ENS de Lyon — 2nd semestre 2018-2019

L3 - Intégration et Probabilités

DM 2

Corrigé.

2 Étude d'une fonction

1. Les fonctions ρ et ρ' sont continues par morceaux sur $\mathbb{R}$, intégrables en $-\infty$, et de valeur absolue dominée en $+\infty$ par $x \mapsto 1/\lfloor x \rfloor!$ qui est intégrable en $+\infty$. Ainsi ρ et ρ' sont L^1. La fonction f est donc une fonction positive, mesurable (car continue) L^1 d'intégrale sur $\mathbb{R}$ égale à 1 : c'est une densité de probabilité.

2. (a) La fonction ρ' admet 0 comme limite à gauche et 1 comme limite à droite en 1 : ρ n'est pas C^1 sur $\mathbb{R}$, et l'on n'a donc pas nécessairement $\widehat{\rho'} = it\widehat{\rho}$.
Soit $t \in \mathbb{R}$,

$$\begin{aligned}\widehat{\rho'}(t) = \int_{\mathbb{R}} \rho'(x)\mathrm{e}^{-itx}\,\mathrm{d}x = \int_1^\infty \rho'(x)\mathrm{e}^{-itx}\,\mathrm{d}x &= [\rho(x)\mathrm{e}^{-itx}]_1^\infty + it\int_1^\infty \rho(x)\mathrm{e}^{-itx}\,\mathrm{d}x \\ &= 0 - \mathrm{e}^{-it} + it\int_0^\infty \rho(x)\mathrm{e}^{-itx}\,\mathrm{d}x - it\int_0^1 \mathrm{e}^{-itx}\,\mathrm{d}x \\ &= it\widehat{\rho}(t) - 1.\end{aligned}$$

L'intégration par parties est légitime du fait que $\rho(x)\mathrm{e}^{-itx}$ admet une limite en $+\infty$ (qui est nulle étant donné que $|\rho(x)| \leq 1/\lfloor x \rfloor!$) et que $\rho(x)\mathrm{e}^{-itx}$ est intégrable sur $[1;+\infty[$.

(b) On cherche à dériver $\widehat{\rho'}$.

- Pour tout $x \in \mathbb{R}$ fixé la fonction $t \mapsto \rho'(x)\mathrm{e}^{-itx}\mathbf{1}_{\{x>1\}}$ est C^1 de dérivée $-ix\rho'(x)\mathrm{e}^{-itx}\mathbf{1}_{\{x>1\}}$
- Pour tout $t \in \mathbb{R}$, $x \mapsto -ix\rho'(x)\mathrm{e}^{-itx}$ est mesurable, et dominée en module par $|x\rho'(x)| = \rho(x-1)$ qui est L^1 et ne dépend pas de t.

Ainsi, le théorème de dérivation sous le signe intégrale assure que $\widehat{\rho'}$ est C^1 sur $\mathbb{R}$ de dérivée $t \mapsto -i\int_1^\infty x\rho'(x)\mathrm{e}^{-itx}\,\mathrm{d}x$.
De même on montre que $\widehat{\rho}$ est C^1 sur $\mathbb{R}$. Ainsi, d'après la question 2a, en dérivant à gauche et à droite l'égalité $it\widehat{\rho}(t) - 1 = \widehat{\rho'}(t)$, on a pour tout $t \in \mathbb{R}$,

$$\begin{aligned}i(t\widehat{\rho}'(t) + \widehat{\rho}(t)) = \widehat{\rho'}'(t) &= -i\int_1^\infty x\rho'(x)\mathrm{e}^{-itx}\,\mathrm{d}x \\ &= i\int_1^\infty \rho(x-1)\mathrm{e}^{-itx}\,\mathrm{d}x = i\mathrm{e}^{-it}\widehat{\rho}(t).\end{aligned}$$

D'où pour tout $t \in \mathbb{R}$,

$$t\widehat{\rho}'(t) = (\mathrm{e}^{-it} - 1)\widehat{\rho}(t). \qquad (*)$$

(c) L'ensemble des fonctions C^1 sur $\mathbb{R}$ satisfaisant l'équation $(*)$ est

$$\mathcal{S} := \{t \mapsto A\exp\Big(\int_0^t \frac{\mathrm{e}^{-iu} - 1}{u}\,\mathrm{d}u\Big) : A \in \mathbb{R}\},$$

et comme $\phi_Z(t) = \widehat{\rho}(-t)$ est telle que $\phi_Z(0) = 1$, on a que pour tout $t \in \mathbb{R}$,

$$\phi_Z(t) = \exp\Big(\int_0^x \frac{\mathrm{e}^{iu} - 1}{u}\,\mathrm{d}u\Big) = \exp\Big(\int_0^1 \frac{\mathrm{e}^{iux} - 1}{u}\,\mathrm{d}u\Big).$$

