QuickSort: Improved right-tail asymptotics for the limiting distribution, and large deviations

James Allen Fill and Wei-Chun Hung

Department of Applied Mathematics and Statistics
The Johns Hopkins University

AofA 2019 (Marseille): June 25, 2019

Abstract

We substantially refine asymptotic logarithmic upper bounds produced by Janson (2015) on the right tail of the limiting `QuickSort` distribution function F and by Fill and H (2018) on the right tails of the corresponding density f and of the absolute derivatives of f of each order.

Using the refined asymptotic bounds on F, we derive right-tail large deviation (LD) results for the distribution of the number of comparisons required by `QuickSort` that substantially sharpen the two-sided LD results of McDiarmid and Hayward (1996).

QuickSort

Let X_n denote the (random!) number of comparisons when sorting n distinct numbers using the `QuickSort` algorithm. It is well known that

$$\mu_n := \mathbb{E}\, X_n = 2\,(n+1)\,H_n - 4n \sim 2n \ln n,$$

where H_n is the nth harmonic number $\sum_{k=1}^{n} k^{-1}$ and (from a simple exact expression) that

$$\mathbb{V} X_n = (1 + o(1)) \left(7 - \tfrac{2\pi^2}{3}\right) n^2.$$

Convergence of Z_n

We center and scale X_n:

$$Z_n = \frac{X_n - \mu_n}{n}.$$

- Rösler (1991) proved that Z_n converges to Z weakly as $n \to \infty$.
- Régnier (1989) proved that Z_n converges to Z in L^p for every finite p (and thus in distribution).

Properties of the limiting QuickSort distribution

- Rösler (1991) proved that Z has everywhere finite moment generating function and satisfies the fixed point equation

$$Z \stackrel{\mathcal{L}}{=} UZ + (1-U)Z^* + g(U).$$

  Here Z and Z^* are independent copies of Z, and U is uniformly distributed on $(0,1)$. The function g satisfies

$$g(u) := 2u \ln u + 2(1-u)\ln(1-u) + 1.$$

- Fill and Janson (2000) proved that Z has a (unique) continuous density f which is everywhere positive and infinitely differentiable, and for every $k \geq 0$ that $f^{(k)}$ is bounded and enjoys superpolynomial decay in both tails.

Knessl and Szpankowski (1999) and Our Contribution

- Knessl and Szpankowski (1999, DMTCS) used the non-rigorous WKB method to establish "exact asymptotics" (i.e., lead-order asymptotics in the probability, not just its log) for the tails of the limiting density and (by integration) of the limiting distribution function.
- We have rigorously come rather close to their results, especially for the right tail. For brevity, talk focuses on UBs for the right tail.
- Later: We use the asymptotics for the distribution of Z to substantially improve LD results for Z_n.

From integral equation for Z to mgf ψ

- K&S begin by considering the mgf ψ for Z, with integral equation

$$(*) \quad \psi(t) = \int_{u=0}^{1} \psi(ut)\psi((1-u)t)e^{tg(u)}\,du, \quad t \in \mathbb{R}.$$

- They find [using $(*)$ and the WKB method) that

$$\psi(t) = \exp[J(t) - t^2 - \alpha t - \ln t + C + o(1)] \text{ as } t \to \infty,$$

where
  - $\alpha := 2\ln 2 + 2\gamma - 1$ (with $\gamma =$ Euler–Mascheroni constant),
  - C is an (unspecified) constant, and
  - $J(t)$ has the following definition and divergent asymptotic expansion:

$$J(t) := \int_{s=1}^{t} 2s^{-1}e^s\,ds \sim 2t^{-1}e^t \sum_{j=0}^{\infty} j!\,t^{-j} \quad (\text{lead term} = 2t^{-1}e^t).$$

