Discrete Mathematics and Theoretical Computer Science DMTCS vol. **2012**, 14:2, 129–154

The Analysis of Find and Versions of it

Diether Knof and Uwe Roesler

Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, Germany

received 12th November 2007, revised 20th January 2010, accepted 12th September 2012.

In the running time analysis of the algorithm Find and versions of it appear as limiting distributions solutions of stochastic fixed points equation of the form

$$X \stackrel{\mathcal{D}}{=} \sum_i A_i X_i \circ B_i + C$$

on the space D of cadlag functions. The distribution of the D-valued process X is invariant by some random linear affine transformation of space and random time change. We show the existence of solutions in some generality via the Weighted Branching Process. Finite exponential moments are connected to stochastic fixed point of supremum type

$$X \stackrel{\mathcal{D}}{=} \sup_i (A_i X_i + C_i)$$

on the positive reals.

Specifically we present a running time analysis of m-median and adapted versions of Find. The finite dimensional distributions converge in L^1 and are continuous in the cylinder coordinates. We present the optimal adapted version in the sense of low asymptotic average number of comparisons. The limit distribution of the optimal adapted version of Find is a point measure on the function

$$[0,1] \ni t \mapsto 1 + \min\{t, 1-t\}.$$

Keywords: Quickselect, Find, running time, analysis of algorithms, perpetuities, cadlag functions, iterated function system, random affine transformation

1 Introduction

We consider fixed points of the operator

$$K(\mu) \stackrel{\mathcal{D}}{=} \sum_{i \in \mathbb{N}} A_i X_i \circ B_i + C. \tag{1}$$

on the set D of cadlag functions on the unit interval $I = [0,1]$. The rvs $((A_i, B_i)_i, C), X_j, j \in \mathbb{N}$ are independent and all X_j rvs have the distribution μ. The rvs A_i, C, X_i take values in the set D and the rvs B_i take values in $D_\uparrow$, mapping the unit interval to itself and piecewise increasing. They serve as a random time change, the A's as a space transformation and the C's as a displacement.

The very first example of such fixed points appeared in the analysis of Find as the limit distribution of the Find algorithm [10]. Grübel and Rösler used a specific representation of the iterates K^n by some rvs with nice properties and showed the convergence of theses rvs in Skorodhod metric on D to a limit, the fixed point. This approach works for the 2-version of Find, splitting into 2 sets, but not for the 3-version, splitting into 3.

The intuition suggests some convergence result also for the 3-version. In his thesis, D. Knof [13][24] used rigorously the contraction method. The contraction method [22],[17] suggests to find a nice complete metric space such that the operator K is a (strict) contraction on it. (The space D with the Skorodhod metric is not complete.) Knof [24] introduced a suitable nice complete metric space. He studied the finite dimensional distributions of the D-valued process and showed convergence of all finite dimensional distributions. Since the family of finite dimensional distributions is consistent, there exists a projective limit, i.e. a probability measure μ on the product space $[0,\infty]^I$ with the given finite dimensional distributions. Another fixed point argument, this time for a fixed point equation of max-type, showed μ is a probability measure on the product space $[0,\infty)^I$. Next he used the contractive branching structure of the underlying Weighted Branching Process (WBP) to verify a condition given by Dubins and Hahn [6] showing D has mass 1 for the outer measure of μ. Therefore the outer measure restricted to D is the limit distribution of Find.

Studying the details and the conditions used there, gave us reasons to prove fixed point results of K by using WBP [23] and to return somehow to rvs. We consider a larger class of operators, obtain stronger results than Knof and have less technically difficulties with right-continuous versions. With the forward view [25] we obtain point wise convergence of appropriate rvs to the fixed point, with the backward view we obtain via the stochastic order convergence of the iterates of K. This interplay of the two points of views simplifies and extends the results of Knof.

Using the monotonicity respectively the absolute convergence in our setting of the rvs R_n we obtain under weak conditions the convergence to a limiting rv R and also continuity in mean. Another fixed point argument shows $\sup_t |R(t) - R_n(t)| \to_n 0$ and so we were able to circumvent Knof's difficulty of showing μ is a probability measure on D.

Interesting is the appearance of a fixed point equation of max-type in all the approaches, Grübel-Rösler, Knof and this paper in order to analyze $\sup_t R(t)$. A fixed point equation of max-type is of the form

$$X \stackrel{\mathcal{D}}{=} \sup_i (A_i X_i + C_i) \tag{2}$$

where $((A_i, C_i)_i), X_j, j \in \mathbb{N}$ are positive real valued and independent rvs and all X_j have the same distribution as X. The first appearance of a stochastic fixed point equation of max-type is in the analysis of Find [10],

$$X \stackrel{\mathcal{D}}{=} \sup\{UX_1, (1-U)X_2\} + 1.$$

The positive rvs U, X_1, X_2 are independent, U is uniformly distributed on the unit interval, $X \sim X_1 \sim X_2$. The argument used was via the contraction method, similar to fixed point equations of sum type, using the contraction condition $E\sum_i A_i^2 < 1$. A more general and systematic approach was carried out by Neininger and Rüschendorf [18]. (See also [1] for some related work.) The basic condition is a contraction constant like $E\sum |A_i|^p < 1$ for some $p > 0$. There are more fixed points, even if the contraction condition fails [12]. (They show a nice connection between fixed points of max-type and sum-type.) If the weights A_i, C_i are all deterministic, we know explicitly the set of all solutions of fixed point of equation (2) [2]. In this paper we use essentially a result by Rüschendorf and Schopp [26] on finite exponential moments. Their approach relies on asymptotic estimates of higher moments and the Taylor expansion of the exponential function, Theorem 8.

We come now to the applications of K-fixed points, the optimal adapted version of Find. The algorithm Find or Quickselect [11] [14] finds the l-th smallest element out of n distinct reals. The procedure is:

- Choose a pivot element by some random procedure;
- Find the list of strictly smaller numbers and the list of strictly larger numbers;
- Recall the appropriate lists till termination.

This version is called the 3-version. The 2-version of Find forms the list of strictly smaller elements than the pivot and the list of larger elements including the pivot.

The divide-and-conquer algorithm Find has a complete running time analysis for the 2-version [10] and uniform distribution of the pivot. Grübel and Rösler took appropriate versions of the running time rvs, which converged in Skorodhod topology on the space D. However this does not work with the 3-version. Moreover it is even shown there, that the standard 3-version with uniform distribution does not converge in Skorodhod topology.

But may be asking for a convergence result of random variables is not the right question. For practical purposes one is mainly interested in the distribution of the running time, for given l and n. As it turns out, we obtain a limiting distribution if the quotient $\frac{l}{n}$ converges to some t. It seems natural to expect some continuous behavior in t of the distribution. As explained below, only the recursive structure (4)(5) is available for the one-dimensional distribution and it determines the one-dimensional distribution. (The one-dimensional distributions are continuous in time with respect to weak convergence.) But that is not enough to construct a nice process $Y = (Y_t)_{t\in I}$ with the right one-dimensional distributions.

We require the stronger recursive equation (6) for the distribution of a whole process (compare [10] [7]). That recursion provides a limiting distribution of a process Y with right continuous paths. Their one-dimensional distributions do the job for the running time including the expected continuity of one-dimensional distributions $t \mapsto \mathfrak{L}(Y_t)$.

Our approach given here does exactly this. The Weighted Branching Process (WBP) is the suitable framework to construct (in a mechanic way) nice versions of the rvs in a general setting, such that the processes converge weakly as processes on D. The main workload in this approach is showing the right continuity of the limiting process Y. (Similar to showing continuous paths for the Brownian motion.) But this property is essential for the continuity of certain functionals and to work with the limiting distribution. It turns out that the one-dimensional distributions of Y are continuous in time.

For the approximation of the Find versions X^n to the limit Y we use the distance

$$d(\mu, \nu) = \inf \sup_{t\in I} \|X_t - Z_t\|_1$$

for probability measures on D. The infimum is taken over all processes (X, Z) with one dimensional distributions μ, ν on some probability space. In this distance we show all finite dimensional distributions of X^n converge to those of Y and all finite dimensional distributions are continuous in the cylinder coordinates. (This does not imply weak convergence on D, since we do not prove tightness or equivalently the convergence of T-distributions with $T \subset I$ dense and countable.)

Eickmeier and Rüschendorf [7] showed L^p-convergence of the processes, averaging the time with the Lebesgue measure, to a limiting process in L^p. This approach allows no conclusion on one-dimensional distributions at some fixed time $t \in I$ or the convergence of finite dimensional distributions. We could put the process to 0 on all rational time points and still have the same process in L^p sense. Even if we choose appropriate weaker formulations, the continuity of the one-dimensional distributions remains questionable. Intuitively, if l/n converges to some t then the asymptotic value we obtain should be independent of how l/n converges to t locally. Therefore we need smoothness of the path, e.g. the path is in D.

Our approach is also superior to the previous one by Grübel and Roesler [10], since we can deal with 3-version and give a very general scheme how to choose the right versions for more general processes.

In detail, let $X^n(l)$ be the number of comparisons in order to find the lth smallest out of n, which is proportional to the running time of the algorithm. The correct normalization turns out to be

$$Y^n\left(\frac{l-1}{n}\right) = \frac{X^n(l)}{n}. \tag{3}$$

If $l = l_n$ is a sequence such that $\frac{l_n}{n}$ converges to some $t \in [0,1]$ we want $Y^n(\frac{l}{n})$ to converge in distribution. For the 2-version the distribution of $Y^n(\frac{\lfloor tn \rfloor}{n})$ will converge to the distribution of the Find process Y at t in the unit interval [10] and as we will show, so does the 3-version. We intend to show convergence, point wise and as a process, for many more versions of Find. (More precisely for all reasonable versions known to us.) We include median versions [3],[9],[16],[20], pick the pivot as the median out of m random chosen elements, or adapted versions [19], choosing the pivot element as the kth smallest in the set of m, where k depends on n and l. Partial Quicksort [15] sorts the lth smallest ones and provides a nice example with non positive costs rv C.

For all these versions, the running time $X^n(l)$ in the 3-version of Find (the 2-version is analogous) satisfies the recursive equation

$$X^n(l) \stackrel{\mathcal{D}}{=} \mathbb{1}_{l<U^n(l)} X_1^{U^n(l)-1}(l) + \mathbb{1}_{l>U^n(l)} X_2^{n-U^n(l)}(l-U^n(l)) + C^n(l). \tag{4}$$

Here $U^n(l)$ is the rank of the pivot element and $U^n(l)-1$ is the cardinality of the set of strictly smaller numbers than the pivot element. The rv $C^n(l)$ gives the number of comparisons to find the pivot element and then to form the two lists. A lower bound of $C^n(l)$ is $n-1$ and a reasonable upper bound $n-1+m^2$. The terms $X_1^{\cdot}, X_2^{\cdot}$ correspond to the running time for the two lists with an index for the size. The rvs $(U^n, C^n), X_1^i, X_2^j, i, j < n$ are assumed to be independent. The distribution of the X_1^j and X_2^j is given recursively by the distribution of X^j.

