A Local Limit Theorem for QuickSort Key Comparisons

Jim Fill

Department of Applied Mathematics and Statistics
The Johns Hopkins University

October 25, 2016

BIRS Workshop in Analytic and Probabilistic Combinatorics

This is joint work with Béla Bollobás and Oliver Riordan.

QuickSort: The Algorithm

- Assume distinct keys (numbers—or perhaps symbol strings—to be sorted).
- Choose "pivot" key uniformly at random.
- Use pivot to partition into two subsets: smaller and larger.
- Apply QuickSort recursively to subsets. Measure runtime by

$$\begin{aligned} K_n &:= \text{number of comparisons, with } K_0 = 0 \\ &\stackrel{\mathcal{L}}{=} K_{U_n - 1} + K^*_{n-U_n} + n - 1, \quad n \geq 1, \end{aligned}$$

  where on the RHS:

$$U_n \sim \text{unif}\{1, \ldots, n\}$$

  and

$$U_n; \; K_0, \ldots, K_{n-1}; \; K^*_0, \ldots, K^*_{n-1}$$

  are all *independent*.
- Applied focus: the law $\mathcal{L}(K_n)$ and asymptotics as $n \to \infty$

The importance of QuickSort

- In a special issue of *Computing in Science & Engineering* (2000), guest editors Jack Dongarra and Francis Sullivan chose QuickSort as one of the ten algorithms "with the greatest influence on the development and practice of science and engineering in the 20th century."
- QuickSort is the standard sorting procedure in Unix systems.
- QuickSort is among "some of the most basic algorithms—the ones that deserve deep investigation."— Ph. Flajolet (1999)
- "probably most widely used sorting algorithm" — U. Rösler (1991)
- "one of the fastest, the best-known, the most generalized, the most completely analyzed, and the most widely used algorithms for sorting" — W. F. Eddy and M. J. Schervish (1995) (But *much* more analysis has followed!)

Conditioning on the pivot index U_n and using the law of total expectation, the distributional recurrence

$$K_n \stackrel{\mathcal{L}}{=} K_{U_n - 1} + K^*_{n-U_n} + n - 1, \quad n \geq 1$$

implies a simple divide and conquer recurrence relation for expected values with explicit solution

$$\kappa_n := \mathbf{E} K_n = 2(n+1)H_n - 4n, \quad n \geq 0.$$

We have

$$\kappa_n \sim 2n \ln n.$$

The law of total variance gives another recurrence relation, with solution

$$\begin{aligned} \mathbf{Var}\, K_n &= 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n \\ &\sim (7 - \tfrac{2}{3}\pi^2)n^2 =: \sigma^2 n^2 \end{aligned}$$

[e.g., Exercise 6.2.2-8 in D. E. Knuth (1973, vol. 3)].

- Higher-order cumulants: Pascal Hennequin (1991 PhD diss.)

2a: Convergence in distribution: heuristic derivation

Is there convergence in distribution?

$$Y_n := \frac{K_n - \kappa_n}{n} \xrightarrow{\mathcal{L}} ?$$

Note

$$K_n \stackrel{\mathcal{L}}{=} K_{U_n - 1} + K^*_{n - U_n} + n - 1, \quad n \geq 1$$

implies

$$Y_n \stackrel{\mathcal{L}}{=} \frac{U_n - 1}{n} Y_{U_n - 1} + \frac{n - U_n}{n} Y^*_{n - U_n} + C_n, \quad n \geq 1,$$

where $Y_0 :=$ arbitrarily, with $C_n := c_n(U_n)$ and

$$c_n(i) := \frac{n-1}{n} + \frac{1}{n}(\kappa_{i-1} + \kappa_{n-i} - \kappa_n).$$

Here $\mathbf{E} Y_n = 0 = \mathbf{E} C_n$.

2a: Convergence in distribution: heuristic derivation

Note: Can get U_n as $\lceil nU \rceil$ with $U \sim \text{unif}(0,1)$.

Can show by calculus:

$$c_n(\lceil nu \rceil) \to c(u) := 2u \ln u + 2(1-u)\ln(1-u) + 1;$$

this suggests

$$Y_n \xrightarrow{\mathcal{L}} Y \stackrel{\mathcal{L}}{=} UY + (1-U)Y^* + c(U)$$

where on the RHS: Y has mean 0 and variance σ^2,

$$U \sim \text{unif}(0,1),$$

and

U, Y, Y^* are *independent*.

2b: Convergence in distribution: three methods of proof

Heuristics have suggested

$$Y_n \xrightarrow{\mathcal{L}} Y \stackrel{\mathcal{L}}{=} UY + (1-U)Y^* + c(U)$$

with

$$\mathbf{E}\,Y = 0 \quad \text{and} \quad \mathbf{Var}\,Y = \sigma^2 = 7 - \tfrac{2}{3}\pi^2 < \infty.$$

This is TRUE!, and at least three methods of proof are possible:

Method 1: $Y_n \xrightarrow{\mathcal{L}} Y$ by *method of moments* [P. Hennequin (1991)]

Method 2: $Y_n \to Y$ a.s. and in L^p $\forall\, 0 < p < \infty$ by *martingale* arguments [M. Régnier (1989)]

Method 3: $Y_n \to Y$ in *Mallows* d_p $\forall p = 1, 2, \ldots$ by *contraction method* [U. Rösler (1991)]

OPEN PROBLEM 6.2 in F and Janson (2002):

Is there a *local* limit theorem for distns.?

