Approximating the Limiting Quicksort Distribution

James Allen Fill[1]

Department of Mathematical Sciences

The Johns Hopkins University

`jimfill@jhu.edu` and `http://www.mts.jhu.edu/~fill/`

and

Svante Janson

Department of Mathematics

Uppsala University

`svante.janson@math.uu.se` and `http://www.math.uu.se/~svante/`

ABSTRACT

The limiting distribution of the normalized number of comparisons used by `Quicksort` to sort an array of n numbers is known to be the unique fixed point with zero mean of a certain distributional transformation S. We study the convergence to the limiting distribution of the sequence of distributions obtained by iterating the transformation S, beginning with a (nearly) arbitrary starting distribution. We demonstrate geometrically fast convergence for various metrics and discuss some implications for numerical calculations of the limiting `Quicksort` distribution. Finally, we give companion lower bounds which show that the convergence is not faster than geometric.

AMS 2000 *subject classifications.* Primary 68W40; secondary 68P10, 60E05, 60E10, 60F05.

Key words and phrases. `Quicksort`, characteristic function, density, moment generating function, sorting algorithm, coupling, Fourier analysis, Kolmogorv–Smirnov distance, total variation distance, integral equation, numerical analysis, d_p-metric.

Date. January 15, 2001.

[1]Research supported by NSF grant DMS–9803780, and by the Acheson J. Duncan Fund for the Advancement of Research in Statistics.

1 Introduction and summary

The `Quicksort` algorithm of Hoare [9] is "one of the fastest, the best-known, the most generalized, ... and the most widely used algorithms for sorting an array of numbers" [4]. `Quicksort` is the standard sorting procedure in `Unix` systems, and in a special issue of *Computing in Science & Engineering*, guest editors Jack Dongarra and Francis Sullivan ([3]; see also [10]) 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." Our goal in this introductory section is to review briefly some of what is known about the analysis of `Quicksort` and to summarize how this paper advances that analysis.

The `Quicksort` algorithm for sorting an array of n numbers is extremely simple to describe. If $n = 0$ or $n = 1$, there is nothing to do. If $n \geq 2$, pick a number uniformly at random from the given array. Compare the other numbers to it to partition the remaining numbers into two subarrays. Then recursively invoke `Quicksort` on each of the two subarrays.

Let X_n denote the (random) number of comparisons required (so that $X_0 = 0$). Then X_n satisfies the distributional recurrence relation

$$X_n \stackrel{\mathcal{L}}{=} X_{U_n - 1} + X^*_{n-U_n} + n - 1, \qquad n \geq 1,$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in law (i.e., in distribution), and where, on the right, U_n is distributed uniformly on the set $\{1, \ldots, n\}$, $X^*_j \stackrel{\mathcal{L}}{=} X_j$, and

$$U_n;\ X_0, \ldots, X_{n-1};\ X^*_0, \ldots, X^*_{n-1}$$

are all independent.

As is well known and quite easily established, for $n \geq 0$ we have

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

where $H_n := \sum_{k=1}^{n} k^{-1}$ is the nth harmonic number and $\sim$ denotes asymptotic equivalence. It is also routine to compute explicitly the standard deviation of X_n (see Exercise 6.2.2-8 in [12]), which turns out to be $\sim n\sqrt{7 - \frac{2}{3}\pi^2}$.

Consider the normalized variate

$$Y_n := (X_n - \mu_n)/n, \qquad n \geq 1. \tag{1.1}$$

Régnier [14] showed using martingale arguments that $Y_n \to Y$ in distribution, with Y satisfying the distributional identity

$$Y \stackrel{\mathcal{L}}{=} UY + (1-U)Y^* + g(U) =: h_{Y,Y^*}(U), \tag{1.2}$$

where

$$g(u) := 2u \ln u + 2(1-u)\ln(1-u) + 1, \tag{1.3}$$

and where, on the right of $\stackrel{\mathcal{L}}{=}$ in (1.2), U, Y, and Y^* are independent, with $Y^* \stackrel{\mathcal{L}}{=} Y$ and $U \sim \mathrm{unif}(0,1)$. Rösler [15] showed that (1.2) characterizes the limiting law $\mathcal{L}(Y)$, in the precise sense that $F := \mathcal{L}(Y)$ is the *unique* fixed point of the operator

$$G = \mathcal{L}(V) \mapsto SG := \mathcal{L}(UV + (1-U)V^* + g(U)) \tag{1.4}$$

(in what should now be obvious notation) subject to

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

[The fixed points of G with finite mean are the translates $\mathcal{L}(Y+c)$ with c constant, but there are other fixed points without mean; see [7] for a complete characterization.]

Rösler [15] showed that the moment generating function of the limiting distribution $\mathcal{L}(Y)$ is everywhere finite. We have studied the limiting distribution further in [6], showing that $\mathcal{L}(Y)$ has a density f which is infinitely differentiable, and that each derivative $f^{(k)}(y)$ is bounded and decays as $y \to \pm\infty$ more rapidly than any power of $|y|^{-1}$. (This improves an earlier result by Tan and Hadjicostas [18].)

The purpose of the present paper is to study the convergence to the limiting distribution $\mathcal{L}(Y)$ of the sequence of distributions obtained by iterating Rösler's operator S in (1.4), beginning with a (nearly) arbitrary starting distribution. To fix notation, we let Z_0 be an arbitrary random variable, and $F_0 := \mathcal{L}(Z_0)$ its distribution. We define, for $n \geq 1$,

$$Z_n := h_{Z_{n-1}, Z^*_{n-1}}(U),$$

with $Z^*_{n-1} \stackrel{\mathcal{L}}{=} Z_{n-1}$ and Z_{n-1}, Z^*_{n-1}, and U independent; in other words,

$$F_n := \mathcal{L}(Z_n) = S^n F_0, \qquad n \geq 0.$$

Let $\|X\|_2 := (\mathbf{E}\, X^2)^{1/2}$ denote the L^2-norm, and let d_2 denote the metric on the space of probability distributions with finite variance defined by

$$d_2(G_1, G_2) := \min \|X_1 - X_2\|_2, \tag{1.5}$$

taking the minimum over all pairs of random variables X_1 and X_2 (defined on the same probability space) with $\mathcal{L}(X_1) = G_1$ and $\mathcal{L}(X_2) = G_2$. Note that, using the coupling with X_1 and X_2 independent, for any G_1 and G_2 each with zero mean and finite variance,

$$d_2(G_1, G_2) \leq (\mathbf{E}\, X_1^2 + \mathbf{E}\, X_2^2)^{1/2} \leq \|X_1\|_2 + \|X_2\|_2, \tag{1.6}$$

when $\mathcal{L}(X_1) = G_1$ and $\mathcal{L}(X_2) = G_2$. Rösler [15] showed that if Z_0 has mean 0 and finite variance, then $F_n \to F$ in the d_2-distance with a geometric rate:

$$d_2(F_n, F) \leq (2/3)^{n/2} d_2(F_0, F) \leq (2/3)^{n/2} (\mathbf{Var}\, Z_0 + \sigma^2)^{1/2}, \tag{1.7}$$

where

$$\sigma^2 := \mathbf{Var}\, Y = 7 - \tfrac{2}{3}\pi^2 \doteq 0.42. \tag{1.8}$$

Our main interest is to show similar estimates for other measures of the distance between F_n and F.

We will show in Section 3, using estimates of the characteristic functions given in [6] and Section 2, that the distribution F_n has a bounded, continuous density function f_n, at least as soon as $n \geq 3$, and that, if Z_0 has mean 0 and finite variance, then f_n converges uniformly to f, with a geometric rate of convergence, as $n \to \infty$. We further show geometrically fast convergence in the total variation and Kolmogorov–Smirnov distances, too.

In Section 4 we give bounds for the moment generating functions of Y and of Z_n.

In Section 5 we show that if Z_0 has mean 0 and a finite moment generating function ψ_0, then the moment generating function ψ_n of F_n is finite and converges uniformly on compact intervals to the moment generating function of Y, again with a geometric rate of convergence. We study in particular the cases $Z_0 = 0$ and Z_0 normally distributed with zero mean and sufficiently large variance; it turns out that in these cases $\psi_n(\lambda)$ converges monotonically.

In Section 6, we discuss some implications for numerical calculations of the limiting `Quicksort` distribution F, showing how explicit and arbitrarily small error bounds can be obtained.

Finally, in Sections 7–8 we give some companion lower bounds, showing that the convergence is not faster than geometrical for several different metrics. We also show geometrically fast convergence in the d_p metric for any finite p.

Remark 1.1. The mode and rate of convergence of the distribution of the actual normalized `Quicksort` variables Y_n of (1.1) to the limit F is a quite different matter, which will be studied in another paper [8].

2 Bounds on the characteristic functions

In [6] we gave bounds on the characteristic function of Y. The same method yields, more generally, bounds on the characteristic function of Z_n for arbitrary Z_0. We write $\phi_X(t) := \mathbf{E}\, e^{itX}$ for any random variable X.

Theorem 2.1. *For every real $p \geq 0$ there is a constant $0 < c_p < \infty$ such that for any Z_0 and any $n > p+1$, the characteristic function $\phi_{Z_n}(t)$ satisfies*

$$|\phi_{Z_n}(t)| \leq c_p |t|^{-p} \quad \text{for all } t \in \mathbf{R}. \tag{2.1}$$

The best possible constants c_p satisfy $c_0 = 1$, $c_{1/2} \leq 2$, $c_{3/4} \leq \sqrt{8\pi}$, $c_1 \leq 4\pi$, $c_{3/2} < 187$, $c_{5/2} < 103215$, $c_{7/2} < 197102280$, and the relation

$$c_{p+1} \leq 2^{p+1} c_p^{1+(1/p)} p/(p-1), \qquad p > 1; \tag{2.2}$$

moreover, at least if we restrict (2.1) *to $n \geq p+2$,*

$$c_p \leq 2^{p^2+6p}, \qquad p > 0. \tag{2.3}$$

[The bounds on the constants c_p obtained here are the same as for the special case $Z_0 \overset{\mathcal{L}}{=} Y$ (whence $Z_n \overset{\mathcal{L}}{=} Y$ for every n) in [6]. However, there is no reason to believe that our method yields the best possible bounds, and the best constants for the special case in [6] may be smaller than the best constants in Theorem 2.1 here.]

Proof. The proof is almost identical to the proof of the special case in [6], so we will omit some details. For any random variable Z, we abuse notation slightly and denote by SZ the random variable $h_{Z,Z^*}(U) = UZ + (1-U)Z^* + g(U)$ where U, Z, and Z^*

are independent, with $Z^* \overset{\mathcal{L}}{=} Z$ and $U \sim \text{unif}(0,1)$; thus SZ is a random variable with the distribution $S\mathcal{L}(Z)$. By conditioning on U, we obtain the fundamental relation

$$\phi_{SZ}(t) = \int_0^1 \phi_Z(ut)\,\phi_Z((1-u)t)\,e^{itg(u)}\,du, \qquad t \in \mathbf{R}, \tag{2.4}$$

and thus the estimate

$$|\phi_{SZ}(t)| \le \int_0^1 |\phi_Z(ut)|\;|\phi_Z((1-u)t)|\;du. \tag{2.5}$$

To complete the proof, we give a series of lemmas.

Lemma 2.2. *For any real numbers y and z, the random variable $h_{y,z}(U)$ defined by* (1.2) *satisfies*

$$|\mathbf{E}\,e^{ith_{y,z}(U)}| \le 2|t|^{-1/2}.$$

Proof. This follows by a method of van der Corput [2, 13, 6], using little more than the fact that $h_{y,z}$ is convex with $h''_{y,z} \ge 8$ on $(0,1)$. □

Lemma 2.3. *For any random variable Z and real t, we have $|\phi_{SZ}(t)| \le 2|t|^{-1/2}$.*

Proof. Lemma 2.2 yields

$$|\phi_{SZ}(t)| = \left|\mathbf{E}\,e^{ith_{Z,Z^*}(U)}\right| \le \mathbf{E}\left|\mathbf{E}\Bigl(e^{ith_{Z,Z^*}(U)} \;\Big|\; Z, Z^*\Bigr)\right| \le 2|t|^{-1/2}. \qquad \square$$

Returning to our sequence (Z_n), the preceding lemma applies to all elements except Z_0, i.e.,

$$|\phi_{Z_n}(t)| \le 2|t|^{-1/2}, \qquad n \ge 1, \tag{2.6}$$

which yields the case $p = 1/2$ of Theorem 2.1. We improve the exponent by induction, using (2.5).

