A statistical view on exchanges in Quickselect

Benjamin Dadoun
Département Informatique
École Normale Supérieure de Cachan
94235 Cachan Cedex
France
Email: benjamin.dadoun@ens-cachan.fr

Ralph Neininger
Institute for Mathematics
J.W. Goethe University
60054 Frankfurt a.M.
Germany
Email: neiningr@math.uni-frankfurt.de

Abstract

In this paper we study the number of key exchanges required by Hoare's FIND algorithm (also called Quickselect) when operating on a uniformly distributed random permutation and selecting an independent uniformly distributed rank. After normalization we give a limit theorem where the limit law is a perpetuity characterized by a recursive distributional equation. To make the limit theorem usable for statistical methods and statistical experiments we provide an explicit rate of convergence in the Kolmogorov–Smirnov metric, a numerical table of the limit law's distribution function and an algorithm for exact simulation from the limit distribution. We also investigate the limit law's density. This case study provides a program applicable to other cost measures, alternative models for the rank selected and more balanced choices of the pivot element such as median-of-$2t+1$ versions of Quickselect as well as further variations of the algorithm.

MSC2010: 60F05, 68P10, 60C05, 68Q25.
Keywords: Quickselect, FIND, key exchanges, limit law, perpetuity, perfect simulation, rate of convergence, coupling from the past, contraction method.

1 Introduction

For selecting ranks within a finite list of data from an ordered set, Hoare [10] introduced the algorithm FIND, also called Quickselect, which is a one sided version of his sorting algorithm Quicksort. The data set is partitioned into two sub-lists by use of a pivot element, then the algorithm is recursively applied to the sub-list that contains the rank to be selected, unless its size is one. Hoare's partitioning procedure is performed by scanning the list with pointers from left and right until misplaced elements are found. They are flipped, what we count as one key exchange. This scanning step is then further performed until the pointers meet within the list. For definiteness, in this paper we consider the version of Hoare's partitioning procedure presented in Cormen, Leiserson and Rivest [3, Section 8.1]. (However, our asymptotic results are robust to small changes in the partitioning procedure, e.g. they also hold for the versions of Hoare's partitioning procedure described in Sedgewick [24, p. 118] or Mahmoud [14, Exercise 7.2].)

We consider the probabilistic model where n distinct data are given in uniformly random order and where the rank to be selected is uniformly distributed on $\{1, \dots, n\}$ and independent of the permutation of the data. In this model the number of key comparisons has been studied in detail in Mahmoud, Moddares and Smythe [17]. For the number Y_n of key exchanges the mean has been identified exactly by means of analytic combinatorics: In Mahmoud [15], for the number of data moves M_n which is essentially (the partitioning procedure used in [15] being slightly different to ours) twice our number of key exchanges it is shown that

$$\mathbb{E}[M_n] = n + \frac{2}{3}H_n - \frac{17}{9} + \frac{2H_n}{3n} - \frac{2}{9n}. \tag{1.1}$$

Note that lower order terms here depend on the particular version of Hoare's partitioning procedure used. Moreover, for the variance, Mahmoud [15] obtained, as $n \to \infty$ that

$$\frac{1}{15}n^2 + \mathrm{O}(n) \leqslant \mathrm{Var}(M_n) \leqslant \frac{41}{15}n^2 + \mathrm{O}(n), \tag{1.2}$$

where the Bachmann–Landau O-notation is used. A different partitioning procedure due to Lomuto is analyzed in Mahmoud [16]. Key exchanges in related but different models are studied in [11, 18]. In the present paper we extend the analysis started in [15] of Quickselect with Hoare's partition procedure. Together with more refined results stated below we identify the asymptotic order of the variance and provide a limit law:

This research was done during an internship of the first mentioned author at J.W. Goethe University from June 2013 to August 2013.

THEOREM 1.1. *For the number Y_n of key exchanges used by Hoare's Quickselect algorithm when acting on a uniformly random permutation of size n and selecting an independent uniform rank we have, as $n \to \infty$, that*

$$\frac{Y_n}{n} \xrightarrow{d} X, \tag{1.3}$$

where the distribution of X is the unique solution of the recursive distributional equation

$$X \stackrel{d}{=} \sqrt{U}X + \sqrt{U}(1-\sqrt{U}), \tag{1.4}$$

where X and U are independent and U is uniformly distributed on $[0,1]$.
Moreover, we have $\mathrm{Var}(Y_n) \sim \frac{1}{60}n^2$ as $n \to \infty$.

Theorem 1.1 follows quite directly from the contraction method and is a corollary to more refined convergence results in our Theorems 3.1 and 3.2. We also obtain $\mathrm{Var}(M_n) \sim \frac{1}{15}n^2$ and $M_n/n \to 2X$ in distribution as $n \to \infty$, cf. (1.2). An interpretation of the coefficients $\sqrt{U}$ and $\sqrt{U}(1-\sqrt{U})$ appearing in (1.4) is given in Remark 2.1 below.

Recursive distributional equations such as (1.4) appear frequently in the asymptotic analysis of random tree models and of complexities of recursive algorithms; they also appear in insurance mathematics as so-called perpetuities and in probabilistic number theory. It should be noted that solutions of recursive distributional equations are typically difficult to access, e.g., with respect to their density if a density exists.

Recall that the original purpose of a limit law, such as our limit law (1.3), consists of being able to approximate the distributions of Y_n by their scaled limit X. However, such an approximation can only be made effective if characteristics of the distribution $\mathcal{L}(X)$ of X are accessible and the distance between $\mathcal{L}(X)$ and $\mathcal{L}(Y_n/n)$ can be bounded explicitly.

For this reason we take a statistician's point of view: To make the limit theorem (1.3) usable for statistical methods and statistical experiments we provide an explicit bound on the rate of convergence in the Kolmogorov–Smirnov metric in Section 3 (Theorem 3.2), a numerical table of the distribution function of $\mathcal{L}(X)$ in Section 4 (Figure 1) and an algorithm for exact simulation from $\mathcal{L}(X)$ in Section 5 (Algorithm 1). The density and further properties of $\mathcal{L}(X)$ are studied in Section 4. In Section 2 the recursive approach our analysis is based on is introduced together with some combinatorial preliminaries.

We consider this paper as a case study with a program applicable to other cost measures, alternative models for the rank selected and more balanced choices of the pivot element such as median-of-$2t+1$ versions of Quickselect as well as further variations of the algorithm.

2 Distributional recurrence and preliminaries

The first call to the partitioning procedure (in the version [3, Section 8.1] we consider here) splits the given uniformly distributed list $A[1..n]$ of size n into two sub-lists of sizes I_n and $n - I_n$ as follows: The first element $p := A[1]$ is chosen as the pivot element, and the list is scanned both forwards and backwards with two indices i and j, looking for elements with $A[i] \geqslant p$ and elements with $A[j] \leqslant p$. Every misplaced pair $(A[i], A[j])$ found is then flipped, unless i has become greater than or equal to j, where we stop (resulting in $I_n = j$). Note that the pivot element is moved to the right sub-list if there is at least one key exchange, and the event $\{I_n = 1\}$ thus occurs if and only if the leftmost element in the array is the smallest or the second smallest element of the whole array. Together with the uniformity assumption we obtain

$$\mathbb{P}(I_n = 1) = \frac{2}{n}, \quad \mathbb{P}(I_n = j) = \frac{1}{n} \text{ for } j = 2, \dots, n-1.$$

The list with elements with value less or equal to the pivot element we call the *left sub-list*, its size is I_n, the other list we call the *right sub-list*. We denote by T_n the number of key exchanges executed during the first call to the partitioning procedure. Note that T_n is random and that I_n and T_n are stochastically dependent (for all n sufficiently large). Since during the first partitioning step comparisons are only done between the elements and the pivot element we have that conditional on the size I_n the left and right sub-list are uniformly distributed and independent of each other. The number Y_n of key exchanges (key swaps) required by Quickselect (when operating on a uniformly permuted list of size n of distinct elements and selecting a rank R_n uniformly distributed over $\{1, \dots, n\}$ and independent of the list) allows a recursive decomposition. The recursive structure of the algorithm, the properties of the partitioning procedure and the model for the rank to be selected imply $Y_1 = 0$ and, for $n \geqslant 2$, the distributional recurrence

$$Y_n \stackrel{d}{=} \mathbb{1}_{\{R_n \leqslant I_n\}} Y_{I_n} + \mathbb{1}_{\{R_n > I_n\}} Y'_{n-I_n} + T_n. \tag{2.5}$$

Here $(Y'_j)_{1 \leqslant j \leqslant n-1}$ is identically distributed as $(Y_j)_{1 \leqslant j \leqslant n-1}$ and we have that $(Y_j)_{1 \leqslant j \leqslant n-1}$, $(Y'_j)_{1 \leqslant j \leqslant n-1}$ and (I_n, T_n) are independent. To make the right hand side of the latter display more explicit we observe that the conditional distribution of T_n given I_n is hypergeometric:

LEMMA 2.1. *Conditional on $I_n = 1$ the number T_n of swaps during the first call to the partitioning procedure* *has the Bernoulli Ber*$(\frac{1}{2})$ *distribution. Conditional on* $I_n = j$ *for* $j \in \{2, \ldots, n-1\}$ *the random variable* T_n *is hypergeometrically Hyp*$(n-1; j, n-j)$ *distributed, i.e.,*

$$\mathbb{P}(T_n = k \mid I_n = j) = \frac{\binom{j}{k}\binom{n-j-1}{n-j-k}}{\binom{n-1}{n-j}},$$

for $\min(1, j-1) \leqslant k \leqslant \min(j, n-j)$.

