Séminaire Lotharingien de Combinatoire XX (2020)
Article #YY, 8 pp.

Proceedings of the 32nd Conference on Formal Power Series and Algebraic Combinatorics (Ramat Gan)

Hydrodynamic limit of Robinson–Schensted–Knuth algorithm

Mikołaj Marciniak*[1]

[1] *Academia Copernicana, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University ul. Chopina 12/18, 87-100 Toruń*

Abstract. We investigate evolution in time of the position of a fixed number in the insertion tableau when the Robinson–Schensted–Knuth algorithm is applied to a sequence of random numbers. When the length of the sequence tends to infinity, a typical trajectory after scaling converges in probability to some deterministic curve.

Keywords: Young tableau, RSK algorithm, bumping route, random Young tableaux, limit shape

1 Introduction

A partition of a natural number n is a break up of n into the sum $n = \lambda_1 + \lambda_2 + \cdots + \lambda_k$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ are positive integer numbers. The vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is usually used to denote a partition. Let $\lambda \vdash n$ denote that λ is a partition of a number n. A Young diagram $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a finite collection of boxes arranged in left-justified row with the row lengths λ_j in the $j-th$ row. Thus the Young diagram λ is a graphical interpretation of the partition λ. A Young tableau is a Young diagram filled with numbers. If the entries strictly decrease along each column from top to bottom and weakly increase along each row from left to right, a tableau is called semistandard. A standard Young tableau is a semistandard Young tableau with n boxes which contains only all numbers $1, 2, \ldots, n$.

Let S_n denote the permutation group of order n. The Robinson–Schensted–Knuth correspondence RSK is a bijective correspondence that assigns to any permutation $\sigma \in S_n$ a pair of standard Young tableaux (P, Q) with the same shape $\lambda \vdash n$. The semistandard tableau P is called an insertion tableau, and the standard tableau Q is called a recording tableau. A detailed description of the RSK algorithm can be found in **[?]**. The RSK algorithm is based on applying the insertion step to successive numbers from a given finite sequence $\{X_j\}_{j=1}^{\infty}$. The insertion step takes as input the previously obtained tableau

*mikolaj@mat.umk.pl. Mikołaj Marciniak was partially supported by Narodowe Centrum Nauki, grant-number 2017/26/A/ST1/00189.

$P(X_1, X_2, \ldots, X_{n-1})$ and the next number X_n from the sequence. Next it produces as output a new tableau $P(X_1, X_2, \ldots, X_n)$ with shape "increased" by one box. We start in the first row with the number $x = X_n$. The insertion step consists of inserting the number x into the first box containing a number y greater than x. Move to the next row with the number $x = y$ and repeat the action. At some row we are forced to insert the number at the end of the row, which will end the insertion step. A collection of rearranged boxes is called the bumping route.

We can say that the boxes with numbers are moved along bumping route using an RSK insertion step. The insertion step of the RSK algorithm can change the position of some numbers. We will consider the problem of the position of the box with the selected number in time.

2 Motivations

The RSK algorithm is an important tool for algebraic combinatorics, especially in the context of Littlewood–Richardson coefficients and plactic monoid. Crucial to understanding the Littlewood–Richardson coefficients is the RSK algorithm and operations on the words. For many years mathematicians have been studying the asymptotic behavior of the insertion tableau when we apply the Robinson–Schensted–Knuth algorithm to random input. The sequence of independent and identically distributed random variables with uniform distribution on interval $(0, 1)$ applied to the RSK algorithm generates the Plancherel measure on Young diagrams. The Plancherel measure is an important element of representation theory because describes how a left regular representation breaks up into irreducible representations.

The Ulam–Hammersley problem concerning typical length of the longest increasing subsequence in random permutations is corresponds to the length of the first row in the Young tableau obtained by the RSK algorithm from the sequence of independent random variables $\{X_j\}_{j=1}^{\infty}$ with uniform distribution $U(0, 1)$ on interval $(0, 1)$.