3 Plus grand cycle d'une permutation

3. Soit $\mathbf{c} \in \mathbb{N}^n$, on a que

$$\begin{aligned}
\mathbb{P}(\mathbf{X}(n) = \mathbf{c},\ T_n = n) &= \mathbb{P}\Big(X_1 = c_1, \dots, X_n = c_n, \sum_{j=1}^n jX_j = n\Big) \\
&= \mathbb{E}\Big[\mathbf{1}_{\{X_1=c_1\}} \times \cdots \times \mathbf{1}_{\{X_n=c_n\}} \mathbf{1}_{\{\sum_{j=1}^n jc_j = n\}}\Big] \\
&= \mathbf{1}_{\{\sum_{j=1}^n jc_j = n\}} \prod_{j=1}^n \mathbb{P}(X_j = c_j) \\
&= \mathbf{1}_{\{\sum_{j=1}^n jc_j = n\}} \prod_{j=1}^n \frac{\mathrm{e}^{-1/j}(1/j)^{c_j}}{c_j!} = \mathbf{1}_{\{\sum_{j=1}^n jc_j = n\}} \mathrm{e}^{-\sum_{j=1}^n \frac{1}{j}} \prod_{j=1}^n \frac{1}{j^{c_j}(c_j!)}.
\end{aligned}$$

La troisième égalité est due au fait que les $(X_j)_{j\geq 1}$ sont indépendants et que $\mathbf{1}_{\{\sum_{j=1}^n jc_j = n\}}$ est déterministe. Ainsi

$$\frac{\mathbb{P}(\mathbf{X}(n) = \mathbf{c},\ T_n = n)}{\mathbb{P}(T_n = n)} = \frac{\exp\big(-\sum_{j=1}^n \frac{1}{j}\big)}{\mathbb{P}(T_n = n)} \mathbb{P}\big(\mathbf{C}(n) = \mathbf{c}\big).$$

Mais comme $\frac{\sum_{\mathbf{c}\in\mathbb{N}^n} \mathbb{P}(\mathbf{X}(n)=\mathbf{c},\ T_n=n)}{\mathbb{P}(T_n=n)} = 1$ et $\sum_{\mathbf{c}\in\mathbb{N}^n} \mathbb{P}\big(\mathbf{C}(n) = \mathbf{c}\big) = 1$, on a

$$\mathbb{P}(T_n = n) = \exp\big(-\sum_{j=1}^n \frac{1}{j}\big)$$

ainsi que l'égalité voulue.

4. La fonction caractéristique de la loi de Poisson de paramètre λ est la fonction $\phi_{\mathcal{P}(\lambda)} : t \in \mathbb{R} \mapsto \exp(\lambda(\mathrm{e}^{it} - 1))$. D'où la fonction caractéristique de T_n/n est déterminée pour tout $t \in \mathbb{R}$ par

$$\begin{aligned}
\Phi_{\frac{T_n}{n}}(t) = \mathbb{E}\Big[\exp(it\sum_{j=1}^n \frac{j}{n}X_j)\Big] = \prod_{j=1}^n \mathbb{E}\big[\exp(i\frac{tj}{n}X_j)\big] &= \prod_{j=1}^n \exp\big(\frac{1}{j}(\mathrm{e}^{it\frac{j}{n}} - 1)\big) \\
&= \exp\big(\frac{1}{n}\sum_{j=1}^n \frac{n}{j}(\mathrm{e}^{it\frac{j}{n}} - 1)\big).
\end{aligned}$$

On reconnaît une somme de Riemann : comme la fonction $x \mapsto \frac{\mathrm{e}^{itx}-1}{x}$ est continue sur $[0;1]$ (en attribuant la valeur it en 0), on a pour tout $t \in \mathbb{R}$

$$\Phi_{\frac{T_n}{n}}(t) = \exp\big(\frac{1}{n}\sum_{j=1}^n \frac{n}{j}(\mathrm{e}^{it\frac{j}{n}} - 1)\big) \underset{n\to\infty}{\longrightarrow} \exp\Big(\int_0^1 \frac{\mathrm{e}^{iut}-1}{u}\,\mathrm{d}u\Big) = \Phi_Z(t).$$

D'où d'après le théorème de convergence de Lévy, T_n/n converge en loi vers Z.

