

Statistiques sur les permutations aléatoires
Groupe de Travail Applicatif 2019/2020

Théophile Baranger
David de Carvalho
Yassine Laamoumri
Exaucé Luweh Adjim Ngarti

30 avril 2020

Enseignant : Bernard Bercu

Table des matières

1	Introduction	2
2	Nombre de descentes	2
2.1	Définition	2
2.2	Espérance	3
2.3	Variance	4
2.4	Simulations	4
3	Nombres de segments alternés	6
3.1	Définition	6
3.2	Espérance	7
3.3	Variance	8
3.4	Simulations	9
4	Plus longue sous-suite alternée	10
4.1	Définition	10
4.2	Lien avec les segments alternés	10
5	Oscillation	10
5.1	Définition	11
5.2	Simulations	11
5.3	Calculs	11
5.4	Probabilités	13
5.5	Espérance	14
	Références	15

1 Introduction

Dans ce rapport, nous allons nous intéresser aux statistiques concernant les permutations aléatoires. On s'intéressera spécifiquement aux variables aléatoires comptant le nombre de descentes, de pics, ou encore à la longueur de la plus grande sous-suite alternée dans une permutation aléatoire. La dernière partie sera dédiée à l'étude de l'oscillation d'une permutation aléatoire.

L'objectif principal de notre projet est d'effectuer des simulations afin de visualiser les convergences, généralement des moyennes empiriques associées à ces variables aléatoires, ainsi que leur normalité asymptotique. Ainsi, les preuves de ces convergences, souvent assez techniques et se reposant principalement sur des outils d'analyse, seront omises mais néanmoins référencées. Nous inclurons également des calculs d'espérances et de variances dans l'objectif de valider ces résultats attendus, et de les confronter aux simulations.

Enfin, l'idée étant de visualiser ces diverses convergences, nous inclurons souvent des graphiques, générés grâce à la bibliothèque Matplotlib sous Python, ainsi que le cas échéant des illustrations construites grâce au logiciel Tulip, développé au Laboratoire Bordelais de Recherche en Informatique (LaBRI), qui nous permettra de représenter de façon intuitive certaines des variables aléatoires mentionnées précédemment.

Ce travail a été effectué collectivement. Nous avons en général utilisé la plateforme Github pour conserver une version à jour de notre travail, et les réseaux sociaux pour communiquer régulièrement, en plus d'entretiens hebdomadaires avec M. Bercu.

2 Nombre de descentes

Dans cette première partie, nous allons nous intéresser à certaines propriétés propres aux descentes dans les permutations aléatoires.

2.1 Définition

Soit S_n l'ensemble des permutations des n premiers entiers, avec $n \geq 1$. Soit π_n une permutation de S_n . On dit que π_n a une descente à la position i si $\pi_n(i+1) < \pi_n(i)$. Dans la figure 1, on représente la permutation 2413 de S_4 , qui contient une descente à la position $i = 2$.

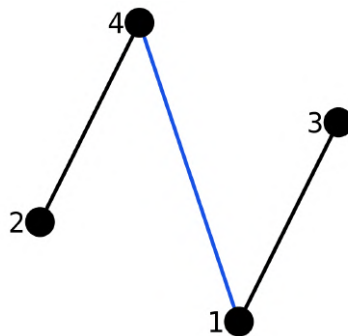


FIGURE 1 – La permutation 2413 contient une descente.

2.2 Espérance

On note D_n la variable aléatoire comptant le nombre de descentes d'une permutation choisie aléatoirement dans S_n . On a alors $D_1 = 0$ et pour tout $n \geq 2$,

$$D_n(\Omega) = \{0, 1, \dots, n-1\}$$

En particulier, on a $\mathbb{P}(D_2 = 0) = \mathbb{P}(D_2 = 1) = 1/2$. On s'intéresse au comportement asymptotique de la suite (D_n) . On peut montrer [6][1], que pour tout $n \geq 1$,

$$\begin{cases} \mathbb{P}(D_{n+1} = D_n + 1 \mid D_n) &= \frac{n - D_n}{n + 1} \\ \mathbb{P}(D_{n+1} = D_n \mid D_n) &= \frac{D_n + 1}{n + 1} \end{cases}$$

On en déduit la décomposition

$$D_{n+1} = D_n + \varepsilon_{n+1}$$

avec

$$\varepsilon_{n+1} = \begin{cases} 1 & \text{avec probabilité } p_n \\ 0 & \text{avec probabilité } 1 - p_n \end{cases}$$

où

$$p_n = \frac{n - D_n}{n + 1}$$

On commence par calculer l'espérance et la variance de D_n , en se reposant sur la formulation récurrente ci-dessus, et qui définit une martingale.

