Limites de permutations aléatoires

(à rendre pour le 4 novembre 2016)

L'objectif de ce devoir est d'étudier un espace $\mathscr{P}$ qui correspond aux limites de permutations aléatoires, au sens de la convergence des densités de motifs. Dans la première partie, on définit cette notion de convergence, et on construit des modèles de permutations aléatoires associés à des paramètres d'un certain espace $\mathscr{P}$, et dont toutes les densités de motifs convergent. Dans la seconde partie, on s'intéresse à la topologie de $\mathscr{P}$, et on montre que c'est un espace métrique compact dont tout point π engendre une famille de déformations aléatoires $(\pi_n)_{n\in\mathbb{N}}$, avec $\pi_n \to_{\mathbb{P}} \pi$.

Partie 1. Motifs de permutations.

On rappelle qu'une *permutation* de taille n est une bijection $\sigma : [\![1,n]\!] \to [\![1,n]\!]$. On identifiera une permutation σ et son mot $\sigma(1)\sigma(2)\cdots\sigma(n)$. L'ensemble des permutations de taille n est le *groupe symétrique* $\mathfrak{S}(n)$, de cardinal $n!$. Si $k \leq n$, $\tau \in \mathfrak{S}(k)$ et $\sigma \in \mathfrak{S}(n)$, on dit que τ est un *motif* de σ s'il existe une partie $\{a_1 < a_2 < \cdots < a_k\}$ de $[\![1,n]\!]$ telle que $\sigma(a_i) < \sigma(a_j)$ si et seulement si $\tau(i) < \tau(j)$, voir la figure ci-dessous avec $\tau = 213$ et $\sigma = 245361$.

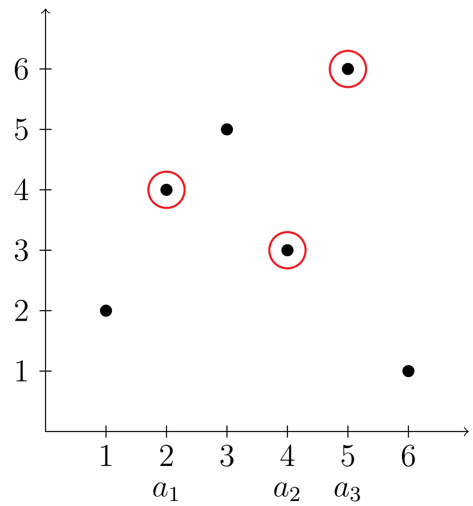


FIGURE 1 – La permutation 213 est un motif dans $\sigma = 245361$.

La *densité* du motif τ dans σ est la quantité

$$t(\tau,\sigma) = \frac{\mathrm{card}\,\{\text{parties } \{a_1 < a_2 < \cdots < a_k\} \text{ qui font apparaître } \tau \text{ comme motif dans } \sigma\}}{\binom{n}{k}}.$$

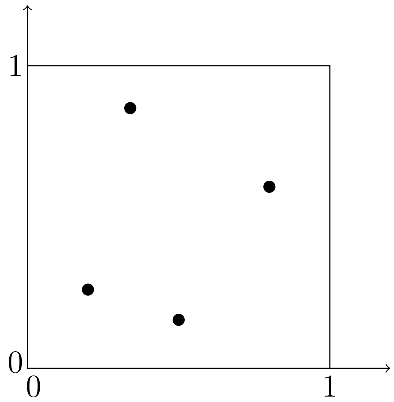
1. Montrer que $t(213, 245361) = \frac{1}{10}$.

On dit qu'une suite de permutations $(\sigma_n)_{n\in\mathbb{N}}$ avec $\sigma_n \in \mathfrak{S}(n)$ converge au sens des motifs si, pour tout $\tau \in \bigsqcup_{k\in\mathbb{N}} \mathfrak{S}(k)$, la suite $(t(\tau, \sigma_n))_{n\in\mathbb{N}}$ admet une limite dans $[0, 1]$. L'objectif de ce devoir est d'étudier en détail cette notion, en particulier lorsque les σ_n sont des permutations aléatoires.

Si $(x_1, y_1), \ldots, (x_k, y_k)$ est une famille de points dans $[0, 1]^2$, on dit qu'ils sont en configuration générale si tous les x_i sont distincts, et si tous les y_i sont distincts. Dans ce cas, on appelle configuration de cette famille de points l'unique permutation $\tau \in \mathfrak{S}(k)$ telle que les deux bijections croissantes

$$\psi_1 : \{x_1, \ldots, x_k\} \to [\![1, k]\!] \quad \text{et} \quad \psi_2 : \{y_1, \ldots, y_k\} \to [\![1, k]\!]$$

vérifient $\psi_2(y_i) = \tau \circ \psi_1(x_i)$ pour tout $i \in [\![1, k]\!]$. Par exemple, la famille de points dessinée ci-dessous



a pour configuration $\mathrm{conf}((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)) = 2413$. D'autre part, on appelle *permuton* une mesure de probabilité π borélienne sur $[0, 1]^2$, telle que si $(X, Y) \sim \pi$, alors X et Y suivent des lois uniformes sur $[0, 1]$ (attention, les variables X et Y ne sont pas supposées indépendantes). On note $\mathscr{P} \subset \mathscr{M}([0, 1]^2)$ l'ensemble des permutons.

