SORTING PROBABILITY OF CATALAN POSETS

SWEE HONG CHAN, IGOR PAK, AND GRETA PANOVA

Abstract. We show that the *sorting probability* of the *Catalan poset* P_n satisfies $\delta(P_n) = O(n^{-5/4})$.

1. Introduction

The *sorting probability* of a poset P, see below, is an interesting measure of independence of linear extensions of P. Originally introduced in connection with sorting under partial information by Kislitsyn (1968) and Fredman (1975), it came to prominence as the subject of the celebrated $\frac{1}{3}$ – $\frac{2}{3}$ *Conjecture*, see [Tro]. The conjecture received further acclaim in 1980s after a remarkable breakthrough by Kahn and Saks [KS], but remains open in full generality. We refer to [CPP, §1.3] for a recent overview of the literature and further references.

In this paper we study the sorting probability $\delta(P_n)$ of a *Catalan poset* P_n on $2n$ elements, which is defined as a product of a chain with 2 elements and with n elements: $P_n := C_2 \times C_n$. The name comes from the fact that the number of linear extensions of P_n is the *Catalan number*:

$$e(P_n) \,=\, \mathrm{Cat}(n) \,:=\, \frac{1}{n+1}\binom{2n}{n}.$$

With over numerous combinatorial interpretations and countless literature, Catalan numbers are extremely well studied, see e.g. [S1, S2]. It is thus remarkable that $\delta(P_n)$ has been out of reach until now.

Formally, for a finite poset $P = (X, \prec)$, let $\mathcal{L}_P$ denote the set of linear extensions of P, and let $e(P) := |\mathcal{L}_P|$. The *sorting probability* $\delta(P)$ is defined as

$$\delta(P) \,:=\, \min_{x,y\in X} \big|\mathbf{P}[L(x) \le L(y)] - \mathbf{P}[L(y) \le L(x)]\big|,$$

where $L \in \mathcal{L}_P$ is a uniform linear extension of P. The $\frac{1}{3}$ – $\frac{2}{3}$ *Conjecture* mentioned above, claims that $\delta(P) \le \frac{1}{3}$ for all finite posets P.

Theorem 1. *For the Catalan poset P_n, we have $\delta(P_n) = O(n^{-5/4})$.*

Until recently, there were very few results in this direction. First, it was shown by Linial, that $\delta(P_n) \le \frac{1}{3}$, and in fact this holds for all posets of width two [Lin]. For indecomposable posets P of width two, this general bound was slightly improved by Sah to $\delta(P) < 0.3225$ [Sah]. In an online discussion about for Catalan posets, the second author improved this bound to $\delta(P_n) < 0.2995$, by comparing $x = (1, 17)$, $y = (2, 3)$ and taking n large enough [P2]. In a different direction, Olson and Sagan showed in [OS], that $\delta(P_\lambda) \le \frac{1}{3}$, for all Young diagrams $\lambda \vdash n$, s.t. $\lambda \ne (n)$, (1^n).

In our recent paper [CPP], we showed that $\delta(P_n) = O(n^{-1/2})$, giving the first bound that $\delta(P_n) \to 0$ as $n \to \infty$. Thus Theorem 1 is a substantial improvement over this result. More

Date: May 31, 2020.
2010 *Mathematics Subject Classification.* 05A16, 06A07, 60C05.
Key words and phrases. 1/3–2/3 conjecture, Catalan numbers, Catalan posets, linear extension, sorting probability.

generally, we showed that $\delta(P_\lambda) = O\big(n^{-1/2}\big)$, for all partitions $\lambda \vdash n$ with bounded length $\ell = \ell(\lambda)$, and such that $\lambda_\ell = \Omega(n)$. The tools in [CPP] rely on technical results in Algebraic Combinatorics. Here we present a more direct computation giving better bounds.

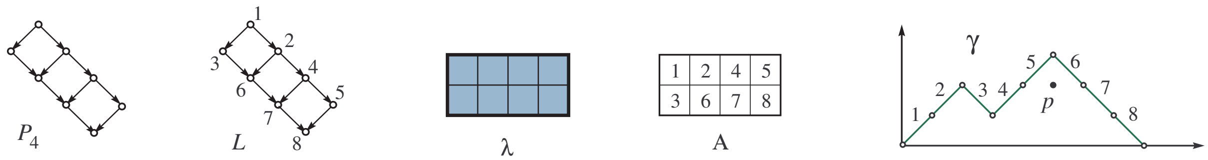

FIGURE 1. Catalan poset P_4, linear extension $L \in \mathcal{L}_{P_4}$, Young diagram $\lambda = (4, 4)$, the corresponding standard Young tableau $A \in \mathrm{SYT}(\lambda)$, and Dyck path $\gamma : (0, 0) \to (8, 0)$.

