Worst-case FIND: Symbol Comparisons

Jim Fill
(based on ongoing jt. work with Ph.D. student Jason Matterer)

Department of Applied Mathematics and Statistics
The Johns Hopkins University

June 13, 2011

AofA 2011 in Będlewo, Poland

Manifesto

Analysis of key-comparisons-based searching and sorting algorithms such as `QuickSelect` in terms of the number of key comparisons cannot fully quantify the complexity of the algorithms.

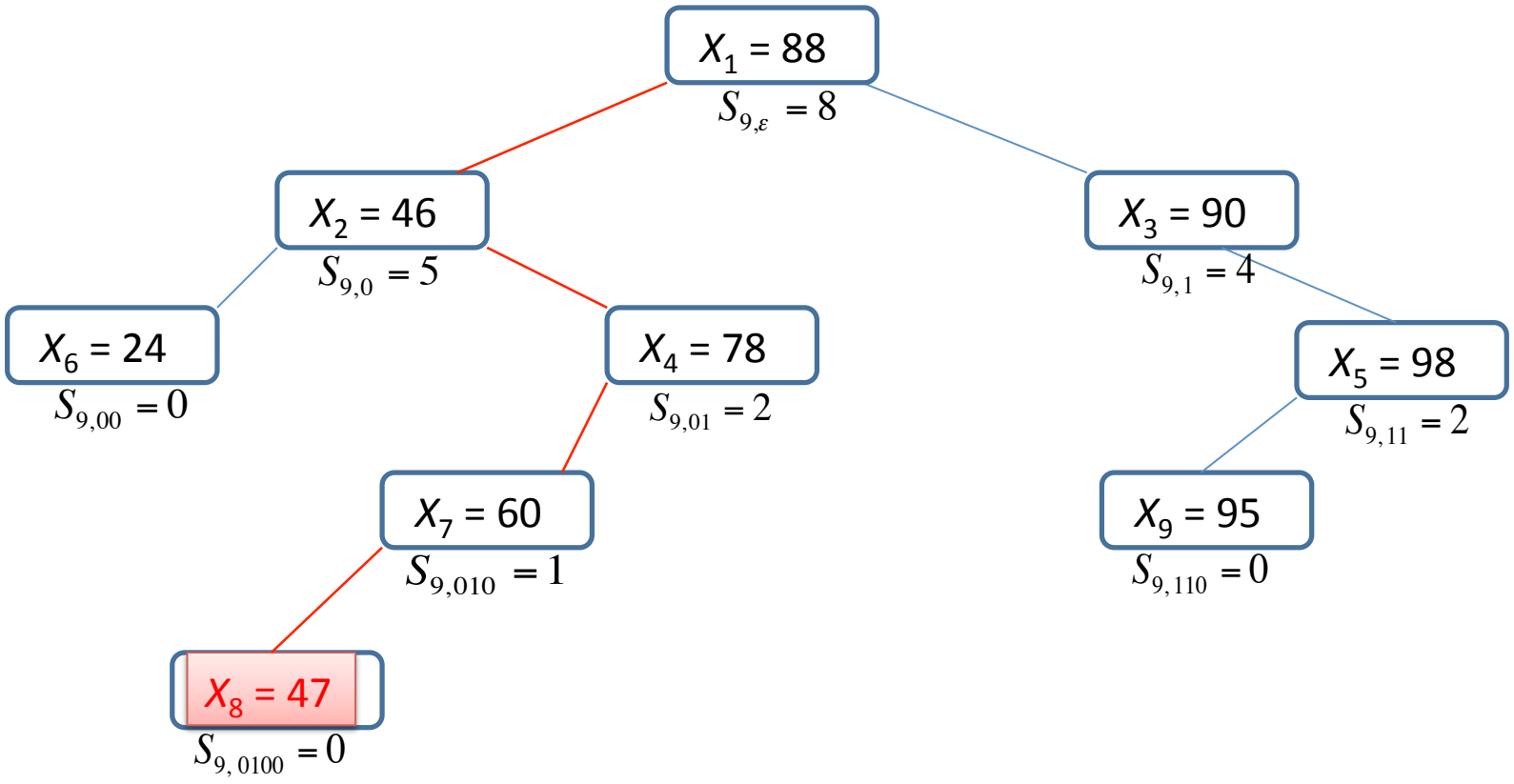
- If keys are represented as symbol strings, then individual symbols of the strings must be compared in order for `QuickSelect` to complete its task.
- Results obtained by analyzing the algorithm with respect to the number of symbol comparisons required to find a target key more accurately reflect actual execution costs.

