Almost sure convergence to the Quicksort process

Uwe Roesler
Mathematisches Seminar
Christian-Albrecht-Universität zu Kiel
Ludewig-Meyn-Strasse 4
24098 Kiel
Germany

February 3, 2016

Abstract

The algorithms Partial Quicksort, introduced by Conrado Martínez, sorts the l smallest reals within a set of n different ones. It uses a splitting like Quicksort, continuing always with the leftmost list. The normalized running time \( Y_n(t) \) converges with \( \frac{1}{n} \rightarrow t \) in distribution to a non degenerate limit. The finite dimensional distributions of the process \( Y_n \) converge to a limit \( [21] \), called the Quicksort process. In this paper we will present the algorithm Quicksort on the fly, a version of Partial Quicksort, showing the almost sure convergence of \( Y_n \) to the Quicksort process in the space of cadlag functions endowed with the Skorodhod topology.

1 Introduction

Quick sort [9] and Quickselect [8] are among the most thoroughly studied algorithms, [12, 17, 18, 20, 6, 7, 11, 10]. Both are divide-and-conquer algorithms, the input list is split by a pivot into two sublists and the algorithm recalled. Optimal versions choosing a pivot element are well known [15, 14].

Partial sorting asks for sorting only the l smallest reals out of n different reals. An obvious algorithm is to combine Quicksort and Quickselect, for example to use a Quickselect variant first in order to find the l smallest and then sort them by a Quicksort version. The algorithm partial_sort (see www.cplusplus.com), implemented in C++, does this in practice. Martínez and Roesler [14] gave optimal versions for given \((n, l)\) in the sense of an asymptotic smallest expectation of comparisons.

The algorithm Partial Quicksort for partial sorting was introduced by Martínez [13]. Let the input \( S \) be a list of different reals. Choose with a uniform distribution an element of \( S \), called pivot, and split \( S \) into a list \( S_\prec \) of strictly smaller, the pivot and \( S_\succ \) of strictly larger elements. Arrange the sublists in the order
Let \( X(S, l) \) be the number of comparisons in order to find the sorted 0 ≤ \( l \leq |S| \) smallest elements in \( S \). The rvs satisfy the recursions

\[
(X(S, l))_{|S|=1} = (|S| - 1 + X(S_\leq, l \cap |S_\leq|) + X(S_\geq, 0 \lor (l - |S_\leq| - 1)))_{|S|=1} \tag{1}
\]

for \( |S| \geq 2 \). We use \( X(S, 0) = 0 \) and \( X(S, l) = 0 \) for \( |S| \leq 1 \). It is tacitly assumed that for every \( S \) our choice of the pivot is independent of everything else. The distribution of \( (X(S, l))_l \) depends only on \( |S| = n \) which follows easily by induction on \( n = |S| \) using (1). It is true due to the internal randomness (within the algorithm) and the uniform distribution on \( S \). We will use a rv \( (X(|S|, l))_l \) with that distribution.

The equation (1) determines recursively the distribution of \( X(n, \cdot) \)

\[
(X(n, l))_l \overset{D}{=} (n - 1 + X_1(I, l \land (I - 1)) + X_2(n - I, 0 \lor (I - l)))_l \tag{2}
\]

\( n \geq 2 \). For given \( n \) the rvs \( I, X_i(j, \cdot), \ i = 1, 2, \ j < n \) are independent. The rv \( I = I_n \) is uniformly distributed on \( \{1, 2, \ldots, n\} \) and \( X_i(j, \cdot) \) has the same distribution as \( X(j, \cdot) \). We use the same starting values \( X(n, 0) = X(1, 1) \) as above. The rv \( I = |S_\leq| + 1 \) has the interpretation of the rank of the pivot within \( S \).

The expectation \( a(n, l) = EX(n, l) \) satisfies the recursion

\[
a(n, l) = n - 1 + \frac{1}{n} \sum_{j=1}^{n} (a(j - 1, l \land (j - 1)) + a(n - j, 0 \lor (l - j))) \tag{3}
\]

for \( n \geq 2, 1 \leq l \leq n \). The solution is explicitly known [13]

\[
a(n, l) = 2n + 2(n + 1)H_n - 2(n + 3 - l)H_{n + 1 - l} - 6l + 6 \tag{4}
\]

for \( n \geq 1, 1 \leq l \leq n \) and the initial conditions \( a(\cdot, 0) = 0 \). \( H_n \) denotes the \( n \)-th harmonic number \( \sum_{i=1}^{n} \frac{1}{i} = \ln n + \gamma + o(1) \), \( \gamma \) the Euler constant. The term \( a(n, n) = -4n + 2(n + 1)H_n \) is the expectation of the comparisons sorting \( n \) numbers by Quicksort [12].

The distributions of

\[
(Y(n, \frac{l}{n}))_l := (\frac{X(n, l) - a(n, l)}{n + 1})_l \tag{5}
\]

satisfy the recursions

\[
(Y(n, \frac{l}{n}))_l \overset{D}{=} (\frac{l}{n + 1} Y_1(I, 1, 1 \land \frac{l}{I - 1}) + \frac{n - I + 1}{n + 1} Y_2(n - I, 0 \lor \frac{l - I}{n - I}) + \frac{C(n, I, l)}{n + 1})_l \tag{6}
\]

for \( n \geq 2 \) where \( I = I_n \) is as above and

\[
C(n, i, l) = n - 1 - a(n, l) + a(i - 1, l \land (i - 1)) + a(n - i, 0 \lor (l - i)) \tag{7}
\]
The distribution of $Y(j,1)$ is called the discrete Quicksort distribution to sort $j$ elements. (The normalization by $n + 1$ is chosen in view of some martingale property [16].)

In a joint paper Martínez and Roesler [14] showed the weak convergence of the distribution of $Y_{n,t_n}$ as $n \to \infty, t_n \to_t t$ to a probability measure $\mu_t$ for all $t$ in the unit interval $(0,1)$. The map $[0,1] \ni t \mapsto \mu_t$ is continuous with respect to weak convergence (and Wasserstein $l_p$-distance). They replaced in (6) the term $Y_1(.,1)$ by an independent rv with the same distribution. That defines a process $Y(n,.)$ with the same one dimensional distributions as $Y(n,.)$. The proof then makes use of an embedding of $Y$ into a weighted branching process and extending $(\tilde{Y}(n,\frac{t}{n}))_{i}$ pathwise to a process with values in the set $D$ of cadlag functions on $[0,1]$. Subsequent papers use the embedding approach, but for $Y$ itself.

Ragab and Roesler [21] considered $l$ as a time index and process convergence after normalizing. Extend $Y(n,.)$ to a cadlag process on $D([0,1])$ left continuous at 1. (Actually they used a right continuous version with existing left limits, which does not matter here since $P(Y(t-) = Y(t+)) = 1$ for all $t$.) The one dimensional distributions converge by the above. They showed $Y(n,.)$ converges in $n$ to some limiting process $Y$ for all finite dimensional distributions. The $D$-valued process $Y$ is called the Quicksort process and is characterized in distribution by the stochastic fixed point equation in $D$

$$(Y(t))_{t \in [0,1]} \overset{D}{=} (UY_1(1 \wedge \frac{t}{U}) + \mathbb{1}_{t\geq U}(1-U)Y_2(\frac{t-U}{1-U})) + C(t,U))_{t \in [0,1]} \quad (8)$$

Here we consider only solutions with a uniformly bounded expectation and variance of all $Y(t)$. Further we require continuity at $t = 1$ a.e.. Here $Y_1, Y_2, U$ are independent, $Y_1, Y_2$ have the same distribution as $Y$ and $U$ has a uniform distribution. The function $C : [0,1] \times [0,1] \to \mathbb{R}$ is given by

$$C(t,s) := \mathbb{1}_{s \geq t} C(s) + \mathbb{1}_{s < t} (-1 + 2s + 2(1-t) \ln(1-t)$$
$$-2(s-t) \ln(s-t) + 2s \ln s) \quad (9)$$

$$C(s) := 1 + 2s \ln s + 2(1-s) \ln(1-s) \quad (10)$$

using continuous extension.

Their approach can be described in 2 steps. First consider the fixed point equation for $t = 1$, which is the Quicksort fixed point equation. Embed the rvs $Y(n,.)$ into a weighted branching process obtaining a recursion of rvs. Second embed the $Y(n,.)$ processes appropriate into a weighted branching process, replacing the trouble making term $Y_1(I - 1,1)$ in equation (5) by a specific rv using the first step. (Details in the section on the discrete Quicksort process.) This provides a nice affine recursion and a representation of $Y(n,.)$ as an uniformly absolute convergent countable sum where all summands converge nicely in $n$.

By the uniformity argument in time $t \in [0,1]$ they obtained convergence of processes in some normed space, which implies convergence of finite dimensional distributions. There exists nice versions of $Y(n,.), Y$ such that $Y(n,.)$ converges uniformly to $Y$ in $L^p(P \times \lambda)$-sense, $p > 1$, more specific $E(\int |Y_n(t) - Y(t)|^p) \to 0$. That is the content of this paper.
Another consequence of their uniformity argument is the existence of versions of the processes \( Y(n, \cdot) \) and \( Y \) on \( D \) such that \( Y(n, \cdot) \) converges to \( Y \) in Skorokhod metric almost surely. The aim of this paper is to present such a version arising from an algorithm.