With the correct normalization (3) we obtain

$$\begin{aligned} Y^n\left(\frac{l-1}{n}\right) \stackrel{\mathcal{D}}{=}\ & \mathbb{1}_{l<U^n(l)} \frac{U^n(l)-1}{n} Y_1^{U^n(l)-1}\left(\frac{l-1}{U^n(l)-1}\right) \\ & + \mathbb{1}_{l>U^n(l)} \frac{n-U^n(l)}{n} Y_2^{n-U^n(l)}\left(\frac{l-1-U^n(l)}{n-U^n(l)}\right) + \frac{C^n(l)}{n}. \end{aligned} \tag{5}$$

Extend Y^n appropriate to a right continuous step function. We shall show under appropriate conditions Y^n converges as a process weakly to a process Y with cadlag values. Further all one-dimensional distributions converge. We obtain a fixed point equation for the distribution,

$$(Y(t))_t \stackrel{\mathcal{D}}{=} (\mathbb{1}_{t<U} U Y_1(\frac{t}{U}) + \mathbb{1}_{t\geq U}(1-U)Y_2(\frac{t-U}{1-U}) + 1)_t. \tag{6}$$

since $\frac{U^n}{n}$ converges weakly to some U. Here U, Y_1, Y_2 are independent and Y has the same distribution as Y_1, Y_2 with values in D.

What about solutions of (6)? The distribution of the original Find process [10] in the 2-version is the smallest positive fixed point of this distributional fixed point equation with U having a uniform distribution. This characterizes the distribution. An existence proof for the three version is given in Knof's PhD-thesis [13] in a more general setting and also in this paper. In the last section we include the convergence of the discrete 3-Find algorithm as an example.

The median m-version was treated by Grübel [9]. He provided pictures of the limiting distribution function and gave more details for the average running time, which becomes better compared to original Find. It seems obvious, that in average the algorithm will perform better, if we can choose a pivot element, which leaves for the next round a smaller list (at least with very high probability). This leads to adapted versions, choosing the kth out of m random ones, such that for $l < n/2$ the rank of the pivot is a little bit larger than l and for $l > n/2$ a little bit smaller with high probability. Martinez, Panario and Viola [19] take that view for various m. They provide also asymptotically best adapted versions for the average.

We come now to asymptotically best adapted Find versions. Choose an $m = m_n$ of the order $\ln n$ and the $k = k_n(l)$th smallest within the m picked ones, where $\frac{k}{m}$ is close to $\frac{l}{n}$. We shall show in Theorem 4, that $(Y^n(t))_t$ converges to the deterministic process $(1 + t \wedge (1-t))_t$.

We can, for example, find the median out of n with asymptotically $\frac{3n}{2}$ comparisons. This is the very best one can do within versions of Find. We need roughly n steps comparing every element to the first pivot element and the remaining appropriate list is of cardinality roughly $\frac{n}{2}$. For that list we need at least $\frac{n}{2}$ comparisons. This sums up to at least $\frac{3n}{2}$ comparisons. And asymptotically that is enough. The point is, that from the remaining list we have to find an element with a given very small rank or a very large rank. A recall of the adapted algorithms leaves us with a very small remaining list, which is asymptotically negligible.

The given argument for the median holds for every l out of n. First we need n comparisons for the reduction and then in the best case remains a list of cardinality $l \wedge (n-l)$ to be searched for an element of very small rank or very large rank. We need at least as many comparisons as the cardinality of the set. This sums up to asymptotically $n + l \wedge (n-l)$ comparisons plus terms of smaller order than n.

Mathematically we show weak convergence of Y^n to the deterministic function $t \mapsto 1 + \min\{t, 1-t\}$. From this and the included result $EY^n(t) \to_n EY(t)$ we obtain well known results for the average number of comparisons of the order $n + l \wedge (1-l) + o(n)$ ([19] and literature cited there). With arbitrary high probability is the distribution of $Y^n(t)$ arbitrary close to that of $Y(t)$ for fixed t and sufficiently large n. Notice, that this is still an average argument and a weaker statement then the worst case behavior of Y^n of the algorithm. For worst case considerations, the algorithm PICK, developed by Blum, Floyd, Pratt, Rivest and Tarjan [21][8], shows an upper bound of 5.4305...n comparisons.

2 Weighted Branching Process

We describe first a weighted branching tree. The underlying graph is the tree $V = \cup_{n=0}^{\infty} \mathbb{N}^n$ of all finite sequences of natural numbers (including the empty sequence) with the directed edges $(v, vi), v \in V, i \in \mathbb{N}$. Let $(G, *)$ be a measurable semi group with a grave $\triangle$ ($\triangle * g = \triangle = g * \triangle$ for all $g \in G$,) and a neutral element g_0 ($g_0 * g = g = g * g_0$ for all $g \in G$). (To every semi group we can add in a measurable way a grave and a neutral element.) Let $(H, +)$ be another measurable semi group with grave and neutral element and let G operate left, transitive and measurable on H ($*_H : G \times H \to H$ is associative and measurable). Every edge (v, vi) carries a random length $L_{v,vi}$ with values in the semi group G and every vertex v a random toll (cost) C_v with value in H. We call $(V, L, C, (G, *, H, +))$ (respectively $(V, L, (G, *))$) a Weighted Branching Process (WBP)[23] (with or without cost function), if the $((L_{v,vi})_{i\in\mathbb{N}}, C_v),\ v \in V$ are iid rvs.

For convenience and simplicity $*_H = *,\ *(f, g) = f * g = fg$ and we suppress the empty set $\emptyset$ and $\omega \in \Omega$, if possible. Notice, the semi group action $*$ needs not to be commutative and the order of the terms may be important.

Define the path weight $L_v = L_{\emptyset,v}$ of the path from $\emptyset$ to v (in graph theoretical sense) recursively by $L_\emptyset$ the neutral element g_0 and

$$L_{vi} = L_v * L_{v,vi}$$

for $v \in V,\ i \in \mathbb{N}$. For the length $L_{v,vw}$ of the (unique) path from v to vw we use the corresponding definition on the tree vV. More formally, $L_{v,v} = g_0,\ L_{v,vwi} = L_{v,vw} * L_{vw,vwi}$. (This definition is compatible with the previous one. If convenient we use $L_v^w = L_{w,wv}$).

In our specific setting described in a moment we will consider the sequence

$$R_n := \sum_{|v|<n} L_v * C_v \qquad R_n^w = \sum_{|v|<n} L_v^w * C_{wv} \tag{7}$$

The length $|v|$ of a vertex v is the number of components. The root $\emptyset$ has length $|\emptyset| = 0$. If everything is positive then R_n increases point wise in n to a limit called R. This process is our main object of interest. The distribution of the limit will be a fixed point (the smallest positive) of K (8) under suitable conditions.

Let $D = D(I, \mathbb{R})$ be the space of right continuous functions on the unit interval $I = [0, 1]$ with existing left limits. We shall also use $\overline{D} = D(I, \overline{\mathbb{R}})$, where $\overline{\mathbb{R}}$ denotes the reals extended by $-\infty, +\infty$ with the Alexandrov one-point compactification. We will use D_+ ($\overline{D}_+$) for the positive functions in D ($\overline{D}$) and $D_\uparrow$ for the subset of all functions $f : I \to I$ in D such that for every $s \in [0, 1)$ the function f is increasing on $[s, t)$ for some $t > s$.

The σ-field $\sigma(D)$ ($\sigma(\overline{D}_+)$) on D ($\overline{D}_+$) is generated by the Skorodhod metric ρ [5]

$$\rho(f, g) = \inf\{\epsilon > 0 \mid \exists\lambda :\ \ \|\lambda - id\|_\infty < \epsilon, \|f - g \circ \lambda\|_\infty < \epsilon\}$$

where λ is a bijective and (strictly) increasing (and continuous) functions from the unit interval to itself. We consider $(D, \sigma(D))$ as a subset of $(\mathbb{R}^T, \mathcal{B}^T),\ T \subset I$ dense in I and containing 1. (Analogous $\overline{D}$.) The σ-field $\mathcal{B}^T$ is the product σ-field of the Borel σ-fields $\mathcal{B}$ on the reals. The induced σ-field on D generated by $\mathcal{B}^T$ is $\sigma(D)$, (Billingsley [5], Theorem 14.5). The σ-fields on D_+ and $D_\uparrow$ are the induced ones.

Let $A = (A_i)_{i\in\mathbb{N}},\ B = (B_i)_{i\in\mathbb{N}},\ C$ be random variables. The A_i, C take values in the set D (respectively $\overline{D}_+$), the rvs B_i values in $D_\uparrow$. The distribution of (A, B, C) is fixed throughout the paper.

Let $M(E)$ be the set of all probability measures on E. Define the map K from the set $M(D)$ (or $M(\overline{D}_+)$) to itself by

$$K(\mu) \stackrel{\mathcal{D}}{=} \sum_{i\in\mathbb{N}} A_i X_i \circ B_i + C. \tag{8}$$

(We use the convention $0\cdot\infty = 0 = \infty\cdot 0$ on $[0,\infty]$.) The rvs $(A,B,C), X_i, i\in\mathbb{N}$ are independent and all X_i rvs have the distribution μ. In more elaborate form,

$$K(\mu) = \mathfrak{L}((\sum_{i\in\mathbb{N}} A_i(t)X_i(B_i(t)) + C(t))_{t\in I}).$$

The symbol $\mathfrak{L}$ denotes the distribution of a rv. The map K is assumed to be well defined.

We speak of the positive case, if all rvs (length, cost and starting distribution) are positive (in $\overline{D}_+$), allowing the value ∞. In the positive case the operator K (8) is always well defined, since the sum of positive measurable terms is point wise well defined and measurable in $[0,\infty]$.

We are interested in fixed points of K obtained by an iteration of K starting with the point measure $\delta_{\underline{0}}$ on the function $\underline{0}\in D$ identical $0\in\mathbb{R}$. With an ad hoc notation the iterates of K applied to μ are

$$\begin{aligned}
K(\mu) &\stackrel{\mathcal{D}}{=} \sum_i A_i\cdot X_i\circ B_i + C\\
K\circ K(\mu) &\stackrel{\mathcal{D}}{=} \sum_i A_i(\sum_j (A_{ij}\circ B_i\cdot X_{ij}\circ B_{ij}\circ B_i) + C_i\circ B_i) + C\\
K^3(\mu) &\stackrel{\mathcal{D}}{=} \sum_i A_i \sum_j A_{ij}\circ B_i(\sum_k A_{ijk}\circ B_{ij}\circ B_i\cdot X_{ijk}\circ B_{ijk}\circ B_{ij}\circ B_i\\
&\qquad + C_{ij}\circ B_{ij}\circ B_i) + \sum_i A_i\cdot C_i\circ B_i + C\\
K^4(\mu) &= \dots
\end{aligned}$$

Let us introduce a suitable notation for the iterates. Actually the iteration of the operator K corresponds to the backward view [25] of a WBP branching process with costs, which we will exploit now.

The semi group structure $(G,*)$ consists of the semi group $G = D\times D_\uparrow$ (or $G=\overline{D}_+\times D_\uparrow$ in the positive case) with the semi group operation

$$(a,b)*(a',b') = (a\cdot(a'\circ b), b'\circ b) = (aa'\circ b, b'\circ b).$$

Here $\circ$ denotes the convolution and $\cdot$ is the point wise multiplication in D (respectively $\overline{D}_+$ with the convention $0\cdot\infty = 0 = \infty\cdot 0$). The neutral element is $(\underline{1}, id)$ with id the identity. The grave $\triangle$ corresponds to $(0,0)$. (More precisely $(0,0)$ is the right sided grave and will serve as left sided grave only after a suitable definition.)

