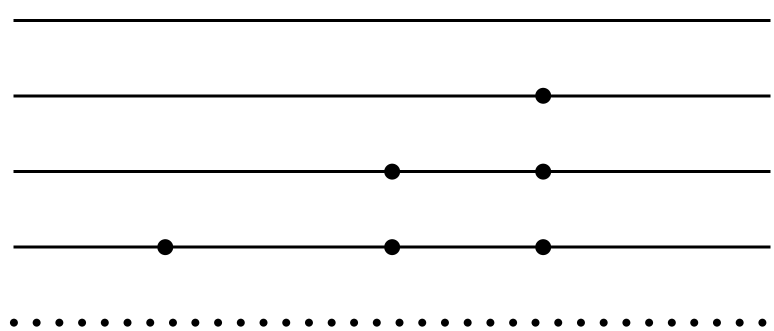
The QUICKSORT PROCESS

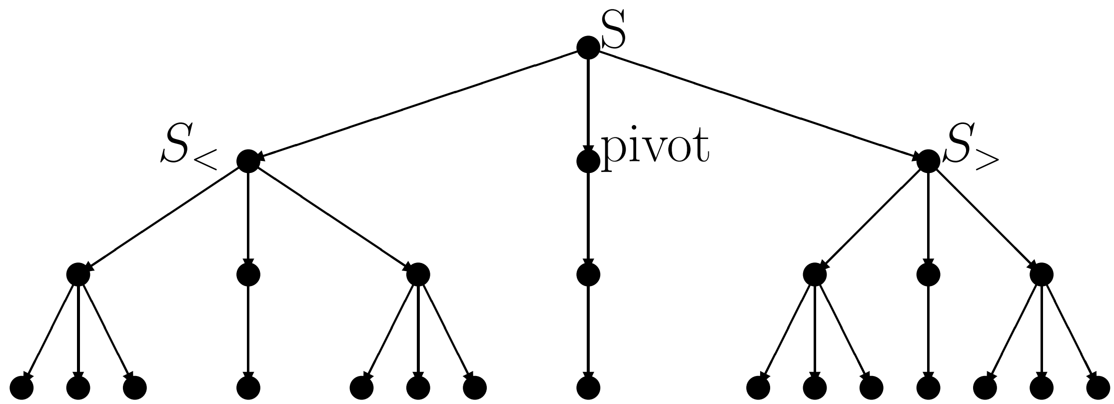
Mahmoud Ragab and Uwe Roesler

– Online Quicksort streaming
– Mathematical formulation
– Results: Quicksort process in D
– Abstraction and Reflection
– Embedding into weighted branching process

QUICKSORT STREAMING

Algorithm: QUICKSORT STREAMING
Input: Set S of n different reals.
Output: Online the smallest, then second smallest and so on.
Procedure: Divide like in Quicksort, continue with left list.
– Choose random pivot within S
– Split S into $S_<, \{ \text{ pivot } \}, S_>$, stored in this order.
– Continue with $S_<$ recursively unless empty.
– If $|S_<| = 1$ output the element. Continue with next list.




