

Étude asymptotique des arbres sous leur mesure de Plancherel

Shengjun ZHANG

Directeur: Pierre-Loïc Méliot (LMO, Orsay)

Résumé

Nous souhaitons étudier des arbres enracinés aléatoires qui ont de nombreuses similarités avec les partitions d'entiers sous la célèbre mesure de Plancherel. Si n est un entier, la somme sur les arbres enracinés à n arêtes du nombre de paires d'étiquetages croissants de l'arbre est égale au produit des n premiers entiers triangulaires $\binom{k+1}{2}$. Cette identité permet de définir une mesure de probabilité sur chaque ensemble $\mathfrak{T}(n)$ des arbres enracinés à n arêtes. Le but est de chercher à décrire les propriétés asymptotiques des arbres aléatoires correspondants : profondeur typique, profil, limite locale et globale, spectre, etc. Les méthodes algébriques et combinatoires employées en probabilités intégrables pour étudier les partitions aléatoires devraient avoir des analogues dans ce contexte.

Table des matières

1	Introduction	2
2	Partitions et tableaux standards	2
2.1	Définitions générales	2
2.2	Détails sur quelques notions	3
3	Arbres enracinés et étiquetages standards	6
3.1	Définitions générales	6
3.2	Détails sur quelques notions	8
4	Questions et outils	10

1 Introduction

Dans ce mémoire, nous nous intéressons aux arbres enracinés, notamment leur comportement asymptotique sous leur mesure de Plancherel, qui a des similarités avec la mesure de Plancherel sur des partitions. La première partie de ce texte rappelle les définitions de la mesure de Plancherel sur les partitions et son intérêt pour l'étude des permutations aléatoires, on résume des notions importants et des résultats connus sur ce objet. La seconde partie introduit une mesure analogue sur les arbres enracinés : c'est l'objet que l'on souhaite étudier, en particulier lorsque la taille des arbres tend vers l'infini. La dernière section liste les principales questions que l'on souhaite aborder, ainsi que les outils que l'on pourrait utiliser pour y répondre.

2 Partitions et tableaux standards

2.1 Définitions générales

Une *partition* de taille n est une suite décroissante d'entiers $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell)$ avec $n = |\lambda| = \sum_{i=1}^{\ell} \lambda_i$. Une telle suite est le plus souvent représentée par son *diagramme de Young*, qui est le tableau avec λ_1 cases sur la première ligne, λ_2 cases sur la seconde ligne, etc.

$$(5, 3, 2) = \begin{array}{|c|c|c|c|c|} \hline & & & & \\ \hline & & & & \\ \hline & & & & \\ \hline & & & & \\ \hline \end{array}$$

Notons $\mathfrak{Y}(n)$ l'ensemble des partitions de taille n . La fonction génératrice des cardinaux des ensembles $\mathfrak{Y}(n)$ est :

$$P(z) = \sum_{n=0}^{\infty} |\mathfrak{Y}(n)| z^n = \prod_{k=1}^{\infty} \frac{1}{1 - z^k} ,$$

ce qui donne une formule implicite pour $|\mathfrak{Y}(n)|$.

Les partitions d'entiers apparaissent souvent comme indices d'objets combinatoires ou algébriques (classes de conjugaison dans le groupe symétrique, représentations irréductibles des groupes classiques, etc.), et dans ces contextes, elles sont associées à des poids non uniformes. La *mesure de Plancherel* est une mesure de probabilité de ce type ; elle peut être définie combinatoirement à l'aide des nombres de *tableaux standards*. Si $\lambda \in \mathfrak{Y}(n)$, un *tableau standard* de forme λ est un remplissage des cases du diagramme de Young de λ par les entiers $1, 2, \dots, n$, de sorte que les lignes et les colonnes sont strictement croissantes. Par exemple,

$$\begin{array}{|c|c|c|c|c|} \hline 6 & 9 & & & \\ \hline 3 & 5 & 8 & & \\ \hline 1 & 2 & 4 & 7 & 10 \\ \hline \end{array}$$

est un tableau standard de forme $\lambda = (5, 3, 2)$. Notons $\text{ST}(\lambda)$ l'ensemble des tableaux standards de forme λ . On peut montrer (voir par exemple [GNW82]) que :

$$|\text{ST}(\lambda)| = \frac{|\lambda|!}{\prod_{\square \in \lambda} h(\square, \lambda)} ,$$

où le dénominateur est le produit sur toutes les cases du diagramme de λ des *longueurs d'équerres* $h(\square, \lambda)$. La longueur d'équerre de \square est le nombre de cases dans le plus grand crochet en L que l'on peut dessiner dans λ avec le coin en \square . Par exemple, $(5, 3, 2)$ a les longueurs d'équerres suivantes :

$$\begin{array}{|c|c|c|c|c|} \hline 2 & 1 & & & \\ \hline 4 & 3 & 1 & & \\ \hline 7 & 6 & 4 & 2 & 1 \\ \hline \end{array}$$

ce qui implique que $|\text{ST}(\lambda)| = 450$.

Pour tout $n \in \mathbb{N}$, on a l'identité fondamentale suivante :

$$n! = \sum_{\lambda \in \mathfrak{Y}(n)} |\text{ST}(\lambda)|^2 .$$

Une preuve de cette identité repose sur l’*algorithme de Robinson–Schensted–Knuth*, qui donne une bijection explicite entre les permutations $\sigma \in \mathfrak{S}(n)$, et les paires (P, Q) de tableaux standards de même forme $\lambda \in \mathfrak{Y}(n)$; voir [Rob38, Sch61, Knu70]. Une autre preuve repose sur la décomposition de l’algèbre de groupe $\mathbb{C}\mathfrak{S}(n)$, qui est semisimple, en somme directe d’algèbres de matrices $\text{End}(V_\lambda)$, avec une algèbre de matrices pour chaque représentation irréductible V_λ de $\mathfrak{S}(n)$. La *mesure de Plancherel* sur les partitions de taille n se déduit de l’identité fondamentale :

