On a random model of forgetting

Noga Alon [*] Dor Elboim [†] Allan Sly [‡]

Abstract

Georgiou, Katkov and Tsodyks considered the following random process. Let $x_1, x_2, \ldots$ be an infinite sequence of independent, identically distributed, uniform random points in $[0,1]$. Starting with $S = \{0\}$, the elements x_k join S one by one, in order. When an entering element is larger than the current minimum element of S, this minimum leaves S. Let $S(1,n)$ denote the content of S after the first n elements x_k join. Simulations suggest that the size $|S(1,n)|$ of S at time n is typically close to n/e. Here we first give a rigorous proof that this is indeed the case, and that in fact the symmetric difference of $S(1,n)$ and the set $\{x_k \geq 1-1/e : 1 \leq k \leq n\}$ is of size at most $\tilde{O}(\sqrt{n})$ with high probability. Our main result is a more accurate description of the process implying, in particular, that as n tends to infinity $n^{-1/2}\big(|S(1,n)| - n/e\big)$ converges to a normal random variable with variance $3e^{-2} - e^{-1}$.

1 Introduction

The following random process, which we denote by P, is considered by Georgiou, Katkov and Tsodyks in [8] motivated by the study of a possible simplified model for the process of forgetting. Let $x_1, x_2, \ldots$ be independent, identically distributed, uniform random points in $[0,1]$. Starting with $S = \{0\} = \{x_0\}$ the elements x_k enter S one by one, in order, and if when x_k enters S it is larger than the minimum member $\min(S)$ of S, this minimum leaves it.

[*]Department of Mathematics, Princeton University, Princeton, NJ 08544, USA and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, Israel. Email: nalon@math.princeton.edu.
Research supported in part by NSF grant DMS-1855464, ISF grant 281/17, BSF grant 2018267 and the Simons Foundation.

[†]Department of Mathematics, Princeton University, Princeton, NJ 08544, USA. Email: delboim@princeton.edu.

[‡]Department of Mathematics, Princeton University, Princeton, NJ 08544, USA. Email: allansly@princeton.edu. Research supported in part by NSF grant DMS-1855527, a Simons Investigator grant and a MacArthur Fellowship.

The set S will be called the memory. We let $P(n)$ be the finite process that stops at time n. Simulations indicate that the expected size of the memory S at time n is $(1+o(1))n/e$ where e is the basis of the natural logarithm, and that in fact with high probability (that is, with probability approaching 1 as n tends to infinity) the size at the end is very close to the expectation. Tsodyks [11] suggested the problem of proving this behavior rigorously. Our first result here is a proof that this is indeed the case, as stated in the following theorem.

Theorem 1. *In the process $P(n)$ the size of the memory S at time n is, with high probability, $(1+o(1))n/e$. Moreover, with high probability, the symmetric difference between the final content of the memory and the set $\{x_k \geq 1-1/e : k \leq n\}$ is of size at most $\tilde{O}(\sqrt{n})$.*

A similar result (without any quantitative estimates) has been proved independently by Friedgut and Kozma [7] using different ideas.

Our main result provides a more accurate description of the process P as stated in the following theorems. To state the next theorem we let $S(1,n)$ be the content of the list at time n and let $s(1,n) := |S(1,n)|$.

Theorem 2. *Let B_t be a standard Brownian motion. We have that*

$$\left\{\frac{s(1,tn)-tn/e}{\sqrt{n}}\right\}_{t>0} \xrightarrow{d} \left\{\frac{\sqrt{3-e}}{e}\cdot B_t\right\}_{t>0}, \quad n\to\infty.$$

Our next theorem describes the final set of points obtained in this process. In particular, it strengthens the estimate in Theorem 1 by removing the poly-logarithmic factor in the bound on the size of the symmetric difference between $S(1,n)$ and the set $\{x_k \geq z_0 : k \leq n\}$ where $z_0 := 1-1/e$. The theorem also provides the limiting distribution of this difference. To state the theorem we define

$$L_n := \big|S(1,n) \setminus \{x_k \geq z_0 : k \leq n\}\big| \quad \text{and} \quad R_n := \big|\{x_k \geq z_0 : k \leq n\} \setminus S(1,n)\big|.$$

Theorem 3. *As n tends to infinity we have*

$$\left\{\left(\frac{L_{tn}}{\sqrt{n}}, \frac{R_{tn}}{\sqrt{n}}\right)\right\}_{t>0} \xrightarrow{d} \left\{\sqrt{2}e^{-1}\big(M_t - B_t, M_t\big)\right\}_{t>0},$$

where B_t is a standard Brownian motion and $M_t := \sup_{s\leq t} B_s$ is the maximum process. In particular, if N is a standard normal random variable, then

$$\frac{L_n}{\sqrt{n}} \xrightarrow{d} \sqrt{2}e^{-1}|N| \quad \text{and} \quad \frac{R_n}{\sqrt{n}} \xrightarrow{d} \sqrt{2}e^{-1}|N|. \tag{1}$$

Moreover, the size of the symmetric difference $R_n + L_n$ has the following limit distribution

$$\frac{L_n+R_n}{\sqrt{n}} \xrightarrow{d} X, \quad \text{where } X \text{ has the density} \quad f(x) = \frac{e^3x^2}{2\sqrt{\pi}}e^{-e^2x^2/4}\,\mathbb{1}\{x>0\}. \tag{2}$$

In fact, in the following theorem, we give a complete description of the evolution of the process in a $n^{-1/2}$ neighbourhood around the critical point $z_0 = 1 - 1/e$. For $0 < z < 1$ and $n \geq 1$ we let $S(z,n)$ be the set of elements in the list that are smaller than z at time n and let $s(z,n) := |S(z,n)|$.

Theorem 4. *As n tends to infinity we have*

$$\left\{ \frac{s\big(z_0 + yn^{-1/2}, tn\big)}{\sqrt{n}} \right\}_{\substack{t>0 \\ y\in\mathbb{R}}} \xrightarrow{d} \left\{ \sqrt{2}e^{-1}B_t + yt - \inf_{x\leq t}\left(\sqrt{2}e^{-1}B_x + yx\right) \right\}_{\substack{t>0 \\ y\in\mathbb{R}}}.$$

Remark 1.1. *The convergence in Theorems 2, 3 and 4 is a weak convergence in the relevant Skorohod spaces. It is equivalent to the following statements. In Theorem 2, we define the random function $f_n(t) := n^{-1/2}\big(s(1, \lfloor tn \rfloor) - tn/e\big)$. Then, for all $M > 0$ there exists a coupling of the sequence f_n and a Brownian motion B_t such that almost surely*

$$\sup_{t\leq M} \left| f_n(t) - \frac{\sqrt{3-e}}{e} \cdot B_t \right| \to 0, \quad n \to \infty.$$

In Theorem 4, we define $f_n(t,y) := n^{-1/2} s\big(z_0 + yn^{-1/2}, \lfloor tn \rfloor\big)$. Then, for all $M > 0$, there exists a coupling such that almost surely

$$\sup_{t,|y|\leq M} \left| f_n(t,y) - \left(\sqrt{2}e^{-1}B_t + yt - \inf_{x\leq t}\left(\sqrt{2}e^{-1}B_x + yx\right)\right) \right| \to 0, \quad n \to \infty.$$

The convergence in Theorem 3 is similar.