This talk will focus on limiting distributions—mainly treating the worst case (in a certain sense) for `FIND` $\equiv$ `QuickSelect`, but also revisiting ordinary `QuickSelect` to find a key of one specified rank. Our assumptions about the costs of comparisons will be broad enough to cover key comparisons, symbol comparisons, and more.

9 keys: X_1=88, X_2=46, X_3=90, X_4=78, X_5=98, X_6=24, X_7=60, X_8=47, X_9=95

sorted: X_6=24, X_2=46, X_8=47, X_7=60, X_4=78, X_1=88, X_3=90, X_9=95, X_5=98

Worst-Case Find = QuickSelect(9, 3):
$S_9 = 8 + 5 + 2 + 1 + 0 = 16$



Setup: Natural coupling

Key observation: Because we assume the keys are iid, we may take the pivot to be the **first** key in the sequence, X_1.

- Thus if $X_1, X_2, \ldots$ is an infinite sequence of keys and we define

$$C_{n,m} := \sum_{i,j \le n} \mathbf{1}(\texttt{QSel}(n,m) \text{ compares } X_i \,\&\, X_j)\, c(X_i, X_j)$$

  using any given cost function $c \ge 0$ for comparing two keys, then we have coupled all the random variables $C_{n,m}$ (all n, m).
- We will assume throughout that this natural coupling of the random variables $C_{n,m}$ has been used.
- The coupling opens up the possibility of establishing stronger forms of convergence than convergence in distribution, such as almost sure convergence and convergence in L^p, for suitably normalized $C_{n,m}$.

Setup: Source model M and cost β

Source model(s) from Vallée, Clément, F, and Flajolet (2009):

- $\Sigma = \{0, 1, \ldots, r-1\}$ or $\{0, 1, \ldots\}$: alphabet of symbols
- $\Sigma^* = \cup_{0 \le k < \infty} \Sigma^k$: set of all "prefixes" (finite-length strings)
- Σ^∞: set of all "words" (infinite-length strings)
- A source generates each word from a seed $u \in (0, 1)$.

$M(u)$: word produced from seed u

- The source preserves order: $M(t) \prec M(u)$ if and only if $t < u$.
- $\beta(u, t)$: cost of comparing $M(u)$ and $M(t)$ to determine the order between them
  - In our main application, $\beta(u, t) = \beta_{\mathrm{symb}}(u, t)$ represents the number of symbol comparisons.
  - To recapture key comparison results, set $\beta = \beta_{\mathrm{key}} \equiv 1$.

Setup: Probabilistic source

- Probability enters: We assume seeds are iid uniform(0,1).
- $p_w :=$ probability that a word generated by the source has prefix w
- $\pi_k := \sup\{p_w : w \in \Sigma^k\}$: biggest probability of any prefix of length k
- We say that the probabilistic source is γ-tamed if there exists A such that $\pi_k \leq A(k+1)^{-\gamma}$ for every k. [This is (easily) true—for all $\gamma < \infty$—for memoryless sources and for any Markovian source such that the supremum of all one-step transition probabilities is strictly smaller than 1.]
- The tameness assumption we make on the cost β will be closely related (when $\beta = \beta_{\mathrm{symb}}$) to tameness of the source.

Setup: Tameness of cost β

- Reminder: We say that the probabilistic source is γ-tamed if there exists A such that $\pi_k \leq A(k+1)^{-\gamma}$ for every k. [This is (easily) true—for all $\gamma < \infty$—for memoryless sources and for most Markovian sources.]
- We say that a symmetric cost function $\beta \geq 0$ is ϵ-tamed for the given source if there exists $c \equiv c_\epsilon$ such that

$$\beta(t,u) \leq c\,(u-t)^{-\epsilon}$$

  for all $0 \leq t < u \leq 1$. (This is always true when $\beta \equiv 1$.)
- **Fact:** The source is Π-tamed with parameters γ and A if and only if the symbols-comparison cost function β_{symb} is ϵ-tamed with $\epsilon = 1/\gamma$ and $c = A^{1/\gamma}$.