We have several motivations for studying the Catalan posets. On the one hand, they are a natural special case of general Young diagram posets P_λ, and Theorem 1 is perhaps an indication how sharp the general bounds in [CPP] are. On the other hand, they are closely related to the behavior of Brownian excursion, via the standard bijection from standard Young tableaux A of shape (n, n) to Dyck paths $\gamma : (0, 0) \to (2n, 0)$, see Figure 1.

Curiously, the sorting probability has a natural probabilistic interpretation in terms of Dyck paths:

$$\mathbf{P}\big[L(1, a) < L(2, b)\big] = \mathbf{P}\big[\gamma \text{ passes above } (a + b - 1, a - b)\big],$$

see Proposition 2. For example, for $A \in \mathrm{SYT}(4, 4)$ as in the figure, let $a = 5$ and $b = 2$. Then we have $A(1, 5) < A(2, 2)$, and the corresponding path γ is above point $p = (5, 2)$. Unfortunately the standard probabilistic tools for the Brownian excursion are too weak to establish Theorem 1, but they do give the right heuristic idea of how to approach the problem (see §5.1). Thus we resort to a direct asymptotic analysis of the sorting probabilities.

Notation. We write $C_1(\varepsilon), C_2(\varepsilon), \ldots$ to denote (effectively computable) positive constants that depend on a fixed parameter $\varepsilon > 0$, but not on n. Similarly, we write $C_1, C_2, \ldots$ to denote (effectively computable) absolute constants that do not depend on ε. In the paper, we identify P_n with Young diagram (n, n), and linear extensions $\mathcal{L}_{P_n}$ with standard Young tableaux $\mathrm{SYT}(n, n)$, see Figure 1. Here we use the matrix coordinates, so e.g. $L(2, 3) = 7$, for L as in the figure.

2. Sorting probability via lattice paths

Throughout the paper, let $I = \big[\frac{n}{10}, \frac{9n}{10}\big]$ and $J = \big[\frac{\sqrt{n}}{10}, 10\sqrt{n}\big]$. Our approach to proving Theorem 1 is to carefully analyze the *sorting probability function* $R_n(h, z) : I \times J \to [0, 1]$, defined as follows:

$$R_n(h, z) := \mathbf{P}\big[L(2, h - z) < L(1, h)\big], \quad \text{where} \quad h \in I,\ z \in J. \tag{2.1}$$

Consider the *lattice paths* γ in $\mathbb{N}^2$ from $(0, 0)$ to (n, n), which move up and to the right and do not go below (Southeast) of the main diagonal $(0, 0)$–(n, n). Denote by $\mathit{Cat}(n)$ the set of such paths. Our intuition comes from the following combinatorial interpretation already mentioned in the introduction.

Proposition 2. *The sorting probability function $R_n(h, z)$ is equal to the probability of a lattice path $\gamma \in \mathit{Cat}(n)$ to pass Southeast* (SE) *of the point $(h - z - \frac{1}{2}, h - \frac{1}{2})$.*

Proof. This follows from the bijection between lattice paths $\gamma \in \mathit{Cat}(n)$ and linear extensions $L \in \mathcal{L}_{P_n}$ via standard Young tableaux $A \in \mathrm{SYT}(n, n)$, as shown in Figure 1. Formally, let up-steps correspond to a square in the first row, and right-steps to squares in the second row. The details are straightforward. □

The next two lemmas describe the local behavior of the sorting probability function $R_n(h,z)$, in essence estimating discrete partial derivatives in both directions. These lemmas are key to the proof of Theorem 1. We prove the theorem in Section 3 and the lemmas in Section 4.

To simplify the notation, we extend this function to all real numbers: $R_n(h,z) := R_n(\lfloor h \rfloor, \lfloor z \rfloor)$.

Lemma 3. *For all $h \in I$ and $z \in J$, the sorting probability function satisfies:*

$$R_n\left(h, \sqrt{n}/10\right) \le \frac{1}{4} \quad \textit{and} \quad R_n(h, 10\sqrt{n}) \ge \frac{3}{4}, \tag{2.2}$$

and

$$\frac{C_1}{\sqrt{n}} \le R_n(h, z+1) - R_n(h,z) \le \frac{C_2}{\sqrt{n}}, \tag{2.3}$$

where $C_1, C_2 > 0$ are universal constants, and n is large enough.

In other words, the function $R_n(h,\cdot)$ is increasing and passing over $1/2$ at some point in the interval J.

Lemma 4. *For all $h \in \left[\frac{n}{10}, \frac{n+z+1}{2}\right] \subset I$ and $z \in J$, the sorting probability function satisfies:*

$$R_n(h,z) = R_n(n+z-h, z), \tag{2.4}$$

and

$$C_3 \frac{n-2h+z}{n^2} \le R_n(h,z) - R_n(h+1,z) \le C_4 \frac{n-2h+z}{n^2}, \tag{2.5}$$

where $C_3, C_4 > 0$ are universal constants, and n is large enough.