Logan and Shepp [] and Vershik and Kerov [] described the limit shape of the insertion tableu $P(X_1, X_2, \ldots, X_n)$ obtained when we apply the RSK algorithm to a random finite sequence.

Romik and Śniady in [**?**] considered the limit shape of bumping routes obtained from the insertion tableau $P(X_1, X_2, \ldots, X_n, w)$, when applying an RSK insertion step with a fixed number to an existing insertion tableau obtained from a random finite sequence. In [**?**] they considered the limit shape of jeu de taquin obtained from the recording tableau $Q(w, X_1, X_2, \ldots, X_n)$ made from a random finite sequence preceded by a fixed number.

What can we say about the evolution over time of the insertion tableau from the point of view of box dynamics, when we apply the RSK algorithm to a sequence of independent random variables with uniform distribution $U(0,1)$? How do the boxes move in the insertion tableau? If investigate the position of a box with a fixed number after scaling, will we get a deterministic limit, when the number of boxes tends to infinity?

More specifically we consider the insertion tableau $P(X_1, X_2, \ldots, X_n, w, X_{n+1} \ldots, X_m)$ obtained by the RSK algorithm applied to a random finite sequence containing a fixed number w at some index. The box with this fixed number w is being bumped by the RSK insertion step along the bumping routes.

3 The main result

The main result of the article is Theorem 1 describing the asymptotic behavior of the box with a fixed number. When the number of boxes tends to infinity then the trajectory of the box converges in probability to the function $G(A) = [willbedefinition]$.

More specifically let $w \in [0,1]$ be a fixed number. Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of independent random variables with uniform distribution $U(0,1)$. For every $n \in \mathbb{N}$ let function $\mathrm{Pos}_n : \{n+1, n+2, \ldots\} \to \mathbb{N}^2$ describe the coordinates of the box with number w in the insertion tableau:

$$\mathrm{Pos}_n\left(j\right) = \mathrm{box}_w\Big(P\left(X_1, \ldots, X_n, w, X_{n+1}, \ldots, X_j\right)\Big)$$

for $j \in \{n+1, n+2, \ldots\}$.

Theorem 1. *The random variable* $\mathrm{Pos}_n(m)$*, after scaling by* $\sqrt{Awn}$ *converges in probability to the limit* $G(A)$*, when* n *tends to infinity. Moreover the convergence is uniform:*

$$\forall_{\epsilon>0} \lim_{n\to\infty} \mathbb{P}\left(\sup_{A\in(1,C]} \left\| \frac{\mathrm{Pos}_n\left(\lfloor An \rfloor\right)}{\sqrt{wn}} - \sqrt{A}G\left(A\right)\right\| < \epsilon\right) = 1$$

Proof. We apply the RSK algorithm to a random sequence of real numbers containing the number w and investigate the position of the box with the number w in the insertion tableau. An insertion step applied to a number greater that w does not change the position of the number w, so it is enough to consider only the subsequence containing numbers no greater than w.

Now we're gonna use this observation in the proof. Let $m = \lfloor An \rfloor$. The probability that the same number will occur twice in the sequence $(w, X_1, X_2, \dots)$ is equal to 0, hence without losing generality we assume that the numbers $w, X_1, X_2, \dots$ are all different. Let the sequence $\{X'_j\}_{j=1}^{\infty}$ be the subsequence of the sequence $\{X_j\}_{j=1}^{\infty}$ containing all elements of the sequence $\{X_j\}_{j=1}^{\infty}$, which are less than w. The sequence $\{X'_j\}_{j=1}^{\infty}$ is a sequence of independent random variables with uniform distribution $U(0,w)$ random variables. Let $n' = n'(n)$ and $m' = m'(m)$ denote the numbers of elements of the sequences $\{X_j\}_{j=1}^{n}$, $\{X_j\}_{j=1}^{m}$ which are smaller than w. Then there is an equality:

$$\begin{aligned} \mathrm{Pos}_n\left(\lfloor An \rfloor\right) &= \mathrm{box}_w\left(P\left(X_1, \dots, X_n, w, X_{n+1}, \dots, X_{m_n}\right)\right) \\ &= \mathrm{box}_w\left(P\left(X_1, \dots, X'_{n'}, w, X'_{n'+1}, \dots, X'_{m'}\right)\right). \end{aligned}$$