$$\mathbb{P}^{(n)}[\lambda] = \frac{|\text{ST}(\lambda)|^2}{n!} .$$

C’est donc la mesure image de la loi uniforme sur $\mathfrak{S}(n)$ par l’application $\sigma \mapsto \lambda(\sigma)$, où $\lambda(\sigma)$ est la forme commune des tableaux $P(\sigma)$ et $Q(\sigma)$ associés à σ par l’algorithme RSK. Les propriétés asymptotiques de $\lambda \sim \mathbb{P}^{(n)}$ ont été étudiées dans de très nombreux articles à partir des années 70, en lien avec le *problème d’Ulam* : la taille de λ_1 sous $\mathbb{P}^{(n)}$ a la loi de la longueur $\ell^{(n)}$ d’une plus longue sous-suite croissante dans une permutation $\sigma^{(n)}$ choisie uniformément dans $\mathfrak{S}(n)$. En particulier, Logan–Shepp et Kerov–Vershik [LS77, KV77] ont montré que le diagramme de Young de $\lambda \sim \mathbb{P}^{(n)}$ avait une forme limite lorsque n tendait vers l’infini. Plus tard, la réinterprétation de la mesure de Plancherel comme un processus ponctuel a permis de comprendre l’asymptotique des plus grandes parts de la partition λ . Les articles [BOO00], [Oko00] et [Joh01] démontrent par des méthodes complémentaires la *correspondance de Baik–Deift–Johansson* :

$$n^{\frac{1}{3}} \left(\frac{\lambda_1}{2\sqrt{n}} - 1, \frac{\lambda_2}{2\sqrt{n}} - 1, \dots \right) \underset{n \rightarrow \infty}{\sim \text{loi}} n^{\frac{2}{3}} \left(\frac{x_1}{2\sqrt{n}} - 1, \frac{x_2}{2\sqrt{n}} - 1, \dots \right) ,$$

où le terme de droite met en jeu les plus grandes valeurs propres x_1, x_2, \dots d’une matrice hermitienne gaussienne de taille $n \times n$ (GUE); le processus limite commun est le processus déterminantal d’Airy. En particulier,

$$n^{\frac{1}{3}} \left(\frac{\lambda_1}{2\sqrt{n}} - 1 \right) \xrightarrow{n \rightarrow \infty} \text{TW} ,$$

le terme de droite désignant la loi dite *de Tracy–Widom* (voir [TW94]).

2.2 Détails sur quelques notions

On présente d’abord l’algorithme RSK qui associe σ une permutation de taille n à deux tableaux standards $P(\sigma)$ et $Q(\sigma)$ de même forme :

- On construit récursivement $P(\sigma)$ et $Q(\sigma)$ en lisant les lettres $\sigma(1), \sigma(2), \dots, \sigma(i)$, et le tableau Q contient les lettres $1, 2, \dots, i$.
- À l’étape i : on essaie d’insérer $\sigma(i)$ au bout de la première ligne de P . Si $\sigma(i)$ est plus grand que toutes les entrées de cette première ligne, c’est bon. Sinon, si $k = \sigma(h)$ est la plus petite entrée plus grande que $\sigma(i)$ dans cette première ligne, on remplace $\sigma(h)$ par $\sigma(i)$ dans la première ligne, et on essaie d’insérer $\sigma(h)$ dans la seconde ligne avec les même règles. On construit ceci récursivement, jusqu’à avoir résolu toutes les insertions.
- Ceci décrit la croissance de P à la i -ième étape. La forme du dit tableau est une partition dont le diagramme de Young a maintenant une case de plus, et on enregistre cette croissance dans le tableau Q avec une case numérotée i .

Par exemple, si $\sigma = (264519378) \in \mathfrak{S}(9)$, alors les tableaux P et Q croissent comme suit :

$$\begin{array}{ccc}
 \boxed{2} & \boxed{1} & ; & \boxed{2} \boxed{6} & \boxed{1} \boxed{2} & ; & \begin{array}{|c|} \hline 6 \\ \hline 2 \ 4 \end{array} & \begin{array}{|c|} \hline 3 \\ \hline 1 \ 2 \end{array} \\
 \\
 \begin{array}{|c|} \hline 6 \\ \hline 2 \ 4 \ 5 \end{array} & \begin{array}{|c|} \hline 3 \\ \hline 2 \ 1 \ 4 \end{array} & ; & \begin{array}{|c|} \hline 6 \\ \hline 2 \\ \hline 1 \ 4 \ 5 \end{array} & \begin{array}{|c|} \hline 5 \\ \hline 3 \\ \hline 1 \ 2 \ 4 \end{array} \\
 \\
 \begin{array}{|c|} \hline 6 \\ \hline 2 \\ \hline 1 \ 4 \ 5 \ 9 \end{array} & \begin{array}{|c|} \hline 5 \\ \hline 3 \\ \hline 1 \ 2 \ 4 \ 6 \end{array} & ; & \begin{array}{|c|} \hline 6 \\ \hline 2 \ 4 \\ \hline 1 \ 3 \ 5 \ 9 \end{array} & \begin{array}{|c|} \hline 5 \\ \hline 3 \ 7 \\ \hline 1 \ 2 \ 4 \ 6 \end{array} \\
 \\
 \begin{array}{|c|} \hline 6 \\ \hline 2 \ 4 \ 9 \\ \hline 1 \ 3 \ 5 \ 7 \end{array} & \begin{array}{|c|} \hline 5 \\ \hline 3 \ 7 \ 8 \\ \hline 1 \ 2 \ 4 \ 6 \end{array} & ; & \begin{array}{|c|} \hline 6 \\ \hline 2 \ 4 \ 9 \\ \hline 1 \ 3 \ 5 \ 7 \ 8 \end{array} & \begin{array}{|c|} \hline 5 \\ \hline 3 \ 7 \ 8 \\ \hline 1 \ 2 \ 4 \ 6 \ 9 \end{array} .
 \end{array}$$

Rappelons que $\lambda(\sigma)$ est la forme commune des deux tableaux $P(\sigma)$ et $Q(\sigma)$; c'est une partition de taille n . Dans l'exemple précédent, $\lambda(\sigma) = (5, 3, 1)$. De façon générale, la première part $\lambda_1(\sigma)$ est égale à la longueur d'un plus long sous-mot croissant de σ ; dans l'exemple précédent, 24578 est un sous-mot croissant de longueur $\lambda_1 = 5$, et on peut vérifier qu'il n'y en a pas de plus grand.