Proof. For simplicity of presentation we identify the elements of the array with their ranks, i.e., we assume the elements are $1, \ldots, n$ in uniformly random order. Conditional on $I_n = 1$ the leftmost element of the array is 1 or 2 resulting in respectively 0 or 1 key exchanges. The uniformity of the array implies the Ber$(\frac{1}{2})$ distribution in the statement of the Lemma. Conditional on $I_n = j$ with $j \in \{2, \ldots, n-1\}$ the pivot element p is moved to the right sub-list and we have $I_n = p - 1$. Thus, we have to count the number of permutations σ of length n such that $\sigma(i) \leqslant I_n$ for exactly k indices $i \in \{I_n+1, \ldots, n\}$, among those having $p = I_n + 1$ as first element. This implies the assertion.

The asymptotic joint behavior of (I_n, T_n) will be crucial in our subsequent analysis:

LEMMA 2.2. *For any* $1 \leqslant p < \infty$ *we have*

$$\left(\frac{I_n}{n}, \frac{T_n}{n}\right) \xrightarrow{\ell_p} (U, U(1-U)) \qquad (n \to \infty),$$

where U *has the uniform distribution on the unit interval* $[0,1]$.

The convergence in ℓ_p (defined below) is equivalent to weak convergence plus convergence of the p-th absolute moments. Lemma 2.2 follows below from Lemma 3.2. The scalings in Lemma 2.2 motivate the normalization

$$X_n := \frac{Y_n}{n}, \quad n \geqslant 1. \tag{2.6}$$

Recurrence (2.5) implies the distributional recurrence

$$\begin{aligned} X_n \stackrel{d}{=} {} & \mathbb{1}_{\{\frac{R_n}{n} \leqslant \frac{I_n}{n}\}} \frac{I_n}{n} X_{I_n} \\ & + \mathbb{1}_{\{\frac{R_n}{n} > \frac{I_n}{n}\}} \frac{n-I_n}{n} X'_{n-I_n} + \frac{T_n}{n}, \end{aligned} \tag{2.7}$$

(for $n \geqslant 2$) where, similarly to (2.5), $(X'_j)_{1 \leqslant j \leqslant n-1}$ is identically distributed as $(X_j)_{1 \leqslant j \leqslant n-1}$ and we have that $(X_j)_{1 \leqslant j \leqslant n-1}$, $(X'_j)_{1 \leqslant j \leqslant n-1}$ and (I_n, T_n) are independent.

The asymptotics of Lemma 2.2 suggest that a limit X of X_n satisfies the recursive distributional equation (RDE)

$$\begin{aligned} X \stackrel{d}{=} {} & \mathbb{1}_{\{V \leqslant U\}} U X \\ & + \mathbb{1}_{\{V > U\}} (1-U) X' + U(1-U), \end{aligned} \tag{2.8}$$

where U, V, X, X' are independent, U and V are uniformly distributed on $[0,1]$ and X' has the same distribution as X.

LEMMA 2.3. *RDE* (2.8) *has a unique solution among all probability distributions on the real line. This solution is also the unique solution (among all probability distributions on the real line) of RDE* (1.4).

Proof. A criterion of Vervaat [26] states that a RDE of the form $X =_d AX + b$ with X and (A, b) independent has a unique solution among all probability distributions on the real line if $-\infty \leqslant \mathbb{E}[\log|A|] < 0$ and $\mathbb{E}[\log^+|b|] < \infty$. These two conditions are satisfied for our RDE (1.4). The full claim of the Lemma hence follows by showing that the solutions of RDE (2.8) are exactly the solutions of RDE (1.4). This can be seen using characteristic functions as follows: Let $\mathcal{L}(Z)$ be a solution of RDE (2.8) and denote its characteristic function by $\varphi_Z(t) := \mathbb{E}[e^{itZ}]$ for $t \in \mathbb{R}$. Conditioning on U and V and using independence we obtain that

$$\varphi_Z(t) = \int_0^1 2u\varphi_Z(tu)e^{itu(1-u)}\mathrm{d}u, \quad t \in \mathbb{R}.$$

Now, for the random variable $Y := \sqrt{U}Z + \sqrt{U}(1 - \sqrt{U})$, where U is uniformly distributed on $[0,1]$ and independent of Z we find that its characteristic function φ_Y satisfies

$$\begin{aligned} \varphi_Y(t) &= \int_0^1 \varphi_Z(t\sqrt{u})e^{it\sqrt{u}(1-\sqrt{u})}\mathrm{d}u \\ &= \int_0^1 2u\varphi_Z(tu)e^{itu(1-u)}\mathrm{d}u = \varphi_Z(t), \quad t \in \mathbb{R}. \end{aligned}$$

This implies that $\mathcal{L}(Z)$ is a solution of RDE (1.4). The same argument shows that every solution of RDE (1.4) is a solution of RDE (2.8).

REMARK 2.1. *Alternatively to recurrence* (2.5) *we have the recurrence*

$$Y_n \stackrel{d}{=} Y_{J_n} + T_n, \quad n \geqslant 2, \tag{2.9}$$

with conditions as in (2.5) *and* J_n *denoting the size of the sub-list where the Quickselect algorithm recurses on. Note that by the uniformity of the rank to be selected* J_n *is a size-biased version of* I_n. *Hence the limit (in distribution) of* J_n/n *is the size-biased version of the limit* U *of* I_n/n. *Since* $\sqrt{U}$ *is a size-biased version of* U, *it appears in the RDE* (1.4). *Moreover, the asymptotic joint behavior of* (J_n, T_n) *is again determined by the concentration of the hypergeometric distribution as in Lemma 2.2 (cf. the proof of Lemma 3.2.) Analogously, we obtain* $(J_n/n, T_n/n) \to (\sqrt{U}, \sqrt{U}(1-\sqrt{U}))$ *which* *explains the occurrences of the additive term* $\sqrt{U}(1-\sqrt{U})$ *in RDE* (1.4). *(Note that this does not contradict Lemma 2.2, since* $U(1-U)$ *and* $\sqrt{U}(1-\sqrt{U})$ *are identically distributed.) We could as well base our subsequent analysis on* (2.9) *but prefer to work with recurrence* (2.5).

3 Convergence and rates

In this section we bound the rate of convergence in the limit law of Theorem 1.1. First, bounds in the minimal ℓ_p-metrics are derived. These imply bounds on the rate of convergence within the Kolmogorov–Smirnov metric. For $1 \leqslant p < \infty$ and probability distibutions $\mathcal{L}(W)$ and $\mathcal{L}(Z)$ with $\mathbb{E}[|W|^p]$, $\mathbb{E}[|Z|^p] < \infty$ the ℓ_p-distance is defined by

$$\ell_p(\mathcal{L}(W),\mathcal{L}(Z)) := \ell_p(W,Z) := \inf\{\|W'-Z'\|_p \,|\, W' \stackrel{d}{=} W, Z' \stackrel{d}{=} Z\}.$$

The infimum is over all vectors (W',Z') on a common probability space with the marginals of W and Z. The infimum is a minimum and such a minimizing pair (W',Z') is called an optimal coupling of $\mathcal{L}(W)$ and $\mathcal{L}(Z)$. For a sequence of random variables $(W_n)_{n\geqslant 1}$ and W we have, as $n\to\infty$, that

$$\ell_p(W_n,W)\to 0 \quad\Longleftrightarrow\quad \begin{cases} W_n \xrightarrow{d} W, \\ \mathbb{E}[|W_n|^p]\to\mathbb{E}[|W|^p]. \end{cases}$$

For these and further properties of ℓ_p see Bickel and Freedman [1, Section 8].