The semi group G operates transitive on the value set $H = D$ (respectively $\overline{D}_+$) of the cost $c\in H$ by

$$(a,b)*c = ac\circ b$$

with common multiplication on D. The grave acts as $\triangle * c = \underline{0}$.

Notice the operation $*$ is not commutative and the order of the terms is important.

In our special situation the commutative operation $+$ on $\mathbb{R}$ ($\overline{\mathbb{R}}_+$) extends to $\mathbb{R}^I$ ($\overline{\mathbb{R}_+}^I$) point wise, the operation $*$ is bilinear in the first coordinate on G,

$$\begin{aligned}(a_1+a_2,b)*(a',b') &= (a_1,b)*(a',b')+(a_2,b)*(a',b')\\ (a,b)*(a'_1+a'_2,b') &= (a,b)*(a'_1,b')+(a,b)*(a'_2,b')\end{aligned}$$

for all $a,a_1,a_2,a',a'_1,a'_2 \in D,\overline{D}_+,\ b,b' \in D_\uparrow$, and the operation $*$ is bilinear on H in the sense

$$\begin{aligned}(a_1+a_2,b)*c &= (a_1,b)*c+(a_2,b)*c\\ (a,b)*(c_1+c_2) &= (a,b)*c_1+(a,b)*c_2\end{aligned}$$

for all a,a_1,a_2,b as above and $c,c_1,c_2 \in D$ ($\overline{D}_+$).

Let $(A^v,B^v,C^v),\ v\in V$ be iid copies of (A,B,C). The edge length $L_{v,vi}$ on the edge (v,vi) is given by (A_i^v,B_i^v) and the vertex weight $C_v = C^v$. Notice an element (a,b) of G corresponds to a map and acts like a change of time (b) and a transformation of space (a).

Define the rvs

$$R_n := \sum_{|v|<n} L_v * C_v$$

$n\in\mathbb{N}_0$ as path weighted cost up to $n-1$th generation. The first values of $(R_n)_n$ are, suppressing the $\emptyset$,

$$R_0=\underline{0},\ R_1=C,\ R_2=\sum_i A_iC^i\circ B_i+C,\ R_3=R_2+\sum_i\sum_j A_iA_j^i\circ B_iC^{i,j}\circ B_j^i\circ B_i$$

and so on.

The connection to the operator K is given in general by the following lemma [25].

Lemma 1 *If the terms are well defined then the distribution of R_n is the n-iterate $K^n(\delta_{\underline{0}})$. The sequence $(R_n)_n$ satisfies the backward recursion [25]*

$$R_{n+1}=\sum_{i\in\mathbb{N}} L_i^\emptyset * R_n^i + C_\emptyset \tag{9}$$

Proof: The sum R_{n+1} is well defined in the positive case and satisfies

$$\begin{aligned}R_{n+1} &= \sum_{|v|<n+1} L_v*C_v = L_\emptyset * C_\emptyset + \sum_{i\in\mathbb{N}}\sum_{|v|<n} L_{iv}*C_{iv}\\ &= C_\emptyset+\sum_{i\in\mathbb{N}}\sum_{|v|<n} L_i*L_v^i*C_{iv} = C_\emptyset+\sum_{i\in\mathbb{N}} L_i*\sum_{|v|<n} L_v^i*C_{iv}\\ &= C_\emptyset+\sum_{i\le N} L_i*R_n^i.\end{aligned}$$

We prove the distributional result on R_n by an induction on n. The induction start $n = 0$ is true since $R_0 \equiv 0$ has the distribution $\delta_{\underline{0}}$. For the induction step n to $n+1$ argue by the backward recursion, notice $K(\mu) \stackrel{\mathcal{D}}{=} \sum_{i\in\mathbb{N}} L_i * X_i + C$

$$R_{n+1} = C_\emptyset + \sum_{i\in\mathbb{N}} L_i * R_n^i \stackrel{\mathcal{D}}{=} K(K^n(\delta_{\underline{0}})) = K^{n+1}(\delta_{\underline{0}}).$$

The distribution of the R_n^i is $K^n(\delta_{\underline{0}})$ by the induction hypothesis. □

3 The limit and continuity mean.

In the positive case the sequence $K^n(\delta_{\underline{0}})$ increases in stochastic order. (The stochastic order on probability measures $\overline{\mathbb{R}}\,(D, \overline{D})$ are defined by $\mu \preceq_s \nu \Leftrightarrow \int f d\mu \leq \int f d\nu$ for all positive (coordinate wise) increasing functions f.) Then by general representation theory, there should be an increasing sequence of rvs to the distributions $K^n(\delta_{\underline{0}})$. The clue is, our sequence $(R_n)_n$ does the job. The limit of the R_n, which exists, should be a fixed point of K by the monotonicity of the operator.

We denote be $\overline{f}$ the supremum norm of a function on $[0,1]$ and $A := \sum_i |A_i|$. We use R_n^v for the R_n rv on the tree vV.

Lemma 2 *In the positive setting the rvs $R_n,\ n \in \mathbb{N}$ increase point wise to $R = \sum_{v\in V} L_v * C_v$. The limit satisfies the equation*

$$R = \sum_{i\in\mathbb{N}} A_i R^i \circ B_i + C \tag{10}$$

where R^i is the point wise limit of the R_n^i. The distribution of R is the smallest fixed point of K in stochastic order, provided R has paths in $\overline{D}$.

In the general setting assume $E\overline{A} < 1,\ E\overline{C} < \infty$. Then $\sup_{n\in\overline{\mathbb{N}}} E\overline{|R_n|} < \infty$ and $E\overline{|R - R_n|}$ converges to 0.

Proof: By definition $R_n(\omega)(t)$ is increasing in n for every realization $\omega \in \Omega$ and every t in the unit interval I. Therefore the limit $R(\omega)(t)$ exists, but might be infinite somewhere. We obtain the equation (10) by the backward view equation (9) going to the limit. Therefore R satisfies the fixed point equation (10) and, provided is $\overline{D}_+$-valued, is a fixed point of K.

Let ν be another fixed point of K. Then $\delta_{\underline{0}} \preceq_s \nu$, $K^n(\delta_{\underline{0}}) \preceq_s K^n(\nu) = \nu$ for all $n \in \mathbb{N}$ and in the limit $\mathfrak{L}(R) \preceq_s \nu$.

Now assume the additional assumptions but for the positive case. The $R_n : \Omega\times I \to \overline{\mathbb{R}}_+$ are measurable maps with respect to the product σ field. (Notice the rvs A_i, B_i, C are all measurable with respect to the product σ-field.) Therefore $E(\overline{R_n}) : \Omega \times I \to \overline{\mathbb{R}}_+$ is measurable and also $\overline{ER_n}$. We will show by induction $\overline{E(R_{n+1} - R_n)} \leq E(\overline{A})\overline{E(R_n - R_{n-1})}$ for $n \in \mathbb{N}_0$ putting $R_{-1} = \underline{0}$ by definition.

The induction start is easy, since $R_1 = C$. Now the step n to $n+1$. The backward view in Lemma 1, notice $R_n(t)$ is finite for all t, provides

$$0 \leq R_{n+1} - R_n = \sum_i L_i * (R_n^i - R_{n-1}^i) = \sum_i A_i(R_n^i - R_{n-1}^i) \circ B_i \tag{11}$$

for $n \in \mathbb{N}$. This implies the recursion, use conditional independence,

$$\begin{aligned} E\overline{(R_{n+1}-R_n)} &= E\overline{\sum_i A_i(R^i_n - R^i_{n-1}) \circ (B_i)} \\ &= \sup_t E(\sum_i A_i(t) E((R^i_n - R^i_{n-1})(B_i(t)) \mid (A_i, B_i)_i)) \\ &\leq \sup_t E(\sum_i A_i(t) E\overline{(R^i_n - R^i_{n-1})}) \\ &\leq E(\overline{A}) E\overline{(R_n - R_{n-1})} \end{aligned}$$

This ends the induction.

An easy consequence is $\sup_{n\in\mathbb{N}} E\overline{R_{n+1}}$ is finite by iteration. This implies $R_n(t),\ n \in \mathbb{N}_0,\ t \in I$ is a uniformly integrable family. Therefore the limits $R(t)$ can be added to the family and $\sup_t ER(t)$ is finite.

Next argue

$$E\overline{(R-R_n)} \leq E\overline{\sum_{i\geq n}(R_{i+1}-R_i)} \leq \sum_{i\geq n} E^i(\overline{A}) E\overline{C} \to_n 0$$

Now to the general case. Consider $\tilde{A}_i := |A_i|, \tilde{B}_i = B_i, \tilde{C} = |C|$ instead of A_i, B_i, C. All rvs are positive now and the Lemma provides finiteness of $\tilde{R}$ and for the first moments of it. The argument is easy observing $|R_{n+1} - R_n| \leq \tilde{R}_{n+1} - \tilde{R}_n$ point wise for all $\omega \in \Omega$. For example, the sum $ER = \sum_n (ER_{n+1} - ER_n)$ converges absolute. We skip the details. □

We turn to show continuity in mean of the limit R.

A process $X = (X_t)_{t\in I}$ is called continuous in mean, if the map $I^2 \ni (s,t) \mapsto E|X_s - X_t|$ is continuous. An equivalent formulation is $I \ni t \mapsto E|X(s) - X(t)|$ is continuous at s for all $s \in I$. Since R is right continuous the Monotone Convergence Theorem ($\overline{E|R|}$) implies right continuity of the function $(s,t) \mapsto E|R(s) - R(t)|$.

Theorem 3 *Assume $E\overline{A} < 1,\ E\overline{C} < \infty$ and the maps $(s,t) \mapsto E\sum_i |A_i - A_i(t)|$ and $(s,t) \mapsto E|C(s) - C(t)|$ are continuous. Further assume*

$$\lim_{y\to 0}\lim_{x\to 0} E \sup_{|s-t|>x} \sum_i A_i(s) \mathbb{1}_{|B_i(s)-B_i(t)|>y} = 0.$$

Then the rv R is continuous in mean.

Proof: Under the above assumptions $R_n,\ n \in \overline{\mathbb{N}}$ is well defined and all sums converge absolutely a.e.. For $s,t \in I$ define $a(s,t) := E\sum_i |A_i(s) - A_i(t)|,\ c(s,t) := E|C(s) - C(t)|,\ e(s,t) := E|R(s) - R(t)|$. Define $\tilde{a}(x) := \sup_{|s-t|\leq x} a(s,t)$ and analogous $\tilde{c}(x),\ \tilde{e}(x)$. The symmetric functions a and c are continuous by assumption. They are uniformly continuous on the compact set I^2 and are 0 on the diagonal. This implies the continuity of the functions $\tilde{a}$ and $\tilde{c}$.