Pour le calcul de l'espérance de D_n , nous allons nous reposer sur la formule d'espérance totale,

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

Si l'on reprend la forme récurrente de D_n , on a

$$\begin{aligned} \mathbb{E}[D_{n+1} \mid D_n] &= \mathbb{E}[D_n + \varepsilon_{n+1} \mid D_n] \\ &= \mathbb{E}[D_n \mid D_n] + \mathbb{E}[\varepsilon_{n+1} \mid D_n] \\ &= D_n + p_n \\ &= D_n + \frac{n - D_n}{n + 1} \\ &= \frac{n}{n + 1} D_n + \frac{n}{n + 1} \end{aligned}$$

Si l'on calcule l'espérance de cette expression, on obtient alors

$$\mathbb{E}[D_{n+1}] = \frac{n}{n + 1} \mathbb{E}[D_n] + \frac{n}{n + 1}$$

Et on peut en déduire le terme général de la suite $\mathbb{E}[D_n]$, grâce à la formule suivante, dans le cas où une suite $(d_n)_{n \in \mathbb{N}^*}$ est définie pour tout $n \in \mathbb{N}^*$ par la relation $d_{n+1} = a_n \cdot d_n + b_n$, et d_1 fixé, on a

$$d_n = d_1 \cdot \prod_{k=1}^{n-1} a_k + \sum_{k=1}^{n-1} \left(\prod_{i=k+1}^{n-1} a_i \right) \cdot b_k$$

En l'occurrence, avec les suite $a_n = b_n = \frac{n}{n+1}$, on obtient (après calculs),

$$d_n = \frac{n-1}{2}, \quad \text{d'où} \quad \mathbb{E} \left(\frac{D_n}{n} \right) \xrightarrow{n \rightarrow +\infty} \frac{1}{2}$$

2.3 Variance

La même méthode que pour le moment d'ordre 1 peut être utilisée pour calculer le moment d'ordre 2, à la différence que les suites qui apparaissent dans la relation de récurrence sont un peu plus compliquées. On en déduira ensuite la variance de D_n (un exemple de calcul détaillé pour un moment d'ordre 2 sera donné dans la section suivante).

Comme pour l'espérance, on obtient une relations de récurrence pour le moment d'ordre 2,

$$\mathbb{E}[D_{n+1}^2] = \frac{n-1}{n+1} \mathbb{E}[D_n^2] + \frac{2n^2 - n + 1}{2(n+1)}$$

On en déduit par la même méthode que précédemment,

$$\text{Var}(D_n) = \frac{n+1}{12}, \quad \text{d'où} \quad \frac{\text{Var}(D_n)}{n} \xrightarrow{n \rightarrow +\infty} \frac{1}{12}$$

2.4 Simulations

On se propose maintenant de visualiser les simulations effectuées sur la base de cette formulation récurrente du nombre de descentes. Nous allons plus particulièrement tenter de visualiser les convergences suivantes [6], démontrées par exemple par Pitman [5] en ce qui concerne la normalité asymptotique (on se réfère à Chatterjee & Diaconis [4] pour un rappel des différentes approches possibles).

$$\frac{D_n}{n} \xrightarrow{\text{p.s.}} \frac{1}{2}$$

$$\sqrt{12n} \left(\frac{D_n}{n} - \frac{1}{2} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Commençons par présenter l'algorithme utilisé pour construire une suite $(\frac{D_n}{n})$.

L'idée est de se reposer sur la martingale définie plus haut plutôt que de générer des permutations aléatoires et d'effectuer des opérations sur celles-ci dans un second temps. En terme de complexité de calcul, cette méthode est beaucoup plus efficace, et l'algorithme traduit aussi cette simplicité. On note que l'indexation du tableau utilisé pour stocker les termes de la suite commence à 1, ce qui nous permet de simplement faire correspondre chaque $\frac{D_n}{n}$ à l'élément $T[n]$.

Algorithm 1 Suite des $(\frac{D_n}{n})$

```
1:  $T \leftarrow \text{creerTableau}(n)$ 
2:  $D_n \leftarrow 0$ 
3:  $T[1] \leftarrow D_n$ 
4: for  $i$  de 2 à  $n$  do
5:    $p_n \leftarrow \text{uniforme}(0, 1)$ 
6:   if  $p_n < (i - D_n)/(i + 1)$  then
7:      $D_n = D_n + 1$ 
8:   end if
9:    $T[i] \leftarrow (D_n)/(i)$ 
10: end for
```

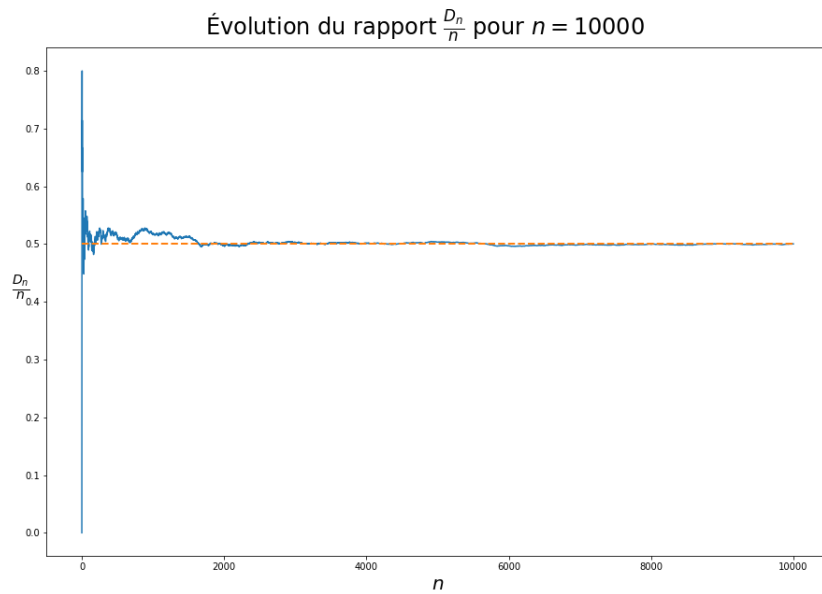


FIGURE 2 – Simulation pour $(\frac{D_n}{n})_{n \geq 1}$ avec $n = 10000$. On observe une convergence vers la valeur attendue, c'est-à-dire $\frac{1}{2}$, ce qui correspond aux énoncés précédents.