Lemma 2.4. *Let $0 < p < 1$. If $|\phi_Z(t)| \le c_p|t|^{-p}$, $t \in \mathbf{R}$, then*

$$|\phi_{SZ}(t)| \le \frac{\bigl[\Gamma(1-p)\bigr]^2}{\Gamma(2-2p)} c_p^2 |t|^{-2p}.$$

Proof. By (2.5) and the hypothesis,

$$|\phi_{SZ}(t)| \le \int_0^1 c_p^2 |ut|^{-p} |(1-u)t|^{-p}\,du = c_p^2 |t|^{-2p} \int_0^1 u^{-p}(1-u)^{-p}\,du,$$

and the result follows by evaluating the beta integral. □

In particular, using (2.6), Lemma 2.4 yields

$$|\phi_{Z_n}(t)| \le 4\pi|t|^{-1}, \qquad n \ge 2. \tag{2.7}$$

This proves (2.1) for $p = 1$, with $c_1 \le 4\pi$. Since $|\phi_{Z_n}(t)| \le 1$, for any $p \le 1$ we trivially have $|\phi_{Z_n}(t)| \le |\phi_{Z_n}(t)|^p$, which by (2.7) establishes (2.1) for all $p \le 1$ with $c_p \le (4\pi)^p$;

applying Lemma 2.4 again, we obtain (2.1) for all $p < 2$. Somewhat better numerical bounds are obtained for $1/2 < p < 1$ by taking a geometric average between the cases $p = 1/2$ and $p = 1$; this yields $c_p \le 2^{2p}\pi^{2p-1}$, $1/2 \le p \le 1$. In particular, we have $c_{3/4} \le \sqrt{8\pi}$, and thus, by Lemma 2.4, $c_{3/2} \le 8\pi^{1/2}\big[\Gamma(1/4)\big]^2 < 186.4 < 187$.

Lemma 2.5. *Let $p > 1$. If $|\phi_Z(t)| \le c_p|t|^{-p}$, $t \in \mathbf{R}$, then*

$$|\phi_{SZ}(t)| \le 2^{p+1}c_p^{1+(1/p)}\frac{p}{p-1}|t|^{-(p+1)}.$$

Proof. This is similar to the proof of Lemma 2.4, substituting the hypothesis (and the trivial $|\phi_Z| \le 1$) into (2.5), but the estimate of the integral is slightly more complicated; for details see [6]. □

Lemma 2.5 completes, by induction, the proof of (2.1) and the estimate (2.2).

The bound for $c_{3/2}$ obtained above and (2.2) now yield (using `Maple`) first $c_{5/2} < 103215$ and then $c_{7/2} < 197102280$. These bounds and (2.2) further yield

$$c_p \le 2^{p^2+5p}, \qquad p = k + \tfrac{3}{2}, \tag{2.8}$$

for integers $k \ge 0$; again see [6] for details. To obtain (2.3) if $p > 1/2$, let $p_1 := \lceil p - \frac{1}{2}\rceil + \frac{1}{2}$. Then, by (2.1) and (2.8), provided $n \ge p + 2 > p_1 + 1$,

$$|\phi_{Z_n}(t)|^{1/p} \le |\phi_{Z_n}(t)|^{1/p_1} \le 2^{p_1+5}|t|^{-1} \le 2^{p+6}|t|^{-1}.$$

The case $p \le 1/2$ follows similarly from (2.6), which completes the proof of Theorem 2.1. □

Remark 2.6. A variety of other bounds are possible. For example, if we begin with the inequality (2.7) and use (2.5), we can easily derive the following result:

$$|\phi_{Z_n}(t)| \le \frac{32\pi^2}{t^2}\Big(\ln\Big(\frac{t}{4\pi}\Big) + 2\Big) \le \frac{32\pi^2 \ln t}{t^2} \quad \text{for all } t \ge 1.72 \text{ and } n \ge 3. \tag{2.9}$$

3 Convergence of densities

It is easily checked that the random variable $h_{y,z}(U)$ is absolutely continuous for every fixed y and z, and thus, by mixing, SZ is absolutely continuous for every Z. In other words, for any Z_0, the random variables Z_n have densities for all $n \ge 1$; cf. [18]. These densities may be unbounded and discontinuous, at least for $n = 1$, as is seen in the case $Z_0 \equiv 0$. However, we now can show that for $n \ge 3$, at least, no such irregularities occur.

Theorem 3.1. *If $n \ge 3$, then Z_n has a bounded continuous density function f_n, for any Z_0. More generally, if $k \ge 0$, then f_n is k times continuously differentiable for all $n \ge k+3$, and there exists a constant C_k independent of Z_0 and n (with $n \ge k+3$) such that $|f_n^{(k)}(x)| \le C_k$, $x \in \mathbf{R}$. Explicitly, $|f_n(x)| \le 16$ when $n \ge 5$, and $|f_n'(x)| \le 2466$ when $n \ge 6$.*

Proof. Theorem 2.1 shows, in particular, that as soon as $n \geq 3$,

$$|\phi_{Z_n}(t)| \leq \min(1, 187|t|^{-3/2}),$$

and thus ϕ_{Z_n} is integrable. This implies, as is well-known (see e.g., [5, Theorem XV.3.3]) that Z_n has a bounded continuous density f_n given by the Fourier inversion formula

$$f_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\, \phi_{Z_n}(t)\, dt, \qquad x \in \mathbf{R}. \tag{3.1}$$

Moreover, using Theorem 2.1 with $p = k + \frac{3}{2}$, we see that $t^k \phi_{Z_n}(t)$ is also integrable when $n \geq k+3$, which by a standard argument shows that f_n is k times differentiable, with

$$f_n^{(k)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^k\, e^{-itx}\, \phi_{Z_n}(t)\, dt, \qquad x \in \mathbf{R}; \tag{3.2}$$

and thus

$$\sup_x |f_n^{(k)}(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |t|^k\, |\phi_{Z_n}(t)|\, dt, \tag{3.3}$$

where the latter integral can be estimated using Theorem 2.1 with $p = k + \frac{3}{2}$.

The argument above yields the bound

$$|f_n(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \min(1, 187|t|^{-3/2})\, dt = \frac{3}{\pi} 187^{2/3} < 31.3, \qquad n \geq 3. \tag{3.4}$$

To obtain better numerical bounds we combine Theorem 2.1 for $p = 0,\ 1/2,\ 3/2,\ 1,\ 5/2,\ 7/2$ and (2.9) (for t in different intervals; see [6] for details); this yields, provided $n \geq 5$, $f_n(x) \leq \frac{1}{2\pi} \int |\phi_n| < 15.3$; similarly, invoking also (2.1) with $p = 9/2$, $f_n'(x) \leq \frac{1}{2\pi} \int |t||\phi_n(t)|\, dt < 2465.9$ for $n \geq 6$. $\square$

Theorem 3.2. *Suppose that* $\mathbf{E}\, Z_0 = 0$ *and* $\mathbf{Var}\, Z_0 < \infty$. *Then the density functions* f_n *of Theorem 3.1 converge uniformly to the (smooth) density function* f *of* Y *at a geometric rate:*

$$\sup_x |f_n(x) - f(x)| = O(r^n) \quad \text{for every fixed } r > (2/3)^{1/2}.$$

Explicitly, for any $p > 1$ *and* $n > p + 1$,

$$\sup_x |f_n(x) - f(x)| \leq \frac{A}{2\pi} \Big(\frac{2c_p}{A}\Big)^{2/(p+1)} \frac{p+1}{p-1} \Big(\frac{2}{3}\Big)^{(\frac{1}{2} - \frac{1}{p+1})n}, \tag{3.5}$$

where $A := (\mathbf{Var}\, Z_0 + \sigma^2)^{1/2}$ *and* c_p *is as in Theorem 2.1. In particular,*

$$\sup_x |f_n(x) - f(x)| \leq 2297A \Big(\frac{2}{3}\Big)^{5n/18} < 2297A(0.8935)^n, \qquad n \geq 5. \tag{3.6}$$

Moreover,

$$\sup_x |f_n(x) - f(x)| \leq \frac{128A}{\pi} \Big(\frac{2}{3}\Big)^{(n/2) - 3.7\sqrt{n}}, \qquad n \geq 3. \tag{3.7}$$

Proof. By the Fourier inversion formula (3.1),

$$|f_n(x) - f(x)| \le \frac{1}{2\pi} \int_{-\infty}^{\infty} |\phi_{Z_n}(t) - \phi_Y(t)| \, dt. \tag{3.8}$$

In order to estimate the right hand side, note that for any random variables X and Y,

$$|\phi_X(t) - \phi_Y(t)| \le \mathbf{E}|e^{itX} - e^{itY}| \le \mathbf{E}|tX - tY| \le |t| \, \|X - Y\|_2;$$

since the characteristic functions here depend on the marginal distributions only, this and the definition (1.5) yield

$$|\phi_X(t) - \phi_Y(t)| \le |t| d_2(\mathcal{L}(X), \mathcal{L}(Y)).$$

In particular, with $\delta_n := d_2(F_n, F)$,

$$|\phi_{Z_n}(t) - \phi_Y(t)| \le |t|\delta_n. \tag{3.9}$$

Further, for any $p > 1$ and $n > p + 1$, Theorem 2.1 yields the estimate

$$|\phi_{Z_n}(t) - \phi_Y(t)| \le |\phi_{Z_n}(t)| + |\phi_Y(t)| \le 2c_p|t|^{-p}.$$

Consequently, for any $T > 0$,

$$\begin{aligned}\int_{-\infty}^{\infty} |\phi_{Z_n}(t) - \phi_Y(t)| \, dt &\le \int_{-T}^{T} \delta_n |t| \, dt + \int_{|t|>T} 2c_p|t|^{-p} \, dt \\ &= \delta_n T^2 + 4\frac{c_p}{p-1} T^{1-p}.\end{aligned}$$

For given n and p, the optimal choice here is $T := (2c_p/\delta_n)^{1/(p+1)}$, giving the bound

$$\int_{-\infty}^{\infty} |\phi_{Z_n}(t) - \phi_Y(t)| \, dt \le \frac{p+1}{p-1} (2c_p)^{2/(p+1)} \delta_n^{1-(2/(p+1))}. \tag{3.10}$$

With (3.8) and the estimate (1.7), this yields (3.5). Choosing $p = 7/2$ and evaluating the constants numerically, using $A \ge \sigma > 0.648$, we obtain (3.6).

To obtain the final estimate, we use (2.3) and observe that, for $p \ge 2$,

$$\Bigl(\frac{2c_p}{A}\Bigr)^{2/(p+1)} \le \Bigl(\frac{2^{p^2+6p+1}}{\sigma}\Bigr)^{2/(p+1)} = 2^{2(p+5)} \Bigl(\frac{2^{-4}}{\sigma}\Bigr)^{2/(p+1)} \le 2^{2(p+5)} \Bigl(1 - \frac{2}{p+1}\Bigr),$$

which by (3.5) yields that for $n \ge p + 2 \ge 4$,

$$\sup_x |f_n(x) - f(x)| \le \frac{A}{2\pi} 2^{2p+10} \Bigl(\frac{2}{3}\Bigr)^{(\frac{1}{2} - \frac{1}{p+1})n}.$$

Choosing the optimal $p := [n \ln(3/2)/(2\ln 2)]^{1/2} - 1$, we find (3.7) [with the constant $(8(\ln 2)/\ln(3/2))^{1/2} < 3.69812$ multiplying $\sqrt{n}$], at least when $n \ge 31$. For $3 \le n \le 30$, (3.7) follows trivially from (3.4), since the right hand side of (3.7) then is larger than 193. $\square$

To test out Theorem 3.2 numerically, choose $Z_0 \equiv 0$, so that $A = \sigma \doteq 0.648$. For $n = 100$, (3.6) yields the bound 0.0192; for $n \geq 177$, (3.7) is better, and yields for example 3.21×10^{-6} for $n = 177$, 2.07×10^{-6} for $n = 180$, and 1.07×10^{-7} for $n = 200$.

Remark 3.3. Similarly, using (3.2), we obtain geometric uniform convergence of the first derivatives, and of any higher derivatives, of the density functions.