We start bounding the rate in the convergence in Lemma 2.2. This can be done using a tail estimate for the hypergeometric distribution derived in Serfling [25, Theorem 3.1], restated here in a slightly weaker form more convenient for our analysis:

Lemma 3.1. *Let* $n\geqslant 2$, $j\in\{1,\dots,n-1\}$ *and* $T_n^{(j)}$ *be a random variable with hypergeometric distribution* Hyp$(n-1;j,n-j)$. *Then for all* $p>0$ *we have*

$$\mathbb{E}\left[\left|\frac{T_n^{(j)}}{n}-\frac{j(n-j)}{n(n-1)}\right|^p\right] \leqslant \frac{\Gamma(p/2+1)}{2^{p/2+1}}\, n^{-p/2},$$

where Γ *denotes Euler's gamma function.*

Lemma 3.2. *For the number* T_n *of key exchanges in the first call to the partitioning procedure of Hoare's Quickselect we have for all* $n\geqslant 2$ *and all* $1\leqslant p<\infty$ *that*

$$\ell_p\left(\frac{T_n}{n},U(1-U)\right)\leqslant(2+\tau_p)n^{-1/2},$$

$$\text{where } \tau_p := \left(\frac{1}{2}+\frac{\Gamma(p/2+1)}{2^{p/2+1}}\right)^{1/p}.$$

Proof. Let U be uniformly distributed over $[0,1]$ and the underlying probability space sufficiently large so that we can also embed the vector (I_n,T_n) such that $I_n=\lfloor nU\rfloor+\mathbb{1}_{\{U\leqslant 1/n\}}$. Let $h(u):=u(1-u)$. The mean value theorem and $|\frac{\mathrm{d}}{\mathrm{d}u}h(u)|=|1-2u|\leqslant 1$ for all $u\in[0,1]$ imply

$$\begin{aligned}\left\|U(1-U)-\frac{I_n(n-I_n)}{n^2}\right\|_p^p &= \sum_{k=0}^{n-1}\int_{\frac{k}{n}}^{\frac{k+1}{n}} \left|h(u)-h\left(\tfrac{k\vee 1}{n}\right)\right|^p \mathrm{d}u \\ &\leqslant \frac{1}{n^p}.\end{aligned}$$

We have $(\frac{1}{n-1}-\frac{1}{n})\|I_n(n-I_n)/n\|_p\leqslant\frac{1}{n}$ since $1\leqslant I_n\leqslant n-1$ a.s. Using Lemma 3.1 we obtain

$$\begin{aligned}\left\|\frac{T_n}{n}-\frac{I_n(n-I_n)}{n(n-1)}\right\|_p^p &= \sum_{j=1}^{n-1}\mathbb{E}\left[\left|\frac{T_n}{n}-\frac{j(n-j)}{n(n-1)}\right|^p \,\middle|\, I_n=j\right]\mathbb{P}(I_n=j) \\ &\leqslant \frac{1}{2n^p}+\frac{n-2}{n}\times\frac{\Gamma(p/2+1)}{2^{p/2+1}}\,n^{-p/2} \\ &\leqslant \tau_p^p n^{-p/2}.\end{aligned}$$

The triangle inequality implies

$$\ell_p\left(\frac{T_n}{n},U(1-U)\right)\leqslant\left\|\frac{T_n}{n}-U(1-U)\right\|_p \leqslant (2+\tau_p)n^{-1/2}, \tag{3.10}$$

the assertion.

We obtain the following bounds on the rate of convergence in Theorem 1.1. For the proof of Theorem 3.1 standard estimates from the contraction method, see [22, 21, 23, 20], are applied.

Theorem 3.1. *For* Y_n *and* X *as in Theorem 1.1 we have for all* $n\geqslant 1$ *and all* $1\leqslant p<\infty$, *that*

$$\ell_p\left(\frac{Y_n}{n},X\right)\leqslant\kappa_p n^{-1/2},\qquad \kappa_p:=\frac{2p+3}{2p-1}(7+\tau_p).$$

Proof. With X_n as defined in (2.6) recall the recurrence (2.7):

$$X_n \stackrel{d}{=} \mathbb{1}_{\{\frac{R_n}{n}\leqslant\frac{I_n}{n}\}}\frac{I_n}{n}X_{I_n} + \mathbb{1}_{\{\frac{R_n}{n}>\frac{I_n}{n}\}}\frac{n-I_n}{n}X'_{n-I_n}+\frac{T_n}{n}, \tag{3.11}$$

For X as in Theorem 1.1 we have, by Lemma 2.3, that

$$X \stackrel{d}{=} \mathbb{1}_{\{V\leqslant U\}}UX + \mathbb{1}_{\{V>U\}}(1-U)X'+U(1-U), \tag{3.12}$$

with conditions as in (2.8). Note that we can embed all random variables appearing on the right hand sides of (3.11) and (3.12) on a common probability space such that we additionally have that $I_n = \lfloor nU \rfloor + \mathbb{1}_{\{U \leqslant 1/n\}}$, $R_n = \lceil nV \rceil$ and that (X_j, X) and (X_j', X') are optimal couplings of $\mathcal{L}(X_j)$ and $\mathcal{L}(X)$ such that (U, V), (X_j, X), (X_j', X') for $j = 1, \ldots, n-1$ are independent.

Now, for $n \geqslant 2$ we define the random variable

$$Q_n := \mathbb{1}_{\{R_n \leqslant I_n\}} \frac{I_n}{n} X + \mathbb{1}_{\{R_n > I_n\}} \frac{n - I_n}{n} X' + \frac{T_n}{n}.$$

The triangle inequality implies

$$\ell_p(X_n, X) \leqslant \ell_p(X_n, Q_n) + \ell_p(Q_n, X). \tag{3.13}$$

The second summand in (3.13) is bounded by

$$\begin{aligned}
&\ell_p(Q_n, X) \\
&\quad\leqslant \left\| \mathbb{1}_{\{V \leqslant U\}} U - \mathbb{1}_{\{R_n \leqslant I_n\}} \frac{I_n}{n} \right\|_p \\
&\qquad + \left\| \mathbb{1}_{\{V > U\}} (1-U) - \mathbb{1}_{\{R_n > I_n\}} \frac{n - I_n}{n} \right\|_p \\
&\qquad + \left\| \frac{T_n}{n} - U(1-U) \right\|_p \\
&\quad\leqslant \frac{2}{n} + \frac{2}{n} + \frac{2 + \tau_p}{\sqrt{n}},
\end{aligned}$$

where we plug in the right hand sides of (3.11) and (3.12), use independence, that $\|X\|_p \leqslant 1$ and the bound in (3.10). For the first summand in (3.13) conditioning on R_n and I_n and using that (X_j, X) and (X_j', X') are optimal couplings of $\mathcal{L}(X_j)$ and $\mathcal{L}(X)$ we have

$$\ell_p(X_n, Q_n) \leqslant \frac{1}{n} \sum_{i=1}^{n-1} \frac{i^p(2i-1) + \mathbb{1}_{\{i=1\}}}{n^{p+1}} \ell_p(X_i, X).$$

The summand $\mathbb{1}_{\{i=1\}} \ell_p(X_1, X)$ is bounded by 1 since $X_1 = 0$ and $\|X\|_p \leqslant 1$. Putting the estimates together we obtain

$$\ell_p(X_n, X) \leqslant \frac{1}{n} \sum_{i=1}^{n-1} \frac{i^p(2i-1)}{n^{p+1}} \ell_p(X_i, X) + \frac{7 + \tau_p}{\sqrt{n}}.$$

Now, by induction, we show $\ell_p(X_n, X) \leqslant \kappa_p n^{-1/2}$. Since $\kappa_p \geqslant 1$ the assertion is true for $n = 1$. For $n \geqslant 2$ using the induction hypothesis we obtain

$$\begin{aligned}
\ell_p(X_n, X) &\leqslant \frac{1}{n} \sum_{i=1}^{n-1} \frac{i^p(2i-1)}{n^{p+1}} \frac{\kappa_p}{\sqrt{i}} + (7 + \tau_p) n^{-1/2} \\
&\leqslant \frac{\kappa_p}{n^{p+2}} \sum_{i=1}^{n-1} \int_i^{i+1} 2x^{p+1/2} \mathrm{d}x + (7 + \tau_p) n^{-1/2} \\
&\leqslant \frac{\kappa_p}{n^{p+2}} \int_0^n 2x^{p+1/2} \mathrm{d}x + (7 + \tau_p) n^{-1/2} \\
&= \left[\frac{4}{2p+3} \kappa_p + (7 + \tau_p) \right] n^{-1/2} \\
&= \kappa_p n^{-1/2}.
\end{aligned}$$

This finishes the proof.

The Kolmogorov–Smirnov distance between $\mathcal{L}(W)$ and $\mathcal{L}(Z)$ is defined by

$$\begin{aligned}
d_{\mathrm{KS}}(\mathcal{L}(W), \mathcal{L}(Z)) &:= d_{\mathrm{KS}}(W, Z) \\
&:= \sup_{x \in \mathbb{R}} |\mathbb{P}(W \leqslant x) - \mathbb{P}(Z \leqslant x)|.
\end{aligned}$$

Bounds for the ℓ_p distance can be used to bound d_{KS} using the following lemma from Fill and Janson [9, Lemma 5.1]:

Lemma 3.3. *Suppose that W and Z are two random variables such that Z has a bounded Lebesgue density f_Z. For all $1 \leqslant p < \infty$, we have*

$$d_{\mathrm{KS}}(W, Z) \leqslant (p+1)^{\frac{1}{p+1}} \Big(\|f_Z\|_\infty \ell_p(W, Z) \Big)^{\frac{p}{p+1}}.$$

Combining Theorem 3.1, Lemma 3.3 and Theorem 4.2 we obtain the following bound:

Theorem 3.2. *For Y_n and X as in Theorem 1.1 we have for all $0 < \varepsilon \leqslant \frac{1}{4}$ and all $n \geqslant 1$ that*

$$\begin{aligned}
d_{\mathrm{KS}}\left(\frac{Y_n}{n}, X \right) &\leqslant \omega_\varepsilon n^{-1/2 + \varepsilon}, \\
\omega_\varepsilon &:= \left(\frac{1}{2\varepsilon} \right)^{2\varepsilon} \left\{ \|f\|_\infty \kappa_{-1+1/(2\varepsilon)} \right\}^{1-2\varepsilon},
\end{aligned}$$

where f denotes the density of X.

Proof. To $0 < \varepsilon \leqslant \frac{1}{4}$ choose $p = -1 + 1/(2\varepsilon)$ in Theorem 3.1.