YES!: this talk, based on B. Bollobás, F, and O. Riordan (2016+)

3: Selected contributions to analysis of QS (2002-)

The following contributions are highlighted because they are relevant to establishment of a LLT!

- K. H. Tan and P. Hadjicostas (1995) — limit distribution is absolutely continuous; density is positive a.e.
- F and S. Janson (*Mathematics and Computer Science: Algorithms, Trees, Combinatorics, and Probabilities*, 2000) — smoothness and decay properties of the limiting density and its derivatives; in particular (Theorems 3.1 and 3.3 and Corollary 4.2 there), there is an infinitely smooth (analytic?) *everywhere* positive limiting density f satisfying

  $$\max_x f(x) < 16 \quad \text{and} \quad \max_x f'(x) < 2466$$

  (with the true bounds probably closer to 1 and 2, resp.).

  Aside: These bounds were one ingredient used by L. Devroye, F, and R. Neininger (2000) to devise an algorithm for perfect simulation from the limiting distribution F for QuickSort.

Plot of the (smooth) density function f for F

Kok Hooi Tan, P. Hadjicostas / Statistics & Probability Letters 25 (1995) 87–94

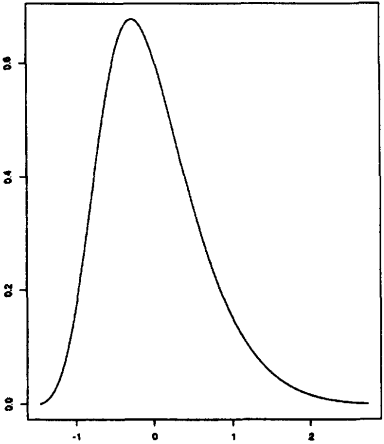

Fig. 4. Density function $f_X(x)$ of limiting distribution.

Thinness of the tails of F (rigorous)

- Invariably, someone asks: How thin are the tails of F?
- A rigorously established answer, sharp to lead order in log-probability, was provided by S. Janson (Electron. J. Probab., 2015), improving significantly on results of F and S. Janson (*RSA*, 2000).
- Let $Y \sim F$. The left tail is doubly-exponentially thin, and the right tail is Poisson(1)-thin:

Theorem (Janson, 2015)

(a) *Let* $\gamma := \left(2 - \frac{1}{\ln 2}\right)^{-1}$. *As* $x \to \infty$ *we have*

$$\exp\left[-e^{\gamma x + \ln\ln x + O(1)}\right] \le \mathbf{P}(Y \le -x) \le \exp\left[-e^{\gamma x + O(1)}\right].$$

(b) *As* $x \to \infty$ *we have*

$$\exp\left[-x \ln x - x \ln\ln x + O(x)\right] \le \mathbf{P}(Y \ge x) \le \exp\left[-x \ln x + O(x)\right].$$

Thinness of the tails of F (non-rigorous)

- Using the non-rigorous WKB method, C. Knessl and W. Szpankowski (*Discrete Math. Theor. Comput. Sci.*, 1999) found very sharp log-asymptotics for both tails:

(a) As $x \to \infty$, for some constant c we have

$$\mathbf{P}(Y \leq -x) = \exp\left[-e^{\gamma x + c + o(1)}\right].$$

(b) As $x \to \infty$ we have

$$\mathbf{P}(Y \geq x) = \exp\left[-x \ln x - x \ln \ln x + (1 + \ln 2)x + o(x)\right].$$

OPEN PROBLEMS: Prove that f is unimodal. Is f in fact *strongly* unimodal? What can one say about changes of signs of the derivatives of f? Is F infinitely divisible?

Let F_n denote the distribution function of $Y_n = (K_n - \kappa_n)/n$. The following two results about convergence of F_n to its limit F are from F and S. Janson (*J. of Algorithms*, 2002), and the upper bounds are proved by first treating *d_p-distances* inductively using the *contraction method* á la Rösler's (1991) proof of convergence and then relating *Kolmogorov–Smirnov distance* to d_p-distances:

- We have $F_n \to F$ at rate $O\left(n^{-(\frac{1}{2}-\epsilon)}\right)$ in the K-S distance (for every $\epsilon > 0$); but our lower bound is only $\Omega(n^{-1})$ (and **improving either bound** in this global theorem remains an **open problem**).
- There is a constant C such that, for any x and any $n \geq 1$,

$$\left| \frac{F_n(x + \frac{\delta_n}{2}) - F_n(x - \frac{\delta_n}{2})}{\delta_n} - f(x) \right| \leq C n^{-1/6},$$

where $\delta_n = 2Cn^{-1/6}$ (w/ no claim of sharpness in the bound).