Remark 3.4. Suppose that Z_0 has finite moments of all orders. Then, by Lemma 7.2 below, $\mathbf{E}|Z_n|^p$ is finite and stays bounded in n, for each real $0 \leq p < \infty$. It follows that the characteristic functions ϕ_{Z_n} are infinitely differentiable with derivatives bounded uniformly in n. If we apply both Theorem 2.1 and (3.9) to $|\phi_{Z_n}(t) - \phi_Y(t)|$ and take the geometric mean of the resulting bounds, we find, for $n > 2p + 2$,

$$|\phi_{Z_n}(t) - \phi_Y(t)| \leq \left[2c_{2p+1}|t|^{-2p} d_2(F_n, F)\right]^{1/2}.$$

It follows easily by induction on k, using [6, Lemma 2.10], that in fact, for every real $p \geq 0$ and integer $k \geq 0$, there is a constant $c_{p,k}$ [depending on $\mathcal{L}(Z_0)$] such that for all $n > 2^{k+1}p + 2$ we have, with $\rho_k := (2/3)^{2^{-k-2}} < 1$,

$$\sup_{t \in \mathbf{R}} |t|^p \left|\phi_{Z_n}^{(k)}(t) - \phi_Y^{(k)}(t)\right| \leq c_{p,k}\rho_k^n.$$

Omitting details, since the Fourier transform is continuous on the Schwartz space [17]

$$\mathcal{S} := \{f : \sup_t |t|^p |f^{(k)}(t)| < \infty \text{ for all } p, k \geq 0\},$$

it follows that for each k and p, $|x|^p f_n^{(k)}(x)$ converges uniformly to $|x|^p f^{(k)}(x)$ with geometric rate.

Theorem 3.2 treats uniform approximation of f by f_n, using the norm $\|f_n - f\|_\infty := \sup_x |f_n(x) - f(x)|$. We now turn to studying the error in the L^1-norm $\|f_n - f\|_1 := \int_{-\infty}^{\infty} |f_n - f|$.

Note first that, because $\int_{-\infty}^{\infty} f_n = \int_{-\infty}^{\infty} f = 1$,

$$\tfrac{1}{2}\int_{-\infty}^{\infty} |f_n(x) - f(x)|\,dx = \int_{-\infty}^{\infty} \bigl(f(x) - f_n(x)\bigr)^+\,dx,$$

and that this coincides with the total variation distance

$$d_{\mathrm{TV}}(F_n, F) := \sup_{A \subseteq \mathbf{R}} |\mathbf{P}(Z_n \in A) - \mathbf{P}(Y \in A)|; \tag{3.11}$$

moreover, it dominates the Kolmogorov–Smirnov distance

$$d_{\mathrm{KS}}(F_n, F) := \sup_{x \in \mathbf{R}} |\mathbf{P}(Z_n \leq x) - \mathbf{P}(Y \leq x)| \leq d_{\mathrm{TV}}(F_n, F). \tag{3.12}$$

Theorem 3.5. *Suppose that* $\mathbf{E}\,Z_0 = 0$ *and* $\mathbf{Var}\,Z_0 < \infty$. *Then the total variation and Kolmogorov–Smirnov distances between* F_n *and* F *converge geometrically to* 0*:* $d_{\mathrm{KS}}(F_n, F) \leq d_{\mathrm{TV}}(F_n, F) = O(r^n)$ *for every fixed* $r > (2/3)^{1/2}$. *Explicitly, for any* $n \geq 1$,

$$d_{\mathrm{KS}}(F_n, F) \leq d_{\mathrm{TV}}(F_n, F) \leq 135An\left(\frac{2}{3}\right)^{(n/2) - 3.7\sqrt{n}}. \tag{3.13}$$

Proof. For any $a \in (0,1)$,

$$d_{\mathrm{TV}}(F_n, F) = \int_{-\infty}^{\infty} \bigl(f(x) - f_n(x)\bigr)^+ \, dx \le \|f_n - f\|_\infty^{1-a} \int_{-\infty}^{\infty} f(x)^a \, dx, \tag{3.14}$$

where $\|f_n - f\|_\infty$ is estimated in Theorem 3.2. The final integral can be estimated by Hölder's inequality: for any $b > 0$

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)^a \, dx &= \int_{-\infty}^{\infty} f(x)^a e^{ab|x|} \cdot e^{-ab|x|} \, dx \\
&\le \Bigl(\int_{-\infty}^{\infty} f(x) e^{b|x|} \, dx\Bigr)^a \Bigl(\int_{-\infty}^{\infty} e^{-ab|x|/(1-a)} \, dx\Bigr)^{1-a} \\
&\le [\psi(b) + \psi(-b)]^a \Bigl(\frac{2(1-a)}{ab}\Bigr)^{1-a} \\
&= \frac{2}{ab}\Bigl(ab\frac{\psi(b) + \psi(-b)}{2}\Bigr)^a (1-a)^{1-a},
\end{aligned} \tag{3.15}$$

where $\psi(\lambda) := \mathbf{E}\, e^{\lambda Y}$ is the moment generating function of Y. Rösler [15] proved that $\psi(\lambda)$ is finite for all λ; thus $\int f^a < \infty$ for every $a \in (0,1)$, and the first claim follows by (3.14) and Theorem 3.2.

For (3.13) we choose $b = 1/3$, for which it will be shown in Theorem 4.1 below that $\psi(\pm b) \le \exp(1/9) < 1.2$, and thus (3.15) implies $\int_{-\infty}^{\infty} f^a < 2/(ab) = 6/a$. Denoting the right hand side of (3.7) by B, we thus obtain from (3.14) and (3.7), observing that $B \ge (3/2)^{-n/2}$,

$$d_{\mathrm{TV}}(F_n, F) \le \frac{6}{a} B^{1-a} \le \frac{6}{a} (3/2)^{an/2} B.$$

We optimize by taking $a := 2/(n \ln(3/2))$ and obtain the following bound (for $n \ge 5$, so that $a < 1$; smaller n are trivial since $d_{\mathrm{TV}} \le 1$):

$$d_{\mathrm{TV}}(F_n, F) \le 3enB \ln(3/2) = \frac{384e \ln(3/2)}{\pi} An \Bigl(\frac{2}{3}\Bigr)^{(n/2) - 3.7\sqrt{n}}. \qquad \square$$

Remark 3.6. If we are content with a weaker explicit bound, we can avoid invoking estimates of ψ by using moments of Y instead. For example,

$$\int_{-\infty}^{\infty} f(x)^{1/2} \, dx \le \Bigl(\int_{-\infty}^{\infty} f(x)(\sigma^2 + x^2) \, dx\Bigr)^{1/2} \Bigl(\int_{-\infty}^{\infty} \frac{dx}{\sigma^2 + x^2}\Bigr)^{1/2} = (2\pi\sigma)^{1/2} < 2.1$$

and thus

$$d_{\mathrm{KS}}(F_n, F) \le d_{\mathrm{TV}}(F_n, F) \le 2.1 \|f_n - f\|_\infty^{1/2}.$$

4 Bounds on moment generating functions

Letting $\psi_Z(\lambda) := \mathbf{E}\, e^{\lambda Z}$ denote the moment generating function of a random variable Z, we find in analogy with (2.4) the relation

$$\psi_{SZ}(\lambda) = \int_0^1 \psi_Z(u\lambda)\, \psi_Z((1-u)\lambda)\, e^{\lambda g(u)} \, du, \qquad \lambda \in \mathbf{R}. \tag{4.1}$$

In particular, it follows that if $\psi_Z(\lambda)$ is finite for all λ, then so is $\psi_{SZ}(\lambda)$.

Rösler [15] proved that the moment generating function ψ_Y is everywhere finite and that for every $L \geq 0$ there is a constant K_L such that

$$\psi_Y(\lambda) \leq e^{K_L\lambda^2}, \qquad |\lambda| \leq L. \tag{4.2}$$

Moreover, it is implicit in the proof that

$$\text{if } \psi_Z(\lambda) \leq e^{K_L\lambda^2} \text{ for } |\lambda| \leq L, \text{ then } \psi_{SZ}(\lambda) \leq e^{K_L\lambda^2} \text{ for } |\lambda| \leq L. \tag{4.3}$$

Note that (4.3) implies by induction that if we choose $Z_0 \equiv 0$, then $\psi_{Z_n}(\lambda) \leq e^{K_L\lambda^2}$, $|\lambda| \leq L$, for every n, and thus (4.2) follows by Fatou's lemma. More generally, if (4.3) holds and $\psi_{Z_0}(\lambda) \leq e^{K_L\lambda^2}$, $|\lambda| \leq L$, then by induction $\psi_{Z_n}(\lambda) \leq e^{K_L\lambda^2}$, $|\lambda| \leq L$, for every n.

Rösler did not give explicit values of the constants K_L, but such values can be obtained from his proof as follows. [Actually, Rösler [15] treated the somewhat more complicated case of the variables Y_n of (1.1); see [8] for explicit constants in that case. In our case there are some simplifications leading to better constants. Moreover, we introduce some deviations from Rösler's proof designed to improve our bounds.]

Theorem 4.1. *Let $L_0 \doteq 5.018$ be the largest root of $e^L = 6L^2$. Then* (4.2) *and* (4.3) *hold with*

$$K_L = \begin{cases} 1, & L \leq 0.42, \\ 12, & 0.42 < L \leq L_0, \\ 2L^{-2}e^L, & L_0 < L, \end{cases}$$

or any larger number. In particular, we can always take $K_L = \max(2L^{-2}e^L, 12)$.

For $\lambda \leq 0$, we can obtain much better estimates. [For (4.3), we restrict to $\lambda \leq 0$ in both the assumption and the conclusion.]

Theorem 4.2. *We have* (4.2) *and* (4.3) *for $\lambda \leq 0$ with*

$$K_L = \begin{cases} 0.5, & L \leq 0.62, \\ 1.25, & 0.62 < L, \end{cases}$$

or any larger number. In particular, we can always take $K_L = 1.25$ for $\lambda \leq 0$.

Proof of Theorems 4.1 and 4.2. If $\psi_Z(\lambda) \leq e^{K\lambda^2}$ for $|\lambda| \leq L$, then by (4.1), for $|\lambda| \leq L$,

$$\psi_{SZ}(\lambda) \leq \int_0^1 e^{K\lambda^2[u^2+(1-u)^2]+\lambda g(u)}\,du = e^{K\lambda^2}\mathbf{E}e^{\lambda g(U)-2K\lambda^2 U(1-U)}.$$

Hence, (4.3) holds with $K_L = K$ if (and only if)

$$f_K(\lambda) := \mathbf{E}e^{\lambda g(U)-2K\lambda^2 U(1-U)} \leq 1, \qquad \text{when } |\lambda| \leq L. \tag{4.4}$$

Similarly, (4.3) holds with $K_L = K$ for $\lambda \geq 0$ (respectively, for $\lambda \leq 0$) if (4.4) holds for $0 \leq \lambda \leq L$ (resp., for $-L \leq \lambda \leq 0$). Clearly, $f_K(\lambda)$ decreases as K increases, and thus if some K satisfies (4.4), then so does any larger K.

Following Rösler, we argue differently for small and large L in order to find a K satisfying (4.4). For small L we use a Taylor expansion. By straightforward differentiations,

$$
\begin{aligned}
f_K(0) &= 1,\\
f_K'(0) &= \mathbf{E}g(U) = 0,\\
f_K''(0) &= \mathbf{E}\bigl(g(U)^2 - 4KU(1-U)\bigr) = \tfrac{1}{3}\sigma^2 - \tfrac{2}{3}K,\\
f_K'''(\lambda) &= \mathbf{E}\Bigl[\Bigl(\bigl(g(U) - 4K\lambda U(1-U)\bigr)^3 - 12KU(1-U)\bigl(g(U) - 4K\lambda U(1-U)\bigr)\Bigr)\\
&\qquad\qquad \times \exp\bigl(\lambda g(U) - 2K\lambda^2 U(1-U)\bigr)\Bigr].
\end{aligned}
$$

We write the last formula as $f_K'''(\lambda) = \mathbf{E}[X(U,\lambda)]$ and note that $0 \le U(1-U) \le 1/4$ and $-\eta \le g(U) \le 1$, where

$$\eta := -g(\tfrac{1}{2}) = 2\ln 2 - 1 \doteq 0.386.$$

Consider first $\lambda \ge 0$. By Taylor's formula, for $0 \le \lambda \le L$,

$$
\begin{aligned}
f_K(\lambda) &\le 1 + \tfrac{1}{2}\lambda^2 f_K''(0) + \tfrac{1}{6}\lambda^3 \sup_{0\le\xi\le L} f_K'''(\xi)\\
&\le 1 + \tfrac{1}{6}\lambda^2\bigl(\sigma^2 - 2K + L \sup_{0\le\xi\le L} f_K'''(\xi)\bigr)
\end{aligned}
$$

so (4.4) is satisfied for $\lambda \ge 0$ provided

$$L \sup_{0\le\xi\le L} f_K'''(\xi) \le 2K - \sigma^2. \tag{4.5}$$

If $g(U) \ge 0$, we find

$$X(U,\lambda) \le (1 + 3K^2L)e^L, \qquad 0 \le \lambda \le L;$$

while if $g(U) \le 0$, we find

$$X(U,\lambda) \le 3K(\eta + KL), \qquad 0 \le \lambda \le L.$$

For $K \ge 1$, in either case, because $3\eta > 1$,

$$X(U,\lambda) \le (3K\eta + 3K^2L)e^L, \qquad 0 \le \lambda \le L,$$

and thus

$$L \sup_{0\le\xi\le L} f_K'''(\xi) \le L(3K\eta + 3K^2L)e^L.$$

It is readily checked that this is less than $2K - \sigma^2$ so that (4.5) holds, for $K = 1$ and $L = 0.42$.