5. (a) Soit $j \geq 1$ et $x_j \in \mathbb{N}$. Si $x_j = 0$, l'égalité est triviale. Sinon,

$$\mathbb{P}(X_j{+}1 = x_j) = \mathbb{P}(X_j = x_j{-}1) = \mathrm{e}^{-1/j}\frac{1}{j^{x_j-1}(x_j-1)!} = jx_j\mathrm{e}^{-1/j}\frac{1}{j^{x_j}(x_j)!} = jx_j\mathbb{P}(X_j = x_j).$$

Montrons la seconde égalité. Soit U_n une variable aléatoire de loi uniforme sur $\{1,\dots,n\}$. Si $k=0$, l'égalité est triviale ; sinon on a pour tout $k\geq 1$

$$\begin{aligned}
\mathbb{P}(T_n+U_n=k) &= \mathbb{P}\Big(\sum_{j=1}^n jX_j+U_n=k\Big)=\sum_{\ell=1}^n \mathbb{P}\Big(\sum_{j=1}^n jX_j+\ell=k,\ U_n=\ell\Big)\\
&=\sum_{\ell=1}^n \mathbb{P}\Big(\ell(X_\ell+1)+\sum_{1\leq j\leq n,\ j\neq \ell} jX_j=k\Big)\times\frac{1}{n}\\
&=\frac{1}{n}\sum_{\ell=1}^n\sum_{\mathbf{c}\in\mathbb{N}^n}\mathbf{1}_{\{\sum_{j=1}^n jc_j=k\}}\mathbb{P}\Big(X_\ell+1=c_\ell \text{ et } \forall j\in[\![1;n]\!], j\neq\ell,\ X_j=c_j\Big)\\
&=\frac{1}{n}\sum_{\ell=1}^n\sum_{\mathbf{c}\in\mathbb{N}^n}\mathbf{1}_{\{\sum_{j=1}^n jc_j=k\}}\mathbb{P}(X_\ell+1=c_\ell)\prod_{j\in[\![1;n]\!],j\neq\ell}\mathbb{P}(X_j=c_j)\\
&=\frac{1}{n}\sum_{\ell=1}^n\sum_{\mathbf{c}\in\mathbb{N}^n}\mathbf{1}_{\{\sum_{j=1}^n jc_j=k\}}\ell c_\ell\prod_{j\in[\![1;n]\!]}\mathbb{P}(X_j=c_j)\\
&=\frac{1}{n}\sum_{\mathbf{c}\in\mathbb{N}^n}\mathbf{1}_{\{\sum_{j=1}^n jc_j=k\}}\big(\sum_{\ell=1}^n \ell c_\ell\big)\prod_{j\in[\![1;n]\!]}\mathbb{P}(X_j=c_j)=\frac{k}{n}\mathbb{P}(\sum_{j=1}^n jX_j=k)=\frac{k}{n}\mathbb{P}(T_n=k).
\end{aligned}$$

Dans la troisième égalité on a utilisé l'indépendance de U_n et $(X_j)_{1\leq j\leq n}$, et dans la quatrième on a distingué selon les valeurs prises par les $(X_j)_{1\leq j\leq n}$.

(b) Soit $k\in[\![1;n]\!]$, on a d'après ce qui précède

$$\begin{aligned}
\mathbb{P}(T_n=k)=\frac{n}{k}\mathbb{P}(T_n+U_n=k)=\frac{n}{k}\sum_{\ell=1}^n\mathbb{P}(T_n=k-\ell,U_n=\ell)&=\frac{n}{k}\sum_{\ell=1}^n\mathbb{P}(T_n=k-\ell)\times\frac{1}{n}\\
&=\frac{1}{k}\mathbb{P}(k-n\leq T_n<k),
\end{aligned}$$

étant donné que l'événement $\{k-n\leq T_n<k\}$ est réunion disjointe des événements $\{T_n=k-\ell\}$ pour $\ell\in[\![1;n]\!]$.