We obtain by the backward view (9) for R

$$\begin{aligned}
e(s,t) &\leq E\sum_i |A_i(s)||R^i(B_i(s)) - R^i(B_i(t))| + E\sum_i |A_i(s)-A_i(t)||R^i(B_i(t))| + E|C(s)-C(t)| \\
&\leq E(\sum_i |A_i(s)|E(|R^i(B_i(s)) - R^i(B_i(t))| \mid A,B,C)) + E(\sum_i |A_i(s)-A_i(t)|E(\overline{R^i})) + c(s,t) \\
&\leq E\sum_i |A_i(s)|e(B_i(s),B_i(t)) + E(\overline{R})a(s,t) + c(s,t)
\end{aligned}$$

and

$$\begin{aligned}
\tilde{e}(x) &\leq \sup_{|s-t|\leq x} E\left(\sum_i |A_i(s)|e(B_i(s),B_i(t))(\mathbb{1}_{|B_i(s)-B_i(t)|\leq y} + \mathbb{1}_{|B_i(s)-B_i(t)|>y})\right) + E(\overline{R})\tilde{a}(x) + \tilde{c}(x) \\
&\leq \tilde{e}(y)E(\overline{A}) + 2E(\overline{R})E(\sup_{|s-t|\leq x}\sum_i |A_i(s)|\mathbb{1}_{|B_i(s)-B_i(t)|>y}) + E(\overline{R})\tilde{a}(x) + \tilde{c}(x)
\end{aligned}$$

for all y. The function $x \mapsto \tilde{e}(x)$ is increasing.

Let $\tilde{e}(0+)$ be the right limit of $\tilde{e}$ at 0. Using the assumptions we obtain for fixed y and then let x tend to 0 and afterwards $y \to 0$

$$\begin{aligned}
\tilde{e}(0+) &\leq \tilde{e}(y)E(\|A\|_\infty) + 2E(\overline{R})\lim_x E(\sup_{|s-t|\leq x}\sum_i |A_i(s)|\mathbb{1}_{|B_i(s)-B_i(t)|>y}) \\
&\quad + E(\overline{R})\lim_x \tilde{a}(x) + \lim_x \tilde{c}(x) \\
&\leq \tilde{e}(0+)E(\overline{A}) + 0.
\end{aligned}$$

This proves $\tilde{e}(0+) = 0 = \tilde{e}(0)$ and consequently continuity of e and mean continuity of R. □

As a consequence, under the above assumptions the maps $t \mapsto ER(t)$ and $t \mapsto E|R(t)|$ are continuous. Further the map from the unit interval to the distributions $t \mapsto \mathfrak{L}(R(t))$ is continuous in the Wasserstein metric l_1 and in weak topology. The Wasserstein (Mallows) l_p-metric $p \geq 1$ [4] on distributions on $\mathbb{R}$ is given by

$$l_p(\mu,\nu) = \inf E\|X - Y\|_p$$

using the L_p-norm. The infimum is taken over all rvs X, Y with distribution μ respectively ν on some probability space. The infimum is attained for the rvs $F^{-1}(U), G^{-1}(U)$ with U uniformly distributed on I, F and G the distribution functions to μ and ν. (Especially for l_1, $l_1(\mu,\nu) = \int |F(x) - G(x)|dx$.) l_p-convergence is equivalent to convergence in distribution and convergence of the pth moment.

4 Values in D

Our next aim and the main task is to show R is a D-valued ($\overline{D}_+$) rv.

We show the result under the additional
Condition F: $A_i(t)A_j(t) = 0$ for all $t \in I$ and $i \neq j$ almost everywhere.

Theorem 4 *Assume condition F and $E\sum_i \overline{A}_i^p < 1,\ E\overline{C}^p < \infty$ for some $p \geq 1$. Then $\overline{R}$ is uniformly p-integrable and R takes values in D. R is a fixed point of K. R_n converges a.e. exponentially fast to R in supremum norm and*

$$\|\overline{R-R_n}\|_p \leq (E\sum_i \overline{A}_i^p)^{n/p}\|C\|_p. \tag{12}$$

In the positive case the distribution of R is the smallest fixed point of K restricted to $M(\overline{D}_+)$ in stochastic order.

Proof: Let us start with the positive case. The $R_n : \Omega \times I \to \overline{\mathbb{R}}_+$ are measurable functions with respect to the product σ-field. (Notice the rvs A_i, B_i, C are all measurable with respect to the product σ-field on $\Omega \times I$.) Therefore $\overline{R}_n : \Omega \to \overline{\mathbb{R}}_+$ is measurable. The rv $\overline{R}_n$ satisfies for every $\omega \in \Omega$, use assumption F,

$$\overline{R_n} \leq \overline{\sum_i |A_i|}\ \overline{R_{n-1}^i} + \overline{C} \leq \sup_i \overline{A_i}\ \overline{R_{n-1}^i} + \overline{C} \tag{13}$$

Let the map $L : M(\overline{\mathbb{R}}_+) \to M(\overline{R}_+)$ be defined by

$$L(\mu) = \mathfrak{L}(\sup_i \overline{A_i}X^i + \overline{C})$$

where $((A_i)_i, C), X^j,\ j \in \mathbb{N}$ are independent and all X^j have the distribution μ. L is well defined, since everything is positive.

With Theorem 4.1 [17] we obtain with the contraction method a fixed point $\mu \in M(\mathbb{R}_+)$ of L. This fixed point has a finite pth moment. The solution is obtained by iteration $\nu_n := L^n(\delta_0) \to_n \mu$ and it is therefore the smallest fixed point in stochastic order.

The sequence ν_n is increasing in stochastic order. Since equation (13) implies the distribution of $\overline{R}_n$ is dominated by $L(\overline{R}_{n-1})$ in stochastic order, we obtain by induction $\overline{R}_n$ is bounded in stochastic order by ν_n for all n. This implies $\overline{R}_n$ is bounded and the limit $\overline{R}$ is bounded in stochastic order by μ and is square integrable. In turn, the rvs $R_n(t)$ and the limit $R(t)$ are finite and p-integrable uniformly in t.

Now consider $a_n = \|\overline{R-R_n}\|_p$ and obtain, using conditional independence

$$\begin{aligned}
a_n &= \|\overline{\sum_i A_i(R^i - R^i_{n-1}) \circ B_i}\|_p \leq \|\sup_t \sum_i |A_i(t)|\overline{R^i - R^i_{n-1}}\|_p \\
&= \|\sup_i \overline{A_i}\ \overline{R^i - R^i_{n-1}}\|_p = (E(\sup_i (\overline{A_i})^p(\overline{R^i - R^i_{n-1}})^p)^{1/p} \leq (E(\sum_i (\overline{A_i})^p(\overline{R^i - R^i_{n-1}})^p)^{1/p} \\
&\leq (E(\sum_i \overline{A_i^p} a_{n-1}^p)^{1/p} = a_{n-1}(E\sum_i \overline{A_i^p})^{1/p} \leq \ldots \leq a_0(E\sum_i \overline{A_i^p})^{n/p} \to_n 0
\end{aligned}$$

Since $\overline{R-R_n}$ decreases and converges to 0 in L_p we conclude it decreases point wise to 0 a.e.. By the Markov inequality we obtain

$$P(\overline{R-R_n} > e^{\lambda n}) \leq \frac{a_n}{e^{\lambda n}} \leq a_0 e^{n(-\lambda + \ln c)} \to 0$$

$c = (E\sum_i \overline{A_i^p})^{1/p}$. Using Borel-Cantelli we obtain exponential fast convergence of $R_n \to R$ in supremum norm for $\ln c < \lambda$.

Since R_n has values in D_+ and converges uniformly to R we obtain the limit R has values in D_+ a.e.. Consequently its distribution is a fixed point of K. It is the smallest one by the monotonicity of the operator K, $K^n(\delta_{\underline{0}})$ increases in stochastic order in n.

Now to the remaining case, $G = D \times D_\uparrow$. Consider $\tilde{A}_i = |A_i|, \tilde{C} = |C|$ instead of A_i, C. The above theorem provides $\tilde{R}_n \to_n \tilde{R}$ a.e. uniformly in sup norm. The sum $R = \sum_n (R_n - R_{n-1})$ is absolute convergent point wise since $|R_n - R_{n-1}| \leq \tilde{R}_n - \tilde{R}_{n-1}$ point wise (for all $\omega \in \Omega$) and the $\tilde{\cdot}$-rvs are finite. Consequently R has values in D and is a square integrable fixed point of K. □

Notice, under condition F the statement $E \sum_i \overline{A_i} < 1$ implies $E\overline{A} < 1$.

$$E^2(\overline{A}) \leq E(\overline{A}^2) = E((\overline{\sum_i A_i})^2) = E\overline{\sum_i A_i^2} \leq E \sum_i \overline{A_i}$$

Corollary 5 *Under the assumptions of the Theorem 3 and Theorem 4 holds $P(R(t-) \neq R(t)) = 0$ for all $t \in I$.*

Proof: The Markov inequality provides

$$P(R(t-) \neq R(t)) \leq \lim_{\epsilon \to 0} \limsup_{s \nearrow t} P(|R(t) - R(s)| < \epsilon) \leq \lim_{\epsilon \to 0} \lim_{s \nearrow t} \frac{E|R(t) - R(s)|}{\epsilon} = 0.$$

□

The process $(R(t-))_t$ with left continuous paths and existing right limits has the same finite dimensional distributions as $(R(t))_t$.

5 Exponential moments of the fixed point.

We are interested in exponential moments of $R_n(t)$ and $R(t)$.

Lemma 6 *Assume $\overline{A} \leq 1$ a.e. and for some $\lambda > 0$ exists a constant $K < \infty$ such that*

$$Ee^{a|C(t)| + aK(A(t)-1)} \leq 1$$

for all $0 \leq a \leq \lambda$ and $t \in I$. Then $Ee^{a|R_n(t)|} \leq e^{aK}$ for all $0 \leq a \leq \lambda,\ t \in I$ and $n \in \mathbb{N}$.

Proof: The proof runs by induction. The case $n = 1$ is easy since $R_1 \stackrel{\mathcal{D}}{=} C$. The induction step is uniformly in t,

$$\begin{aligned}
Ee^{a|R_{n+1}(t)|} &= Ee^{a|\sum_i A_i(t) R_n^i(B_i(t)) + C(t)|} \\
&\leq E(e^{a|C(t)|} E(e^{a \sum_i |A_i(t)||R_n^i(B_i(t))|} \mid (A_i, B_i)_i, C)) \\
&= E(e^{a|C(t)|} \prod_i E(e^{a|A_i(t)||R_n^i(B_i(t))|} \mid (A_i, B_i)_i, C)) \\
&\leq E(e^{a|C(t)|} \prod_i E(e^{aK|A_i(t)|} \mid (A_i, B_i)_i, C)) \\
&= e^{aK} Ee^{a|C(t)| + aK(A(t)-1)} \\
&\leq e^{aK}
\end{aligned}$$

$\square$

Theorem 7 *Assume* $\overline{A} < 1$ *a.e. and* $\sup_t Ee^{\lambda|C(t)|} < \infty$ *for some* $\lambda > 0$ *and assume* $R(t)$ *is finite a.e. for all* $t \in I$*. Then* $e^{aR_n(t)}$ *converges to* $e^{aR(t)} \in L_1$ *in* L_1*-norm for all* $0 \leq a < \lambda$ *and* $t \in I$*.*

Proof: Let $a_0 < \lambda$.

• There exists a constant $K = K(a_0, \lambda) > 0$ such that $Ee^{aK(\overline{A}-1)+|C(t)|} \leq 1$ for all $0 \leq a \leq a_0,\ t \in I$.

