Convergence of the Population Dynamics algorithm in the Wasserstein metric

Mariana Olvera-Cravioto

May 26, 2017

Abstract

We study the convergence of the population dynamics algorithm, which produces sample pools of random variables having a distribution that closely approximates that of the *special endogenous solution* to a stochastic fixed-point equation of the form:

$$R \stackrel{\mathcal{D}}{=} \Phi(Q, N, \{C_i\}, \{R_i\}),$$

where $(Q, N, \{C_i\})$ is a real-valued random vector with $N \in \mathbb{N}$, and $\{R_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d. copies of R, independent of $(Q, N, \{C_i\})$; the symbol $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. Specifically, we show its convergence in the Wasserstein metric of order p ($p \geq 1$) and prove the consistency of estimators based on the sample pool produced by the algorithm.

Keywords: Population dynamics; iterative bootstrap; Wasserstein metric; distributional fixed-point equations.

1 Introduction

We study an iterative bootstrap algorithm, known as the "population dynamics" algorithm, that can be used to efficiently generate samples of random variables whose distribution closely approximates that of the so-called special endogenous solution to a stochastic fixed-point equation (SFPE) of the form:

$$R \stackrel{\mathcal{D}}{=} \Phi(Q, N, \{C_i\}, \{R_i\}), \tag{1.1}$$

where $(Q, N, \{C_i\})$ is a real-valued random vector with $N \in \mathbb{N} = \{0, 1, 2, \dots\}$, and $\{R_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d. copies of R, independent of $(Q, N, \{C_i\})$. These equations appear in a variety of problems, ranging from computer science to statistical physics, e.g.: in the analysis of divide and conquer algorithms such as Quicksort [25, 14, 26] and `FIND` [13], the analysis of Google's PageRank algorithm [28, 18, 9], the study of queueing networks with synchronization requirements [21, 24], and the analysis of the Ising model [12], to name a few. In general, SPFEs of the form in (1.1) can have multiple solutions, but in most cases we are interested in computing those that can be explicitly constructed on a weighted branching process, known as *endogenous* solutions. In some cases, even the endogenous solution is not unique [5], but characterizing all endogenous solutions can be done using the *special endogenous* solution, which is the only attracting solution, and can be constructed by iterating (1.1) starting from some well-behaved initial distribution.

This work focuses on the analysis of a simulation algorithm that can be used to generate samples from a distribution that closely approximates that of the special endogenous solution to a variety of SFPEs. The need for such an approximate algorithm lies on the numerical complexity of simulating even a few generations of a weighted branching process using naive Monte Carlo methods. The population dynamics algorithm, described in §14.6.4 in [23] and §8.1 in [1], circumvents this problem by resampling with replacement from previously computed iterations of (1.1), i.e., by using an iterative bootstrap technique. However, as is the case with the standard bootstrap algorithm, the samples obtained are neither independent nor exactly distributed according to the target distribution, which raises the need to study the convergence properties of the algorithm.

Before presenting the algorithm and stating our main results, it may be helpful to describe in more detail some of the examples mentioned above. Throughout the paper, we use $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$ to denote the maximum and the minimum, respectively, of x and y.

- The linear SFPE or "smoothing transform":

$$R \stackrel{\mathcal{D}}{=} Q + \sum_{i=1}^{N} C_i R_i, \tag{1.2}$$

  appears in the analysis of the number of comparisons required by the sorting algorithm Quicksort [25, 14, 26], and can also be used to describe the distribution of the ranks computed by Google's PageRank algorithm on directed complex networks [28, 18, 9].

- The maximum SFPE or "high-order Lindley equation":

$$R \stackrel{\mathcal{D}}{=} Q \vee \bigvee_{i=1}^{N} C_i R_i, \qquad \text{equivalently,} \qquad X \stackrel{\mathcal{D}}{=} T \vee \bigvee_{i=1}^{N} (\xi_i + X_i), \tag{1.3}$$

  arises as the limiting waiting time distribution on queueing networks with parallel servers and synchronization requirements [21, 24] and in the analysis of the branching random walk [1].