For larger L, we begin by another crude estimate. Let $W \overset{\mathcal{L}}{=} U/2$ be uniformly distributed on $(0,1/2)$. Then, by $|g(U)| \le 1$ and symmetry,

$$\begin{aligned} f_K(\lambda) &\le e^{|\lambda|}\mathbf{E}\exp\bigl(-2K\lambda^2 U(1-U)\bigr) = e^{|\lambda|}\mathbf{E}\exp\bigl(-2K\lambda^2 W(1-W)\bigr) \\ &\le e^{|\lambda|}\mathbf{E}\exp\bigl(-K\lambda^2 W\bigr) = e^{|\lambda|}\int_0^1 \exp\bigl(-K\lambda^2 u/2\bigr)\,du \\ &= e^{|\lambda|}\frac{1-\exp\bigl(-K\lambda^2/2\bigr)}{K\lambda^2/2} =: g_K(\lambda). \end{aligned} \tag{4.6}$$

Note that that g_K, too, decreases if K is increased. Taking the logarithmic derivative, we find for $\lambda > 0$,

$$\begin{aligned} \bigl(\ln g_K(\lambda)\bigr)' &= 1 - \frac{2}{\lambda} + K\lambda e^{-K\lambda^2/2}\bigl(1-\exp\bigl(-K\lambda^2/2\bigr)\bigr)^{-1} \\ &= 1 - \frac{2}{\lambda} + \frac{K\lambda}{e^{K\lambda^2/2}-1}. \end{aligned} \tag{4.7}$$

For $\lambda \ge 2$, this is evidently positive, and thus g_K then is increasing. Hence, if $K \ge \tilde{K} := 2L^{-2}e^L$, then

$$g_K(\lambda) \le g_K(L) \le g_{\tilde{K}}(L) = 1 - \exp(-e^L) < 1, \qquad 2 \le \lambda \le L.$$

For smaller λ, we take $K = 12$, and check numerically that $g_{12}(0.42) < 1$. Moreover,

$$e^{K\lambda^2/2} - 1 = e^{6\lambda^2} - 1 \ge 6\lambda^2 + 18\lambda^4$$

and further, if $1/3 \le \lambda \le 1$,

$$(1 - \tfrac{\lambda}{2})^{-1} \le 1 + \lambda \le 1 + 3\lambda^2$$

and thus

$$\frac{K\lambda}{e^{K\lambda^2/2}-1} \le \frac{12\lambda}{6\lambda^2(1+3\lambda^2)} \le \frac{2}{\lambda}\Bigl(1-\frac{\lambda}{2}\Bigr) = \frac{2}{\lambda} - 1.$$

Hence, (4.7) shows that g_{12} is decreasing on $[1/3,1]$, and thus

$$g_{12}(\lambda) \le g_{12}(0.42) < 1, \qquad 0.42 \le \lambda \le 1.$$

Finally,

$$g_{12}(\lambda) \le \frac{1}{6}\frac{e^\lambda}{\lambda^2} \le \frac{e}{6} < 1, \qquad 1 \le \lambda \le 2.$$

Combining these estimates we find that if $K \ge \max(12, 2L^{-2}e^L)$, then $f_K(\lambda) \le g_K(\lambda) < 1$ whenever $0.42 \le \lambda \le L$, while $f_K(\lambda) \le f_1(\lambda) \le 1$ for $0 \le \lambda \le 0.42$, and thus (4.4) holds when $\lambda \ge 0$.

We have also shown that $K = 12$ will do for $L \le 2$ and $\lambda \ge 0$; since $2L^{-2}e^L$ is increasing for $L \ge 2$, and thus less than 12 for $2 \le L < L_0$ but larger than 12 for $L > L_0$, Theorem 4.1 for $\lambda \ge 0$ follows.

For $\lambda \le 0$, we again use Taylor's formula for small $|\lambda|$; arguing as above we see that (4.4) holds for $\lambda \le 0$ provided

$$L \sup_{-L\le\xi\le 0} \bigl(-f_K'''(\xi)\bigr) \le 2K - \sigma^2. \tag{4.8}$$

It is easily checked numerically that $\max_{0\le u\le 1} u(1-u)g(u) < 0.033$. It follows that

$$X(u,\lambda) \ge (-\eta^3 - 0.396K - 3K^2L)e^{\eta L}, \qquad -L \le \lambda \le 0.$$

Hence, (4.8) holds and (4.4) is satisfied for $\lambda \le 0$ provided

$$(\eta^3 + 0.396K + 3K^2L)Le^{\eta L} \le 2K - \sigma^2.$$

It is readily checked that this holds for $K = 0.5$ and $L = 0.62$.

For larger L we argue as follows. The function $h(u) := g(u) + 4\eta u(1-u)$ satisfies

$$h''(u) = \frac{2}{u(1-u)} - 8\eta \ge 8 - 8\eta > 0, \qquad 0 < u < 1.$$

Hence h is convex, and since $h'(1/2) = 0$,

$$h(u) \ge h(\tfrac12) = 0, \qquad 0 \le u \le 1.$$

Consequently, if $\lambda \le 0$ and $K|\lambda| \ge 2\eta$, then

$$\lambda g(U) - 2K\lambda^2 U(1-U) \le \lambda h(U) \le 0$$

and thus $f_K(\lambda) \le 1$. Choosing $K = 2\eta/0.62 < 1.247$, this shows $f_K(\lambda) \le 1$ for $\lambda \le -0.62$, while $f_K(\lambda) \le f_{0.5}(\lambda) \le 1$ for $-0.62 \le \lambda \le 0$ by the preceding case.

This completes the proof of both theorems. □

If we just want a bound on ψ_Y, (4.2) and Theorems 4.1 and 4.2 can be stated more simply as follows (ignoring the better bounds obtained for small λ).

Corollary 4.3. *With L_0 as in Theorem 4.1,*

$$\psi_Y(\lambda) \le \begin{cases} e^{1.25\lambda^2}, & \lambda \le 0,\\ e^{12\lambda^2}, & 0 \le \lambda \le L_0,\\ e^{2e^\lambda}, & \lambda \ge L_0. \end{cases}$$

In particular, $\psi_Y(\lambda) \le \exp(\max(12\lambda^2, 2e^\lambda))$. □

The bound e^{2e^λ} is very large even for moderately large λ, but the next result shows that $\psi_Y(\lambda)$ really is of essentially this size. In particular, it follows that $\ln\ln\psi_Y(\lambda) \sim \lambda$ as $\lambda \to +\infty$.

Theorem 4.4. *If $\gamma < 2/e$, then for sufficiently large λ,*

$$\psi_Y(\lambda) \ge \exp(\gamma\lambda^{-1}e^\lambda).$$

Proof. Since a moment generating function is convex and $\psi_Y'(0) = \mathbf{E}Y = 0$, ψ_Y is increasing on $[0,\infty)$. Moreover, g is decreasing on $[0,1/2]$. Hence, if $0 \le \delta \le 1/2$, the integrand in (4.1) with $Z = Y$ is for $0 \le u \le \delta$ at least $\psi_Y(0)\psi_Y((1-\delta)\lambda)e^{\lambda g(\delta)}$ and the same holds for $1-\delta \le u \le 1$ by symmetry. Consequently,

$$\psi_Y(\lambda) \ge 2\int_0^\delta \psi_Y(u\lambda)\,\psi_Y((1-u)\lambda)\,e^{\lambda g(u)}\,du \ge 2\delta\psi_Y((1-\delta)\lambda)e^{\lambda g(\delta)}, \qquad 0 \le \delta \le 1/2. \tag{4.9}$$

Let $a > 1/2$ be a constant to be determined later and choose $\delta := ae^{-\lambda}$. Then $g(\delta) = 1 - O(\lambda e^{-\lambda})$ and thus by (4.9), for $\lambda \ge \ln(2a)$,

$$\psi_Y(\lambda) \ge 2ae^{-O(\lambda^2 e^{-\lambda})}\psi_Y(\lambda - a\lambda e^{-\lambda}).$$

If $0 < \varepsilon < 2a$, there thus exists A such that for $\lambda \ge A$,

$$\psi_Y(\lambda) \ge (2a-\varepsilon)\psi_Y(\lambda - a\lambda e^{-\lambda}).$$

Given $\lambda \ge A$, let $\lambda_0 := \lambda$ and define inductively $\lambda_{n+1} := \lambda_n - a\lambda_n e^{-\lambda_n}$, $n \ge 0$. Let N be the smallest integer with $\lambda_N < A$. Then $\psi_Y(\lambda_n) \ge (2a-\varepsilon)\psi_Y(\lambda_{n+1})$, $n = 0,\dots,N-1$, and thus

$$\psi_Y(\lambda) = \psi_Y(\lambda_0) \ge (2a-\varepsilon)^N\psi_Y(\lambda_N) \ge (2a-\varepsilon)^N.$$

It remains to estimate N from below. Since e^x is increasing,

$$\int_{\lambda_{n+1}}^{\lambda_n} e^x\,dx \le e^{\lambda_n}(\lambda_n - \lambda_{n+1}) = a\lambda_n \le a\lambda$$

and thus

$$Na\lambda \ge \int_{\lambda_N}^{\lambda_0} e^x\,dx \ge \int_A^\lambda e^x\,dx = e^\lambda - e^A.$$

Consequently,

$$\ln\psi_Y(\lambda) \ge N\ln(2a-\varepsilon) \ge \frac{\ln(2a-\varepsilon)}{a}\lambda^{-1}(e^\lambda - e^A), \qquad \lambda \ge A.$$

We choose $a = e/2$, which maximizes $\ln(2a)/a$. Then $\ln(2a)/a = 2/e$, we may choose ε so small that $\ln(2a-\varepsilon)/a > \gamma$, and the result follows. ∎

As is well known, bounds on the moment generating function yield bounds on the tails of the distribution.

Theorem 4.5. *If $y \ge 2e^{L_0} = 12L_0^2 \doteq 302.1$, then*

$$\mathbf{P}(Y \ge y) \le \exp\bigl(-y(\ln y - 1 - \ln 2)\bigr).$$

Proof. For $\lambda \ge L_0$, by Corollary 4.3,

$$\mathbf{P}(Y \ge y) \le e^{-\lambda y}\mathbf{E}e^{\lambda Y} \le \exp(2e^\lambda - y\lambda),$$

and the result follows by taking $\lambda = \ln(y/2)$. ∎

Remark 4.6. The same estimate holds for every Z_n provided, say, $\psi_{Z_0}(\lambda) \le \exp(12\lambda^2)$; for example, when $Z_0 \equiv 0$.

Theorem 4.4 suggests that the true size of $\mathbf{P}(Y \ge y)$ is (for large y) not much smaller than the upper bound in Theorem 4.5. Indeed, Knessl and Szpankowski [11] have found (assuming an as yet unverified regularity hypothesis) a much more precise formula for the asymptotics of $\mathbf{P}(Y \ge y)$ which is of the order $\exp\bigl(-y[\ln y + \ln\ln y + O(1)]\bigr)$.

For the left tail, Corollary 4.3 similarly implies $\mathbf{P}(Y \le y) \le \exp(-y^2/5)$ for $y \le 0$, but this result is much weaker than the doubly exponential decay found by Knessl and Szpankowski [11].