2. Soit π un permuton. Montrer que si $(X_1, Y_1), \ldots, (X_k, Y_k)$ sont des variables indépendantes de loi π, alors cette famille de points est en configuration générale avec probabilité 1.

On peut donc définir sans ambiguïté $\mathrm{conf}((X_1, Y_1), \ldots, (X_k, Y_k))$, et si $\tau \in \mathfrak{S}(k)$, on pose :

$$t(\tau, \pi) = \int_{([0,1]^2)^k} 1_{\mathrm{conf}(p_1,\ldots,p_k)=\tau}\, \pi^{\otimes k}(dp_1, \ldots, dp_k) = \mathbb{P}_\pi[\mathrm{conf}((X_1, Y_1), \ldots, (X_k, Y_k)) = \tau].$$

3. Pour toute permutation τ, calculer $t(\tau, \upsilon)$, où υ est la mesure de Lebesgue sur $[0, 1]^2$.

Étant donné un permuton π, on note $\sigma_n(\pi) \in \mathfrak{S}(n)$ la permutation aléatoire définie par :

$$\sigma_n(\pi) = \mathrm{conf}((X_1, Y_1), \ldots, (X_n, Y_n)), \quad \text{avec } ((X_1, Y_1), \ldots, (X_n, Y_n)) \sim \pi^{\otimes n}.$$

4. Soit $\tau \in \mathfrak{S}(k)$, et $n \geq k$. Si $A = \{a_1 < a_2 < \cdots < a_k\}$ est une partie de taille k dans $[\![1, n]\!]$, on pose

$$C_{A,\tau} = \begin{cases} 1 & \text{si } \tau = \mathrm{conf}((X_{a_1}, Y_{a_1}), \ldots, (X_{a_k}, Y_{a_k})), \\ 0 & \text{sinon.} \end{cases}$$

Exprimer $t(\tau, \sigma_n(\pi))$ en fonction des variables $C_{A,\tau}$. Montrer que $\mathbb{E}[t(\tau, \sigma_n(\pi))] = t(\tau, \pi)$.

5. Montrer que si $A \cap B = \emptyset$, alors $C_{A,\tau}$ et $C_{B,\tau}$ sont indépendantes. En déduire que

$$\mathrm{var}(t(\tau, \sigma_n(\pi))) \leq \frac{\binom{n}{k} - \binom{n-k}{k}}{\binom{n}{k}} \leq \frac{k^2}{n}.$$

6. Montrer que la suite $(t(\tau, \sigma_n(\pi))_{n\in\mathbb{N}}$ converge en probabilité vers $t(\tau, \pi)$ pour toute permutation π.

Étant donnée une permutation $\sigma \in \mathfrak{S}(n)$, on peut définir réciproquement un permuton π_σ par la formule

$$d\pi_\sigma(x, y) = n\, 1_{\sigma(\lceil nx \rceil) = \lceil ny \rceil}\, dx\, dy,$$

où $\lceil t \rceil$ désigne l'entier approchant t par valeurs supérieures : $\lceil t \rceil - 1 < t \leq \lceil t \rceil$. Par exemple, le permuton associé à $\sigma = 245361$ est :

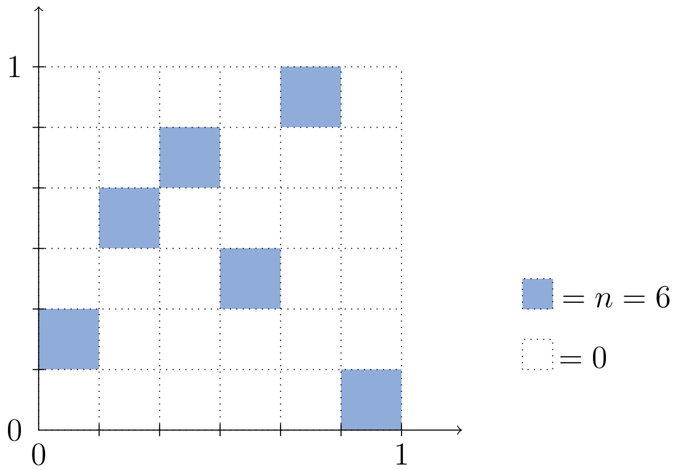


FIGURE 2 – Densité du permuton π_σ associé à la permutation $\sigma = 245361$.

7. Vérifier que pour toute permutation σ, π_σ est un permuton. Montrer que pour toutes permutations $\sigma \in \mathfrak{S}(n)$ et $\tau \in \mathfrak{S}(k)$,

$$|t(\tau, \sigma) - t(\tau, \pi_\sigma)| \leq \frac{1}{n} \binom{k}{2}.$$

Étant données des variables indépendantes (X_i, Y_i) de loi π_σ, notant $n_i = \lceil nX_i \rceil$, on pourra introduire les événements

$$A = \{\mathrm{conf}((X_1, Y_1), \ldots, (X_k, Y_k)) = \tau\};$$
$$B = \{\forall 1 \leq i < j \leq k,\ n_i \neq n_j\},$$

et utiliser l'inégalité $|\mathbb{P}[A|B] - \mathbb{P}[A]| \leq \mathbb{P}[B^c]$.