In other words, the function $R_n(\cdot, z)$ is symmetric, bimodal, and attains its minimum value at $h = \lfloor \frac{n+z}{2} \rfloor$. See Figure 2 and 5 for an illustration.

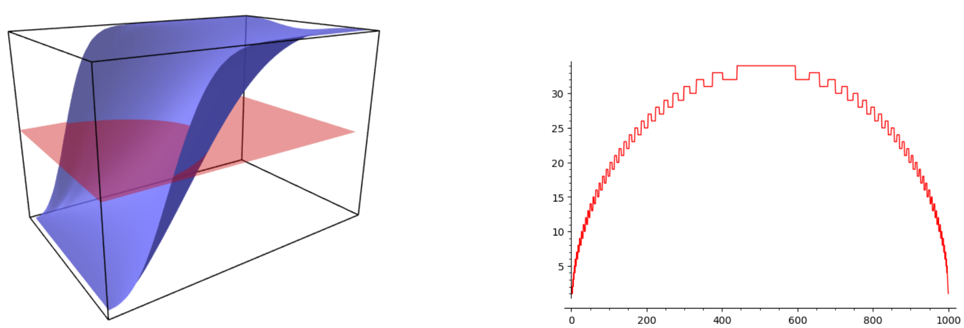

FIGURE 2. Left: Graph of the function $S(r,t)$ defined in §5.1, which coincides with the limit surface of the sorting probability $R_n(h,z)$ when $n \to \infty$. We also draw the red plane at height $\frac{1}{2}$ to indicate positions of the best sorting pairs $x = (1,h)$ and $y = (1, h-z)$. Right: The intersection between the plane and the surface on the left picture when $n = 1000$. The plot is the function $z = f(h)$ that minimizes $\left|R_n(h,z) - \frac{1}{2}\right|$ for $h \in [0,n]$.

In fact, the symmetry follows from the central symmetry of the Catalan poset P_n:

$$\mathbf{P}\left[\, L(2,b) < L(1,a) \,\right] = \mathbf{P}\left[\, L(2, n-a) < L(1, n-b) \,\right],$$

which proves (2.4).

3. Proof of Theorem 1

By Lemma 3, there exists $z \in J$, so that

$$\frac{1}{2} - \frac{C_2}{\sqrt{n}} \leq R_n\big(n/2, z\big) \leq \frac{1}{2}. \tag{3.1}$$

Let $h_0 := n/2 - K n^{3/4}$, where the constant $K > 0$ will be determined later. We have:

$$\begin{aligned} R_n(h_0, z) &= R_n\big(n/2, z\big) + \sum_{k=0}^{\lfloor K n^{3/4} \rfloor} R_n(h_0 + k, z) - R_n(h_0 + k + 1, z) \\ &\geq_{(2.5)} R_n\big(n/2, z\big) + C_3 \sum_{k=0}^{\lfloor K n^{3/4} \rfloor} \frac{|k - z|}{n^2} \\ &\geq R_n\big(n/2, z\big) + C_3 \frac{\big(K n^{3/4}\big)^2}{4 n^2} \\ &\geq_{(3.1)} \frac{1}{2} - \frac{C_2}{\sqrt{n}} + \frac{C_3 K^2}{4\sqrt{n}}. \end{aligned}$$

Taking $K := 2\sqrt{\frac{C_2}{C_3}}$, we get

$$R_n(h_0, z) \geq \frac{1}{2}. \tag{3.2}$$

It then follows from (3.1) and (3.2), that

$$R_n\big(n/2, z\big) \leq \frac{1}{2} \leq R_n\big(h_0, z\big).$$

Hence, there exists an integer $h_1 \in [h_0, n/2]$, such that $R_n(h_1 + 1, z) \leq \frac{1}{2} \leq R_n(h_1, z)$. We conclude:

$$\frac{1}{2} - R_n(h_1 + 1, z) \leq R_n(h_1, z) - R_n(h_1 + 1, z) \leq_{(2.5)} C_4 \frac{2K n^{3/4} + 10\sqrt{n}}{n^2} = O\big(n^{-5/4}\big).$$

This completes the proof of the theorem. □

Example 5. The construction in the proof is quite delicate, as it is fundamentally discrete rather than continuous. In Figure 3, we show the graph of $R_n(h, z)$ with $n = 1000$ and two values: $z = 33$ and $z = 34$. In the former case, the function intersects $\frac{1}{2}$, and $h_1 = 439$ as in the proof. In the latter case, the function is always above $\frac{1}{2}$.