For fixed t consider the function $g(a) = Ee^{aK(\overline{A}-1)+|C(t)|)}$. The function is well defined, twice differentiable and convex. (The second derivative is positive

$$g''(a) = E((K(\overline{A}-1) - |C(t)|)^2 e^{a(K(\overline{A}-1)-|C(t)|}) \geq 0.$$

Since $g(0) = 1$ and $g'(0) \leq K(E\overline{A}-1) - \sup_t E|C(t)|$ is negative for K sufficiently large, it remains to show $g(a_0) \leq 1$. Use the Hölder duals $p = \frac{\lambda}{a_0}, q = \frac{p}{p-1} = \frac{\lambda}{\lambda - a_0}$, to show

$$g(a_0) \leq \|e^{a_0 K(\overline{A}-1)}\|_q \|e^{a_0|C(t)|}\|_p \to_{K\to\infty} 0$$

decreases in $K \to \infty$ to 0 uniformly in $t \in I$.

• $Ee^{a|R_n(t)|} \leq e^{aK}$ for all $t \in I,\ n \in \mathbb{N},\ a \in [0, a_0]$.

Our Lemma 6 applies.

• The family $e^{a|R_n(t)|},\ t \in I,\ n \in \mathbb{N},\ 0 \leq a \leq a_1$ is uniformly integrable for all $0 < a_1 < a_0$.

We have to show $\sup_{t,n,a} Ee^{a|R_n(t)|}\mathbb{1}_{e^{a|R_n(t)|}>c}$ converges to 0 as $c \to \infty$. By monotonicity it suffices to take $a = a_1$. Using Hoelder with $p = \frac{a_0}{a_1}$ and $q = \frac{a_0}{a_0 - a_1}$ we obtain, using the Markov inequality,

$$\begin{aligned} Ee^{a_1|R_n(t)|}\mathbb{1}_{e^{a_1|R_n(t)|>c}} &\leq \|e^{a_1|R_n(t)|}\|_p \|\mathbb{1}_{e^{a_1|R_n(t)|}>c}\|_q \\ &\leq e^{a_1 K}(\frac{a_1 K}{c})^{1/q} \to_{c\to\infty} 0. \end{aligned}$$

• $Ee^{aR_n(t)} \to_{n\to\infty} Ee^{aR(t)}$ for all $0 \leq a < \lambda$ and $s \in I$.

The uniform integrability as stated above implies $Ee^{aR_n(t)} \to_{n\to\infty} Ee^{aR(t)} < \infty$ for all $0 \leq a \leq a_1$ and $t \in I$. The convergence of the moments implies L_1-convergence. Since a_1 can be arbitrary but strictly smaller than λ we are done. $\square$

6 Exponential moments of the supremum

Now we consider $\overline{R}$ and explain first the structure of the proof. If R satisfies the stochastic fixed point equation

$$R = \sum_i A_i R^i \circ B + C$$

then the supremum $\overline{R}$ satisfies under condition F

$$\overline{R} \leq \sup_t (\sum_i |A_i(t)| \overline{R^i} + \overline{C}) \leq \sup_i \overline{A_i}\ \ \overline{R^i} + \overline{C}.$$

Our following results will rely on this estimate and a theorem [26] concerning stochastic fixed points of max-type.

Define the operator $L : M(\overline{\mathbb{R}}_+) \to M(\overline{\mathbb{R}}_+)$ of max-type by

$$L(\mu) = \mathfrak{L}(\sup_{i\in\mathbb{N}}(Y_i X_i + Z_i))$$

where $(Y_i, Z_i)_i, X_j,\ j \in \mathbb{N}$ are independent and all X_j have distribution μ. The rvs Y_i, Z_i take values in $\overline{\mathbb{R}}_+$. The fixed point equation $L(\mu) = \mu$

$$X \stackrel{\mathcal{D}}{=} \sup_{i\in\mathbb{N}}(Y_i X_i + Z_i)$$

is called stochastic fixed point equation of max-type.

For $q > 1$ define

$$f(q) = \frac{\|\sup_i Z_i\|_q}{1 - (E\sum_i Y_i^q)^{1/q}}.$$

For $a, p > 0$ let $b = b(a,p) = (pea)^{-1/p}$. Let $M_q(\mathbb{R}_+)$ be the set of probability measures on $\mathbb{R}_+$ with finite q moment.

Theorem 8 (Rueschendorf-Schopp) *Assume $E\sum_i Y_i^q < 1$ for all $q \geq q_0$ and for some $a, p > 0$*

$$\limsup_{q\to\infty} \frac{f(q)}{bq^{1/p}} \leq 1.$$

Then L has a unique fixed point $\mu = \mathfrak{L}(X)$ in the set $M_{q_0}(\mathbb{R}_+)$, $\|X\|_q \leq f(q)$ and

$$\limsup_{x\to\infty} \frac{P(X > x)}{e^{-ax^p}} \leq 1.$$

Remark: If $p > 1$ then

$$\limsup_{\lambda\to\infty} \frac{Ee^{\lambda X}}{c\lambda^{\frac{p}{p-1}}} \leq 1,$$

where $c = \frac{(p-1)}{p^{\frac{p}{p-1}} a^{\frac{1}{p-1}}}$.

Remark The condition $E\sum_i Y_i^q < 1$ for all $q \geq q_0$ implies $\sum_i Y_i \leq 1$ a.e..

Theorem 9 *Assume condition F and the conditions of Theorem 8 using $Y_i = \overline{A}_i$ and $Z_i = \overline{C}$. Then $\overline{R}$ is dominated in stochastic order by the smallest fixed point of L and*

$$\limsup_{x\to\infty} \frac{P(\overline{R} > x)}{e^{-ax^p}} \leq 1.$$

If $p > 1$ then for all $t > 0$

$$\limsup_{\lambda\to\infty} \frac{\ln Ee^{\lambda\overline{R}}}{c\lambda^{\frac{p}{p-1}}} \leq 1$$

where $c = \frac{(p-1)}{p^{\frac{p}{p-1}} a^{\frac{1}{p-1}}}$.

Proof: The operator L is well defined. L is a monotone operator and therefore $L^n(\delta_0)$ increases to the smallest fixed point ν of L. That is the unique fixed point in $M_{q_0}(\mathbb{R}_+)$.

Then

$$\overline{R_n} \leq \sup_t \sum_i |A_i(t)| \overline{R^i_{n-1}} + \overline{C} \leq \sup_i \overline{A_i} \ \ \overline{R^i_{n-1}} + \overline{C}.$$

By induction we conclude the distribution of $\overline{R_n}$ is dominated in stochastic order by $L^n(\delta_{\underline{0}})$ which in turn is dominated by $\nu = L(\nu)$. In the positive case R_n is increasing to R and $\overline{R_n}$ is increasing to some rv, which has to be $\overline{R}$. This provides the existence and finiteness of R and $\overline{R}$. (The argument in the general case follows the same way, since the telescope sum $R = \sum(R_n - R_{n-1})$ converges absolutely and uniformly point wise.)

The limit $\overline{R}$ of the $\overline{R_n}$ is dominated in stochastic order by ν. We obtain the statements by Theorem 8 and the remark after. □

Example: Find The maximum of Find satisfies the fixed point equation

$$\overline{R} \stackrel{\mathcal{D}}{=} U\overline{R_1} + (1-U)\overline{R_2} + 1,$$

where U is uniform on I. Then $\|U\|_q = (q+1)^{1/q}$ and using some asymptotic for $q \to \infty$

$$\|\overline{R}\|_q \leq f(q) = \frac{1}{1 - e^{\frac{1}{q} \ln \frac{2}{q+1}}} \approx \frac{q}{\ln q}.$$

We can choose $p = 1$ and any $a > 0$. It follows

$$P(\overline{R} > x) \leq e^{-ax}$$

for x sufficiently large. All exponential moments of $\overline{R}$ are finite.

7 Discrete approximations

We consider recursive equations in distribution of the form

$$(X(n,t))_{t\in I} \stackrel{\mathcal{D}}{=} (\sum_{i\in\mathbb{N}} A_i(n,t) X(I_i(n,t), B_i(n,t)) + C(n,t))_{t\in I}. \tag{14}$$

This gives raise to operators K_n on distributions on D or $\overline{D}_+$ defined by

$$K_n(\mu_0, \ldots, \mu_{n-1}) = \mathfrak{L}((\sum_i A_i(n,t) X(I_i(n,t), B_i(n,t)) + C(n,t))_{t\in I}) \tag{15}$$

for $n \in \mathbb{N}$. Here $((A_i(n,\cdot), B_i(n,\cdot), I_i(n,\cdot))_i, C(n,\cdot)), X_j(m,\cdot), m \in \mathbb{N}_0, m < n, j \in \mathbb{N}$ are independent and $X_j(m,\cdot)$ are distributed like μ_m. The $A_i(n,\cdot), C(n,\cdot)$ take values in D (respectively $\overline{D}_+$), the $B_i(n,\cdot)$ in $D_\uparrow$. The $I_i(n,\cdot)$ are rvs with values in $D(I, \{0,1,\ldots,n-1\})$.

Consider the sequence $\mu_n, n \in \mathbb{N}$ recursively given by

$$\mu_0 = \delta_{\underline{0}},\ \mu_1 = K_1(\mu_0),\ \mu_2 = K_2(\mu_0, \mu_1), \ldots, \mu_n = K_n(\mu_0, \mu_1, \ldots, \mu_{n-1}), \ldots.$$

(We do not need (without loss of generality) starting distributions up to n_0 and start the induction afterwards, since we put the noisy part into C.) Our aim is to show under suitable conditions μ_n converges to $\mathfrak{L}(R)$. As the example of 3-Find indicates, we can not do this on the level of rvs. So on the level of distributions we shall show convergence of all finite dimensional distributions, using the Mallows l_1-distance. This is however done by choosing appropriate rvs which will converge in L_1-sense and therefore also in probability. A general way to pick appropriate rvs is via WBP.

Let $((A_i^v, B_i^v, I_i^v)_i, C^v),\ v \in V$ be iid copies of $((A_i, B_i, I_i)_i, C),\ v \in V$. Here A_i, C are rvs with values in $\overline{\mathbb{N}} \times D$ (respectively $\overline{\mathbb{N}} \times \overline{D}_+$), B_i are rvs with values in $\overline{\mathbb{N}} \times D_\uparrow$ and I_i is a rv with values in $F := \{(n, f) \in \overline{\mathbb{N}} \times D(I, \{0, 1, \ldots, n-1\}) \mid f(0, t) = 0$ and $f(n, t) < n$ for $n \in \mathbb{N},\ t \in I\}$. The rvs $A_i(0, \cot), B_i(0, \cdot), C(0, \cdot)$ are $\underline{0}$.

The semi group $(G, *)$ is $G = F \times D \times D_\uparrow$ (respectively $F \times \overline{D}_+ \times D_\uparrow$) with the concatenation

$$((f, a, b) * (f', a', b'))(n, t) := (f'(f(n, t), b(n, t)), a(n, t)a'(f(n), b(n, t)), b'(f(n, t), b(n, t))).$$

The set H is $\overline{N} \times D$ (respectively $\overline{\mathbb{N}} \times \overline{D}_+$) with the transitive operation

$$((f, a, b) * c)(n, t) := a(n, t)c(f(n, t), b(n, t)).$$

Define the edge length $L_{v,vi}$ of the edge (v, vi) by (I^v, A_i^v, B_i^v) and the vertex weight C_v of the vertex v by C^v. Let L_v^w be the recursively defined path length, where we use $L_\emptyset^v = (id, 1, id)$ with id the corresponding identity as the neutral element. Define as before

$$R_m = \sum_{|v|<m} L_v * C_v$$

Then, supposing μ_n is well defined, $R_m(n, \cdot)$ converges for fixed n to some limiting rv $R(n, \cdot)$ with distribution μ_n.