4 Density and distribution function

In this section we derive properties of the limit X in Theorem 1.1 mainly concerning its density and distribution function. In particular, in Theorem 4.2 we obtain a bound for the density f of X as required for

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.000	0.0000	0.0054	0.0268	0.0811	0.2044	0.4400	0.7655	0.9768
0.005	0.0000	0.0060	0.0285	0.0853	0.2133	0.4550	0.7811	0.9809
0.010	0.0000	0.0067	0.0303	0.0896	0.2224	0.4703	0.7963	0.9844
0.015	0.0001	0.0074	0.0322	0.0942	0.2318	0.4858	0.8112	0.9874
0.020	0.0002	0.0081	0.0341	0.0989	0.2415	0.5016	0.8256	0.9900
0.025	0.0003	0.0089	0.0362	0.1038	0.2516	0.5175	0.8396	0.9922
0.030	0.0004	0.0097	0.0383	0.1089	0.2619	0.5337	0.8531	0.9939
0.035	0.0006	0.0106	0.0405	0.1142	0.2726	0.5500	0.8661	0.9954
0.040	0.0008	0.0115	0.0428	0.1198	0.2835	0.5665	0.8784	0.9965
0.045	0.0010	0.0125	0.0453	0.1255	0.2948	0.5831	0.8902	0.9975
0.050	0.0012	0.0135	0.0478	0.1314	0.3064	0.5999	0.9014	0.9982
0.055	0.0015	0.0146	0.0505	0.1376	0.3184	0.6167	0.9120	0.9987
0.060	0.0018	0.0157	0.0533	0.1440	0.3306	0.6335	0.9218	0.9991
0.065	0.0021	0.0169	0.0562	0.1506	0.3432	0.6503	0.9310	0.9994
0.070	0.0025	0.0181	0.0593	0.1575	0.3561	0.6672	0.9396	0.9996
0.075	0.0029	0.0194	0.0626	0.1647	0.3693	0.6839	0.9474	0.9997
0.080	0.0033	0.0208	0.0660	0.1721	0.3829	0.7006	0.9546	0.9998
0.085	0.0038	0.0222	0.0695	0.1797	0.3967	0.7171	0.9611	0.9999
0.090	0.0043	0.0237	0.0732	0.1877	0.4109	0.7335	0.9670	0.9999
0.095	0.0049	0.0252	0.0770	0.1959	0.4253	0.7496	0.9722	1.0000

Figure 1: Distribution function of the solution of $X =_d \sqrt{U}X + \sqrt{U}(1-\sqrt{U})$, which is the limit distribution in Theorem 1.1. All values are exact up to 10^{-4}. The value at, e.g., 0.355 can be found in column labelled 0.3 and row labelled 0.055 as $\mathbb{P}(X \leqslant 0.355) \approx 0.1376$.

Theorem 3.2. Most results in this section are derived along the lines of [12, Section 5], where the related RDE

$$X \stackrel{d}{=} UX + U(1-U), \tag{4.14}$$

discovered in Hwang and Tsai [11], is studied. We start with moments:

Lemma 4.1. *For the limit X in Theorem 1.1 for all $k \geqslant 1$, we have*

$$\mathbb{E}[X^k] = 2(k+2)!(k-1)! \sum_{i=0}^{k-1} \frac{\mathbb{E}[X^i]}{(2k-i+2)!i!}.$$

In particular, $\mathbb{E}[X] = \frac{1}{2}$, $\mathbb{E}[X^2] = \frac{4}{15}$ *and* $\mathrm{Var}(X) = \frac{1}{60}$.

Proof. We raise left and right hand side of equation (1.4) to the power of k and take expectations. This implies

$$\begin{aligned}\mathbb{E}[X^k] &= \mathbb{E}\left[\left(\sqrt{U}X + \sqrt{U}(1-\sqrt{U})\right)^k\right] \\ &= \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}\left[\sqrt{U}^k (1-\sqrt{U})^{k-i}\right] \mathbb{E}[X^i] \\ &= 2(k+1)!k! \sum_{i=0}^{k} \frac{\mathbb{E}[X^i]}{(2k-i+2)!i!},\end{aligned}$$

where we used that

$$\mathbb{E}\left[\sqrt{U}^k (1-\sqrt{U})^{k-i}\right] = \frac{2(k+1)!(k-i)!}{(2k-i+2)!}.$$

This implies the assertion.

Lemma 4.2. *For the limit X in Theorem 1.1 we have $X \in [0,1]$ almost surely. For all $\varepsilon > 0$ and all $k \geqslant 1$,*

$$\mathbb{P}(X \geqslant 1-\varepsilon) \leqslant 2^{\frac{k(k+3)}{4}} \varepsilon^{\frac{k}{2}}. \tag{4.15}$$

Proof. For the first claim set $Z_0 := 0$ and

$$Z_{n+1} := \sqrt{U_{n+1}} Z_n + \sqrt{U_{n+1}}(1-\sqrt{U_{n+1}}),$$

where, for all $n \geqslant 0$, U_{n+1} is uniformly distributed on $[0,1]$ and independent of Z_n. This construction implies that $\mathbb{P}(Z_n \in [0,1]) = 1$ for all $n \geqslant 0$. Since Z_n tends to X in law we obtain $\mathbb{P}(X \in [0,1]) = 1$.

For the second claim note that (1.4) implies

$$\mathbb{P}(X \geqslant 1-\varepsilon) = \mathbb{P}\left(\sqrt{U}X + \sqrt{U}(1-\sqrt{U}) \geqslant 1-\varepsilon\right).$$

On the event $\{\sqrt{U}X + \sqrt{U}(1-\sqrt{U}) \geqslant 1-\varepsilon\}$ we have

$$X \geqslant 2\sqrt{1-\varepsilon} - 1 \quad \text{and}$$
$$\sqrt{U} \geqslant \frac{1 + X - \sqrt{(1+X)^2 - 4(1-\varepsilon)}}{2}.$$

Using that $0 \leqslant X \leqslant 1$ almost surely we obtain $X \geqslant 1-2\varepsilon$, and $\sqrt{U} \geqslant \sqrt{1-\varepsilon} - \sqrt{\varepsilon}$, hence $U \geqslant 1 - 2\sqrt{\varepsilon}$. By independence this implies

$$\begin{aligned}\mathbb{P}(X \geqslant 1-\varepsilon) &\leqslant \mathbb{P}(X \geqslant 1-2\varepsilon, U \geqslant 1-2\sqrt{\varepsilon}) \\ &= 2\sqrt{\varepsilon}\mathbb{P}(X \geqslant 1-2\varepsilon).\end{aligned}$$

Iterating the latter inequality $k \geqslant 1$ times yields

$$\begin{aligned}\mathbb{P}(X \geqslant 1-\varepsilon) &\leqslant (2\sqrt{\varepsilon})^k \mathbb{P}(X \geqslant 1-2^k\varepsilon)\sqrt{2}\sqrt{4}\cdots\sqrt{2^{k-1}} \\ &\leqslant 2^{\frac{k(k+3)}{4}}\varepsilon^{\frac{k}{2}}.\end{aligned}$$

We turn to the density of X:

THEOREM 4.1. *The limit X in Theorem 1.1 has a Lebesgue density f satisfying $f(t) = 0$ for $t < 0$ or $t > 1$, and, for $t \in [0,1]$,*

$$f(t) = 2\int_{p_t}^{t} g(x,t)f(x)\mathrm{d}x + \int_t^1 (g(x,t)-1)f(x)\mathrm{d}x, \tag{4.16}$$

where $p_t := 2\sqrt{t}-1$.
Here, for $x \in [0,1]$ and $t < ((1+x)/2)^2$,

$$g(x,t) := \frac{1+x}{\sqrt{(1+x)^2-4t}}. \tag{4.17}$$

Proof. Let $\mu := \mathcal{L}(X)$ denote the law of X and $B \subset \mathbb{R}$ any Borel set. By (1.4) we obtain

$$\begin{aligned}\mathbb{P}(X \in B) &= \mathbb{P}\left(\sqrt{U}X + \sqrt{U}(1-\sqrt{U}) \in B\right) \\ &= \int_0^1 \mathbb{P}\left(\sqrt{U}x + \sqrt{U}(1-\sqrt{U}) \in B\right)\mathrm{d}\mu(x) \\ &= \int_0^1\int_B \varphi(x,t)\mathrm{d}t\,\mathrm{d}\mu(x) \\ &= \int_B\left(\int_0^1 \varphi(x,t)\mathrm{d}\mu(x)\right)\mathrm{d}t,\end{aligned}$$

where $\varphi(x,\cdot)$ denotes the Lebesgue density of $\sqrt{U}x + \sqrt{U}(1-\sqrt{U})$ for $x \in [0,1]$. Hence, X has a Lebesgue density f satisfying $f(t) = \int_0^1 \varphi(x,t)\mathrm{d}\mu(x)$, thus

$$f(t) = \int_0^1 \varphi(x,t)f(x)\mathrm{d}x.$$

It remains to identify $\varphi(x,\cdot)$: The distribution function F_x of $\sqrt{U}x + \sqrt{U}(1-\sqrt{U})$ is given by

$$F_x(t) = \begin{cases} 0, & \text{if } t < 0, \\ \left(\frac{1+x-\sqrt{(1+x)^2-4t}}{2}\right)^2, & \text{if } t < x, \\ 1-(1+x)\sqrt{(1+x)^2-4t}, & \text{if } t < \left(\frac{1+x}{2}\right)^2, \\ 1, & \text{otherwise.}\end{cases}$$

Thus

$$\varphi(x,t) = \begin{cases} 2g(x,t), & \text{if } p_t < x \leqslant t, \\ g(x,t)-1, & \text{if } t < x \leqslant 1, \\ 0, & \text{otherwise,}\end{cases} \tag{4.18}$$

which implies the assertion.