5 Geometric rate of convergence for moment generating functions

Theorem 5.1. *Suppose that Z_0 has mean zero and an everywhere finite moment generating function ψ_{Z_0}. Then $\psi_{Z_n}(\lambda) \to \psi_Y(\lambda)$ at a geometric rate for every fixed $\lambda \in \mathbf{R}$. Explicitly, if $L \ge 0$ and K_L are such that* (4.2) *and* (4.3) *hold, and if moreover*

$$\psi_{Z_0}(\lambda) \le e^{K_L\lambda^2}, \qquad |\lambda| \le L, \tag{5.1}$$

then, for every $n \ge 0$ and $|\lambda| \le L/2$,

$$\begin{aligned} |\psi_{Z_n}(\lambda) - \psi_Y(\lambda)| &\le (\mathbf{Var}\, Z_0 + \sigma^2)^{1/2}|\lambda|\bigl(\psi_{Z_n}(2\lambda) + \psi_Y(2\lambda)\bigr)^{1/2}(2/3)^{n/2} \\ &\le 2^{1/2}(\mathbf{Var}\, Z_0 + \sigma^2)^{1/2}|\lambda|e^{2K_L\lambda^2}(2/3)^{n/2}. \end{aligned}$$

Of course, if the hypotheses in the first sentence of the theorem's statement are met, then, given $L \ge 0$, (5.1) holds for some $K_L < \infty$, which by Theorem 4.1 may be chosen so large that (4.2) and (4.3) hold, too.

Proof. By (4.3) and induction, the estimate (5.1) holds for every ψ_{Z_n}. Fix $n \ge 0$ and consider the optimal d_2-coupling of (the laws of) Z_n and Y. Then for $\lambda \in [-L/2, L/2]$ we have, using the mean value theorem and the Cauchy–Schwarz inequality,

$$\begin{aligned} \left|\mathbf{E}\, e^{\lambda Z_n} - \mathbf{E}\, e^{\lambda Y}\right| &\le \mathbf{E}\left|e^{\lambda Z_n} - e^{\lambda Y}\right| \\ &\le \mathbf{E}\left(|\lambda||Z_n - Y|e^{\max(\lambda Z_n, \lambda Y)}\right) \\ &\le |\lambda|\bigl(\mathbf{E}|Z_n - Y|^2\bigr)^{1/2}\left(\mathbf{E}\, e^{2\max(\lambda Z_n, \lambda Y)}\right)^{1/2} \\ &\le |\lambda|\bigl(\mathbf{E}|Z_n - Y|^2\bigr)^{1/2}\left(\mathbf{E}\, e^{2\lambda Z_n} + \mathbf{E}\, e^{2\lambda Y}\right)^{1/2}. \end{aligned}$$

By the optimality of the coupling and (1.7),

$$\bigl(\mathbf{E}|Z_n - Y|^2\bigr)^{1/2} = d_2(F_n, F) \le (\mathbf{Var}\, Z_0 + \sigma^2)^{1/2}(2/3)^{n/2},$$

and by (4.2) and (5.1) for ψ_{Z_n}

$$\mathbf{E}\, e^{2\lambda Z_n} + \mathbf{E}\, e^{2\lambda Y} \le 2e^{K_L(2\lambda)^2},$$

whence the result follows. □

Note further that the operator $\psi_Z \mapsto \psi_{SZ}$ given by (4.1) is monotone, in the specific sense that if $\psi_Z(\lambda) \le \psi_W(\lambda)$ for $|\lambda| \le L$, then also $\psi_{SZ}(\lambda) \le \psi_{SW}(\lambda)$ for $|\lambda| \le L$. In particular, by induction, if $\psi_{Z_0}(\lambda) \le \psi_{Z_1}(\lambda)$ for $|\lambda| \le L$, then $\psi_{Z_n}(\lambda)$ increases monotonically to its limit $\psi_Y(\lambda)$ for $|\lambda| \le L$. Likewise, if $\psi_{Z_0}(\lambda) \ge \psi_{Z_1}(\lambda)$ for $|\lambda| \le L$, then $\psi_{Z_n}(\lambda)$ decreases to $\psi_Y(\lambda)$ for $|\lambda| \le L$.

We give two simple special cases.

Corollary 5.2. *Suppose that $Z_0 \equiv 0$. Then $\psi_{Z_n}(\lambda)$ increases monotonically to $\psi_Y(\lambda)$ for every fixed λ. If $L \ge 0$ and K_L are such that* (4.2) *and* (4.3) *hold, then, for every $n \ge 0$ and $|\lambda| \le L/2$,*

$$\begin{aligned} 0 \le \psi_Y(\lambda) - \psi_{Z_n}(\lambda) &\le 2^{1/2}\sigma|\lambda| \left(\psi_Y(2\lambda)\right)^{1/2} (2/3)^{n/2} \\ &\le 2^{1/2}\sigma|\lambda| e^{2K_L\lambda^2} (2/3)^{n/2}. \end{aligned}$$

Proof. Since $\mathbf{E}\, Z_1 = 0$, by Jensen's inequality $\psi_{Z_1}(\lambda) \ge 1 = \psi_{Z_0}(\lambda)$, and the monotonicity follows. In particular, $\psi_{Z_n}(2\lambda) \le \psi_Y(2\lambda)$, and since (5.1) trivially is satisfied, the result follows from Theorem 5.1. $\square$

Corollary 5.3. *Suppose $L \ge 0$ and K_L are such that* (4.2) *and* (4.3) *hold, and let $Z_0 \sim N(0, 2K_L)$. Then $\psi_{Z_n}(\lambda)$ decreases monotonically to $\psi_Y(\lambda)$ for every fixed λ with $|\lambda| \le L$, and, for every $n \ge 0$ and $|\lambda| \le L/2$,*

$$\begin{aligned} 0 \le \psi_{Z_n}(\lambda) - \psi_Y(\lambda) &\le (4K_L + 2\sigma^2)^{1/2}|\lambda| \left(\psi_{Z_n}(2\lambda)\right)^{1/2} (2/3)^{n/2} \\ &\le (4K_L + 2\sigma^2)^{1/2}|\lambda| e^{2K_L\lambda^2} (2/3)^{n/2}. \end{aligned}$$

Proof. Since $\psi_{Z_0}(\lambda) = e^{K_L\lambda^2}$, (4.3) yields $\psi_{Z_1}(\lambda) \le \psi_{Z_0}(\lambda)$, and the monotonicity follows. The estimate thus follows from Theorem 5.1. $\square$

6 On numerical calculations

The preceding results make it possible, in principle at least, to calculate the density, distribution, characteristic, and moment generating functions of Y numerically, with provable arbitrarily high accuracy.

To begin, the results of earlier sections show that it suffices to start with a suitable $\mathcal{L}(Z_0)$, for example unit mass at 0 or a normal distribution, and then calculate the corresponding quantity for Z_n, for a large n that can be determined. The distribution of Z_n can be calculated recusively; for the characteristic and moment generating functions we have the recurrence relations (2.4) and (4.1), while for the density functions we have the following recursion:

Theorem 6.1. *If $n \ge 0$ is arbitrary and Z_0 has a bounded continuous density function f_0, or if Z_0 is arbitrary and $n \ge 3$, then Z_n and Z_{n+1} have bounded continuous density functions f_n and f_{n+1} satisfying the identity*

$$f_{n+1}(x) = \int_{u=0}^{1} \int_{z \in \mathbf{R}} f_n(z)\, f_n\left(\frac{x - g(u) - (1-u)z}{u}\right) \frac{1}{u}\, dz\, du, \qquad x \in \mathbf{R}, \tag{6.1}$$

with $g(\cdot)$ given by (1.3).

Proof. Our proof (similar to that of Theorem 4.1 in [6]) is by induction on $n \geq 0$ in the first case and on $n \geq 3$, using Theorem 3.1 to get started, in the second case. We may therefore assume as our induction hypothesis that f_n is bounded and continuous. It is easily checked that, for each $0 < u < 1$, the inner integral

$$h_u(x) := \int_{z \in \mathbf{R}} f_n(z)\, f_n\left(\frac{x - g(u) - (1-u)z}{u}\right) \frac{1}{u}\, dz$$

is a density function for the random variable

$$uZ_n + (1-u)Z_n^* + g(u), \tag{6.2}$$

and, using dominated convergence, that h_u is bounded and continuous. Indeed, $h_u(x) \leq (\sup f_n)/u$, and since $h_u = h_{1-u}$ by symmetry in (6.2), $h_u(x) \leq 2 \sup f_n$, uniformly in u and x. It follows, by dominated convergence again, that $x \mapsto f_{n+1}(x) = \int_0^1 h_u(x)\, du$ is a bounded continuous density for Z_{n+1}. $\square$

The integrals in (2.4), (4.1), or (6.1) have to be computed numerically—as does the integral of f_n to get F_n—but that can be done with arbitrary precision since the results above provide bounds for the integrands and their derivatives. [The function $g(u)$ has an unbounded derivative as $u \to 0$ or $u \to 1$, but that can be handled by truncating the interval.] Consequently, to calculate $\phi_{Z_n}(t)$ with given precision for a given t, it suffices to know $\phi_{Z_{n-1}}(t_k)$ with another given precision for a finite number of points t_k, which can be done recursively. (However, a brute force recursion along these lines seems to require too many numerical integrations to be practical if we want reasonably high provable accuracy.)

Remark 6.2. To calculate the density f_n numerically, it might be better to compute ϕ_{Z_n} recursively by (2.4) and then use (3.1), instead of using the recursion (6.1) directly. This is both because (6.1) is a double integral and because we have the simple bounds $|\phi_n| \leq 1$ and $|\phi_n^{(k)}| \leq \mathbf{E}|Z_n|^k$, $k \geq 1$.

7 The metrics d_p and a lower bound on $d_2(F_n, F)$

At (1.7) we recalled Rösler's fundamental result

$$d_2(F_n, F) = O(\rho^n)$$

with $\rho := (2/3)^{1/2}$. The question naturally arises as to whether there is a lower bound that matches at least to the extent that

$$d_2(F_n, F) = \Omega(r^n)$$

for some $r > 0$. Of course, the answer is negative without any further restrictions, since if $F_0 = F$ then $F_n = F$ for every n. However, our main result of this section asserts that this is the only exception, at least among distributions F_0 with finite moments of all orders:

Theorem 7.1. *If $F_0 \neq F$ has finite moments of all orders, then there exists $r > 0$ (depending on F_0) so that*

$$d_2(F_n, F) = \Omega(r^n).$$

Our arguments for Theorem 7.1 will require use of metrics d_p generalizing (1.5). So we will warm up in Section 7.1 by recalling the definition of, and two useful facts about, d_p and in Section 7.2 by extending the upper bound result (1.7) to d_p for $p \geq 1$. Then in Section 7.3 we will prove a sharpened version of Theorem 7.1 (namely, Theorem 7.7).

7.1 The metrics d_p

For real $1 \leq p < \infty$, let $\|X\|_p := (\mathbf{E}|X|^p)^{1/p}$ denote the L^p-norm, and let d_p denote the metric on the space of probability distributions with finite L^p-norm defined by

$$d_p(F, G) := \min \|X - Y\|_p,$$

taking the minimum, as at (1.5), over all couplings of $\mathcal{L}(X) = F$ and $\mathcal{L}(Y) = G$. It is worth noting that there is a coupling [namely, $X = F^{-1}(U)$ and $Y = G^{-1}(U)$ for U uniform and a suitable definition of the inverse probability transform F^{-1}] that achieves the minimum simultaneously for each $1 \leq p < \infty$ (assuming F and G have finite moments of all orders): see [1].

We begin with two elementary facts that will be useful later. The proof of the first fact (Lemma 7.2) shows that S is a contraction for the d_p-metric.

Lemma 7.2. *Consider real $1 \leq p < \infty$. The d_p-distance from the limiting* `Quicksort` *distribution F does not increase when the operator S of* (1.4) *is applied. Therefore, $d_p(F_n, F)$ is nonincreasing, and hence bounded, in n if $\mathbf{E}\,|Z_0|^p < \infty$.*

Proof. With a slight abuse of notation, we find, for Z with any law,

$$d_p(SZ, Y) = d_p(SZ, SY) \leq \|U(Z - Y) + (1 - U)(Z^* - Y^*)\|_p, \tag{7.1}$$

coupling (Y, Z) optimally and (Y^*, Z^*) optimally and choosing U, (Y, Z), and (Y^*, Z^*) to be independent. In calculating the L^p-norm value on the right in (7.1), condition on U and then use subadditivity of L^p-norm together with independence to bound that value by $\|Z - Y\|_p = d_p(Z, Y)$. This establishes the first assertion: $d_p(SZ, Y) \leq d_p(Z, Y)$.