The rest of the paper is organized as follows. In the next section we describe the proof of Theorem 1. It is based on subadditivity and martingale concentration applied to the natural Doob martingale associated with the process. The concentration part resembles the basic approach in [1]. In Section 3 we prove Theorem 2, Theorem 3 and Theorem 4. The main idea in the proofs is to consider the process

$$W(z,n) := \sum_{x\in S(z,n)} \frac{1}{1-x} \tag{3}$$

which we call "the branching martingale." As long as $S(z,n)$ is not empty, this process is a martingale when $z = 1 - 1/e$, a supermartingale when $z < 1 - 1/e$ and a submartingale when $z > 1 - 1/e$. The intuition behind this magical formula comes from a certain multi-type branching process. In a multi-type branching process, there are individuals of k different types. For each type i, an offspring distribution μ_i on $\mathbb{N}^k$ is given. In each step, every individual of type i gives birth to a random number of individuals of each type

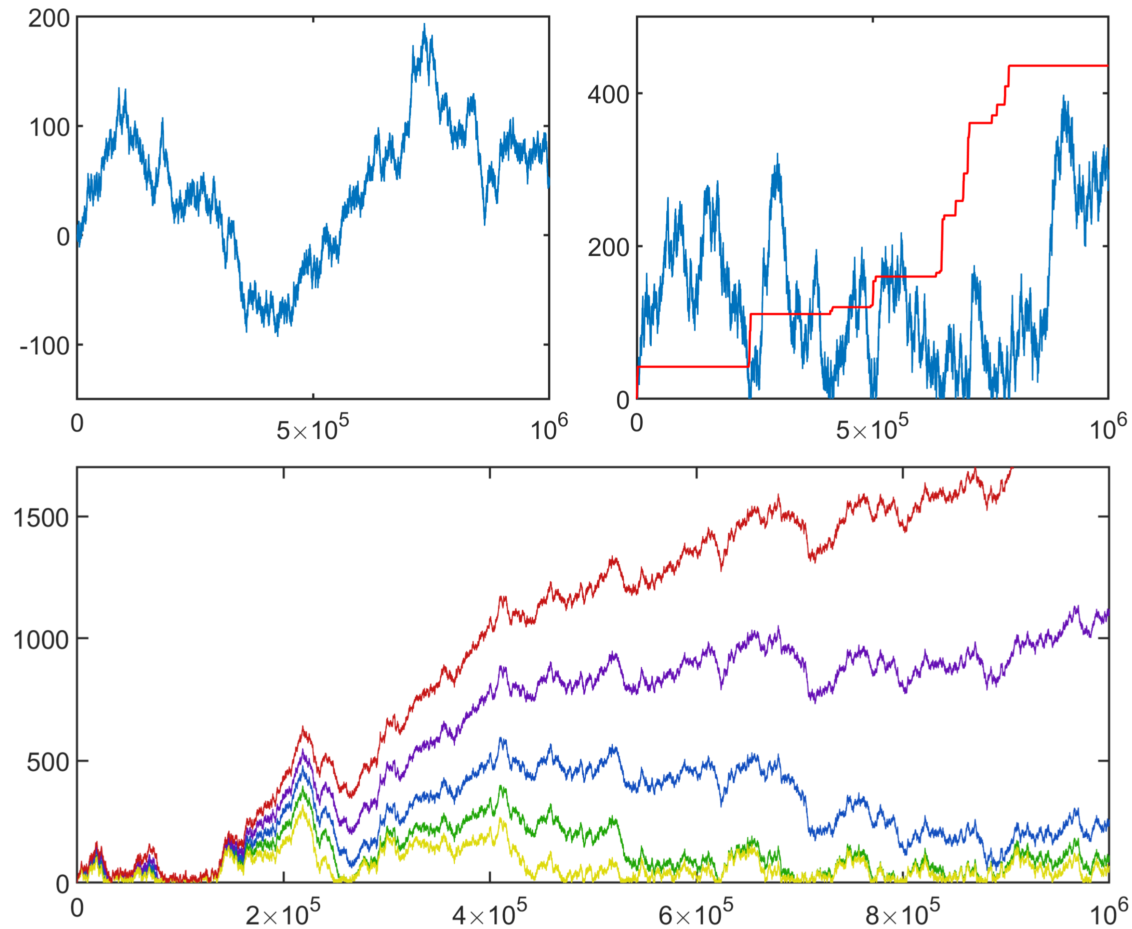

Figure 1: A computer simulation of the process running for $n = 10^6$ steps. Each picture is an independent sample of the process. The first picture shows the centered size of the list $s(1,k) - k/e$ as a function of $k \le n$. Theorem 2 states that this is a rescaled Brownian motion. The second picture shows the processes R_k in red and L_k in blue. These processes scale to the maximum process M_t and to $M_t - B_t$ respectively by Theorem 3. Note that R_k can grow only when L_k is zero and similarly, in the limit, M_t can grow only when $M_t - B_t$ is zero. The third picture shows $s(z_0 + yn^{-1/2}, k)$ as a function of k for $y = -2$ in yellow, $y = -1$ in green, $y = 0$ in blue, $y = 1$ in purple and $y = 2$ in red. Theorem 4 gives the scaling limit of these processes.

according to the distribution μ_i. Recall that a single-type branching process is critical when the expected number of offspring is 1 and in this case the size of the generation is a martingale. Similarly, for a multi-type branching process one defines the expectation matrix M where $M_{i,j} := \int x_j d\mu_i(x)$ is the expected number of children of type j of an individual of type i. The process is then critical when the maximal eigenvalue of M is one and in this case the process

$$N_t := u \cdot Z(t) \tag{4}$$

is a martingale where u is the eigenvector of M with eigenvalue 1 and $Z(t) = (Z_1(t), \dots, Z_k(t))$ is the number of individuals of each type at time t. For more background on multi-type branching processes see chapters V and VI in [3].

We think of our process as a multi-type branching process in which the type space is continuous. Given $0 < z < 1$, we define a branching process with the type space $[0, z]$. In this branching, the offspring distribution of an individual of type $x < z$ is given as follows. With probability $(1-z)$, the individual has 0 offspring, with probability $(z-x)$ it has one offspring of type that is uniformly distributed in $[x, z]$ and with probability x it has two offspring, one of type x and one of type that is uniformly distributed in $[0, x]$.

We now explain the connection between this branching and our process. Suppose we start the process with a list S such that the minimum of S is z, and we want to study the time until z is removed from the list. We think of each point $x < z$ that is added to the list as an individual of type x in the branching process and ignore points that fall to the right of z. We claim that the number of individuals that were born before the extinction of the branching described above is exactly the time until z is removed in the minimum process. Indeed, instead of exposing the branching in the usual way according to the generations, we expose the offspring of an individual when it becomes the current minimum of the list. When the element x becomes the minimum, with probability $1-z$ it is removed and $S(z,n)$ decreases by 1, with probability $z-x$ we remove x from $S(z,n)$ and replace it with a uniform element in $[x, z]$ and with probability x we keep x and add another element to $S(z,n)$ that is uniformly distributed in $[0, x]$.

The branching described above can be shown to be subcritical when $z < z_0$, critical when $z = z_0$ and supercritical when $z > z_0$, where $z_0 = 1 - 1/e$. Now, in the critical case we look for a martingale of the form (4) which translates in the continuous case to the process

$$N_n := \sum_{x \in S(z_0,n)} f(x) \tag{5}$$

where f is an eigenfunction of the expectation operator with eigenvalue 1. In order to find

f we let $m_n < z_0$ be the minimum of the list at time n and write

$$\mathbb{E}\big[N_{n+1} - N_n \mid \mathcal{F}_n\big] = \int_0^{z_0} f(x)dx - (1-m_n)f(m_n). \tag{6}$$

The last expression is zero only when $f(x) = 1/(1-x)$ which leads to (3).

2 Proof of Theorem 1

For any integer $0 \le \ell \le n$ and any $x \in [0,1]$ let $S(x,\ell)$ denote the set of numbers that are in S and lie in $[0,x]$ after the first ℓ steps, that is, after $x_1, x_2, \ldots, x_\ell$ have been inserted. Put $s(x,\ell) = |S(x,\ell)|$. Therefore $s(1,n)$ is the size of S at the end of the process. Let m_i denote the value of $\min(S)$ just before x_i enters it. Thus $m_1 = 0$ and $m_2 = x_1$.

We start with the following lemma showing that the expected value of the size of S at the end is linear in n.