8. Pour tout permuton π, on pose $\pi_n(\pi) = \pi_{\sigma_n(\pi)}$. Montrer que pour tout motif τ, $(t(\tau, \pi_n(\pi)))_{n\in\mathbb{N}}$ converge en probabilité vers $t(\tau, \pi)$.

Partie 2. L'espace des permutons.

1. Montrer que $\mathscr{P}$ est une partie compacte de $\mathscr{M}([0,1]^2)$ pour la topologie de la convergence en loi.
2. Si π est un permuton, on note $F_\pi(x,y)$ sa fonction de répartition bivariée, c'est-à-dire
$$F_\pi(x,y) = \mathbb{P}_\pi[X \leq x,\, Y \leq y] = \int_{s=0}^{x} \int_{t=0}^{y} d\pi(s,t).$$
Montrer que si π et π' sont deux permutons avec $F_\pi(x,y) = F_{\pi'}(x,y)$ pour tout $(x,y) \in [0,1]^2$, alors $\pi = \pi'$ dans $\mathscr{P} \subset \mathscr{M}([0,1]^2)$. On pourra montrer que $\pi(f) = \pi'(f)$ pour toute fonction continue bornée sur $[0,1]^2$.
3. Soit π, π' deux permutons tels que $t(\tau,\pi) = t(\tau,\pi')$ pour toute permutation $\tau \in \bigsqcup_{k\in\mathbb{N}} \mathfrak{S}(k)$. Montrer que $\sigma_n(\pi)$ et $\sigma_n(\pi')$ ont même loi pour tout entier n. On pose
$$F_{n,\pi}(x,y) = \frac{1}{n} \sum_{i=1}^{\lceil nx \rceil} 1_{\sigma_n(\pi)(i) \leq \lceil ny \rceil},$$
et on définit de même $F_{n,\pi'}(x,y)$. Montrer que $F_{n,\pi}(x,y) \to_{\mathbb{P}} F_\pi(x,y)$, et en déduire que $\pi = \pi'$. Étant donné $((X_1,Y_1),\ldots,(X_n,Y_n)) \sim \pi^{\otimes n}$, on pourra introduire les réordonnements croissants
$$X_{\theta_1(1)} < X_{\theta_1(2)} < \cdots < X_{\theta_1(n)}$$
$$Y_{\theta_2(1)} < Y_{\theta_2(2)} < \cdots < Y_{\theta_2(n)}$$
et comparer à l'aide du théorème de Glivenko-Cantelli les quantités
$$F_{n,\pi}(x,y) \quad ; \quad \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq X_{\theta_1(k)},\, Y_i \leq Y_{\theta_2(l)}} \quad ; \quad \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq \frac{k}{n},\, Y_i \leq \frac{l}{n}}.$$
avec $k = \lceil nx \rceil$ et $l = \lceil ny \rceil$.
4. On équipe $\mathscr{P}$ de la restriction de la topologie de $\mathscr{M}([0,1])^2$ (convergence en loi). Étant donnée une suite de permutons $(\pi_n)_{n\in\mathbb{N}}$, montrer que $\pi_n \rightharpoonup \pi$ si et seulement si, pour toute observable $t(\tau,\cdot)$ avec $\tau \in \bigsqcup_{k\in\mathbb{N}} \mathfrak{S}(k)$, $\lim_{n\to\infty} t(\tau,\pi_n) = t(\tau,\pi)$. On pourra utiliser la compacité de $\mathscr{P}$ pour le sens réciproque.
5. Montrer que pour tout permuton π, $(\pi_n(\pi))_{n\in\mathbb{N}}$ converge en probabilité vers π dans l'espace $\mathscr{P}$.
6. Soit $(\sigma_n)_{n\in\mathbb{N}}$ une suite de permutations, avec $\sigma_n \in \mathfrak{S}(n)$ pour tout n. On suppose que $(\sigma_n)_{n\in\mathbb{N}}$ converge au sens de la densité des motifs, et on note $t(\tau,\sigma_\infty) = \lim_{n\to\infty} t(\tau,\sigma_n)$. Montrer qu'il existe un unique permuton π tel que $t(\tau,\sigma_\infty) = t(\tau,\pi)$ pour tout motif τ.
7. (Bonus) On fixe $\pi \in \mathscr{P}$ et τ un motif. Calculer la limite de
$$\mathrm{var}(\sqrt{n}(t(\tau,\sigma_n(\pi)) - t(\tau,\pi))).$$
Montrer que $(\sqrt{n}(t(\tau,\sigma_n(\pi)) - t(\tau,\pi)))_{n\in\mathbb{N}}$ converge en loi vers une gaussienne.

Corrigé.