- The discounted tree-sum SFPE:

$$R \stackrel{\mathcal{D}}{=} Q + \bigvee_{i=1}^{N} C_i R_i \tag{1.4}$$

  appears in the worst-case analysis of the `FIND` algorithm [13] and the analysis of the "discounted branching random walk" [6].

- The "free-entropy" SFPE:

$$R \stackrel{\mathcal{D}}{=} Q + \sum_{i=1}^{N} \operatorname{arctanh}(\tanh(\beta) \tanh(R_i)) \tag{1.5}$$

  characterizes the asymptotic free-entropy density in the ferromagnetic Ising model on locally tree-like graphs [12]. In this case, $C_i \equiv \tanh(\beta)$ for all $i \geq 1$, $\beta \geq 0$ represents the "inverse temperature", and Q the magnetic field.

- Although the analysis presented here does not directly apply to this case, we mention that the population dynamics algorithm can also be used to simulate the fixed points of the belief propagation equations on random graphical models [23]:

$$R \stackrel{\mathcal{D}}{=} \Phi\left(Q, N, \{C_i\}, \{\tilde{R}_i\}\right) \qquad \text{and} \qquad \tilde{R} \stackrel{\mathcal{D}}{=} \Psi\left(\tilde{Q}, \tilde{N}, \{\tilde{C}_i\}, \{R_i\}\right),$$

  where the $\{\tilde{R}_i\}$ are i.i.d. copies of $\tilde{R}$ independent of the vector $(Q, N, \{C_i\})$ and the $\{R_i\}$ are i.i.d. copies of R independent of the vector $(\tilde{Q}, \tilde{N}, \{\tilde{C}_i\})$, with Φ and Ψ potentially different.

We refer the reader to [1] for even more examples, including some involving minimums.

The existence and uniqueness of solutions to any of these SFPEs is in itself a non-trivial problem. We refer the reader again to [1] for a broad survey of known results and open problems on this topic. The most well-studied equations are the linear (1.2) and maximum (1.3) SFPEs, which have been extensively studied in [22, 16, 2, 4, 5, 3, 17] and [7, 20], respectively. However, to provide some context to where the population dynamics algorithm fits in, we briefly mention that the existence of solutions is often established by showing that the transformation T that maps the distribution μ on $\mathbb{R}$ to the distribution of

$$\Phi\left(Q, N, \{C_i\}, \{X_i\}\right),$$

where the $\{X_i\}$ are i.i.d. random variables distributed according to μ, independent of the vector $(Q, N, \{C_i\})$, is strictly contracting under some suitable metric. Note that in this case, we have that the sequence of probability measures $\mu_{n+1} = T(\mu_n)$ converges as $n \to \infty$ to a fixed point of (1.1). Moreover, as long as the initial distribution μ_0 has sufficiently light tails, one can show that $\{\mu_n\}$ converges to the special endogenous solution to (1.1), and the contracting nature of T provides an upper bound of the form

$$d(\mu_n, \mu) \leq d(T(\mu_{n-1}), T(\mu)) \leq c d(\mu_{n-1}, \mu) \leq c^n d(\mu_0, \mu), \qquad n = 1, 2, \ldots,$$

for some constant $0 < c < 1$, where d is the distance under which T is a contraction. It follows that from a computational point of view, it suffices to have an algorithm for computing μ_k for a fixed number of iterations $k \in \mathbb{N}$. The population dynamics algorithm produces a sample of observations approximately distributed according to μ_k. We now describe how to obtain an exact sample of μ_k, which will also make clear then need for a computationally efficient method.

1.1 Constructing endogenous solutions on weighted branching processes

As mentioned earlier, the attracting endogenous solution to (1.1), provided it exists, can be constructed on a structure known as a weighted branching process [25]. We now elaborate on this point.