(c) Le second théorème de Dini assure qu'étant donné que $T_n/n\to Z$ en loi, et que les fonctions de répartition sont croissantes, et celle de Z continue (Z étant une v.a. à densité au vu de sa définition dans la question 1),

$$\sup_{x\in\mathbb{R}}|\mathbb{P}(T_n\leq x)-\mathbb{P}(Z\leq x)|\to 0. \qquad (**)$$

D'où, soit $(k_n)_{n\geq 1}$ telle que $k_n/n\to_n y$, on a

$$\begin{aligned}
&|k_n\mathbb{P}(T_n=k_n)-\mathbb{P}(y-1\leq Z<y)|=\Big|\mathbb{P}(\frac{k_n-n}{n}\leq\frac{T_n}{n}<\frac{k_n}{n})-\mathbb{P}(y-1\leq Z<y)\Big|\\
&\leq\Big|\mathbb{P}(\frac{k_n-n}{n}\leq\frac{T_n}{n}<\frac{k_n}{n})-\mathbb{P}(\frac{k_n-n}{n}\leq Z<\frac{k_n}{n})\Big|+\Big|\mathbb{P}(\frac{k_n-n}{n}\leq Z<\frac{k_n}{n})-\mathbb{P}(y-1\leq Z<y)\Big|.
\end{aligned}$$

Le premier terme de cette somme tend vers 0 d'après $(**)$, et le second également par continuité de la fonction de répartition de Z (Z étant une v.a. à densité au vu de sa définition dans la question 1) étant donné que $k_n/n\to y$. Ainsi,

$$\mathbb{P}(T_n=k_n)\sim\frac{1}{ny}\mathbb{P}(y-1\leq Z<y)=\frac{1}{n}f(y),$$

d'après l'affirmation de la question 1, et où l'on rappelle que f est la densité de Z.

6. (a) D'après la question 3 sur la loi de $\mathbf{C}(n)$, on a

$$\begin{aligned}\mathbb{P}(G_n \le m) = \mathbb{P}(C_{m+1}(n) = 0, \dots, C_n(n) = 0) &= \frac{\mathbb{P}\big(X_{m+1} = 0, \dots, X_n = 0,\ \sum_{j=1}^n jX_j = n\big)}{\mathbb{P}(T_n = n)\big)} \\ &= \frac{\mathbb{P}\big(X_{m+1} = 0, \dots, X_n = 0,\ \sum_{j=1}^m jX_j = n\big)}{\mathbb{P}(T_n = n)\big)} \\ &= \mathbb{P}(X_{m+1} = 0) \times \cdots \times \mathbb{P}(X_n = 0) \frac{\mathbb{P}(T_m = 0)}{\mathbb{P}(T_n = 0)}.\end{aligned}$$

La dernière égalité est obtenue par indépendance de $X_{m+1}, \dots, X_n$ et de T_m (qui est $\sigma(X_1, \dots, X_m)$-mesurable).

(b) Par définition, $G_n/n \in [0;1]$; et pour tout $x \in]0;1]$,

$$\mathbb{P}\big(\frac{G_n}{n} \le x\big) = \mathbb{P}(G_n \le \lfloor nx \rfloor) = \Big(\prod_{k=\lceil nx \rceil}^{n} \mathrm{e}^{-1/k} \frac{1}{k^0 \times 0!} \Big) \frac{\mathbb{P}(T_{\lfloor nx \rfloor} = n)}{\mathbb{P}(T_n = n)} = \exp\Big(- \sum_{k=\lceil nx \rceil}^{n} \frac{1}{k} \Big). \frac{\mathbb{P}(T_{\lfloor nx \rfloor} = n)}{\mathbb{P}(T_n = n)}.$$

Or, $\sum_{k=\lceil nx \rceil}^n \frac{1}{k} = \ln(n) - \ln(\lfloor nx \rfloor) + o(1) = -\ln(x) + o(1)$, et $n/\lfloor nx \rfloor \to 1/x$ donc d'après la question précédente,

$$\mathbb{P}\big(\frac{G_n}{n} \le x\big) = \exp(\ln(1/x) + o(1)) \big(\frac{(1/\lfloor nx \rfloor) f(1/x) + o(1)}{(1/n) f(1) + o(1)} \big) \underset{n\to\infty}{\longrightarrow} \frac{f(\frac{1}{x})}{f(1)}.$$

En tant que limite de fonction croissante, $F(x) := f(1/x)/f(1)$ est croissante, telle que $F(1) = 1$, et se prolonge par continuité en 0 en posant $F(0) = 0$. Ainsi, c'est la fonction de répartition d'une loi à support dans $[0;1]$, et la convergence de la fonction de répartition de G_n/n vers F assure sa convergence vers une telle loi.

