Dynamic 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, Poland*

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

1 Introduction

1.1 Notations

A *partition* of a natural number n is a break up of n into a 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 rows with the row length λ_j of 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 all numbers $1, 2, \ldots, n$. Figure 1 show examples of Young diagram and standard Young tableau.

The *Robinson–Schensted–Knuth algorithm RSK* is a bijective algorithm which takes a finite sequence of numbers as the input and returns a pair of 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*. In particular the RSK algorithm assigns to any permutation σ a pair of standard Young tableaux (P, Q). A detailed description of the RSK algorithm can be found in [Rom15]. The RSK algorithm is based on
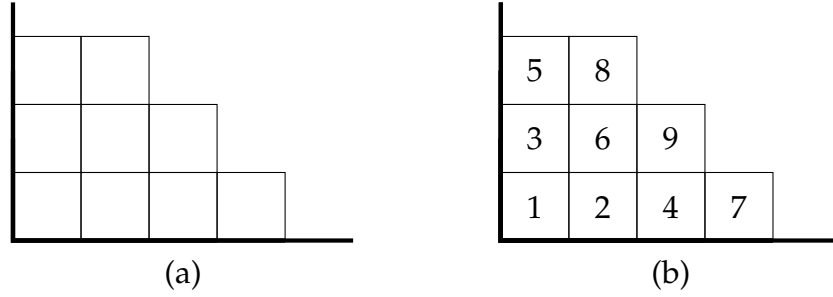
Figure 1: (a) The Young diagram of shape $(4,3,2) \vdash 9$ and (b) a standard Young tableau of shape $(4,3,2) \vdash 9$.

applying the *insertion step* to successive numbers from a given finite sequence $\{X_j\}_{j=1}^n$. The insertion step takes as input the previously obtained tableau $P(X_1, X_2, \ldots, X_{j-1})$ and the next number X_j 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. The RSK-insertion step starts in the first row with the number $x := X_n$. The insertion step consists of inserting the number x into the first box from left 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. The collection of rearranged boxes is called the *bumping route*.

We can say that the boxes with numbers are moved along bumping route during an RSK insertion step. The insertion step of the RSK algorithm can change the position of some numbers. We will consider the position of the box with a selected number as it changes with time.

1.2 Motivations

The RSK algorithm is an important tool for algebraic combinatorics, especially in the context of Littlewood–Richardson coefficients and plactic monoid [FH91]. The RSK algorithm and operations on the words are crucial to understanding the Littlewood–Richardson coefficients.

For many years mathematicians have been studying the asymptotic behavior of the insertion tableau when we apply the Robinson–Schensted–Knuth algorithm to a random input. In the following paragraphs we will see several examples of such considerations. The RSK algorithm applied to the sequence of independent and identically distributed random variables with the uniform distribution on interval $(0,1)$ generates the

Plancherel measure on Young diagrams [RS15]. The Plancherel measure is an important element of representation theory because it describes how the left regular representation decomposes into irreducible representations [FH91].

The Ulam–Hammersley problem [BDJ99] concerning typical length of the longest increasing subsequence in a random permutation corresponds to the problem concerning typical length of the first row in the Young tableau obtained by the RSK algorithm from the sequence of independent random variables $\{X_j\}_{j=1}^n$ with the uniform distribution $U(0,1)$ on the interval $(0,1)$.

Logan and Shepp [LS77] and Vershik and Kerov [KV86] 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 [RS15] 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 w to an existing insertion tableau obtained from a random finite sequence. In [RS15] they considered also 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 w.

1.3 Main problem

This paper also concerns the asymptotic behavior of the insertion tableau when we apply the RSK algorithm to a random input. 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 the uniform distribution $U(0,1)$? How do the boxes move in the insertion tableau? If we investigate the scaled position of a box with a fixed number, 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. We will describe the scaled limit position of the box with the number w depending on the ratio of the numbers m and n.
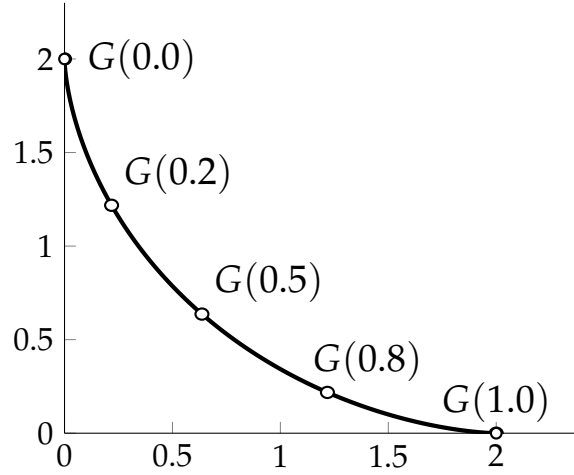
Figure 2: The graph of the function $G(x)$ with specified values for numbers $x = 0.0$, $x = 0.2$, $x = 0.5$, $x = 0.8$ and $x = 1.0$.