Let $\mathbb{N}_+ = \{1, 2, 3, \ldots\}$ be the set of positive integers and let $U = \bigcup_{k=0}^{\infty} (\mathbb{N}_+)^k$ be the set of all finite sequences $\mathbf{i} = (i_1, i_2, \ldots, i_n)$, $n \geq 0$, where by convention $\mathbb{N}_+^0 = \{\emptyset\}$ contains the null sequence $\emptyset$. To ease the exposition, we will use $(\mathbf{i}, j) = (i_1, \ldots, i_n, j)$ to denote the index concatenation operation. Next, let $(Q, N, \{C_i\}_{i \geq 1})$ be a real-valued vector with $N \in \mathbb{N}$. We will refer to this vector as the generic branching vector. Now let $\{(Q_{\mathbf{i}}, N_{\mathbf{i}}, \{C_{(\mathbf{i},j)}\}_{j \geq 1}\}_{\mathbf{i} \in U}$ be a sequence of i.i.d. copies of the generic branching vector. To construct a weighted branching process we start by defining a tree as

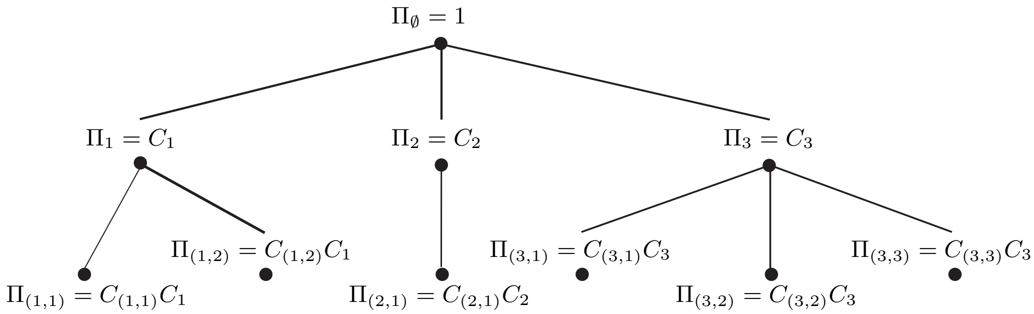

Figure 1: Weighted branching process

follows: let $A_0 = \{\emptyset\}$ denote the root of the tree, and define the nth generation according to the recursion

$$A_n = \{(\mathbf{i}, i_n) \in U : \mathbf{i} \in A_{n-1}, 1 \le i_n \le N_{\mathbf{i}}\}, \quad n \ge 1.$$

Now, assign to each node $\mathbf{i}$ in the tree a weight $\Pi_{\mathbf{i}}$ according to the recursion

$$\Pi_\emptyset \equiv 1, \qquad \Pi_{(\mathbf{i},i_n)} = C_{(\mathbf{i},i_n)} \Pi_{\mathbf{i}}, \qquad n \ge 1,$$

see Figure 1. If $P(N < \infty) = 1$ and $C_i \equiv 1$ for all $i \ge 1$, the weighted branching process reduces to a Galton-Watson process.

To generate a sample from μ_k we first need to fix the initial distribution μ_0, e.g., by letting μ_0 be the probability measure of a constant, say zero or one. Now construct a weighted branching process with k generations, and let $\{R_{\mathbf{i}}^{(0)}\}_{\mathbf{i} \in A_k}$ be i.i.d. random variables having distribution μ_0. Next, define recursively for each $\mathbf{i} \in A_{k-r}$, $1 \le r \le k$,

$$R_{\mathbf{i}}^{(r)} = \Phi \left(Q_{\mathbf{i}}, N_{\mathbf{i}}, \{C_{(\mathbf{i},j)}\}_{j \ge 1}, \{R_{(\mathbf{i},j)}^{(r-1)}\}_{j \ge 1} \right).$$

The random variable $R_\emptyset^{(k)}$ is distributed according to μ_k, and its generation requires on average $(E[N])^k$ i.i.d. copies of the generic branching vector $(Q, N, \{C_i\}_{i \ge 1})$. It follows that if the goal was to obtain an i.i.d. sample of size m from distribution μ_k, one would need to generate on average $m(E[N])^k$ copies of the generic branching vector. However, in applications one typically has $E[N] > 1$, e.g., $N \equiv 2$ for Quicksort, $E[N] \approx 30$ in the analysis of PageRank on the WWW graph, and $E[N]$ can be in the hundreds for MapReduce implementations related to the maximum SFPE. This makes the exact simulation of $R^{(k)}$ using a weighted branching process impractical.