1.1. Il y a deux parties qui font apparaître 213 comme motif de 245361, à savoir, $\{2,4,5\}$ et $\{3,4,5\}$. La densité de motif est donc

$$t(213, 245361) = \frac{2}{\binom{6}{3}} = \frac{2}{20} = \frac{1}{10}.$$

1.2. Si $(X_1,Y_1),\dots,(X_k,Y_k)$ sont des variables indépendantes de loi $\pi \in \mathscr{P}$, alors les variables $X_1,\dots,X_k$ sont indépendantes, et toutes de loi uniforme sur $[0,1]$. Par Fubini,

$$\mathbb{P}[X_i = X_j] = \iint_{[0,1]^2} 1_{x=y}\, dx\, dy = \int_{[0,1]} 0\, dx = 0,$$

donc toutes ces variables sont distinctes avec probabilité 1. Il en va de même pour les variables $Y_1,\dots,Y_k$ qui vérifient les mêmes hypothèses. Donc, $((X_1,Y_1),\dots,(X_k,Y_k))$ est en configuration générale avec probabilité 1.

1.3. Si υ est la loi uniforme sur le carré $[0,1]^2$, alors $((X_1,Y_1),\dots,(X_k,Y_k)) \sim \upsilon^{\otimes k}$ est un vecteur de $2k$ variables indépendantes de loi uniforme sur $[0,1]$. Notons θ_1 et θ_2 les deux permutations aléatoires telles que $X_{\theta_1(1)} < X_{\theta_1(2)} < \cdots < X_{\theta_1(k)}$ et $Y_{\theta_2(1)} < Y_{\theta_2(2)} < \cdots < Y_{\theta_2(k)}$. Ces deux permutations indiquent l'ordre relatif des variables X_i et Y_i ; elles sont donc indépendantes, car les vecteurs $(X_1,\dots,X_k)$ et $(Y_1,\dots,Y_k)$ le sont. De plus, pour tout $\sigma \in \mathfrak{S}(k)$,

$$\mathbb{P}[\theta_1 = \sigma] = \int_{[0,1]^k} 1_{x_{\theta_1(1)} < x_{\theta_1(2)} < \cdots < x_{\theta_1(k)}} dx_1 \cdots dx_k = \int_{[0,1]^k} 1_{u_1 < u_2 < \cdots < u_k} du_1 \cdots du_k$$

ne dépend pas de σ, donc vaut la même chose pour tout $\sigma \in \mathfrak{S}(k)$, c'est-à-dire $\frac{1}{k!}$. Il en va de même pour θ_2, donc θ_1 et θ_2 sont deux permutations indépendantes de loi uniforme sur $\mathfrak{S}(k)$. Finalement, la configuration $\mathrm{conf}((X_1,Y_1),\dots,(X_k,Y_k))$ est l'unique permutation τ telle que $\theta_1 = \theta_2 \circ \tau$. Donc, $\tau = (\theta_2)^{-1} \circ \theta_1$ est le produit de deux permutations uniformes indépendantes, et elle suit donc une loi uniforme sur $\mathfrak{S}(k)$. On conclut que

$$t(\tau, \upsilon) = \frac{1}{k!} \quad \forall \tau \in \mathfrak{S}(k).$$

1.4. Le nombre de parties A qui font apparaître τ comme motif de la permutation $\sigma_n(\pi)$ est $\sum_{|A|=k,\, A \subset [\![1,n]\!]} C_{A,\tau}$. On a donc

$$t(\tau, \sigma_n(\pi)) = \frac{1}{\binom{n}{k}} \sum_{|A|=k,\, A \subset [\![1,n]\!]} C_{A,\tau}.$$

Pour montrer que $\mathbb{E}[t(\tau, \sigma_n(\pi))] = t(\tau, \pi)$, il suffit donc de montrer qu'on a la même égalité avec $\mathbb{E}[C_{A,\tau}]$. Or,

$$\begin{aligned} \mathbb{E}[C_{A,\tau}] &= \mathbb{P}_\pi[\mathrm{conf}((X_{a_1},Y_{a_1}),\dots,(X_{a_k},Y_{a_k})) = \tau] \\ &= \mathbb{P}_\pi[\mathrm{conf}((X_1,Y_1),\dots,(X_k,Y_k)) = \tau] = t(\tau, \pi). \end{aligned}$$

1.5. Si les parties A et B vérifient $A \cap B = \emptyset$, alors les vecteurs $((X_{a_1}, Y_{a_1}), \dots, (X_{a_k}, Y_{a_k}))$ et $((X_{b_1}, Y_{b_1}), \dots, (X_{b_k}, Y_{b_k}))$ sont indépendants, donc les variables $C_{A,\tau}$ et $C_{B,\tau}$ le sont également. Par conséquent,

$$\mathrm{var}(t(\tau, \sigma_n(\pi))) = \frac{1}{\binom{n}{k}^2} \sum_{A,B} \mathrm{cov}(C_{A,\tau}, C_{B,\tau}) = \frac{1}{\binom{n}{k}^2} \sum_{A,B \,|\, A \cap B \neq \emptyset} \mathrm{cov}(C_{A,\tau}, C_{B,\tau}),$$