4: From a semi-local LT to a local LT

Recall that F_n denotes the distribution function of

$$Y_n = (K_n - \kappa_n)/n$$

and that we have the following semi-local limit theorem:

- There is a constant C such that for any x and any $n \geq 1$ we have

$$\left| \frac{F_n(x + \frac{\delta_n}{2}) - F_n(x - \frac{\delta_n}{2})}{\delta_n} - f(x) \right| \leq C n^{-1/6},$$

  where $\delta_n = 2Cn^{-1/6}$.

Our **new result** is that δ_n can be decreased to n^{-1}, with the error bound increasing from $O(n^{-1/6})$ to a little more than $O(n^{-1/18})$. Expressed in terms of the distribution of the unnormalized comparisons-count K_n, the result is as follows (again with no claim of sharpness in the bound).

Main theorem: A LLT for QuickSort

The following local limit theorem is our main result and gives a positive answer to Open Problem 6.2 in F and S. Janson (*J. of Algorithms*, 2002):

Theorem (Local Limit Theorem for QuickSort)

Let K_n denote the (random) number of key comparisons required by QuickSort to sort a file of n distinct keys, let $\kappa_n = \mathbf{E} K_n$, and let f denote the continuous density of the limiting distribution for the normalized random variable $Y_n = (K_n - \kappa_n)/n$. Then there is a constant C such that for any integer k and any $n \geq 1$ we have

$$\left| \mathbf{P}(K_n = k) - n^{-1} f((k - \kappa_n)/n) \right| \leq n^{-1} \times C n^{-1/18} \log n.$$

We obtain the LLT from the semi-LLT by multiple rounds of a smoothing argument. The proof is a bit too complex even for a 60-minute talk, but I will give some ideas.

5: Obtaining the LLT from the semi-LLT: main ideas

We obtain the LLT from the semi-LLT by multiple rounds of a smoothing argument. The basic idea of strengthening a distributional (often normal) limit theorem to a local one by smoothing is by now quite old, but we find it necessary (and sufficient) to do this smoothing in multiple rounds. [Multi-round smoothing has recently been used independently by Diaconis and Hough (2015) in a different context.]

- We know that there is global (and, indeed, semi-local) convergence to a well-behaved distribution. (GLT)
- To deduce a LLT from the GLT it would suffice to show that "nearby" values for K_n have probabilities that are close, namely: If $k_n, k'_n = \kappa_n + O(n)$ and $|k_n - k'_n| = o(n)$, then

$$\mathbf{P}(K_n = k_n) = \mathbf{P}(K_n = k'_n) + o(n^{-1}). \quad (1)$$

- For this, in turn, we might (as in D. R. McDonald (1979)) try to find a "smooth part" within the distribution of K_n. More precisely, . . .

Obtaining the LLT from the semi-LLT: main ideas

- We might try to find a "smooth part" within the distribution of K_n. More precisely, we might try to write
$$K_n = A_n + B_n$$
where, for some σ-field $\mathcal{F}_n$, we have that A_n is $\mathcal{F}_n$-measurable and the conditional distribution of B_n given $\mathcal{F}_n$ (wvhp) obeys a "closeness" relation corresponding to (1).
- Then it follows easily (by first considering conditional probabilities given $\mathcal{F}_n$) that (1) holds.
- One idea is to choose $\mathcal{F}_n$ so that B_n has a very well understood conditional distribution, such as binomial.
- In some contexts, this approach works directly. Here it does not. We can get such a decomp. with (conditionally) $B_n \sim \text{Bin}(\Theta(n), 2/3)$. But then the variance of B_n grows linearly, while the variance of K_n grows quadratically. Roughly speaking, this allows us to prove (1) when $|k_n - k'_n| = o(\sqrt{n})$, but we need (1) for $|k_n - k'_n| = o(n)$ [or, in light of the semi-LLT, at least for $|k_n - k'_n| = O(n^{5/6})$].

6: The tree-exploration lemma

The key idea, then, is not to try to jump straight from the GLT (or from the semi-LLT) to the LLT, but to proceed in stages (rounds). Rather than outline the entire argument (which makes use of the CLT for sums of independent random variables after suitable exponential tilting, among many other things), I will discuss a tree-exploration lemma (along with its proof using martingale arguments) that lies at the heart of the argument. Let $c := 1/100$.

Lemma (tree-exploration lemma)

Let $r \geq 2$ be even, and assume that $r \leq n/75$. Then setting $s = \lceil cn/r \rceil$, we may write $K_n = A + B$ where, for some σ-field $\mathcal{F}$:

- *A is $\mathcal{F}$-measurable, and*
- *with probability at least $1 - e^{-s}$, the conditional distribution of B given $\mathcal{F}$ is the sum of s independent random variables $B_1, \ldots, B_s$ with each B_i having the distribution K_{r_i} for some r_i with $r/2 \leq r_i \leq r$.*

7: Proof of the tree-exploration lemma

The remainder of the talk is devoted to the proof of the tree-exploration lemma. Before any claims, much preliminary discussion:

- T_n = random binary search tree for n nodes labeled $1, \ldots, n$.
- $T(v)$ denotes the subtree of T_n consisting of a given node v and its descendants (the "fringe" subtree of T_n rooted at v). We consider two counts, X_n (primary) and Y_n (auxiliary):
- $X_n :=$ number of nodes v such that (i) $r/2 \le |T(v)| \le r$ and (ii) either v is the root or $|T(\text{parent of } v)| > r$. Call such a node *special*.
- Observe: If v and w are distinct special nodes, then $T(v)$ and $T(w)$ are disjoint.
- We will show that, with probability at least $1 - e^{-s}$, there are at least s special nodes in T_n, i.e.: $\mathbf{P}(X_n \ge s) \ge 1 - e^{-s}$.
- $Y_n :=$ number of fringe trees with at least $r + 1$ nodes.