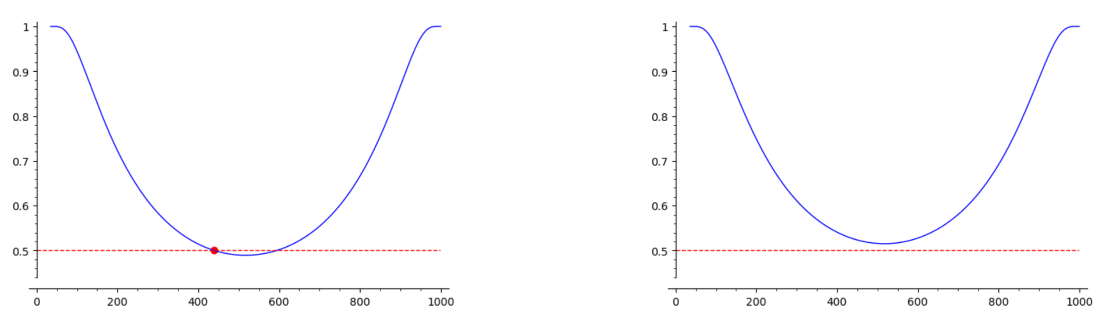

Figure 3. Functions $R_{1000}(h, 33)$ and $R_{1000}(h, 34)$.

4. Proof of lemmas

4.1. **Preliminaries.** For $0 \le a \le b$, denote by $f(a,b)$ the number of paths $\gamma : (0,0) \to (a,b)$ above diagonal $y = x$.

Lemma 6 (The ballot theorem, see e.g. [Fel, §III.1]). *For* $0 \le a \le b$,

$$f(a,b) \;=\; \binom{a+b}{a} - \binom{a+b}{a-1} \;=\; \binom{a+b}{a}\frac{b-a+1}{b+1}.$$

Let $p_n(a,b)$ be the probability that the lattice path $\gamma \in \mathit{Cat}(n)$ passes through the point (a,b). By definition,

$$p_n(a,b) \;=\; \frac{f(a,b)\cdot f(n-b,n-a)}{\mathrm{Cat}(n)}\,.$$

Lemma 7 ([MP2, Thm 3.3]). *Fix* $\varepsilon, K > 0$. *We have:*

$$p_n(h-z,h) \;\le\; \frac{C_1(\varepsilon)\cdot(z+1)^2}{n^{3/2}}\, e^{-z^2/n}, \tag{4.1}$$

for $\varepsilon n \le h \le (1-\varepsilon)n$, *and* $0 \le z < h$, *and* $C_1(\varepsilon)$ *a constant independent of* n. *Furthermore,*

$$C_2(\varepsilon,K)\,\frac{(z+1)^2}{n^{3/2}} \;\le\; p_n(h-z,h) \;\le\; C_3(\varepsilon,K)\,\frac{(z+1)^2}{n^{3/2}} \quad \textit{for} \quad 1 \le z \le K\sqrt{n}\,, \tag{4.2}$$

where h, z *as above, and* $C_2(\varepsilon,K)$, $C_3(\varepsilon,K) > 0$ *are constants independent of* n.

In fact, when $n \to \infty$, the constants in the theorem are computed explicitly in [MP2], but only the upper and lower bounds are needed in the proof of Lemmas 3 and 4.

Let $q_n(a,b)$ denote the probability that the lattice path $\gamma \in \mathit{Cat}(n)$ passes through both $(a,b-1)$ and (a,b). Similarly, let $r_n(a,b)$ denote the probability that the lattice path $\gamma \in \mathit{Cat}(n)$ passes through points $(a,b-1)$, (a,b) and $(a,b+1)$. From Lemma 7, we immediately have:

$$r_n(h-z,h) \;\le\; q_n(h-z,h) \;\le\; p_n(h-z,h) \;\le\; \frac{C_1(\varepsilon)\cdot(z+1)^2}{n^{3/2}}\, e^{-z^2/n} \tag{4.3}$$

This immediately gives the upper bound in

$$C_4(\varepsilon,K)\,\frac{(z+1)^2}{n^{3/2}} \;\le\; r_n(h-z,h) \;\le\; q_n(h-z,h) \;\le\; C_3(\varepsilon,K)\,\frac{(z+1)^2}{n^{3/2}} \quad \text{for} \;\; 1 \le z \le K\sqrt{n}\,, \tag{4.4}$$

The lower bound in (4.4) follows from

$$r_n(a,b) = \frac{f(a,b-1)\cdot f(n-b-1,n-a)}{\mathrm{Cat}(n)} \;=\; p_n(a,b)\,\frac{(n-b)(b-a)(b-a+2)(b+1)}{(2n-a-b)(b-a+1)^2(a+b)}\,.$$

Indeed, for $b = h$ and $a = h - z = b - o(b)$, one can take $C_4(\varepsilon,K) = C_2(\varepsilon,K)/5$ for $h > \varepsilon n$ large enough.