Lemma 2.1. *The expected value of the random variable $s(1,n)$ is at least $(n+1)/4 > n/4$.*

Proof: For each $1 \le i \le n$ let X_i be the indicator random variable whose value is 1 if $x_i < m_i$ (and 0 otherwise). Let Y_i be the indicator random variable with value 1 if $x_i > m_i$ and $m_i \le 1/2$. Note that $\sum_{i=1}^n Y_i$ is at most the number of x_i, $0 \le i < n$ whose values are in $[0,1/2]$. Indeed, whenever $Y_i = 1$ an element $x_i \le 1/2$ leaves S. The expected value of this number is $1 + (n-1)/2$, as $x_0 = 0 < 1/2$ and any other x_i lies in $[0,1/2]$ with probability $1/2$. Therefore the expectation of $\sum_{i=1}^n Y_i$ is at most $(n+1)/2$. We claim that for every $1 \le i \le n$, the expected value of $2X_i + Y_i$ is at least 1. Indeed, if $m_i > 1/2$ the expected value of $2X_i$ is $2m_i > 1$ (and the expectation of Y_i is 0). If $m_i \le 1/2$, the expected value of $2X_i$ is $2m_i$, and that of Y_i is $1 - m_i$. By linearity of expectation in this case the expected value of $2X_i + Y_i$ is $2m_i + (1-m_i) = 1 + m_i \ge 1$. This proves the claim. Using again linearity of expectation we conclude that the expectation of $\sum_{i=1}^n (2X_i + Y_i)$ is at least n. Since the expectation of $\sum_{i=1}^n Y_i$ is at most $(n+1)/2$ it follows that the expectation of $\sum_{i=1}^n X_i$ is at least $(n-1)/4$. Note that the size of S at the end is exactly $1 + \sum_{i=1}^n X_i$, as it has size 1 at the beginning, its size never decreases, and it increases by 1 in step i if and only if $X_i = 1$. The assertion of the lemma follows. $\square$

We also need the following simple deterministic statement.

Lemma 2.2. *Let T_1, T_2 be arbitrary finite disjoint subsets of $[0,1]$. Let $P = P(n)$ be the process above with the sequence $x_1, x_2, \ldots, x_n \in [0,1]$ starting with $S = T_1$. Let P' denote the process with the same sequence x_i, starting with $S = T_1 \cup T_2$. For $x \in [0,1]$*

and $0 \leq \ell \leq n$, let $S(x,\ell)$ denote the set of numbers in S that lie in $[0,x]$ after the first ℓ steps in the process P. Similarly, let $S'(x,\ell)$ denote the set of numbers in S that lie in $[0,x]$ after the first ℓ steps in the process P'. Then $S'(x,\ell)$ contains $S(x,\ell)$ and $|S'(x,\ell)| - |S(x,\ell)| \leq |T_2|$.

Proof: We first describe the proof for $x = 1$. Apply induction on ℓ. The result is trivial for $\ell = 0$. Assuming it holds for ℓ consider step number $\ell + 1$ in the two processes, when $x = x_{\ell+1}$ enters the memory S. If x is smaller than the minimum of both sets $S(1,\ell)$ and $S'(1,\ell)$ then it joins the memory in both processes and the desired properties clearly hold. If it is larger than the minimum in $S(1,\ell)$ then it is also larger than the minimum in $S'(1,\ell)$, it joins the memory in both cases, and both minima leave. It is easy to check that the desired properties hold in this case too. Finally, if x is smaller than the minimum in $S(1,\ell)$ but larger than the minimum in $S'(1,\ell)$ then x joins both memories and the minimum of $S'(1,\ell)$ leaves. In this case S' still contains S, but the difference between their sizes decreases by 1. This also satisfies the desired properties, completing the proof of the induction step for $x = 1$. The proof for general x is essentially identical. When an arriving element x_i is smaller than x then the above argument can be repeated, and if it is larger it is also easy to check that the desired properties are maintained. $\square$

The above lemma has two useful consequences which are stated and proved in what follows.

Corollary 2.3. *For each $x \in [0,1]$ and any integer ℓ, let $E(s(x,\ell))$ denote the expectation of the random variable $s(x,\ell)$. Then for each fixed x the function $f(\ell) = E(s(x,\ell))$ is subadditive.*

Proof: We have to show that for every p, q, $f(p+q) \leq f(p) + f(q)$. Consider the process P with the points $x_1, x_2, \ldots x_p, y_1, y_2, \ldots, y_q$. The process starts with the first p steps, let T be the content of the memory S after these p steps. We can now compare the process P with the points $y_1, y_2, \ldots y_q$ starting with the empty set S (which is identical to starting with $S = \{0\}$), and the process P' with the same points y_i starting with $S = T$. By Lemma 2.2, throughout the two processes, the number of elements in the process P' that lie in $[0,x]$ is always at most the number of elements in P that lie in $[0,x]$, plus $|T|$. Taking expectations in both sides of this inequality implies the desired subadditivity. $\square$

Corollary 2.4. *For each $x \in [0,1]$ and any integer ℓ, the random variable $s(x,\ell)$ is 2-Lipschitz, that is, its value changes by at most 2 if we change the value of a single x_i. Therefore for any $\lambda > 0$ the probability that it deviates from its expectation by more than $\lambda\sqrt{\ell}$ is at most $2e^{-\lambda^2/8}$.*

Proof: The effect that changing the value of x_i to x'_i has on the content of S after step i is a change of x_i to x'_i in S, and possibly a removal of the minimum element of S before step i in one scenario and not in the other. In any case this means that one process can be converted to the other by adding an element to the memory and then by removing one or two elements from it. By Lemma 2.2 for each fixed choice of all other x_j the first modification can only increase the value of $s(x, \ell)$, and can increase it by at most 1. The second modification can then only decrease this value, decreasing it by at most 2. This implies that the value changes by at most 2, that is, the function is 2-Lipschitz. The concentration inequality thus follows from Azuma's Martingale Inequality, c.f. [2], Theorem 7.4.2. □

Proof of Theorem 1: Without any attempt to optimize the absolute constants, let $y \in [0, 1]$ be a real number so that the expected value $E(s(y, n/2))$ of the number of elements in the memory S that lie in $[0, y]$ at the end of the process with the first $n/2$ elements x_i is $5\sqrt{n \log n}$. Since $E(s(0, n/2)) = 0$, $E(s(1, n)) > n/8$ by Lemma 2.1, and the function $f(y) = E(s(y, n/2))$ is continuous, there is such a value (provided $n/8 \geq 5\sqrt{n \log n}$.) We claim that the expected number of indices i so that $n/2 < i \leq n$ and the minimum in the memory satisfies $m_i > y$, is smaller than $1/n^2$. Indeed, by the fact (which follows from Lemma 2.2) that the function $E(s(y, m))$ is monotone increasing in m, for every $\ell \in (n/2, n]$, $E(s(y, \ell)) \geq 5\sqrt{n \log n}$. By Corollary 2.4 this implies that for every fixed ℓ in this range the probability that m_ℓ is larger than y is at most $2e^{-25n \log n/(8\ell)} < 1/n^3$. Therefore, by linearity of expectation, the expected number of steps $i \in (n/2, n]$ in which $m_i > y$ is smaller than $1/n^2$, proving the claim.

By this claim, the expected number of steps $i \in (n/2, n]$ so that $m_i > y$ and $x_i > m_i$ is, of course, also at most $1/n^2$.

On the other hand, by the subadditivity established in Corollary 2.3, the expected value of $s(y, n)$ is at most $10\sqrt{n \log n}$.

We next show that y deviates from $1 - 1/e$ by $O(\frac{\sqrt{\log n}}{\sqrt{n}})$. To this end we compute the expectation of the random variable Y counting the number of indices $i \in (n/2, n]$ for which $x_i \leq y$ and x_i is being removed at a step in which the arriving element exceeds y. By definition $Y = \sum_{n/2 < i \leq n} X_i$ where X_i denotes the indicator random variable whose value is 1 if $x_i \leq y$ and x_i is being removed at a step in which the arriving element is larger than y. For each $i \in (n/2, n]$ let Z_i be the random variable whose value is 1 with probability $\frac{1-y}{1-x_i}$ if $x_i \leq y$ and is 0 otherwise. Note that if x_i is being removed during the process, then the probability it is being removed by an arriving element exceeding y is $\frac{1-y}{1-x_i}$. Therefore, in this case the expectation of X_i equals that of Z_i. The expected number of $i \in (n/2, n]$ for which $x_i < y$ is not removed is at most $10\sqrt{n \log n}$, as all these elements belong to $S(1, n)$. Therefore $|\sum_{n/2<i\le n} E(X_i) - \sum_{n/2<i\le n} E(Z_i)| \le 10\sqrt{n \log n}$.