L'algorithme qui consiste à générer une suite des $\sqrt{12n}(\frac{D_n}{n} - \frac{1}{2})$ est très similaire à l'algorithme précédent, à la différence que nous allons répéter cette simulation un nombre suffisamment important de fois afin de produire une distribution satisfaisante. On prend par exemple une suite de $n = 10000$ éléments, itérée $m = 1000$ fois.

Algorithm 2 Suite des $(\sqrt{12n}(\frac{D_n}{n} - \frac{1}{2}))$

```
1:  $T \leftarrow \text{creerTableau}(n)$ 
2:  $D_n \leftarrow 0$ 
3:  $T[1] \leftarrow D_n$ 
4: for  $i$  de 2 à  $n$  do
5:    $p_n \leftarrow \text{uniforme}(0, 1)$ 
6:   if  $p_n < (i - D_n)/(i + 1)$  then
7:      $D_n = D_n + 1$ 
8:   end if
9:    $T[i] \leftarrow (D_n)/(i)$ 
10: end for
```

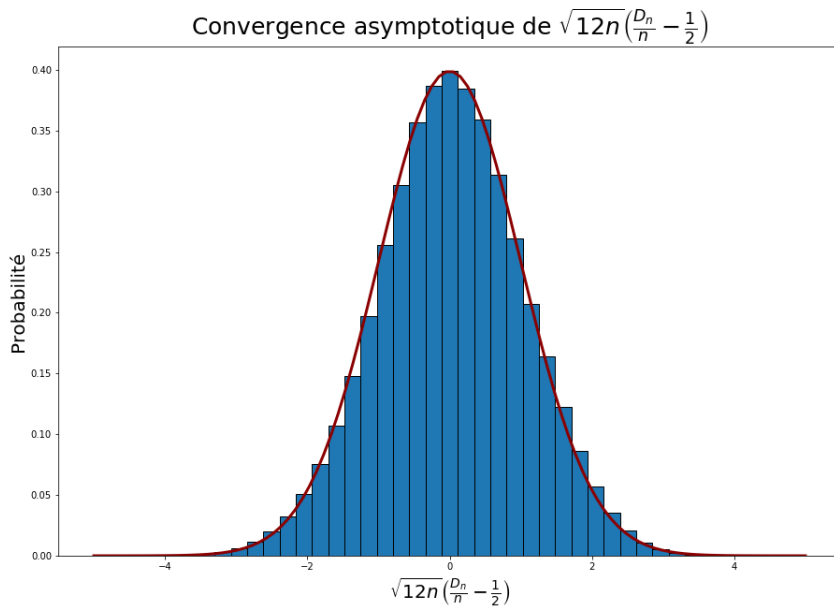


FIGURE 3 – Simulation pour $\sqrt{12n}(\frac{D_n}{n} - \frac{1}{2})$ avec $n = 10000$ itérations et $m = 1000$ répétitions. On observe une convergence vers la loi normale centrée réduite. Il s'agit en fait de la conséquence d'un théorème central limite valable pour les martingales.

3 Nombres de segments alternés

L'introduction mentionne le nombre de pics d'une permutation aléatoire. En fait, il semblerait que dans la littérature, il soit plus commun de s'intéresser au nombre de segments alternés, *alternating runs* en anglais. Il ne s'agit que d'un problème de définition, comme nous allons le voir, et les résultats restent les mêmes.

3.1 Définition

On dit qu'une permutation aléatoire π_n dans S_n possède un pic à la position i si $\pi_n(i-1) < \pi_n(i) > \pi_n(i+1)$ ou si $\pi_n(i-1) > \pi_n(i) < \pi_n(i+1)$. On dit qu'une permutation aléatoire possède k segments alternés si elle possède $k-1$ pics. Il arrive

également que l'on distingue les pics des vallées (*peaks* et *troughs* en anglais), mais cette distinction ne nous intéresse pas ici.

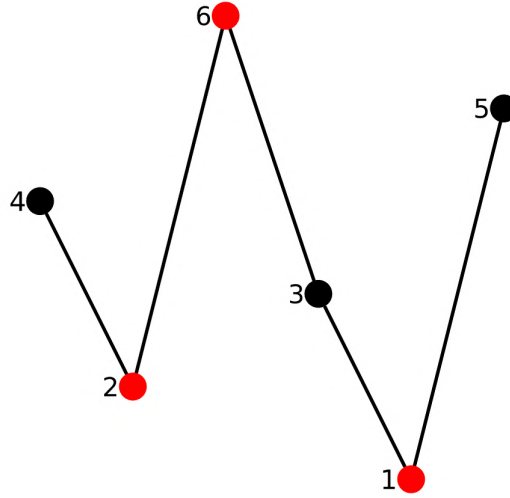


FIGURE 4 – La permutation 426315 contient trois pics, donc 4 segments alternés.