4 Quelques remarques

1. La fonction ρ s'appelle fonction de Dickman, et a été introduite en théorie analytique des nombres (voir question bonus 14).
2. La méthode suivie dans ce DM est un très grand classique de l'étude probabiliste d'objets discrets de grande taille uniformes :
   (a) Écrire un objet uniforme de taille n comme un processus stochastique classique de taille aléatoire, conditionné à avoir taille n. (question 3 ici)
   (b) Obtenir l'asymptotique de la probabilité de l'évènement de conditionnement (question 5 ici). Elle ne doit pas tendre vers 0 trop vite.
   (c) En déduire des propriétés sur l'objet uniforme

   Par exemple,
   (a) un arbre de Galton-Watson de loi Geom(1/2) conditionné à avoir n sommets est un arbre (ordonné, enraciné) à n sommets uniforme. La probabilité de l'évènement de conditionnement est $\approx n^{-3/2}$.
   (b) Un graphe à n sommets et m arêtes uniforme s'écrit comme le graphe d'Erdös-Rényi $G(n, m/\binom{n}{2})$, conditionné à avoir m arêtes. La probabilité de l'évènement de conditionnement est (quand m et ou $\binom{n}{2} - m$ tendent vers l'infini) de l'ordre de $n^{-1/2}$.

5 Questions bonus

7. Les élèves décident de la stratégie suivante : ils choisissent d'attribuer à chacun d'entre eux un numéro entre 1 et n, de manière à ce que tous aient un numéro distinct. Ces numéros sont attribués uniformément au hasard : ainsi, le Professeur K. ne peut pas deviner leur numéro. Si l'on note $g(k)$ le numéro de l'élève auquel appartient la copie se trouvant dans l'enveloppe k, g est donc une permutation de loi uniforme.

   Puis, chaque élève qui rentrera dans la salle suivra la procédure suivante. Il commence par ouvrir l'enveloppe correspondant à son numéro : s'il tombe sur sa copie, c'est gagné ; sinon il tombe sur la copie d'un autre élève auquel correspond un numéro. Il va donc ouvrir l'enveloppe numérotée par celui-ci, et à nouveau, soit trouver sa copie, soit trouver une copie auquel correspond un numéro. Il procède ainsi de suite jusqu'à éventuellement tomber sur sa copie, ou bien jusqu'à avoir ouvert $n/2$ enveloppes.

   L'événement $\{$l'élève k a ouvert $n/2$ enveloppes sans trouver sa copie$\}$ est donc égal à $\{k$ est dans un cycle de longueur $\geq n/2\}$, et l'événement $\{$au moins un élève n'a pas trouvé sa copie$\}$ est égal à l'événement $\{g$ admet un cycle de longueur$\geq n/2\}$. Or, pour n grand, $\mathbb{P}(G_n < n/2) \to f(2)/f(1) = \rho(2)$. Au vu de l'équation différentielle satisfaite par ρ, on a $\rho'(x) = -1/x$ pour $x \in [1;2]$ et par continuité de ρ, $\rho(x) = 1 - \ln(x)$ d'où $\mathbb{P}(G_n < n/2) \approx 1 - \ln(2) \approx 0,306852$. La probabilité que leur stratégie réussisse est donc pour n grand supérieure à 30%.

   On peut en fait montrer que cette stratégie est optimale [1]

8. Soit $\mathbf{c}$ fixé, tel que $\sum_{j=1}^{n} jc_j = n$. On veut compter le cardinal de l'ensemble $\mathfrak{S}(\mathbf{c})$ des permutations qui ont c_j cycles de taille j pour tout $1 \leq j \leq n$.