- Janson (2015, ECP) proves that $\psi(t) \le \exp[e^t + O(t)]$.
- We now match K&S to quadratic term with linear remainder:

$$\psi(t) = \exp[J(t) - t^2 + O(t)].$$

From integral equation for Z to mgf ψ: Recap

- Knessl and Szpankowski (1999):

$$\psi(t) = \exp[J(t) - t^2 - \alpha t - \ln t + C + o(1)] \text{ as } t \to \infty,$$

where

- $\alpha := 2\ln 2 + 2\gamma - 1$,
- C is an (unspecified) constant, and
- $J(t)$ has the following definition and divergent asymptotic expansion:

$$J(t) := \int_{s=1}^{t} 2s^{-1}e^s\, ds \sim 2t^{-1}e^t \sum_{j=0}^{\infty} j!\, t^{-j} \quad (\text{lead term} = 2t^{-1}e^{-t}).$$

Theorem (Fill & H, 2019)

As $t \to \infty$, the mgf ψ for limiting `QuickSort` *satisfies*

$$\psi(t) = \exp[J(t) - t^2 + O(t)].$$

Remark. With proof technique of Janson and of Fill and H, little hope of improvement.

From mgf ψ to density f and distribution function F

- Knessl and Szpankowski (1999) (non-rigorously) derive asymptotics for the density f by the "standard saddle point approximation" from their mgf expansion (for the same C):

$$f(x) \sim (2\pi \times 2w^{-1}e^{w})^{-1/2} e^{-xw}\psi(w)$$
$$\sim \exp[-xw + J(w) - w^2 - (\alpha + \tfrac{1}{2})w - \tfrac{1}{2}\ln w + C - \ln(2\sqrt{\pi})],$$

where $w \equiv w(x)$ is the larger of the two sols. in $\mathbb{R}$ to $x = 2w^{-1}e^{w}$:

$$w = \ln(x/2) + \ln\ln(x/2) + (1 + o(1))\frac{\ln\ln(x/2)}{\ln(x/2)} \sim \ln x \text{ as } x \to \infty.$$

- They then derive asymptotics for the d.f. F by integrating:

$$\overline{F}(x) := 1 - F(x) =$$
$$\exp\left[-xw + J(w) - w^2 - (\alpha + \tfrac{1}{2})w - \tfrac{3}{2}\ln w + C - \ln(2\sqrt{\pi}) + o(1)\right].$$

From mgf ψ to distribution function F (cont.)

- Kn. & Szp. (1999) recap for F, with $x = 2w^{-1}e^w$ (and $w \sim \ln x$):

$$\overline{F}(x) = 1 - F(x) = \exp\left[-xw + J(w) - w^2 - (\alpha + \tfrac{1}{2})w - \tfrac{3}{2}\ln w + C - \ln(2\sqrt{\pi}) + o(1)\right].$$

- Janson (2015) via Chernoff bound $\overline{F}(x) \le e^{-tx}\psi(t)$:

$$\overline{F}(x) \le \exp[-x\ln x + O(x)].$$

Theorem (Fill &H, 2019, via Chernoff bound, $t = w$)

As $x \to \infty$, the limiting `QuickSort` *distribution function F satisfies*

$$\overline{F}(x) \le \exp[-xw + J(w) - w^2 + O(\log x)].$$

Remark: no improvement from using the optimal t for the bound on $\psi(t)$

From d.f. F to density f

- Recall Kn. & Szp. (1999) for f, with $x = 2w^{-1}e^w$ (and $w \sim \ln x$):

$$f(x) \sim \exp[-xw + J(w) - w^2 - (\alpha + \tfrac{1}{2})w - \tfrac{1}{2}\ln w + C - \ln(2\sqrt{\pi})].$$

- Janson (2015) does not treat the density function f.
- Fill and H (2018 AofA & full) extend Janson's bounds on $\overline{F}$:

$$\exp[-x\ln x - x\ln\ln x + O(x)] \le f(x) \le \exp[-x\ln x + O(x)].$$

Theorem (Fill and H, 2019)