Setup: Uniform seeds and the infinite binary search tree

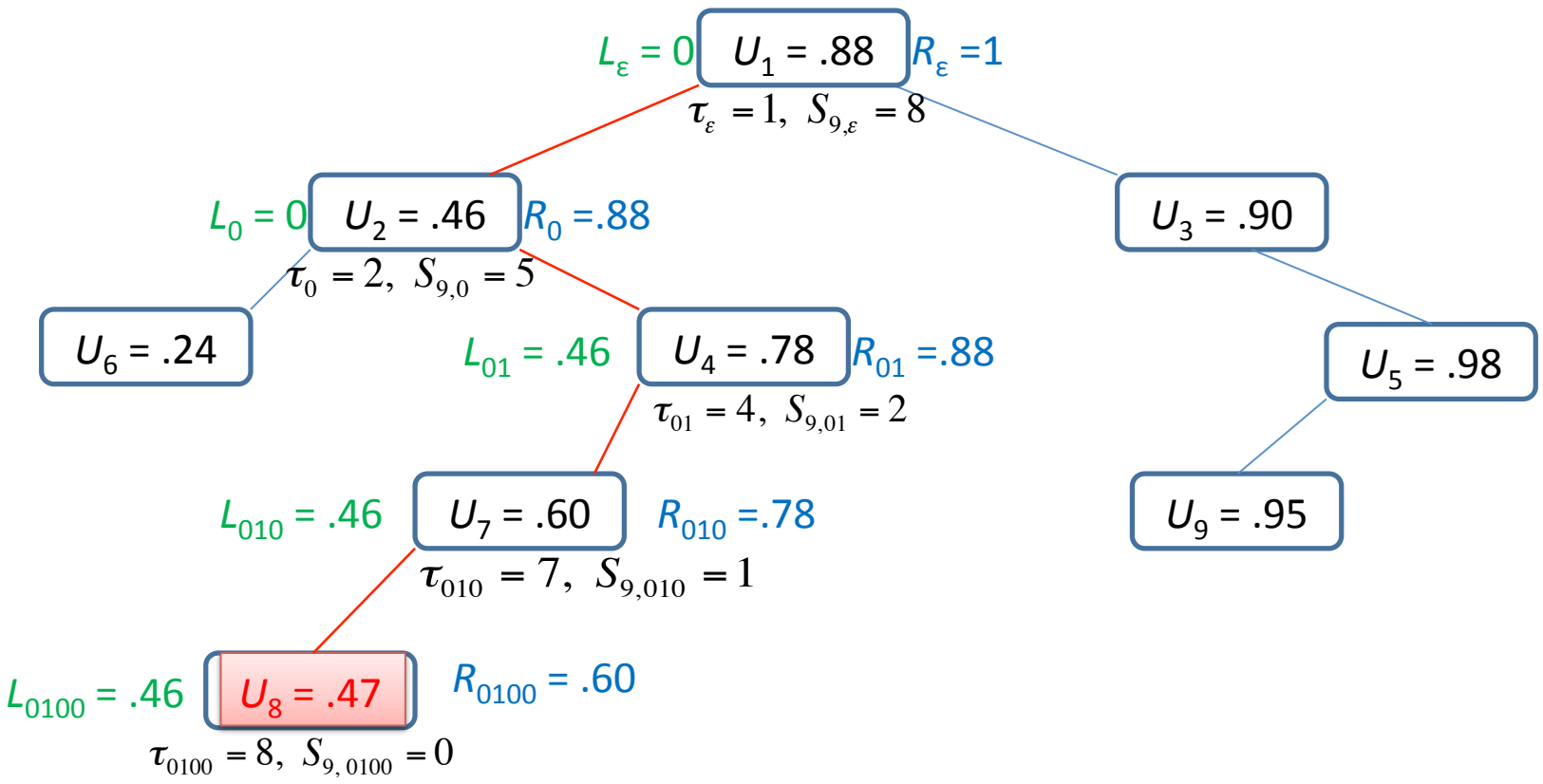
It is natural and convenient to define all $C_{n,m}$ (for all cost functions!) in terms of a single infinite sequence $(U_i)_{i\geq 1}$ of seeds that are iid uniform(0, 1).

