

Hydrodynamic limit of Robinson–Schensted–Knuth algorithm

Mikołaj Marciniak^{*1}

¹*Nicolaus Copernicus University in Torun, Academia Copernicana*

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, algorithm RSK, position, bumping route, trajectory, convergence

1 Introduction

A partition of a natural number n is a breaks up of the number n into the sum $n = \lambda_1 + \lambda_2 + \dots + \lambda_k$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$. The vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ is usually used to denote a partition. Let $\lambda \vdash n$ denotes that λ is the partition of number n . The Young diagram $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ is a finite collection of boxes arranged in left-justified rows 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 contains only the numbers $1, 2, \dots, n$.

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 $\lambda \vdash n$. P is called an insertion tableau, and Q is called a recording tableau. A detailed description of the RSK algorithm can be found in [?]. The RSK algorithm is based on applying of the insertion step to successive numbers from a given finite sequence. The insertion step takes as input the previously obtained tableau T and next number x from the sequence. Next produces as output a new tableau increased by one box. The insertion step consists in

^{*}mikolaj@mat.umk.pl. Mikołaj Marciniak was partially supported by Polish Ministry of Higher Education research Grant NCN Maestro nr 2017/26/A/ST1/00189.

inserting the number x into first box containing number y greater than x . Move to the next row with the number y and repeat the action. At some row we are forced to insert the number at the end of the row. 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. Perhaps it is worth considering the position of the box with the selected number in time.

2 Motivations

Mathematicians using representation theory tried to prove that Littlewood–Richardson coefficients are non-negative integers. Finally, after 40 years and a lot of incorrect proofs, the hypothesis was proved. Crucial to understanding the Littlewood–Richardson coefficients is the RSK algorithm and operations on the words. The RSK algorithm is also related to the plactic monoid.

The RSK algorithm is usually used for the sequence of independent and identically distributed (i.i.d.) random variables with uniform distribution on interval $(0, 1)$, because such an input 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 described by the length of the first row in the Young Tableau obtained by the RSK algorithm from the sequence of i.i.d $U(0, 1)$ random variables $\{X_j\}_{j=1}^{\infty}$. The solution of the Ulam–Hammersley problem is given by the Tracy–Widom distribution.

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

Romik and Śniady in [?] considered limit shape of bumping routes obtained from insertion tableau $P(X_1, X_2, \dots, X_n, w)$, when applying an RSK insertion step with a fixed number to an existing insertion tableau obtained from the random finite sequence.

In [?] they considered the limit shape of jeu de taquin obtained from recording tableau $Q(w, X_1, X_2, \dots, X_n)$ made from the random finite sequence preceded by a fixed number.

It therefore seems natural to also consider the insertion tableau $P(X_1, X_2, \dots, X_n, w, X_{n+1}, \dots, X_m)$ obtained by using the RSK algorithm applied to a random finite sequence containing a fixed number at some index. The box with fixed number can be sliding by the RSK insertion step along the bumping routes.

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

3 The main result

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

Let $w \in [0, 1]$ be a fixed number. Let $\{X_j\}_{j=1}^\infty$ be a sequence of i.i.d. $U(0, 1)$ random variables. For every $n \in \mathbb{N}$ let function $\text{Pos}_n : \{n+1, n+2, \dots\} \rightarrow \mathbb{N}^2$ describe coordinates of the box with number w in the insertion tableau:

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

for $j \in \{n+1, n+2, \dots\}$. For each number n we define the number $m = m(n)$. We consider a sequence of natural numbers $\{m\}_{n=1}^\infty$ satisfying the condition $\lim_{n \rightarrow \infty} \frac{m}{n} = A \in (1, \infty)$. For example, let m be equal to $\lfloor An \rfloor$.

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

$$\forall \epsilon > 0 \lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{1 \leq A \leq 100} \left\| G(A) - \frac{\text{Pos}_n(m)}{\sqrt{wn}} \right\| > \epsilon \right) = 0$$

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. Numbers greater than w do not affect the position of the box with the number w , so it is enough to consider only the subsequence containing numbers no greater than

w .

Now we will give the formal proof. 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 all numbers w, X_1, X_2, \dots are 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 i.i.d. $U(0, w)$ random variables. Let $n' = n'(n)$ and $m' = m'(n)$ denote the numbers of element of sequences $\{X_j\}_{j=1}^n, \{X_j\}_{j=1}^m$ smaller than w . Then there is an equality:

$$\begin{aligned} \text{Pos}_n(m) &= \text{box}_w(P(X_1, \dots, X_n, w, X_{n+1}, \dots, X_m)) \\ &= \text{box}_w(P(X_1, \dots, X'_{n'}, w, X'_{n'+1}, \dots, X'_{m'})) . \end{aligned}$$

In addition the random variable n' count how many numbers from the sequence $\{X_j\}_{j=1}^n$ are less than w , so n' is the random variable with binomial distribution with parameters n and w : $n' \sim B(n, w)$. Likewise the random variable $m' - n'$ count 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 exists the limits:

$$\begin{cases} \lim_{n \rightarrow \infty} \frac{n'}{n} &= w, \\ \lim_{n \rightarrow \infty} \frac{m' - n'}{m - n} &= w. \end{cases}$$

Therefore, with probability 1 also exist the limit:

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

where the sequence $\{X'_j\}_{j=1}^\infty$ is a sequence of i.i.d. $U(0, z)$.

We define the function $z : \{1, 2, \dots, m'\} \cup \{m', n' + \frac{1}{2}\} \rightarrow [0, 1]$, that assigns to the number t , the t -th largest number among $X'_1, X'_2, \dots, X'_{m'}$, and additionally $z(m' + 1) = w$

$$\text{and } z\left(n' + \frac{1}{2}\right) = \frac{z(n') + z(n'+1)}{2}.$$

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 a 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 act on Young tableau by acting on each box individually. Then

$$\begin{aligned} \text{Pos}_n(m) &= \text{box}_w(P(X'_1, \dots, X'_{n'}, w, X'_{n'+1}, \dots, X'_{m'})) \\ &= \text{box}_w(P(z(\Pi_1), \dots, z(\Pi_{n'}), w, z(\Pi_{n'+1}), \dots, z(\Pi_{m'}))) \\ &= \text{box}_w(z \circ P(\Pi_1, \dots, \Pi_{n'}, m' + 1, \Pi_{n'+1}, \dots, \Pi_{m'})) \\ &= \text{box}_{m'+1}(P(\Pi_1, \dots, \Pi_{n'}, m' + 1, \Pi_{n'+1}, \dots, \Pi_{m'})). \end{aligned}$$

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

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

Therefore

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

Now, using the function z , we will try to get the sequence of i.i.d. $U(0, 1)$

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

Π is the random permutation with uniform distribution, so Π^{-1} is also random permutation with uniform distribution, then the sequence $z \circ \Pi^{-1} = (z(\Pi_1^{-1}), \dots, z(\Pi_{m'}^{-1}))$ is a sequence of i.i.d. $U(0, w)$. We define the random variable S_n and the sequence $\{Y_j\}_{j=1}^{m'}$.

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

The sequence $\{Y_j\}_{j=1}^{m'}$ is a sequence of i.i.d $U(0, 1)$. In additional random variable S_n converges with probability 1 to $\frac{1}{A}$:

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

Therefore

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

□

References

- [1] Dan Romik and Piotr Śniady. Jeu de taquin dynamics on infinite young tableaux and second class particles. 2015.
- [2] Dan Romik and Piotr Śniady. Limit shapes of bumping routes in the robinson-schensted correspondence. *Random Struct. Algorithms*, 48:171–182, 2016.