As x_i is uniform in $[0, y]$ given that it is at most y, the expectation of Z_i is exactly

$$\int_{x=0}^{y} \frac{1-y}{1-x} dx = (1-y)\ln(\frac{1}{1-y}).$$

By linearity of expectation and the discussion above, the expectation of $Y = \sum_{i\in(n/2,n]} X_i$ is

$$(n/2 - O(\sqrt{n \log n}))(1-y)\ln(\frac{1}{1-y}).$$

On the other hand, the expected number of steps $j \in (n/2, n]$ in which an element $x_j > y$ arrived and caused the removal of an element $x_i = m_j < y$ for some $i \in (n/2, n]$ is $(n/2)(1-y) - O(\sqrt{n \log n})$. Indeed, the expected number of $j \in (n/2, n]$ with $x_j > y$ is $(n/2)(1-y)$, and almost each such x_j removes an element m_j, where the expected number of such m_j that exceed y is $o(1)$. In some of these steps the removed element may be x_i for some $i \le n/2$, but the expected number of such indices i is at most the expectation of $s(y, n/2)$ which is only $O(\sqrt{n \log n})$.

It follows that

$$(n/2 - O(\sqrt{n \log n}))(1-y)\ln(\frac{1}{1-y}) = (n/2)(1-y) - O(\sqrt{n \log n}).$$

Dividing by $n(1-y)/2$ (noting that $1-y$ is bounded away from 0) we conclude that $y = 1 - 1/e + O(\frac{\sqrt{\log n}}{\sqrt{n}})$.

Since the expected number of elements x_i for $i \in (n/2, n]$ that fall in the interval $[y - O(\frac{\sqrt{\log n}}{\sqrt{n}}), y + O(\frac{\sqrt{\log n}}{\sqrt{n}})]$ is $O(\sqrt{n \log n})$ we conclude that the expected number of steps $i \in (n/2, n]$ satisfying $x_i > 1 - 1/e$ that leave the memory during the process is $O(\sqrt{n \log n})$. In addition, since $E(s(y, n)) \le 10\sqrt{n \log n}$ it follows that the expected number of steps $i \in (n/2, n]$ so that $x_i < 1 - 1/e$ and x_i stays in the memory at the end of the process is also at most $O(\sqrt{n \log n})$.

By splitting the set of all n steps into dyadic intervals we conclude that the expected size of the symmetric difference between the final content of the memory and the set of all elements x_i larger than $z = 1 - 1/e$ is $O(\sqrt{n \log n})$. This clearly also implies that the expected value of $s(1, n)$ deviates from n/e by $O(\sqrt{n \log n})$. Finally note that by Corollary 2.4, for any positive λ the probability that either $s(1, n)$ or the above mentioned symmetric difference deviate from their expectations by more than $\lambda\sqrt{n}$ is at most $O(e^{-\Omega(\lambda^2)})$. $\square$

3 The branching martingale

In this section we prove Theorem 2, Theorem 3 and Theorem 4. The main idea in the proof is the following martingale which we call the branching martingale.

Let $0 < z < 1$ and let $z_0 := 1 - 1/e$ be the critical point. Recall that $S(z, n)$ is the set of elements in the list at time n that are smaller than z. Define the processes

$$W(z,n) := \sum_{x \in S(z,n)} \frac{1}{1-x} \qquad Z(z,n) := \sum_{k=1}^{n} W(z,k) \cdot \mathbb{1}\{W(z,k-1) = 0\}$$

and let $X(z, n) := W(z, n) - Z(z, n)$. The following claim is the fundamental reason to consider these processes.

Claim 3.1. *On the event $\{W(z, n) \neq 0\}$ we have that*

$$\mathbb{E}\big[W(z, n+1) - W(z, n) \,|\, \mathcal{F}_n\big] = -\log(1-z) - 1$$

and therefore $X(z_0, n)$ is a martingale. Moreover, Z is roughly minus the minimum process of X. More precisely, we have almost surely

$$0 \le Z(z,n) + \min_{k \le n} X(z,k) \le \frac{1}{1-z}. \tag{7}$$

Remark 3.2. *Claim 3.1 essentially explains all the results in this paper. See Theorem 4 for example. Since $X(z_0, n)$ is a martingale it is a Brownian motion in the limit and $Z(z_0, n)$ is minus the minimum process of this Brownian motion. Thus, $W(z_0, n)$, which is closely related to the number of elements in the list that are smaller than z_0 is the difference between the Brownian motion and its minimum.*

Proof of Claim 3.1. On the event $\{W(z, n) \neq 0\}$ we have

$$W(z,n+1) - W(z,n) = \mathbb{1}\{x_{n+1} \le z\} \frac{1}{1 - x_{n+1}} - \mathbb{1}\{x_{n+1} \ge m_n\} \frac{1}{1 - m_n}, \tag{8}$$

where m_n is the minimum of the list at time n and x_{n+1} is the uniform variable that arrives at time $n + 1$ in the process. Thus, on the event $\{W(z, n) \neq 0\}$ we have

$$\mathbb{E}\big[W(z,n+1) - W(z,n) \;\big|\; \mathcal{F}_n\big] = \int_0^z \frac{1}{1-x} dx - 1 = -\log(1-z) - 1.$$

It follows that $X(z_0, n)$ is a martingale. Indeed on the event $\{W(z, n) \neq 0\}$ we have

$$\mathbb{E}\big[X(z,n+1) - X(z,n) \;\big|\; \mathcal{F}_n\big] = \mathbb{E}\big[W(z,n+1) - W(z,n) \;\big|\; \mathcal{F}_n\big] = -\log(1-z_0) - 1 = 0.$$

Whereas on the event $\{W(z,n)=0\}$ we have that $X(n+1,z)=X(n,z)$.

We next turn to prove the second statement in the claim. For all $k\le n$ we have that

$$X(z,k)=W(z,k)-Z(z,k)\ge -Z(z,k)\ge -Z(z,n),$$

and therefore $\min_{k\le n}X(z,k)\ge -Z(z,n)$. Next, if $W(z,n)=0$ then

$$\min_{k\le n}X(z,k)\le X(z,n)=-Z(z,n).$$

If $W(z,n)\neq 0$, we let $n'\le n$ be the last time before n for which $W(z,n)=0$. We have that

$$\min_{k\le n}X(z,k)\le X(z,n')=-Z(z,n')\le -Z(z,n)+\frac{1}{1-x_{n'+1}}\le -Z(z,n)+\frac{1}{1-z}.$$

This finishes the proof of the claim. □

In the following section we give some rough bounds on the process that hold with very high probability. We will later use these a priori bounds in order to obtain more precise estimates and in order to prove the main theorems.