- Label the nodes of the complete infinite binary tree as usual using finite-length binary strings σ.
- Insert $U_1, U_2, \ldots$ into the tree in the usual search-tree way.
- Let $\tau_\sigma :=$ time at which node σ becomes filled. Note that seed U_{τ_σ} is stored at node σ.
- Let $L_\varepsilon := 0$ and $R_\varepsilon := 1$. Inductively define (L_σ, R_σ) to be the interval of possible values for U_{τ_σ}, as follows.
  Let σ' denote the parent node of σ.
  If σ is a left child, let $(L_\sigma, R_\sigma) := (L_{\sigma'}, U_{\tau_{\sigma'}})$;
  if σ is a right child, let $(L_\sigma, R_\sigma) := (U_{\tau_{\sigma'}}, R_{\sigma'})$.

9 keys: U_1=.88, U_2=.46, U_3=.90, U_4=.78, U_5=.98, U_6=.24, U_7=.60, U_8=.47, U_9=.95

sorted: U_6=.24, U_2=.46, U_8=.47, U_7=.60, U_4=.78, U_1=.88, U_3=.90, U_9=.95, U_5=.98

Worst-Case Find = QuickSelect(9, 3):
$S_9 = 8 + 5 + 2 + 1 + 0 = 16$



Preliminaries: A crucial conditional iid observation

- Let $S_{n,\sigma}$ denote the total cost using `QuickSort` of comparisons of U_{τ_σ} with descendant nodes that are filled by time n:

$$S_{n,\sigma} := \sum_{i:\, \tau_\sigma < i \le n} \mathbf{1}(L_\sigma < U_i < R_\sigma)\, \beta(U_{\tau_\sigma}, U_i).$$

  Note that these costs will be charged to `QuickSelect`(n, m) if and only if node σ is visited by the algorithm.

- Define

$$I_\sigma := \int_{L_\sigma}^{R_\sigma} \beta(U_{\tau_\sigma}, u)\, du.$$

- Crucial obs.!: Conditionally given $C_\sigma := (\tau_\sigma, U_{\tau_\sigma}, L_\sigma, R_\sigma)$, the r.v.s $U_{\tau_\sigma+1}, U_{\tau_\sigma+2}, \ldots$ are iid unif$(0,1)$ and $S_{n,\sigma}$ is the sum of $(n - \tau_\sigma)^+$ iid r.v.s with mean I_σ.

Preliminaries: Conditional use of LLN

- Crucial obs.!: Conditionally given $C_\sigma := (\tau_\sigma, U_{\tau_\sigma}, L_\sigma, R_\sigma)$, the r.v.s $U_{\tau_\sigma+1}, U_{\tau_\sigma+2}, \ldots$ are iid unif$(0,1)$ and $S_{n,\sigma}$ is the sum of $(n-\tau_\sigma)^+$ iid r.v.s with mean I_σ.
- By the SLLN,
  conditionally given C_σ: $\frac{S_{n,\sigma}}{n} \xrightarrow{\text{a.s.}} I_\sigma$
  provided only that $I_\sigma < \infty$. So, unconditionally: if $I_\sigma < \infty$ a.s., then
  $$\frac{S_{n,\sigma}}{n} \xrightarrow{\text{a.s.}} I_\sigma.$$
- An only slightly less easy argument gives, for $1 \le p < \infty$, the result
  $$\frac{S_{n,\sigma}}{n} \xrightarrow{L^p} I_\sigma.$$
  provided $\mathbf{E}\, I_\sigma^p < \infty$. Later I'll discuss, in terms of tameness, when this condition holds.

The two main theorems

`QuickQuant`(n, α) : `QuickSelect`(n, m_n) for rank $m_n = \alpha n + o(n)$ (quantile approximately α) $(0 \leq \alpha \leq 1)$

Theorem (F and Nakama, for `QuickQuant`(n, α) with α fixed)

If $p \geq 1$ and the cost function β is ϵ-tame with $0 \leq \epsilon < 1/p$, then

$$\frac{1}{n} \times \text{cost of } \texttt{QuickQuant}(n, \alpha) \xrightarrow{L^p} I_\gamma \equiv \sum_{\sigma \in \gamma} I_\sigma,$$

where $\gamma \equiv \gamma(\alpha)$ is the infinite path from the root to seed α.