On note alors G_n la variable aléatoire qui compte le nombre de segments alternés d'une permutation, et on observe que $G_1 = 0$, $G_2 = 1$, $\mathbb{P}(G_3 = 1) = 1/3$ et $\mathbb{P}(G_3 = 2) = 2/3$.

Comme dans le cas des descentes, on peut déterminer [6] les probabilités conditionnelles suivantes, pour tout $n \geq 1$,

$$\begin{cases} \mathbb{P}(G_{n+1} = G_n + 2 \mid G_n) &= \frac{n - 1 - G_n}{n + 1} \\ \mathbb{P}(G_{n+1} = G_n + 1 \mid G_n) &= \frac{2}{n + 1} \\ \mathbb{P}(G_{n+1} = G_n \mid G_n) &= \frac{G_n}{n + 1} \end{cases}$$

On en déduit là encore une décomposition de la forme

$$G_{n+1} = G_n + \varepsilon_{n+1}$$

avec

$$\varepsilon_{n+1} = \begin{cases} 2 & \text{avec probabilité } p_n \\ 1 & \text{avec probabilité } q_n \\ 0 & \text{avec probabilité } 1 - p_n - q_n \end{cases}$$

où

$$p_n = \frac{n - 1 - G_n}{n + 1} \quad q_n = \frac{2}{n + 1}$$

3.2 Espérance

Les probabilités conditionnelles mentionnées ci-dessus nous permettent d'établir la relation de récurrence suivante pour l'espérance de G_n . On obtient ainsi

$$\mathbb{E}[G_{n+1}] = \frac{n - 1}{n + 1} \cdot \mathbb{E}[G_n] + \frac{2n}{n + 1}$$

D'où l'on déduit

$$E[G_n] = \frac{2n-1}{3}$$

On note par ailleurs que ce résultat se devine facilement en notant [6] que la probabilité de changer de direction à chaque position est en général $2/3$. Pour s'en convaincre, on peut visualiser chaque segment de taille 3 comme une permutation de taille 3, et le résultat devient alors évident.

3.3 Variance

On commence par calculer $E[G_n^2]$ par la même méthode que précédemment. Nous allons cette fois-ci détailler les calculs.

$$\begin{aligned} E[G_{n+1}^2 \mid G_n] &= E[(G_n + \varepsilon_{n+1})^2 \mid G_n] \\ &= (G_n^2 + 4G_n + 4)p_n + (G_n^2 + 2G_n + 1)q_n + G_n^2(1 - p_n - q_n) \\ &= \frac{1}{n+1} [(G_n^2 + 4G_n + 4)(n-1-G_n) + 2(G_n^2 + 2G_n + 1) + G_n^3] \\ &= \frac{1}{n+1} [(n-3)G_n^2 + (4n-4)G_n + 4n-2] \end{aligned}$$

En appliquant la **formule d'espérance totale**, on obtient

$$\begin{aligned} E[G_{n+1}^2] &= \frac{n-3}{n+1} E[G_n^2] + \frac{4n-4}{n+1} \cdot \frac{2n-1}{3} + \frac{4n-2}{n+1} \\ &= \frac{n-3}{n+1} E[G_n^2] + \frac{2}{3} \cdot \frac{4n^2-1}{n+1} \end{aligned}$$

On arrive donc ensuite à obtenir, pour tout $n \geq 4$

$$\begin{aligned} E[G_n^2] &= \frac{24(n-4)!}{n!} \left[\frac{35}{6} + \sum_{k=4}^{n-1} \left(\frac{2}{3} \cdot \frac{4k^2-1}{k+1} \cdot \frac{(k+1)k(k-1)(k-2)}{24} \right) \right] \\ &= \frac{2}{3} \cdot \frac{(n-4)!}{n!} \left[\sum_{k=3}^{n-1} (2k-1)(2k+1)k(k-1)(k-2) \right] \\ &= \frac{1}{90} (8n(5n-3) - 19) \end{aligned}$$

On peut maintenant calculer la variance de G_n en appliquant la définition

$$\begin{aligned} \text{Var}(G_n) &= E[G_n^2] - E[G_n]^2 \\ &= \frac{1}{90} (8n(5n-3) - 19) - \left(\frac{2n-1}{3} \right)^2 \\ &= \frac{8n}{45} - \frac{29}{90} \end{aligned}$$

3.4 Simulations

Les algorithmes utilisés pour générer les suites de variables aléatoires étant très similaires à ceux utilisés dans le cas des descentes, on propose de simplement visualiser les résultats obtenus, et qui illustrent les convergences

$$\frac{G_n}{n} \xrightarrow{\text{p.s.}} \frac{2}{3}$$

$$\sqrt{n} \left(\frac{D_n}{n} - \frac{1}{2} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{8}{45} \right)$$