We consider the deterministic algorithm Quicksort on the fly with random input list \( U_1, U_2, \ldots, U_n \). Quicksort on the fly, the version of Martínez [13] Partial Quicksort with external randomness, works as follows:

- choose always the first element in the list as pivot
- compare any other element sequentially with the pivot
- form the list of strictly smaller, the list containing only the pivot and the list of strictly larger reals in this order, preserving the sequential order of comparisons within a list
- continue always with the leftmost list containing at least 2 elements

This algorithm terminates with the sorted input as output.

If we stop the algorithm Quicksort on the fly when the \( l \)-th smallest element is identified then we have performed (a version of) the algorithm Partial Quicksort to \((n, l)\) [13], find the ordered \( l \) smallest numbers out of \( n \). And we did only the necessary comparisons so far, no more.

Let \( X(u, l) \) denote the number of comparisons necessary to find and sort the \( 0 \leq l \leq n \in \mathbb{N}_0 \) smallest numbers out of the finite input \( u \in (0, 1)^*_\mathbb{N} \). The Quicksort on the fly recursion is,

\[
X(u, l) = X(u, l \wedge |u^1|) + X(u^2, 0 \lor (l - |u^1| - 1)) + n - 1 \quad (11)
\]

for \( n \geq 2 \) according to the above splitting procedure. Here \( u^1, u^2 \) denote the lists of all \( u_i < u_1 \) respectively \( u_i > u_1 \) for \( i \leq n \).

Let \( U = (U_i)_{i \in \mathbb{N}} \) be a vector of iid random variables with a uniform distribution on the open unit interval and let \( U_{|n} := (U_1, U_2, \ldots, U_n) \). The expectation \( E(X(U_{|n}, l)) \) satisfies the same recursion (3) as the \( a(n, l) \). Since the initial values are the same we obtain

\[
a(n, l) = E(X(U_{|n}, l))
\]

Define

\[
Y(u, \frac{l}{n}) = \frac{X(u, l) - a(|u|, l)}{|u| + 1}
\]

as before. Now extend \( Y \).

We prefer to work on the space \( D = D([0, 1]) \) of left continuous functions with existing right limit instead of the common space \( D([0, 1]) \) of cadlag function [3]. The existing literature prefers more cadlag version for historic reasons and its importance as a standard setting for time continuous Markov processes. However the space of left continuous functions with existing right limits, as used before in [21], would be more natural in view of the recursive combinatorial structure and also the distributional fixed point equations. It provides a more elegant normalization and easier notation of the equations. Both views are equivalent, since for our processes \( Z \) we have \( P(Z(t-) = Z(t+) = 1 \) for all \( t \). (The points 1 for cadlag functions and 0 for caglad function play a special role.)
The theoretic foundation [3] for both is the same. A quick argument is the time
reversion \( t \to 1 - t \).

Extend \( Y(u, \cdot) \) to a left continuous step function in \( D \)

\[ Y(u, t) = \frac{X(u, l) - a(|u|, l)}{|u| + 1} \]

(13)

\( t \in (0, 1], \quad |u| \geq 2 \) where \( 1 \leq l \leq |u| \) is determined by \( l \leq |u| t < l \).

Here is the major result.

**Theorem 1** Let \( U_i, \quad i \in \mathbb{N}, \) be iid rvs with a uniform distribution. Then the
\( D \)-valued process \( Y_n = Y(U_n, \cdot) \) as defined above for deterministic Quicksort on
the fly with random input \( U_n \) converges path wise almost surely to the Quicksort
process \( R_V(\infty) \) as given in section (5) in the Skorohod \( J_1 \)-topology.

Equivalent is the statement for a right continuous extension of the process
\( Y(U_n, \cdot) \) taking special care of the point \( t = 0 \) and \( t = 1 \).

Notice our version of \( Y(U_n, \cdot) \) arises from a deterministic algorithm with ran-
dom input. Quicksort itself is often programmed as a deterministic algorithm,
where the randomness comes exclusively from the input. For the analysis of
the asymptotic distributional behavior of Quicksort [17] it is somewhat easier
to work with a random version of the algorithm Quicksort, picking the pivot
element always independently random. The result for the random version is the
same asymptotic distributional behavior of the running time for every input.
This is of course not true for the deterministic version.

The main result was obtained by a rigorous embedding of the processes
into weighted branching processes (WBP) [19][20], see following sections for
explanation and notation. The existence of the Quicksort process itself is shown
as the limit of the sum

\[ R_W = \sum_{v \in W} L_v \ast_v C_v \quad (14) \]

\( W \subset V = \{1, 2\}^* \) for \( W \to V \) in the norm \( \|\| \cdot \|_{\infty} \|_p \) for \( p > 1 \) on the space
\( D \). The limit exists since \( R_{V,m} \) is a Cauchy sequence and the space \( D \) with the
norm \( \|\| \cdot \|_{\infty} \|_p \) for \( p > 1 \) is complete.

Using our specific deterministic limiting Quicksort version with random in-
put the process \( Y(U_n, \cdot) \) has a natural embedding into a WBP and the corre-
sponding \( R_W(n) \) as above indexed by the length \( n \) of the input \( U_n \) will converge
in \( D \) to \( R_V \). We use a random time shift \( \tau_n \) such that the shifted process \( Y_n(\tau_n) \)
has only jumps at \( U_1, U_2, \ldots, U_n \), the same as the Quicksort process \( Y \). Then
the shifted \( Y_n(\tau_n) \) is close to \( Y \) in the supremum metric pathwise almost surely.
(The supremum metric on \( D \) is complete.) It remains to show the time shift is
small in the sense \( \sup_l |\tau_n(l) - t| \to_n 0 \) almost surely. This amounts to the well
known fact, that the empirical distribution function of the \( U_i, \quad i \leq n \) converges
a.s. uniformly to the continuous distribution function of \( U_1 \).
2 Definitions and setting

**Binary Ulam-Harris trees:** Let \( V = \{1, 2\}^* = \cup_{n=0}^{\infty} \{1, 2\}^n \) be the set of all finite \( 1 \rightarrow 2 \) sequences, the empty sequence denoted by \( \emptyset \). We consider \( V \) as a rooted tree, a graph with (directed) edges \((v, vi), v \in V, i \in \{1, 2\}\) and root \( \emptyset \). The vertex \( v = (v_1, v_2, \ldots, v_n) = v_1 v_2 \ldots v_n \) has length \( |v| = n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\} \). We use \( V_n, V_{<n}, \ldots \) in the natural sense for all vertices with length \( n \), or strictly smaller than \( n \) and so on.

The genealogical order \( \preceq_g \) on the infinite binary tree \( V \) is defined by \( v \preceq_g w \) if and only if \( v \) is a prefix of \( w \), \( v = (v_1, v_2, \ldots, v_m) \preceq_g w = (w_1, w_2, \ldots, w_n) \Leftrightarrow m \leq n \) and \( \forall i \leq m : v_i = w_i \).

We call \( v \) an ancestor of \( w \) iff there exists a path from \( v \) to \( w \neq v \).

The traverse order \( \preceq_t \) on the infinite binary tree \( V \) is defined by traversing from left to the right. Formally define \( h(v) = \sum_{|v|} 3^{-i} 2(v_i - 1) \) and put \( v \preceq_t w \Leftrightarrow h(v) \leq h(w) \).

Notice the traverse order is a total order on \( V \).

The shift operator on functions \( f \) with domain \( V \) is defined by \( (S_v(f))(w) = f(vw) \). We use throughout the notation \( f_v = S_v(f) \) with an upper index. In analogy for any subset \( W \) of \( V \) we use \( W^v := S_v(\mathbb{1}_W) = \{w \in V \mid vw \in W\} \).

A binary Ulam-Harris subtree is a subset \( W \) of the infinite binary tree \( V \) including the induced graph structure and closed below under the genealogical order \( w \in W \Rightarrow \forall v \preceq_g w : v \in W \). We call these objects just trees. The smallest tree is the empty set. Let \( W \) be the set of all trees and \( W_f \) be the subset of finite trees. On trees we use the induced genealogical and traverse order.

For a finite tree \( W \in W_f \) and \( 0 \leq l \leq |W| \) define

\[
X(W, l) := \sum_{\{w \in W \mid w \preceq_g v \text{ or } w \preceq_t v\}} (|W^w| - 1) \tag{15}
\]

with \( v \) the \( l \)-th smallest element of \( W \) relative to the traverse order. By (15) the function \( X \) satisfies the recurrences,

\[
X(W, l) = X(W^1, l \wedge |W^1|) + X(W^2, 0 \lor (l - |W^1| - 1)) + |W| - 1 \tag{16}
\]

for finite non empty trees \( W \) and \( 1 \leq l \leq |W| \). The symbol \( \wedge \) denotes the infimum. (The proof is omitted due to its simplicity.) Notice \( X(W, 0) = 0 \).

For a finite non empty tree \( W \) and \( 0 \leq l \leq |W| \) define

\[
Y \left(W, \frac{l}{|W|}\right) := \frac{X(W, l) - a(|W|, l)}{|W| + 1} \tag{17}
\]

The \( a(n, l) \) are explicitly given in [14] (3).
The recurrences (16) imply for $|W| \geq 2$, $1 \leq l \leq |W|

\begin{align*}
Y(W, \frac{l}{|W|}) &= \frac{|W^1| + 1}{|W| + 1} Y(W^1, 1 \wedge \frac{l}{|W^1|}) \\
&+ \frac{|W^2|}{|W| + 1} Y(W^2, 0 \vee \frac{l - |W^1| - 1}{|W^2|}) \\
&+ C(|W|, |W^1| + 1, l) \frac{1}{|W| + 1}
\end{align*}
(18)

where $C$ is defined in (7). The boundary conditions are $Y(W, 0) = 0$ for all $W$ and additionally $Y(W, 1) = 0$ for $|W| = 1$. Our interest centers on the $D$-valued objects $Y(W, l) = Y(W, \frac{|W| |l|}{|W|})$ for random input of search trees.