Therefore, $d_p(F_n, F) = d_p(S^n Z_0, Y)$ is nonincreasing, and hence bounded by $d_p(Z_0, Y)$, which is bounded by $\|Z_0\|_p + \|Y\|_p < \infty$ if $\mathbf{E}\,|Z_0|^p < \infty$. □

Remark 7.3. Conversely, if $\|SZ\|_p < \infty$, then $\mathbf{E}|uZ + (1 - u)Z^*|^p < \infty$ for some $u \in (0, 1)$, and thus $\mathbf{E}|Z|^p < \infty$ too. Hence, if $\mathbf{E}|Z_0|^p = \infty$, then $\mathbf{E}|Z_n|^p = \infty$ and $d_p(Z_n, Y) = \infty$ for all n.

Lemma 7.4. *For real $2 < p < q < \infty$ we have, for any F and G,*

$$d_p(F, G) \leq d_2^{\frac{2(q-p)}{p(q-2)}}(F, G) \times d_q^{\frac{q(p-2)}{p(q-2)}}(F, G).$$

Proof. Using the common optimal coupling for d_2 and d_q, this is immediate from the inequality

$$\|X\|_p \le \|X\|_2^{\frac{2(q-p)}{p(q-2)}} \|X\|_q^{\frac{q(p-2)}{p(q-2)}},$$

which in turn follows from the fact [16, Exercise 4(b) of Chapter 3] that $\ln \|X\|_p^p$ is convex in $p \in (0, \infty)$. $\square$

7.2 Geometric rate of convergence in each metric d_p

Under suitable conditions, we can establish a geometric rate of convergence for $d_p(F_n, F)$ for any real $1 \le p < \infty$. We begin with an elementary lemma.

Lemma 7.5. *If $p \ge 2$, then for all $x, y \ge 0$,*

$$(x+y)^p \le x^p + y^p + c_p(x^{p-1}y + xy^{p-1}),$$

where $c_p := p(p-1)2^{p-2}$.

Proof.

$$\begin{aligned}(x+y)^p - x^p - y^p &= \int_0^y p\big((x+t)^{p-1} - t^{p-1}\big)\,dt = \int_{t=0}^y \int_{u=0}^x p(p-1)(t+u)^{p-2}\,du\,dt \\ &\le p(p-1)xy(x+y)^{p-2} \le p(p-1)xy2^{p-2}(x^{p-2} + y^{p-2}).\end{aligned}$$

$\square$

Theorem 7.6. *Let $p_0 \doteq 6.557$ be the largest positive solution to*

$$\Big(\frac{2}{p_0+1}\Big)^{1/p_0} = \Big(\frac{2}{3}\Big)^{1/2} \tag{7.2}$$

and let, for any $\varepsilon > 0$,

$$\beta_p := \begin{cases} \big(\frac{2}{3}\big)^{1/2}, & 1 \le p < p_0, \\ \big(\frac{2}{3}\big)^{1/2} + \varepsilon, & p = p_0, \\ \big(\frac{2}{p+1}\big)^{1/p} & p > p_0. \end{cases} \tag{7.3}$$

Thus, for $p \ge 2$ except $p = p_0$, $\beta_p = \max\big((2/3)^{1/2}, (2/(p+1))^{1/p}\big)$. Then, for any Z_0 with zero mean and finite variance, and every $p \ge 1$ such that $\mathbf{E}|Z_0|^p < \infty$, there exists a constant $\alpha_p < \infty$ [depending on $\mathcal{L}(Z_0)$] such that

$$d_p(Z_n, Y) \le \alpha_p \beta_p^n. \tag{7.4}$$

Proof. First we note that (7.2) can be written $(3/2)^{p_0/2} = (p_0+1)/2$. One root of this equation is 2, and since $(3/2)^{p/2}$ is convex, with derivative less than 1/2 at $p = 2$, it follows that the equation has two positive roots, 2 and $p_0 > 2$, and that $(2/3)^{p/2} > 2/(p+1)$ for $2 < p < p_0$, while $(2/3)^{p/2} < 2/(p+1)$ for $p > p_0$.

Next we note that (7.4) holds for $p \le 2$, with

$$\alpha_p := (\mathbf{Var}\, Z_0 + \sigma^2)^{1/2}, \qquad p \le 2,$$

by (1.7) and the inequality $d_p \le d_2$, $p \le 2$. We then proceed by induction on $\lfloor p \rfloor$. For the induction step, suppose that $p > 2$ and that Z_0 has zero mean and satisfies $\|Z_0\|_p < \infty$. By the induction hypothesis, there exist constants $0 < \alpha_q < \infty$, $1 \le q \le p-1$, such that

$$d_q(Z_n, Y) \le \alpha_q \beta_q^n \quad \text{for all } q \le p-1 \text{ and } n \ge 0. \tag{7.5}$$

Using our usual coupling of Z_n and Y in terms of the optimal coupling of Z_{n-1} and Y, we find easily by Lemma 7.5, for $n \ge 1$ [with (Z_{n-1}, Y), (Z^*_{n-1}, Y^*), and U independent],

$$\begin{aligned}
d_p^p(Z_n, Y) &\le \mathbf{E}\big|U(Z_{n-1} - Y) + (1-U)(Z^*_{n-1} - Y^*)\big|^p \\
&\le \mathbf{E}\big(U|Z_{n-1} - Y| + (1-U)|Z^*_{n-1} - Y^*|\big)^p \\
&\le \mathbf{E}\big(U^p|Z_{n-1} - Y|^p\big) + \mathbf{E}\big((1-U)^p|Z^*_{n-1} - Y^*|^p\big) \\
&\qquad + c_p\mathbf{E}\big(U^{p-1}(1-U)|Z_{n-1} - Y|^{p-1}|Z^*_{n-1} - Y^*|\big) \\
&\qquad + c_p\mathbf{E}\big(U(1-U)^{p-1}|Z_{n-1} - Y||Z^*_{n-1} - Y^*|^{p-1}\big) \\
&= \frac{2}{p+1} d_p^p(Z_{n-1}, Y) + \frac{2c_p}{p(p+1)} d_1(Z_{n-1}, Y)\, d_{p-1}^{p-1}(Z_{n-1}, Y).
\end{aligned}$$

So by induction on n it follows that

$$d_p^p(Z_n, Y) \le \Big(\frac{2}{p+1}\Big)^n d_p^p(Z_0, Y) + \frac{c_p}{p} \sum_{i=0}^{n-1} \Big(\frac{2}{p+1}\Big)^{n-i} d_1(Z_i, Y) d_{p-1}^{p-1}(Z_i, Y).$$

By the induction hypothesis (7.5) this yields, for some $a_1, a_2 < \infty$ (depending on p),

$$d_p^p(Z_n, Y) \le a_1 \Big(\frac{2}{p+1}\Big)^n + a_2 \sum_{i=0}^{n-1} \Big(\frac{2}{p+1}\Big)^{n-i} \big(\beta_1 \beta_{p-1}^{p-1}\big)^i. \tag{7.6}$$

Let $\gamma := \beta_1 \beta_{p-1}^{p-1}$. We break our treatment into three cases:

(i) If $\gamma > 2/(p+1)$, we write the sum in (7.6) as

$$\gamma^n \sum_{i=0}^{n-1} \left(\frac{2}{p+1}\gamma^{-1}\right)^{n-i} < \left(1 - \frac{2}{p+1}\gamma^{-1}\right)^{-1} \gamma^n,$$

and thus (7.6) shows that (7.4) holds with $\beta_p^p = \gamma$.

(ii) If $\gamma < 2/(p+1)$, we write the sum in (7.6) as

$$\Big(\frac{2}{p+1}\Big)^n \sum_{i=0}^{n-1} \left(\gamma \frac{p+1}{2}\right)^i < \Big(1 - \gamma\frac{p+1}{2}\Big)^{-1} \Big(\frac{2}{p+1}\Big)^n,$$

and thus (7.4) holds with $\beta_p^p = 2/(p+1)$.

(iii) If $\gamma = 2/(p+1)$, the sum in (7.6) equals $n\big(2/(p+1)\big)^n$. Consequently, (7.4) holds with any $\beta_p > \big(2/(p+1)\big)^{1/p}$.

It remains to verify that this yields the β_p given in (7.3).

First, if $2 < p < p_0$, then the induction hypothesis yields $\gamma = \beta_1 \beta_{p-1}^{p-1} = (2/3)^{p/2} > 2/(p+1)$, so case (i) gives $\beta_p^p = (2/3)^{p/2}$. Similarly, for $p = p_0$, $\gamma = (2/3)^{p/2} = 2/(p+1)$ and (iii) shows that any $\beta_p > \bigl(2/(p+1)\bigr)^{1/p} = (2/3)^{1/2}$ will do.

For $p_0 < p < p_0 + 1$, we have $\gamma = \beta_1 \beta_{p-1}^{p-1} = (2/3)^{p/2} < 2/(p+1)$, so case (ii) yields $\beta_p^p = 2/(p+1)$. The same applies for $p = p_0 + 1$, since again $(2/3)^{p/2} < 2/(p+1)$ and we thus may choose ε so small that $\gamma = (\frac{2}{3})^{1/2}\bigl((\frac{2}{3})^{1/2} + \varepsilon\bigr)^{p-1} < 2/(p+1)$.

Finally, for $p > p_0 + 1$, we have

$$\gamma = \beta_1 \beta_{p-1}^{p-1} = \left(\frac{2}{3}\right)^{1/2} \frac{2}{p} < \frac{2}{p+1},$$

since $p/(p+1)$ is increasing and equals $(2/3)^{1/2}$ when $p = (\sqrt{3/2} - 1)^{-1} = 2(\sqrt{6} - 2)^{-1} = \sqrt{6} + 2 < 5 < p_0$. Hence case (ii) applies. $\square$

7.3 Lower bounds

The main goal of this subsection is to establish Theorem 7.1, or rather the sharper Theorem 7.7 below. Since (as noted in Section 4) the limiting `Quicksort` distribution F has everywhere finite moment generating function, it is uniquely determined by its moments. Hence if $F_0 \neq F$ has finite moments of all orders, then $\mathbf{E}\, Z_0^j \neq \mathbf{E}\, Y^j$ for some integer $j \geq 1$.

Theorem 7.7. *Suppose $F_0 \neq F$ has finite moments of all orders, and let p be the smallest positive integer such that $\mathbf{E}\, Z_0^p \neq \mathbf{E}\, Y^p$. Then, for any $0 < r < \bigl(\frac{2}{p+1}\bigr)^{p/2}$,*

$$d_2(F_n, F) = \Omega(r^n).$$

(The implicit multiplicative constant depends on both F_0 and r.)

Remark 7.8. The cases $p = 1$ and $p = 2$ are a bit special, and in these cases we claim that Theorem 7.7 holds even with $r = \bigl(\frac{2}{p+1}\bigr)^{p/2}$, i.e., with $r = 1$ and $r = 2/3$, respectively, and without the assumption that F_0 has finite moments.

We may and shall assume that $\mathbf{E}\, Z_0^2 < \infty$, since otherwise $d_2(F_n, F) = \infty$.

First, $p = 1$ when $\mathbf{E}\, Z_0 \neq \mathbf{E}\, Y = 0$; in this case Z_n converges in distribution to $Y + \mathbf{E}\, Z_0$ and not to Y, and thus $\inf d_2(Z_n, Y) > 0$, i.e., the theorem holds with $r = 1$. Indeed, we have the sharper result that

$$d_2(Z_n, Y) = d_2(Y + \mathbf{E}\, Z_0, Y) + O\bigl(d_2(Z_n, Y + \mathbf{E}\, Z_0)\bigr) = |\mathbf{E}\, Z_0| + O\bigl((2/3)^{n/2}\bigr).$$

Next, $p = 2$ when $\mathbf{E}\, Z_0 = 0$ but $\mathbf{Var}\, Y \neq \mathbf{Var}\, Z_0$; in this case Theorem 7.9 shows that the result holds with $r = 2/3$. Even in this case we have a gap between the lower bound $\Omega\bigl((2/3)^n\bigr)$ and Rösler's upper bound $O\bigl((2/3)^{n/2}\bigr)$; it is an open problem to find the rate of approximation more precisely.

We prove Theorem 7.7 using the following Theorem 7.9, which is a similar lower bound for the d_p-metric.