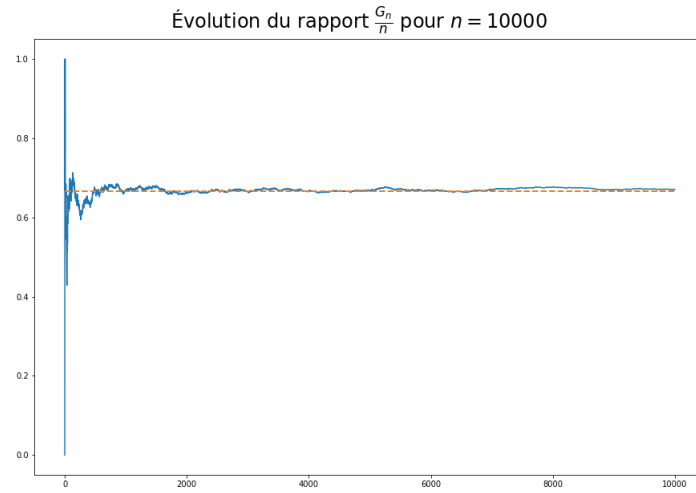


FIGURE 5 – Simulation pour $(\frac{G_n}{n})_{n \geq 1}$ avec $n = 10000$. On observe une convergence vers la valeur attendue, c'est-à-dire $\frac{2}{3}$.

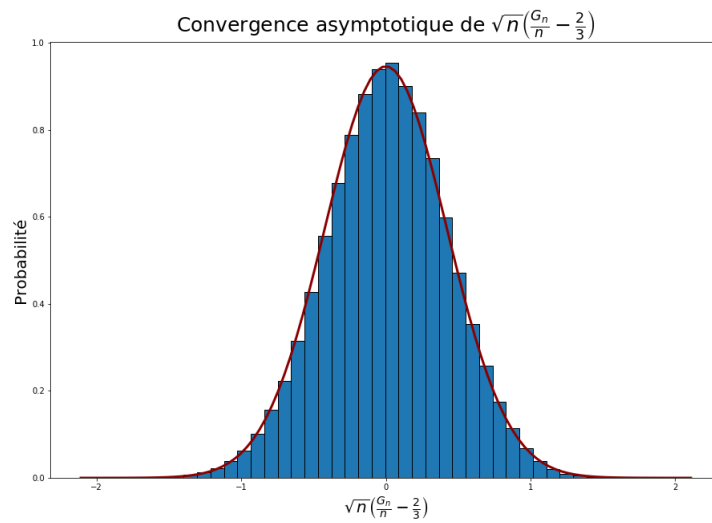


FIGURE 6 – Simulation pour $\sqrt{n}(\frac{G_n}{n} - \frac{2}{3})$ avec $n = 10000$ itérations et $m = 1000$ répétitions.

4 Plus longue sous-suite alternée

4.1 Définition

On définit une sous-suite alternée [1] dans une permutation aléatoire π_n de S_n comme une sous-suite de la forme $\pi_n(i_1)\pi_n(i_2)\cdots\pi_n(i_k)$ telle que

$$\pi_n(i_1) > \pi_n(i_2) < \pi_n(i_3) > \pi_n(i_4) < \cdots$$

On note que l'on distingue alors les sous-suite alternées des sous-suites alternées inverses dans ce sens que les dernières commencent par une montée, alors que les sous-suites alternées commencent par une descente. On se réfère à la section 1.3 dans l'ouvrage de Miklós Bóna [1] pour une discussion intéressante sur les propriétés des sous-suites alternées.

4.2 Lien avec les segments alternés

Les sous-suites alternées sont très proches du nombre de segments alternés tel que défini dans la section précédente. C'est pourquoi on passera l'expression récurrente de L_n , dont les probabilités conditionnelles sont quasi-identiques à celle du nombre de segments alternés.

Néanmoins, il est important de mentionner l'origine de ces similarités. En fait, on peut démontrer [1] [6] qu'il existe un lien fort entre ces deux variables. En notant $L_{n,k}$ le nombre de permutation de taille n dont la plus longue sous-suite alternée est de longueur k , et $G_{n,k}$ le nombre de permutations de taille n ayant k segments alternés, alors on a la relation suivante :

$$L_{n,k} = \frac{1}{2}(G_{n,k-1} + G_{n,k})$$

Les résultats étant ainsi très similaires au cas des segments alternés, on évitera la redondance des simulations, ce qui nous laissera l'opportunité de parler d'une quatrième statistique intéressante, l'oscillation d'une permutation aléatoire.

5 Oscillation

La dernière section est attribuée à l'étude d'une statistique qui nous a été suggérée par M. Bercu, appelée oscillation d'une permutation aléatoire. Des résultats ont été démontrés en 1993 par Chao, Bai et Liang [3], puis en 1997 par Chao [2]. Il va être question ici de les répliquer par simulation dans un premier temps, puis d'explorer la possibilité de représenter le phénomène par une martingale, telle que dans les sections précédentes.

Le travail présenté ici n'est pas tout à fait abouti, comme il va apparaître dans les sous-sections suivantes. L'une des raisons est que malgré des propriétés intéressantes, l'oscillation est moins facile à comprendre en terme de probabilités conditionnelles que les statistiques que nous avons vues précédemment.