où les sommes portent sur les parties de taille k dans $[\![1, n]\!]$. Chaque covariance est bornée en valeur absolue par 1, car les variables $C_{A,\tau}$ le sont. D'autre part, le nombre de paires (A, B) avec $A \cap B \neq \emptyset$ est

$$\binom{n}{k}\left(\binom{n}{k} - \binom{n-k}{k}\right).$$

En effet, il y a $\binom{n}{k}$ choix possibles pour A, puis $\binom{n-k}{k}$ choix possibles pour une partie B telle que $A \cap B = \emptyset$, donc $\binom{n}{k} - \binom{n-k}{k}$ choix possibles pour B telle que $A \cap B \neq \emptyset$. On conclut que

$$\mathrm{var}(t(\tau, \sigma_n(\pi))) \leq \frac{\binom{n}{k} - \binom{n-k}{k}}{\binom{n}{k}} = 1 - \frac{(n-k)(n-k-1)\cdots(n-2k+1)}{n(n-1)\cdots(n-k+1)}.$$

Le terme de droite est la probabilité pour qu'un arrangement aléatoire $(a_1, \dots, a_k)$ dans $[\![1, n]\!]$ rencontre $\{1, 2, \dots, k\}$. Cette probabilité est plus petite que la somme $\sum_{i,j=1}^{k} \mathbb{P}[a_i = j] = \frac{k^2}{n}$, d'où la borne sur la variance.

1.6. Par l'inégalité de Bienaymé-Chebyshev,

$$\forall \varepsilon > 0,\;\; \mathbb{P}[|t(\tau, \sigma_n(\pi)) - t(\tau, \pi)| \geq \varepsilon] \leq \frac{\mathrm{var}(t(\tau, \sigma_n(\pi)))}{\varepsilon^2} \leq \frac{k^2}{n\varepsilon^2} \to_{n \to \infty} 0,$$

d'où la convergence en probabilité de la suite $(t(\tau, \sigma_n(\pi)))_{n \in \mathbb{N}}$ vers $t(\tau, \pi)$.

1.7. Pour tout $x \in [0, 1]$, l'ensemble des y pour lesquels la densité $f(x, y)$ de π_σ vaut n est un intervalle $(\frac{\sigma(\lceil nx \rceil)-1}{n}, \frac{\sigma(\lceil nx \rceil)}{n}]$ de taille $\frac{1}{n}$, donc si $A \subset [\![0, 1]\!]$ est une partie mesurable, alors

$$\pi_\sigma(A \times [0, 1]) = \int_{x \in A} \int_{\frac{\sigma(\lceil nx \rceil)-1}{n}}^{\frac{\sigma(\lceil nx \rceil)}{n}} n \, dx \, dy = \int_{x \in A} dx$$

est la mesure de Lebesgue de A. La première marginale de π_σ est donc bien la mesure uniforme sur $[0, 1]$, et de même pour la seconde marginale, car

$$\pi_\sigma([0, 1] \times A) = \int_{y \in A} \int_{\frac{\sigma^{-1}(\lceil ny \rceil)-1}{n}}^{\frac{\sigma^{-1}(\lceil ny \rceil)}{n}} n \, dx \, dy = \int_{y \in A} dy.$$

Donc, π_σ est un permuton pour toute permutation $\sigma \in \mathfrak{S}(n)$.

Si $A = \{\mathrm{conf}((X_1, Y_1), \dots, (X_k, Y_k)) = \tau\}$, alors par définition $\mathbb{P}[A] = t(\tau, \pi_\sigma)$. Dans ce qui suit, on voit la probabilité donnée par le permuton π_σ comme la réunion de n carrés de taille $\frac{1}{n} \times \frac{1}{n}$, chacun étant de probabilité $\frac{1}{n}$. Si l'événement $B = \{\forall i \neq j,\; \lceil nX_i \rceil \neq \lceil nX_j \rceil\}$ est réalisé, alors les points $p_i = (X_i, Y_i)$ tombent tous dans des carrés différents de taille $\frac{1}{n} \times \frac{1}{n}$. Conditionnellement à cet événement, leur configuration est celle donnée par la permutation σ restreinte à la partie

$$\{a_1 < a_2 < \cdots < a_k\} = \{\lceil nX_i \rceil,\ i \in [\![1, k]\!]\},$$

et chacune de ces parties a la même probabilité (y compris conditionnellement à B). Donc,

$$\mathbb{P}[A|B] = \frac{1}{\binom{n}{k}} \sum_{A \subset [\![1,n]\!],\, |A|=k} 1_{A \text{ fait apparaître } \tau \text{ comme motif de } \sigma} = t(\tau, \sigma).$$

On a alors :

$$|t(\tau, \pi_\sigma) - t(\tau, \sigma)| = |\mathbb{P}[A] - \mathbb{P}[A|B]| = \mathbb{P}[A|B]\,(1 - \mathbb{P}[B]) \leq 1 - \mathbb{P}[B] = \mathbb{P}[B^{\mathrm{c}}].$$

Finalement, $\mathbb{P}[B^{\mathrm{c}}] \leq \sum_{1 \leq i < j \leq k} \mathbb{P}[\lceil nX_i \rceil = \lceil nX_j \rceil] = \sum_{1 \leq i < j \leq k} \frac{1}{n} = \frac{1}{n}\binom{k}{2}$.