S
$S_<$
pivot
$S_>$

$$78, 90, 46, 60, 88, 24, 95, 47, 98$$

$$78, 46, 60, 24, 47, \mathbf{88}, 90, 95, 98$$

$$24, \mathbf{46}, 78, 60, 47, \mathbf{88}, 90, 95, 98$$

publish 24 and 46

$$60, 47, \mathbf{78}, \mathbf{88}, 90, 95, 98$$

$$\mathbf{47}, 60, \mathbf{78}, \mathbf{88}, 90, 95, 98$$

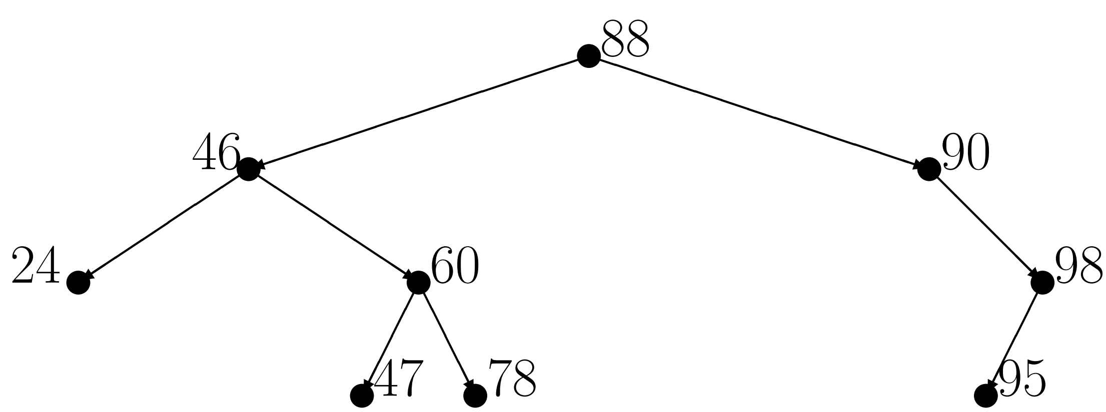
publish 47, 60, 78, 88,

$$\mathbf{90}, 95, 98$$

publish 90

$$95, \mathbf{98}$$

publish 95 and 98



RECURSION

Let $X(S,l)$ be number of comparisons up to l-th smallest

$$\begin{aligned}(X(S,l))_l &= (n-1+\mathbb{1}_{I<l}(X(S_<,I-1)+X(S_>,l-I)) \\ &+ \mathbb{1}_{l<I}X_1(S_<,l))_l\end{aligned}$$

$l=1,2,\ldots,|S|$. I uniform distribution and independent of all X-rvs. $I=|S_<|+1$ denotes rank of pivot after comparisons.

$X(S,|S|)$ has Quicksort distribution.

Replace S by permutation.

Internal randomness: $X(S,\cdot)$ has distribution $X(|S|,\cdot)$ depending on cardinality.

External randomness: Input permutation with uniform distribution or iid rvs taking the first as pivot. Randomness in input, performance deterministic.

In either case $(X(S,l))_l \stackrel{\mathcal{D}}{=} (X(|S|,l))_l$ and

$$\begin{aligned}(X(n,l))_l &\stackrel{\mathcal{D}}{=} (n-1+\mathbb{1}_{I\le l+1}(X_1(I-1,I-1)+X_2(n-I,l-I)) \\ &+ \mathbb{1}_{I>l+1}X_1(I-1,l))_l\end{aligned}$$

where $I=I(n)$. This determines the distribution.

EXPECTATION

$a(n,l) = EX(n,l)$

$$a(n,l) = fct(n, (a_i)_{i<n}).$$

C. Martinez, Partial Quicksort

$$a(n,l) = 2n + 2(n+1)H_n - 2(n+3-l)H_{n+1-l} - 6l + 6.$$

H_n n-th harmonic number $\Sigma_{i=1}^{n} \frac{1}{i}$.

DISTRIBUTION

$$Y(n, \frac{l}{n}) := \frac{X(n,l) - a(n,l)}{n+1}$$

$I = I(n)$

$$\begin{aligned} Y(n,\cdot) \overset{\mathcal{D}}{=} \ & (\mathbb{1}_{I \le l+1}(\frac{I}{n+1} Y_1(I-1, I-1) + \frac{n-I+1}{n+1} Y_2(n-I, \frac{l-I}{n-I})) \\ + \ & \mathbb{1}_{I > l+1} \frac{I}{n+1} Y_1(I-1, \frac{l}{I-1}) + C(n, \frac{l}{n}, I))_l \end{aligned}$$

where

$$\begin{aligned} C(n, \frac{l}{n}, I) \ = \ & \frac{1}{n+1} (\mathbb{1}_{I \le l+1}(a(I-1, I-1) + a(n-I, l-I)) \\ + \ & \mathbb{1}_{I > l+1} a(I-1, l) - a(n,l) + n - 1 \end{aligned}$$

WISHFUL THINKING

Extend $Y(n,\cdot)$ nicely (piece wise constant or linear) to a rv Y_n with values in D

If $Y_n \to Y$ in some sense

$$(Y(t))_{t\in[0,1]} \stackrel{\mathcal{D}}{=} (\mathbb{1}_{U\le t}(UY_1(1)+(1-U)Y_2(\frac{t-U}{1-U}))+\mathbb{1}_{U>t}UY_1(\frac{t}{U})+C(U,t)$$

on D. Y_1, Y_2, U independent, U uniformly on $[0,1]$.

$$\begin{aligned} C(x,t) &= 1+2x\ln x+2(1-x)\ln(1-x) \\ &\quad +2\mathbb{1}_{x\ge t}((1-t)\ln(1-t)+(1-x)\ln(1-x)+1) \\ &\quad -(x-t)\ln(x-t) \end{aligned}$$

where $C(n,\cdot,I) \to_n C(U,\cdot)$

QUICKSORT PROCESS

Theorem Ragab-R.
There exists a fixed point Y as above with values in D.

Y is Quicksort process. Observation is a path, ω fixed

$$[0,1] \ni t \mapsto Y(\omega)(t)$$

Proof via weighted branching process (plus some trick) on $(D, \|\cdot\|_\infty)$ and L_2. Neininger via Zolotarev ζ_2.