4.2. **Proof of Lemma 3.** By Proposition 2, the sorting probability function $R_n(h,z)$ is the probability that the vertical step at height h of a random lattice path $\gamma \in \mathit{Cat}(n)$ happens at $x \ge h - z$. This gives:

$$R_n(h,z) \;=\; \sum_{k=1}^{z} q_n(h-k,h)\,. \tag{4.5}$$

Since $q_n(h-k,h) \ge 0$, it then follows that $R_n(h,\cdot)$ is an increasing function for every h.

Now set $\varepsilon = \frac{1}{10}$, $K = 10$, and let $\varepsilon n \le h \le (1-\varepsilon)n$. We have:

$$R_n\big(h, \sqrt{n}/10\big) \;=\; \sum_{k=1}^{\sqrt{n}/10} q_n(h-k,h) \;\le_{(4.4)}\; \sum_{k=1}^{\sqrt{n}/10} C_3(\varepsilon,K)\,\frac{(k+1)^2}{n^{3/2}}$$
$$\le\; C_3(\varepsilon,K)\,\frac{\big(\sqrt{n}/10\big)^3}{n^{3/2}} \;=\; \frac{C_3(\varepsilon,K)}{1000}\,.$$

A direct computer calculation shows that

$$\frac{C_3\big(\frac{1}{10},10\big)}{1000} \;<\; \frac{1}{4}\,.$$

This proves the first inequality in (2.2).

On the other hand, we have:

$$R_n\,\big(h, K\sqrt{n}\big) \;=\; \sum_{k=1}^{K\sqrt{n}} q_n(h-k,h) \;=\; 1 \;-\; \sum_{k>K\sqrt{n}} q_n(h-k,h)$$
$$\ge_{(4.3)}\; 1 \;-\; C_1(\varepsilon) \sum_{k>K\sqrt{n}} \frac{(z+1)^2}{n^{3/2}}\, e^{-z^2/n}$$
$$\gtrsim\; 1 \;-\; C_1(\varepsilon) \int_K^\infty x^2\, e^{-x^2}\, dx\,.$$

A direct computer calculation shows that for $\varepsilon = \frac{1}{10}$ and $K = 10$, we have:

$$C_1\big(0.1\big) \int_{10}^\infty x^2\, e^{-x^2}\, dx \;<\; \frac{1}{4}\,.$$

This proves the second inequality in (2.2).

For (2.3), let $h \in I$, $z \in J$ be as in the lemma. We have:

$$R_n(h,z+1) \;-\; R_n(h,z) \;=_{(4.5)}\; q_n(h-(z+1),h)$$

and the bounds now follow from (4.2). This completes the proof of the Lemma 3. $\square$

4.3. **Bimodality.**

The following lemma is used in the proof of Lemma 4 in the next section.

Lemma 8. *Let $h \in [1, n-1]$ and $z \in [1, h-1]$. Then $R_n(h,z) > R_n(h+1,z)$ if and only if $h \le \frac{1}{2}(n+z)$.*

Proof. Let $A = (h-z+1/2, h+1/2)$ and $B = (h-z-1/2, h-1/2)$ be two points in the plane. By Proposition 2, the sorting probabilities $R_n(h+1,z)$ and $R_n(h,z)$ are probabilities that the lattice path $\gamma \in Cat(n)$ passes to SE of the points A and B, respectively. Denote by N_1 and N_2, respectively, the numbers of these paths. Then we have:

$$R_n(h,z) \;-\; R_n(h+1,z) \;=\; \frac{1}{\mathrm{Cat}(n)}\,\big(N_2 \;-\; N_1\big).$$

Let $x := h+1-z$. Denote by M_1 and M_2 the number of paths $\gamma \in Cat(n)$ which contain segments $(x-2,h) \to (x,h)$ and $(x-1,h-1) \to (x-1,h+1)$, respectively. Note that $N_2 - N_1$ is exactly the difference between the numbers of paths passing below point A but above B, and the paths passing left of A but right of B. Thus, $N_2 - N_1 = M_2 - M_1$.

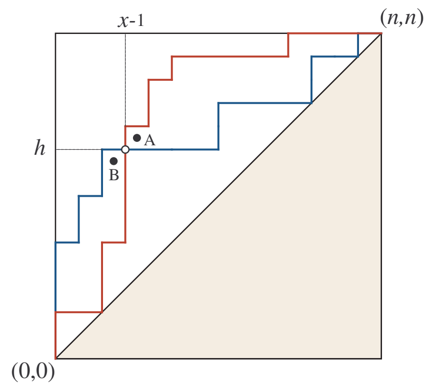

FIGURE 4. The two types of paths in the proof of Lemma 8, with M_1 counting blue paths and M_2 red paths.