1.8. Comme $t(\tau, \sigma_n(\pi))$ tend en probabilité vers $t(\tau, \pi)$, et $|t(\tau, \pi_n(\pi)) - t(\tau, \sigma_n(\pi))| \leq \frac{1}{n}\binom{k}{2}$ tend également en probabilité vers 0, la somme

$$t(\tau, \pi_n(\pi)) = t(\tau, \sigma_n(\pi)) + (t(\tau, \pi_n(\pi)) - t(\tau, \sigma_n(\pi)))$$

converge en probabilité vers $t(\tau, \pi)$ (la convergence en probabilité est compatible avec les opérations continues, en particulier la somme).

2.1. Comme $[0,1]^2$ est compact, $\mathscr{M}^1([0,1]^2)$ est également un espace métrique compact. Pour montrer que l'ensemble des permutons $\mathscr{P}$ est compact, il suffit donc de montrer qu'il est fermé. Notons υ la mesure de Lebesgue sur $[0,1]$, et considérons une suite $(\pi_n)_{n\in\mathbb{N}}$ de permutons qui converge en loi vers une mesure de probabilité π. Comme les projections

$$f_1 : (x, y) \in [0,1]^2 \mapsto x \in [0,1];$$
$$f_2 : (x, y) \in [0,1]^2 \mapsto y \in [0,1]$$

sont des applications continues, $(f_1)_*\pi_n \rightharpoonup (f_1)_*\pi$ et $(f_2)_*\pi_n \rightharpoonup (f_2)_*\pi$. Or, $(f_1)_*\pi_n = (f_2)_*\pi_n = \upsilon$ pour tout n, donc $(f_1)_*\pi = (f_2)_*\pi = \upsilon$ et π est un permuton.

2.2. Si π est un permuton, alors pour tout x, y, η, ε,

$$\begin{aligned} F_\pi(x+\eta, y+\varepsilon) &= \mathbb{P}_\pi[X \leq x+\eta,\, Y \leq y+\varepsilon] \\ &\leq \mathbb{P}_\pi[X \leq x,\, Y \leq y] + \mathbb{P}_\pi[x < X \leq x+\varepsilon] + \mathbb{P}_\pi[y < Y \leq y+\eta] \\ &\leq F_\pi(x,y) + \eta + \varepsilon. \end{aligned}$$

Ceci implique que pour tout (x, y),

$$\begin{aligned} F_\pi(x,y) = \mathbb{P}_\pi[X \leq x,\, Y \leq y] &= \mathbb{P}_\pi[X \leq x,\, Y < y] \\ = \mathbb{P}_\pi[X < x,\, Y \leq y] &= \mathbb{P}_\pi[X < x,\, Y < y]. \end{aligned}$$

Ensuite, si $[a,b] \times [c,d]$ est un pavé dans $[0,1]^2$, alors

$$\pi([a,b] \times [c,d]) = F_\pi(b,d) + F_\pi(a,c) - F_\pi(a,d) - F_\pi(b,c)$$

et on a la même formule pour les pavés ouverts ou semi-ouverts. Par conséquent, si $F_\pi = F'_\pi$, alors π et π' ont la même valeur sur les pavés.

Soit f une fonction continue sur $[0,1]^2$, et $\varepsilon > 0$. Comme $[0,1]^2$ est un compact, f est uniformément continue, et il existe $n \geq 1$ tel que

$$\left| f(x,y) - f\left(\frac{\lceil nx \rceil}{n}, \frac{\lceil ny \rceil}{n}\right)\right| \leq \varepsilon$$

pour tout $(x,y) \in [0,1]^2$. Notons $f_n(x,y) = f(\frac{\lceil nx \rceil}{n}, \frac{\lceil ny \rceil}{n})$ l'approximation de f qui est constante sur chaque petit carré $[\frac{k-1}{n}, \frac{k}{n}] \times [\frac{l-1}{n}, \frac{l}{n}]$. On a

$$|\pi(f) - \pi'(f)| \leq |\pi(f_n) - \pi'(f_n)| + 2\varepsilon.$$

Or,

$$\pi(f_n) = \sum_{k,l=1}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right) \pi\left(\left[\frac{k-1}{n}, \frac{k}{n}\right] \times \left[\frac{l-1}{n}, \frac{l}{n}\right]\right)$$

et on obtient le même résultat pour $\pi'(f_n)$ si $F_\pi = F_{\pi'}$. Ainsi, si $F_\pi = F_{\pi'}$, alors $\pi(f) = \pi'(f)$ pour toute fonction f continue sur $[0,1]^2$. On sait alors que ceci implique $\pi = \pi'$.

2.3. Notons que pour toute permutation $\sigma \in \mathfrak{S}(n)$,

$$t(\sigma, \pi) = \mathbb{E}[t(\sigma, \sigma_n(\pi))] = \mathbb{P}[\sigma_n(\pi) = \sigma].$$

Par conséquent, si $t(\tau, \pi) = t(\tau, \pi')$ pour tout $\tau \in \bigsqcup_{k \in \mathbb{N}} \mathfrak{S}(k)$, alors les modèles de permutations aléatoires $(\sigma_n(\pi))_{n \in \mathbb{N}}$ et $(\sigma_n(\pi'))_{n \in \mathbb{N}}$ sont les mêmes.