As $x \to \infty$, the limiting `QuickSort` *density f satisfies*

$$f(x) \le \exp[-xw + J(w) + O(\sqrt{x\log x})].$$

Derivatives of the density f

- Knessl and Szpankowski (1999) do not treat derivatives of f.
- Fill and H (ANALCO 2019) have UBs and LBs on derivatives of f of each order that match those on f (and on $\overline{F}$), with limitations.
- A key limitation: We don't even know that f or any of its derivatives is ultimately of constant sign in either tail!
- So, to deal with lower bounds: For a function $h : \mathbb{R} \to \mathbb{R}$, write

$$\|h\|_x := \sup_{t \geq x} |h(t)|$$

  for the right-tail sup-norm.
- Our proofs (omitted in this talk) make use of the Landau–Kolmogorov inequality (among other tools).

Derivatives of the density f (cont.)

- For a function $h : \mathbb{R} \to \mathbb{R}$, write

$$\|h\|_x := \sup_{t \geq x} |h(t)|.$$

- Fill and H (2018 AofA & full) nearly extend Janson's bounds on $\overline{F}$:

$$\exp[-x \ln x - (k \vee 1) x \ln \ln x + O(x)] \leq \|\overline{F}^{(k)}\|_x \leq \exp[-x \ln x + O(x)].$$

Theorem (Fill and H, 2019)

As $x \to \infty$, the k^{th} derivative of the limiting `QuickSort` *d.f. $\overline{F}$ satisfies*

$$\overline{F}^{(k)}(x) \leq \|\overline{F}^{(k)}\|_x \leq \exp[-xw + J(w) + O(\sqrt{x \log x})].$$

Conjecture about derivatives of the QuickSort density f

We conjecture that repeated formal differentiation of the Knessl and Szpankowski (1999) result for f gives the correct exact asymptotics:

As $x \to \infty$,

$$f^{(k)}(x) \sim (-1)^{k-1} \times$$
$$\exp[-xw + J(w) - w^2 - (\alpha + \tfrac{1}{2})w + (k - \tfrac{1}{2}) \ln w + C - \ln(2\sqrt{\pi})].$$

A quick word about left tails: doubly-exponentially thin

The left tail of F is doubly exponentially thin, as are its derivatives of each order. Let

$$\underline{F}(x) := F(-x).$$

Theorem (Janson, 2015 for $k = 0$; Fill and H, 2018)

Given an integer $k \geq 0$, as $x \to \infty$ the k^{th} derivative of the limiting `QuickSort` *distribution function F satisfies*

$$\exp\left[-e^{\Gamma x + \ln\ln x + O(1)}\right] \leq \|\underline{F}^{(k)}\|_x \leq \exp\left[-e^{\Gamma x + O(1)}\right].$$

where $\Gamma := (2 - \frac{1}{\ln 2})^{-1}$.

Non-rigorous Knessl and Szpankowski (1999) exact asymptotics are consistent with upper bound here.

In brief, **left tail: Gumbel-like** and **right tail: Poisson-like**.

Right-tail large deviations for QuickSort

Recall that X_n is the number of comparisons used by QuickSort to sort n distinct numbers, with expectation $\mu_n = 2(n+1)H_n - 4n \sim 2n \ln n$, and

$$Z_n := (X_n - \mu_n)/n \xrightarrow{\mathcal{L}} Z.$$

There are two ways that (right-tail) asymptotics for Z can be used to study (right-tail) large deviations for QuickSort Z_n itself.

1. Utilize the following Berry–Esseen-type lemma:

Lemma (Fill and Janson, 2002, J. Algo.)

We have

$$\sup_x |\mathbb{P}(Z_n > x) - \mathbb{P}(Z > x)| \le \exp\left[-\tfrac{1}{2}\ln n + O\left((\log n)^{1/2}\right)\right].$$

Right-tail large deviations for QuickSort (1. cont.)