REMARK 4.1. *Note that, w.r.t. x, g defined in (4.17) admits the simple primitive*

$$G(x,t) = \sqrt{(1+x)^2-4t}$$

which is 0 at $x = p_t$. Moreover, $g(x,\cdot)$ is increasing for fixed x and $g(\cdot,t)$ is decreasing for fixed t. Indeed,

$$\frac{\partial g}{\partial t}(x,t) = \frac{1}{2}(1+x)\left((1+x)^2-4t\right)^{-3/2} > 0,$$

and

$$\frac{\partial g}{\partial x}(x,t) = -4t\left((1+x)^2-4t\right)^{-3/2} \leqslant 0,$$

with equality if and only if $t = 0$.

COROLLARY 4.1. *The version of the density f of X with (4.16) satisfies $f(0) = 0$, $f(1) = 0$ and is increasing on $[0, \frac{1}{4}]$.*

Proof. Since $f(x) = 0$ for all $x \in (p_0, 0) = (-1,0)$ and $g(x,0) = 1$ for all $x \in (0,1)$, we obtain $f(0) = 0$ from (4.16). Since $p_1 = 1$, we also get $f(1) = 0$.

For the monotonicity from (4.16) we obtain, for $0 \leqslant s < t \leqslant \frac{1}{4}$, that

$$\begin{aligned}f(t) - f(s) = &\int_0^1 \underbrace{[g(x,t)-g(x,s)]}_{>0} f(x)\mathrm{d}x \\ &+ \int_0^s \underbrace{[g(x,t)-g(x,s)]}_{>0} f(x)\mathrm{d}x \\ &+ \int_s^t \underbrace{[g(x,t)+1]}_{>0} f(x)\mathrm{d}x \\ > 0,&\end{aligned}$$

using that $g(x,\cdot)$ is increasing for any fixed x (Remark 4.1).

THEOREM 4.2. *The density f of X in Theorem 1.1 is bounded with $\|f\|_\infty \leqslant 109$.*

Proof. We bound $f(t)$ for $t \in (0,1)$ since $f(t) = 0$ elsewhere. For $t < \frac{1}{4}$, using (4.16) and the monotonicity in Remark 4.1 we have the bound

$$\begin{aligned}f(t) &\leqslant 2\int_{p_t}^1 g(x,t)f(x)\mathrm{d}x \leqslant 2g(0,t)\int_0^1 f(x)\mathrm{d}x \\ &= \left(\frac{1}{4}-t\right)^{-\frac{1}{2}}.\end{aligned} \tag{4.19}$$

Subsequently, we split the first integral in (4.19) into a left part where we will bound f and a right part where we will bound g: For any $\gamma \in (p_t, 1]$, we split

$$f(t) \leqslant 2\int_{p_t}^{\gamma} g(x,t)f(x)\mathrm{d}x + 2\int_{\gamma}^1 g(x,t)f(x)\mathrm{d}x. \tag{4.20}$$

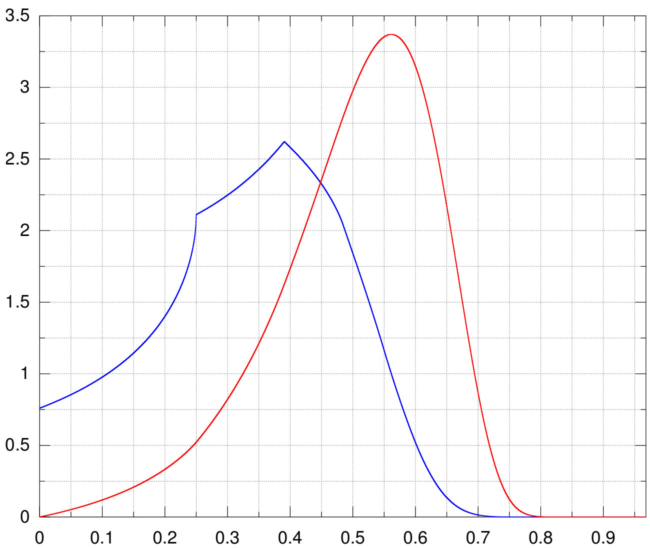

Figure 2: Approximated densities of RDE (1.4) (red) and RDE (4.14) (blue).

Let

$$\gamma = \gamma_t := \frac{p_t + t}{2} \in (p_t, 1],$$
$$\mu_t := \sup\{f(\tau) \mid \tau \in (p_t, \gamma_t)\}.$$

From (4.20) we obtain, for all $t \in [0, 1]$,

$$\begin{aligned} f(t) &\leqslant 2\mu_t \int_{p_t}^{\gamma_t} g(x,t)\mathrm{d}x + 2g(\gamma_t, t) \int_{p_t}^{1} f(x)\mathrm{d}x. \\ &= 2\mu_t G(\gamma_t, t) \qquad (4.21) \\ &\quad + 2g(\gamma_t, t)\mathbb{P}(X \geqslant 1 - 2(1-\sqrt{t})), \end{aligned}$$

with $G(\cdot,\cdot)$ as in Remark 4.1. Hence

$$G(\gamma_t, t) = \frac{1}{2}(1-\sqrt{t})\sqrt{1+6\sqrt{t}+t}, \tag{4.22}$$

and

$$\begin{aligned} g(\gamma_t, t) &= \frac{1 + \frac{2\sqrt{t}-1+t}{2}}{G(\gamma_t, t)} \\ &= \frac{(1+\sqrt{t})^2}{(1-\sqrt{t})\sqrt{1+6\sqrt{t}+t}}. \end{aligned} \tag{4.23}$$

We have $\gamma_{\frac{1}{4}} = \frac{1}{8}$, so, by (4.19), $\mu_{\frac{1}{4}} \leqslant 2\sqrt{2}$. Therefore

$$\begin{aligned} f\left(\frac{1}{4}\right) &\leqslant 4\sqrt{2}G\left(\frac{1}{8}, \frac{1}{4}\right) + 2g\left(\frac{1}{8}, \frac{1}{4}\right)\mathbb{P}(X \geqslant 0) \\ &= \sqrt{\frac{17}{2}} + \frac{18}{\sqrt{17}} \leqslant 8 =: M_0. \end{aligned}$$

Since f is increasing on $I_0 := [0, \frac{1}{4}]$, see Corollary 4.1, we obtain

$$f(t) \leqslant M_0, \quad 0 \leqslant t \leqslant \frac{1}{4}.$$

We now bound f on $(\frac{1}{4}, 1)$. To do so we decompose this interval into subintervals I_n where, for each I_n, we will deduce a bound M_n. Define $b_0 := 0$, and, for $i \geqslant 1$, $k \geqslant 1$,

$$b_i := \left(\frac{b_{i-1}+1}{2}\right)^2,$$
$$I_{2k-1} := \left(b_k, \frac{b_k + b_{k+1}}{2}\right], \ I_{2k} := \left(\frac{b_k + b_{k+1}}{2}, b_{k+1}\right].$$

We have $b_1 = \frac{1}{4}$ and b_i increases towards 1 as $i \to \infty$, so that

$$\left(\frac{1}{4}, 1\right) = \bigcup_{n=1}^{\infty} I_n.$$

Let $I_{-1} := \emptyset$. In a first step we show for all $n \geqslant 1$ that

$$(p_t, \gamma_t) \subseteq I_{n-2} \cup I_{n-1} \ \text{ for all } \ t \in I_n. \tag{4.24}$$

We denote $I_n =: (\alpha_n, \beta_n]$. If $n = 2k-1, k \geqslant 1$, then $p_t > p_{\alpha_n} = p_{b_k} = b_{k-1}$, and $\gamma_t \leqslant \gamma_{\beta_n} \leqslant b_k$ since

$$\gamma_{\beta_n} := \frac{2\sqrt{\frac{b_k + b_{k+1}}{2}} - 1 + \frac{b_k + b_{k+1}}{2}}{2} \leqslant b_k$$

since $(17-b_k)(1-b_k)^3 \geqslant 0$. Hence $(p_t, \gamma_t) \subseteq (b_{k-1}, b_k] = I_{2k-3} \cup I_{2k-2}$. In the other case $n = 2k, k \geqslant 1$, we have $p_t \leqslant p_{\beta_n} = p_{b_{k+1}} = b_k$, so $\gamma_t := \frac{p_t+t}{2} \leqslant \frac{b_k+b_{k+1}}{2}$, and $p_t > p_{\alpha_n} \geqslant \frac{b_{k-1}+b_k}{2}$ since

$$p_{\alpha_n} := 2\sqrt{\frac{b_k + b_{k+1}}{2}} - 1 \geqslant \frac{b_{k-1} + b_k}{2},$$

which holds because of $(b_{k-1} - 1)^4 \geqslant 0$. Thus $(p_t, \gamma_t) \subseteq I_{2k-2} \cup I_{2_k-1}$, and (4.24) is proved.