Si $\sigma^{(n)} \sim \text{Unif}(\mathfrak{S}(n))$, alors $\lambda^{(n)} = \lambda(\sigma^{(n)})$ suit la mesure de Plancherel sur l'ensemble $\mathfrak{P}(n)$ des partitions de tailles n :

$$\mathbb{P}[\lambda^{(n)} = \lambda] = \mathbb{P}[P(\sigma^{(n)}) \text{ et } Q(\sigma^{(n)}) \text{ ont pour forme } \lambda] = \frac{|\text{ST}(\lambda)|^2}{n!} .$$

En effet, $\sigma \mapsto (P(\sigma), Q(\sigma))$ est une bijection de $\mathfrak{S}(n)$ vers l'ensemble des paires de tableaux standards de même forme et de taille n . Notons $\ell(\sigma) = \lambda_1(\sigma)$ la longueur d'un plus long sous-mot croissant dans une permutation σ et $\ell^{(n)} = \ell(\sigma^{(n)})$.

Théorème 2.1. (Logan-Shepp, Kerov-Vershik). *On a la convergence en probabilité :*

$$\frac{\ell^{(n)}}{2\sqrt{n}} \xrightarrow{\mathbb{P}} 1 .$$

La preuve de cette loi des grands nombres pour la longueur du plus long sous-mot croissant peut être décomposée en deux parties : montrer que $\ell^{(n)} \geq (2 - \varepsilon)\sqrt{n}$ avec grande probabilité, et montrer que $\ell^{(n)} \leq (2 + \varepsilon)\sqrt{n}$ aussi avec grande probabilité. Cette seconde partie découle de la première et de l'inégalité élémentaire suivante :

Théorème 2.2. (Kerov-Vershik). *Pour tout $n \geq 1$, $\mathbb{E}[\ell^{(n)}] \leq 2\sqrt{n}$.*

Démonstration. On montre $\mathbb{E}[\ell^{(n)} - \ell^{(n-1)}] \leq \frac{1}{\sqrt{n}}$. Ceci implique que

$$\mathbb{E}[\ell^{(n)}] \leq \sum_{k=1}^n \frac{1}{\sqrt{k}} \leq \int_0^n \frac{1}{\sqrt{x}} dx = 2\sqrt{n} .$$

Soit $\sigma^{(n)} \sim \text{Unif}(\mathfrak{S}(n))$, et $\sigma^{(n,n-1)}$ le sous-mot formé par les $n - 1$ premières lettres de $\sigma^{(n)}$. Ces lettres ne forment pas forcément une permutation de $\llbracket 1, n \rrbracket$ (ce n'est pas le cas que si $\sigma^{(n)}(n) = n$), mais leur ordre relatif est une permutation uniforme de taille $n - 1$, donc le plus long sous-mot croissant dans $\sigma^{(n,n-1)}$ a la loi de $\ell^{(n-1)}$. Cette argument permet de coupler $\ell^{(n)}$ et $\ell^{(n-1)}$: $\ell^{(n)}$ est la longueur de la première ligne de $P(\sigma^{(n)})$ et $Q(\sigma^{(n)})$, et $\ell^{(n-1)}$ peut être représenté en loi par la longueur de la première ligne des tableaux P et Q obtenus en lisant $\sigma^{(n)}(1) \dots \sigma^{(n)}(n - 1)$. On a donc :

$$\begin{aligned} & \mathbb{E}[\ell^{(n)} - \ell^{(n-1)}] \\ &= \mathbb{P}[\text{on augmente la taille de la première ligne en lisant } \sigma^{(n)}(n) \text{ l'algorithme RSK}] \\ &= \mathbb{P}[n \text{ apparaît sur la première ligne de } Q(\sigma^{(n)})] . \end{aligned}$$

Notons $n \in Q_1(\sigma^{(n)})$ pour la condition obtenue. Alors, en combinant la formule des probabilités totales et l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} (\mathbb{E}[\ell^{(n)} - \ell^{(n-1)}])^2 &= \left(\sum_{\lambda \in \mathfrak{P}(n)} \mathbb{P}[\lambda^{(n)} = \lambda] \mathbb{P}[n \in Q_1(\sigma^{(n)}) | \lambda(\sigma^{(n)}) = \lambda] \right)^2 \\ &\leq \sum_{\lambda \in \mathfrak{P}(n)} \mathbb{P}[\lambda^{(n)} = \lambda] (\mathbb{P}[n \in Q_1(\sigma^{(n)}) | \lambda(\sigma^{(n)}) = \lambda])^2 \\ &= \sum_{\lambda \in \mathfrak{P}(n)} \mathbb{P}[\lambda^{(n)} = \lambda] \mathbb{P}[n \in P_1(\sigma^{(n)}), n \in Q_1(\sigma^{(n)}) | \lambda(\sigma^{(n)}) = \lambda] \\ &= \mathbb{P}[n \in P_1(\sigma^{(n)}), n \in Q_1(\sigma^{(n)})] . \end{aligned}$$

En effet, sachant que $\lambda^{(n)}$, les deux tableaux $P(\sigma^{(n)})$ et $Q(\sigma^{(n)})$ sont indépendants et de loi uniforme sur l'ensemble $\text{ST}(\lambda)$ des tableaux standards de forme λ (parce que RSK est une bijection). Maintenant, on voit facilement que n apparaît sur la première ligne de $P(\sigma^{(n)})$ et $Q(\sigma^{(n)})$ si et seulement si $\sigma^{(n)} = n$ (on essaie alors d'insérer n sur la première ligne du tableau P : c'est toujours possible). Donc,

$$(\mathbb{E}[\ell^{(n)} - \ell^{(n-1)}])^2 \leq \mathbb{P}[\sigma^{(n)}(n) = n] = \frac{1}{n} ,$$

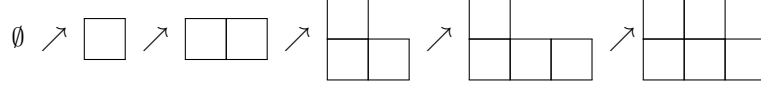
ce que l'on voulait démontrer. □

Ensuite, on présente une démonstration de la formule des équerres pour le nombre des tableaux standards d'une forme donnée, qui se base sur une marche aléatoire sur le diagramme de Young :

Théorème 2.3. (Frame-Robinson-Thrall). *Pour une partition $\lambda \in \mathfrak{Y}(n)$, le nombre des tableaux standards de forme λ est donné par la formule des équerres :*

$$|\text{ST}(\lambda)| = \frac{n!}{\prod_{\square \in \lambda} h(\square, \lambda)} .$$

Soit $\Lambda \in \mathfrak{Y}(n+1)$, on dit qu'une partition $\lambda \in \mathfrak{Y}(n)$ satisfait $\lambda \nearrow \Lambda$, si après avoir enlevé une case au bord supérieur droite du diagramme de Young de Λ , on obtient le diagramme de Young de λ . Par exemple, on a :