1. Combining the Berry–Esseen-type lemma with our right-tail asymptotics for Z yields the following theorem (and also a more refined upper bound):

Theorem (Fill and H, ANALCO2019)

Let (ω_n) be any sequence diverging to $+\infty$ as $n \to \infty$ and let $c > 1$. For integer $n \geq 3$, consider the interval

$$I_n := \left[c, \tfrac{1}{2}\tfrac{\ln n}{\ln\ln n}\left(1 - \tfrac{\omega_n}{\ln\ln n}\right)\right].$$

If $x_n \in I_n$ for all large n, then

$$\mathbb{P}(Z_n > x_n) \sim \mathbb{P}(Z > x_n) = \exp[-x_n \ln x_n - x_n \ln\ln x_n + O(x_n)].$$

Remark. Roughly speaking (suppressing technical details), our theorem has a slightly narrower range of applicability than McDiarmid and Hayward (1996, J. Algo.) but one more term for $\ln \mathbb{P}(Z_n > x_n)$.

[and left tail: out to $x_n = O(\log\log n)$]

Right-tail large deviations for QuickSort (2.)

- If we let $N := n + 1$ and study the slight modification

$$\widehat{Z}_n := (X_n - \mu_n)/N = [n/(n+1)]Z_n,$$

then large deviation upper bounds based on tail estimates of the limiting F have broader applicability and are easier to derive, too.

- The reason is that
  - (i) our upper bound on $\overline{F}$ was derived by establishing an upper bound on the limiting mgf ψ and using a Chernoff bound, and
  - (ii) according to Fill and Janson (2002, Thm. 7.1), ψ majorizes the mgf $\widehat{\psi}_n$ of $\widehat{Z}_n$ for every n.
- It follows immediately that $\mathbb{P}(\widehat{Z}_n > x)$ is bounded above uniformly in n by

$$\exp[-xw + J(w) - w^2 + O(\log x)] \tag{1}$$

$$= \exp[-x \ln x - x \ln\ln x + (1 + \ln 2)x + o(x)] \tag{2}$$

there is *no restriction at all* on how large x can be in terms of n.

Example of 2.: *Extremely* large deviation

- Here is an example of a *very* large value of x for which the tail probability is nonzero and the aforementioned bound still matches logarithmic asymptotics to lead order of magnitude, albeit not to lead-order term. Let lg denote binary log.
- The largest possible value of X_n is $\binom{n}{2}$ (corresponding to any binary search tree which is a path), which occurs with probability $2^{n-1}/n!$.
- Correspondingly, the largest possible value of $\widehat{Z}_n$ is

$$\lambda_n := \tfrac{n(n+7)}{2(n+1)} - 2H_n = (1 + o(1))\tfrac{1}{2}N.$$

The bound (2) on $\mathbb{P}(\widehat{Z}_n > \lambda_n)$ is in fact also (by the same proof) a bound on the larger probability $\mathbb{P}(\widehat{Z}_n \geq \lambda_n)$, and equals

$$\exp\left\{-\tfrac{1}{2}N[\ln N + \ln\ln N - (2\ln 2 + 1) + o(1)]\right\},$$

whereas (using Stirling's formula) the truth is

$$\mathbb{P}(\widehat{Z}_n \geq \lambda_n) = \exp[-N\ln N + (1 + \ln 2)N + O(\log N)].$$

- We can handle the smallest possible value of X_n similarly.

$\psi(t) \le \exp[J(t) - t^2 + at]$ for some $a \ge 0$ and all $t \ge 0$

We outline a proof that

$$(**) \quad \psi(t) \le \exp[J(t) - t^2 + at] \text{ for some } a \ge 0 \text{ and all } t \ge 0;$$

a matching lower bound (with linear term $-at$) is proved similarly.

The proof will require the following lemma. Recall

$$(*) \quad \psi(t) = 2 \int_{u=0}^{1/2} \psi(ut)\psi((1-u)t)e^{tg(u)}\, du, \quad t \in \mathbb{R}.$$

and define $\widehat{\psi}(t) := 1$ if $t \le 1$ and

$$\widehat{\psi}(t) := (1 - e^{-t/2}) \exp[J(t) - t^2 - \alpha t - \ln t] \text{ if } t > 1.$$

Lemma

For all sufficiently large t we have the strict inequality

$$2 \int_{u=0}^{1/2} \widehat{\psi}(ut)\widehat{\psi}((1-u)t) \exp[tg(u)]\, du < \widehat{\psi}(t).$$

Proof. This lemma is the heart of the proof of $(**)$, but it's technical! □

Proof of $(**)$ $\psi(t) \le \exp[J(t) - t^2 + at]$ (cont.)