We have:

$$M_1 \;=\; f(x-2,h)\, f(n-h,n-x) \;=\; \binom{x+h-2}{x-2}\binom{2n-x-h}{n-h}\frac{(h-x+3)(h-x+1)}{(h+1)(n-x+1)}\,,$$
$$M_2 \;=\; f(x-1,h-1)\, f(n-h-1,n-x+1) \;=\; \binom{x+h-2}{x-1}\binom{2n-x-h}{n-h-1}\frac{(h-x+3)(h-x+1)}{h(n-x+2)}$$

and therefore:

$$M_2 \,-\, M_1 \;=\; \binom{x+h-2}{x-1}\binom{2n-x-h}{n-h-1}\frac{(h-x+1)(h-x+2)(h-x+3)(n-x-h+1)}{h(h+1)(n-h)(n-x+2)}.$$

The last expression is ≥ 0 if and only if $h+x\leq n+1$, and the result follows. □

4.4. **Proof of Lemma 4.** Equation (2.4) is proved earlier. For (2.5), from the proof of Lemma 8 we have:

$$R_n(h,z) \,-\, R_n(h+1,z) \;=\; \frac{M_2}{\mathrm{Cat}(n)}\;\frac{(z+1)(n-2h+z)}{(h+1)(n-h)} \;=\; r_n(h-z,z)\,\frac{(z+1)(n-2h+z)}{(h+1)(n-h)}\,.$$

Since $z\in J$, we have:

$$r_n(h-z,z) \;=_{(4.2)}\; \Theta\left(\frac{1}{\sqrt{n}}\right).$$

On the other hand, since $h\in I$ and $z\in J$, we have:

$$\frac{(z+1)(n-2h+z)}{(h+1)(n-h)} \;=\; \Theta\left(\frac{\sqrt{n}}{n^2}\,(n-2h+z)\right).$$

Combining these two asymptotics, we conclude:

$$R_n(h,z) \,-\, R_n(h+1,z) \;=\; \Theta\left(\frac{n-2h+z}{n^2}\right).$$

This proves (2.5) and completes the proof of Lemma 4. □

5. Final remarks and open problems

5.1. The sorting probability function $R_n(h,z)$ is the discrete version of the continuous function

$$S(t,r) \,:=\, \mathbf{P}\big[B_0^+(t) \geq r\big],$$

where B_0^+ is the *Brownian excursion* on $[0,1]$, defined as the standard Brownian motion conditioned on the event $B_0^+(0)=B_0^+(1)=0$ and $B_0^+(t)>0$, for all $t\in(0,1)$. It has the following explicit density formula (see e.g., [IM, Pit]):

$$S(t,r) \;=\; \frac{2}{\sqrt{2\pi t^3(1-t^3)}}\int_0^r x^2 \exp\left(\frac{-x^2}{2t(1-t)}\right)dx,$$

see Figure 2. It is shown by Kaigh [Kai] that $R_n(h,z)$ converges to $S\left(\frac{h}{n},\frac{z}{\sqrt{2n}}\right)$ as $n\to\infty$. Unfortunately, the error terms of this convergence are too weak to imply Theorem 1. See Figure 5 for a plot comparing functions S and R_{200} side by side, and note that these graphs appear nearly identical on this scale.

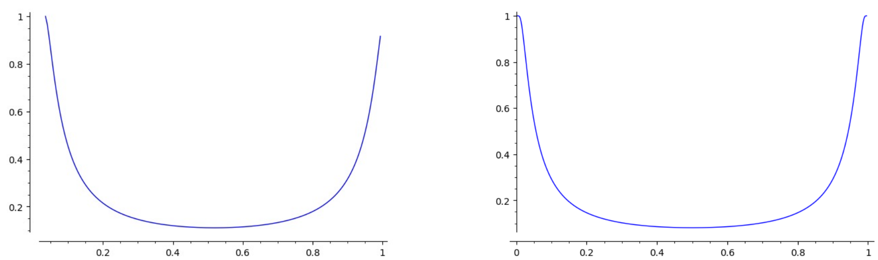

Figure 5. Left: The plot of $R_n(tn,\sqrt{n}/2)$, where $t\in(0,1)$ and $n=200$. Right: The plot of the probability $1-S(t,r)$, where $t\in(0,1)$ and $r=\sqrt{2}/4$.