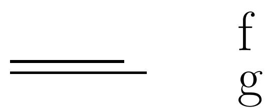
DISCRETE QP

Theorem Ragab-R.

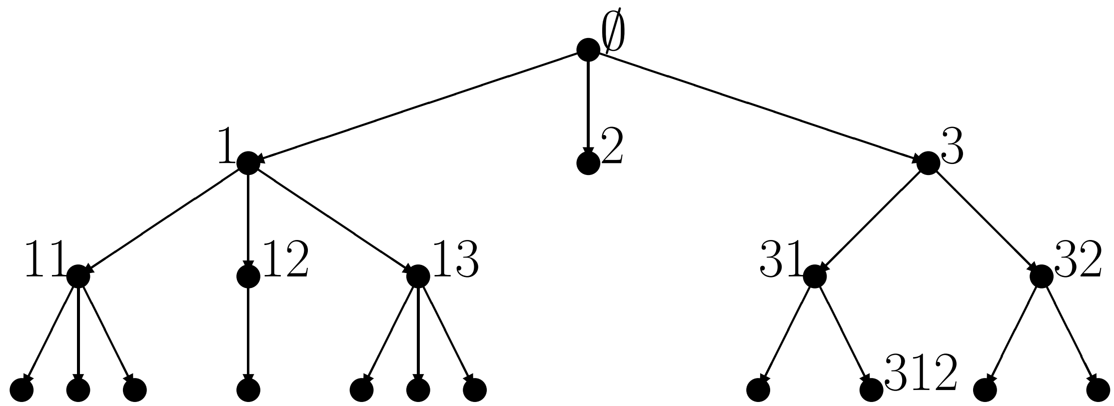
There are nice versions of Y_n converging in Skorodhod metric on D to Y. a.e.

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

where Λ is the set of all bijective increasing functions $\lambda : [0,1] \to [0,1]$.



Convergence: $E(f(Y_n)) \to E(f(Y))$ for all bounded continuous functions on D.


∅
1
2
3
11
12
13
31
32
312

MEASURES versus FUNCTIONS

Structure: V rooted tree

$v \mapsto Z^v$ defined via tree vV.

Transformation φ^v such that

$$Z^v = \varphi^v((Z^{vi})_i)$$

Objects $(Z^v, \varphi^v)_v$ as tree indexed process

$$(Z_n^v, \varphi_n^v)_v \to_n (Z^v, \varphi^v)_v$$

Measure theory: Z^v a measure μ^v,

– Dynamics φ^v deterministic function

– Methods: contraction method, generating function

– Convergence $\stackrel{\mathcal{D}}{=}$,

In many examples $\mu_n^\emptyset \to_n \mu^\emptyset$ like contraction method

Probability theory: Z^v a random variable

– Dynamics φ^v is itself a random function

– Methods: martingales, functional analysis

– Convergence: almost everywhere, stochastic convergence, L_p

In many examples $Z_n^\emptyset \to_n Z^\emptyset$

like in FIND Gruebel-Roesler 96, Monday Fill, Gruebel, Meiners,

COMPLEMENTARY

Which method is better?

Both help to analyse and are **complementary.**

Example: Fern $X_n = A_n X_{n-1} + B_n$ in $\mathbb{R}^2$

plotting of $(X_n)_n$ provides picture Fern.

Fern

Analysis, X_n does not converge in probabilistic sense.

But $\mu_n = X_n$ converges weakly to μ.

Argument by version $Y_n \to Y$ a.e. as rvs. (Kesten 73)

Fern is support of μ, a fixed point of stochastic fixed point equation, Burton-R 95

WEIGHTED BRANCHING PROCESS

WBP Roesler 93

A weighted branching process (WBP) is tupel $(V, (T^v, C^v)_{v\in V}, (G, *, H$

- V a rooted tree. Ulam-Harris notation.
- $v \mapsto C^v, \qquad (v, vi) \mapsto T_i^v$
- $(T^v = (T_1^v, T_2^v, \ldots), C^v),\ v \in V$, independent with values in $G \times H$.
- $(G, *)$ is measurable semi group ($* : G\times G \to G,\ *(g,h) = g*h$ associative and measurable) with neutral element e and grave $\triangle$
- $(G, *)$ operates transitive and measurable on H via $*_1 : G \times H \to H$.

Define the path weights $(v, vw) \mapsto L_v^w$ with values in G recursively by $L_\emptyset^v = e$ and

$$L_{wi}^v = L_w^v * T_i^{vw}.$$

Suppress $\emptyset$.