**Search trees:** A binary search tree is a tuple $(W, \psi)$ consisting of a tree $W$ and an injective map $\psi : W \rightarrow R$ which is strictly isotone (= strictly order preserving) with respect to the traverse order.

The set of all binary search trees is an ordered set by inclusion, $(W, \psi) \leq (W', \psi')$ if $W \subset W'$ and $\psi'$ restricted to $W$ is $\psi = \psi'_{|W}$. Any increasing sequence $(W_n, \psi_n)$ of binary search trees forms a projective system and has a projective limit $(\cup_n W_n, \psi = \lim_n \psi_n)$.

**Vector input:** Let $u = (u_i)_i \in R^*_+ \cup \mathbb{R}^N$ be a finite or infinite sequence of different reals. We use $u_{|n} = (u_1, u_2, \ldots, u_n)$ and the length $|u|$ as number of coordinates. Define $u^1 = (u_{i_1}, u_{i_2}, \ldots)$ where $i_1 < i_2 < \ldots$ are the ordered elements of $\{2 \leq i \leq |u| \mid u_i < u_1\}$. Analogous define $u^2 = (u_{i_1}, u_{i_2}, \ldots)$ where $i_1 < i_2 < \ldots$ are the ordered elements of $\{2 \leq i \leq |u| \mid u_i > u_1\}$. The vector $u^v, v \in V$ is defined by $u^0 = u$ and recursively

\[ u^{vi} = (u^v)^i. \]

**Proposition 2** There is an injective map between finite or infinite sequences $u$ of different reals and and binary search trees $(W, \psi)$. There is a one-to-one map between $u$ and increasing sequences $(W_{u,n}, \psi_{u,n})$, $N_0 \geq n \leq |u|$, of binary search trees, such that $\psi_{u,n}(W_{u,n}) = \{u_1, u_2, \ldots, u_n\}$.

Proof: In algorithmic wording, start with $u$ at the root. Then put $u_1$ to the root and $u^1$ to the vertex 1 and $u^2$ to the vertex 2. Recurse recursively the procedure now for vertex $v$ and $u^v$.

Alternatively, put $u_1$ to the root and take $W_{u_1}$ consisting of the root and $\psi_{u_1}(0) = u_1$. Next put $u_2$ either to the vertex 1 in case $u_2 < u_1$ and otherwise to the vertex 2. We obtain a search tree containing the first one and having one more vertex. By recursion, compare the next $u$ coordinate value with the value at the root, go left or right, continue going left or right until a free vertex is found. In that case put the coordinate value down to that vertex. Then $u^v$ are the coordinates values of $u$ which reached or passed the vertex $v$ ordered in time of appearance.

7
This way we obtain a one-to-one map from real valued sequences to increasing sequences of search trees.

The proposition remains true, if we take sequence with values in a subset $A$ of the reals, instead of the reals. We use the notation $A^a_n = \bigcup_{n \in \mathbb{N}_0} A^n$ where $A^n = \{ u \in A^a | u_i \neq u_j \text{ for } i \neq j \}$.

We use the notation $(W_u, \psi_u)$ for the binary search tree generated by $u$. Traversing the finite search tree $(W_u, \psi_u)$ from left to right provides the elements of $\psi_u(W_u) = \{ u_1, u_2, \ldots, u_n \}$ in natural order. The number of comparisons in order to find the $l$ smallest coordinate of finite $u$ by constructing the binary search trees only as far as necessary is $X(W_u, l)$.

For given finite input $u$ we will use freely

$$X(u, \cdot) := X(W_u, \cdot) \quad Y(u, \cdot) := Y(W_u, \cdot)$$

where $u$ generates the tree $W_u$ as search tree. (For this part $\psi$ is not necessary.) Notice that this implies for $v \in \mathbb{V}$

$$X(u^v, \cdot) = X(W^v_u, \cdot) \quad Y(u^v, \cdot) = Y(W^v_u, \cdot)$$

$W^v_u := (W_u)^v = W_u^v$. We neglect the symbol $\emptyset$ whenever possible. Our interest centers on these objects for random input $u$.

Random Input: Let $U_i$, $i \in \mathbb{N}$, be independent rvs with a uniform distribution on $(0, 1)$ on some abstract probability space $(\Omega, \mathcal{A}, P)$ sufficiently rich. Without loss of generality we may assume (and do so for notational reasons) that for every fixed $\omega$ in the realization space $\Omega$ all values $U_n(\omega), n \in \mathbb{N}$, are different. We will use the same notation as above replacing the input $u$ by $U_n(\omega)$ for $n \in \mathbb{N}_0$ and $\omega \in \Omega$. However we suppress the $\omega$ whenever possible, as is common in probability theory. The following properties are elementary and found in standard text books.

Proposition 3 Let $U = (U_n)_{n \in \mathbb{N}}$ be a sequence of iid rvs with a uniform distribution on $(0, 1)$.

- The rvs $\frac{U^{(1)}_i}{U^{(1)}_n}$, $i \geq 2$ are iid with a uniform distribution on $(0, 1)$.
- The rvs $\frac{U^{(j)}_i}{U^{(j)}_n}$, $j \geq 2$ are iid with a uniform distribution on $(0, 1)$.
- The vectors $U^{(1)}_1, U^{(2)}_1$ conditioned to the cardinality $U^{(1)}_n$ consists both of iid rvs for every $n \in \mathbb{N}$.
- $\frac{U^{(1)}_i}{U^{(1)}_n}, \frac{U^{(2)}_i-U^{(1)}_i}{U^{(1)}_n}, (R_{1:n})_n$ are independent, where $R_{1:n}$ is the rank of $U_1$ under $U_1, U_2, \ldots, U_n$.
- $E(U^{(1)}_1(1-U_1)^{j}) = \frac{k^{(j)}_1}{(k+j+1)}$ using $k, j \in \mathbb{N}_0$ for the power function and not as an index.

8
Proposition 4 Let $U_n$, $n \in \mathbb{N}$, be independent rvs with a uniform distribution, $U = (U_n)_n$. The distribution of $X(U_{[n]} \cdot)$ depends only on $n$. The expectation of $X(U_{[n]}, l)$ is $a(n, l)$ as given in (3) for $0 \leq l \leq n$.

Proof: The statement is true for $n = 0$ and $n = 1$. We use now induction from $n - 1 \rightarrow n$. Let $\mu_m$ be the distribution of $X(U_{[m]} \cdot)$ for $m \leq n - 1$. Since $n \geq 2$ we can use the recurrence (16). If we condition on $|U_{[n]}^1| = k$ it suffices to show the distribution of

$$
(X(U_{[n]}, l))_{l=1}^n = (X(U_{[n]}^1, l \land k) + X(U_{[n]}^2, 0 \lor l - k - 1))_{l=1}^n
$$

(20)

depends not on the actual realization $U_{[n]}^1, U_{[n]}^2$.

Under $k$ the vector $\frac{U_{[n]}^1}{E}$ consists of $k$ independent rvs with a uniform distribution and the vector $\frac{U_{[n]}^2 - U_{[n]}^1}{E}$ of $n - k - 1$ independent rvs with a uniform distribution. As a consequence $X(U_{[n]}^1 \cdot)$ and $X(U_{[n]}^2 \cdot)$ are independent and have the distribution $\mu_k, \mu_{n-k-1}$ by induction hypothesis. Therefore the distribution of (20) is determined by the probability measures $\mu_k$ and $\mu_{n-k-1}$. This proves the induction step.

The expectations $E(X(U_{[n]}, l)) = b(n, l)$ satisfy the recurrences

$$
b(n, l) = \sum_{k=0}^{n-1} P(|U_{[n]}^1| = k) (E(X(U_{[n]}^1, l \land k) \mid |U_{[n]}^1| = k)) + \mathbb{I}_{l > k+1} E(E(X(U_{[n]}^2, l - k - 1) \mid |U_{[n]}^1| = k)) + n - 1
$$

(21)

for $n \geq 2$. These are the same recurrences as in Martínez-Roesler [14] for $a(n, l)$ with the same starting values $a(1, 1) = 0 = a(n, 0), n \in \mathbb{N}_0$. The explicit solution is given in (3). The uniqueness of the solution for given starting values implies $b = a$.

q.e.d.

3 The weighted branching process

For the analysis of $Y$ (17) and the Quicksort process we need the weighted branching process (WBP) [17, 18, 20, 19]. The WBP is similar to a tree-indexed stochastic process like a discrete, branching dynamical system or iterated function system. The paths of the tree obtain some weights, represented by transformations forward or backward.

Let $V$ be the set $\mathbb{N}^*$ of finite sequences of natural number. $V$ is a directed graph with edges $(v, w), v \in V, i \in \mathbb{N}$. A WBP is a tuple $(V, L, G, \ast)$. Here $L = (L_{v, v'w})_{v, v'w \in V}$ is a stochastic process, where $L_{v, v'w}$ has values in $G$. $(G, \ast)$ is a measurable semi group with identity and a grave. $L$ preserves the structure of path concatenation for the directed graph and the composition in $G$ via

$$
L_{v, v'w} = L_{v, v'} \ast L_{v', v'w}
$$
Further $(L_{v,vi})_{i \in \mathbb{N}}$, $v \in V$ are iid rvs. Notice the independence of families, but we do not require independence within a family.