Proof of the tree-exploration lemma (cont.): exploration

For convenience in presenting the arguments that follow, attach (unlabeled) external nodes as necessary to each node of T_n so that as a result each of the original nodes of T_n has two children; these external nodes do not have labels and do not contribute to the sizes of subtrees.

Consider the following discrete-time procedure for exploring the original nodes of T_n in order to learn their search-tree labels.

- At time 0 nothing is known.
- At time 1, the label of the root is revealed, thus also revealing the sizes of the left and right subtrees.

Now let $2 \leq t \leq n$. ...

Proof of the tree-exploration lemma: exploration (cont.)

Now let $2 \leq t \leq n$. At time $t - 1$, it will be true inductively that nodes with revealed labels will have fringe subtree size at least $r + 1$ and the subtree sizes of their two children will be known.

- At time t, choose an unrevealed child v of a revealed node such that $|T(v)| \geq r + 1$ and reveal its label, if such a v exists.
  - For definiteness, among such nodes v with smallest level, choose the leftmost one.
  - If there are no such nodes v, then nothing is revealed at time t (nor at later times).
  - At any time $t \in [1, n]$, call the unrevealed children (including all external nodes) of revealed nodes "leaves".
- Note that each time the label of a node is revealed, two new leaves (namely, the children of that node) are created; for later technical reasons, we consider the left child to be created as a leaf before the right child.

Proof of the tree-exploration lemma (cont.): martingales

- $\tau := Y_n$ is the random time at which the procedure ends, i.e., the first time at which all of the leaves have subtree sizes $\leq r$.
- $\mathcal{F}_t :=$ the σ-field corresponding to the labels that have been revealed through time t.
- **Observe!:** For any random variable W with finite expectation, the stochastic process $(\mathbf{E}[W|\mathcal{F}_t])_{0 \leq t \leq n}$ is a (Doob's) martingale with values $\mathbf{E}W$, W at times $t = 0, n$, resp.
- We will apply this observation taking W to be X_n [with martingale (M_t)] and Y_n [with martingale (N_t)]. Clearly $N_\tau = Y_n = \tau$.
- The respective means $\xi_n := \mathbf{E}X_n$ and $\eta_n := \mathbf{E}Y_n$ of the martingales (M_t) and (N_t) can be computed by solving standard **divide-and-conquer recurrence relations**.
- From those computations it is simple to deduce that

$$-2 < N_t - N_{t-1} \leq 1, \qquad -\tfrac{1}{2} < M_t - M_{t-1} < 1$$

for every t.

Proof of the tree-exploration lemma (cont.): bound on τ

First claim: $\mathbf{P}(\tau < \frac{3n}{r+1}) \geq 1 - \exp\left[-\frac{2}{27}\left(\frac{n}{r+1}\right)\right]$.

Proof: Recalling that $\tau = N_\tau$, we have

$$\mathbf{P}\left(\tau \geq \frac{3n}{r+1}\right) \leq \mathbf{P}\left(N_t = t \text{ for some } t \geq \frac{3n}{r+1}\right).$$

By the generalized Azuma inequality we have

$$\begin{aligned}
&\mathbf{P}\left(\tau \geq \frac{3n}{r+1}\right) \\
&\leq \mathbf{P}\left(N_t - \eta_n = t - \left[\frac{2(n+1)}{r+2} - 1\right] \text{ for some } t \geq \frac{3n}{r+1}\right) \\
&\leq \mathbf{P}\left(N_t - \eta_n \geq \frac{1}{3}t \text{ for some } t \geq \frac{3n}{r+1}\right) \\
&\leq \exp\left[-\frac{2}{81}\left[\frac{3n}{r+1}\right]\right] \\
&\leq \exp\left[-\frac{2}{27}\left(\frac{n}{r+1}\right)\right] =: \varepsilon. \qquad \square
\end{aligned}$$

First claim: $\mathbf{P}(\tau < \frac{3n}{r+1}) \geq 1 - \exp\left[-\frac{2}{27}\left(\frac{n}{r+1}\right)\right]$. □

Second claim: $\mathbf{P}\left(X_n \geq \frac{(1/100)n}{r}\right) \geq 1 - \exp\left[-\left\lceil\frac{1}{100}\left(\frac{n}{r}\right)\right\rceil\right]$.