Fixons $\varepsilon > 0$, et $(x,y) \in [0,1]^2$; on note $k = \lceil nx \rceil$ et $l = \lceil ny \rceil$. La fonction $F_{n,\pi}(x,y)$ est égale à $\frac{1}{n}$ fois le nombre de points (X_i, Y_i) tels que :

(a) X_i fait partie des k plus petites valeurs de $(X_1, \ldots, X_n)$;

(b) Y_i fait partie des l plus petites valeurs de $(Y_1, \ldots, Y_n)$.

On peut donc réécrire :

$$F_{n,\pi}(x,y) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq X_{\theta_1(k)},\, Y_i \leq Y_{\theta_2(l)}}.$$

Comme $(X_1, \ldots, X_n)$ est un vecteur de variables indépendantes uniformes sur $[0,1]$, $X_{\theta_1(k)} - \frac{k}{n}$ tend en probabilité vers 0. En effet,

$$\begin{aligned}\mathbb{P}\left[\left|X_{\theta_1(k)} - \frac{k}{n}\right| \geq \varepsilon\right] \leq \mathbb{P}&\left[\frac{1}{n}\sum_{i=1}^{n} 1_{X_i \geq \frac{k}{n} + \varepsilon} \geq 1 - \frac{k-1}{n}\right] \\ &+ \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq \frac{k}{n} - \varepsilon} \geq \frac{k}{n}\right]\end{aligned}$$

qui tend vers 0 : par le théorème de Glivenko-Cantelli, la fonction de répartition empirique $\frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq x}$ tend presque sûrement en norme infinie vers la fonction de répartition de la loi uniforme (on peut même montrer des bornes exponentiellement petites sur les probabilités ci-dessus). De même, $Y_{\theta_2(l)} - \frac{l}{n}$ tend en probabilité vers 0, donc pour $n \geq n_0$,

$$\frac{k}{n} - \varepsilon \leq X_{\theta_1(k)} \leq \frac{k}{n} + \varepsilon \quad \text{et} \quad \frac{l}{n} - \varepsilon \leq Y_{\theta_2(l)} \leq \frac{l}{n} + \varepsilon$$

avec probabilité plus grande que $1-\varepsilon$. Ceci implique, avec probabilité supérieure à $1-\varepsilon$,

$$\frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq \frac{k}{n}-\varepsilon,\, Y_i \leq \frac{l}{n}-\varepsilon} \leq F_{n,\pi}(x,y) \leq \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq \frac{k}{n}+\varepsilon,\, Y_i \leq \frac{l}{n}+\varepsilon}.$$

Pour n assez grand, l'approximation $\frac{k}{n} \pm \varepsilon$ est proche de x à 2ε près, et de même pour $\frac{l}{n} \pm \varepsilon$ par rapport à y. On conclut qu'avec probabilité supérieure à $1-\varepsilon$ et pour $n \geq n_0$,

$$\frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq x-2\varepsilon,\, Y_i \leq y-2\varepsilon} \leq F_{n,\pi}(x,y) \leq \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq x+2\varepsilon,\, Y_i \leq y+2\varepsilon}.$$

Par la loi (faible) des grands nombres, le terme de droite tend en probabilité vers $F_\pi(x-2\varepsilon, y-2\varepsilon)$, et le terme de gauche vers $F_\pi(x+2\varepsilon, y+2\varepsilon)$, donc pour n assez grand, avec probabilité plus grande que $1-2\varepsilon$,

$$F_\pi(x,y)-5\varepsilon \leq F_\pi(x-2\varepsilon,y-2\varepsilon)-\varepsilon \leq F_{n,\pi}(x,y) \leq F_\pi(x+2\varepsilon,x+2\varepsilon)+\varepsilon \leq F_\pi(x,y)+5\varepsilon,$$

en utilisant les propriétés de croissance et de régularité de la fonction F_π. On a donc bien $F_{n,\pi}(x,y) \to_{\mathbb{P}} F_\pi(x,y)$, et on peut reconstruire π à partir des modèles $(\sigma_n(\pi))_{n\in\mathbb{N}}$, et donc à partir des observables $t(\tau,\pi)$. En particulier, si $t(\tau,\pi) = t(\tau,\pi')$ pour tout motif τ, alors $\pi = \pi'$.