The $L_{v,vi}$, $v \in V$, $i \in \mathbb{N}$ determine completely $L_{v,vw}$ for any path. If the rvs are degenerate then we face a discrete dynamical system indexed by a tree structure. We use $L_{\emptyset,v} = L_v$.

Without loss of generality, we take $G$ as a subset of $H^H$ for some measurable space $H$. $H = G$ will always do it, but other spaces may be more suitable. Formally $G$ acts right on $H$ via $* : H \times G \to H$ or acts left via $* : G \times H \to H$. In the first case we have $h \ast g \ast g' = g'(g(h))$ and in the second case $g \ast g' \ast h = g(g'(h))$.

In many cases and without loss of generality we assume $H$ has a grave called $\triangle$ (once in a grave you stay in the grave). In this manner we include branching processes [1] considering only vertices $v$ such that $L_v(h) \neq \triangle$ for starting with $h$ at the root.

For simplicity and a special application in mind we restrict us to $V = \{1, 2\}^*$, the set of finite binary sequences. Here are some examples and objects used in the sequel for analyzing Quicksort with the forward dynamic ($\ast G$ acts right).

The input for Quicksort is an element $x \in H_1 := (0,1)^*_x$, a finite sequence of different reals in the open unit interval. Define $H_2 := \{(c,d) \mid 0 \leq c \leq d \leq 1\}$ and $H_3 = \{x = (x_v)_{v \in W} \mid x_v \in (0,1), \text{ } W \in W_f\}$. If necessary extend $u$ by $u_v = \triangle$ for $v \notin W$.

**Sequences:** Take $H = H_1$ or the set of infinite sequences $(0,1)^\mathbb{N}$ and let $G$ be generated by the $g_i(u) = u^i$ (and the identity and a grave) acting right on $H$. Putting $L_{v,vi} = g_i$ we obtain $L_{v,vw}(u) := u^w$. (See example vector input.)

The binary search tree $(W_u, \psi_u)$ is recovered from $u^w$, $v \in V$, as the set of all $w$ such that $u^w$ is not empty and $\psi_u(w)$ is the first element in $u^w$.

**Trees:** Let $H = W_f$ or $W$ and let $G$ be generated by the $g_i(W) = W^i$ acting right on $H$. Putting $L_{v,vi} = g_i$ for $v \in V$ we obtain $L_{v,vw}(W) = W^w$. For $u \in H_1$ holds $L_{v,vw}(W_u) = W^w_u$ where $(W_u, \psi_u)$ is the binary search tree for $u$.

**Intervals:** Let $H = H_1 \times H_2$ with the grave $\triangle = (\emptyset, \emptyset)$. Let $G$ be generated by the $g_i : H \to H$ by

$$
g_1(u,(c,d)) = (u^1,(c,d + (d-c)u_1]) \quad g_2(u,(c,d)) = (u^2,((d-c)u_1,d])
$$

for $u \neq \emptyset, c \neq d$ and otherwise the grave $\triangle$. Put $L_{v,vi} = g_i$ for all $v \in V$. We will use in the sequel the notation $I_u(u,I) := (L_v(u,I))_2$ and $J_v(u) = I_v(u,(0,1]) = (c_v(u),d_v(u))$.

Notice for $\emptyset \neq u \in H_1$ both maps $c_v(u), d_v(u)$ on outer leaves $v$ of $W_u (v \notin W_u$ and $\forall w \sim_0 v : w \in W_u)$ are strictly isotone in traverse order. If we traverse the outer leaves of $W_u$ observing the $c$-values ($d$) then we obtain the coordinates of $u$ in increasing order and the additional values $0$ for the $c$-sequence respectively $1$ for the $d$-sequence. The intervals $J_v(u)$ for outer leaves $v$ of $W_u$ form a partition of $(0,1]$.

**Interval length:** The interval length of the interval $I_u(u,I)$ from above is an example of a WBP.
Intervals Revisited: Let $\tilde{H} = H_3 \times H_2$ extended by the grave. $\tilde{G}$ is generated by

$$\tilde{g}_1(x,(c,d]) = (S_1(x),(c,c+(d-c)x_{\tilde{g}}]) \quad \tilde{g}_2(x,(c,d]) = (S_2(x),(c+(d-c)x_{\tilde{g}},d])$$

whenever $x$ has non empty domain and $\Delta$ otherwise. $S_v$ denotes the shift by $v$. Put $\tilde{L}_{v,vi} = \tilde{g}_i$.

There is a connection between the last two examples. Define the map $H_1 \ni u \mapsto x(u) = (x_v(u))_{v \in W_v} \in H_3$ by $x_v(u) := \frac{|J_v(u)|}{|J_v(u)|}$ if $J_v(u)$ is non empty. Then we claim $x(u)$ is well defined and $J_v(u) = \tilde{J}_v(x(u))$ whenever non empty, where the right side is defined by $\tilde{L}_v(x,(0,1]) =: (S_v(x),\tilde{J}_v(x))$. The proof is easy by induction on the length of $u$ and therefore skipped.

In the next two examples $G$ acts left on $H$.

**Sum:** Let $H = \mathbb{R}^{H_3}$ and $G \subset H^H$ be generated by $g_1,g_2$

$$(g_1(h))(x) := u_1h(S_1(x)) \quad (g_2(h))(x) := (1-x_1)h(S_2(x))$$

The operations are $g * g' * r_* h = g(g'(h))$. Put $L_v, vi = g_i$. For given $C_v \in H$, $v \in V$ with $C_v(\Delta) = 0$ consider the map $R_{v,vw} \in H$

$$R_{v,vw}(x) := (L_v,vw * r_* C_{vw})(x) = |\tilde{J}_w(x)|C_{vw}(S_w(x))$$

the $\tilde{J}$ as above. Then $R_{v,vW} := \sum_{w \in W} R_{v,vw}$ converges pointwise to $R_{v,vV}$ as $W \to V$ and the limit satisfies

$$R_{v,vV} = \sum_{i=1}^2 L_i * r_* R_{v,viV} + C_v$$

for all $v \in V$.

Now an example for a non deterministic WBP.

**Quicksort [17]:** Let $H = G$ be the unit interval $[0,1]$ with multiplication and $U_v, v \in V$ be iid rvs with a uniform distribution on $(0,1)$. Define

$$L_{v,v1} = U_v \quad L_{v,v2} = (1-U_v)$$

The 0 plays the role of the grave. Let $C : [0,1] \to \mathbb{R}$ be the continuously extended map

$$C(x) := 1 + 2x\ln x + 2(1-x)\ln(1-x) \quad (22)$$

and put $C_v = C(U_v)$. Then

$$R_{v,vW} := \sum_{w \in W} L_{v,vw}C_{vw}$$

converges as $W \to V$ almost surely and in any $L^p$, $p > 1$ to a limit called $Q^v$ [20]. $Q$ has the Quicksort distribution and satisfies, with obvious notation,

$$Q^v = U_vQ^v + (1-U_v)Q^{v2} + C(U_v)$$

11
almost surely simultaneously for all $v \in V$.

For later reference define a map $\varphi$ from finite binary search trees $(W, \psi)$ with values of $\psi$ in $(0, 1)$ to maps $(0, 1)^{|W|} \in H_3$

$$
\varphi(W, \psi)(v) = \frac{\psi(v) - c_v}{d_v - c_v}
$$

for $v \in W$, where

$$
c_v := \sup \{0, \psi(w) \mid W \ni w \prec v\} \quad d_v = \inf \{1, \psi(w) \mid v \prec w \in W\}
$$

The interval $(c_v, d_v]$ is $J_v(x)$ as above. The map $u \mapsto x(u)$ from example 'Intervals Revisited' satisfies $x(u) = \varphi(W_u, \psi_u)$.

Extend the map $\varphi$ to infinite binary search trees $(W, \psi)$ via

$$
\varphi(W, \psi)(v) = \lim_{n \to \infty} \varphi(W_n, \psi_u)(v)
$$

for any increasing sequence $(W_n, \psi_n)$ of finite binary search trees to $(W, \psi)$.

**Proposition 5** The map $\varphi$ on finite search trees $(W, \psi)$ with values of $\psi$ in $(0, 1)$ to $(0, 1)^{|W|} \in H_3$ is well defined and is a bijection. The extension to infinite search trees is well defined. It holds $x(u) = \varphi(W_u, \psi_u))$ for finite or infinite sequence $u$ of $(0, 1)$ values.

The proof is standard and omitted.

For input $u$ we obtain via $(W_u, \psi_u)$ an $x(u) = \varphi(W_u, \psi_u)$. For given $x \in (0, 1)^{|W|}$ we can only recover the underlying binary search tree $(W, \psi)$. We obtain the set of values of the coordinates of $u$ but not the vector $u$ itself and not the sequence $(W_{u_n}, \psi_{u_n})_n$, $n \leq |u|$.

**Proposition 6** If $U$ is a finite sequence of uniformly distributed rvs, then for every tree $W$ the rvs $\varphi(W_U, \psi_U)(v)$, $v \in W$, conditioned on $W \subseteq W_U$, are independent and have a uniform distribution.

If the input is an infinite sequence $U$ of independent rvs uniformly distributed on the unit interval $(0, 1)$ then $\varphi(W_U, \psi_U)(v)|_{v \in V}$ are independent rvs with a uniform distribution. For an infinite sequence

$$
\lim_{n \to \infty} \frac{|W_{U_n}^1|}{|W_{U_n}^2|} = \varphi(W_U, \psi_U)(v) \quad \lim_{n \to \infty} \frac{|W_{U_n}^2|}{|W_{U_n}^3|} = 1 - \varphi(W_U, \psi_U)(v)
$$

almost surely.