Notice, that we use partly the same symbols for the new weighted branching process. The old one of the previous section is contained as $(A_i(\infty, \cdot), B_i(\infty, \cdot), \infty)_i, C(\infty, \cdot))$ with an additional index ∞. We shall also use $R(\infty, \cdot)$ for the old $R(\cdot)$. If there is no confusion, we will drop the ∞, like in $A_i(t), B_i(t), C(t)$ and so on.

Now to the basic conditions on the rvs.

Assumption **Approx**: $((A_i(n, \cdot), B_i(n, \cdot), I_i(n, \cdot))_i, C(n, \cdot)), n \in \overline{\mathbb{N}}$ satisfy

$$\limsup_n \sup_{t \in I} E|C(n, t) - C(t)| = 0$$

$$\limsup_n \sup_{t \in I} E \sum_{i \in \mathbb{N}} |A_i(n, t) - A_i(t)| = 0$$

$$\lim_{y \to 0} \limsup_n \sup_{t \in I} E \sum_{i \in \mathbb{N}} A_i(n, t) \mathbb{1}_{|B_i(n,t) - B_i(t)| > y} = 0$$

$$\lim_{x \to \infty} \limsup_n \sup_t E \sum_i A_i(n, t) \mathbb{1}_{I_i(n,t) < x} = 0.$$

Theorem 10 *Assume R exists with path in D and R is continuous in mean. Further assume $E\overline{A} < 1$, $E\overline{C} < \infty$ and the assumption Approx.*

Then the finite dimensional distributions of $R(n,\cdot)$ converge to the ones of R. Further $\sup_t E|R(t) - R(n,t)|$ is finite and converges to 0 as $n \to \infty$. The family $R_m(n,t),\ t \in I, m,n \in \overline{\mathbb{N}}$ is uniformly integrable.

Proof: Let $d_n = \sup_t E|R(n,t) - R(t)|$. Then

$$\begin{aligned}
d_n &\leq \sup_t E|\sum_i A_i(n,t)R^i(I_i(n,t), B_i(n,t)) + C(n,t) - \sum_i A_i(t)R^i(B_i(t)) - C(t)| \\
&\leq \sup_t E|C(n,t) - C(t)| + \sup_t E\sum_i |A_i(n,t)||R^i(I_i(n,t), B_i(n,t)) - R^i(B_i(n,t))| \\
&\quad + \sup_t E\sum_i |A_i(n,t) - A_i(t)||R^i(B_i(n,t))| + \sup_t E\sum_i |A_i(t)||R^i(B_i(n,t)) - R^i(B_i(t))| \\
&= I + II + III + IV.
\end{aligned}$$

By our assumptions the first term and the third

$$III \leq E(\overline{R}) \sup_t E\sum_i |A_i(n,t) - A_i(t)|$$

converge to 0 as $n \to \infty$. The fourth, using the e-notation of Theorem 3,

$$\begin{aligned}
IV &\leq \sup_t E(\sum_{i\in\mathbb{N}} |A_i(t)|\tilde{e}(B_i(n,t), B_i(t))\mathbb{1}_{|B_i(n,t)-B_i(t)|\leq y}) \\
&+ 2E(\overline{R})\sup_t E(\sum_i |A_i(t)|\mathbb{1}_{|B_i(n,t)-B_i(t)|>y}) \\
&\leq \tilde{e}(y)E(\overline{A}) + 2E(\overline{R})E(\sum_i |A_i(t) - A_i(n,t)|\mathbb{1}_{|B_i(n,t)-B_i(t)|>y}) + E(\sum_i |A_i(n,t)|\mathbb{1}_{|B_i(n,t)-B_i(t)|>y})
\end{aligned}$$

is arbitrary small for y sufficiently small and then n sufficiently large.

It remains the term II. Define $d_{m,n} = \sup_{m<i<n} d_i$ and $r_{m,n} := d_{0,m} \sup_t E(\sum_i |A_i(n,t)|\mathbb{1}_{I_i(n,t)\leq m})$.

$$\begin{aligned}
II &\leq \sup_t E(\sum_i |A_i(n,t)|E(|R^i(I_i(n,t), B_i(n,t)) - R^i(I_i(n,t), B_i(n,t))| \mid A(n,\cdot), B(n,\cdot))) \\
&\leq \sup_t E\sum |A_i(n,t)|d_{I_i(n,t)} \\
&= \sup_t E(\sum_i |A_i(n,t)|d_{I_i(n,t)}(\mathbb{1}_{I_i(n,t)>m} + \mathbb{1}_{I_i(n,t)\leq m})) \\
&\leq d_{m,n}\sup_t E\sum_i |A_i(n,t)| + r_{m,n} \\
&\leq d_{m,n}\sup_t E\sum_i |A_i(t)| + d_{m,n}\sup_t E\sum_i |A_i(n,t) - A_i(t)| + r_{m,n}
\end{aligned}$$

We obtain together

$$d_n \le d_{m,n} E\overline{A} + d_{m,n} \sup_t E \sum_i |A_i(n,t) - A_i(t)| + r_{m,n} + o(1).$$

• The sequence $(d_n)_n$ is bounded.
Here we take $m = 0$. The above inequality has the form

$$0 \le d_n \le d_{0,n}(E\overline{A} + o(1)) + o(1).$$

From this it is easy by some epsilontic to show $(d_n)_n$ is bounded.
• $d_\infty = \limsup_n d_n = 0$.
We obtain for fixed m

$$d_\infty \le \limsup_n d_{m,n} E\overline{A} + o(1)).$$

Since d_∞ is a positive real and $E\overline{A} < 1$ we conclude $d_\infty = 0$.
As a consequence, the family $E|R(n,t)|,\ n \in \overline{\mathbb{N}},\ t \in I$ is uniformly integrable.
We come now to finite dimensional distributions. Since

$$E \sum_{s \in S} |R(n,s) - R(s)| \le |S| \sup_s E|R(n,s) - R(s)| \to_n 0$$

converges to 0 for all finite subsets S of I, we obtain convergence of the measure μ_n to μ in terms of finite dimensional distributions. Since both measure μ_n and μ are on D ($\overline{D}_+$) the sequence μ_n converges weakly to μ. □

For an application of Theorem 10 we have to find good version $((A_i(n,\cdot), B_i(n,\cdot), I_i(n,\cdot))_i, C(n,\cdot))$ of $((A_i^n, B_i^n, I_i^n)_i, C^n)$.

As explained in the introduction we will now consider processes which arise in the study of versions of the Find algorithm. Since we have a discrete recursive structure, in the mathematical sense (3),(5), we consider sequences $Y^n(\frac{l}{n})$, $l = 0, \dots, n$ of rvs satisfying $Y^0 \equiv 0$ and $\forall n \in \mathbb{N}$

$$Y^n\left(\frac{l}{n}\right) \stackrel{\mathcal{D}}{=} \sum_{i<N} A_i^n\left(\frac{l}{n}\right) Y_i^{I_i^n(\frac{l}{n})}\left(B_i^n\left(\frac{l}{I^n(\frac{l}{n})}\right)\right) \tag{16}$$

$$+C^n\left(\frac{l}{n}\right) \tag{17}$$

where N is a natural number or infinity. $C^n(\frac{l}{n}) : \Omega \to \mathbb{R}$, $A_i^n(\frac{l}{n}) : \Omega \to \mathbb{R}$, $I_i^n(\frac{l}{n}) : \Omega \to \{0, 1, 2, \dots, n-1\}$, $B_i^n(\frac{l}{n}) : \Omega \to \frac{1}{I_i^n(\frac{l}{n})}\{0, 1, \dots, I_i^n(\frac{l}{n})\}$, are measurable, Y_i^j are copies of $Y^j, j < n$ and $\big((A_i^n, B_i^n, I_i^n)_i, C^n\big), Y_k^j, k \le N, j \le n-1$ are independent.

We shall extend Y^n suitable to a map with values in D. Extend first the A_i^n, B_i^n, C^n appropriate (as step functions or by linear extension) to values D. Then define Y^n recursively point wise in $t \in [0,1]$ by

$$Y^n(t) \stackrel{\mathcal{D}}{=} \sum_i A_i^n(t) Y_i^{I_i^n(t)}(B_i^n(t)) + C^n(t) \tag{18}$$

(This procedure is compatible, i.e. does not change $Y_n(\frac{l}{n})$.) A more intuitive and explicit approach would be, to require the equation (16) as given only for $l = 0, 1, \ldots, n-1$, change the values of $B_i^n(\frac{l}{n})$ to have values in $\frac{1}{I_i^n(\frac{l}{n})}\{0, 1, \ldots, I_i^n(\frac{l}{n})-1\}$, and set $Y^n(t) := Y^n(\frac{l}{n})$ for $t \in [\frac{l}{n}, \frac{l+1}{n})$ and $Y^n(1) = Y^n(\frac{n-1}{n})$. Extend the A_i^n, B_i^n, C^n similar as step functions with values in D. Then equation (18) remains true point wise in $t \in [0, 1]$

We shall also consider the process equation

$$Y^n \stackrel{\mathcal{D}}{=} \sum_i A_i^n Y_i^{I_i^n} \circ B_i^n + C^n. \tag{19}$$

The process equation (19) gives a recursion involving also the marginal distributions. The distributional equations (18) (and also the Find algorithm) makes no statement on the joint distributions. Notice that therefore the process equation (19) is stronger than (18). It is an extension (or specific version) of the Find recursion.

For exponential moments we follow the same way as for R_n and R using condition F. The task broils down to recursive equations of max-type. All necessary tools and results for those are in the paper by Rüschendorf and Schopp [26].

8 Find and Versions

Standard Find

We will now consider versions of Find with one pivot element. (It would suffice to consider a binary tree. We take L_i as the grave for $i \geq 3$ and stay on V.) For simplicity we concentrate on one version of Find, the 3-Find. The analysis of the 2-Find is similar. (Appropriately chosen realizations for the 2-Find process will converge in the Skorodhod topology [10]. This is not true for 3-Find.) We are in the positive case.

As described in the introduction, the Find algorithm consists of two main steps:
– The choice of the pivot element;
– The splitting into two subsets.

Let $C^{n,1}(l)$ be the number of comparisons needed for choosing the pivot element and let $C^{n,2}(l)$ be the number of comparisons needed for the splitting. Let $I^n(l)$ be the rank of the pivot element in the considered list.

Then the recursive formula for the number $X^n(l)$ of comparisons in order to find the lth out of n is

$$X^n(l) \stackrel{\mathcal{D}}{=} C^n(l) + \mathbb{1}_{l<I^n(l)} X_1^{I^n(l)-1}(l) + \mathbb{1}_{l>I^n(l)} X_2^{n-I^n(l)}(l - I^n(l))$$

where $C^n(l) = C^{n,1}(l) + C^{n,2}(l)$. (In a rigorous proof, one has to show first by induction, that finding the lth smallest out of a list is a rv with a distribution depending only on l and on the size of the list, but not on the actually given input (=list). We skip this easy task of induction.)