Proof: By the first claim, because $t \geq \tau$ implies $M_t = X_n$, we have

$$\mathbf{P}\left(X_n < \frac{(1/2)(n+1)}{r+1}\right) \leq \mathbf{P}\left(\tau \geq \frac{3n}{r+1}\right) + \mathbf{P}\left(M_{\lceil 3n/(r+1)\rceil} < \frac{(1/2)(n+1)}{r+1}\right).$$

Further, by Azuma's inequality we have

$$\begin{aligned}
&\mathbf{P}\left(X_n < \frac{(1/100)n}{r}\right)\\
&\leq \mathbf{P}\left(X_n < \frac{(1/2)(n+1)}{r+1}\right) \leq \varepsilon + \mathbf{P}\left(M_{\lceil 3n/(r+1)\rceil} < \frac{(1/2)(n+1)}{r+1}\right)\\
&= \varepsilon + \mathbf{P}\left(M_{\lceil 3n/(r+1)\rceil} - \xi_n < -\frac{(1/2)(n+1)}{r+1}\right)\\
&\leq \varepsilon + \exp\left[-\frac{8}{9}\left[\frac{(1/2)(n+1)}{r+1}\right]^2 \Big/ \lceil 3n/(r+1)\rceil\right] \leq \varepsilon + \exp\left[-\frac{1}{18}\left(\frac{n}{r+1}\right)\right]\\
&\leq \exp\left[-\frac{4}{81}\left(\frac{n}{r}\right)\right] + \exp\left[-\frac{1}{27}\left(\frac{n}{r}\right)\right] \leq \exp\left[-\left\lceil\frac{1}{100}\left(\frac{n}{r}\right)\right\rceil\right], \quad \text{where}
\end{aligned}$$

Pf. of T-E L: whp, X_n is not too small rel. to n/r (cont.)

- the orange ineq. holds because $\lceil 3n/(r+1) \rceil \le 4n/(r+1)$,
- the blue inequality holds because $r + 1 \le (3/2)r$, and
- the green inequality holds because

  $e^{-4x/81} + e^{-x/27} \le e^{-(x/100)+1} \le e^{-\lceil x/100 \rceil}$ for $x \ge 75$.

We have proven the second claim that the number X_n of special nodes is at least $s = \lceil (1/100)n/r \rceil$ wp at least $1 - e^{-s}$.

The proof of the tree-exploration lemma, and hence of the LLT, concludes on the next slide.

Proof of the tree-exploration lemma (conclusion)

- Let $\sigma \le \tau$ denote
  - the stopping time at which the exploration process has discovered either s or (by means of having simultaneously discovered two special-leaf children with a single reveal) $s + 1$ special leaves, in the event E that such a time exists;
  - otherwise set $\sigma = n$.
- If the event E occurs, fully explore the subtrees rooted at all leaves except the first s special leaves that have been discovered by time σ.
- Take the desired σ-field $\mathcal{F}$ to be (informally stated) the σ-field corresponding to all the information uncovered as we have just described.
- Taking B_i, $i = 1, \dots, s$, to be (when E occurs) the total (internal) path length (i.e., number of key comparisons for QuickSort) of the ith special leaf discovered by time σ, the tree-exploration lemma follows. □

Thanks for attending this talk!

Method 1: $Y_n \xrightarrow{\mathcal{L}} Y$ by *method of moments* [Hennequin (1991)]

- Hennequin (who more generally studied s-ary QuickSort with pivots chosen as medians of samples of size $2t+1$) "pumped moments" and developed asymptotics for them.
- He proved that for $r = 1, 2, \ldots$ we have

$$\mathbf{E}\, Y_n^r \to m_r \in (-\infty, \infty),$$

  where no general expression for m_r seems possible, but a recurrence relation can be established.
- From the recurrence relation he proved that the radius of convergence $[\limsup_r (|m_r|/r!)^{1/r}]^{-1}$ of $\sum_r (m_r/r!) t^r$ is positive. (Rosler: Mgf_Y is finite *everywhere*!) So $Y_n \xrightarrow{\mathcal{L}} Y$, where $\mathcal{L}(Y)$ is the unique distribution with moments m_r.
- Hennequin's recurrence relation for (m_r) states *precisely* that (i) $\mathbf{E}\, Y = 0$ and (ii) Y and $UY + (1-U)Y^* + c(U)$ have the same moments, so the distributional identity follows, too.

An interesting sidelight from Hennequin

A mathematically interesting sidelight from Hennequin (1991) is that the cumulants k_r of Y (that is, the sequence defined by $\sum_{r=1}^{\infty}(k_r/r!)t^r = \ln \mathbf{E}\, e^{tY}$) are for every $r \geq 2$ of the form

$$k_r = (-1)^r 2^r [a_r - (r-1)! \zeta(r)],$$

where the constants a_r are all rational and $\zeta(\cdot)$ is Riemann's zeta function.