   Regardons l'action de $\mathfrak{S}_n$ sur $\mathfrak{S}_n$ par conjugaison. σ agit sur une permutation τ en remplaçant i par $\sigma(i)$ dans l'écriture en cycles de τ. En particulier, ses orbites sont les $\mathfrak{S}(\mathbf{c})$. On donc la formule des classes : pour $\tau \in \mathfrak{S}(\mathbf{c})$,

$$\#\mathfrak{S}(\mathbf{c}) = \frac{\#\mathfrak{S}_n}{\#\mathrm{Stab}(\tau)}$$

   Comptons donc les réétiquetages σ préservant $\tau \in \mathfrak{S}(\mathbf{c})$. Soit $1 \leq i \leq n$. On a $\tau^k(\sigma(i)) = \sigma(\tau^k(i))$ pour tout k. Ceci implique que si i est dans un cycle de taille j dans τ, alors $\sigma(i)$ aussi. De plus le choix de $\sigma(i)$ détermine σ sur tout le cycle de i (qui est envoyé sur le cycle de $\sigma(i)$. On en déduit qu'il y a jc_j choix pour σ sur le cycle de i.

   Maintenant on doit déterminer $\sigma_{|[n]\backslash \mathrm{cyc}(i)}$, ce qui se ramène au même problème pour une permutation de $\mathfrak{S}(c_1, \ldots, c_j - 1, \ldots, c_n)$.

   Par induction on a donc que pour $\tau \in \mathfrak{S}(\mathbf{c})$, $\#\mathrm{Stab}(\tau) = \prod_{j=1}^{n} c_j! j_j^c$. Par la formule des classes on conclut.

9. Pour $j \geq n/2$, le nombre de permutations ayant un cycle de taille j (il n'y en a alors q'un !) est (choix ordonné des indices à mettre dans le cycle)$\times$(correction du surcomptage par k vu qu'un cycle est invariant par décalage cyclique)$\times$(détermination de la permutation hors du cycle)$\frac{n!}{(n-j)!} \times \frac{1}{j} \times (n-j)! = \frac{n!}{j}$.

   Ce qui est remarquable, c'est que pour $j \geq n/2$, les évènements "avoir un cycle de taille j" sont tous disjoints. Ainsi

$$\mathbb{P}(\text{avoir un cycle de taille } \geq k) = \sum_{j=k}^{n} \mathbb{P}(\text{avoir un cycle de taille } j) = \sum_{j=k}^{n} 1/j = H_n - H_k$$

1. Voir cet article du Mathematical Intelligencer `https://www.cl.cam.ac.uk/~gw104/Locker_Puzzle.pdf` ou (en français) dans l'article suivant de Pour La Science `http://cristal.univ-lille.fr/~jdelahay/pls/2016/272.pdf`

Alors pour $y \geq 1/2$, $\mathbb{P}(\text{avoir un cycle de taille } \geq ny) = H_n - H_{\lceil ny \rceil} = \log(n) + \gamma + o(1) - \log(\lceil ny \rceil) - \gamma - o(1) = \log(y) + o(1)$. On retrouve la fonction de répartition limite sur $[1/2, 1]$. De plus, la suite $H_n - H_{\lceil ny \rceil}$ est croissante ! Donc pour tout n la proba de réussir est au dessus de 30%.

10. On a obtenu que $\mathbb{P}(T_n = n) = \exp\left(-\sum_{j=1}^n \frac{1}{j}\right) = \exp(-\log n - \gamma - o(1)) = e^{-\gamma}(1+o(1))/n$. Mais la question 5c nous dit que $\mathbb{P}(T_n = n) \sim f(1)/n = \frac{1}{n \int_0^\infty \rho(x)dx}$. En identifiant les deux asymptotiques, on a $\int_0^\infty \rho(x)dx = e^\gamma$.

11. On considère le processus du restaurant chinois et on définit $B_n = \#\{\text{table de 1 au temps } n\}$ et $R_n = \#\{\text{autres tables au temps } n\} + 1$. On a que $(B_1, R_1) = (1, 1)$ et par construction du restaurant chinois, on voit que ce processus est une **urne de Polyà** : à chaque étape, on augmente B avec probabilité $B_n/(n+1)$ et R avec probabilité $R_n/(n+1)$. On a vu en TD que B_n/n converge presque sûrement vers une variable uniforme.

    Remarque 1 On a voit que le cycle qui contient 1 est de taille d'ordre n (comme le cycle de taille maximale) Ceci semble contre-intuitif car le nombre de cycles est presque sûrement équivalent à $\log(n)$ et donc la taille moyenne des cycles de σ_n est presque sûrement équivalente à $n/\log(n)$. Mais le cycle qui contient 1 n'est choisi uniformément au hasard dans l'ensemble des cycles, mais de manière biaisée par la taille (plus on est grand, plus on a de chances de contenir 1), ce qui le rend franchement plus grand. C'est un exemple de paradoxe des autobus [2].