Proof: The distributional statement is shown by induction on the length of the input $U$ using Proposition (3). For an infinite sequence $U$ we have $W_U = V$ and the strong law of large numbers implies the last statement. q.e.d.

With our problems in mind we are only interested in functions of $(W_U, \psi_U)$ for random input $U$ of iid rvs with a uniform distribution. Sometimes it seems easier to work with the corresponding random variable $\varphi(W_U, \psi_U)(v)$ as input instead of $U$. We use this in the next section.
4 The Quicksort process

We present the limiting Quicksort process [21] as another example of a WBP. The proof given here follows basically the original one [21] and is essential for the sequel. We take the Quicksort process for a random algorithm and deterministic input.

Let \( D = D([0,1]) \) [3] be the set of all caglad (continue à gauche, limite à droite) functions \( f : [0,1] \to \mathbb{R} \) (left continuous functions with existing left limits). The value 0 plays a special role, since there is no limit from the left. We consider only functions with \( f(0) = 0 \). \( D \) is a metric space with the Skorodhod \( J_1 \)-metric

\[
d(f,g) = \inf_{\lambda} \| \lambda - \text{id} \|_{\infty} \lor \| f - g \circ \lambda \|_{\infty}
\]

(23)

The infimum is over all bijective increasing functions \( \lambda : [0,1] \to [0,1] \) and \( \cdot \|_{\infty} \) is the usual supremum norm. The \( \sigma \)-field on \( D \) is the Borel-\( \sigma \)-field via the Skorodhod metric.

On the space \( D \) consider the space-time transformations \( T_{a,b} : D \to D \), \( a \in D \), \( b \in D \)

\[
T_{a,b}(f) = af \circ b
\]
on \( D \). The set \( D_t \) consist of increasing caglad functions : \([0,1] \to [0,1] \). The composition of such space-time transformation is

\[
T_{a,b} \circ T_{c,d} = T_{ac,bd}
\]

Let \( U_v \), \( v \in V \), be independent rvs with a uniform distribution on \((0,1)\). (We assume without loss of generality \( U_i(\omega) \neq U_j(\omega) \) for all \( i \neq j \) and \( \omega \in \Omega \).

Define the random edge values \( L_{v,v_1} = T_{A_{v,v_1}} \) with values in \( D \to D \)

\[
(A_{v,v_1}(t), B_{v,v_1}(t)) = (\mathbb{I}_{(0,U_v]}(t)U_v, t/ U_v \lor 0)
\]

\[
(A_{v,v_2}(t), B_{v,v_2}(t)) = (\mathbb{I}_{(U_v,1]}(t)(1 - U_v), t - U_v)/(1 - U_v) \lor 0)
\]
t \( \in [0,1] \). Let \( L_{v,vw} = T_{A_{v,vw}} \) be the random path weight for the path \((v,vw)\), \( w \neq 0 \) and \( L_{v,v} = T_1 \circ \text{id} \) the identity. Notice \( A_{v,vw}(t) = \mathbb{I}_{L_{v,vw}}(t)|L_{v,vw}| \) and \( B_{v,vw}(t) = t|L_{v,vw}|/L_{v,vw}(t) \) where \( L_{v,vw} = [c_{v,vw}, d_{v,vw}] \) is from example 'Intervals Revisited' now for the infinite input \( U_i(\omega) \), \( v \in V \) for an \( \omega \in \Omega \). We face a backward weighted branching process \((V,L,D^P,\circ)\).

Define rvs \( R_{v,vw} := L_{v,vw}(C_{vw}) \), \( v, w \in V \) with values in \( D \) where \( C_v = C(U_v,\cdot) + \mathbb{I}_{(U_v,1]}QU_{w}^{-1} \). The function \( C : [0,1] \times [0,1] \to \mathbb{R} \) is given in (9) and \( Q \) is the Quicksort rvs from the example Quicksort above.

Our aim is to show that \( R_{v,vW} = \sum_{w \in W} R_{v,vw} \) converges point wise to a process \( R^v \) as \( W \to V \). The limit \( R = R_0 \) respectively \( R^v \) is called the Quicksort process.

We work on an abstract probability space \((\Omega, \mathcal{A}, P)\) rich enough for all our purposes. We use the supremum norm \( \| f \|_{\infty} = \sup_{0 \leq t \leq 1} |f(t)| \) on \( D \) and the
$L^p$-norm $\| \cdot \|_p$ for $1 \leq p < \infty$ for real valued rvs. Let $F$ be the space of all measurable functions $X : \Omega \rightarrow D$. For $1 \leq p < \infty$ let $F_p$ be the subspace of all rvs $X \in F$ with finite value

$$\|X\|_{\infty,p} := \|\|X\|_\infty\|_p$$

The map $\| \cdot \|_{\infty,p}$ is a pseudo metric on $F_p$. By the usual procedure using the equivalence relation $X \sim Y \Leftrightarrow P(X \neq Y) = 0$ we obtain a metric on the set $F_p$ of equivalence classes $[X] = \{Y \in F_p \mid X \sim Y\}$ for $X \in F_p$. $(F_p, \| \cdot \|_{\infty,p})$ is a Banach space with the usual addition and multiplication. We will not distinguish between random variables and equivalence classes in the sequel.

**Lemma 7** For any $1 < p < \infty$, $n \in \mathbb{N}$ holds

$$\| \sum_{w \in V_m} |R_{v,vw}| \|_{\infty,p} \leq (c + \|Q\|_p)(\frac{2}{1 + p})^{n/p} < \infty$$

$$\| \sum_{w \in V} |R_{v,v}| \|_{\infty,p} \leq \frac{c + \|Q\|_p}{1 - (\frac{2}{1 + p})^{1/p}} < \infty$$

for all $v \in V$ with $c = \sup_{x,t} |C(x,t)| < \infty$.

Proof: We skip the argument $c < \infty$. Notice the distribution of $(R_{v,vw})_{w \in V}$ is the same as of $(R_{w})_{w \in V}$. Define $e_m = E \sum_{v \in V_m} |I_v|^p$ for input $U$.

- $e_m \leq \frac{2}{1 + p} e_{m-1}$ for $m \in \mathbb{N}$

$$e_m = E(\sum_{i=1}^{2} \sum_{v \in V_{m-1}} |I_i|^p |I_{i,v}|^p)$$

$$= E(|I_i|^p (\sum_{v \in V_{m-1}} |I_{i,v}|^p))$$

$$= E(|I_i|^p) E(\sum_{v \in V_{m-1}} |I_{i,v}|^p)$$

$$= \frac{2}{1 + p} e_{m-1}$$

- $\| \sum_{v \in V_m} |R_v| \|_{\infty} \leq \sup_{v \in V_m} |I_v| |C_v|.$

Notice $R_v = \mathbb{I}_{I_v} |I_v| C_v$ and the $I_v$, $v \in V_m$ are a partition of $[0, 1)$. Therefore the left side is $\leq \sup_{v \in V_m} |I_v| |C_v|$.

- $E\| \sum_{v \in V_m} |R_v| \|_{\infty} \leq (c + \|Q\|_p) e_m$
Then the $L^p$-inequality and independence under the $\sigma$-field $F_m$ generated by $U_v$, $v \in V_{<m}$ provides
\[
E\|\sum_{v \in V_m} |R_v|\|_{\infty}^p \leq E(\sup_{v \in V_m} |I_v|^p | C_v|^p) 
\leq \sum_{v \in V_m} E(|I_v|^p | C_v|^p) = \sum_{v \in V_m} E(|I_v|^p E(|C_v|^p | F_m)) 
= \sum_{v \in V_m} E(|I_v|^p)E(|C_v|^p) 
= E(|C_v|^p)e_m \leq (c + \|Q\|_p)^p e_m
\]

\bullet E\|\sum_{v \in V} |R_v|\|_{\infty, p} \leq (c + \|Q\|_p) \frac{1}{1 - (\frac{2}{1 + p})^{1/p}}

\|
\sum_{v \in V} |R_v|\|_{\infty, p} = \|
\sum_{m \in N_0} \sum_{v \in V_m} |R_v|\|_{\infty, p}
\leq \sum_{m \in N_0} \|
\sum_{v \in V_m} |R_v|\|_{\infty, p}
\leq \sum_{m \in N_0} (c + \|Q\|_p)e_m^{1/p} \leq (c + \|Q\|_p)e_0^{1/p}(\frac{2}{1 + p})^{m/p}
\leq (c + \|Q\|_p) \frac{1}{1 - (\frac{2}{1 + p})^{1/p}}
\]

q.e.d.

As a consequence $R_W$ converges absolute and uniformly in $t \in [0, 1]$ for $W \to V$ almost surely.

**Theorem 8** Let $1 < p < \infty$. Then almost surely holds for all $v \in V$ and all $W \subset V$:
- The sum $R_{v,vW}$ is absolute convergent in $F_p$.
- For any increasing sequence $W_n$ to $W$ the sequence $R_{v,vW_n}$ converges absolute in supremum metric almost surely to a limit called $R_{v,vV}$ with values in $D$.

Proof: $Q$ has even finite exponential moments [17]. The theorem is standard by the above lemma and monotonicity. q.e.d.