2.4. Si $\pi_n \rightharpoonup \pi$, alors $(\pi_n)^{\otimes k} \rightharpoonup \pi^{\otimes k}$ pour tout $k \geq 1$. En effet, comme $([0,1]^2)^k$ est compact, on sait qu'il existe des sous-suites convergentes en loi $((\pi_{n_l})^{\otimes k})_{l\in\mathbb{N}}$ dans l'espace compact $\mathscr{M}^1(([0,1]^2)^k)$. Pour montrer que $(\pi_n)^{\otimes k} \rightharpoonup \pi^{\otimes k}$, il suffit donc de montrer que l'unique limite possible d'une sous-suite convergente est $\pi^{\otimes k}$. Or, si $(\pi_{n_l})^{\otimes k} \rightharpoonup M$, alors pour tout pavé $P = ([a_1,b_1]\times[c_1,d_1]) \times \cdots \times ([a_k,b_k]\times[c_k,d_k])$,

$$\lim_{l\to\infty} (\pi_{n_l})^{\otimes k}(P) = \lim_{l\to\infty} \prod_{j=1}^{k} \pi_{n_l}([a_i,b_i]\times[c_i,d_i]) = \prod_{j=1}^{k} \pi([a_i,b_i]\times[c_i,d_i]) = \pi^{\otimes k}(P).$$

On en déduit que $M = \pi^{\otimes k}$ sur les pavés, et on peut ensuite utiliser le même raisonnement qu'à la question 2.2. pour montrer que ceci implique $M = \pi^{\otimes k}$.

Supposons maintenant $\pi_n \rightharpoonup \pi$. Alors, pour tout $\tau \in \mathfrak{S}(k)$, l'ensemble P_τ des familles $((x_1,y_1),\ldots,(x_k,y_k))$ de $([0,1]^2)^k$ qui ont pour configuration τ s'écrit

$$P_\tau = \bigsqcup_{\theta\in\mathfrak{S}(k)} \left\{x_{\theta(1)} < x_{\theta(2)} < \cdots < x_{\theta(k)} \text{ et } y_{\theta\circ\tau^{-1}(1)} < y_{\theta\circ\tau^{-1}(2)} < \cdots < y_{\theta\circ\tau^{-1}(k)}\right\}.$$

En particulier, la frontière topologique de P_τ est incluse dans l'ensemble S des familles $((x_1,y_1),\ldots,(x_k,y_k))$ de $([0,1]^2)^k$ telles que $x_i = x_j$ pour un couple $i<j$, ou $y_i = y_j$ pour un certain couple $i<j$ (familles qui ne sont pas en configuration générale). Pour tout permuton π, $\pi^{\otimes k}(S) = 0$, car les lois uniformes sur $[0,1]^k$ donnent avec

probabilité 1 des k-uplets de points distincts. Par conséquent, $\pi^{\otimes k}(\partial P_\tau) = 0$, donc, par le théorème de Portmanteau,

$$t(\tau, \pi_n) = (\pi_n)^{\otimes k}(P_\tau) \to \pi^{\otimes k}(P_\tau) = t(\tau, \pi).$$

Réciproquement, supposons qu'on ait la convergence de toutes les observables $t(\tau, \pi_n)$ vers celles de π. Comme $\mathscr{P}$ est compact, il existe des sous-suites convergentes $(\pi_{n_l})_{l\in\mathbb{N}}$, et il suffit donc de montrer que la limite ϖ d'une telle sous-suite convergente est π. Or, par le sens direct de l'implication,

$$t(\tau, \varpi) = \lim_{l\to\infty} t(\tau, \pi_{n_l}) = t(\tau, \pi),$$

donc $\varpi = \pi$ par la question 2.3.

2.5. Fixons un voisinage ouvert V de π dans $\mathscr{P}$ (qui est métrisé par exemple par la restriction de la métrique de Lévy-Prohorov). La question précédente montre que ce voisinage contient un ouvert pour la topologie de la convergence des observables : il existe $\tau_1, \ldots, \tau_k$ et $\varepsilon_1, \ldots, \varepsilon_k > 0$ tels que

$$V \supset \{\varpi \in \mathscr{P} \,|\, \forall i \in [\![1, k]\!]\,,\ |t(\tau_i, \varpi) - t(\tau_i, \pi)| < \varepsilon_i\}.$$

D'après 1.8., $t(\tau, \pi_n(\pi))$ converge en probabilité vers $t(\tau, \pi)$ pour toute permutation τ, donc la probabilité pour que $\pi_n(\pi)$ appartienne au terme de droite tend vers 1 lorsque n tend vers l'infini. *A fortiori*, $\lim_{n\to\infty} \mathbb{P}[\pi_n(\pi) \in V] = 1$, donc $\pi_n(\pi) \to_{\mathbb{P}} \pi$.

2.6. Comme $\mathscr{P}$ est compact, il existe des sous-suites convergentes à la suite de permutons $(\pi_{\sigma_n})_{n\in\mathbb{N}}$. De plus, pour tout motif τ, si ϖ est une limite d'une telle suite, alors

$$t(\tau, \varpi) = \lim_{l\to\infty} t(\tau, \pi_{\sigma_{n_l}}) = \lim_{l\to\infty} t(\tau, \sigma_{n_l}) = t(\tau, \sigma_\infty),$$

en utilisant l'inégalité de la question 1.7. pour la seconde égalité. Les limites possibles ont donc toutes les mêmes observables, donc sont toutes égales par 2.3. On conclut que $\pi_{\sigma_n} \to \pi$ pour un certain permuton π, et que $t(\tau, \sigma_\infty) = t(\tau, \pi)$.