3.1 First control of the process

Definition 3.1. *We say that an event $\mathcal{A}$ holds with very high probability (WVHP) if there are absolute constants $C,c>0$ such that $\mathbb{P}(\mathcal{A})\ge 1-C\exp(-n^c)$.*

Let $\epsilon:=0.01$ and recall that $z_0:=1-1/e$. Define the events

$$\mathcal{A}_1:=\big\{\forall z<z_0,\ s(z,n)\le n^\epsilon/(z_0-z)\big\}$$

and

$$\mathcal{A}_2:=\big\{\forall z\ge z_0,\ s(z,n)\le (z-z_0)n+n^{1/2+\epsilon}\big\}.$$

Lemma 3.3. *The event $\mathcal{A}_1\cap\mathcal{A}_2$ holds WVHP.*

Proof. Let $z\le z_0$ and note that $W(z,n)\ge s(z,n)$. Thus, we have that

$$\big\{s(z,n)\ge n^\epsilon/(z_0-z)\big\}\subseteq\big\{W(z,n)\ge n^\epsilon/(z_0-z)\big\}\subseteq\bigcup_{k\le n}\mathcal{C}_k \tag{9}$$

where the events $\mathcal{C}_k$ are defined by

$$\mathcal{C}_k=\big\{0<W(z,k)\le e,\ \forall k\le m\le n, W(z,m)\neq 0 \text{ and } W(z,n)\ge n^\epsilon/(z-z_0)\big\}.$$

Next, we show that each of these events has a negligible probability. To this end, let $k \le n$ and define

$$M_m := W(z,m) + (m-k)(\log(1-z)+1), \quad \tau := \min\{m \ge k : W(z,m) = 0\}.$$

By Claim 3.1, the process $M_{m\wedge\tau}$ is a martingale. Moreover, since $\log(1-z)+1 \ge 1.1(z_0-z)$, we have that

$$\begin{aligned}\mathcal{C}_k &\subseteq \big\{M_{n\wedge\tau} - M_{k\wedge\tau} \ge n^\epsilon/(z_0-z) - e + (n-k)(\log(1-z)+1)\big\}\\ &\subseteq \big\{M_{n\wedge\tau} - M_{k\wedge\tau} \ge n^\epsilon/(z_0-z) + (n-k)(z_0-z)\big\}.\end{aligned}$$

Thus, using that $M_{m\wedge\tau}$ has bounded increments, we get from Azuma's inequality that

$$\mathbb{P}\big(\mathcal{C}_k\big) \le \exp\Big(-c\,\frac{n^{2\epsilon}(z_0-z)^{-2} + (n-k)^2(z_0-z)^2}{n-k}\Big) \le C\exp(-cn^c),$$

where the last inequality clearly holds both when $n-k \le n^\epsilon(z_0-z)^{-2}$ and when $n-k \ge n^\epsilon(z_0-z)^{-2}$. It follows from (9) that

$$\mathbb{P}\big(s(z,n) \ge n^\epsilon/(z_0-z)\big) \le C\exp(-n^c).$$

In order to obtain this inequality simultaneously for all $z \le z_0$ we first use a union bound to get it simultaneously for all $z \in \{i/n : i \in \mathbb{N}, i/n \le z_0\}$ and then argue that $s(z,n)$ does not change by much w.v.h.p when z changes by $O(n^{-1})$. The details are omitted. It follows that $\mathbb{P}(\mathcal{A}_1) \ge 1 - C\exp(-n^c)$.

We turn to bound the probability of $\mathcal{A}_2$. To this end, let $z \ge z_0$ and $z_1 := z_0 - n^{-1/2}$. The event

$$\big\{s(z_1,n) \le n^{1/2+\epsilon}/2\big\} \cap \big\{\big|\{k \le n : x_k \in [z_1,z]\}\big| \le (z-z_0)n + n^{1/2+\epsilon}/2\big\}$$

holds w.v.h.p by the first part of the proof and Chernoff inequality for the Binomial random variable. On this event we have that $s(z,n) \le (z-z_0)n + n^{1/2+\epsilon}$ and therefore, using the discretization argument as above we obtain $\mathbb{P}(\mathcal{A}_2) \ge 1 - C\exp(-n^c)$. □

Next, define the random variable

$$M_n(z) := \big|\{k \le n : m_k \ge z\}\big|$$

and the events

$$\mathcal{B}_1 := \big\{\forall z > z_0,\ M_n(z) \le n^\epsilon/(z-z_0)\big\}$$

and

$$\mathcal{B}_2 := \big\{\forall z \le z_0,\ \big|M_n(z) - (\log(1-z)+1)n\big| \le n^{1/2+\epsilon}\big\}.$$

Lemma 3.4. *The event $\mathcal{B}_1 \cap \mathcal{B}_2$ holds WVHP.*

Proof. We start by bounding the probability of $\mathcal{B}_2$. Let $z \le z_0$. By Claim 3.1 we have that

$$\mathbb{E}\big[X(z,k+1) - X(z,k) \bigm| \mathcal{F}_k\big] = \mathbb{1}\{W(z,k) \neq 0\} \cdot \big(-\log(1-z) - 1\big)$$

and therefore the process

$$N(z,n) := X(z,n) + (\log(1-z)+1)n - (\log(1-z)+1)M_n(z) \tag{10}$$

is a martingale. Thus, by Azuma's inequality WVHP $|N(z,n)| \le n^{1/2+\epsilon}$. Next, we have that

$$\mathbb{E}\big[Z(z,k+1) - Z(z,k) \mid \mathcal{F}_k\big] = \mathbb{1}\{m_k \ge z\} \int_0^z \frac{1}{1-x} = -\mathbb{1}\{m_k \ge z\} \log(1-z)$$

and therefore by Azuma's inequality we have WVHP

$$\big|Z(z,n) + \log(1-z)M_n(z)\big| \le 1 + \big|Z(z,n) - \log(1-z)M_{n-1}(z)\big| \le n^{1/2+\epsilon}.$$

Thus, on the event $\mathcal{A}_2$ we have that

$$\begin{aligned}\big|X(z,n) - \log(1-z)M_n(z)\big| &\le \big|X(z,n) - Z(z,n)\big| + n^{1/2+\epsilon} = W(z,n) + n^{1/2+\epsilon} \\ &\le e \cdot s(z,n) + n^{1/2+\epsilon} \le e \cdot s(z_0,n) + n^{1/2+\epsilon} \le n^{1/2+2\epsilon}.\end{aligned}$$

Substituting this back into (10) we obtain that WVHP

$$\big|M_n(z) - (\log(1-z)+1)n\big| \le n^{1/2+3\epsilon}.$$

This shows that $\mathcal{B}_2$ holds WVHP using the discretization trick. Indeed, WVHP each $z \in [0,1]$ is the minimum at most n^ϵ times and therefore WVHP $M_n(z)$ won't change by more than $n^{2\epsilon}$ when z changes by $O(n^{-1})$.

We turn to bound the probability of $\mathcal{B}_1$. Let $z_0 < z \le 3/4$ and let $n_0 := n^\epsilon (z-z_0)^{-2}$. By the same arguments as in the first part of the proof we have with very high probability for all $n_0 \le k \le n$

$$M_k(z) \le M_k(z_0) \le k^{1/2+\epsilon} \quad \text{and} \quad |N(z,k)| \le k^{1/2+\epsilon}.$$

Substituting these estimates back into (10) and using that $X(z,n) \le W(z,n)$ and that $\log(1-z) + 1 \le -e(z-z_0)$ we obtain WVHP for all $n_0 \le k \le n$

$$W(z,k) \ge e(z-z_0)k - 4k^{1/2+\epsilon} > 0.$$

Thus, WVHP, for all $n_0 \le k \le n$ we have that $m_k \le z$ and therefore WVHP

$$M_n(z) = M_{n_0}(z) \le M_{n_0}(z_0) \le n_0^{1/2+\epsilon} \le n^\epsilon/(z - z_0).$$

Of course the same bound holds when $3/4 \le z \le 1$ and therefore, using the discretization trick we get that $\mathcal{B}_1$ holds WVHP.

□

3.2 Proof of Theorem 4

The main step toward proving the theorem is the following proposition that determines the scaling limit of $X(z,n)$ when z is in a small neighborhood around $z_0 = 1 - 1/e$.

Proposition 3.5. *We have that*

$$\Big\{\frac{X(z_0 + yn^{-1/2}\,,\, tn)}{\sqrt{n}}\Big\}_{\substack{t>0\\ y\in\mathbb{R}}} \xrightarrow{d} \big\{\sqrt{2}B_t + eyt\big\}_{\substack{t>0\\ y\in\mathbb{R}}}, \qquad n\to\infty.$$

We start by proving Theorem 4 using Proposition 3.5.