Notice the set of all $\omega \in \Omega$ such that $\|
\sum_{v \in V} |R_{v,vU}|\|_{\infty}$ is infinite for some $v \in V$ has probability 0. The convergence in supremum metric implies also convergence for every $t \in [0, 1]$ and the convergence in Skorokhod metric on $D$.

Consider the distributional fixed point equation
\[
(X(t))_{t} \overset{D}{=} (UX_1(t \frac{U}{U} \land 1) + (1 - U)X^2(0 \lor t - \frac{U}{1 - U}) + C(U, t))_{t}
\]

on $D$. The rvs $X^1, X^2, U$ are independent. The rvs $X, X^1, X^2$ have values in $D$ and all have the same distribution. The rv $U$ has a uniform distribution on the
unit interval, the function $C : [0, 1] \rightarrow D$ given in (9). The symbol $\overset{D}{=}\}$ denotes equality in distribution in $D$. The right side is well defined.

We may choose versions of the processes $X^1, X^2$ and of $U$ defining $X$ that way. We obtain that way point wise equality in (24). For $t = 1$ we obtain

$$X(1) = UX^1(1) + X^2(1) + C(U)$$

(25)

where $C$ is given in (10). If $X(1)$ is integrable \cite{[5]} then $X(1)$ has the Quicksort distribution $Q$.

We write the stochastic fixed point equation (24) in the form

$$\begin{align*}
(X(t))_t \overset{D}{=} \mathbb{1}_{t \leq U}UX^1(1) + \mathbb{1}_{t \geq U}((1-U)X^2(\frac{t-U}{1-U}) + UX^1(1)) + C(U,t)
\end{align*}$$

(26)

treating the term $\mathbb{1}_{t \geq U}UX^1(1)$ like an additional cost function.

**Theorem 9** The family $R^v = R^v,v \in V$, $v \in V$, satisfies

$$R^v = L_{v,1} \ast_r R^{v,1} + L_{v,2} \ast_r R^{v,2} + C_v$$

(27)

for all $v \in V$ almost surely. The distribution of the rv $R$ is the unique solution of the distributional fixed point equation (24) in $F_p$ for every $p > 1$.

Proof: Establish

$$R_{v,vW} = \sum_{w \in W} R_{v,vw} = \mathbb{1}_{\emptyset \in W}C_v + \sum_{i=1}^2 \sum_{w \in W_i} L_{v,vi} \ast_r L_{vi,viw} \ast_r C_{viw}$$

$$= \mathbb{1}_{\emptyset \in W}C_v + \sum_{i=1}^2 L_{v,vi} \ast_r R_{vi,viW}$$

and then take $W$ converges to $V$. This proves equation (27). $R$ is a solution of the stochastic fixed point equation.

For the uniqueness consider the map $K$ on the set $M_p$ of probability measures $\mu$ on $D$ satisfying $\int \|f\|_\infty^p \mu(df) < \infty$ defined by

$$K(\mu) = \mathcal{L}(\sum_{i=1}^2 T_i(X_i) + C)$$

$\mathcal{L}$ denotes distribution. $X_1, X_2, U$ are independent rvs. The distribution of $X_1$ and $X_2$ is $\mu$ and the distribution of $U$ is uniform. The random variables $T_1, T_2$ are $D^D$-valued, $C$ is $D$-valued

$$T_1(f)(t) := \mathbb{1}_{f(t) \leq \frac{t}{U}} \land 1 \quad T_2(f)(t) := \mathbb{1}_{t \geq U(1-U)}f(\frac{t-U}{1-U}) \quad C(t) = C(U,t)$$

The fixed point of $K$ are the solution of (24) and vice versa.

16
$K$ is a strict contraction on the space $M_p$ for the metric

$$d_p(\mu, \nu) := \inf \|X - Y\|_{\infty, p}$$

The infimum is over all $D$-valued rv $(X, Y)$ on any probability space with marginal distribution $\mu$ and $\nu$. The space $(M_p, d_p)$ is a complete metric space [2].

- The contraction constant is $\leq \frac{2}{2p+1} < 1$.

For this let $(X_1, Y_1)$ and $(X_2, Y_2)$ be independent copies of each other with values in $D$, marginals $\mu, \nu$ and $\|X - Y\|_{\infty, p} \leq d_p(\mu, \nu) + \epsilon$. Choose the versions $\sum_{i=1}^{2} L_i X_i + C$ respectively $\sum_{i=1}^{2} L_i Y_i + C$ with the same $L_i, C$ for $K(\mu)$ and $K(\nu)$. Then

$$d_p(K(\mu), K(\nu)) \leq \|T_1(X^1) + T_2(X^2) + C - T_1(Y_1) - T_2(Y_2) - C\|_{\infty, p}$$

$$\leq \|T_1(C)\|_{\infty, p} + \|T_2(C)\|_{\infty, p}$$

$$\leq 2\|T_1\|_p(d_p(\mu, \nu) + \epsilon) \rightarrow_{\epsilon \rightarrow 0} \frac{2}{p+1}d_p(\mu, \nu)$$

The rest is easy using Banach Fixed Point Theorem [4]. q.e.d.

Remark: Here we took the view of a weighted branching process with $H = D, G \subset H^H$. The transformations $L_u, u$ and the cost functions $C_u$ depend on the values $U^n$, $v \in V$ at the knots. This corresponds to the Quicksort process with internal randomness of the algorithm giving the same distributional result for any deterministic input. In the next section we introduce another version of the Quicksort process, corresponding to a deterministic algorithm and a random input by iid rvs.

## 5 Convergence in Skorohod metric.

This section is on the almost sure convergence in Skorohod metric of the discrete Quicksort process $Y(U_{\mathbb{N}})$ to the Quicksort process for iid input $(U_i)_{i \in \mathbb{N}}$. We copy the Quicksort process approach, now modeling by a deterministic WBP with random input. The term $Y(U_{\mathbb{N}})$ reduces to a sum of finitely many terms. A random time change to the summands and a uniformity result will be the key to almost sure convergence in Skorohod metric.

Throughout this section $U$ is a sequence of iid rvs with a uniform distribution. The set $(H_2, \ast_2)$ of half open intervals is a monoid with the operation

$$(c, d) \ast_2 (c', d') = (c + c'(d - c), d + d'(d - c))$$

Consider the deterministic WBP $(V, L, G, \ast)$ on the infinite binary tree $V = \{1, 2\}^*$. Let $H = H_1 \times H_2$ and define $g_1, g_2 \in H^H$

$$g_1(u, I) := (u^1, I \ast_2 [0, \frac{|u^1|}{|u^1| + 1}])$$

$$g_2(u, I) := (u^2, I \ast_2 [\frac{|u| - |u^1| - 1}{|x| + 1}, 1])$$

17
if $|u| \geq 1$ and $I$ are non empty and the grave $\Delta = (\emptyset, \emptyset)$ otherwise. Let $G \subset H^2$ be generated by the functions $g_1, g_2$ adding the identity as neutral element and the function identical $\triangle$ acting as a grave. The operation on $G$ is the composition $g \circ g' = g' \circ g$ and $G$ operates right on $H$ via $h \circ I = g \circ h$. Put $L_{v, v} = g_v$ for all $v \in V$.

Define $\Psi : H_2 \to D^D$ by

$$\Psi(I)(f) = \Psi_I(f) = a_I \circ f \circ b_I$$

$$a_I := \mathbb{I}_I|I| \quad b_{(c, d)}(t) = \mathbb{I}_{d > c} 0 \lor \left(\frac{t - c}{d - c}\right) \land 1$$

Define $J_v : H_1 \to H_2$ by

$$L_v(u, [0, 1)) =: (u^v, J_v(u))$$

and $R_{v, w} : H_1 \to D$

$$R_{v, w}(u) := \Psi_{J_w(u)}(C_w(u^w))$$

$$C_w(u) := C(|w^w|, |w^1|, \cdot) + \mathbb{I}_{\frac{|w^1|}{|w^1| + 1}} \left[\frac{|u^1|}{|u|} + 1\right] Y(u^1, 1)$$

for $|u| \geq 1$, $I$ non empty and the grave otherwise. The function $C$ is given in (7) with extension $C(n, k, t) := C(n, k, \frac{nt}{n})$ in (7).

**Proposition 10** The map $H_2 \ni I \mapsto b_I$ is injective. For $I, J \in H_2$ and $v \in V$ holds

- $|I||J| = |I \circ J| = |J \circ I|$
- $b_I \circ b_J = b_{J \circ I}$
- $a_I \circ a_J \circ b_I = a_{J \circ I}$
- $\Psi_I \circ \Psi_J = \Psi_{J \circ I}$
- $L_{v, w}(u, I) = (u^w, I \circ J_v(u))$
- $J_{v, w}(u) = J_v(u) \circ J_w(u^v)$
- $R_{v, w} = R_w$
- $R_{v, w} = \sum_{w \in W} R_{v, w} \to_{V \to V} R_{v, v}$ point wise
- $R_v(u) = \sum_{i=1}^2 \Psi_{J_i(u)}(R_i(u^i)) + C_{\emptyset}(u)$
Proof: The first statement is easy. We obtain for $I = [c, d]$, $J = [c', d']$ and $I, J \neq \emptyset$
\[|I \ast_2 J| = (d - c)(d' - c') = |I||J| = |J \ast_2 I|\]
\[(b_I \circ b_J)(t) = b_J(0 \vee (t - c')) \wedge 1 = 0 \vee \frac{t - c'}{d - c} \wedge 1 = b_{J \ast_2 I}\]
\[a_I a_J(b_I(t)) = \mathbb{I}_I(t)|I|\mathbb{I}_J(0 \vee \frac{t - c}{d - c} \wedge 1) = \mathbb{I}_I(t)|I|\mathbb{I}_J\left(\text{min}\left(\frac{t - c}{d - c}, 1\right)\right) = a_{I \ast_2 J}(t)\]
\[\Psi_I \circ \Psi_J(f) = \Psi_I(a_J \circ b_J) = a_I a_J(b_I \circ b_J) = \Psi_{I \ast_2 J}(f)\]
\[L_{v,vw}(u, I) = L_w(u, I) = (u^w, I \ast_2 L_w(u, (0, 1])) = (u^w, I \ast_2 J_w(u))\]
\[(u^w, J_{vw}) = L_{v,vw}(L_v(u, (0, 1))) = L_{v,vw}(u^w, J_v(u)) = (u^{w^w}, \Psi_{J}(u) \ast_2 J_w(u^v))\]

Next show the result for $I$ or $J$ the empty set.