(corollary: convergence in law for multiple QuickQuant)

Theorem (F and Matterer, for worst-case `FIND`)

If $p > 5/2$ and the cost fcn. β is ϵ-tame with $0 \leq \epsilon < 1/p$, then

$$\frac{1}{n} \times \max_{1 \leq m \leq n} [\text{cost of } \texttt{QuickSelect}(n, m)] \xrightarrow{L^p} \sup_{\gamma \in \Gamma} I_\gamma,$$

where the sup *is taken over all infinite paths from the root.*

I-moment lemma

Recall our concern with the relationship between the condition $\mathbf{E}\, I_\sigma^p < \infty$ and tameness, with $I_\sigma := \int_{L_\sigma}^{R_\sigma} \beta(U_{\tau_\sigma}, u)\, du$. More generally, for $0 \le r < \infty$ define

$$I_{r,\sigma} := \int_{L_\sigma}^{R_\sigma} \beta^r(U_{\tau_\sigma}, u)\, du.$$

We'll need

Lemma (I-moment lemma)

Suppose that the cost function β is ϵ-tame with $0 \le \epsilon < 1/r$, and let $0 \le s < \infty$. Then there exists a constant $c_{\epsilon,r,s}$ such that for every σ in Λ_k (i.e., level k) we have

$$\mathbf{E}\, I_{r,\sigma}^s \le c_{\epsilon,r,s}(s + 1 - sr\epsilon)^{-k}.$$

Proof of I-moment lemma

Using $\beta(t,u) \le c_\epsilon |u-t|^{-\epsilon}$ (the ϵ-tameness condition),

$$\begin{aligned} I_{r,\sigma} &\le c_\epsilon^r \int_{L_\sigma}^{R_\sigma} |u - U_{\tau_\sigma}|^{-r\epsilon}\, du \\ &\le \tfrac{c_\epsilon^r}{1-r\epsilon} \left[(U_{\tau_\sigma} - L_\sigma)^{1-r\epsilon} + (R_\sigma - U_{\tau_\sigma})^{1-r\epsilon} \right]. \\ &\le \tfrac{c_\epsilon^r 2^{r\epsilon}}{1-r\epsilon} (R_\sigma - L_\sigma)^{1-r\epsilon} \text{ by concavity,} \end{aligned}$$

so

$$\mathbf{E}\, I_{r,\sigma}^s \le \left(\tfrac{c_\epsilon^r 2^{r\epsilon}}{1-r\epsilon} \right)^s \mathbf{E} (R_\sigma - L_\sigma)^{s(1-r\epsilon)}.$$

But for $\sigma \in \Lambda_k$ we have $R_\sigma - L_\sigma \stackrel{\mathcal{L}}{=} U_1 \cdots U_k$, so

$$\mathbf{E} (R_\sigma - L_\sigma)^{s(1-r\epsilon)} = \left(\mathbf{E}\, U^{s(1-r\epsilon)} \right)^k = (s+1-sr\epsilon)^{-k}. \qquad \square$$

Proof for worst-case-FIND theorem

Recall that our claim is

$$\frac{1}{n} \times \max_{1 \le m \le n} [\text{cost of } \texttt{QuickSelect}(n, m)] \xrightarrow{L^p} \sup_{\gamma \in \Gamma} I_\gamma,$$

Write this assertion as

$$T_n \xrightarrow{L^p} T \overset{\text{a.s.}}{=} \lim_\ell \max_{\gamma \in \Gamma(\ell)} I_\gamma,$$

where $\Gamma(\ell)$ is the set of all 2^ℓ paths from the root to a node at level ℓ. Note

$$T_n(\ell) \le T_n \le T_n(\ell) + V_n(\ell)$$
$$T(\ell) \le T \le T(\ell) + V(\ell)$$

with

$$T(\ell) := \max_{\gamma \in \Gamma_\ell} I_\gamma, \qquad V(\ell) := \sum_{k > \ell} \max_{\sigma \in \Lambda_k} I_\sigma;$$

$$T_n(\ell) := \max_{\gamma \in \Gamma_\ell} \frac{S_{n,\gamma}}{n}, \quad V_n(\ell) := \sum_{k > \ell} \max_{\sigma \in \Lambda_k} \frac{S_{n,\sigma}}{n}.$$

Proof for worst-case-FIND theorem (cont.)