As normalization we choose $Y^n\left(\frac{l-1}{n}\right) = \frac{1}{n}X^n(l)$. (Also $Y^n\left(\frac{l}{n}\right) = \frac{1}{n}X^n(l)$ would do the job, see the discussion at equation (18).) We obtain

$$\begin{aligned} Y^n\left(\frac{l-1}{n}\right) &:= \frac{1}{n}X^n(l) \\ &\stackrel{\mathcal{D}}{=} \frac{C^n(l)}{n} + \mathbb{1}_{l<I^n(l)} \frac{I^n(l)-1}{n} \frac{X_1^{I^n(l)-1}(l)}{I^n(l)-1} + \mathbb{1}_{l>I^n(l)} \frac{n-I_1^n(l)}{n} \frac{X_1^{n-I^n(l)}(l-I^n(l))}{n-I^n(l)}. \end{aligned}$$

We obtain the equation (16) with the coefficients

$$\begin{array}{llll} I_1^n\left(\frac{l-1}{n}\right) & := I^n(l)-1 & I_2^n\left(\frac{l-1}{n}\right) & := n-I^n(l) \\ A_1^n\left(\frac{l-1}{n}\right) & := \mathbb{1}_{l<I^n(l)}\frac{I^n(l)-1}{n} & A_2^n\left(\frac{l-1}{n}\right) & := \mathbb{1}_{l>I^n(l)}\frac{n-I^n(l)}{n} \\ B_1^n\left(\frac{l-1}{n}\right) & := \frac{l-1}{I^n(l)-1} & B_2^n\left(\frac{l-1}{n}\right) & := \frac{l-1-I^n(l)}{n-I^n(l)} \end{array}$$

Notice: $I_1^n + I_2^n \equiv n-1$.

The maximum $\overline{R}$ of Find was treated in the Section 6. It follows

$$P(\overline{R} > x) \le e^{-ax}$$

for x sufficiently large and all exponential moments of $\overline{R}$ are finite. This argument works also for the approximation.

Median versions

Median versions of Find fix some natural number m and take as pivot the median out of m randomly chosen elements of the list. For simplicity only we take m an odd number. The case $m = 1$ is the basic Find. Also for simplicity only we will draw the m numbers with replacement and uniform distribution. (Usually median versions are introduced with drawing without replacement. These versions are obtained adding few approximating arguments to the given proof. We skip the more technical, but standard details.)

In order to find the median out of m we apply any reasonable algorithm, which uses no more than m^2 comparisons. We obtain the bounds $n-1 \le C^n \le n+m^2$.

We need some notation on order statistics. The empirical distribution function F_n of iid rvs $U_1,\dots,U_n$ with distribution function F is

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{U_i\le x}.$$

The empirical distribution function converges point wise to the underlying distribution function F. The expectation $EF_n(x)$ is $F(x)$ and the variance $\operatorname{var} F_n(x) = \frac{F(x)(1-F(x))}{n}$. The rank of U_i within $U_1,\dots,U_n$ is $nF_n(U_i)$. The k largest within $U_1,\dots,U_n$ is denoted by $U_{k:n}$. Notice $k = nF_n(U_{k:n})$ for continuous F. If the U are uniformly distributed, then the expectation of the kth smallest is $EU_{k:n} = \frac{k}{n+1}$ and the variance $\operatorname{var} U_{k:n} = \frac{k(k+1)}{(n+1)^2(n+2)} \le \frac{1}{n}$. We will use F_m^n for the empirical distribution of $U_i, m < i \le n$.

Coming back to Find, in order to construct nice versions of the rvs, let $U_1,U_2,U_3,\dots,$ be independent rvs with uniform distribution. Let $U_{k:m}, 1 \le k \le m$ be the order statistic of $(U_1,\dots,U_m)$. Define $U = U_{\frac{m+1}{2}:m}$ and $I^n(l) = I^n = nF_n(U) = \frac{m+1}{2} + (n-m)F_m^n(U)$. The U corresponds to the pivot element of the median m-version of Find and I^n is the rank of the pivot element for a list of length n. Notice the rank depends not on l. The quotient $\frac{I^n}{n}$ converges almost surely to U. The expectation of $(\frac{I^n}{n} - U)^2$ converges to 0. (Argue by a.s. convergence and boundedness of the rvs or do the calculation.)

With these coefficients (8) we expect for the limiting fixed point equation (8) the following parameters:

$$C(t) = 1 \quad A_1(t) := \mathbb{1}_{t<U}U \quad A_2(t) = \mathbb{1}_{t\ge U}(1-U) \quad B_1(t) = \frac{t}{U} \quad B_2(t) = \frac{t-U}{1-U}$$

Corollary 11 *The above version of the Find process Y^n converges weakly to a limit Y. All finite-dimensional distributions converge and are uniformly integrable. The distribution of the limit Y is the fixed point of the operator K as given above.*

Proof: We have to show the assumptions of Theorem 10. The costs C are easily controlled,

$$\sup_{t\in I} E|C^n(t)-C(t)| \leq \frac{1}{n} \to_n 0$$

We give now only the estimates involving the A_1^n part. The argument for the A_2 part follows the same line. (Notice the symmetry of the uniform distribution around $1/2$ and for the algorithm, the lth smallest is the $(n-l+1)$th largest.)

We use the notation as above. Define $l^n(t)=\lfloor tn\rfloor+1$. Estimate for some $\epsilon>0$

$$\begin{aligned}
\sup_{t\in I} E|A_1^n(t)-A_1(t)| &= \sup_t E|\mathbb{1}_{l^n(t)<I^n(l^n(t))}\frac{I^n(l^n(t))-1}{n} - \mathbb{1}_{t<U}U| \\
&\leq \sup_t E|\mathbb{1}_{\lfloor tn\rfloor+1<I^n}\frac{I^n}{n} - \mathbb{1}_{t<U}U| + \frac{1}{n} \\
&\leq P(\lfloor tn\rfloor+1<I^n,\ t\geq U) + P(\lfloor tn\rfloor+1\geq I^n,\ t<U) + \frac{1}{n} \\
&\leq \sup_t (2P(|U-t|\leq\epsilon) + P(\lfloor tn\rfloor+1<I^n,\ t\geq U, |U-t|>\epsilon) + \frac{1}{n} \\
&\quad +P(\lfloor tn\rfloor+1\geq I^n,\ t<U, |U-t|>\epsilon)) + \frac{1}{n} \\
&\leq 4\epsilon + \sup_t (P(tn<I^n, t-U>\epsilon) + P(tn+1\geq I^n,\ U-t>\epsilon) + \frac{1}{n} \\
&\leq 4\epsilon + \sup_t (P(|\frac{I^n}{n}-U|>\epsilon) + P(|U-\frac{I^n-1}{n}|>\epsilon) + \frac{1}{n} \\
&\leq 4\epsilon + \frac{E(\frac{I^n}{n}-U)^2}{\epsilon^2} + \frac{E(U-\frac{I^n-1}{n})^2}{\epsilon^2} + \frac{1}{n} \\
&\to_{n\to\infty} 4\epsilon \to_{\epsilon\to 0} 0
\end{aligned}$$

For the third condition argue for $\epsilon>0$ and sufficiently large n

$$\begin{aligned}
\sup_{t\in I} EA_1^n(t)\mathbb{1}_{|B_1^n(t)-B^n(t)|>y} &= \sup_t E\mathbb{1}_{l^n(t)<I^n}\frac{I^n-1}{n}\mathbb{1}_{|\frac{l^n(t)-1}{I^n-1}-\frac{t}{U}|>y} \\
&\leq \sup_t E\mathbb{1}_{\epsilon<U}\mathbb{1}_{|\frac{l^n(t)-1}{I^n(l^n(t))-1}-\frac{t}{U}|>y} + P(U\leq\epsilon) \\
&\leq \sup_l P(\epsilon<U,\ 2l|U-I^n|>y\epsilon^2)+\epsilon \\
&\to_{n\to\infty} \epsilon \to_{\epsilon\to 0} 0
\end{aligned}$$

The last condition $\lim_{x\to\infty}\limsup_n \sup_t E\sum_i A_i^n(t)\mathbb{1}_{I_i^n(t)<x}=0$ of Theorem 10 is obviously satisfied and Theorem 10 provides the statement. □

The last corollary covers also the standard 3-version. Notice, the choice of I^n depends not on l.

Optimal adapted Find

Adapted versions of the Find algorithm take a smarter approach for selecting the pivot element. This time the choice of $I^n(l)$ will depend on l. Choose a sequence $m_n \to \infty$ of natural numbers and choose the pivot element as the $k^n(l)$ smallest element of m_n random elements chosen with uniform distribution. For simplicity only we shall take m_n is odd and draw the m_n elements without replacement. (Asymptotically other settings are equivalent and can be treated following our setting.)

We give now one adapted, asymptotically optimal version in the sense of average comparisons. Choose m_n odd of the order $\ln n$ by $|m_n - \ln n| \leq 1$. With any reasonable algorithms we can obtain the bounds $n - 1 \leq C^n \leq n + (m_n)^2$. Let $U_i, i \in \mathbb{N}$ be independent rvs with a uniform distribution on the unit interval. Choose $\epsilon^n(l), n \in \mathbb{N}, 1 \leq l \leq n$, satisfying $\lim_{n\to\infty} \sup_{1\leq l\leq n} \epsilon^n(l) = 0$ and $\lim_{n\to\infty} \sup_l \frac{1}{m_n(\epsilon^n(l))^2} = 0$ such that $k^n(l)$ defined by $0 < \frac{k^n(l)}{m_n+1} - \frac{l}{n} =: \epsilon^n(l)$ for $l \leq \frac{n}{2}$ and $0 \leq \frac{l}{n} - \frac{k^n(l)}{m_n+1} =: \epsilon^n(l)$ for $l > \frac{n}{2}$, are natural numbers. Such values exists.

Let $I^n(l)$ be the rank of $U = U_{k^n(l):m_n}$ in the sequence $U_1, U_2, \ldots, U_n$. (This is the rank of the pivot element in the considered list.) The coefficients are as in (8) and the limiting fixed point equation (8) will have the following parameters:

$$C(t) \equiv 1 \quad A_1(t) \equiv \mathbb{1}_{t<\frac{1}{2}} t \quad A_2(t) \equiv \mathbb{1}_{t\geq\frac{1}{2}}(1-t) \quad B_1(t) \equiv 1 \quad B_2(t) \equiv 0$$

Corollary 12 *Corollary 11 is true for this adapted Find version. The limit Y is a degenerate process and takes the deterministic function*

$$t \mapsto 1 + \min\{t, 1-t\}$$

as value.

Proof: Apply Theorem 10. We verify the assumptions of Theorem 10.

$$\sup_t E\left|\frac{C^n(t)}{n} - C(t)\right| \leq \frac{1 + (m_n)^2}{n} \to_n 0.$$

We give now only the estimates involving the A_1^n part and $l \leq \frac{n}{2}$. The argument for the A_2^n part follows the same line. (Notice again the symmetry of the distribution of the uniform distribution around $1/2$ and also the symmetry of the algorithm around $n/2$.)