The first statement on $R$ is easy. For the second notice $R_{v,vW}$ has an increasing number of non zero summands for increasing $W$ and $R_{v,vV}(u)$ has only finitely many non zero summands for every $u \in H_1$. The third statement is also easy. The last follows by induction on the length of the input $u$. We skip this. q.e.d.

Now let $U = (U_i)_{i \in \mathbb{N}}$ be a sequence of independent rvs with a uniform distribution on $(0, 1)$. The input is now $U_{|n}(\omega)$ at the root $\emptyset$ of length $n$. From that we obtain at any vertex $(U_{|n})^v$ as starting value. Our objects of interest are $R_{v,vU}((U_{|n})^v) =: R_{v,vU}(n)$ (a slight abuse of notation). Notice the $n$ is for fixed input $U_{|n}$ at the root and not relative to the starting vertex $v$ of the path as used before. Further $R_{v,vV}(n) = Y((U_{|n})^v)$ with recursion
\[R_{v,vV}(n) = \sum_i \Psi_{J_i((U_{|n})^v)} R_{v_i,v_iV}(n) + C_i((U_{|n}))\]

We now introduce the limiting Quicksort process $R(\infty)$ of $R(n)$. We consider a weighted branching process $(V, \tilde{L}, \tilde{G}, \ast)$ with $\tilde{H} = (0, 1)^V \times H_2$ and $\tilde{G}$ generated by $\tilde{g}_1$, $\tilde{g}_2$, a grave and the identity
\[\tilde{g}_1(u, I) := (u^1)_{u_1} \ast_2 [0, u_1) \ast_2 \tilde{g}_2(u, I) := (u^2 - u_1, I \ast_2 (u^2_1 - u_1, 1))\]
Put $\tilde{L}_{v,v1} = \tilde{g}_1$ and define $\tilde{u}^v \in (0, 1)^V \times H_2$.

Notice $\tilde{u}^v = \frac{u^v}{d - c}$ for $\tilde{J}_v(u) = [c, d] \neq \emptyset$ and appropriate otherwise. Define
\[\tilde{R}_{v,vU}(u) := \Psi_{\tilde{J}_v(u)} \tilde{C}_{vU}(\tilde{u}^v)\]
where $\tilde{C}_v : (0, 1)^V \times H_2 \rightarrow D$, $v \in V$ is given by
\[\tilde{C}_v(u) := C(u_1, \cdot) + \mathbb{I}_{u_1} u_1 Q(u^v)\]
Let \( U = (U_n)_{n \in \mathbb{N}} \) be a sequence of independent rvs with a uniform distribution on \((0,1)\). Then \( \overline{R}_v(n) \) is a majorant of \( R_v(n) \) for every \( n \). For every \( p > 1 \) and \( n \in \mathbb{N} \) holds

\[
\| \overline{R}_{V_n}(n) \|_{\infty, p} \leq (c + \|Q^*\|_p)(2^{\frac{2}{1+p}})^{m/p} < \infty
\]

\[
\| \overline{R}_V(n) \|_{\infty, p} \leq \frac{c + \|Q^*\|_p}{1 - (\frac{1}{1+p})^1/p} < \infty
\]

\[\sup_{n \in \mathbb{N}} \overline{R}_V(n) < \infty \text{ and } \sup_{n \in \mathbb{N}} \overline{R}_{V_{m,n}}(n) \to_{m \to \infty} 0 \text{ a.e.}\]

Proof: We show \( c < \infty \) in Proposition 12 at the end of this paper. Since all moments of the Quicksort distribution \( Q = Q(U) \) are finite [17] and \( Y(U_{[n]}, 1) \) is a martingale [16] converging to \( Q \) we obtain finite moments of all order of the supremum \( Q^* \).

Notice the independence of \( \tilde{U}^i \), \( i = 1, 2 \) from the ranks \( (R_{1,v})_{n \in \mathbb{N}} \) where \( R_{1,v} = \sum_{j=1}^{n} \mathbb{1}_{U_{1} \leq U_{j}} \) denotes the rank of \( U_{1} \) under \( U_1, U_2, \ldots, U_n \). This implies \( Q^*(U^{v^1}) \) is independent of \( J_v(U_{[n]}) \) for every \( n \in \mathbb{N}, v \in V \). This is due to the fact that \( Q^*(U^{v^1}) = Q^*(\tilde{U}^{v^1}) \). Further \( \overline{C}(U) \) is independent of \( J_v(U) \).

- \( \overline{R} \) is a majorant of \( R \)

This follows by \( |C_v(U_{[n]})| \leq \overline{C}_v(U) \) for all \( n \) from the martingale argument.

- ||\( \overline{R}_{V_n}(n) \)\|_p || \( \leq (c + \|Q^*(U)\|_p)(\sum_{v \in V_{m}} E(|J_v(n)|^p))^{1/p} \)

\[
\| \overline{R}_{V_n}(n) \|_{\infty} = \| \sum_{v \in V_{m}} \mathbb{1}_{J_v(U_{[n]})} J_v(U_{[n]}) \overline{C}_v(U) \|_{\infty} \\
= \sup_{v \in V_{m}} |J_v(U_{[n]})| \overline{C}_v(U)
\]
\[
E(\|R_{V_m}(n)\|_p) = E(\sup_{v \in V_m} |J_v(U_{v,n})|^p E_\nu^p(U)) = E(\sum_{v \in V_m} |J_v(U_{v,n})|^p E_\nu^p(U)) = \sum_{v \in V_m} E(|J_v(U_{v,n})|^p E_\nu^p(U))) \leq E(\nu^p(U)) \sum_{v \in V_m} E(|J_v(U_{v,n})|^p) = (c + \|Q^*(U)\|_p)^p \sum_{v \in V_m} E(|J_v(U_{v,n})|^p)
\]

- \(E(|(U_{v,n})|^p) \leq \frac{n^p}{1+1+p}\) 
  \(|(U_{v,n})|^1 + 1\) is the rank of \(U_1\) under \(U_1, U_2, \ldots, U_n\). The rank has a uniform distribution on \(\{1, \ldots, n\}\). We have
  
  \[E(|(U_{v,n})|^p) = \sum_{k=0}^{n-1} \frac{1}{n^k} \leq \int_0^n x^p dx = \frac{n^p}{1+p}\]

- \(E(|J_v(U_{v,n})|^p) \leq (\frac{1}{1+p})^{|v|}\) for all \(n, v\)
  
  For \(v \in V_m\)

\[
\begin{align*}
E(|J_1(U_{v,n})|^p) & \leq E(U_1)^p = \int_0^1 x^p dx = \frac{1}{p+1} \\
E(|J_2(U_{v,n})|^p) & = E(|J_1(U_{v,n})|^p) \\
E(|J_v(U_{v,n})|^p) & = E(|J_{v,n-1}(U_{v,n}) * J_{v,n})(U_{v,n})|^{v(m-1)}|^p) \\
& = E(|J_{v,n-1}(U_{v,n})|^p| J_{v,n}(U_{v,n})^{v(m-1)}|^p) \\
& = E(|J_{v,n-1}(U_{v,n})|^p| E(|J_{v,n}(U_{v,n})^{v(m-1)}|^p) \\
& \leq E(|J_{v,n-1}(U_{v,n})|^p|) \frac{1}{1+p} \leq \cdots \leq \left(\frac{1}{1+p}\right)^m
\end{align*}
\]

- \(\sum_{v \in V_m} E(|J_v(U_{v,n})|^p) \leq (\frac{2}{1+p})^m\) for \(m \in \mathbb{N}\)
  Easy.

- \(\|R_{V}(n)\|_{\infty,p} \leq \frac{c + \|Q^*\|_p}{(1+\frac{2}{1+p})^{1/p}}\)
For the last statement we need estimates uniformly in $n$.

- There exists a $p > 1$ with $d := E(\sum_n \sup_n |J_v(U_n)|^p) < 1$.
- It suffices to show $\sup_{n \in \mathbb{N}} |J_1(U_n)| < 1$ a.e.

$J_1(U_n)$ is $\frac{R_{1,n} - 2}{n - 1} < 1$ for $n \geq 1$, where $R_{1,n}$ is the rank of $U_1$ under $U_1, \ldots, U_n$. Since $\frac{R_{1,n}}{n}$ converges a.e. to $U_1 \in (0, 1)$ we obtain the partial result for $i = 1$. The argument for $i = 2$ are analogue.