- Recall $\widehat{\psi}(t) := 1$ if $t \le 1$ and

$$\widehat{\psi}(t) := (1 - e^{-t/2}) \exp[J(t) - t^2 - \alpha t - \ln t] \text{ if } t > 1.$$

- We carry out the proof by showing that there exists $a' \ge 0$ such that

$$(***) \quad \psi(t) \le e^{a't} \widehat{\psi}(t)$$

  for every $t > 0$.
- To begin, we compare asymptotics of $\psi(t)$ and $\widehat{\psi}(t)$ as $t \to 0$. Because Z has zero mean and finite variance, we have $\psi(t) = 1 + O(t^2)$. On the other hand, $\widehat{\psi}(t) = 1$ for all $0 < t \le 1$.
- We can thus choose $t_1 > 0$ and $a'' > 0$ such that $(***)$ holds for $t \in [0, t_1]$ and any $a' \ge a''$.
- Let $t_2 > 1$ be such that the strict inequality in the lemma holds for all $t \ge t_2$, and choose $a' \ge a''$ so that $(***)$ holds for $t \in [t_1, t_2]$.
- Assuming for the sake of contradiction that $(***)$ fails for some $t > 0$, let $T := \inf\{t > 0 : (***) \text{ fails}\}$.

Proof of $(**)$ $\psi(t) \le \exp[J(t) - t^2 + at]$ (cont.)

- Recall the definition and desired inequality (respectively)

$$\widehat{\psi}(t) := (1 - e^{-t/2}) \exp[J(t) - t^2 - \alpha t - \ln t] \text{ if } t > 1,$$

$$(***) \quad \psi(t) \le e^{a' t} \widehat{\psi}(t) \quad \text{(We've shown } (***) \text{ for } t \in [0, t_2].)$$

- We have assumed for the sake of contradiction that

$$T := \inf\{t > 0 : (***) \text{ fails}\} < \infty.$$

- Then $T \ge t_2$, and continuity gives

$$\psi(T) = e^{a' T} \widehat{\psi}(T).$$

- Further, if $0 < u < 1$, then $(***)$ holds for $t = uT$ and $t = (1-u)T$, and thus, using our standard integral equation $(*)$ for ψ, we have

$$\psi(T) \le e^{a' T} \times 2 \int_{u=0}^{1/2} \widehat{\psi}(uT) \widehat{\psi}((1-u)T) \exp[tg(u)] \, du,$$

which is strictly smaller than $e^{a' T} \widehat{\psi}(T)$ by applying the lemma with $t = T \ge t_2$.

Proof of $(**)$ $\psi(t) \leq \exp[J(t) - t^2 + at]$ (conclusion)

- The resulting strict inequality $\psi(T) < e^{a'T}\widehat{\psi}(T)$ contradicts the definition of T.
- Hence $(***)$ holds for all $t \geq 0$.

(Additional) Tools: 1. An integral equation for f

Lemma (Integral equation, Fill and Janson (2000))

The density function f of the limiting `QuickSort` *distribution satisfies the integral equation*

$$f(x) = 2 \int_{u=0}^{1/2} \int_{z \in \mathbb{R}} f(z)\, f\left(\frac{x - g(u) - (1-u)z}{u} \right) \frac{1}{u}\, dz\, du.$$

This integral equation is the starting point for our lower bounds on f. The rough idea behind the proof for the right tail (for example) is to shrink the region of integration (strategically!) to lower-bound $f(x)$ for $x \in [0, kb]$ (for suitable b and every $k \geq 2$) in terms of the values of $f(z)$ for $z \in [0, (k-1)b]$. This gives a recurrence inequality which can be "solved".