Proof of Theorem 4. By Proposition 3.5 and (7) we have that the joint distribution

$$\Big(\frac{X(z_0 + yn^{-1/2}\,,\, tn)}{\sqrt{n}}\;,\;\frac{Z(z_0 + yn^{-1/2}\,,\, tn)}{\sqrt{n}}\Big)$$

converges to $\big(\sqrt{2}B_t + eyt\,,\, -\inf_{s\le t}(\sqrt{2}B_t + eys)\big)$. Thus, using that $W(z,n) = X(z,n) + Z(z,n)$ we obtain

$$\Big\{\frac{W(z_0 + yn^{-1/2}\,,\, tn)}{\sqrt{n}}\Big\}_{\substack{t>0\\ y\in\mathbb{R}}} \xrightarrow{d} \big\{\sqrt{2}B_t + eyt - \inf_{s\le t}\big(\sqrt{2}B_t + eys\big)\big\}_{\substack{t>0\\ y\in\mathbb{R}}}. \tag{11}$$

For the proof of Theorem 4 it suffices to show that $W(z_0 + yn^{-1/2}\,,\,k)$ in the equation* above can be replaced by $e\cdot s(z_0 + yn^{-1/2}, k)$. Intuitively, it follows as most of the terms in the sum in the definition of W are close to e. Formally, let $z_1 := z_0 - n^{-1/4}$ and $z_2 := z_0 + n^{-1/2+\epsilon}$. By Lemma 3.3 we have WVHP that for all $z_1 \le z \le z_2$ and $k \le n^{1+\epsilon}$

$$\begin{aligned}\big|W(z,k) - e\cdot s(z,k)\big| &\le Cs(z_1,k) + \sum_{x\in S(z,k)\setminus S(z_2,k)} \Big|\frac{1}{1-x} - e\Big| \\ &\le Cn^{1/4+\epsilon} + Cn^{-1/4}s(z_2,k) \le n^{1/4+2\epsilon}.\end{aligned} \tag{12}$$

This shows that W can be replaced by $e\cdot s$ in equation (11) and therefore finishes the proof of the theorem. □

We turn to prove Proposition 3.5. The proposition follows immediately from the following two propositions.

Proposition 3.6. *We have that*

$$\Big\{\frac{X(z_0,tn)}{\sqrt{n}}\Big\}_{t>0} \xrightarrow{d} \big\{\sqrt{2}B_t\big\}_{t>0}, \quad n\to\infty.$$

Proposition 3.7. *We have WVHP for all z with $|z-z_0|\le n^{1/2+\epsilon}$ that*

$$\big|X(z,n)-X(z_0,n)-en(z-z_0)\big|\le n^{1/4+3\epsilon}.$$

We start by proving Proposition 3.6. The main tool we use is the following martingale functional central limit theorem [5, Theorem 8.2.8] .

Theorem 3.8. *Let M_n be a martingale with bounded increments. Recall that the predictable quadratic variation of M_n is given by*

$$V_n := \sum_{k=1}^{n}\mathbb{E}\big[\big(M_k-M_{k-1}\big)^2 \mid \mathcal{F}_{k-1}\big].$$

Suppose that for all $t>0$ we have that

$$\frac{V_{tn}}{n}\xrightarrow{p}\sigma^2 t, \quad n\to\infty.$$

where the convergence is in probability. Then

$$\Big\{\frac{M_{tn}}{\sqrt{n}}\Big\}_{0<t<1} \xrightarrow{d} \big\{\sigma B_t\big\}_{0<t<1},$$

where B_t is a standard Brownian motion.

In order to use Theorem 3.8 we need to estimate the predictable quadratic variation of the martingale $X(z_0,n)$. To this end we need the following lemma.

Lemma 3.9. *For any bounded function $f:[0,1]\to\mathbb{R}$ that is differentiable on $[0,z_0]$ we have*

$$\mathbb{P}\Big(\Big|\sum_{k=1}^{n}f(m_k)-n\int_0^{z_0}\frac{f(x)}{1-x}dx\Big|\le n^{3/4+4\epsilon}\Big)\ge 1-C\exp(-n^c),$$

where the constants C,c may depend on the function f.

Proof. For any $z \leq z_0$ let

$$I_n(z) := \big\{ m_k : k \leq n \text{ and } z \leq m_k \leq z + n^{-1/4} \big\}.$$

We have that $|I_n(z)| = M_n(z) - M_n(z + n^{1/4})$ and therefore, using the expansion

$$\log(1 - (z + n^{-1/4})) = \log(1 - z) + n^{-1/4}/(1 - z) + O(n^{-1/2})$$

and Lemma 3.4 we obtain that w.v.h.p for all $z \leq z_0 - n^{-1/4}$

$$\big| |I_n(z)| - n^{3/4}/(1 - z) \big| \leq n^{1/2+2\epsilon}. \tag{13}$$

Next, Let $z_i := (i - 1)n^{-1/4}$ for any $1 \leq i \leq i_0$ where $i_0 := \lfloor n^{1/4} z_0 \rfloor$. We have WVHP

$$\begin{aligned}
&\Big| \sum_{k=1}^{n} f(m_k) - n \int_0^{z_0} \frac{f(x)}{1 - x} dx \Big| \\
&\leq \sum_{i=1}^{i_0 - 1} \Big| \sum_{m \in I_n(z_i)} f(m) - n \int_{z_i}^{z_{i+1}} \frac{f(x)}{1 - x} dx \Big| + \sum_{k=1}^{n} f(m_k) \cdot \mathbb{1}\{m_k \geq z_{i_0}\} + n \int_{z_{i_0}}^{z_0} \frac{f(x)}{1 - x} dx \\
&\leq C \sum_{i=1}^{i_0 - 1} \Big(\big| |I_n(z_i)| - n^{3/4}/(1 - z_i) \big| + n^{-1/4} |I_n(z_i)| + n^{1/2} \Big) + C M_n(z_0 - n^{-1/4}) + C n^{3/4} \\
&\leq C n^{3/4+2\epsilon} + C n^{-3/4} + C n^{3/4+\epsilon} \leq n^{3/4+3\epsilon},
\end{aligned}$$

where in the second inequality we used the approximations $f(m) = f(z_i) + O(n^{-1/4})$ for all $m \in I_n(t_i)$ and $f(x)/(1 - x) = f(z_i)/(1 - z_i) + O(n^{-1/4})$ for all $x \in [z_i, z_{i+1}]$. In the last inequality we used (13). $\square$

We can now prove Proposition 3.6.

Proof of Proposition 3.6. In order to use Theorem 3.8 we compute the predictable quadratic variation of $X(z_0, k)$. Clearly, on the event $\{m_k \geq z_0\}$ we have that $\mathbb{E}\big[(X_{k+1} - X_k)^2 \bigm| \mathcal{F}_k\big] = 0$. On the event $\{m_k \leq z_0\}$ we have

$$\begin{aligned}
\mathbb{E}\big[(X(z_0, k+1) - X(z_0, k))^2 \bigm| \mathcal{F}_k\big] &= \mathbb{E}\big[(W(z_0, k+1) - W(z_0, k))^2 \bigm| \mathcal{F}_k\big] \\
&= \int_0^{z_0} \frac{1}{(1 - x)^2} dx - \frac{2}{1 - m_k} \int_{m_t}^{z_0} \frac{1}{1 - x} dx + \frac{1}{1 - m_k} \\
&= \frac{1}{1 - x}\Big|_0^{z_0} + \frac{2}{1 - m_k} \log(1 - x)\Big|_{m_k}^{z_0} + \frac{1}{1 - m_k} = e - 1 - \frac{2\log(1 - m_k) + 1}{1 - m_k}.
\end{aligned}$$

Thus, the predictable quadratic variation of $X(z_0, k)$ is given by

$$V_n := \sum_{k=0}^{n-1} \mathbb{E}\big[(X(z_0, k) - X(z_0, k))^2 \bigm| \mathcal{F}_k\big] = \sum_{k=1}^{n-1} f(m_k), \tag{14}$$

where

$$f(x) := \Big(e - 1 - \frac{2\log(1-x)+1}{1-x}\Big) \cdot \mathbb{1}\{x \le z_0\}.$$

Next, we have that

$$\int_0^{z_0} \frac{f(x)}{1-x}dx = \int_0^{z_0} \frac{1}{1-x}\Big(e - 1 - \frac{2\log(1-x)+1}{1-x}\Big)dx$$
$$= (1-e)\log(1-x)\Big|_0^{z_0} - \frac{2\log(1-x)+3}{1-x}\Big|_0^{z_0} = e-1-e+3 = 2$$

and therefore, by Lemma 3.9 we have WVHP that $|V_n - 2n| \le Cn^{3/4+3\epsilon}$. It follows that for all $t > 0$ we have $V_{tn}/n \to 2t$ in probability and therefore by Theorem 3.8 we have

$$\Big\{\frac{X(z_0, tn)}{\sqrt{n}}\Big\}_{t>0} \xrightarrow{d} \big\{\sqrt{2}B_t\big\}_{t>0}$$

as needed. □

We turn to prove Proposition 3.7. The main tool we use is the following martingale concentration result due to Freedman [6]. See also [4, Theorem 1.2].