Theorem 7.9. *Let $p \geq 1$ be an integer, and suppose that $\mathbf{E}\, Z_0^j = \mathbf{E}\, Y^j$ for integers $1 \leq j \leq p-1$, and that $\mathbf{E}\, Z_0^p$ exists and is finite but fails to equal $\mathbf{E}\, Y^p$. Then*

$$d_p(F_n, F) = \Omega\left(\left(\tfrac{2}{p+1}\right)^n\right).$$

Theorem 7.9 is, in turn, a simple consequence of the following two elementary lemmas.

The first lemma demonstrates a sense in which the value of p in Theorems 7.7 and 7.9 persists from F_0 to each F_n; the second gives a general lower bound on d_p in terms of discrepancy in pth moments.

Lemma 7.10. *Let $p \geq 1$ be an integer, and suppose for $n = 0$ that $\mathbf{E}\, Z_n^j = \mathbf{E}\, Y^j$ for integers $1 \leq j \leq p-1$, and that $\mathbf{E}\, Z_n^p$ exists and is finite but fails to equal $\mathbf{E}\, Y^p$. Then for every $n \geq 0$ the same is true and, moreover,*

$$\mathbf{E}\, Z_n^p - \mathbf{E}\, Y^p = \left(\tfrac{2}{p+1}\right)^n (\mathbf{E}\, Z_0^p - \mathbf{E} Y^p).$$

Proof. If $\mathbf{E}|Z|^m < \infty$, then, with Z, $Z^* \stackrel{\mathcal{L}}{=} Z$, and U independent, by (1.4) and a trinomial expansion we have

$$\begin{aligned}
\mathbf{E}(SZ)^m &= \sum_{j+k \leq m} \frac{m!}{j!\,k!\,(m-j-k)!} \mathbf{E}\big(U^j Z^j (1-U)^k (Z^*)^k g(U)^{m-j-k}\big) \\
&= \sum_{j+k \leq m} \frac{m!}{j!\,k!\,(m-j-k)!} \mathbf{E}\big(U^j (1-U)^k g(U)^{m-j-k}\big)\, \mathbf{E} Z^j\, \mathbf{E} Z^k.
\end{aligned}$$

We apply this with $m = 1, \ldots, p$ for both $Z = Z_{n-1}$ and $Z = Y$, and note that by induction on n, all terms in the sum with $j \leq p-1$ and $k \leq p-1$ coincide for the two choices of Z. Hence, $\mathbf{E} Z_n^m = \mathbf{E} Y^m$ for $1 \leq m \leq p-1$ and

$$\mathbf{E} Z_n^p - \mathbf{E} Y^p = \big(\mathbf{E} U^p + \mathbf{E}(1-U)^p\big)\big(\mathbf{E} Z_{n-1}^p - \mathbf{E} Y^p\big) = \frac{2}{p+1}\big(\mathbf{E} Z_{n-1}^p - \mathbf{E} Y^p\big),$$

and the result follows. $\square$

Lemma 7.11. *Let $p \geq 1$ be an integer. Then, for any F and G,*

$$d_p(F, G) \geq \frac{|\mathbf{E}\, X^p - \mathbf{E}\, Y^p|}{\sum_{j=0}^{p-1} \|X\|_p^{p-1-j}\, \|Y\|_p^j}$$

with $X \sim F$ and $Y \sim G$ (and $0^0 := 1$).

Proof. Let (X, Y) be an optimal coupling of F and G. If $p = 1$, then

$$d_1(F, G) = \|X - Y\|_1 = \mathbf{E}\,|X - Y| \geq |\mathbf{E}\, X - \mathbf{E}\, Y|,$$

as desired. If $p \geq 2$, we employ the factorization

$$X^p - Y^p = (X - Y) \sum_{j=0}^{p-1} X^{p-1-j} Y^j,$$

whence

$$\begin{aligned}
|\mathbf{E}\,X^p - \mathbf{E}\,Y^p| &\le \mathbf{E}|X^p - Y^p| \\
&\le \|X-Y\|_p \left\| \sum_{j=0}^{p-1} X^{p-1-j} Y^j \right\|_{p/(p-1)} \\
&\le d_p(F,G) \sum_{j=0}^{p-1} \left\| X^{p-1-j} Y^j \right\|_{p/(p-1)}, \qquad (7.7)
\end{aligned}$$

where at the second inequality we have employed Hölder's inequality and at the third we have invoked the optimality of the coupling. Another application of Hölder's inequality, this time with conjugate exponents $\frac{p-1}{p-1-j}$ and $\frac{p-1}{j}$, yields

$$\|X^{p-1-j} Y^j\|_{p/(p-1)} \le \|X\|_p^{p-1-j}\, \|Y\|_p^j \qquad (7.8)$$

for $1 \le j \le p-2$, and (7.8) is trivially an equality when $j=0$ or $j=p-1$. Combining (7.7) and (7.8) and rearranging, we obtain the desired result. □

Proof of Theorem 7.9. By Lemma 7.2, we have the bound

$$\|Z_n\|_p \le d_p(Z_n,Y) + \|Y\|_p \le d_p(Z_0,Y) + \|Y\|_p \le \|Z_0\|_p + 2\|Y\|_p.$$

Thus Lemmas 7.11 and 7.10 yield the explicit bound

$$d_p(Z_n,Y) \ge \frac{|\mathbf{E}\,Z_0^p - \mathbf{E}\,Y^p|}{\sum_{j=0}^{p-1} (\|Z_0\|_p + 2\|Y\|_p)^{p-1-j}\, \|Y\|_p^j} \left(\frac{2}{p+1}\right)^n. \qquad \square$$

Proof of Theorem 7.7. The cases $p=1$ and $p=2$ follow immediately from Theorem 7.9; see also Remark 7.8.

When $p \ge 3$, fix $q > p$. By Lemmas 7.4 and 7.2 (the latter applied to d_q), for some C_q we have

$$d_p(F_n,F) \le C_q\, d_2^{\frac{2(q-p)}{p(q-2)}}(F_n,F)$$

and thus Theorem 7.9 implies Theorem 7.7 with $r = \left(\frac{2}{p+1}\right)^{\frac{p(q-2)}{2(q-p)}}$. By taking q sufficiently large, we obtain the result for any $r < \left(\frac{2}{p+1}\right)^{p/2}$. □

We have assumed in Theorems 7.1 and 7.7 that F_0 has finite moments of all orders. What happens if this fails? If $\mathbf{E}\,|Z_0|^p = \infty$ for some $p > 2$, then $d_p(Z_n,Y) = \infty$ for all n, but what can be said about $d_2(Z_n,Y)$? It seems reasonable to conjecture that we have at least as large d_2-distance in this case as in the nicer case with all moments finite, and that Theorems 7.1 and 7.7 hold for all $F_0 \ne F$. Unfortunately, we have not been able to prove this, but we offer the following partial result.

Theorem 7.12. *If $d_2(F_n,F) = O(r^n)$ for every $r > 0$, then $F_0 = F$. Consequently, if $F_0 \ne F$, there exists $r > 0$ such that $d_2(F_n,F) > r^n$ for infinitely many values of n.*

Proof. We may assume that $\mathbf{E}Z_0 = 0$ and $\mathbf{E}Z_0^2 = \mathbf{E}Y^2$ (in particular, $\mathbf{E}Z_0^2 < \infty$), because otherwise $d_2(F_n, F) = \Omega(r^n)$ with $r = 2/3$: see Remark 7.8. By induction then $\mathbf{E}Z_n = 0$ and $\mathbf{E}Z_n^2 = \mathbf{E}Y^2 < 1$ for every n: see Lemma 7.10.

As usual, let Z, $Z^* \stackrel{\mathcal{L}}{=} Z$, and U be independent. If $|Z| \geq 2x$, $|Z^*| \leq 2$, and $\frac{2}{3} \leq U \leq 1$, where $x \geq 5$, then

$$|SZ| = |UZ + (1-U)Z^* + g(U)| \geq \tfrac{2}{3}|Z| - \tfrac{1}{3}|Z^*| - 1 \geq \tfrac{4}{3}x - \tfrac{5}{3} \geq x.$$

Thus,

$$\mathbf{P}(|SZ| \geq x) \geq \mathbf{P}(|Z| \geq 2x) \cdot \mathbf{P}(|Z| \leq 2) \cdot \tfrac{1}{3}, \qquad x \geq 5.$$

If further $\mathbf{E}Z^2 \leq 1$, and thus by Chebyshev's inequality $\mathbf{P}(|Z| \leq 2) = 1 - \mathbf{P}(|Z| > 2) \geq 1 - \frac{1}{4} = \frac{3}{4}$, this yields

$$\mathbf{P}(|SZ| \geq x) \geq \tfrac{1}{4}\mathbf{P}(|Z| \geq 2x), \qquad x \geq 5.$$

Hence, by induction on n and our assumption on the first two moments of Z_0, for any $x \geq 5$,

$$\mathbf{P}(|Z_n| \geq x) \geq 4^{-n}\mathbf{P}(|Z_0| \geq 2^n x), \qquad n \geq 0,$$

and in particular

$$\mathbf{P}(|Z_n| \geq 2^n) \geq 4^{-n}\mathbf{P}(|Z_0| \geq 4^n), \qquad n \geq 3. \tag{7.9}$$

Now suppose that $d_2(Z_n, Y) = O(r^n)$. Using an optimal coupling between Z_n and Y, and the fact that Y has moments of all orders, we find

$$\begin{aligned}\mathbf{P}(|Z_n| \geq 2^n) &\leq \mathbf{P}(|Z_n - Y| \geq 2^{n-1}) + \mathbf{P}(|Y| \geq 2^{n-1}) \\ &\leq 2^{2-2n} d_2^2(Z_n, Y) + \mathbf{P}(|Y| \geq 2^{n-1}) = O(2^{-2n} r^{2n}).\end{aligned}$$

Combining this with (7.9), we obtain, for $n \geq 3$,

$$\mathbf{P}(|Z_0| \geq 4^n) \leq 4^n \mathbf{P}(|Z_n| \geq 2^n) = O(r^{2n}),$$

which implies that $\mathbf{E}|Z_0|^p < \infty$ for every $p > 0$ such that $4^p r^2 < 1$.

Consequently, if $d_2(Z_n, Y) = O(r^n)$ for every $r > 0$, then $\mathbf{E}|Z_0|^p < \infty$ for every $p > 0$, and Theorem 7.7 applies to yield $F_0 = F$. $\square$

Remark 7.13. Our proof of Theorem 7.12, combined with the proof of Theorem 7.7, shows that if $F_0 \neq F$ and p (assumed ≥ 3 here) is the smallest positive integer such that either $\mathbf{E}\, Z_0^p$ does not exist or $\mathbf{E}\, Z_0^p \neq \mathbf{E}\, Y^p$, then $d_2(Z_n, Y) > r^n$ for infinitely many values of n for any $0 < r < r_p$, with $r_p := 2^{-q}$, where q is the unique solution in (p, ∞) to $2^{\frac{2q(q-p)}{p(q-2)}} = \frac{p+1}{2}$.

8 Other lower bounds

In Section 7 we showed that convergence of the iterates F_n to F in the d_2-metric is not faster than geometric. In this final section we show likewise that the convergence is not faster than geometric in the other metrics we have considered in this paper. We again assume that $F_0 \neq F$ has finite moments of all orders. (Without this hypothesis, we can prove partial results by the method used in the proof of Theorem 7.12, but we do not know whether the full results hold.)

8.1 Kolmogorov–Smirnov and total variation distances

Recall the definitions of Kolmogorov–Smirnov distance and total variation distance given at (3.11) and (3.12), respectively. We begin with a simple lemma.

Lemma 8.1. *Let $p > 0$. For any $X \sim F$ and $Y \sim G$ each with finite* pth *absolute moment, if $K := d_{\mathrm{KS}}(F, G)$, then, for any $0 \le M < \infty$,*

$$|\mathbf{E}\,(X^p; X > 0) - \mathbf{E}\,(Y^p; Y > 0)| \le KM^p + \mathbf{E}\,(X^p; X > M) + \mathbf{E}\,(Y^p; Y > M)$$

and, if p is an integer,

$$|\mathbf{E}\,X^p - \mathbf{E}\,Y^p| \le 2KM^p + \mathbf{E}\,(|X|^p; |X| > M) + \mathbf{E}\,(|Y|^p; |Y| > M)\,.$$

Proof. Define $X_M := \min(X^+, M)$, where $X^+ = \max(X, 0)$, and similarly Y_M. Then

$$0 \le \mathbf{E}(X^p; X > 0) - \mathbf{E}X_M^p \le \mathbf{E}(X^p; X > M)$$

and similarly for Y, while

$$\begin{aligned} \left|\mathbf{E}\,X_M^p - \mathbf{E}\,Y_M^p\right| &= \left|\int_0^M px^{p-1}\mathbf{P}(X > x)\,dx - \int_0^M px^{p-1}\mathbf{P}(Y > x)\,dx\right| \\ &\le \int_0^M px^{p-1}\left|\mathbf{P}(X > x) - \mathbf{P}(Y > x)\right| dx \le KM^p. \end{aligned}$$

Together, these yield the first inequality.