The population dynamics algorithm, described below, uses a bootstrap approach to produce a sample of size m of random variables that are approximately distributed according to μ_k, and that although not independent, can be used to obtain consistent estimators for moments, quantiles and other functions of μ_k.

1.2 The population dynamics algorithm

The population dynamics algorithm is based on the bootstrap, i.e., in the idea of sampling with replacement random variables from a common pool. As described above, the algorithm starts by generating a sample of i.i.d. random variables having distribution μ_0, with the difference that when computing the next level of the recursion, it samples with replacement from this pool as needed by the map Φ. In other words, to obtain a pool of approximate copies of $R^{(j)}$ we bootstrap from the pool previously obtained of approximate copies of $R^{(j-1)}$. The approximation lies in the fact that we are not sampling from $R^{(j-1)}$ itself, but from a finite sample of conditionally independent observations that are only approximately distributed as $R^{(j-1)}$. The algorithm is described below.

Let $(Q, N, \{C_r\})$ denote the generic branching vector defining the weighted branching process. Let k be the depth of the recursion that we want to simulate, i.e., the algorithm will produce a sample of random variables approximately distributed according to μ_k. Choose $m \in \mathbb{N}_+$ to be the bootstrap sample size. For each $0 \leq j \leq k$, the algorithm outputs $\mathscr{P}^{(j,m)} \triangleq \left(\hat{R}_1^{(j,m)}, \hat{R}_2^{(j,m)}, \ldots, \hat{R}_m^{(j,m)}\right)$, which we refer to as the sample pool at level j.

a.) *Initialize*: Set $j = 0$. Simulate a sequence $\{R_i^{(0)}\}_{i=1}^m$ of i.i.d. random variables distributed according to some initial distribution μ_0. Let $\hat{R}_i^{(0,m)} = R_i^{(0)}$ for $i = 1, \ldots, m$. Output $\mathscr{P}^{(0,m)} = \left(\hat{R}_1^{(0,m)}, \hat{R}_2^{(0,m)}, \ldots, \hat{R}_m^{(0,m)}\right)$ and update $j = 1$.

b.) While $j \leq k$:

  (a) Simulate a sequence $\{(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}_{r\geq 1})\}_{i=1}^m$ of i.i.d. copies of the generic branching vector, independent of everything else.

  (b) Let
$$\hat{R}_i^{(j,m)} = \Phi\left(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}, \{\hat{R}_{(i,r)}^{(j-1,m)}\}\right), \qquad i = 1, \ldots, m, \tag{1.6}$$
  where the $\hat{R}_{(i,r)}^{(j-1,m)}$ are sampled uniformly with replacement from the pool $\mathscr{P}^{(j-1,m)}$.

  (c) Output $\mathscr{P}^{(j,m)} = \left(\hat{R}_1^{(j,m)}, \hat{R}_2^{(j,m)}, \ldots, \hat{R}_m^{(j,m)}\right)$ and update $j = j + 1$.

We conclude this section by pointing out that the complexity of the algorithm described above is of order km, while the naive Monte Carlo approach described earlier, which consists on sampling m i.i.d. copies of a weighted branching process up to the kth generation, has order $(E[N])^k m$. Our main results establish the convergence of the algorithm in the Wasserstein metric of order p ($p \geq 1$), as well as the consistency of estimators constructed using the pool $\mathscr{P}^{(k,m)}$. The following section contains all the statements, and the proofs are given in Section 3.

2 Main results

We start by defining the Wasserstein metric of order p.