    Remarque 2 On peut faire le même argument sur le processus privé de la table 1, pour montrer que la suite des longueurs des tables divisées par n, converge p.s. vers $(U_1, (1-U_1)U_2, (1-U_1)(1-U_2)U_3, \ldots)$, où $(U_i)_i$ est une suite de variables uniformes. On nomme ce dernier processus le "stick-breaking process". Attention : cette suite n'est pas forcément décroissante. C'est la limite des cycles triés dans l'ordre d'apparition, et pas dans l'ordre de taille. La limite de la suite des cycles triés par taille est le réordonnement du stick-breaking process.

    Remarque 3 On peut en déduire (preuve laissée au lecteur) que G_n/n converge vers le maximum (atteint !) G du stick-breaking process. On a l'identité en loi suivante : $G \stackrel{\mathcal{L}}{=} \max(U, (1-U)G)$, où $U \perp\!\!\!\perp G$, qui permet de retrouver la fonction de répartition de G (exercice !)

12. Notons $S = U_1 + U_1U_2 + \cdots$. Déjà cette série est p.s. finie car son espérance (qui vaut 1 !) l'est. Ensuite on a l'identité en loi $S = U(1+S)$, où $U \perp\!\!\!\perp S$, qui implique $\phi_S(t) \stackrel{\mathcal{L}}{=} \int_0^1 e^{itu}\phi_S(ut)du = \frac{1}{t}\int_0^t e^{ix}\phi_S(x)dx$. En dérivant on obtient une équadiff sur ϕ_S qui est la même que celle sur ϕ_Z.

13. Comme déjà dit, la fonction ρ s'appelle fonction de Dickman. Vous trouverez des graphes partout sur le web. Pour simuler les tailles des cycles dans une grande permutation, allez voir cette implémentation du processus des restaurants chinois : `http://topicmodels.west.uni-koblenz.de/ckling/tmt/crp.html?parameters=1&dp=1#`.

14. Les longueurs de cycles dans σ_n forment un ensemble aléatoire de nombres qui somment à n. Ce DM montre que le plus grand de ces nombres, divisé par la somme, tend en loi vers Z.

    En théorie des nombres, on a que les logarithmes des facteurs premiers d'un nombre n somment à $\log(n)$. Et le résultat, dû à Dickman et De Bruijn [3] est le suivant : notant U_n

2. Voir `https://jakevdp.github.io/blog/2018/09/13/waiting-time-paradox/` et `https://en.wikipedia.org/wiki/Length_time_bias`

3. Pour l'anecdote, Dickman était un actuaire qui n'a écrit qu'un seul papier de mathématiques dans sa vie. De Bruijn était un mathématicien/informaticien/logicien de renom. Voir `https://en.wikipedia.org/wiki/Dickman_function` et les liens à l'intérieur.

un entier uniforme entre 1 et n, et P_n son plus grand facteur premier, on a $\frac{\log P_n}{\log U_n} \xrightarrow{\mathcal{L}} Z$. En fait l'analogie ne s'arête pas là. Notant $p_{1,n}, p_{2,n}, \ldots$ la suite décroissante des facteurs premiers de U_n, on a que la suite $(\frac{\log p_{1,n}}{\log U_n}, \frac{\log p_{2,n}}{\log U_n}, \ldots)$ converge vers le réordonnement du stick-breaking process. La convergence est due aux informaticiens Knuth [4] et Trabb Pardo, et l'identification du processus limite aux probabilistes Billingsley, Donnelly et Grimmett. Tout ceci est joliment raconté dans un article du blog de Terry Tao : `https://terrytao.wordpress.com/2013/09/21/the-poisson-dirichlet-process-and-large-prime-factors-of-a-ran`

4. Que vient faire Knuth, le pionnier de l'informatique théorique et de l'algorithmique, là dedans ? Eh bien de manière générale, pour analyser la complexité en moyenne d'algorithmes, il est intéressant de connaître les lois de diverses statistiques d'une entrée uniforme de l'algo (cela fait partie des grandes motivations de l'étude probabiliste de grands objets discrets). Dans ce cas il s'agissait d'étudier la complexité en moyenne d'un algo de factorisation d'entiers.