Étant donnée une partition Λ de taille $n+1$, on choisit une partition aléatoire de taille n parmi celles qui satisfont $\lambda \nearrow \Lambda$, en suivant la procédure suivante :

- Commençons par tirer uniformément une case aléatoire \square_1 du diagramme de Young de Λ . Chaque case a la probabilité $\frac{1}{n+1}$ d'être tirée.
- Si \square_1 est au bord supérieur droite de Λ , on peut l'enlever pour obtenir une partition λ telle que $\lambda \nearrow \Lambda$.
- Sinon, la longueur d'équerres $h(\square_1, \Lambda)$ est supérieure à 2, on peut donc choisir uniformément une nouvelle case aléatoire \square_2 dans l'équerre de \square_1 , à part \square_1 . Chaque case dans l'équerre de \square_1 a la probabilité $\frac{1}{h(\square_1, \Lambda)-1}$ d'être choisie.
- Si \square_2 est au bord supérieur droite de Λ , on l'enlève et obtient un $\lambda \in \mathfrak{Y}(n)$ telle que $\lambda \nearrow \Lambda$. Sinon, on continue la procédure en choisissant uniformément une case aléatoire \square_3 dans l'équerre de \square_2 , puis une case aléatoire \square_4 dans l'équerre de \square_3 , etc. jusqu'à ce qu'on obtienne une case à enlever.

À chaque étape, la longueur d'équerres d'une nouvelle case \square_i est strictement inférieure à celle de \square_{i-1} , donc cette procédure terminera presque sûrement.

Lemme 2.4. *Notons $p(\Lambda \rightarrow \lambda)$ la probabilité d'obtenir λ en enlevant une case aléatoire de Λ en suivant la procédure ci-dessus, et $f(\lambda)$ le terme de droite de la formule des équerres, alors on a :*

$$p(\Lambda \rightarrow \lambda) = \frac{f(\lambda)}{f(\Lambda)} .$$

Démonstration. Notons (x, y) les coordonnées de la case du diagramme de Young de Λ qui est enlevée pour obtenir λ . On a :

$$\frac{f(\lambda)}{f(\Lambda)} = \frac{1}{n+1} \frac{\prod_{\square \in \Lambda} h(\square, \Lambda)}{\prod_{\square \in \lambda} h(\square, \lambda)} ,$$

les cases qui ont une longueur d'équerres différentes dans λ et dans Λ sont celles dans la même ligne ou la même colonne que (x, y) . Donc,

$$\begin{aligned} \frac{f(\lambda)}{f(\Lambda)} &= \frac{1}{n+1} \prod_{i=1}^{x-1} \frac{h(i, y)}{h(i, y)-1} \prod_{j=1}^{y-1} \frac{h(x, j)}{h(x, j)-1} \\ &= \frac{1}{n+1} \prod_{i=1}^{x-1} \left(1 + \frac{1}{h(i, y)-1} \right) \prod_{j=1}^{y-1} \left(1 + \frac{1}{h(x, j)-1} \right) \\ &= \frac{1}{n+1} \sum_{\substack{I \subset \llbracket 1, x-1 \rrbracket \\ J \subset \llbracket 1, y-1 \rrbracket}} \frac{1}{\prod_{i \in I} (h(i, y)-1) \prod_{j \in J} (h(x, j)-1)} , \end{aligned}$$

où $h(a, b) = h((a, b), \Lambda)$.

Ensuite, on montre que l'on peut obtenir la même formule en suivant la procédure aléatoire de la suppression d'une case au bord. Soit

$$(\square_1, \square_2, \dots, \square_r = (x, y)) = ((x_1, y_1), (x_2, y_2), \dots, (x_r, y_r) = (x, y))$$

une suite des cases dans le diagramme de Young de Λ , telle que \square_i soit dans l'équerre de \square_{i-1} et que $\square_r = (x, y)$ soit au bord supérieure droite du diagramme de Young. En conséquence, cela peut être le chemin des cases qu'on suit dans la procédure de la suppression aléatoire d'une case au bord dans Λ . Soit $I = \{x_1, x_2, \dots, x_r\} \setminus \{x\}$ et $J = \{y_1, y_2, \dots, y_r\} \setminus \{y\}$. Remarquons qu'il est possible qu'il existe des répétitions dans les suites (x_1, x_2, \dots, x_r) et (y_1, y_2, \dots, y_r) , donc les ensembles I et J peuvent avoir des cardinaux différents et leurs cardinaux sont souvent plus petits que $r - 1$.

Notons $p(\Lambda \rightarrow \lambda, I, J)$ la probabilité d'obtenir $\lambda = \Lambda \setminus (x, y)$ et les projections de chemins dans la procédure aléatoires sont les ensembles I et J . On montrera que

$$p(\Lambda \rightarrow \lambda, I, J) = \frac{1}{n+1} \frac{1}{\prod_{i \in I} (h(i, y) - 1) \prod_{j \in J} (h(x, j) - 1)},$$

ce qui termine la preuve, en prenant la somme des paires (I, J) des ensembles possibles.

On raisonne par récurrence sur $s = |I| + |J|$. Si $s = 0$, alors la suite des cases contient seulement une case $\square_1 = (x, y)$, la probabilité de choisir cette case est $\frac{1}{n+1}$, c'est ce qu'on voudrais obtenir. Supposons que le résultat est vrai jusqu'à $s-1 \geq 0$. Remarquons que la première case $\square_1 = (x_1, y_1)$ d'un chemin est déterminé par deux ensembles I et J :

$$x_1 = \min(I \sqcup \{x\}) \quad ; \quad y_1 = \min(J \sqcup \{y\}) .$$

Ensuite, il y a deux possibilité pour \square_2 : soit $\square_2 = (x_2, y_1)$ avec $x_2 = \min(I \sqcup \{x\} \setminus \{x_1\})$, soit $\square_2 = (x_1, y_2)$ avec $y_2 = \min(J \sqcup \{y\} \setminus \{y_1\})$. En conséquence :

$$p(\Lambda \rightarrow \lambda, I, J) = \frac{1}{h(x_1, y_1) - 1} (p(\Lambda \rightarrow \lambda, I \setminus \{x_1\}, J) + p(\Lambda \rightarrow \lambda, I, J \setminus \{y_1\})) .$$