5.1 Définition

On appelle l'oscillation d'une permutation π_n dans S_n la valeur suivante

$$\text{Osc}(\pi_n) = \sum_{i=1}^n |\pi_n(i) - \pi_n(i+1)|, \text{ avec } \pi_n(n+1) = \pi_n(1)$$

On note ainsi \mathcal{O}_n la variable aléatoire correspondant à l'oscillation d'une permutation aléatoire de longueur n . On a $\mathcal{O}_2 = 2$, et $\mathcal{O}_3 = 4$. Il est utile de voir qu'en réfléchissant de manière cyclique sur les permutations, le dernier terme ne modifiant pas le comportement asymptotique d'une suite (\mathcal{O}_n) normalisée, l'oscillation est symétrique et invariante par translation. Les résultats obtenus par Chao en 1993 sont les suivants [3] : $E(\mathcal{O}_n) = \frac{1}{3}n(n+1)$ et $\text{Var}(\mathcal{O}_n) = \frac{2}{45}n^3 + o(n^3)$.

Le résultat concernant l'espérance de \mathcal{O}_n est présenté comme élémentaire, mais la méthode n'est pas explicitée, donc nous allons tenter ici de le retrouver.

5.2 Simulations

Nous avons commencé par observer par simulation les résultats annoncés.

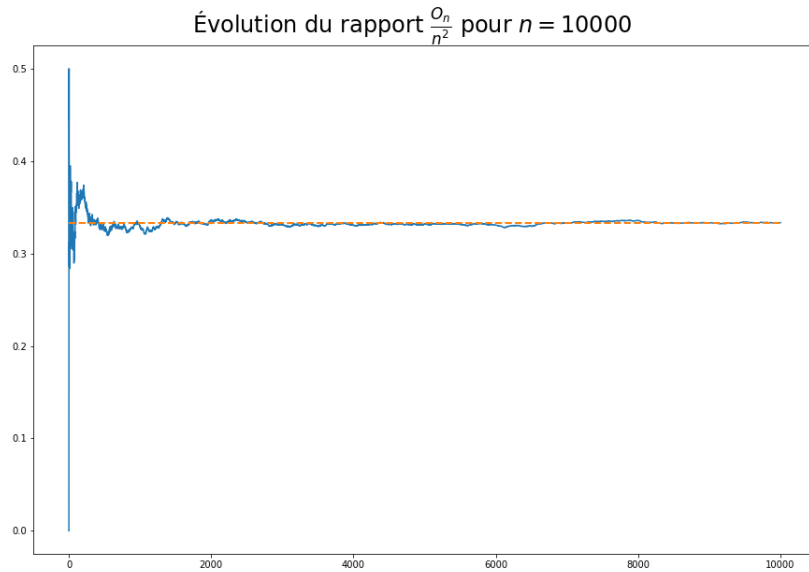


FIGURE 7 – Simulation pour $(\frac{\mathcal{O}_n}{n^2})_{n \geq 1}$ avec $n = 10000$. On observe une convergence vers la valeur attendue, c'est-à-dire $\frac{1}{3}$.

Les simulations confirment bien à priori les résultats, et donc par la suite nous avons tenté de les comprendre, pour tenter de les retrouver. En particulier, nous avons retrouvé le résultat concernant l'espérance de \mathcal{O}_n .

5.3 Calculs

Pour essayer de comprendre le comportement de la variable aléatoire \mathcal{O}_n , nous nous sommes intéressés à la valeur $\mathcal{O}_{n+1} - \mathcal{O}_n$. On pose donc $\varepsilon_{n+1} = \mathcal{O}_{n+1} - \mathcal{O}_n$, et on

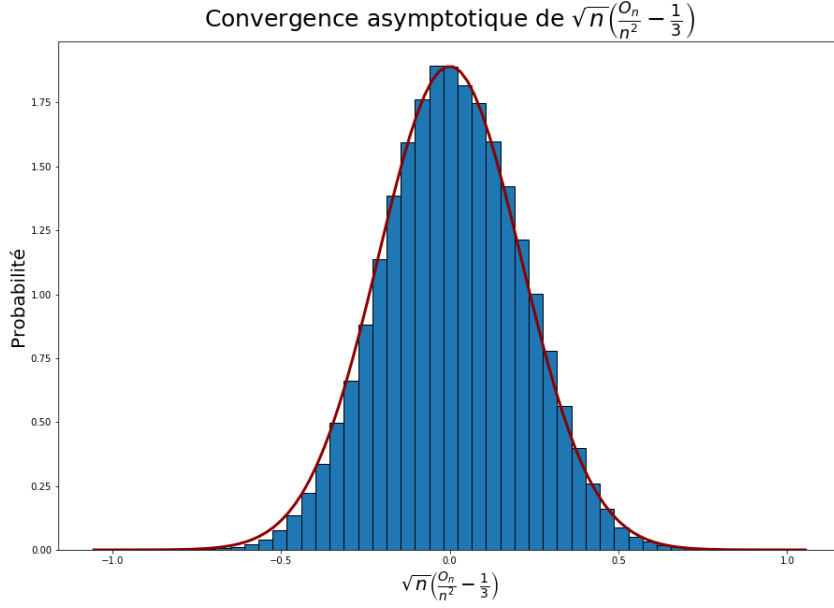


FIGURE 8 – Simulation pour $\sqrt{n}(\frac{\mathcal{O}_n}{n^2} - \frac{1}{3})$ avec $n = 5000$ itérations et $m = 500$ répétitions.