Tools: 2. The Landau–Kolmogorov inequality

For an overview of the Landau–Kolmogorov inequality, see Mitrinović et al. (1991, Chapter 1). Here we state a version of the inequality well-suited to our purposes; see Matorin (1955) and Stechkin (1967).

Lemma

Let $n \geq 2$, and suppose $h : \mathbb{R} \to \mathbb{R}$ has n derivatives. If h and $h^{(n)}$ are both bounded, then for $1 \leq k < n$ so is $h^{(k)}$. Moreover, there exist constants $c_{n,k}$ (not depending on h) such that, for every $x \in \mathbb{R}$, the tail-supremum norm $\|\cdot\|_x$ satisfies

$$\|h^{(k)}\|_x \leq c_{n,k} \, \|h\|_x^{1-(k/n)} \, \|h^{(n)}\|_x^{k/n}, \quad 1 \leq k < n.$$

Further, for $1 \leq k \leq n/2$ the best constants $c_{n,k}$ satisfy

$$c_{n,k} \leq \left(\frac{e^2 n}{4k}\right)^k.$$

Tools: 3. Constant upper bounds for absolute derivatives

We also make use of the following result, easily proved from Fill and Janson (2000)

Lemma

For every integer $k \geq 0$ we have

$$\sup_{x \in \mathbb{R}} |f^{(k)}(x)| \leq 2^{k^2+10k+17}.$$

Proof of left tail lower bound on f

Reminder: $g(u) = 2u \ln u + 2(1-u)\ln(1-u) + 1$.

Lemma

Given $\epsilon \in (0, \frac{1}{2})$, *let* $a \equiv a(\epsilon) := -g(\frac{1}{2} - \epsilon)$ *and (for* $z \geq 0$*)*

$m_z := \left(\min_{x \in [-z,0]} f(x) \right) \wedge 1$. *Then provided* ϵ *is sufficiently small that* $a > 0$, *for any integer* $k \geq 2$ *we have*

$$m_{ka} \geq \left(2\epsilon^3 m_{2a}\right)^{2^{k-2}}.$$

This lemma comes from a recurrence relation $m_{ka} \geq 2\epsilon^3 m_{(k-1)a}^2$, which in turn is derived using the integral equation. Therefore, we can lower-bound the density function: For $x > a$ we have

$$f(-x) \geq m_x \geq m_{\lceil \frac{x}{a} \rceil a} \geq \left(2\epsilon^3 m_{2a}\right)^{2^{\lceil \frac{x}{a} \rceil - 2}} \geq \left(2\epsilon^3 m_{2a}\right)^{2^{\frac{x}{a}}}.$$

The left tail lower bound is achieved by choosing $\epsilon = x^{-\frac{1}{2}}$.

Proof of right tail lower bound on f

Lemma

Let $m_z := \min_{x\in[0,z]} f(x)$ *(for* $z \geq 0$*) and* $c := 2 \cdot \mathbb{P}(0 \leq Z \leq 1)$*. Then provided* x *is sufficiently large, with* $b := 1 - \frac{2}{\ln x}$ *and* $\delta := \frac{1}{x \ln x}$*, for any integer* $k \geq 1$ *satisfying* $2 + (k-1)b \leq \frac{(g(\delta)-b)}{\delta}$ *we have*

$$m_{2+kb} \geq (c\delta)^{k-1} m_3.$$

Using the integral equation, we can derive a recurrence relation $m_{2+kb} \geq c\delta m_{2+(k-1)b}$. The above lemma is a consequence of this recurrence relation. By choosing $k = \lceil (x-2)/b \rceil \geq 2$, the density function can thus be lower-bounded by

$$f(x) \geq m_{2+kb} \geq (c\delta)^{k-1} m_3 \geq (c\delta)^{\frac{x-2}{b}} m_3,$$

which leads easily to the desired right-tail asymptotic lower bound on f.