Par l'hypothèse de récurrence,

$$\begin{aligned} p(\Lambda \rightarrow \lambda, I \setminus \{x_1\}, J) &= \frac{1}{n+1} \frac{1}{\prod_{i \in I \setminus \{x_1\}} (h(i, y) - 1) \prod_{j \in J} (h(x, j) - 1)} \\ &= \frac{h(x_1, y) - 1}{n+1} \frac{1}{\prod_{i \in I} (h(i, y) - 1) \prod_{j \in J} (h(x, j) - 1)} , \end{aligned}$$

et de même,

$$p(\Lambda \rightarrow \lambda, I, J \setminus \{y_1\}) = \frac{h(x, y_1) - 1}{n+1} \frac{1}{\prod_{i \in I} (h(i, y) - 1) \prod_{j \in J} (h(x, j) - 1)} .$$

Pour obtenir le résultat avec $|I| + |J| = s$, il suffit de montrer que $h(x, y_1) + h(x_1, y) - 2 = h(x_1, y_1) - 1$, c'est un calcul simple. \square

Maintenant, on peut donner une démonstration de la formule d'équerres. Comme $p(\Lambda \rightarrow \cdot)$ est une probabilité sur l'ensemble des partitions λ telle que $\lambda \nearrow \Lambda$, on a

$$1 = \sum_{\lambda | \lambda \nearrow \Lambda} \frac{f(\lambda)}{f(\Lambda)} .$$

C'est-à-dire que $f(\cdot)$ satisfait une relation de récurrence : $f(\Lambda) = \sum_{\lambda | \lambda \nearrow \Lambda} f(\lambda)$. Cependant, $|\text{ST}(\lambda)|$ satisfait la même relation de récurrence, en conséquence, $|\text{ST}(\lambda)| = f(\lambda)$.

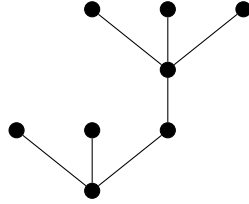
3 Arbres enracinés et étiquetages standards

3.1 Définitions générales

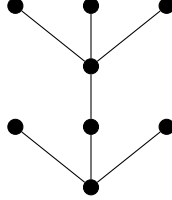
Un *arbre enraciné* t est la donnée d'une racine \bullet , et d'un multi-ensemble (ensemble avec des multiplicités) d'autres arbres enracinés $\{(t_1)^{m_1}, (t_2)^{m_2}, \dots, (t_s)^{m_s}\}$. Ce multi-ensemble de sous-arbres peut être vide. On peut représenter t par un graphe, en reliant la racine aux sous-arbres t_1 apparaissant m_1 fois, t_2 apparaissant m_2 fois, etc. Par exemple,

$$t = (\bullet, \{(\bullet, \{\})^2, (\bullet, \{(\bullet, \{\})^3\})\})$$

peut être représenté par :



On n'accorde pas d'importance à l'ordre des sous-arbres, donc l'arbre t ci-dessus peut aussi être représenté par :



La taille de l'arbre est son nombre d'arêtes, donc un de moins que son nombre de sommets. Si $t = (\bullet, \{(t_1)^{m_1}, (t_2)^{m_2}, \dots, (t_s)^{m_s}\})$, on a donc :

$$|t| = m_1 + m_2 + \dots + m_s + |t_1| + |t_2| + \dots + |t_s| .$$

Notons $\mathfrak{T}(n)$ l'ensemble des arbres enracinés de taille n . La fonction génératrice des cardinaux des ensembles $\mathfrak{T}(n)$ vérifie l'équation :

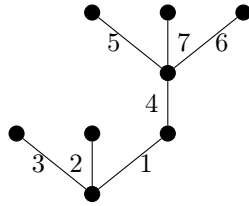
$$T(z) = \sum_{n=0}^{\infty} |\mathfrak{T}(n)| z^n = \prod_{n \geq 0} \frac{1}{(1 - x^{n+1})^{|\mathfrak{T}(n)|}} ,$$

ce qui permet de calculer par récurrence $|\mathfrak{T}(n)|$.

La mesure uniforme sur les arbres de taille n n'est pas forcément la plus naturelle, en particulier si l'on pense à la construction d'un arbre comme un processus récursif de greffage de feuilles. Dans [Ful09], Fulman définit un analogue de la mesure de Plancherel des partitions sur $\mathfrak{T}(n)$. Si $t \in \mathfrak{T}(n)$, un *étiquetage standard* de t est une numérotation des arêtes de l'arbre par les entiers $1, 2, \dots, n$, de sorte que tout chemin allant de la racine vers les feuilles soit strictement croissant. Autrement dit, c'est une bijection croissante de l'ensemble E_t des arêtes de l'arbre t vers $\llbracket 1, n \rrbracket$, l'ordre sur E_t étant

$$(e \prec f) \iff (\text{il existe un chemin dans } t \text{ partant de la racine et empruntant } e \text{ puis } f) .$$

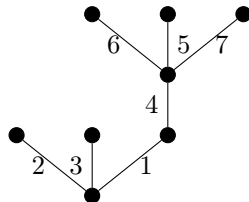
Par exemple,



est un étiquetage standard. On a ainsi défini l'analogue pour les arbres de la notion de tableau standard. Cependant, comme les arbres sont considérés à symétrie près, il y a en fait deux notions pertinentes d'étiquetages standards. Si $t = (V_t, E_t)$ est vu comme un graphe, alors un étiquetage standard est une bijection croissante $L : E_t \rightarrow \llbracket 1, n \rrbracket$, mais on pourrait aussi considérer les étiquetages standards à symétrie près, c'est-à-dire les classes d'équivalences de bijections $L : E_t \rightarrow \llbracket 1, n \rrbracket$ pour la relation :

$$(L_1 \sim L_2) \iff (\text{il existe un isomorphisme de graphes enracinés } \psi : V_t \rightarrow V_t \text{ tel que } L_2(\psi(x), \psi(y)) = L_1(x, y) \text{ pour toute arête } (x, y)) .$$

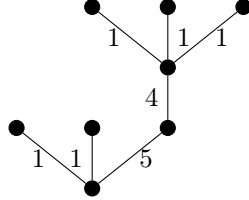
Ainsi, l'étiquetage standard précédent est équivalent à



Notons $\text{SL}(t)$ l'ensemble des étiquetages standards d'un arbre enraciné t , $d(t) = |\text{SL}(t)|$ son cardinal, et $u(t)$ le nombre d'étiquetages standards à symétrie près de t . On a alors $d(t) = |\text{Aut}(t)| u(t)$, avec $\text{Aut}(t)$ groupe des automorphismes de l'arbre enraciné t . Il existe une formule des équerres pour $d(t)$:

$$d(t) = \frac{|t|!}{\prod_{e \in E_t} |t_e|},$$

où t_e désigne le sous-arbre de t situé après l'arête e (cette arête étant incluse). Par exemple, les tailles des sous-arbres de notre exemple sont :



donc $d(t) = 252$ et $u(t) = \frac{252}{12} = 21$.