Lemma (immediate max-lemma)

If F is a finite set and $X_f \geq 0$ for $f \in F$, then

$$\left\| \max_{f \in F} X_f \right\|_p \leq |F|^{1/p} \max_{f \in F} \|X_f\|_p. \qquad \square$$

Our next claim has an easy but instructional proof. Recall

$$V(\ell) := \sum_{k>\ell} \max_{\sigma \in \Lambda_k} I_\sigma.$$

Lemma

If $1 \leq p < \infty$ and β is ϵ-tamed with $0 \leq \epsilon < (p-1)/p$, then

$$\|V(\ell)\|_p \to 0 \text{ as } \ell \to \infty.$$

Proof.

Since

$$\|V(\ell)\|_p \leq \sum_{k>\ell} \left\| \max_{\sigma \in \Lambda_k} I_\sigma \right\|_p \leq \sum_{k>\ell} 2^{k/p} \max_{\sigma \in \Lambda_k} \|I_\sigma\|_p,$$

it is sufficient to show that $\|\!|(I_\sigma)|\!\|_p < \infty$, where the norm

Proof for worst-case-FIND theorem (cont.)

$$|||(X_\sigma)|||_p := \sum_{k=0}^{\infty} 2^{k/p} \max_{\sigma\in\Lambda_k} \|X_\sigma\|_p$$

makes the linear space of tree-indexed stochastic processes (X_σ) into a Banach space (exercise). But for $\epsilon < 1$ we know

$$\forall \sigma \in \Lambda_k : \quad \|I_\sigma\| \le c_{\epsilon,p}\,(p+1-p\epsilon)^{-k/p},$$

so it suffices to have $0 \le \epsilon < (p-1)/p$. □

- Using our fixed-σ L^{p}LLN result $\frac{S_{n,\sigma}}{n} \xrightarrow{L^p} I_\sigma$ and a smidgen more, we know [recalling the definitions that $T_n(\ell) := \max_{\gamma\in\Gamma_\ell} \frac{S_{n,\gamma}}{n}$ and $T(\ell) := \max_{\gamma\in\Gamma_\ell} I_\gamma$] for each fixed ℓ that $T_n(\ell) \xrightarrow{L^p} T(\ell)$ as $n \to \infty$ when $0 \le \epsilon < 1$.
- From here it's easy to check that to finish the proof for worse-case FIND that it's enough to show

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \|V_n(\ell) - V(\ell)\|_p = 0.$$

Proof for worst-case-FIND theorem (conclusion)

- To finish the proof for worse-case FIND it's enough to show

$$\lim_{\ell \to \infty} \limsup_{n \to \infty} \| V_n(\ell) - V(\ell) \|_p = 0.$$

But

$$\begin{aligned} \| V_n(\ell) - V(\ell) \|_p &= \| \textstyle\sum_{k>\ell} \max_{\sigma \in \Lambda_k} \frac{S_{n,\sigma}}{n} - \sum_{k>\ell} \max_{\sigma \in \Lambda_k} I_\sigma \|_p \\ &\le \textstyle\sum_{k>\ell} 2^{k/p} \max_{\sigma \in \Lambda_k} \| \frac{S_{n,\sigma}}{n} - I_\sigma \|_p \\ &\le \| \| \left(\frac{S_{n,\sigma}}{n} \right) - (I_\sigma) \| \|_p, \quad \text{independent of } \ell \ . \end{aligned}$$

- So our worst-case-FIND theorem follows from the following process-convergence theorem. □

Process-convergence theorem

Recall

$$|||(X_\sigma)|||_p := \sum_{k=0}^{\infty} 2^{k/p} \max_{\sigma \in \Lambda_k} \|X_\sigma\|_p.$$

Theorem (Process-convergence theorem)

If $p > 5/2$ and the cost fcn. β is ϵ-tame with $0 \le \epsilon < 1/p$, then

$$||| \left(\frac{S_{n,\sigma}}{n} \right) - (I_\sigma) |||_p \to 0 \text{ as } n \to \infty.$$

Proof (sketch).