Definition 2.1 *Let $M(\mu, \nu)$ denote the set of joint probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν. Then, the Wasserstein metric of order p $(1 \leq p < \infty)$ between μ and ν is given by*
$$d_p(\mu, \nu) = \inf_{\pi \in M(\mu,\nu)} \left(\int_{\mathbb{R}\times\mathbb{R}} |x - y|^p \, d\pi(x, y) \right)^{1/p}.$$

An important advantage of working with the Wasserstein metrics is that on the real line they admit the explicit representation

$$d_p(\mu, \nu) = \left(\int_0^1 |F^{-1}(u) - G^{-1}(u)|^p du \right)^{1/p}, \tag{2.1}$$

where F and G are the cumulative distribution functions of μ and ν, respectively, and $f^{-1}(t) = \inf\{x \in \mathbb{R} : f(x) \geq t\}$ denotes the generalized inverse of f. It follows that the optimal coupling of two real random variables X and Y is given by $(X, Y) = (F^{-1}(U), G^{-1}(U))$, where U is uniformly distributed in $[0, 1]$.

With some abuse of notation, we use the notation $d_p(F, G)$ to denote the Wasserstein distance of order p between the probability measures μ and ν, where $F(x) = \mu((-\infty, x])$ and $G(x) = \nu((-\infty, x])$ are their corresponding cumulative distribution functions.

The Wasserstein metric has proven to be very useful in the analysis of distributional fixed-point equations, since it is often easy to find a value of $p \geq 1$ under which T is a contraction. For example, for the linear (1.2) and maximum (1.3) SFPEs, a sufficient condition guaranteeing that T is a strict contraction under d_1 is that $E\left[\sum_{i=1}^N |C_i|^\beta\right] < 1$ for some $0 < \beta \leq 1$; if $E[Q] = 0$ and $E\left[\sum_{i=1}^N C_i^2\right] < 1$, then the linear SFPE defines a contraction under d_2.

Our main results establish the convergence of $d_p(\hat{F}_{k,m}, F_k)$ as $m \to \infty$, both in mean and almost surely, where

$$\hat{F}_{k,m}(x) = \frac{1}{m} \sum_{i=1}^m 1(\hat{R}_i^{(k,m)} \leq x) \qquad \text{and} \qquad F_k(x) = \mu_k((-\infty, x]), \qquad k \in \mathbb{N},$$

and $\mathscr{P}^{(k,m)} = \left(\hat{R}_1^{(k,m)}, \ldots, \hat{R}_m^{(k,m)}\right)$ is the pool generated by the population dynamics algorithm. The theorems are proven under two different assumptions, the first one imposing a Lipschitz condition on the mean of Φ, and the second one requiring Φ to be Lipschitz continuous almost surely.

Assumption 2.2 *For some $p \geq 1$ there exist a constant $0 < H_p < \infty$ such that if $\{(X_i, Y_i)\}$ is a sequence of i.i.d. random vectors such that X_i has marginal distribution μ and Y_i has marginal distribution ν, then*

$$E\left[|\Phi(Q, N, \{C_r\}, \{X_r\}) - \Phi(Q, N, \{C_r\}, \{Y_r\})|^p\right] \leq H_p E[|X_1 - Y_1|^p].$$

Assumption 2.3 *Suppose that for any vector $(q, n, \{c_r\})$, with $n \in \mathbb{N} \cup \{\infty\}$, and any sequences of numbers $\{x_r\}$ and $\{y_r\}$ for which $\Phi(q, n, \{c_r\}, \{x_r\})$ and $\Phi(q, n, \{c_r\}, \{x_r\})$ are well defined, there exists a function $\varphi : \mathbb{R} \to \mathbb{R}_+$ such that*

$$|\Phi(q, n, \{c_r\}, \{x_r\}) - \Phi(q, n, \{c_r\}, \{y_r\})| \leq \sum_{r=1}^n \varphi(c_r)|x_r - y_r|.$$