Theorem 3.10 (Freedman's inequality)**.** *Let M_n be a martingale with increments bounded by M and let*

$$V_n := \sum_{k=1}^{n} \mathbb{E}\big[(M_{k+1} - M_k)^2 \mid \mathcal{F}_k\big]$$

be the predictable quadratic variation. Then,

$$\mathbb{P}\big(|M_n| \ge x,\ V_n \le y\big) \le \exp\Big(-\frac{x^2}{2y + 2Mx}\Big).$$

We can now prove Proposition 3.7.

Proof of Proposition 3.7. Let z such that $|z - z_0| \le n^{-1/2+\epsilon}$ and recall that the process

$$N(z,k) := X(z,k) + (\log(1-z)+1)k - (\log(1-z)+1)\cdot M_k(z)$$

defined in (10) is a martingale. Also, define the difference martingale $L(z,k) := N(z,k) - N(z_0,k)$ and note that

$$\big|L(k+1,z) - L(k,z)\big| \le C|z-z_0| + C\mathbb{1}\{x_{k+1} \in [z_0,z]\cup[z,z_0]\} + C\mathbb{1}\{m_k \ge z_0 - n^{-1/2+\epsilon}\}.$$

Thus, by Lemma 3.4 we have WVHP

$$V_n := \sum_{k=1}^{n} \mathbb{E}\big[(L(z,k) - L(z,k-1))^2 \mid \mathcal{F}_{k-1}\big] \le Cn^{1/2+\epsilon} + CM_n(z_0 - n^{-1/2+\epsilon}) \le n^{1/2+2\epsilon}.$$

We obtain using Theorem 3.10 that

$$\mathbb{P}\big(\big|L(z,n)\big| \geq n^{1/4+2\epsilon}\big) \leq \mathbb{P}\big(\big|L(z,n)\big| \geq n^{1/4+2\epsilon},\ V_n \leq n^{1/2+2\epsilon}\big) + \mathbb{P}\big(V_k \geq n^{1/2+2\epsilon}\big)$$
$$\leq C\exp(-n^c).$$

Finally, using the expansion $\log(1-z)+1 = -e(z-z_0) + O(n^{-1+2\epsilon})$ we get that WVHP

$$\big|X(z,n) - X(z_0,n) - en(z-z_0)\big| \leq \big|L(z,k)\big| + Cn^{2\epsilon} + C\cdot|z-z_0|\cdot M_n(z) \leq n^{1/4+3\epsilon}.$$

This finishes the proof of the proposition using the discretization trick. □

3.3 Proof of Theorem 3

The main statement in Theorem 3 is the scaling limit result

$$\Big\{\Big(\frac{L_{tn}}{\sqrt{n}}\,,\,\frac{R_{tn}}{\sqrt{n}}\Big)\Big\}_{t>0} \xrightarrow{d} \big\{\sqrt{2}e^{-1}\big(M_t - B_t, M_t\big)\big\}_{t>0}. \tag{15}$$

We start by proving (15) and then briefly explain how the other statements in the Theorem 3 follow.

By Proposition 3.6 and Claim 3.1 we have that

$$\big\{n^{-1/2}\big(W(z_0,tn)\ ,\ Z(z_0,tn)\big)\big\}_{t>0} \xrightarrow{d} \big\{\sqrt{2}\big(B_t - \inf_{x<t} B_x\ ,\ -\inf_{x<t} B_x\big)\big\}_{t>0}. \tag{16}$$

Moreover, by symmetry

$$\big\{\sqrt{2}\big(B_t - \inf_{x<t} B_x\ ,\ -\inf_{x<t} B_x\big)\big\}_{t>0} \overset{d}{=} \big\{\sqrt{2}\big(M_t - B_t\ ,\ M_t\big)\big\}_{t>0}. \tag{17}$$

Next, recall that $L_n = s(z_0,n)$ and therefore, using the same arguments as in (12) we obtain that WVHP

$$\big|e\cdot L_n - W(z_0,n)\big| \leq n^{1/3}. \tag{18}$$

The convergence in (15) clearly follows from (16), (17), (18) and the following lemma

Lemma 3.11. *We have WVHP*

$$\big|e\cdot R_n - Z(z_0,n)\big| \leq n^{1/3}.$$

Proof. We write

$$\big|e\cdot R_n - Z(z_0,n)\big| \leq e\big|R_n - R_n'\big| + \big|e\cdot R_n' - M_{n-1}(z_0)\big| + \big|M_{n-1}(z_0) - Z(z_0,n)\big|, \tag{19}$$

where $R_n' := \sum_{k=1}^{n-1} \mathbb{1}\{m_k \geq z_0\}(1-m_k)$ and bound each one of the terms in the right hand side of (19).

We start with the first term. Note that $R_n = \big|\{2 \le k \le n : m_k \ge m_{k-1} \ge z_0\}\big|$ and therefore

$$\mathbb{E}\big[R_{n+1} - R_n \mid \mathcal{F}_n\big] = \mathbb{1}\big\{m_{n-1} \ge z_0\big\}(1 - m_{n-1}).$$

It follows that $N_n := R_n - R'_n$ is a martingale. The predictable quadratic variation of N_n is given by

$$\sum_{k=1}^{n-1} \mathbb{E}\big[(N_{k+1} - N_k)^2 \mid \mathcal{F}_k\big] \le e \sum_{k=1}^{n-1} \mathbb{1}\{m_{k-1} \ge z_0\} \le e \cdot M_n(z_0) \le n^{1/2+\epsilon},$$

where the last inequality holds WVHP by Lemma 3.4. Thus, using the same arguments as in the proof of Proposition 3.7 and by Theorem 3.10 we have that $|R_n - R'_n| \le n^{1/4+\epsilon}$ WVHP.

We turn to bound the second term in the right hand side of (19). We have

$$\begin{aligned}\big|e \cdot R'_n - M_{n-1}(z_0)\big| &\le \sum_{k=1}^{n} \mathbb{1}\{m_k \ge z_0\}\big(e(1 - m_k) - 1\big) \\ &\le M_n(z_0) n^{-1/4} + M_n\big(z_0 + n^{-1/4}\big) \le n^{1/4+2\epsilon},\end{aligned}$$

where in the last inequality we used Lemma 3.4.

The lest term in the right hand side of (19) is bounded by $n^{1/4+\epsilon}$ WVHP using the same arguments as in the first term. Indeed, $Z(z_0, n) - M_{n-1}(z_0)$ is clearly a martingale and its quadratic variation is bounded by $n^{1/4+\epsilon}$ WVHP. $\square$

We turn to prove the other results in Theorem 3. By (15) we have that

$$\frac{L_n}{\sqrt{n}} \xrightarrow{d} \sqrt{2}e^{-1}(M_1 - B_1), \qquad \frac{R_n}{\sqrt{n}} \xrightarrow{d} \sqrt{2}e^{-1}M_1 \tag{20}$$

and

$$\frac{L_n + R_n}{\sqrt{n}} \xrightarrow{d} \sqrt{2}e^{-1}(2M_1 - B_1).$$

It is well known (see [9]) that $M_1 - B_1 \overset{d}{=} M_1 \overset{d}{=} |N|$. Substituting these identities into (20) finishes the proof of (1). In order to prove (2) is suffices to compute the density of $2M_1 - B_1$. The following argument was given to us by Iosif Pinelis in a Mathoverflow answer [10]. By the reflection principle for all $a > b > 0$

$$\mathbb{P}\big(M_1 > a\ ,\ B_1 < b\big) = \mathbb{P}\big(B_1 \ge 2a - b\big) = 1 - F(2a - b). \tag{21}$$

Equation (21) gives the joint distribution of M_1 and B_1. From here it is straightforward to compute the joint density of (M_1, B_1) and the density of $2M_1 - B_1$. The details of the computation are omitted.