For our purposes, H has additional structure $+$

$$R_n = \sum_{|v|<n} L_v *_1 C^v$$

and $\varphi : G^{\mathbb{N}} \times H \times H^{\mathbb{N}} \to H$ like

$$\varphi(t, c, r) = \sum_i t_i *_1 r_i + c$$

affine linear function.

WEIGHTED BRANCHING PROCESS

Quicksort: G is multiplicative semi group $\mathbb{R}$ with neutral element $e = 1$ and grave $\triangle = 0$. G operates transitive on $H = \mathbb{R}$ by multiplication. Let $U^v, v \in V$ be independent rvs with a uniform distribution on $[0,1]$. Put

$$T_1^v = U^v, \quad T_2^v = 1 - U^v, \quad T_3^v \equiv 0 \equiv T_4^v \dots \quad C^v = C(U^v)$$

where

$$C(x) = 1 + 2x \ln x + 2(1-x)\ln(1-x).$$

Then $R_n := \Sigma_{|v|<n} L_v C^v$ is L_2-martingale.

$$R_n \to_n Q$$

a.e. and in L_2. Quicksort distribution

$$Q = UQ^1 + (1-U)Q^2 + C(U)$$

with expectation 0 and finite variance.

$R_n^v \to Q^v$ a.e. on tree vV and

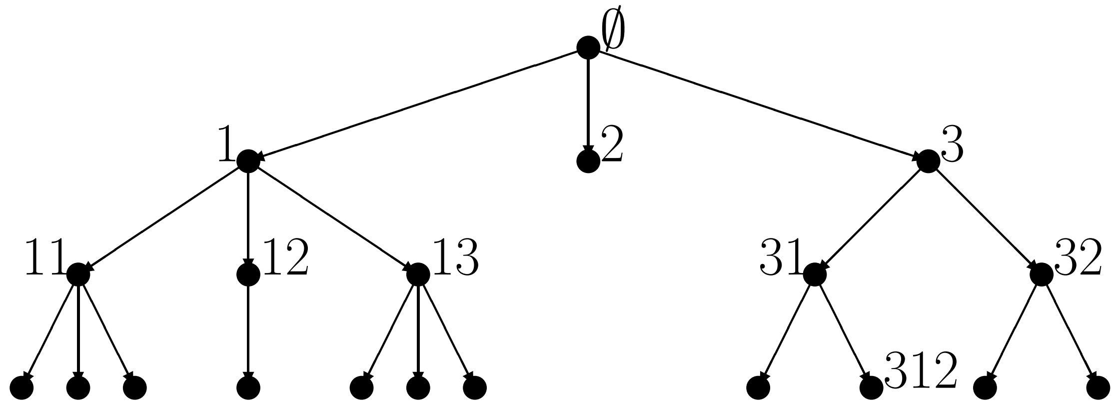
$$Q^v = U^v Q^{v1} + (1-U^v)Q^{v2} + C(U^v)$$

for every $v \in V$.

Distribution:

$$K(\mu) \stackrel{\mathcal{D}}{=} UX_1 + (1-U)X_2 + C(U))$$

$K^n(\delta_0) = \mathcal{L}(R_n) \to \mathcal{L}(Q)$


∅
1
2
3
11
12
13
31
32
312

QUICKSORT PROCESS as WBP

$G = D \times D_{\uparrow}$, neutral element $(1, \mathrm{id})$ and grave identically 0.

$(f_1, g_1) * (f_2, g_2) := (f_1 \cdot f_2 \circ g_1, g_1 \circ g_2)$.

$H = D$ and $(f, g) *_1 h := f \cdot h \circ g$.

Let (U^v, Q^v), $v \in V$ be as in the Quicksort example. Define $T = (T_1, T_2, \ldots,)$, $T_i = (A_i, B_i)$, C by

$$A_1 = U \mathbb{1}_{[0,U)}, \quad A_2 = (1-U)\mathbb{1}_{[U,1)}, \quad A_3 \equiv 0 \equiv A_4 \equiv 0$$

$$B_1(t) = 1 \wedge \frac{t}{U}, \quad B_2(t) = 0 \vee \frac{t-U}{1-U}, \quad B_3 = id = B_4 \ldots$$

$$C = A_1 Q + C(U, \cdot)$$

$C(x, t)$ as before.

(T^v, C^v) via (U^v, Q^v)

Define

$$R_n^v := \sum_{w \in V_{<n}} L_w^v *_1 C^{vw}$$

and

$$\varphi(t, c, r) = \sum_{i \in I\!N} t_i r_i + c.$$

Then $R_n \to_n R$ in sense $\| \| R_n - R \|_\infty \|_2$.

DISCRETE QP as WBP

Choose $I^v(n) = \lfloor nU^v \rfloor + 1$. That's it