Method 2: $Y_n \to Y$ a.s. and in L^p $\forall\, 0 < p < \infty$ by *martingale* arguments [Régnier (1989)]

- There are $n + 1$ external nodes in the binary search tree built from $X_1, \ldots, X_n$.
- $K_n = \text{IPL}_n =$ internal path length in the binary search tree built from $X_1, \ldots, X_n$.
- By induction, $\text{XPL}_n =$ external path length $= \text{IPL}_n + 2n$. [When $n = 0$ this says $0 = 0$. When an external node at distance d from the root is converted to an internal node, IPL increases by d and XPL increases by $2(d + 1) - d = d + 2$; so XPL − IPL increases by $(d + 2) - d = 2$.]
- So, conditionally given the evolution of the random binary search tree through time n, the expected increase in $K = \text{IPL}$ at time $n + 1$ is $\text{XPL}_n/(n + 1) = (K_n + 2n)/(n + 1)$.

Martingale approach: application of L^p Martingale Convergence Theorem

Method 2: $Y_n \to Y$ a.s. and in L^p $\forall\, 0 < p < \infty$ by *martingale* arguments [Régnier (1989)]

- Summarizing from preceding slide,
$$\mathbf{E}(K_{n+1} \,|\, \mathcal{F}_n) = K_n + \tfrac{K_n + 2n}{n+1}.$$
This is easily rearranged to reveal that the process
$\left(\widehat{Y}_n := \frac{n}{n+1} Y_n = \frac{K_n - \kappa_n}{n+1}\right)$ is a martingale with respect to $(\mathcal{F}_n)$.
- **L^p Martingale Convergence Theorem:** Fix $p \in [1, \infty)$. A martingale (Z_n) such that $(|Z_n|^p)_{n=1,2,\dots}$ is uniformly integrable converges both almost surely and in L^p to a (common) limit Z.
- We get the desired conclusions by checking that, for each fixed r, the sequence of absolute moments of order r is bounded. This check does not require the delicacy of Hennequin!

Convergence in distribution: contraction method

Method 3: $Y_n \to Y$ in *Mallows* d_p $\forall p = 1, 2, \ldots$ by *contraction method* [Rösler (1991)]

Rösler's method also shows that $F := \mathcal{L}(Y)$ is the *unique* fixed point of the distributional transformation

$$G = \mathcal{L}(V) \mapsto SG := \mathcal{L}(UV + (1 - U)V^* + c(U))$$

subject (only!) to

$$\mathbf{E}\,V = 0, \quad \mathbf{Var}\,V < \infty.$$

Recall that the method of moments used by Hennequin only gave uniqueness subject to $\mathbf{E}\,V = 0$ and finiteness of *all* moments of V.

Other fixed points

NOTE. (F & S. Janson, *Electron. Commun. Probab.*, 2000a)

$F := \mathcal{L}(Y)$ is the *unique* fixed point of

$$G = \mathcal{L}(V) \mapsto SG := \mathcal{L}(UV + (1-U)V^* + c(U))$$

subject to

$$\mathbf{E}V = 0 \qquad \text{(only!!)}.$$

It's easy to check that the convolution of F with *any* Cauchy (any location, any scale) is also a fixed point.

F & Janson (2000a): Those are *all* the fixed points!

- Due to time constraints, let's limit attention to d_2 here.

Rösler's proof of the existence and uniqueness of a d_2-fixed point for S: Let $d \equiv d_2$. Then

$$\begin{aligned} d(G, H) &:= \inf_{\text{couplings with } V \sim G,\ W \sim H} \|V - W\|_2 \\ &= \|G^{-1}(U) - H^{-1}(U)\|_2. \end{aligned}$$

This is a metric on

$$D := \{\text{distns. with mean 0 and finite variance}\}.$$

Note

(D, d) is a Polish space (i.e., a complete separable metric space);

$\to$ in d iff $\stackrel{\mathcal{L}}{\to}$ and conv. of 2nd moments.

Contraction

Existence & uniqueness of a fixed point for S follows directly from

Proposition

S is a strict contraction on (D, d).

Proof. Given G and H in D, let

$$\begin{aligned} U &\sim \text{unif}(0,1) \\ (V, W) &\sim \text{coupling achieving } d(G, H) \\ (V^*, W^*) &\sim \text{coupling achieving } d(G, H), \end{aligned}$$

with these three independent. Then $S : D \to D$ and

$d^2(S(G), S(H))$

$$\begin{aligned} &\leq \| [UV + (1-U)V^* + c(U)] - [UW + (1-U)W^* + c(U)] \|_2^2 \\ &= \| U(V - W) + (1-U)(V^* - W^*) \|_2^2 \\ &= \|U\|_2^2 \, \|V - W\|_2^2 + \|1 - U\|_2^2 \, \|V^* - W^*\|_2^2 = \tfrac{2}{3} d^2(G, H). \quad \blacksquare \end{aligned}$$

Approximating the fixed point F

Rösler also showed that

$$\mathcal{L}(Y_n) \to \text{ the fixed point } F \text{ of } S.$$

Trusting this for now, we can thus get approximations that converge **geometrically rapidly** in d_2 to the limiting QuickSort distribution by choosing any G with mean zero and finite variance (e.g., $G = \delta_0$) and using $S^n G$ with n large. Indeed, for any such G, Rösler showed that

$$d_2(S^n G, F) \leq \left(\frac{2}{3}\right)^{n/2} (\sigma^2 + \mathbf{Var}\, G)^{1/2}.$$

(Other metrics have been treated, but we don't have time today!)