- $E(\sup_n |J_v(U_n)|^p) \leq (\frac{d}{2})^{[v]}$ for all $v \in V$.
- $\sum_{v \in V_m} E(\sup_n |J_v(n)|^p) \leq d^m$ for $m \in \mathbb{N}$
- $\|\sup_{n \in \mathbb{N}} \|\!\!\!\!\|\overline{R}_{V_m}(n)\|_\infty^p \leq (c + \|Q^*(U)\|_p)d^{m/p}$

$E(\sup_n \|\!\!\!\!\|\overline{R}_{V_m}(n)\|_\infty^p) = E(\sup_n \|\!\!\!\!\|J_v(U_n)\|\!\!\!\!\|\overline{C}_v(U)\|_\infty^p) \leq E(\sup_n \|\!\!\!\!\|J_v(U_n)\|\!\!\!\!\|\overline{C}_v(U)\|_\infty^p) = E(\sup_{n \in \mathbb{N}} \|\!\!\!\!\|J_v(U_n)\|\!\!\!\!\|\overline{C}_v(U)\|_\infty^p) = (c + \|Q^*(U)\|_p)d^{m/p}$

$\|\sup_n \|\!\!\!\!\|\overline{R}_V(n)\|_\infty\|_p \leq \frac{c + 3\|Q^*_p\|_p}{1 - d}d^{m/p} < \infty$
The last statement follows from \( \| \sup_n \| V_\infty(n) \|_1 \leq \| \sup_n \| V_{\|}(n) \|_\infty \|_p < \infty \).

**Proof of Theorem (1)** Estimate path wise for some \( m \in \mathbb{N} \)

\[
d(R_V(n), R_V(\infty)) \leq d(R_V(n), R_{V_{\leq m}}(n)) + d(R_{V_{\leq m}}(n), R_{V_{\leq m}}(\infty)) + d(R_{V_{\leq m}}(\infty), R(\infty)) = I + II + III
\]

\[
\sup_n I \leq \sup_n \| R_{V_{\leq m}}(n) \|_\infty \to_{m \to \infty} 0
\]

\[
\sup_n III \leq \sup_n \| R_{V_{\leq m}}(\infty) \|_\infty \to_{m \to \infty} 0
\]

\[
\lim_{n \to \infty} II \leq \sum_{v \in V_{\leq m}} \lim d(R_v(n), R_v(\infty))
\]

Therefore it suffices to show the \( \lim_{n} d(R_v(n), R_v(\infty)) = 0 \) for every \( v \in V \).

Define for \( n \in \mathbb{N} \) and \( v \in V \) the function \( \varphi = \varphi_{v,n} : [0,1] \to [0,1] \) by

\[
\varphi(0) = 0, \quad \varphi(A_v(\infty)) = A_v(n), \quad \varphi(B_v(\infty)) = B_v(n), \quad \varphi_{v,n}(1) = 1
\]

and linear between, where \( (A_v(n), B_v(n)] = J_v(U_{v,n}) =: J_v(n) \neq \emptyset \).

\[
d(R_v(n), R_v(\infty)) \leq d(\| J_v(n) \| C_v(n), \| J_v(n) \| C_v(\infty)) + d(\| J_v(n) \| C_v(n), \| J_v(n) \| C_v(\infty)) + d(\| J_v(n) \| C_v(\infty), \| J_v(n) \| C_v(\infty)) = IV + V + VI
\]

\[
\lim_{n \to \infty} \sup IV \leq \lim_{n \to \infty} \sup \| \varphi(t) - t \| \leq \lim_{n \to \infty} \sup |A_v(n) - A_v(\infty)| + \lim_{n \to \infty} \sup |B_v(n) - B_v(\infty)| = 0
\]

\[
\lim_{n \to \infty} \sup V \leq \lim_{n \to \infty} \sup |C_v(n, \varphi(t)) - C_v(\infty, \varphi(t)) = 0
\]

\[
\lim_{n \to \infty} \sup VI \leq \lim_{n \to \infty} \sup |A_v(n) - A_v(\infty)| + \lim_{n \to \infty} \sup |B_v(n) - B_v(\infty)| = 0
\]

q.e.d.

**Proposition 12**

\[
\sup_{n \in \mathbb{N}, 1 \leq n \leq k \leq n-1} \frac{|C(n,k,l)|}{n} \leq c < \infty.
\]

If \( \frac{t_n}{n} \to t \in [0,1], \frac{t_n}{n} \to s \in [0,1] \) and \( t \neq s \) then

\[
\frac{C(n,i_n, s_i)}{n} \to_C C(t,s)
\]

where the \( C \)-functions are given in (7) and (9).
Proof: The proof is straightforward. Notice $a(0, \cdot) = 0$.

The asymptotics of the harmonic numbers are

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{10}{120n^4} + O(n^{-6})$$

with $\gamma = 0.577215...$ the Euler constant. We will use $H_n = \ln n + \gamma + b_n$ with $b_n = O\left(\frac{1}{n}\right)$.

For $n \geq 2$ and $1 \leq l = l_n \leq n, 1 < i = i_n < n - 1$

$$\frac{a(n,l)}{n} = \frac{2 + 2 l \gamma - 2(1 - \frac{l}{n}) \ln(1 - \frac{l}{n}) + 2 \frac{l}{n} \ln n - 6 \frac{l}{n} + O\left(\frac{\ln n}{n}\right)}{n}$$

$$= O(1) + 2 \frac{l}{n} \ln n$$

$$\frac{C(n,l,i)}{n} = \left\{ \begin{array}{ll}
I & := \frac{n - 1 - a(n,l) + a(i - 1, l)}{n} \\
II & := \frac{n - 1 - a(n,l) + a(i - 1, i - 1)}{n} \\
III & := \frac{n - 1 - a(n-l) + a(i-1, i-1) + a(n-i, l-i)}{n}
\end{array} \right.$$

$$|I_{l \leq i - 1}| \leq I_{l \leq i - 1}(O(1) + \left|2 \frac{l}{n} \ln n - 2 \frac{l}{n} \ln(i - 1)\right|)$$

$$\leq I_{l \leq i - 1}(O(1) + 2 \frac{l}{n} \ln \frac{i - 1}{n})$$

$$\leq O(1) + 2 \frac{i - 1}{n} \ln \frac{i - 1}{n}$$

$$\leq O(1) + 2 \sup_{0 \leq x \leq 1} |x \ln x| = O(1)$$

$$|I_{i - 1 \leq l \leq i}| \leq I_{i - 1 \leq l \leq i}(O(1) + \left|2 \frac{l}{n} \ln n - 2 \frac{i - 1}{n} \ln(i - 1)\right|)$$

$$\leq O(1) + 2 \frac{i}{n} \ln n - \frac{i - 1}{n} \ln(i - 1)$$

$$\leq O(1)$$

$$|I_{l > i}| \leq O(1) + \left|2 \frac{l}{n} \ln n - 2 \frac{i - 1}{n} \ln(i - 1) - 2 \frac{l - i}{n} \ln(n - i)\right|$$

$$\leq O(1) + 2 \frac{i - 1}{n} - \frac{i - 1}{n} - \frac{l - i}{n} \ln\left(n - \frac{i - 1}{n}\right)$$

$$\leq O(1) + 2 \frac{l - n}{n} \ln \frac{n - i - 1}{n}$$

$$\leq O(1) + 2 \frac{l - n - i - 1}{n} \ln\left(1 - \frac{n - i - 1}{n}\right) = O(1)$$

The cases $i = 1$ and $i = n - 1$ have to be done separately, but provide no difficulty. (Otherwise $\ln 0$ would appear in the above argument.)

24
Now assume $\lim_{n \to \infty} \frac{t_n}{n} = t \in (0, 1)$ and $\lim_{n \to \infty} \frac{i_n}{n} = s \in (0, 1)$. We obtain

$$\lim_{n \to \infty} \mathbb{1}_{t < i - 1} I = \mathbb{1}_{t < s} (\lim_{n} (\frac{-1 - 2t \gamma - 2t \ln n + 2(1 - t) \ln(1 - t) + 6t}{n}) + s(2 + \frac{t}{s} \gamma + 2 \frac{t}{s} \ln ns - 2(1 - \frac{t}{s}) \ln(1 - \frac{t}{s}) - 6 \frac{t}{s}))$$

$$= \mathbb{1}_{t < s} (-1 + 2s + 2(1 - t) \ln(1 - t) - 2(s - t) \ln(s - t) + 2s \ln s)$$

and

$$\lim_{n \to \infty} \mathbb{1}_{t > i} III = \mathbb{1}_{t > s} (\lim_{n} (\frac{1 - 2t \gamma + 2t \ln n - 2(1 - t) \ln(1 - t) - 6t}{n}) + s(2 + \frac{2t \gamma + 2 \ln ns - 6}{(1 - s)})(2 + \frac{t - s}{1 - s} \gamma)$$

$$+ \frac{2}{1 - s} \ln(n(1 - s)) - 2(1 - \frac{t - s}{1 - s}) \ln(1 - \frac{t - s}{1 - s}) - 6 \frac{t - s}{1 - s}))$$

$$= \mathbb{1}_{t > s}(1 + 2s \ln s + 2(1 - s) \ln(1 - s))$$

This implies the statement. q.e.d.

Notice for $t = s$ there might be no convergence of $C(n, i_n, i_n)$. Conditioned to $l_n < i_n - 1$ the above converges to $-1 + 2s + 2s \ln s + 2(1 - s) \ln(1 - s)$. But conditioned to $l_n > i_n$ the above converges to $1 + 2s \ln s + 2(1 - s) \ln(1 - s)$. In our setting the probability of being in that case is 0 and the boundedness of $C$ is sufficient for our purpose.

References


26