1.4   The result of Romik and Śniady

In evidence, we will need the following theorem proved by Romik and Śniady in the theorem.

Theorem 1. *Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of independent random variables with the uniform distribution $U(0,1)$. Let $\square_n(x) \in \mathbb{N}^2$ denote a position of the new box in the in the insertion tableau $P(X_1), \ldots, X_n, x)$ when we apply the RSK insertion step for the number $z \in [0,1]$ to the previous obtained tableau $P(X_1, \ldots, X_n)$. Then this position, after scaling, converges in probability to the some limit curve:*

$$\frac{\square_n(x)}{\sqrt{n}} \xrightarrow{p} G(x),$$

where $G(x) : [0,1] \to [0,2]^2$ is the Logan–Sheep–Vershik–Kerov curve rotated by 45 degrees. The figure 2 shows the graph of the function $G(x)$. Moreover the Logan–Sheep–Vershik–Kerov curve is a Lipschitz function with constant 1. After rotated it means that for each $x \geq 0$ and $\epsilon > 0$ where $x + \epsilon \leq 1$ occurs:

$$G(x) - \epsilon(1, -1) < G(x + \epsilon) < G(x) + \epsilon(1, -1),$$

where the relation $<$ is an order in the plane defined as follows: $(x_1, y_1) < (x_2, y_2)$ only and only if $x_1 < x_2$ and $y_1 > y_2$. The relation $<$ retains the order of the points in plane along the line $y = x$. In addition, the function $G(x)$ is also monotonic with respect to the relation $<$ in plane.
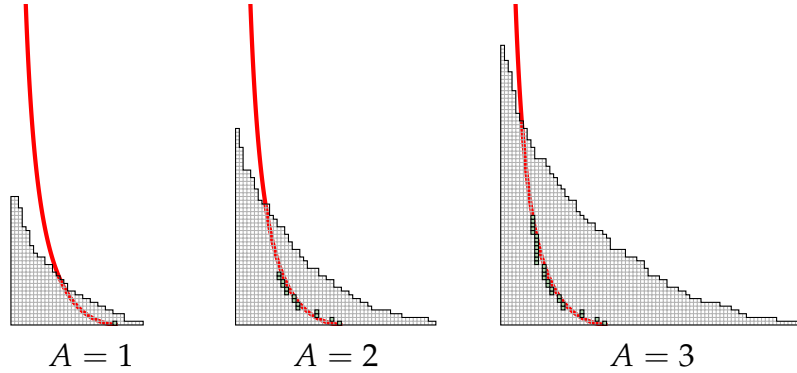
Figure 3: The trajectory describing the position of the box with the fixed number w at the moment $\lfloor An \rfloor$ for $n = 400$ and numbers $A = 1$, $A = 2$ and $A = 3$.

2   The main result

The main result is Theorem 2 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

$$H(A) = \sqrt{A} G(\frac{1}{A}).$$

The curve $H(A)$ is also limit shapes of bumping routes [RS16] in the RSK algorithm. Figure 3 shows the graph of the function $H(x)$ and the experimentally determined position of the box with the number $w = 0.5$ for $n = 400$. 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 the uniform distribution $U(0,1)$. For every $n \in \mathbb{N}$ let function $\text{Pos}_n : \{n+1, n+2, \ldots\} \to \mathbb{N}^2$ describe the coordinates of the box with the number w in the insertion tableau:

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

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

Theorem 2. *For each number $A \in (1, \infty)$ the random variable $\text{Pos}_n(\lfloor An \rfloor)$, after scaling by $\sqrt{wn}$ converges in probability to the limit $H(A)$, when n tends to infinity:*

$$\forall_{\epsilon > 0} \lim_{n \to \infty} \mathbb{P}\left(\left\| \frac{\text{Pos}_n(\lfloor An \rfloor)}{\sqrt{wn}} - H(A) \right\| < \epsilon \right) = 1$$

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 will 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 the uniform distribution $U(0, w)$. Let $n' = n'(n)$ and $m' = m'(m)$ denote the numbers of element respectively of the sequences $\{X_j\}_{j=1}^{n}$, $\{X_j\}_{j=1}^{m}$ which are smaller than w. Then there is an equality:

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

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. We denote it $n' \sim B(n, w)$. 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 the limits exist with probability 1:

$$\lim_{n \to \infty} \frac{n'}{n} = w,$$
$$\lim_{n \to \infty} \frac{m' - n'}{m - n} = w.$$

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 a number $t \in \{1,2,\dots,m'\}$, the $t-th$ smallest 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 the 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

$$
\left\{ X'_j \right\}_{j=1}^{m'} = \left\{ z(\Pi_j) \right\}_{j=1}^{m'} = z \circ \Pi,
$$

where z is the function that acts pointwise on every element of the permutation Π. Similarly, the function z acts on a Young tableau by acting on each box individually. Then