However, note that computation of $S^n G$ requires numerical integration (e.g., of densities).

Rösler proved $Y_n \to Y$ (i.e., $F_n \to F$) in d;

rate of convergence? From F & Janson (2002, J. Algo.):

Theorem

$d(F_n, F) < \frac{2}{\sqrt{n}}$ for $n \geq 1$.

Remark: By considering $n = 2$, the numerator 2 here is optimal within a factor of about 2.

Outline of **Proof**: Follow Rösler's outline for $d(F_n, F) \to 0$, but be more quantitative. Recall

$$Y_n \stackrel{\mathcal{L}}{=} \frac{\lceil nU \rceil - 1}{n} Y_{\lceil nU \rceil - 1} + \frac{n - \lceil nU \rceil}{n} Y^*_{n - \lceil nU \rceil} + c_n(\lceil nU \rceil),$$

$$Y \stackrel{\mathcal{L}}{=} UY + (1 - U)Y^* + c(U).$$

Proceed as in proof that S is contraction to get

$$\begin{aligned} d^2(F_n, F) &\le \mathbf{E}\left[\frac{\lceil nU \rceil - 1}{n} Y_{\lceil nU \rceil - 1} - UY\right]^2 \\ &\quad + \mathbf{E}\left[\frac{n - \lceil nU \rceil}{n} Y_{n - \lceil nU \rceil} - (1-U)Y\right]^2 \\ &\quad + \mathbf{E}\left[c_n(\lceil nU \rceil) - c(U)\right]^2 \\ &=: \mathbf{(I)} + \mathbf{(II)} + \mathbf{(III)} \end{aligned}$$

using law of total variance. (Condition on U.)
Can show (using calculus)

$$|c_n(\lceil nu \rceil) - c(u)| \le c_1 \frac{\ln n}{n},$$

so

$$\mathbf{(III)} \le c_1^2 \left(\frac{\ln n}{n}\right)^2.$$

Next,

$$\begin{aligned}
\textbf{(I)} &= \mathbf{E}\left[\frac{\lceil nU \rceil - 1}{n} Y_{\lceil nU \rceil - 1} - UY\right]^2 \\
&= \mathbf{E}\left[\frac{\lceil nU \rceil - 1}{n}\left(Y_{\lceil nU \rceil - 1} - Y\right) + \left(\frac{\lceil nU \rceil - 1}{n} - U\right) Y\right]^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i-1}{n} d(F_{i-1}, F) + \frac{\sigma}{n}\right)^2 .
\end{aligned}$$

[Get this by Minkowski's inequality for $\| \cdot \mid U \|_2$;
factor of **2** in $(a+b)^2 \leq 2(a^2+b^2)$ would be disastrous.]

Handle **(II)** similarly.

Find

$$d^2(F_n, F) =: a_n^2 \le \frac{2}{n^3} \sum_{j=1}^{n-1} j^2 a_j^2 + \frac{4\sigma}{n^3} \sum_{j=1}^{n-1} j a_j + c_1^2 \left(\frac{\ln n}{n} \right)^2 + \frac{2\sigma^2}{n^2}.$$

Show

$a_n \le c_2 n^{-1/4}$ using induction;

then

$$\sum_{j=1}^{n-1} j a_j \le c_3 n^{7/4}$$

and so

$$
\begin{aligned}
a_n^2 &\le \frac{2}{n^3}\sum_{j=1}^{n-1} j^2 a_j^2 + c_4 n^{-5/4} + c_1^2\left(\frac{\ln n}{n}\right)^2 + \frac{2\sigma^2}{n^2} \\
&\le \frac{2}{n^3}\sum_{j=1}^{n-1} j^2 a_j^2 + c_5 n^{-5/4} \\
&=: \frac{2}{n^3}\sum_{j=1}^{n-1} j^2 a_j^2 + b_n.
\end{aligned}
$$

By standard **divide and conquer** recurrence (for $n^2 a_n^2$),

$$
a_n^2 \le b_n + 2\frac{n+1}{n^2}\sum_{k=1}^{n-1}\frac{k^2 b_k}{(k+1)(k+2)} \le \frac{c_6}{n},
$$

which gives the result up to a constant multiple. By both tuning up the argument and keeping track of constants, the result follows. ∎

Is $n^{-1/2}$ the right rate for $d(F_n, F)$?

We don't know.

A trivial lower bound:

$$d(F_n, F) = \inf_{\text{couplings}} \|Y_n - Y\|_2$$

$$\geq \left| \|Y_n\|_2 - \|Y\|_2 \right| \geq c_7 \frac{\ln n}{n}.$$

Exact rate of convergence in the Zolotarev metric ζ_3

- R. Neininger and L. Rüschendorf (2002) — **exact** rate $\Theta(\frac{\log n}{n})$ in the *Zolotarev metric* ζ_3

  The *Zolotarev metric* ζ_3 is defined as follows: If $V \sim G$ and $W \sim H$, then

  $$\zeta_3(G, H) := \sup_{f \in \mathcal{F}_3} |\mathbf{E}\, f(V) - \mathbf{E}\, f(W)| \qquad \text{where}$$

  $$\mathcal{F}_3 := \{f : |f''(x) - f''(y)| \le |x - y| \text{ for all } x, y\}$$

  is the class of functions having a Lipschitz-continuous second derivative with Lipschitz constant equal to 1.