Using only $\epsilon < 1/p$, we already know $\forall \sigma : \| \frac{S_{n,\sigma}}{n} - I_\sigma \|_p \to 0$, so by the DCT it suffices to produce a bound

$$\forall \sigma \in \Lambda_k : \| \frac{S_{n,\sigma}}{n} - I_\sigma \|_p \le b_k \quad \text{satisfying} \quad \sum_{k=0}^{\infty} 2^{k/p} b_k < \infty$$

to finish.

Proof (sketch) of process-convergence theorem (cont.)

Primarily by making use of H. P. Rosenthal's inequality (requiring $p \geq 2$) conditionally given C_σ, one finds rather routinely that

$$\| \frac{S_{n,\sigma}}{n} - I_\sigma \|_p^p \leq c_p \left\{ \mathbf{E} \left[\mathbf{1}(\tau_\sigma < n) \left(n^{-(p-1)} I_{p,\sigma} + n^{-p/2} I_{2,\sigma}^{p/2} \right) \right] + \mathbf{E}\, I_\sigma^p \right\}.$$

Here, only the term involving $I_{p,\sigma}$ causes any real trouble, so we'll ignore the other terms in producing b_k. The reason there's trouble with this term is that our I-moment lemma gives (for $0 \leq \epsilon < 1/p$) only

$$2^k \mathbf{E}\, I_{p,\sigma} \leq c_{\epsilon,p} \left(\frac{2}{2-p\epsilon} \right)^k,$$

and $2/(2-p\epsilon) > 1$. However, we can put the additional factors $\mathbf{1}(\tau_\sigma < n) n^{-(p-1)}$ to good use!

Proof (sketch) of process-convergence theorem (cont.)

Summarizing progress thus far, we need a bound b'_k on

$$e_{n,\sigma,p} := \mathbf{E}\left[\mathbf{1}(\tau_\sigma < n) n^{-(p-1)} I_{p,\sigma}\right]$$

for $\sigma \in \Lambda_k$ such that $\sum_k 2^k b'_k < \infty$. Now

$$e_{n,\sigma,p} \le \mathbf{E}\left[\tau_\sigma^{-(p-1)} I_{p,\sigma}\right] \le \left\|\tau_\sigma^{-(p-1)}\right\|_2 \|I_{p,\sigma}\|_2.$$

We can get enough geometric decay using the first of the two L^2-norms here to obtain b'_k. Indeed, for any $\sigma \in \Lambda_k$ the law of τ_σ is that of the kth record epoch τ_k in an iid sequence from a continuous distribution. Bounding the second L^2-norm using the I-moment lemma and using the following steps to bound the first, the proof of the process-convergence theorem is not hard to complete for $p > 5/2$:

Proof (sketch) of process-convergence theorem (conclusion)

- Write the expected value of τ_k as an integral of tail probabilities.
- Use a "switching relation" to convert the tail probabilities to tail probabilities for R_m (for various values of m), where R_m is the number of records collected through epoch m.
- Recall that R_m is distributed as a "Poisson binomial sum ", where the success probabilities are $1, 1/2, 1/3, \ldots, 1/m$ and sum to H_m.
- Use a standard Chernoff bound for Poisson binomial sums. □

F–Nakama QuickQuant thm. as cor. of process-conv.

The F–Nakama QuickQuant theorem giving convergence in L^p for a fixed quantile α is, like the worst-case-FIND theorem, a corollary of the process-convergence theorem (at least for $p > 5/2$). (In fact, uniformity in α can also be established this way, but I won't show it here.) First, the following intermediary result is straightforward to prove from process-convergence .

Corollary

If the process-convergence

$$\left\| \left(\frac{S_{n,\sigma}}{n} \right) - (I_\sigma) \right\|_p \to 0 \text{ as } n \to \infty.$$

holds, then

$$\sup_\gamma \left\| \frac{S_{n,\gamma}}{n} - I_\gamma \right\|_p \xrightarrow{L^p} 0 \text{ as } n \to \infty,$$

where the sup *is over all finite or infinite paths from the root.*

Proof of $\xrightarrow{L^p}$ for fixed quantile α

Suppose $p > 5/2$. Recall that the preceding corollary gave

$$\sup_\gamma \| \tfrac{S_{n,\gamma}}{n} - I_\gamma \|_p \xrightarrow{L^p} 0 \text{ as } n \to \infty,$$