3.4 Proof of Theorem 2

Consider the martingale

$$N_n := X(z_0, n) + e\big|\{k \le n : x_k \ge z_0\}\big| - n.$$

Theorem 2 clearly follows from the following two lemmas.

Lemma 3.12. *We have that*

$$\Big\{\frac{N_{tn}}{\sqrt{n}}\Big\}_{t>0} \xrightarrow{d} \big\{\sqrt{e-3}\,B_t\big\}_{t>0}$$

Lemma 3.13. *We have WVHP*

$$\big|s(1,n) - n/e - N_n/e\big| \le n^{1/3+\epsilon}.$$

We start by proving Lemma 3.12.

Proof of Lemma 3.12. On the event $\{m_k \le z_0\}$ we have that

$$N_{k+1} - N_k = \mathbb{1}\{x_{k+1} \le z_0\}\frac{1}{1-x_{k+1}} - \mathbb{1}\{x_{k+1} \ge m_k\}\frac{1}{1-m_k} + e \cdot \mathbb{1}\{x_{k+1} \ge z_0\} - 1.$$

Thus, expanding the 10 terms in $(N_{k+1} - N_k)^2$ we get that on this event

$$\begin{aligned}
&\mathbb{E}\big[(N_{k+1} - N_k)^2 \mid \mathcal{F}_k\big] \\
&= \int_0^{z_0} \frac{dx}{(1-x)^2} - \frac{1}{1-m_k}\int_{m_k}^{z_0} \frac{2\,dx}{1-x} - \int_0^{z_0} \frac{2\,dx}{1-x} + \frac{1}{1-m_k} - \frac{2}{1-m_k} + 2 + e - 2 + 1 \\
&= \frac{1}{1-x}\Big|_0^{z_0} + \frac{2\log(1-x)}{1-m_k}\Big|_{m_k}^{z_0} + 2\log(1-x)\Big|_0^{z_0} - \frac{1}{1-m_k} + 1 + e \\
&= e - 1 - \frac{2}{1-m_k} - \frac{2\log(1-m_k)}{1-m_k} - 2 - \frac{1}{1-m_k} + 1 + e = \frac{-2\log(1-m_k) - 3}{1-m_k} + 2e - 2.
\end{aligned}$$

Thus, we can write $\mathbb{E}\big[(N_{k+1} - N_k)^2 \mid \mathcal{F}_k\big] = f(m_k)$ where f is given by the right hand side of the last equation when $m_k \le z_0$. Moreover, we have that

$$\int_0^{z_0} \frac{f(x)}{1-x}dx = \frac{-2\log(1-x) - 5}{1-x}\Big|_0^{z_0} - (2e-2)\log(1-x)\Big|_0^{z_0} = -3e + 5 + 2e - 2 = 3 - e.$$

Thus, by Corollary 3.9 we have with very high probability that

$$V_n := \sum_{k=1}^{n-1} \mathbb{E}\big[(N_{k+1} - N_k)^2 \mid \mathcal{F}_k\big] = \sum_{k=1}^{n-1} f(m_k) = (3-e)n + O(n^{5/6}).$$

This finishes the proof of the lemma using Theorem 3.8. □

We turn to prove Lemma 3.13.

Proof of Lemma 3.13. Define

$$E_n(z) := \big|\{k \le n : x_k \ge n\}\big|$$

and let $z_1 := z_0 + n^{-1/3}$. By Claim 3.1 the process

$$N'_k := X(z_1, k) + (\log(1-z_1)+1)k - (\log(1-z_1)+1)\cdot M_k(z_1) + eE_n(z_1) - e(1-z_1)n$$

is a martingale. It is straightforward to check that

$$e \cdot s(1,n) - n - N_n = A_1 + A_2 + A_3 + A_4 + A_5 + A_6$$

where

$$\begin{aligned}
&A_1 := e\big(s(1,n) - s(z_1,n) - E_n(z_1)\big), \quad A_2 := e\cdot s(z_1,n) - W(z_1,n),\\
&A_3 := -Z(z_1,n), \quad A_4 := N'_n - N_n, \quad A_5 := e(z_1-z_0)n - (\log(1-z_1)+1)n\\
&\text{and} \quad A_6 = (\log(1-z_1)+1)M_n(z_1).
\end{aligned}$$

Next, we bound each of the A_i WVHP. Any uniform point $x \ge z_1$ that arrived before time n such that $x \notin S(1,n)$ can be mapped to the time $k \le n$ in which it was removed. Thus, using also Lemma 3.4 we obtain that WVHP

$$0 \le E_n(z_1) - \big(s(1,n) - s(z_1,n)\big) \le M_n(z_1) \le n^{1/3+\epsilon}.$$

This shows that WVHP $|A_1| \le en^{1/3+\epsilon}$.

Using the same arguments as in (12) we get that $|A_2| \le n^{1/3+2\epsilon}$.

By Lemma 3.4 we have WVHP $|A_3| \le 3M_n(z_1) \le 3n^{1/3+\epsilon}$.

In order to bound A_4 we use the same arguments as in the proof of Proposition 3.7. Define the martingale $L_k := N'_k - N_k$. We have that

$$\big|L_{k+1} - L_k\big| \le Cn^{-1/3} + C\mathbb{1}\{x_{k+1} \in [z_0, z_1]\} + C\mathbb{1}\{m_k \ge z_0\}.$$

Thus, using Theorem 3.10 we obtain that WVHP $|N_n - N'_n| = |L_n| \le n^{1/3+3\epsilon}$.

It follows from a straightforward Taylor expansion that $|A_5| \le Cn^{1/3}$.

Lastly, by Lemma 3.4 we have WVHP $|A_6| \le Cn^{-1/3}M_n(z_1) \le Cn^{\epsilon}$.

$\square$

Acknowledgment: We thank Ehud Friedgut and Misha Tsodyks for helpful comments, we thank Ron Peled, Sahar Diskin and Jonathan Zhang for fruitful discussion and we thank Iosif Pinelis for proving in [10] that the density function of $2M_1 - B_1$ is given by $\frac{2x^2}{\sqrt{2\pi}}e^{-x^2/2}$.

References

[1] N. Alon, C. Defant and N. Kravitz, The runsort permuton, arXiv:2106.14762, 2021.

[2] N. Alon and J. H. Spencer, The Probabilistic Method, Fourth Edition, Wiley, 2016, xiv+375 pp.

[3] K. B. Athreya and P. E. Ney, Branching processes, Courier Corporation, 2004.

[4] B. Bercu and A. Touati, Exponential inequalities for self-normalized martingales with applications, The Annals of Applied Probability, 18(5), 1848–1869, 2008.

[5] R. Durrett, Probability: theory and examples, volume 49, Cambridge university press, 2019.

[6] D. A. Freedman, On tail probabilities for martingales, The Annals of Probability, 3(1), 100–118, 1975.

[7] E. Friedgut and G. Kozma, Private communication.

[8] A. Georgiou, M. Katkov and M. Tsodyks, Retroactive interference model of forgetting, Journal of Mathematical Neuroscience (2021) 11:4.

[9] P. Mörters and Y. Peres, Brownian motion, volume 30, Cambridge University Press, 2010.

[10] I. Pinelis, What is the distribution of $2M_1 - B_1$ where M_t is the maximum process of the Brownian motion B_t. MathOverflow question 409729.

[11] M. Tsodyks, Private Communication.