Inductively we now define bounds M_n for f on I_n for all $n \geqslant 0$. We already have $M_0 = 8$ and set $M_{-1} := 0$. For each $n \geqslant 1$ we use (4.15) with $\varepsilon = 2(1-\sqrt{t})$ and $k = 2$, and obtain

$$\mathbb{P}\left(X \geqslant 1 - 2(1-\sqrt{t})\right) \leqslant 8\sqrt{2}(1-\sqrt{t}).$$

Plugging this into (4.21), and substituting expressions (4.22) and (4.23), we have, for all $t \in I_n$,

$$\begin{aligned} f(t) &\leqslant (1-\sqrt{t})\sqrt{1+6\sqrt{t}+t}\max\{M_{n-2}, M_{n-1}\} \\ &\quad + \frac{16\sqrt{2}(1+\sqrt{t})^2}{\sqrt{1+6\sqrt{t}+t}} \\ &\leqslant \lceil v(\alpha_n)\max\{M_{n-1}, M_{n-2}\}\rceil + 32 =: M_n \end{aligned}$$

since the map $t \mapsto v(t) := (1-\sqrt{t})\sqrt{1+6\sqrt{t}+t}$ is decreasing on $(\frac{1}{4},1)$. We obtain $M_0 = 8$, $M_1 = 41$, $M_2 = 71$, $M_3 = 93$, $M_4 = 106$, $M_5 = 109$, $M_6 = 106$, and since $v(t) < \frac{77}{109}$ for $t > b_4$, $M_n \leqslant 109$ for $n > 6$. This completes the proof of Theorem 4.2.

The bound of 109 in Theorem 4.2 appears to be poor as the plot in Figure 2 indicates $\|f\|_\infty \leqslant 3.5$.

THEOREM 4.3. *The version of the density f of X with* (4.16) *has a right derivative at* 0 *with*

$$f_r'(0) = \mathbb{E}\left[\frac{2}{(1+X)^2}\right] \approx 0.911364.$$

Hence, f is not differentiable at 0*, for $f_\ell'(0) = 0$. We have $\mathbb{E}[X^{-2+\varepsilon}] < \infty$ for all $\varepsilon > 0$.*

Proof. Let $t \in (0, \frac{1}{4}]$. From (4.16) we have

$$f(t) = \int_0^1 2\mathbb{1}_{\{x<t\}} g(x,t) f(x)\mathrm{d}x + \int_0^1 \mathbb{1}_{\{x>t\}}(g(x,t)-1)f(x)\mathrm{d}x \tag{4.25}$$

where $g(\cdot,\cdot)$ is given in (4.17). For $x \in (0,1]$ we have

$$g(x,t) \leqslant \sqrt{\frac{2}{x}} =: \varphi(x).$$

Hence $0 \leqslant g(x,t)f(x) \leqslant \|f\|_\infty \varphi(x)$ for all $(x,t) \in (0,1] \times (0,\frac{1}{4}]$. Since $g(x,t) \to 1$ as $t \downarrow 0$ for all $x \in (0,1]$ and φ is integrable (on $(0,1]$) Lebesgue's dominated convergence theorem allows to interchange integration with the limit $t \downarrow 0$. This implies $f(t) \to 0$ as $t \downarrow 0$, thus f is continuous at 0.

Now, substituting x with xt in the first integral in (4.25), we obtain

$$\frac{f(t)}{\sqrt{t}} = \int_0^1 2\sqrt{t} g(xt,t) f(xt)\mathrm{d}x + \int_0^1 \mathbb{1}_{\{x>t\}} \frac{g(x,t)-1}{\sqrt{t}} f(x)\mathrm{d}x \tag{4.26}$$

for all $t \in (0,\frac{1}{4}]$. The first integrand in the latter display tends to 0 as $t \downarrow 0$ and, using that f is increasing on $(0,\frac{1}{4}]$, see Corollary 4.1, we obtain for all $x \in (0,1]$ that

$$0 \leqslant \sqrt{t} g(xt,t) f(xt) \leqslant \frac{2f(xt)}{\sqrt{x}} \leqslant \frac{2f(x)}{\sqrt{x}} =: \psi(x).$$

Note that ψ is integrable since, using (1.4),

$$\begin{aligned}
\int_0^1 \psi(x)\mathrm{d}x &= 2\mathbb{E}\left[\frac{1}{\sqrt{X}}\right] \\
&= 2\mathbb{E}\left[\left(\sqrt{U}X + \sqrt{U}(1-\sqrt{U})\right)^{-1/2}\right] \\
&\leqslant 2\mathbb{E}\left[\left(\sqrt{U}(1-\sqrt{U})\right)^{-1/2}\right] \\
&= 2\pi.
\end{aligned}$$

Hence, by dominated convergence, the first integrand in (4.26) tends to 0 as $t \downarrow 0$. For the second integrand in (4.26), plugging in (4.17), we find

$$\frac{g(x,t)-1}{\sqrt{t}} \to 0 \quad \text{as} \quad t \downarrow 0$$

and this fraction is dominated by φ uniformly in $t \in (0,\frac{1}{4}]$. Hence, altogether we obtain $f(t)/\sqrt{t} \to 0$ as $t \downarrow 0$. In particular, $f(t)/\sqrt{t}$ is bounded by some constant C. Finally,

$$\frac{f(t)}{t} = \int_0^1 2g(xt,t)f(xt)\mathrm{d}x + \int_0^1 \mathbb{1}_{\{x>t\}} \frac{g(x,t)-1}{t} f(x)\mathrm{d}x$$

where the first integrand is dominated by $\sqrt{8}C$ and tends to 0 as $t \downarrow 0$ (since $f(xt) \to 0$). The second integrand is dominated by $2\|f\|_\infty\varphi$ and tends to $2f(x)/(1+x)^2$ as $t \downarrow 0$. With the limit $t \downarrow 0$ and dominated convergence we obtain that f has a right derivative at 0 with

$$\begin{aligned}
f_r'(0) &= \mathbb{E}\left[\frac{2}{(1+X)^2}\right] \\
&= 2\sum_{k=0}^{\infty}(-1)^k(k+1)\mathbb{E}[X^k].
\end{aligned}$$

The interchange of summation and expectation in the latter display is justified by the fact that, for $0 < \eta < 1$,

$$\begin{aligned}
&\left|\int_\eta^1 \sum_{k=n+1}^{\infty} (-1)^k(k+1)x^k f(x)\mathrm{d}x\right| \\
&\leqslant \int_\eta^1 \sum_{k=n+1}^{\infty} (k+1)x^k f(x)\mathrm{d}x \\
&= \sum_{k=n+1}^{\infty} (k+1)\mathbb{P}(X^k \geqslant \eta),
\end{aligned}$$

where we used Levi's monotone convergence theorem and this is further bounded using Lemma 4.2 and denoting $\lambda := -\log\eta$ by

$$\begin{aligned}&\sum_{k=n+1}^{\infty}(k+1)\mathbb{P}(X \geqslant 1-\tfrac{\lambda}{k})\\&\leqslant 2^{10}\lambda^{5/2}\sum_{k=n+1}^{\infty}(k+1)k^{-5/2}\\&\to 0, \quad \text{as} \quad n\to\infty\end{aligned}$$

and the series $\sum(-1)^k(k+1)f(x)x^k$ is normally convergent on $[0,\eta]$.

The approximation for $f'_r(0)$ in the statement of Theorem 4.3 is obtained using (4.27) and Lemma 4.1.

Finally, since $t \mapsto f(t)/t$ remains bounded we obtain $\mathbb{E}\big[X^{-2+\varepsilon}\big] < \infty$ for all $\varepsilon > 0$.

THEOREM 4.4. *For all $0 < \varepsilon < 1$, the version of the density f with* (4.16) *is Hölder continuous on $[0, 1-\varepsilon]$ with Hölder exponent $\frac{1}{2}$: if $0 \leqslant s < t \leqslant 1-\varepsilon$, then*

$$|f(t)-f(s)| \leqslant (9+6\varepsilon^{-3/2})\|f\|_\infty\sqrt{t-s}.$$

Proof. Let $0 \leqslant s < t \leqslant 1$. From the integral equation (4.16), we deduce that

$$\begin{aligned}&|f(t)-f(s)|\\&\leqslant 2\left|\int_{p_t}^{t}g(x,t)f(x)\mathrm{d}x-\int_{p_s}^{s}g(x,s)f(x)\mathrm{d}x\right|\\&\quad+\left|\int_{t}^{1}g(x,t)f(x)\mathrm{d}x-\int_{s}^{1}g(x,s)f(x)\mathrm{d}x\right|\\&\quad+\int_{s}^{t}f(x)\mathrm{d}x\\&=: C_1+C_2+C_3.\end{aligned}$$