s'intéresse à la distribution de cette variable aléatoire. Une façon efficace de calculer cette valeur est de mettre à jour la valeur de l'oscillation d'une permutation lors de l'insertion de $n + 1$ de la façon suivante :

Algorithm 3 Calcul de $\varepsilon_{n+1} = \mathcal{O}_{n+1} - \mathcal{O}_n$

Entrée : Permutation au rang $n + 1$, Position de $n + 1$

Sortie : ε_{n+1}

1: **return** $\text{abs}(\text{P}[\text{m}]-\text{P}[\text{m}-1]) + \text{abs}(\text{P}[\text{m}]-\text{P}[\text{m}+1]) - \text{abs}(\text{P}[\text{m}-1]-\text{P}[\text{m}+1])$

On a volontairement omis pour une question de lisibilité de le préciser dans l'algorithme ci-dessus, mais les positions dans π_n sont calculées de manière cyclique, c'est à dire modulo $n + 1$. Dans la suite, $\pi_n(0) = \pi_n(n)$ et $\pi_n(n + 1) = \pi_n(1)$.

L'idée pour étudier la distribution des ε_{n+1} est de générer un nombre important de permutations de taille n , puis d'observer les valeurs prises par ε_{n+1} grâce à l'algorithme ci-dessus. Nous avons été au premier abord assez surpris d'observer une grande régularité, comme en témoigne la simulation représentée dans la figure 9.

Sur la base de cette distribution, avec $\varepsilon_n(\Omega) = 2, 4, \dots, 2(n - 2)$, on peut supposer que la probabilité $\mathbb{P}(\varepsilon_n = 2(n - 2))$ satisfait l'équation

$$\sum_{k=1}^{n-2} k \cdot \mathbb{P}(\varepsilon = 2(n - 2)) = 1$$

Et ainsi, on aurait

$$\mathbb{P}(\varepsilon_n = 2(n - 2)) = \frac{2}{(n - 1)(n - 2)}$$

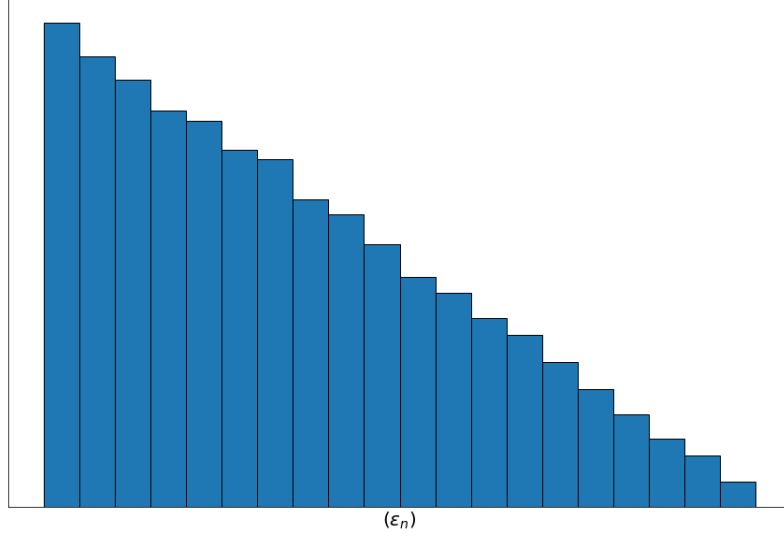


FIGURE 9 – Simulation pour ε_n avec $n = 50000$ itérations sur des permutations aléatoires de tailles $l = 22$.

D'où l'on déduirait géométriquement que la loi de la variable ε_n est la suivante

$$\forall n \geq 2, \forall k \in \{1, \dots, n-2\}, \quad \mathbb{P}(\varepsilon_n = 2k) = \frac{2(n-1-k)}{(n-1)(n-2)}$$

Les simulations valident cette hypothèse pour toute valeur de n , ce qui nous a d'abord satisfait, mais nous avons depuis compris l'origine de ces probabilités, et il se trouve qu'on peut les déduire d'un argument combinatoire assez simple.

5.4 Probabilités

On sait que, si n se trouve à la position i dans π_n , alors on a

$$\varepsilon_n = (n - \pi_n(i-1)) + (n - \pi_n(i+1)) - |\pi_n(i-1) - \pi_n(i+1)|$$

Maintenant, si l'on note $a = \max\{\pi_n(i-1), \pi_n(i+1)\}$, c'est équivalent à dire

$$\varepsilon_n = 2n - 2a$$

On obtient donc une condition sur ε_n qui dépend simplement de a . C'est-à-dire, $\forall k \in \{1, 2, \dots, n-2\}$

$$\varepsilon_n = 2k \iff a = n - k$$

Autrement dit, le maximum des voisins de n dans π_n doit être égal à $n - k$. Et on peut exprimer cela de manière combinatoire assez simplement, de la façon suivante. On commence par noter qu'il y a $2n$ façons pour n et $n - k$ d'être côte à côte dans π_n du point de vue de l'oscillation, puisque l'on considère les permutations de manière cyclique (normalement, il y aurait $n - 1$ positions possibles pour un couple dans une

permutation de taille n). Ensuite, si $n - k$ doit être le maximum des voisins de n , le second voisin ne doit être pas être compris dans $\{n - k + 1, \dots, n - 1\}$.