The random variable n' counts how many numbers from the sequence $\{X_j\}_{j=1}^{n}$ are less than w, so n' is a random variable with the binomial distribution with parameters n and w: $n' \sim B(n,w)$. Likewise the random variable $m' - n'$ counts how many numbers from the sequence $\{X_j\}_{j=n'+1}^{m'}$ are less than w, so $m' - n' \sim B(m-n, w)$. Moreover, the random variables n' and $m' - n'$ are independent, because the random variables $X_1, X_2, \dots$ are independent.

From the Khinchin–Kolmogorov–Etemadi Strong Law of Large Numbers theorem, we know that if n tends to infinity, then with probability 1 exist the limits:

$$\begin{cases} \lim_{n\to\infty} \frac{n'}{n} &= \mathbb{E}(X_1) = & w, \\ \lim_{n\to\infty} \frac{m'-n'}{m-n} &= \mathbb{E}(X_1) = & w. \end{cases}$$

Therefore, with probability 1 also exist the limits:

$$\begin{aligned} \lim_{n\to\infty} \frac{m'}{n'} &= \lim_{n\to\infty} 1 + \frac{m'-n'}{n'} \\ &= 1 + \lim_{n\to\infty} \frac{m'-n'}{m-n} \frac{1}{\frac{n'}{n}} \frac{m-n}{n} \\ &= 1 + \lim_{n\to\infty} \frac{m'-n'}{m-n} \frac{1}{\frac{n'}{n}} \left(\frac{m}{n} - 1\right) \\ &= 1 + \frac{w}{w}(A-1) \\ &= A. \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty} \frac{m'}{n} &= \lim_{n\to\infty} \frac{n'}{n} \frac{m'}{n'} \\ &= wA. \end{aligned}$$

We define the function $z : \{1, 2, \dots m'\} \cup \{m'+1, n'+\frac{1}{2}\} \to [0,1]$ that assigns to the number $t \in \{1, 2, \dots, m'\}$, the $t-th$ largest number among $X_1', X_2', \dots, X_{m'}'$ and additionally $z(m'+1) = w$ and $z\left(n'+\frac{1}{2}\right) = \frac{z(n')+z(n'+1)}{2}$.

The sequence $\{z(j)\}_{j=1}^{m'}$ contains all elements of the sequence $\{X_j'\}_{j=1}^{m'}$ in the ascending order. In addition the sequence $\{X_j'\}_{j=1}^{m'}$ is a random permutation Π with uniform distribution of the sequence $\{z(j)\}_{j=1}^{m'}$. The sequence $\{z(j)\}_{j=1}^{m'}$ will be called a random increasing sequence with the uniform distribution on the interval $[0,1]$. Let $\Pi = (\Pi_1, \Pi_2, \dots, \Pi_{m'})$. Then

$$\{X_j'\}_{j=1}^{m'} = \{z(\Pi_j)\}_{j=1}^{m'} = z \circ \Pi,$$

where z is the function that acts separately on every element of the permutation Π.

Similarly the function z acts on a Young tableau by acting on each box individually. Then

$$\begin{aligned}
\mathrm{Pos}_n(m) &= \mathrm{box}_w\left(P\left(X_1', \dots, X_{n'}', w, X_{n'+1}', \dots, X_{m'}'\right)\right) \\
&= \mathrm{box}_w\left(P\left(z(\Pi_1), \dots, z\left(\Pi_{n'}\right), w, z\left(\Pi_{n'+1}\right), \dots, z\left(\Pi_{m'}\right)\right)\right) \\
&= \mathrm{box}_w\left(z \circ P\left(\Pi_1, \dots, \Pi_{n'}, m'+1, \Pi_{n'+1}, \dots, \Pi_{m'}\right)\right) \\
&= \mathrm{box}_{m'+1}\left(P\left(\Pi_1, \dots, \Pi_{n'}, m'+1, \Pi_{n'+1}, \dots, \Pi_{m'}\right)\right).
\end{aligned}$$