5.2. Lemma 7 proved in [MP2] is one of many results in the context of the limit shape of *pattern avoiding permutations*, see e.g. [Kit] for an extensive overview of pattern avoidance. Many strongly related results are obtained in this direction, too many to list. Let us single out papers [AM, MP1] which are independent of [MP2], but cover the same pattern avoidance problem which translates into asymptotics of Dyck paths. Let us also mention two followup papers [HRS1, HRS2] which rederives and extends results in [MP1, MP2] via Brownian excursions.

5.3. In answering the second author's question [P2], Richard Stanley found the following curious limit formulas:

$$\lim_{n\to\infty}\mathbb{E}\big[L(1,k-1)\big] \,=\, 2k-\frac{k\binom{2k}{k}}{4^{k-1}}\,,\quad \lim_{n\to\infty}\mathbb{E}\big[L(2,k)\big] \,=\, 2k+\frac{k\binom{2k}{k}}{4^{k-1}}\,, \tag{5.1}$$

where the expectation is over random $L\in\mathcal{L}(P_n)$. The limits for probabilities $\mathbf{P}\big[L(1,a)<(2,b)\big]$ for fixed $a>b\geq 1$ also exist, but much less elegant. Stanley asked whether there are elegant expectation formulas similar to (5.1), for other partitions $\lambda=n\alpha$.

In principle, using the technology in [KS, Saks], one can use (5.1) to show that $\delta(P_n)<\frac{1}{e}+\varepsilon$ for all $\varepsilon>0$ and n large enough. Note that in [CPP] we already showed that $\delta(P_\lambda)=O(1/\sqrt{n})$ for the general *TVK case* $\lambda=n\alpha$.

5.4. It would be interesting to see how tight Theorem 1 is. Let

$$\alpha := \liminf_{n\to\infty} \frac{\log \delta(P_n)}{\log n} \quad \text{and} \quad \beta := \limsup_{n\to\infty} \frac{\log \delta(P_n)}{\log n}. \tag{5.2}$$

We conjecture that

$$-\infty < \alpha < \beta = -\frac{5}{4}. \tag{5.3}$$

In other words, we believe that our upper bound is asymptotically tight. On the other hand, we believe that the lower bound is substantially smaller, but still polynomial. This has to do with the fact that lim inf depends on number theoretic properties of n governing the position of $\frac{1}{2}$ in the interval $\big[R(h+1,z), R(h,z)\big]$. At the moment, we cannot even prove that $\delta(P_n) > 0$ for all $n \geq 3$. Finally, most speculatively, we conjecture that

$$\delta(P_n) = o\big(n^{-5/4}\big). \tag{5.4}$$

We refer to [CPP, §12-13] for further discussions and conjectures on the sorting probability.

5.5. Our computer calculations show that the sorting probability $\delta(P_n)$ has an erratic behavior, but seem to fit well Theorem 1 and the conjectures above. The first graph in Figure 6 shows that $\delta(n)n^{5/4}$ is always less than 3, but frequently greater than 1, and greater than $\frac{1}{3}$ at least half the time. While this may seem to point against (5.4), we believe it holds since a simple regression does indicate a very slow trend downwards.

Similarly, the second graph in Figure 6 shows that $\log_n \delta(n)$ is frequently smaller than $-\frac{5}{4}$, but is never too small, suggesting that $-3 < \alpha < \frac{3}{2}$, in the notation of (5.2). Perhaps, going far beyond $n = 1000$ would give further evidence in support or against the conjectures above. See the full sequences $\delta(P_n)\,\mathrm{Cat}(n)$ and $\frac{1}{2}(1-\delta(P_n))\,\mathrm{Cat}(n)$ at [OEIS, A335212] and [OEIS, A335213], respectively.

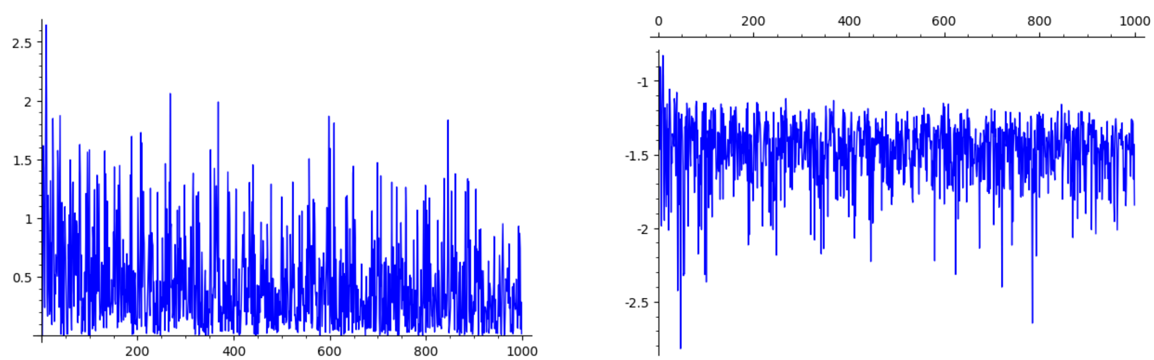

FIGURE 6. Graphs of $\delta(P_n)\,n^{5/4}$ and $\log_n \delta(P_n)$, for $3 \leq n \leq 1000$.