Si l'on cherche à compter le nombre de permutations de taille n qui satisfont la propriété que n est voisin (d'un point de vue cyclique) de $n - k$, et que $n - k = a$ au sens de la définition précédente, alors il y en a exactement $2n(n - 2 - (k - 1))(n - 3)!$. Le terme $(n - 2 - (k - 1))$ traduit la restriction des valeurs du second voisin de n aux valeurs inférieures à $n - k$.

Enfin, ce raisonnement justifie entièrement les observations concernant $\varepsilon_n(\Omega)$. On obtient alors les probabilités suivantes, $\forall n \geq 3, \forall k \in \{1, 2, \dots, n - 2\}$,

$$\begin{aligned}\mathbb{P}(\varepsilon_n = 2k) &= \frac{2n(n - 2 - (k - 1))(n - 3)!}{n!} \\ &= \frac{2(n - 1 - k)}{(n - 1)(n - 2)}\end{aligned}$$

On retrouve bien les probabilités que nous avons devinées géométriquement, et dont on avait supposé pour les construire que leur somme donnait 1.

5.5 Espérance

Par définition de ε_{n+1} et par le raisonnement précédent, on a

$$\mathcal{O}_{n+1} = \mathcal{O}_n + \varepsilon_{n+1}$$

Et un calcul direct sur la base des probabilités calculés précédemment donne

$$\begin{aligned}\mathbb{E}[\varepsilon_n] &= \sum_{k=1}^{n-2} 2k \cdot \frac{2(n - 1) - 2k}{(n - 1)(n - 2)} \\ &= \frac{4}{(n - 1)(n - 2)} \left((n - 1) \sum_{k=1}^{n-2} k - \sum_{k=1}^{n-2} k^2 \right) \\ &= \frac{4}{(n - 1)(n - 2)} \left((n - 1) \frac{(n - 1)(n - 2)}{2} - \frac{(n - 2)(n - 1)(2n - 3)}{6} \right) \\ &= 2 \left((n - 1) - \frac{(2n - 3)}{3} \right) \\ &= \frac{2n}{3}\end{aligned}$$

On a donc immédiatement par linéarité de l'espérance, $\forall n \geq 3$,

$$\mathbb{E}[\mathcal{O}_{n+1}] = \mathbb{E}[\mathcal{O}_n] + \frac{2(n + 1)}{3}$$

Et avec $\mathcal{O}_2 = 2$, on obtient tout de suite, $\forall n \geq 2$,

$$\mathbb{E}[\mathcal{O}_n] = \frac{1}{3}n(n + 1)$$

Enfin, le calcul de la variance de \mathcal{O}_n n'est pas aussi simple. Par ailleurs, les \mathcal{O}_n ne sont bien sûr pas indépendantes, ce qui aurait rendu le calcul trivial. Il nécessite donc de faire appel à des techniques liées à l'étude des martingales, que nous ne maîtrisons pas (encore !) tout à fait.

Pour illustrer cette approche, si l'on revient sur le calcul de l'espérance, et en utilisant le formalisme propre aux martingales, M. Bercu nous a guidés pour montrer que l'on pouvait établir la relation suivante, en notant $\mathcal{F}_n = \sigma(\pi_1, \dots, \pi_n)$, grossièrement l'information au temps n (la connaissance seule de \mathcal{O}_n n'est pas suffisante, on veut par exemple fixer la position de $n+1$ dans π_{n+1} , et celle-ci étant choisie de manière uniforme par hypothèse, ou pourra généraliser ensuite),

$$\mathbb{E}[\varepsilon_{n+1} \mid \mathcal{F}_n] = n + 1 - \frac{1}{n}\mathcal{O}_n$$

Et alors, grâce à l'égalité obtenue précédemment

$$\mathcal{O}_{n+1} = \mathcal{O}_n + \varepsilon_{n+1}$$

On obtient

$$\mathbb{E}[\mathcal{O}_{n+1} \mid \mathcal{F}_n] = \frac{n-1}{n}\mathcal{O}_n + n + 1$$

D'où l'on déduit que

$$\mathbb{E}[\mathcal{O}_{n+1}] = \frac{n-1}{n}\mathbb{E}[\mathcal{O}_n] + n + 1$$

Et on peut alors utiliser la même méthode que celle utilisée dans les sections précédentes, pour obtenir le résultat

$$\mathbb{E}[\mathcal{O}_n] = \frac{1}{3}n(n+1)$$

Références

- [1] Miklós Bóna. *Combinatorics of permutations*. CRC Press, 2012.
- [2] Chern-Ching Chao. A note on applications of the martingale central limit theorem to random permutations. *Random Structures & Algorithms*, 10(3) :323–332, 1997.
- [3] Chern-Ching Chao, Zhidong Bai, and Wen-Qi Liang. Asymptotic normality for oscillation of permutation. *Probability in the Engineering and Informational Sciences*, 7(2) :227–235, 1993.
- [4] Sourav Chatterjee and Persi Diaconis. A central limit theorem for a new statistic on permutations, 2016.
- [5] Jim Pitman. Probabilistic bounds on the coefficients of polynomials with only real zeros. *Journal of Combinatorial Theory, Series A*, 77(2) :279–303, 1997.
- [6] Alperen Y. Özdemir. Martingales and descent statistics, 2019.