We will use $EI^n(l) = \frac{m_n+1}{2} + (n - m_n)EU$ and the variance is

$$\text{var } I^n(l) = (n - m_n)^2 \text{ var } U + (n - m_n)EU(1-U) \leq \frac{n^2}{m_n} + n$$

where we used $U = U_{k^n(l):m_n}$. Notice

$$\begin{aligned} E\frac{I^n(l)}{n} - \frac{l}{n} &= \epsilon^n(l) + \frac{m_n + 1}{2n} - \frac{m_n}{n}\frac{k^n(l)}{m_n + 1} \\ &= \epsilon^n(l)(1 - \frac{m_n}{n}) + \frac{(n-2l)m_n + n}{2n^2} \geq \frac{\epsilon^n(l)}{2} + \frac{1}{n} > 0 \end{aligned}$$

for $n \geq 3$. Therefore

$$\sup_{l \leq \frac{n}{2}} P(I^n(l) < l) \leq \sup_{l \leq \frac{n}{2}} \frac{\text{var } I^n(l)}{(EI^n(l) - l)^2} \leq \sup_{l \leq \frac{n}{2}} \frac{4(1 + \frac{m_n}{n})}{m_n(\epsilon^n(l) + \frac{2}{n})^2} \rightarrow_n 0. \tag{20}$$

Define $l^n(t) = \lfloor tn \rfloor + 1$ and notice $l^n(t) \leq \frac{n}{2}$ when $t < \frac{1}{2}$ and $l^n(t) > \frac{n}{2}$ when $t \geq \frac{1}{2}$.

$$\begin{aligned}
\sup_{t \in I} E|A_1^n(t) - A_1(t)| &\leq \sup_t E|\mathbb{1}_{l^n(t) \leq \frac{n}{2}} \mathbb{1}_{l^n(t) < I^n(l^n(t))} \frac{I^n(l^n(t)) - 1}{n} - \mathbb{1}_{t < \frac{1}{2}} t| \\
&\leq \sup_{l \leq \frac{n}{2}} E \left| \mathbb{1}_{l < I^n(l)} \frac{I^n(l) - 1}{n} - \frac{EI^n(l)}{n} \right| + \sup_{t < 1/2} E \left| t - \frac{EI^n(l^n(t))}{n} \right| \\
&\leq \sup_{l \leq \frac{n}{2}} E\mathbb{1}_{l < I^n(l)} \left| \frac{I^n(l) - EI^n(l)}{n} \right| + \sup_{t < 1/2} E|t - l^n(t)| + \sup_{l \leq \frac{n}{2}} \left| E\frac{I^n(l)}{n} - \frac{l}{n} \right| + \frac{1}{n} \\
&\leq \sup_{l \leq \frac{n}{2}} \left(E\mathbb{1}_{l < I^n(l)} \frac{I^n(l) - EI^n(l)}{n} + E\mathbb{1}_{l \geq I^n(l)} \frac{EI^n(l)}{n} \right) + \frac{2}{n} + \sup_l \epsilon^n(l) \\
&\leq \sup_{l \leq \frac{n}{2}} \left(\sqrt{\text{var } \frac{I^n(l)}{n}} + E\mathbb{1}_{l \geq I^n(l)} \right) + \frac{2}{n} + \sup_l \epsilon^n(l) \\
&\leq \frac{1}{\sqrt{m_n}} + \sup_{l \leq \frac{n}{2}} P(l \geq I^n(l)) + \frac{3}{n} + \sup_l \epsilon^n(l) \\
&\rightarrow_n 0.
\end{aligned}$$

For the third assumption use the convention $\frac{0}{0} = 1$,

$$\begin{aligned}
\sup_t EA_1^n(t)\mathbb{1}_{|B_1^n(t) - B^n(t)| > y} &= \sup_{1 \leq l \leq \frac{n}{2}} E\mathbb{1}_{l < I^n(l)} \frac{I^n(l) - 1}{n} \mathbb{1}_{|\frac{l-1}{I^n(l)-1} - 1| > y} \\
&\leq \sup_{1 \leq l < \epsilon n} E\frac{I^n(l) - 1}{n} + \sup_{\epsilon n < l \leq \frac{n}{2}} E\mathbb{1}_{l < I^n(l)} \mathbb{1}_{|l - I^n(l)| > y(I^n(l) - 1)} \\
&\leq E\frac{I^n(\epsilon n)}{n} + \sup_{\epsilon n < l \leq \frac{n}{2}} P(l < I^n(l),\ I^n(l) - l > y(I^n(l) - 1)) \\
&\leq \frac{m_n + 1}{2n} + \frac{n - m_n}{n} \frac{k^n(\epsilon n)}{m_n + 1} + \sup_{\epsilon n \leq l \leq \frac{n}{2}} P\left(I^n(l) - l > \frac{y(l-1)}{1-y} \right) \\
&\leq \frac{2m_n}{n} + \frac{n - m_n}{n} \left(\frac{l}{n} + \epsilon^n(\epsilon n) \right) + \sup_{\epsilon n \leq l \leq \frac{n}{2}} \frac{E(I^n(l) - l)^+(1-y)}{y(l-1)} \\
&\leq \frac{2m_n}{n} + \epsilon^n(\epsilon n) + \frac{1-y}{y\epsilon} \sup_{\epsilon n \leq l \leq \frac{n}{2}} \left(E\frac{I^n(l) - l}{n} + E\mathbb{1}_{l > I^n(l)} \frac{l}{n} \right) \\
&\leq \frac{4m_n}{n} + \frac{1-y}{y\epsilon} \sup_{\epsilon n \leq l \leq \frac{n}{2}} (\epsilon^n(l) + P(l > I^n(l))) \\
&\rightarrow_n 0
\end{aligned}$$

Further for all $x > 0$

$$\sup_{t<\frac{1}{2}} EA_1^n(t)\mathbb{1}_{I_i^n(t)<x} \leq \sup_{1\leq l\leq \frac{n}{2}} E\mathbb{1}_{I^n(l)-1<x} = E\mathbb{1}_{I^n(1)-1<x} \to_n 0.$$

Since the assumptions of Theorem 10 are satisfied, the finite dimensional distributions of Y^n of the adapted version of Find converge to those of Y with

$$(Y(t))_t \sim \mathcal{D}{=}(\mathbb{1}_{t<\frac{1}{2}} tY(0) + \mathbb{1}_{t\geq\frac{1}{2}}(1-t)Y(1)+1)_t.$$

The process equation implies $Y(0) = Y(1) = 1$ and we obtain the limiting process $(Y(t))_t$ is degenerate to a constant value, the function $t \mapsto \mathbb{1}_{t<\frac{1}{2}} t + \mathbb{1}_{t\geq\frac{1}{2}}(1-t) + 1 = 1 + \min\{t, 1-t\}$. □

Remark 13 *Here we face as additional properties, that the A_i, B_i and C are degenerate random variables. The B_i and C are also independent of s.*

References

[1] David J. Aldous and Antar Bandyopadhyay. A survey of max-type recursive distributional equations. *Annals of Applied Probability*, 15:1047–1110, 2005.

[2] Gerold Alsmeyer and Uwe Roesler. A stochastic fixed point equation for weighted minima and maxima. Preprintreihe angewandte mathematik und informatik 01/05-s, Universität Münster, 2005. `http://wwwmath.uni-muenster.de/inst/statistik/publ/Preprintlisten/index.html` `http://www.math.uni-kiel.de/stochastik/roesler/research.html`.

[3] D. Anderson and R. Brown. Combinatorial aspects of C.A.R. Hoare's find algorithm. *Australian Journal of Combinatorics*, 5:109–119, 1992.

[4] Peter J. Bickel and David A. Freedman. Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9:1196–1217, 1981.

[5] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, 1968.

[6] Lester Dubins and Marjorie Hahn. A characterization of the families of finite-dimensional distributions associated with countable additive stochastic processes whose sample paths are in d. *Zeitschrift Wahrscheinlichkeitstheorie verwandte Gebiete*, 43:97–100, 1978.

[7] Kord Eickmeyer and Ludger Rüschendorf. A limit theorem for recursively defined processes in l^p.. *Statistics & Decisions*, 28:1001–1099, 2007.

[8] Robert W. Floyd and Ronald L. Rivest. Expected time bounds for selection. *Communications of the ACM*, 8:165–173, 1975.

[9] R. Grübel. On the median-of-k version of Hoare's selection algorithm. *Theoretical Informatics and Applications*, 33(2):177–192, 1999.

[10] Rudolf Grübel and Uwe Rösler. Asymptotic distribution theory for Hoare's selection algorithm. *Advances in Applied Probability*, 28:252–269, 1996.

[11] Charles A.R. Hoare. PARTITION (Algorithm 63);QUICKSORT (Algorithm 64);FIND (Algorithm 65). *Communication of the Association for Computing Machinery*, 4:321–322, 1961.

[12] Peter Jagers and Uwe Roesler. Stochastic fixed points for the maximum. In M. Drmota, P. Flajolet, D. Gardy, and B. Gittenberger, editors, *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and probabilities*, pages 325–338. Birkhäuser Verlag Basel, trend in mathematics edition, 2004.

[13] Diether Knof. *Struktur von Fixpunkten aus Prozessgleichungen.* PhD thesis, University of Kiel, 2007.

[14] Donald E. Knuth. Mathematical analysis of algorithms. *Information Processing*, 1:19–27, 1972.

[15] Conrado Martínez and Uwe Roesler. Partial quicksort and quickpartitionsort. Submitted to *CRM*, see Citeseer, 2009.

[16] Conrado Martínez and Salvador Roura. Optimal sampling strategies in quicksort and quickselect. *SIAM Journal Computing*, 31(3):683–705, 2001.

[17] Ralf Neininger and Ludger Rueschendorf. Analysis of algorithms by the contraction method: additive and max-rexursive sequences. In A. Greven J.D. Deuschel, editor, *Interacting stochastic Systems*, pages 435–450. Springer, Heidelberg, 2005.

[18] Ralf Neininger and Ludger Rueschendorf. On stochastic recursive equations of sum- and max-type. *Journal of Applied Probability*, 43:687–701, 2006.

[19] Conrado Martínez; Daniel Panario and Alfredo Viola. Adaptive sampling strategies for quickselect. *ACM Trans. on Algorithms*, 2008. accepted.

[20] Helmut Prodinger Peter Kirschenhofer and Conrado Martínez. Analysis of hoare's find algorithm with median-of-three partition. *Random Structures and Algorithms*, 10:143–156, 1997.

[21] M. Blum; R.W. Floyd; V. Pratt; R.L. Rivest and R.E. Tarjan. Time bounds for selection. *Journal of Computer and System Sciences*, 7:448–461, 1973.

[22] Uwe Roesler. A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 42:195–214, 1992.

[23] Uwe Roesler. *Dynamics of complex and irregular systems, Bielefeld Encount. Math. Phys. VIII, Bielefeld 1991*, chapter The weighted branching process., pages 154–165. World Sci. Publishing, River Edge, NJ, 1993.

[24] Uwe Roesler. Quickselect revisited. *Journal of the Iranian Statistical Institute*, 3:271–296, 2004.

[25] Uwe Roesler and Ludger Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29:3–33, 2001.

[26] Ludger Rüschendorf and Eva-Maria Schopp. Exponential tail bounds for max-recursive sequences. *Discrete Mathematics and Theoretical Computer Science DMTCS*, 9:333–352, 2007.