Remarks 2.4 *(i) Note that Assumption 2.3 combined with Lemma 4.1 in [19] give that*

$$E\left[\left(\sum_{r=1}^{N}\varphi(C_r)|X_r-Y_r|\right)^p\right] \le E\left[\sum_{r=1}^{N}\varphi(C_r)^p|X_r-Y_r|^p\right] + E\left[\left(\sum_{r=1}^{N}\varphi(C_r)\right)^p\right]\left(E[|X_1-Y_1|^{p-1}]\right)^{p/(p-1)} \le 2E\left[\left(\sum_{r=1}^{N}\varphi(C_r)\right)^p\right]E\left[|X_1-Y_1|^p\right],$$

which implies that Assumption 2.2 holds with $H_p = 2E\left[\left(\sum_{r=1}^N \varphi(C_r)\right)^p\right]$*, provided the expectation is finite.*

(ii) Similarly, if $E\left[\left(\sum_{r=1}^N \varphi(C_r)\right)^p + |R^{(0)}|^p\right] < \infty$*, Assumption 2.3 combined with Lemma 4.1 in [19] yield*

$$\begin{aligned}||R^{(k)}||_p &\le \left|\left|\Phi(Q,N,\{C_r\},\{R_r^{(k-1)}\}) - \Phi(Q,N,\{C_r\},\{X_r\})\right|\right|_p + ||\Phi(Q,N,\{C_r\},\{X_r\})||_p \\ &\le \left|\left|\sum_{r=1}^{N}\varphi(C_r)|R_r^{(k-1)}-X_r|\right|\right|_p + ||\Phi(Q,N,\{C_r\},\{X_r\})||_p \\ &\le 2^{1/p}\left|\left|\sum_{r=1}^{N}\varphi(C_r)\right|\right|_p ||R_1^{(k-1)}-X_1||_p + ||\Phi(Q,N,\{C_r\},\{X_r\})||_p\end{aligned}$$

for any sequence $\{X_r\}$ *of i.i.d. random variables;* $||X||_p = (E[|X|^p])^{1/p}$ *denotes the* L_p *norm. In particular, choosing* $X_r \equiv 0$ *for all* $r \ge 1$ *gives*

$$||R^{(k)}||_p \le 2^{1/p}\left|\left|\sum_{r=1}^{N}\varphi(C_r)\right|\right|_p ||R^{(k-1)}||_p + ||R^{(0)}||_p,$$

which implies through induction that $E[|R^{(k)}|^p] < \infty$ *for all* $k \in \mathbb{N}$.

Before stating the main theorems establishing the convergence of the algorithm in the Wasserstein metric, we point out how Assumption 2.3 is satisfied by most of the SFPEs encountered in the literature, including all those mentioned in the introduction.

Examples 2.5

- *The linear SFPE* (1.2) *clearly satisfies Assumption 2.3 with* $\varphi(t) = |t|$.
- *Using the inequality*

$$\left|\max_{1\le i\le n}\{x_i\} - \max_{1\le i\le n}\{y_i\}\right| \le \max_{1\le i\le n}|x_i-y_i| \le \sum_{i=1}^{n}|x_i-y_i|$$

for any real numbers $\{x_i, y_i\}$ *and any* $n \ge 1$*, we obtain that both the maximum SFPE* (1.3) *and the discounted tree sum SFPE* (1.4) *satisfy Assumption 2.3 with* $\varphi(t) = |t|$ *as well.*

- *To see that* (1.5) *also satisfies Assumption 2.3 with* $\varphi(t) = |t|$ *(in this case* $C_i \equiv \tanh(\beta)$ *for all* $i \geq 1$*), let* $c = \tanh(\beta) \in [0, 1)$ *(since* $\beta \geq 0$*) and note that the function*

$$f(x) = \operatorname{arctanh}(c\tanh(x)) = \frac{1}{2}\ln\left(\frac{1 + c(e^{2x} - 1)/(e^{2x} + 1)}{1 - c(e^{2x} - 1)/(e^{2x} + 1)}\right) = \frac{1}{2}\ln\left(\frac{e^{2x}(1 + c) + 1 - c}{e^{2x}(1 - c) + 1 + c}\right)$$

has derivative

$$f'(x) = \frac{4c}{2(1 + c^2) + (e^{2x} + e^{-2x})(1 - c^2)} = \frac{2c}{1 + c^2 + \cosh(2x)(1 - c^2)},$$

and therefore satisfies

$$|f(x) - f(y)| = |f'(\xi)||x - y| \leq c|x - y|, \qquad \text{for some } \xi \text{ between } x \text{ and } y.$$