We have $C_3 \leqslant \|f\|_\infty(t-s) \leqslant \|f\|_\infty\sqrt{t-s}$. Using the primitive of $g(\cdot,t)$ given in Remark 4.1 and the monotonicity of $g(x,\cdot)$,

$$\begin{aligned}C_1 &\leqslant 2\int_{p_t}^{t}(g(x,t)-g(x,s))f(x)\mathrm{d}x\\&\quad+2\int_{s}^{t}g(x,s)f(x)\mathrm{d}x+2\int_{p_s}^{p_t}g(x,s)f(x)\mathrm{d}x\\&\leqslant 2\|f\|_\infty\left(\int_{p_t}^{t}g(\cdot,t)+\int_{p_s}^{p_t}g(\cdot,s)-\int_{p_t}^{s}g(\cdot,s)\right)\\&\leqslant 2\|f\|_\infty(4\sqrt{t-s}-(t-s))\\&\leqslant 8\|f\|_\infty\sqrt{t-s}.\end{aligned}$$

Finally, with $u_{x,s} := \sqrt{(1+x)^2-4s} \geqslant u_{x,t} \geqslant \sqrt{\varepsilon}$ for all $t \leqslant x \leqslant 1$, and using that $g(\cdot,s)$ is decreasing,

$$\begin{aligned}C_2 &\leqslant \int_t^1(g(x,t)-g(x,s))f(x)\mathrm{d}x+\int_s^t g(x,s)f(x)\mathrm{d}x\\&=\int_t^1\frac{4(1+x)(t-s)f(x)}{u_{x,t}u_{x,s}(u_{x,t}+u_{x,s})}\mathrm{d}x+\int_s^t g(x,s)f(x)\mathrm{d}x\\&\leqslant \|f\|_\infty(t-s)\left(\int_t^1 4\varepsilon^{-3/2}\mathrm{d}x+g(s,s)\right)\\&\leqslant (4\varepsilon^{-3/2}+2\varepsilon^{-1})\|f\|_\infty(t-s)\\&\leqslant 6\varepsilon^{-3/2}\|f\|_\infty\sqrt{t-s}.\end{aligned}$$

This completes the proof.

For the distribution function of the limit X in Theorem 1.1 we can apply a variant of a numerical approximation developed in [12] for which a rigorous error analysis shows all values in the table of Figure 1 being exact up to 10^{-4}.

5 Perfect simulation

We construct an algorithm for perfect (exact) simulation from the limit X in Theorem 1.1. We assume that a sequence of independent and uniformly on $[0,1]$ distributed random variables is available and that elementary operations of and between real numbers can be performed exactly; see Devroye [4] for a comprehensive account on non-uniform random number generation. Methods based on coupling from the past have been developed and applied for the exact simulation from perpetuities in [8, 6, 7, 13, 2]; see also [5]. Our perpetuity $X =_d \sqrt{U}X+\sqrt{U}(1-\sqrt{U})$ shares properties of $X =_d UX+U(1-U)$ considered in [13] which simplify the construction of an exact simulation algorithm considerably compared to the examples of the Vervaat perpetuities and the Dickman distribution in Fill and Huber [8] and Devroye and Fawzi [6]. Most notably the Markov chain underlying $X =_d \sqrt{U}X+\sqrt{U}(1-\sqrt{U})$ is positive Harris recurrent which allows to directly construct a multigamma coupler as developed in Murdoch and Green [19, Section 2.1]. The design of the following algorithm `Simulate[`$X =_d \sqrt{U}X+\sqrt{U}(1-\sqrt{U})$`]` is similar to the construction in [13]: We construct an update function $\Phi : [0,1]\times\{0,1\}\times[0,1] \to [0,1]$ such that first for all $x \in [0,\infty)$ we have that $\sqrt{U}x+\sqrt{U}(1-\sqrt{U})$ and $\Phi(x,B,U)$ are identically distributed, where U is uniformly distributed on $[0,1]$ and B is an independent Bernoulli distributed random variable, and second coalescence of the underlying Markov chains is supported.

Recall the densities $\varphi(x,\cdot)$ of $\sqrt{U}x+\sqrt{U}(1-\sqrt{U})$ given explicitly in (4.18). Fix $t \in (\frac{1}{8},\frac{1}{4})$. For all $x \in [0,t]$, we have $\varphi(x,t) \geqslant \varphi(t,t) = \frac{2(1+t)}{1-t} \geqslant \varphi(\frac{1}{8},\frac{1}{8})$, and for $x \in (t,1]$, we obtain $\varphi(x,t) \geqslant \varphi(1,t) = \frac{1}{\sqrt{1-t}} -$

$1 \geqslant \varphi(1, \frac{1}{8}) \geqslant \varphi(\frac{1}{8}, \frac{1}{8})$. Thus, noting $\alpha := \varphi(\frac{1}{8}, \frac{1}{8}) = \sqrt{8/7} - 1 \approx 0.069$,

$$\varphi(x,t) \geqslant r(t) := \alpha \mathbb{1}_{(\frac{1}{8},\frac{1}{4})}(t) \quad \text{for all } (x,t) \in [0,1]^2.$$

Consequently, we can write $\varphi(x, \cdot) = r + g_x$ for some nonnegative functions g_x for all $x \in [0,1]$. Note that $1 = \|r\|_1 + \|g_x\|_1$, with $\|r\|_1 := \int_{\mathbb{R}} r(t)\mathrm{d}t = \frac{\alpha}{8}$.

Let R, Y^x, B be random variables with R having density $r/\|r\|_1$, Y^x having density $g_x/\|g_x\|_1$, and B with Bernoulli($\|r\|_1$) distribution and independent of (R, Y^x). Then we have

$$BR + (1-B)Y^x \stackrel{d}{=} \sqrt{U}x + \sqrt{U}(1-\sqrt{U}).$$

Hence we can use the update function

$$\Phi(x,b,u) = b\left(\frac{1}{8}u + \frac{1}{8}\right) + (1-b)G_x^{-1}(u).$$

We construct our Markov chains from the past using Φ as an update function. In each transition there is a probability of $\|r\|_1 = \alpha/8$ that all chains couple simultaneously. In other words, we can just start at a geometric Geom($\alpha/8$) distributed time τ in the past, the first instant of $\{B = 1\}$ when moving back into the past. At this time $-\tau$ we couple all chains via $X_{-\tau} := \frac{1}{8}U + \frac{1}{8}$ and let the chains run from there until time 0 using the updates $G^{-1}_{X_{-k}}(U_{-k+1})$ for $-k = -\tau, \ldots, -1$. It is shown in Murdoch and Green [19, Section 2.1] that this is a valid implementation of the coupling from the past algorithm in general.

Hence, it remains to invert the distribution functions $G_x\colon [0,1] \to [0,1]$ of Y^x. We have $\|g_x\|_1 = 1 - \|r\|_1 = \frac{8-\alpha}{8}$, and

$$\begin{aligned} G_x(t) &:= \frac{8}{8-\alpha}\int_0^t (\varphi_x(u) - r(u))\mathrm{d}u \\ &= \frac{8}{8-\alpha}\left(F_x(t) - \frac{\alpha}{8}\max\{0, \min\{t - \tfrac{1}{8}, \tfrac{1}{8}\}\}\right), \end{aligned}$$

with F_x obtained in the proof of Theorem 4.1. The inversions of the functions G_x can be computed explicitly and lead to the functions G_x^{-1} stated below.

With the sequence $(U_{-k})_{k\geqslant 0}$ of independent uniformly on $[0,1]$ distributed random variables and an independent Geom($\frac{\alpha}{8}$) geometrically distributed random variable we obtain the following algorithm:

Algorithm 1: `Simulate[`$X \stackrel{d}{=} \sqrt{U}X + \sqrt{U}(1-\sqrt{U})$`]`

$\tau \leftarrow \mathrm{Geom}\left(\frac{1}{2\sqrt{14}} - \frac{1}{8}\right)$
$X \leftarrow \frac{1}{8}U_{-\tau} + \frac{1}{8}$
for k **from** $-\tau + 1$ **to** 0 **do**
  $X \leftarrow G_X^{-1}(U_k)$
return X

The function G^{-1} is given by

$$G_x^{-1}(u) = \begin{cases} {}_1G_x^{-1}(u), & \text{if } x \in [0, 1/8), u \in [0, \mathrm{a}_x), \\ {}_2G_x^{-1}(u), & \text{if } x \in [0, 1/8), u \in [\mathrm{a}_x, \mathrm{b}_x), \\ {}_3G_x^{-1}(u), & \text{if } x \in [0, 1/8), u \in [\mathrm{b}_x, \mathrm{c}_x), \\ {}_4G_x^{-1}(u), & \text{if } x \in [0, 1/8), u \in [\mathrm{c}_x, 1], \\ {}_1G_x^{-1}(u), & \text{if } x \in [1/8, 1/4), u \in [0, \mathrm{d}_x), \\ {}_5G_x^{-1}(u), & \text{if } x \in [1/8, 1/4), u \in [\mathrm{d}_x, \mathrm{e}_x), \\ {}_3G_x^{-1}(u), & \text{if } x \in [1/8, 1/4), u \in [\mathrm{e}_x, \mathrm{c}_x), \\ {}_4G_x^{-1}(u), & \text{if } x \in [1/8, 1/4), u \in [\mathrm{c}_x, 1], \\ {}_1G_x^{-1}(u), & \text{if } x \in [1/4, 1], u \in [0, \mathrm{d}_x), \\ {}_5G_x^{-1}(u), & \text{if } x \in [1/4, 1], u \in [\mathrm{d}_x, \mathrm{f}_x), \\ {}_6G_x^{-1}(u), & \text{if } x \in [1/4, 1], u \in [\mathrm{f}_x, \mathrm{g}_x), \\ {}_4G_x^{-1}(u), & \text{if } x \in [1/4, 1], u \in [\mathrm{g}_x, 1], \end{cases}$$