The second follows by applying the first to (X, Y) and to $(-X, -Y)$ and summing or subtracting, depending on the parity of p. □

Theorem 8.2. *Suppose $F_0 \ne F$ has finite moments of all orders, and let p be defined as in Theorem 7.7. Then, for any $0 < r < 2/(p+1)$,*

$$d_{\mathrm{TV}}(F_n, F) \ge d_{\mathrm{KS}}(F_n, F) = \Omega(r^n).$$

(The implicit multiplicative constant depends on both F_0 and the choice of r.)

Proof. Let $K_n := d_{\mathrm{KS}}(F_n, F) > 0$. If we apply Lemma 8.1 and then use Lemma 7.10, we find, for any $q \ge p$,

$$\begin{aligned} |\mathbf{E}\,Z_0^p - \mathbf{E}\,Y^p|\left(\tfrac{2}{p+1}\right)^n &\le 2K_nM^p + \mathbf{E}\,(|Z_n|^p; |Z_n| > M) + \mathbf{E}\,(|Y|^p; |Y| > M) \\ &\le 2K_nM^p + M^{-(q-p)}\mathbf{E}|Z_n|^q + M^{-(q-p)}\mathbf{E}|Y|^q, \end{aligned} \tag{8.1}$$

for any $0 \le M < \infty$. It follows from Lemma 7.2 that $\mathbf{E}|Z_n|^q \le C_q$, for some C_q not depending on n. Choosing $M = K_n^{-1/q}$ thus gives, with $c := |\mathbf{E}\,Z_0^p - \mathbf{E}\,Y^p| > 0$,

$$c\left(\tfrac{2}{p+1}\right)^n \le 2K_n^{1-(p/q)} + 2C_qK_n^{1-(p/q)},$$

and thus $K_n = \Omega(r^n)$ with $r = \left(\frac{2}{p+1}\right)^{q/(q-p)}$. The result follows, since $r \to 2/(p+1)$ as $q \to \infty$. □

8.2 Density functions and characteristic functions

We immediately obtain results for the density functions f_n, which by Theorem 3.1 exist at least for $n \geq 3$ and by Theorem 3.2 converge uniformly, at a geometric rate, to the density f of Y.

Corollary 8.3. *Suppose $F_0 \neq F$ has finite moments of all orders, and let p be defined as in Theorem 7.7. Then, for any $0 < r < 2/(p+1)$,*

$$\int_{-\infty}^{\infty} |f_n(x) - f(x)| \, dx = \Omega(r^n) \tag{8.2}$$

and

$$\sup_x |f_n(x) - f(x)| = \Omega(r^n). \tag{8.3}$$

Proof. The estimate (8.2) follows from Theorem 8.2, because (whenever f_n exists) $\int_{-\infty}^{\infty} |f_n(x) - f(x)| \, dx = 2d_{\mathrm{TV}}(F_n, F)$.

The estimate (8.3) follows from Theorem 8.2 using inequality (3.14) and the discussion following it. □

Similarly, we have a geometric lower bound for the $L^1(\mathbf{R})$ and $L^\infty(\mathbf{R})$ distances of the characteristic functions.

Corollary 8.4. *Suppose $F_0 \neq F$ has finite moments of all orders, and let p be defined as in Theorem 7.7. Then, for any $0 < r < 2/(p+1)$,*

$$\int_{-\infty}^{\infty} |\phi_{Z_n}(t) - \phi_Y(t)| \, dt = \Omega(r^n). \tag{8.4}$$

and

$$\sup_t |\phi_{Z_n}(t) - \phi_Y(t)| = \Omega(r^n). \tag{8.5}$$

Proof. The estimate (8.4) is immediate from Corollary 8.3 and inequality (3.8). Next, (8.4) for some $r = r_0$ and Theorem 2.1 imply (8.5) for any $r < r_0$ by the argument used to show (3.10) in the proof of Theorem 3.2. □

It is not too hard to extend Corollaries 8.3 and 8.4 to any $L^q(\mathbf{R})$ distance, $1 \leq q \leq \infty$.

8.3 Moment generating functions

Finally, we consider lower bounds for the convergence of moment generating functions. We assume for simplicity that Z_0 has an everywhere finite moment generating function, and know by Theorem 5.1 that then ψ_{Z_n} converges to ψ_Y pointwise, and uniformly on compact sets, with geometric rate. For lower bounds we first note that if $F_0 \neq F$ and p is as in Theorem 7.7, then the derivatives of ψ_{Z_n} and ψ_Y at the origin, which equal the corresponding moments of Z_n and Y, by Lemma 7.10 coincide up to order $p-1$, while the pth derivatives differ by $c\left(\frac{2}{p+1}\right)^n$ with $c = \mathbf{E}Z_0^p - \mathbf{E}Y^p \neq 0$. For λ close to the origin, this and a Taylor expansion shows that $|\psi_{Z_n}(\lambda) - \psi_Y(\lambda)| = \Omega\left(|\lambda|^p \left(\frac{2}{p+1}\right)^n\right)$, but the range of λ where we can prove that this is valid depends on n.

Indeed, there is no general lower bound for $|\psi_{Z_n}(\lambda) - \psi_Y(\lambda)|$ for a fixed λ, since there may be points λ where $\psi_{Z_n}(\lambda)$ and $\psi_Y(\lambda)$ coincide "accidentally". For example, suppose that Z_0 is bounded with $\mathbf{E}\, Z_0 = 0$ and $\mathbf{Var}\, Z_0 > \mathbf{Var}\, Y$. By induction, the same holds for each Z_n: see Lemma 7.10 and note that $g(U)$ is bounded. Consequently, for each n, Taylor's formula shows that $\psi_{Z_n}(\lambda) > \psi_Y(\lambda)$ for small positive λ, while $\psi_{Z_n}(\lambda) = \exp\bigl(O(\lambda)\bigr)$ and thus Theorem 4.4 shows that $\psi_{Z_n}(\lambda) < \psi_Y(\lambda)$ for large λ. Hence there exists for every n at least one positive $\lambda = \lambda_n$ such that $\psi_{Z_n}(\lambda) = \psi_Y(\lambda)$. Nevertheless, such points have to be isolated, and if we consider the maximum deviation over an interval, we have a geometric lower bound.

Theorem 8.5. *Suppose $F_0 \neq F$ has everywhere finite moment generating function, and let (a, b) be a nonempty interval. Then there exists $r > 0$ such that $\sup_{a \le \lambda \le b} |\psi_{Z_n}(\lambda) - \psi_Y(\lambda)| = \Omega(r^n)$.*

Proof. We use the fact that the moment generating functions ψ_{Z_n} and ψ_Y are entire analytic functions in the complex plane $\mathbf{C}$.

Let $R := |a| + |b| + 1$. There exists a (unique) function ω which is continuous on $D_R := \{z \in \mathbf{C} : |z| \le R\}$ and analytic in $\Omega := \{z : |z| < R\} \setminus [a, b]$ such that $\omega(z) = 0$ for $|z| = R$ and $\omega(z) = 1$ for $z \in [a, b]$; this function is called harmonic measure and is probabilistically given by the probability that a Brownian motion starting at z hits $[a, b]$ before it hits $\{z : |z| = R\}$.

Let $f_n(z) := \psi_{Z_n}(z) - \psi_Y(z)$ and $u_n(z) := \ln |f_n(z)| \ge -\infty$. For $z \in D_R$,

$$|f_n(z)| \le |\psi_{Z_n}(z)| + |\psi_Y(z)| \le \psi_{Z_n}(R) + \psi_Y(R) + \psi_{Z_n}(-R) + \psi_Y(-R),$$

which by Theorem 5.1 is bounded by some constant $A < \infty$ (depending on Z_0 but not on n). Let further $\delta_n := \max_{a \le \lambda \le b} |f_n(\lambda)|$; we may of course restrict attention to those values of n satisfying $\delta_n < 1$. Now $u_n(z) \le \ln A$ for $|z| = R$ and $u_n(z) \le \ln \delta_n$ for $z \in [a, b]$; thus (since $A \ge 1$)

$$u_n(z) \le \ln A + (\ln \delta_n)\omega(z) \tag{8.6}$$

for every $z \in \partial\Omega$. Since u_n is subharmonic and the right hand side is harmonic in Ω and continuous on its closure, (8.6) holds for every $z \in \overline{\Omega} = D_R$, cf. [16, Theorems 17.3 and 17.4]. In particular, setting $\varepsilon := \inf_{|z| \le 1} \omega(z) > 0$, we have

$$u_n(z) \le \ln A + \varepsilon \ln \delta_n, \qquad |z| \le 1,$$

or

$$|f_n(z)| \le A\delta_n^{\varepsilon}, \qquad |z| \le 1. \tag{8.7}$$

Let p be as in Theorem 7.7. By (8.7) and Cauchy's estimates [16, Theorem 10.26],

$$|f_n^{(p)}(0)| \le p!\, A\delta_n^{\varepsilon}.$$

Since by Lemma 7.10

$$|f_n^{(p)}(0)| = |\mathbf{E} Z_n^p - \mathbf{E} Y^p| = \Omega\bigl((\tfrac{2}{p+1})^n\bigr),$$

it follows that $\delta_n = \Omega(r^n)$ with $r = \bigl(\frac{2}{p+1}\bigr)^{1/\varepsilon}$. $\square$

References

[1] Cambanis, S., Simons, G., and Stout, W. Inequalities for $\mathbf{E}k(X, Y)$ when the marginals are fixed. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **36** (1976), 285–294.

[2] van der Corput, J. G. Zahlentheoretische Abschätzungen. *Math. Ann.* **84** (1921), 53–79.

[3] Dongarra, J. and Sullivan, F. Guest editors' introduction: the top 10 algorithms. *Computing in Science & Engineering* **2** (2000).

[4] Eddy, W. F. and Schervish, M. J. How many comparisons does Quicksort use? *J. Algorithms* **19** (1995), 402–431.

[5] Feller, W. *An Introduction to Probability Theory and its Applications. Vol. II.* Second edition. Wiley, New York, 1971.

[6] Fill, J. A. and Janson, S. Smoothness and decay properties of the limiting Quicksort density function. In D. Gardy and A. Mokkadem, editors, *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics, pages 53–64. Birkhäuser Verlag, 2000. Refereed article, available from `http://www.mts.jhu.edu/~fill/` or `http://www.math.uu.se/~svante/`.

[7] Fill, J. A. and Janson, S. A characterization of the set of fixed points of the Quicksort transformation. *Electron. Comm. Probab.* **5** (2000), 77–84 (electronic).

[8] Fill, J. A. and Janson, S. Quicksort asymptotics. In preparation.

[9] Hoare, C. A. R. Quicksort. *Comput. J.* **5** (1962), 10–15.

[10] JaJa, J. A perspective on quicksort. *Computing in Science & Engineering* **2** (2000).

[11] Knessl, C. and Szpankowski, W. Quicksort algorithm again revisited. *Discrete Math. Theor. Comput. Sci.* **3** (1999), 43–64.

[12] Knuth, D. E. *The Art of Computer Programming. Volume 3. Sorting and searching.* Second edition. Addison–Wesley, Reading, Mass., 1998.

[13] Montgomery, H. L. *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis.* CBMS Reg. Conf. Ser. Math. 84, AMS, Providence, R.I., 1994.

[14] Régnier, M. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23** (1989), 335–343.

[15] Rösler, U. A limit theorem for 'Quicksort'. *RAIRO Inform. Théor. Appl.* **25** (1991), 85–100.

[16] Rudin, W. *Real and Complex Analysis.* Second edition. McGraw–Hill, New York, 1974.

[17] Schwartz, L. *Théorie des Distributions.* Second edition. Hermann, Paris, 1966.

[18] Tan, K. H. and Hadjicostas, P. Some properties of a limiting distribution in Quicksort. *Statist. Probab. Lett.* **25** (1995), 87–94.