5.6. By the proof of Lemma 8, the integer $N_2 - N_1 \geq 0$ for $h \leq \frac{1}{2}(n+z)$. This is a fundamentally combinatorial statement about the difference in the number of certain lattice paths, somewhat similar in nature to the *super Catalan numbers*, see e.g. [P1, §4.5] for the references. It would be interesting to find an explicit combinatorial interpretation for $(N_2 - N_1)$.

Acknowledgements. We are grateful to Sam Hopkins, Han Lyu and Richard Stanley for interesting discussions and useful comments, and to MathOverflow for providing a convenient platform for such discussions. The last two authors were partially supported by the NSF.

References

[AM] M. Atapour and N. Madras, Large deviations and ratio limit theorems for pattern-avoiding permutations, *Combin. Probab. Comput.* **23** (2014), 161–200.

[CPP] S. H. Chan, I. Pak and G. Panova, Sorting probability for large Young diagrams, `arXiv:2005.08390`, 38 pp.

[Fel] W. Feller, *An introduction to probability theory and its applications*, Vol. I (third ed.), John Wiley, New York, 1968, 509 pp.

[HRS1] C. Hoffman, D. Rizzolo and E. Slivken, Pattern-avoiding permutations and Brownian excursion Part I: shapes and fluctuations, *Random Structures Algorithms* **50** (2017), 394–419.

[HRS2] C. Hoffman, D. Rizzolo and E. Slivken, Pattern-avoiding permutations and Brownian excursion Part II: fixed points, *Probab. Theory Related Fields* **169** (2017), 377–424.

[IM] K. Itô and H. McKean, *Diffusion processes and their sample paths*, Springer, Berlin, 1974, 321 pp.

[KS] J. Kahn and M. Saks, Balancing poset extensions, *Order* **1** (1984), 113–126.

[Kai] W. D. Kaigh, An invariance principle for random walk conditioned by a late return to zero, *Ann. Probab.* **4** (1976), 115–121.

[Kit] S. Kitaev, *Patterns in permutations and words*, Springer, Heidelberg, 2011, 494 pp.

[Lin] N. Linial, The information-theoretic bound is good for merging, *SIAM J. Comput.* **13** (1984), 795–801.

[MP1] N. Madras and L. Pehlivan, Structure of random 312-avoiding permutations, *Random Structures Algorithms* **49** (2016), 599–631.

[MP2] S. Miner and I. Pak, The shape of random pattern-avoiding permutation, *Adv. Appl. Math.* **55** (2014), 86–130.

[OS] E. J. Olson and B. E. Sagan, On the $\frac{1}{3}$–$\frac{2}{3}$ conjecture, *Order* **35** (2018), 581–596.

[P1] I. Pak, Complexity problems in enumerative combinatorics, `arXiv:1803.06636`, 31 pp., an expanded version of the paper in *Proc. ICM Rio de Janeiro*, Vol. 3, 2018, 3139–3166.

[P2] I. Pak, *Coincidences between average Catalan tableaux*, `MathOverflow`, Question 331579 (May 15, 2019), https://tinyurl.com/ybhvh4hv.

[Pit] J. Pitman, *Combinatorial stochastic processes*, Springer, Berlin, 2006, 256 pp.

[Sah] A. Sah, Improving the $\frac{1}{3}$–$\frac{2}{3}$ conjecture for width two posets, 20 pp.; `arXiv:1811.01500`.

[Saks] M. Saks, Balancing linear extensions of ordered sets, *Order* **2** (1985), 327–330.

[OEIS] N. J. A. Sloane, The Online Encyclopedia of Integer Sequences, `oeis.org`.

[S1] R. P. Stanley, *Enumerative Combinatorics*, vol. 1 (second ed.) and vol. 2, Cambridge Univ. Press, 2012 and 1999.

[S2] R. P. Stanley, *Catalan numbers*, Cambridge Univ. Press, 2015, 215 pp.

[Tro] W. T. Trotter, Partially ordered sets, in *Handbook of combinatorics*, Vol. 1, Elsevier, Amsterdam, 1995, 433–480.

(Swee Hong Chan) Department of Mathematics, UCLA, Los Angeles, CA 90095.
Email address: `sweehong@math.ucla.edu`

(Igor Pak) Department of Mathematics, UCLA, Los Angeles, CA 90095.
Email address: `pak@math.ucla.edu`

(Greta Panova) Department of Mathematics, USC, Los Angeles, CA 90089.
Email address: `gpanova@usc.edu`