On peut donner une preuve de cette formule inspirée de celle de [GNW82]; cette preuve est liée à un algorithme permettant de choisir un étiquetage standard uniformément dans $\text{SL}(t)$.

Étant donné un arbre enraciné t , le produit $d(t) u(t)$ correspond aux paires d'étiquetages standards de t à symétrie près. On a alors l'identité fondamentale :

$$\prod_{k=1}^n \binom{k+1}{2} = \sum_{t \in \mathfrak{T}(n)} d(t) u(t).$$

La mesure de Plancherel sur les arbres de taille n est donc définie par :

$$\mathbb{P}_{\text{arbres}}^{(n)}[t] = \frac{d(t) u(t)}{\prod_{k=1}^n \binom{k+1}{2}}.$$

Il semble que cette mesure n'apparaisse que dans [Ful09], et aucune étude des arbres aléatoires correspondants n'a jusqu'ici été menée. Il y a plusieurs preuves de l'identité fondamentale pour les arbres; l'une d'entre elle s'appuie sur un algorithme de type RSK qui est décrit dans [KP12].

3.2 Détails sur quelques notions

On présente d'abord l'algorithme de type RSK, qui prouve l'identité fondamentale $\sum_{t \in \mathfrak{T}(n)} d(t) u(t) = \prod_{k=1}^n \binom{k+1}{2}$.

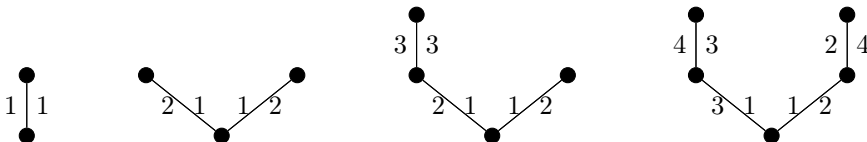
Dans le numérateur de la mesure de Plancherel sur les arbres de taille n , chaque terme $\binom{i+1}{2} = i + \binom{i}{2}$ est le nombre d'ensembles de la forme $\{u, v\}$ avec $0 \leq u \leq v \leq i - 1$:

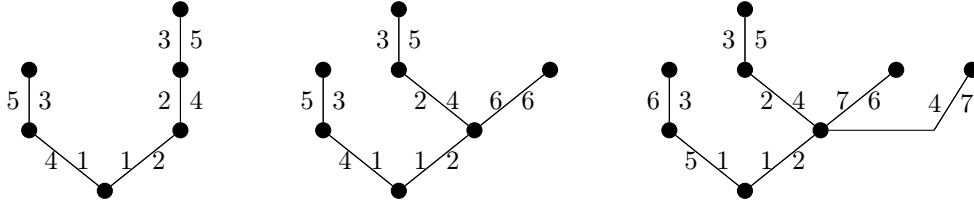
- cas $u = v$: on a i singletons $\{u\}$;
- cas $u < v$: on a $\binom{i}{2}$ paires $\{u, v\}$.

Il doit donc y avoir une bijection entre les arbres de taille n munis d'une paire d'étiquetages standards, et les suites $(\{u_1, v_1\}, \dots, \{u_n, v_n\})$ avec pour chaque indice i les inégalités $0 \leq u_i \leq v_i \leq i - 1$. Expliquons cette bijection de type RSK :

- On construit récursivement l'arbre doublement étiqueté de taille n en lisant la suite $\{u_1, v_1\}, \dots, \{u_n, v_n\}$. Après lecture de $\{u_i, v_i\}$, l'arbre est de taille i .
- À l'étape i , si $u_i = v_i$, alors on greffe au-dessus du nœud qui termine l'arête étiquetée (u_i, p) une nouvelle arête (i, i) . Par convention, il y a une arête virtuelle étiquetée $(0, 0)$ en dessous de la racine.
- Sinon, il y a deux arêtes étiquetées (u_i, p) et (v_i, q) . On remplace toutes les entrées gauches de l'arbre étiqueté $r \geq v_i$ par $r + 1$, et on greffe au-dessus du nœud qui termine l'arête étiquetée (u_i, p) une nouvelle arête (v_i, i) .

Par exemple, supposons que la suite soit $\{0\}, \{0, 1\}, \{2\}, \{1, 2\}, \{2, 3\}, \{1\}, \{1, 4\}$, les arbres doublement étiquetés successivement obtenus sont :





Ensuite, on présente une démonstration de la formule des équerres pour le nombre des étiquetages standards d'un arbre enraciné. La preuve est similaire que la formule des équerres pour des partitions, qui se base sur une marche aléatoires sur le arbre enraciné.

Théorème 3.1. *Pour un arbre enraciné $t \in \mathfrak{T}(n)$, le nombre des étiquetages standards de cet arbre est donné par la formule des équerres :*

$$|\text{SL}(t)| = \frac{n!}{\prod_{e \in E_t} |t_e|} .$$

On dit une arête est une arête au bord si elle est liée à une feuille, qui est un sommet dont le degré est 1. Soit $t \in \mathfrak{T}(n+1)$, on dit qu'un arbre enraciné $t' \in \mathfrak{T}(n)$ satisfait $t' \nearrow t$, si après avoir enlevé une arête au bord de t , on obtient t' .

Étant donné un arbre enraciné t de taille $n+1$, on choisit un arbre enraciné aléatoire de taille n parmi ceux qui satisfont $t' \nearrow t$, en suivant la procédure suivante :

- Commençons par tirer uniformément une arête aléatoire e_1 de l'arbre t . Chaque arête a la probabilité $\frac{1}{n+1}$ d'être tirée.
- Si e_1 est une arête au bord de t , on peut l'enlever pour obtenir un arbre t' tel que $t' \nearrow t$.
- Sinon, la taille du sous-arbre t_{e_1} (l'arête e_1 étant incluse) est supérieure à 2, on peut donc choisir uniformément une nouvelle arête aléatoire e_2 dans sous-arbre t_{e_1} , à part e_1 . Chaque arête dans ce sous-arbre a la probabilité $\frac{1}{|t_{e_1}|-1}$ d'être choisie.
- Si e_2 est au bord de t , on l'enlève et obtient un $t' \in \mathfrak{T}(n)$ tel que $t' \nearrow t$. Sinon, on continue la procédure en choisissant uniformément une arête aléatoire e_3 dans le sous-arbre t_{e_2} , puis une arête aléatoire e_4 dans le sous-arbre t_{e_3} , etc. jusqu'à ce qu'on obtienne une arête à enlever.

