autres applications. On pourra également proposer d'autres méthodes.
- Comme application du premier chapitre, imaginons qu'on veuille simuler une permutation sans point fixe, aléatoire et uniformément distribuée sur l'ensemble des permutations sans point fixe. Une méthode est de tirer une permutation au hasard

et de la garder si elle n'a pas de point fixe. Sinon, on la rejette et on tire une nouvelle permutation au hasard, etc jusqu'à tomber sur une permutation sans point fixe. Combien de permutations devra-t-on tirer en moyenne pour obtenir ce qu'on cherche ?

- Plus généralement, comment faire pour simuler une permutation aléatoire ayant exactement k points fixes, $k \leqslant n$ fixé ?
- La méthode proposée au dernier chapitre permet-elle de construire des permutations de $\{1,\dots,3n\}$ dont les cycles ont toujours une longueur multiple de 3 ?

— *Etude numérique.* On pourra simuler des permutations aléatoires suivant les différentes méthodes décrites et illustrer les convergences en loi du nombre de points fixes et de la longueur du cycle contenant 1. On pourra aussi simuler directement la variable C_n et observer, au-delà de la moyenne, une éventuelle convergence en loi (après renormalisation).

— *Développements mathématiques.* On peut étudier les démonstrations des théorèmes énoncés, notamment :

- prouver que les variables sont bien de loi uniforme sur S_n ;
- détailler la démonstration des théorèmes 1, 2 et 3.