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

We denote $\Pi^{-1} = \left(\Pi_1^{-1}, \ldots, \Pi_{m'}^{-1} \right)$ and we define the permutation

$$\Pi\!\uparrow = \left(\Pi_1, \ldots, \Pi_{n'}, m' + 1, \Pi_{n'+1}, \ldots, \Pi_{m'} \right)$$

as a natural extension of the permutation Π. Then $\Pi\!\uparrow^{-1} = \left(\Pi\!\uparrow_1^{-1}, \ldots, \Pi\!\uparrow_{m'+1}^{-1} \right)$ where

$$\Pi\!\uparrow_j^{-1} = \begin{cases} \Pi_j^{-1} & j \leq n \\ \Pi_j^{-1} + 1 & n < j \leq m' \\ n + 1 & j = m' + 1 \end{cases}$$

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 $\Pi\!\uparrow^{-1}$:

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

Therefore

$$\mathrm{Pos}_n(m) = \mathrm{box}_{m'+1}\left(P\left(\Pi_1, \ldots, \Pi_{n'}, m' + 1, \Pi_{n'+1}, \ldots, \Pi_{m'}\right) \right)$$

$$= \mathrm{box}_{m'+1}\left(Q\left(\Pi\!\uparrow^{-1}\right) \right)$$

$$= \mathrm{box}_{m'+1}\left(Q\left(\Pi\!\uparrow_1^{-1}, \ldots, \Pi\!\uparrow_{m'}^{-1}, n' + 1\right) \right).$$

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

$$\mathrm{Pos}_n(m) = \mathrm{box}_{m'+1}\left(Q\left(\Pi\!\uparrow_1^{-1}, \ldots, \Pi\!\uparrow_{m'}^{-1}, n' + 1\right) \right)$$

$$= \mathrm{box}_{m'+1}\left(Q\left(\Pi\!\uparrow_1^{-1}, \ldots, \Pi\!\uparrow_{m'}^{-1}, n' + \frac{1}{2}\right) \right)$$

$$= \mathrm{box}_{m'+1}\left(Q\left(\Pi_1^{-1}, \ldots, \Pi_{m'}^{-1}, n' + \frac{1}{2}\right) \right)$$

$$= \mathrm{box}_{m'+1}\left(z \circ Q\left(\Pi_1^{-1}, \ldots, \Pi_{m'}^{-1}, n' + \frac{1}{2}\right) \right)$$

$$= \mathrm{box}_{m'+1}\left(Q\left(z\left(\Pi_1^{-1}\right), \ldots, z\left(\Pi_{m'}^{-1}\right), z\left(n' + \frac{1}{2}\right)\right) \right).$$

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 a random

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

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

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

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

Therefore

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

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

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

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

$$T_n \to T$$

and for each x occurs:

$$G_{m'}(x) \xrightarrow{p} G(x).$$

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

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

From Theorem 1 we have the relation $<$ in the plane. 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:

$$\|G_{m'}(T_n) - G(T)\| < 2\sqrt{2}\epsilon < 4\epsilon.$$

Then for each $A \in (1,\infty)$ with probability 1 for each $\epsilon > 0$ exists n_0 that for each $n > n_0$ the following inequality is true:

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

$$\left\|\frac{\mathrm{Pos}_n\left(\lfloor An \rfloor\right)}{\sqrt{wn}} - \sqrt{A}G\left(\frac{1}{A}\right)\right\| < \epsilon.$$

Therefore for each $A \in (1,\infty)$:

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

$\square$

References

[BDJ99] Jinho Baik, Percy Deift, and Kurt Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. **12** (1999), no. 4, 1119–1178. MR 1682248 (2000e:05006)

[FH91] William Fulton and Joe Harris, *Representation theory*, Graduate Texts in Mathematics, vol. 129, Springer-Verlag, New York, 1991, A first course, Readings in Mathematics. MR 1153249

[KV86] Sergei V. Kerov and Anatol M. Vershik, *The characters of the infinite symmetric group and probability properties of the Robinson-Schensted-Knuth algorithm*, SIAM J. Algebraic Discrete Methods **7** (1986), no. 1, 116–124. MR 819713 (87e:22014)

[LS77] B. F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Advances in Math. **26** (1977), no. 2, 206–222. MR 1417317 (98e:05108)

[Rom15] Dan Romik, *The surprising mathematics of longest increasing subsequences*, Institute of Mathematical Statistics Textbooks, vol. 4, Cambridge University Press, New York, 2015. MR 3468738

[RS15] Dan Romik and Piotr Śniady, *Jeu de taquin dynamics on infinite Young tableaux and second class particles*, Ann. Probab. **43** (2015), no. 2, 682–737. MR 3306003

[RS16] _____, *Limit shapes of bumping routes in the Robinson-Schensted correspondence*, Random Structures Algorithms **48** (2016), no. 1, 171–182. MR 3432576