À chaque étape, la taille du sous-arbre t_{e_i} est strictement inférieure à celle de $t_{e_{i-1}}$, donc cette procédure terminera presque sûrement.

Lemme 3.2. *Notons $p(t \rightarrow t')$ la probabilité d'obtenir t' en enlevant une arête aléatoire de t en suivant la procédure ci-dessus, et $g(t) = \frac{|t|!}{\prod_{e \in E_t} |t_e|}$ le terme de droite de la formule des équerres, alors on a :*

$$p(t \rightarrow t') = \frac{g(t')}{g(t)} .$$

Démonstration. Notons e_* l'arête de t qui est enlevée pour obtenir t' . On a :

$$\frac{g(t')}{g(t)} = \frac{1}{n+1} \frac{\prod_{e \in E_t} |t_e|}{\prod_{e \in E_{t'}} |t'_e|} ,$$

les arêtes après lesquelles les tailles du sous-arbre sont différentes dans t et t' sont celles dans le chemin qui lie e_* avec la racine. Notons C_{e_*} l'ensemble des arêtes qui sont dans le chemin liant e_* et la racine : $C_{e_*} = \{e_1, e_2, \dots, e_{r-1}, e_r = e_*\}$. Donc,

$$\begin{aligned} \frac{g(t')}{g(t)} &= \frac{1}{n+1} \prod_{i=1}^{r-1} \frac{|t_{e_i}|}{|t_{e_i}|-1} \\ &= \frac{1}{n+1} \prod_{i=1}^{r-1} \left(1 + \frac{1}{|t_{e_i}|-1} \right) \\ &= \frac{1}{n+1} \sum_{I \subset \llbracket 1, r-1 \rrbracket} \frac{1}{\prod_{i \in I} (|t_{e_i}|-1)} . \end{aligned}$$

Ensuite, on montre que l'on peut obtenir la même formule en suivant la procédure aléatoire de la suppression d'une arête au bord. Soit $I \subset \llbracket 1, r-1 \rrbracket$ un sous-ensemble, donc dans la procédure aléatoire, un chemin possible de choisir une arête est donné par $C_I = \{e_i \mid i \in I\} \sqcup \{e_r\}$.

Notons $p(t \rightarrow t', I)$ la probabilité d'obtenir $t' = t \setminus \{e_*\}$ quand la procédure suit C_I . On montrera que

$$p(t \rightarrow t', I) = \frac{1}{n+1} \frac{1}{\prod_{i \in I} (|t_{e_i}| - 1)},$$

ce qui termine la preuve, en prenant la somme des I des sous-ensembles possibles de $\llbracket 1, r-1 \rrbracket$.

Cependant, la preuve de cette identité est directe en suivant la procédure, car $\frac{1}{n+1}$ vient de la probabilité du premier choix de l'arête, et $\frac{1}{|t_{e_i}| - 1}$ correspond à la probabilité d'un choix uniforme de l'arête dans un sous-arbre. \square

Maintenant, on peut donner une démonstration de la formule des équerres. Comme $p(t \rightarrow \cdot)$ est une probabilité sur l'ensemble des arbres enracinés t' tel que $t' \nearrow t$, on a

$$1 = \sum_{t' | t' \nearrow t} \frac{g(t')}{g(t)}.$$

C'est-à-dire que $g(\cdot)$ satisfait une relation de récurrence : $g(t) = \sum_{t' | t' \nearrow t} g(t')$. Cependant, $|\text{SL}(t)|$ satisfait la même relation de récurrence, en conséquence, $|\text{SL}(t)| = g(t)$.

4 Questions et outils

La similarité évidente entre les deux mesures de Plancherel (sur les partitions et sur les arbres) s'étend aux questions que l'on peut se poser sur ces objets, et aux outils que l'on peut employer pour les aborder. Dans ce qui suit, nous listons les principales questions que l'on peut se poser sur les arbres $t^{(n)} \sim \mathbb{P}_{\text{arbres}}^{(n)}$; les questions analogues pour les partitions sous leurs mesures de Plancherel ont toutes déjà été résolues.

- **problème d'Ulam.** Inspiré par le problème d'Ulam qui étudie la longueur d'un plus long sous-mot croissant d'une permutation aléatoire, on peut considérer :

$$\ell_{\text{arbres}}^{(n)} = \ell(t^{(n)}) = \text{profondeur maximale d'une arête de l'arbre } t^{(n)}.$$

Dans l'exemple vu à la section précédente, $n = 7$ et $\ell(t^{(7)}) = 3$. On souhaite alors trouver une fonction $f(n)$ telle que

$$\frac{\ell_{\text{arbres}}^{(n)}}{f(n)} \xrightarrow{\mathbb{P}} 1.$$

Après quelques réflexions, notre conjecture est que $f(n) = c \ln(n)$, où c est une constante. Une première étape consisterait à évaluer $\mathbb{E}[\ell_{\text{arbres}}^{(n)} - \ell_{\text{arbres}}^{(n-1)}]$ et $\mathbb{E}[\ell_{\text{arbres}}^{(n)}]$ grâce à l'algorithme de type RSK de Kuba–Panholzer. Pour le problème analogue avec les partitions, l'algorithme RSK montre que $\mathbb{E}[\ell^{(n)} - \ell^{(n-1)}] \leq \frac{1}{\sqrt{n}}$, ce qui implique $\mathbb{E}[\ell^{(n)}] \leq 2\sqrt{n}$.

- **limite globale.** Plus généralement, on souhaiterait comprendre la géométrie globale des arbres aléatoires $t^{(n)}$. Le problème analogue pour des arbres sous la mesure uniforme sur $\mathfrak{T}(n)$ a été résolu au début des années 2010; voir par exemple [HM12]. On peut associer à tout arbre enraciné de taille n un quadruplet (X, d, μ, x) formé :
 - d'un espace métrique compact (X, d) : en l'occurrence, $X = \llbracket 0, n \rrbracket$ est l'ensemble des sommets de t , et d est la distance de graphe sur X .
 - d'un point distingué $x \in X$: la racine de l'arbre (on peut choisir l'indexation pour que ce soit 0).
 - d'une mesure de probabilité sur X : la mesure uniforme.