where

$${}_1G_x^{-1}(u) := \left(\frac{\alpha}{8} - 1\right)u + \frac{1+x}{4}\sqrt{2(8-a)}\sqrt{u},$$

$${}_2G_x^{-1}(u) := \frac{64(1+x)^4 - (8(1-u) + \alpha u)^2}{256(1+x)^2},$$

$$\begin{aligned} {}_3G_x^{-1}(u) := \frac{1}{8\alpha^2}\Big[&2\alpha^2 + \alpha(8-\alpha)(1-u) - 16(1+x)^2 \\ &+ 4(1+x)\sqrt{4(4+\alpha^2)(1+x)^2 - 2\alpha(8-\alpha)(1-u) - 4\alpha^2}\Big], \end{aligned}$$

$${}_4G_x^{-1}(u) := \frac{64(1+x)^4 - (8-\alpha)^2(1-u)^2}{256(1+x)^2},$$

$$\begin{aligned} {}_5G_x^{-1}(u) := \frac{1}{8(1+\alpha)^2}\Big[&4\alpha(1+x)^2 + (1+\alpha)(\alpha u + \alpha - 8u) \\ &+ 2(1+x)\sqrt{4\alpha^2(1+x)^2 - 2(1+\alpha)(\alpha u + \alpha - 8u)}\Big], \end{aligned}$$

$${}_6G_x^{-1}(u) := \left(\frac{\alpha}{8} - 1\right)u + \frac{1+x}{4}\sqrt{2}\sqrt{8u + \alpha(1-u)} - \frac{\alpha}{8},$$

and

$$\mathrm{a}_x := \frac{8x^2}{8-\alpha}, \quad \mathrm{b}_x := \frac{4(2-(1+x)\sqrt{4x^2+8x+2})}{8-\alpha},$$
$$\mathrm{c}_x := \frac{8(1-(1+x)\sqrt{x^2+2x}) - \alpha}{8-\alpha},$$

$$\mathrm{d}_x := \frac{(2+2x-\sqrt{4x^2+8x+2})^2}{16-2\alpha}, \quad \mathrm{e}_x := \frac{8x^2 + (1-8x)\alpha}{8-\alpha},$$

$$\mathrm{f}_x := \frac{4x^2 + 8x + 2 - \alpha - 4(1+x)\sqrt{x^2+2x}}{8-\alpha},$$
$$\mathrm{g}_x := \frac{8x^2 - \alpha}{8-\alpha}.$$

Copyable versions of the latter expressions are given below (`G`k denotes ${}_kG_x^{-1}$ for $k = 1, \ldots, 6$ and `a`, `b`, `c`, `d`, `e`, `f`, `g` respectively denote a_x, b_x, c_x, d_x, e_x, f_x, g_x).

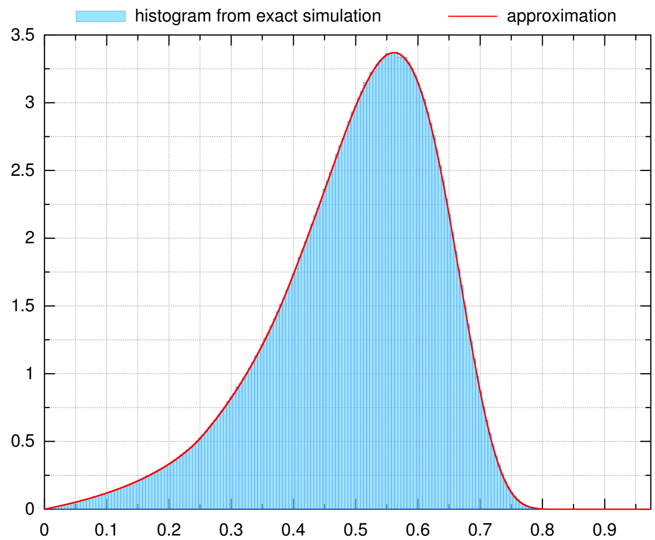

Figure 3: Normalized histogram of exact simulations (10,000,000 samples) of RDE (1.4) with Algorithm 1.

```
A=sqrt(8./7)-1
G1=(A/8-1)*u+(1+x)/4*sqrt(2*(8-A))*sqrt(u)
G2=(64*(1+x)^4-(8*(1-u)+A*u)^2)/(256*(1+x)^2)
G3=(2*A*A+A*(8-A)*(1-u)-16*(1+x)^2+4*(1+x)*sqrt
     (4*(4+A*A)*(1+x)^2-2*A*(8-A)*(1-u)-4*A*A))
     /(8*A*A)
G4=(64*(1+x)^4-(8-A)^2*(1-u)^2)/(256*(1+x)^2)
G5=(4*A*(1+x)^2+(1+A)*(A*u+A-8*u)+2*(1+x)*sqrt
     (4*A*A*(1+x)^2-2*(1+A)*(A*u+A-8*u)))/(8*(1+
     A)^2)
G6=(A/8-1)*u+(1+x)*sqrt(2*(8*u+A*(1-u)))/4-A/8

a=8*x*x/(8-A)
b=4*(2-(1+x)*sqrt(4*x*x+8*x+2))/(8-A)
c=(8*(1-(1+x)*sqrt(x*x+2*x))-A)/(8-A)
d=(2+2*x-sqrt(4*x*x+8*x+2))^2/(16-2*A)
e=(8*x*x+(1-8*x)*A)/(8-A)
f=(4*x*x+8*x+2-A-4*(1+x)*sqrt(x*x+2*x))/(8-A)
g=(8*x*x-A)/(8-A)
```

References

[1] Bickel, P.J. and Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.

[2] Blanchet, J.H. and Sigman, K. (2011) On exact sampling of stochastic perpetuities. *J. Appl. Probab.* **48A**, 165–182.

[3] Cormen, T.H., Leiserson, C.E. and Rivest, R.L. (1990) Introduction to algorithms. The MIT Electrical Engineering and Computer Science Series. *MIT Press, Cambridge, MA; McGraw-Hill Book Co., New York.*

[4] Devroye, L. (1986) Nonuniform random variate generation. *Springer-Verlag, New York.*

[5] Devroye, L. (2001) Simulating perpetuities. *Methodol. Comput. Appl. Probab.* **3**, 97-115.

[6] Devroye, L. and Fawzi, O. (2010) Simulating the Dickman distribution. *Statist. Probab. Lett.* **80**, 242–247.

[7] Devroye, L. and James, L. (2011) The double CFTP method. *ACM Trans. Model. Comput. Simul.* **21**, 1–20.

[8] Fill, J.A. and Huber, M.L. (2010) Perfect simulation of Vervaat perpetuities. *Electron. J. Probab.* **15**, 96–109.

[9] Fill, J.A. and Janson, S. (2002) Quicksort asymptotics. *J. Algorithms* **44**, 4–28.

[10] Hoare, C.A.R. (1961) Algorithm 65, FIND *Comm. Assoc. Comput. Mach.* **4**, 321–322.

[11] Hwang, H.-K. and Tsai, T.-H. (2002) Quickselect and the Dickman function. *Combin. Probab. Comput.* **11**, 353–371.

[12] Knape, M. and Neininger, R. (2008) Approximating perpetuities. *Methodol. Comput. Appl. Probab.* **10**, 507–529.

[13] Knape, M. and Neininger, R. (2013) Appendix to "Approximating perpetuities". *Methodol. Comput. Appl. Probab.* **15**, 707–712.

[14] Mahmoud, H.M. (2000) Sorting. A distribution theory. Wiley-Interscience Series in Discrete Mathematics and Optimization. *Wiley-Interscience, New York.*

[15] Mahmoud, H.M. (2009) Average-case Analysis of Moves in Quick Select. *Proceedings of the Sixth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, 35–40.

[16] Mahmoud, H.M. (2010) Distributional analysis of swaps in Quick Select. *Theoret. Comput. Sci.* **411**, 1763–1769.

[17] Mahmoud, H.M., Modarres, R. and Smythe, R.T. (1995) Analysis of QUICKSELECT: an algorithm for order statistics. *RAIRO Inform. Théor. Appl.* **29**, 255–276.

[18] Martínez, C., Panholzer, A. and Prodinger, H. (2011) The analysis of Range Quickselect and related problems. *Theoret. Comput. Sci.* **412**, 6537–6555.

[19] Murdoch D.J. and Green P.J. (1998) Exact sampling from a continuous state space. *Scand. J. Statist.* **25**, 483–502

[20] Neininger, R. and Rüschendorf, L. (2004) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14**, 378–418.

[21] Rachev, S.T. and Rüschendorf, L. (1995) Probability metrics and recursive algorithms. *Adv. in Appl. Probab.* **27**, 770–799.

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

[23] Rösler, U. and Rüschendorf, L. (2001) The contraction method for recursive algorithms. *Algorithmica* **29**, 3–33.

[24] Sedgewick, R. (1990) Algorithms in C. Addison-Wesley Publishing Company.

[25] Serfling, R.J. (1974) Probability inequalities for the sum in sampling without replacement. *Ann. Statist.* **2**, 39–48.

[26] Vervaat, W. (1979) On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.* **11**, 750–783.