which implies

$$\left\| \tfrac{1}{n} \times \text{cost of } \texttt{QuickQuant}(n, \alpha) - I_{\gamma_n} \right\|_p \to 0$$

where $\gamma_n(\alpha) \equiv \gamma_n = (\sigma_{n0}, \sigma_{n1}, \ldots, \sigma_{n,|\gamma_n|})$ is the path to finite level $|\gamma_n|$ used by $\texttt{QuickQuant}(n, \alpha)$, while the desired result is

$$\left\| \tfrac{1}{n} \times \text{cost of } \texttt{QuickQuant}(n, \alpha) - I_{\gamma} \right\|_p \to 0$$

where $\gamma(\alpha) \equiv \gamma = (\sigma_0, \sigma_1, \ldots)$ is the infinite path from the root to seed α. So we need only show that

$$\|I_{\gamma_n} - I_\gamma\|_p \le \textstyle\sum_{k=0}^{\infty} \|\mathbf{1}(k \le |\gamma_n|) I_{\sigma_{nk}} - I_{\sigma_k}\|_p \to 0.$$

Proof of $\xrightarrow{L^p}$ for fixed quantile α (conclusion)

Reminder: We need only show that

$$\sum_{k=0}^{\infty} \|\mathbf{1}(k \le |\gamma_n|) I_{\sigma_{nk}} - I_{\sigma_k}\|_p \to 0.$$

To see this, first note as an easy consequence of the SLLN that for every k we have a.s. that

$$\mathbf{1}(k \le |\gamma_n|) I_{\sigma_{nk}} - I_{\sigma_k} = 0 \text{ for all large } n;$$

then we'll use the DCT (nestedly) to complete the proof. Indeed,

$$|\mathbf{1}(k \le |\gamma_n|) I_{\sigma_{nk}} - I_{\sigma_k}| \le 2 \max_{\sigma \in \Lambda_k} I_\sigma$$

and by our immediate max-lemma and I-moment lemma

$$\| \max_{\sigma \in \Lambda_k} I_\sigma \|_p \le 2^{k/p} \max_{\sigma \in \Lambda_k} \|I_\sigma\|_p \le c_{\epsilon,p} (p/2)^{-k/p} < \infty;$$

so also

$$\|\mathbf{1}(k \le |\gamma_n|) I_{\sigma_{nk}} - I_{\sigma_k}\|_p \le 2 \times c_{\epsilon,p} (p/2)^{-k/p},$$

which sums since $p > 5/2 > 2$. □

Previous literature (key comparisons) on (worst-case) FIND

- Worst-case FIND was studied by Grübel and Rösler (1996) (see their Theorem 12 and the sentence preceding it) using a rather different sort of process-convergence and by Devroye (2001); see also Grübel (1998, Section 4.3). Our approach is rather different from previous approaches.
- Grübel and Rösler (1996) also used their process-convergence to find a limiting distribution for QuickQuant(n, α) for each α. In our notation, $\beta \equiv 1$ and the limit random variable is $I_{\gamma(\alpha)} = \sum_{\sigma \in \gamma(\alpha)} (R_\sigma - L_\sigma)$. If *also* $\alpha = 0$ (minimum-finding), then

$$I_{\gamma(\alpha)} = 1 + \sum_{k \geq 1} U_{\tau_{0k}} \stackrel{\mathcal{L}}{=} 1 + \sum_{k=1}^{\infty} U_1 \cdots U_j,$$

which is the well-known perpetuity representation of the Dickman distribution. See also Mahmoud, Modarres, and Smythe (1995) and Hwang and Tsai (2002).

Literature on QuickSelect for symbol comparisons

The following have treated the number of symbol comparisons required by QuickQuant(n, α) for fixed α.

- F and Nakama (*Algorithmica*, 2009): expected number of symbol comparisons for standard binary source
- Vallée, Clément, F, and Flajolet (*ICALP*, 2009): expected number of symbol comparisons for a wide variety of probabilistic sources
- F and Nakama (in preparation; see Nakama dissertation, 2009): limiting distributions for a wide variety of probabilistic sources; first to get a limiting distribution for the number of symbol comparisons for any algorithm