Le quadruplet (X, d, μ, x) vit dans l'espace \mathcal{M} des (classes d'isomorphismes d')espaces métriques compacts pointés mesurés, muni de la distance de Gromov–Hausdorff–Prohorov. Considérons alors l'espace renormalisé

$$\left(\llbracket 0, n \rrbracket, \frac{d_{t^{(n)}}}{\ell_{\text{arbres}}^{(n)}}, \mu_{t^{(n)}}, 0 \right)$$

avec $t^{(n)} \sim \mathbb{P}_{\text{arbres}}^{(n)}$. Y a-t-il une limite en loi dans \mathcal{M} à cet espace métrique mesuré pointé ? Dans le cas des arbres uniformes, la limite est l'arbre continu d'Aldous, mais on pense que le résultat est différent pour la mesure de Plancherel.

- **profil et observables.** Une autre façon d’aborder la géométrie de l’arbre est de considérer le *profil*

$$p^{(n)} = (\text{nombre d’arêtes de profondeur 1 dans } t^{(n)}, \text{ nombre d’arêtes de profondeur 2, } \dots) .$$

Les mesures de Plancherel sur les arbres sont probablement les marginales de processus markoviens sur l’analogie $\mathfrak{T} = \bigsqcup_{n \in \mathbb{N}} \mathfrak{T}(n)$ du graphe de Young $\mathfrak{Y} = \bigsqcup_{n \in \mathbb{N}} \mathfrak{Y}(n)$. Dans le contexte des partitions aléatoires, des observables $f(\lambda)$ des partitions appliquées à ces processus donnent des variables aléatoires qui capturent la géométrie des diagrammes de Young, et qui permettent d’en déterminer l’asymptotique (par des méthodes de moments ou de martingales). Pour les arbres, la construction d’une algèbre d’observables capturant le profil et compatible avec les mesures de Plancherel est un problème combinatoire très intéressant ; cette construction est probablement reliée à l’algèbre des arbres de Connes–Kreimer (voir par exemple [Kre99]).

- **mesures centrales.** Les mesures de Plancherel sur les partitions dérivent d’une *mesure centrale* sur les tableaux standards infinis : ce sont les mesures de probabilité sur les suites croissantes de partitions $(\emptyset \nearrow \lambda^{(1)} \nearrow \lambda^{(2)} \nearrow \dots \nearrow \lambda^{(n)} \nearrow \dots)$ telles que conditionnellement à $\lambda^{(n)} = \lambda$, le début de la suite $(\emptyset \nearrow \lambda^{(1)} \nearrow \lambda^{(2)} \nearrow \dots \nearrow \lambda^{(n)})$ est distribuée comme un tableau standard uniforme dans $\text{ST}(\lambda)$. Une classification de toutes les mesures centrales sur les tableaux standards infinis a été obtenue par Kerov et Vershik dans [VK81] ; elles sont liées aux spécialisations Schur-positives de l’algèbre des fonctions symétriques. On peut de même définir les mesures centrales sur les arbres standards infinis : ce sont les mesures de probabilité sur les suites croissantes d’arbres $(\bullet \nearrow t^{(1)} \nearrow t^{(1)} \nearrow \dots \nearrow t^{(n)} \nearrow \dots)$ telles que conditionnellement à $t^{(n)} = t$, le début de la suite $(\bullet \nearrow t^{(1)} \nearrow t^{(2)} \nearrow \dots \nearrow t^{(n)})$ est distribué comme un étiquetage standard uniforme dans $\text{SL}(t)$. On voit facilement que les mesures de Plancherel sur les arbres sont les marginales d’une certaine mesure centrale sur les arbres standards infinis. Peut-on plus généralement décrire toutes les mesures centrales sur les arbres standards infinis ? Y a-t-il un lien avec les spécialisations de l’algèbre de Connes–Kreimer ?

Références

- [BOO00] A. Borodin, A. Okounkov, and G. Olshanski. Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.*, 13 :491–515, 2000.
- [Ful09] J. Fulman. A sharp analysis of the mixing time for random walk on rooted trees. *Electronic Journal of Combinatorics*, 16(1) :R139, 2009.
- [GNW82] C. Greene, A. Nijenhuis, and H. S. Wilf. A probabilistic proof of a formula for the number of Young tableaux of a given shape. In *Young tableaux in combinatorics, invariant theory, and algebra*, pages 17–22. Elsevier, 1982.
- [HM12] B. Haas and G. Miermont. Scaling limits of Markov branching trees with applications to Galton–Watson and random unordered trees. 40(6) :2589–2666, 2012.
- [Joh01] Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. Math.*, 153 :259–296, 2001.
- [Knu70] D. Knuth. Permutations, matrices, and generalized Young tableaux. *Pacific Journal of Mathematics*, 34(3) :709–727, 1970.
- [KP12] M. Kuba and A. Panholzer. Bilabelled increasing trees and hook-length formulae. *European Journal of Combinatorics*, 33(2) :248–258, 2012.
- [Kre99] D. Kreimer. Chen’s iterated integral represents the operator product expansion. *Advances in Theoretical and Mathematical Physics*, 3 :627–670, 1999.
- [KV77] S. V. Kerov and A. M. Vershik. Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux. *Soviet Mathematics Doklady*, 18 :527–531, 1977.
- [LS77] B. F. Logan and L. A. Shepp. A variational problem for random Young tableaux. *Adv. Math.*, 26 :206–222, 1977.
- [Oko00] A. Okounkov. Random matrices and random permutations. *Intern. Math. Res. Notices*, 2000(20) :1043–1095, 2000.
- [Rob38] G. de B. Robinson. On the representations of the symmetric group. *American Journal of Mathematics*, pages 745–760, 1938.
- [Sch61] C. Schensted. Longest increasing and decreasing subsequences. *Canadian Journal of mathematics*, 13 :179–191, 1961.
- [TW94] Craig A Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159 :151–174, 1994.
- [VK81] A. M. Vershik and S. V. Kerov. Asymptotic theory of characters of the symmetric group. *Functional analysis and its applications*, 15(4) :246–255, 1981.