We define the permutation $\Pi\uparrow = \left(\Pi_1, \dots, \Pi_{n'}, m'+1, \Pi_{n'+1}, \dots, \Pi_{m'}\right)$ as a natural extension of the permutation Π. In addition, we will use the fact that for any permutation $\Pi\uparrow$ the insertion tableau of $\Pi\uparrow$ is equal to the recording tableau of the inverse permutation $\Pi\uparrow^{-1}$:

$$P(\Pi\uparrow) = Q\left(\Pi\uparrow^{-1}\right).$$

Therefore

$$\begin{aligned}
\mathrm{Pos}_n(m) &= \mathrm{box}_{m'+1}\left(P\left(\Pi_1, \dots, \Pi_{n'}, m'+1, \Pi_{n'+1}, \dots, \Pi_{m'}\right)\right) \\
&= \mathrm{box}_{m'+1}\left(Q\left(\Pi\uparrow^{-1}\right)\right) \\
&= \mathrm{box}_{m'+1}\left(Q\left(\Pi_1^{-1}\uparrow, \dots, \Pi_{m'}^{-1}\uparrow, n'+1\right)\right).
\end{aligned}$$

Now, using the function z, we will get the sequence of independent random variables with uniform distribution $U(0,1)$

$$\begin{aligned}
\mathrm{Pos}_n(m) &= \mathrm{box}_{m'+1}\left(Q\left(\Pi_1^{-1}\uparrow,\dots,\Pi_{m'}^{-1}\uparrow,n'+1\right)\right)\\
&= \mathrm{box}_{m'+1}\left(Q\left(\Pi_1^{-1}\uparrow,\dots,\Pi_{m'}^{-1}\uparrow,n'+\frac{1}{2}\right)\right)\\
&= \mathrm{box}_{m'+1}\left(Q\left(\Pi_1^{-1},\dots,\Pi_{m'}^{-1},n'+\frac{1}{2}\right)\right)\\
&= \mathrm{box}_{m'+1}\left(z\circ Q\left(\Pi_1^{-1},\dots,\Pi_{m'}^{-1},n'+\frac{1}{2}\right)\right)\\
&= \mathrm{box}_{m'+1}\left(Q\left(z\left(\Pi_1^{-1}\right),\dots,z\left(\Pi_{m'}^{-1}\right),z\left(n'+\frac{1}{2}\right)\right)\right).
\end{aligned}$$

The permutation Π is a random permutation with the uniform distribution, so Π^{-1} is also a random permutation with the uniform distribution. If we act with random permutation on a random increasing sequence with the uniform distribution we will get the sequence of independent random variables with uniform distribution, then the sequence $z\circ\Pi^{-1} = \left(z\left(\Pi_1^{-1}\right),\dots,z\left(\Pi_{m'}^{-1}\right)\right)$ is a sequence of independent random variables with uniform distribution $U(0,w)$. We define the random variable T_n and the sequence $\{Y_j\}_{j=1}^{m'}$:

$$\begin{cases} Y_j = \frac{z\left(\Pi_j^{-1}\right)}{w} & \text{for} \quad j \in 1,2,\dots,m' \\ T_n = \frac{z\left(n'+\frac{1}{2}\right)}{w} \end{cases}$$