  The precise statement of their theorem matters!

Perfect simulation from limiting QuickSort distribution

- Devroye, F, and R. Neininger (2000) — an algorithm for perfect simulation from the limiting distribution F for QuickSort. **Brief summary:** Combining
  - explicit integer bounds on density f and $|f'|$ from F and S. Janson (2000),
  - an explicit integer bound on $\mathbf{E}\, Y^4$, and
  - standard arguments from the book of Devroye (1986) on simulation from distributions,

  we can find a function g such that
  - g is of the form $g(y) = \min(c_1, c_2 y^{-2})$ (and hence integrable),
  - perfect simulation from density normalized-g is elementary, and
  - $f \leq g$.

  Then we can use the rejection method to sample from f, if we can also find a sequence of explicitly computable approximations to f with explicitly computable error bounds. But F and S. Janson (2001) also supplies such a sequence.

1. $\mathbf{E}\,Y = 0$, $\mathbf{Var}\,Y = \sigma^2 = 7 - \frac{2}{3}\pi^2$.
2. Mgf M is finite everywhere (Rösler, 1991) and satisfies

$$M(\lambda) = \int_{u=0}^{1} M(u\lambda)M((1-u)\lambda)e^{\lambda c(u)}\,du, \quad \lambda \in \mathbf{R}.$$

3. Moments of all orders can be "pumped." (Hennequin, Rösler)
4. Characteristic function (ch.f.) ϕ satisfies

$$\phi(t) = \int_{u=0}^{1} \phi(ut)\phi((1-u)t)e^{itc(u)}\,du, \quad t \in \mathbf{R}.$$

5. Method of successive substitutions "works" both for M and for ϕ.

Knowledge about $F = \mathcal{L}(Y)$: absolute continuity

6. F has a density f, which is > 0 a.e. (Tan & Hadjicostas, 1995)

Proof of existence of density: For fixed y and z,

$$h_{y,z}(U) := Uy + (1 - U)z + c(U)$$

is absolutely continuous, say with density $g_{y,z}$. Now mix densities:

$$f(t) := \int_{(y,z)\in\mathbf{R}^2} g_{y,z}(t)\, dF(y)\, dF(z), \quad t \in \mathbf{R}$$

gives density for F, satisfies integral equation

$$f(t) := \int_{(y,z)\in\mathbf{R}^2} g_{y,z}(t) f(y) f(z)\, dy\, dz, \quad t \in \mathbf{R}. \quad \blacksquare$$

But singularity of $g_{y,z}(\cdot)$ at the left endpoint

$$\beta_{y,z} := 1 - 2\ln\left(e^{-y/2} + e^{-z/2}\right)$$

of its support makes proving more about f *challenging*.

Knowledge about F from F and S. Janson (2000b):

- There is a density f that looks like the density plot (Fig. 4) in K. H. Tan and P. Hadjicostas (1995):
  - f is positive everywhere (not just a.e.)
  - f is infinitely differentiable [Is it analytic?]
  - f and its derivatives are bounded (We proved $f < 16$ and $|f'| < 2466$ — *far* from sharp!)
  - $f^{(k)}(t) \to 0$ as $t \to \pm\infty$, superpolynomially quickly, for each k
- f satisfies this integral equation for $x \in \mathbf{R}$:

$$f(x) = \int_{u=0}^{1} \int_{y \in \mathbf{R}} f(y)\, f\left(\frac{x - c(u) - (1-u)y}{u}\right) \frac{1}{u}\, dy\, du$$

OPEN PROBLEMS: Prove that f is unimodal. Is f in fact *strongly* unimodal? What can one say about changes of signs of the derivatives of f? Is F infinitely divisible?

Knowledge about $F = \mathcal{L}(Y)$

More knowledge about F from F and S. Janson (2000b):

- Ch.f. ϕ decays superpolynomially quickly. (NOTE: We proved this first!, with explicit bounds.) Thus, in particular, $\phi \in L^1$(Lebesgue measure), so f is continuous and given by the Fourier inversion formula

$$f(y) \equiv \frac{1}{2\pi} \int_{t \in \mathbf{R}} e^{-iyt} \phi(t)\, dt.$$

[*Note.* Several authors had used this without proof or comment.]

Knowledge about $F = \mathcal{L}(Y)$

An example of knowledge about F contained in F and S. Janson (2001):

- Applying the method of successive substitutions to the integral equation for f gives a sequence (f_n) that converges uniformly to f (method for Figure 4—authors have **disclaimer**), with a geometric rate of convergence, provided we start with a density having zero mean and finite variance (such as standard normal).
- We also have geometrically fast convergence of f_n to f in Kolmogorov–Smirnov and total variation distances.