The sequence $\{Y_j\}_{j=1}^{m'}$ is a sequence of independent random variables with uniform distribution $U(0,1)$. In addition the random variable T_n converges with probability 1 to $\frac{1}{A}$:

$$\begin{aligned}
\lim_{n\to\infty} T_n &= \lim_{n\to\infty}\frac{z\left(n'+\frac{1}{2}\right)}{w}\\
&= \lim_{n\to\infty}\frac{1}{w}\frac{z(n')+z(n'+1)}{2}\\
&= \lim_{n\to\infty}\frac{z(n')}{w} + \lim_{n\to\infty}\frac{z(n'+1)-z(n')}{2w}\\
&= \lim_{n\to\infty}\frac{z(n')}{w}\\
&= \lim_{n\to\infty}\frac{n'}{m'}\\
&= \frac{1}{A}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{Pos}_n(m) &= \mathrm{box}_{m'+1}\left(Q\left(z\left(\Pi_1^{-1}\right),\dots,z\left(\Pi_{m'}^{-1}\right),z\left(n'+\frac{1}{2}\right)\right)\right)\\
&= \mathrm{box}_{m'+1}\left(Q\left(\frac{z\left(\Pi_1^{-1}\right)}{w},\dots,\frac{z\left(\Pi_{m'}^{-1}\right)}{w},\frac{z\left(n'+\frac{1}{2}\right)}{w}\right)\right)\\
&= \mathrm{box}_{m'+1}\left(Q\left(Y_1,\dots,Y_{m'},T_n\right)\right).\\
&= \square_{m'}(T_n),
\end{aligned}$$

where with probability one $T_n \to T = \frac{1}{A} \in (0,1)$, when n tends to infinity.
From Romik and Śniady for each $x \in (0,1)$ we have:

$$\frac{\square_{m'}(x)}{\sqrt{m'}} \xrightarrow{p} F(x)$$

Let $G_{m'}(x) = \frac{\square_{m'}(x)}{\sqrt{Anw}}$. Therefore

$$G_{m'}(x) = \sqrt{\frac{m'}{n}\frac{1}{Aw}}\ \frac{\square_{m'}(x)}{\sqrt{m'}} \xrightarrow{p} \sqrt{\frac{Aw}{Aw}}\ G(x) = G(x)$$

We know that if n tends to infinity then with probability 1 for each x occurs:

$$\begin{cases} G_{m'}(x) \overset{p}{\rightrightarrows} G(x) \\ T_n \to T \end{cases}$$

Then with probability 1 for each $\epsilon > 0$ exists n_0 that for each $n > n_0$ and for each $x \in (\frac{1}{C}, 1)$ the following inequalities occur:

$$\begin{cases} G(x) + \epsilon(1,-1) < G_{m'}(x) < G(x) + \epsilon(1,-1) \\ T - \epsilon < T_n < T + \epsilon \end{cases}$$

In addition the function $G(x)$ is monotonic and Lipschitz function with constant 1. Then:

$$G_{m'}(T_n) < G_{m'}(T+\epsilon) < G_{m'}(T) + \epsilon(1,-1) < G(T) + 2\epsilon(1,-1)$$

and

$$G_{m'}(T_n) > G_{m'}(T-\epsilon) > G_{m'}(T) - \epsilon(1,-1) > G(T) - 2\epsilon(1,-1)$$

Thereby:

$$\|G_{m'}(T_n) - G(A)\| < 2\sqrt{2}\epsilon < 4\epsilon$$

. Then with probability 1 for each $\epsilon > 0$ exists n_0 that for each $n > n_0$ and for each $x \in (\frac{1}{C}, 1)$ the following inequality is true:

$$\sup_{A\in(1,C)} \left\| \frac{\mathrm{Pos}_n\left(\lfloor An \rfloor\right)}{\sqrt{Awn}} - G\left(A\right) \right\| < 4\frac{\epsilon}{4\sqrt{A}}$$

Therefore:

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{A\in(1,C)} \left\| \frac{\mathrm{Pos}_n\left(\lfloor An \rfloor\right)}{\sqrt{wn}} - \sqrt{A}G\left(A\right) \right\| < \epsilon \right) = 1$$

□

References

[1] Dan Romik and Piotr Śniady. Jeu de taquin dynamics on infinite Young tableaux and second class particles. *Ann. Probab.*, 43(2):682–737, 2015.

[2] Dan Romik and Piotr Śniady. Limit shapes of bumping routes in the Robinson-Schensted correspondence. *Random Structures Algorithms*, 48(1):171–182, 2016.