Our first result establishes the convergence in mean of $d_p(\hat{F}_{k,m}, F_k)$ under the "optimal" moment conditions, that is, assuming only that $\max_{0 \leq j \leq k} E[|R^{(j)}|^p] < \infty$. In view of Remark 2.4(ii), this is implied in all our examples by $E\left[\left(\sum_{i=1}^N \varphi(C_i)\right)^p\right] < \infty$. This result was previously proven in [10] for the linear SFPE (1.2) for $p = 1$.

Theorem 2.6 *Fix* $1 \leq p < \infty$ *and suppose that* Φ *satisfies Assumption 2.2 for* p*. Assume further that for any fixed* $k \in \mathbb{N}$, $\max_{0 \leq j \leq k} E[|R^{(j)}|^p] < \infty$*. Let* $\{R_1^{(j)}, \ldots, R_m^{(j)}\}$ *be an i.i.d. sample from distribution* F_j*, and let* $F_{j,m}$ *denote their corresponding empirical distribution function. Then,*

$$E\left[d_p(\hat{F}_{k,m}, F_k)^p\right] \leq \left(\sum_{r=0}^{k} (H_p^{1/p})^r\right)^{p-1} \sum_{j=0}^{k} (H_p^{1/p})^{k-j} E\left[d_p(F_{j,m}, F_j)^p\right],$$

where $0 < H_p < \infty$ *is the same from Assumption 2.2. Moreover, if* $\max_{0 \leq j \leq k} E[|R^{(j)}|^q] < \infty$ *for* $q > p \geq 1$, $q \neq 2p$*, then*

$$E\left[d_p(\hat{F}_{k,m}, F_k)^p\right] \leq m^{-\min\{(q-p)/q,\, 1/2\}} K \left(\sum_{r=0}^{k} (H_p^{1/p})^r\right)^{p-1} \sum_{j=0}^{k} (H_p^{1/p})^{k-j} (E[|R^{(j)}|^q])^{p/q},$$

where $K = K(p, q)$ *is a constant that only depends on* p *and* q*.*

Remarks 2.7 *(i) Note that Assumption 2.2 does not require that* $H_p < 1$*, i.e., it is not necessary for* Φ *to define a contraction for the algorithm to work. However, if* $H_p \geq 1$ *the bound given above may grow with the level of the recursion, i.e., the value of* k*, and the convergence of the sequence* $\{\mu_k\}$ *as* $k \to \infty$ *may not be guaranteed.*

(ii) We also point out that the first inequality in Theorem 2.6 implies that the rate at which $E\left[d_p(\hat{F}_{k,m}, F_k)^p\right]$ *converges to zero is determined by* $\max_{0 \leq j \leq k} E[d_p(F_{j,m}, F_j)]$*. Since* $d_p(F_{j,m}, F_j)$ *corresponds to implementing the population dynamics algorithm by sampling without replacement from a "perfect" i.i.d. pool of observations from* μ_{j-1}*, this convergence rate is in some sense optimal.*

We now turn our attention to the almost sure convergence of $d_p(\hat{F}_{k,m}, F_k)$, for which we provide two different results. The first one holds under Assumption 2.2 as above, but under rather strong moment conditions.

Theorem 2.8 *Fix $1 \leq p < \infty$ and suppose that Φ satisfies Assumption 2.2 for both p and $2p$. Assume further that for any fixed $k \in \mathbb{N}$, $\max_{0 \leq j \leq k} E[(R^{(j)})^{2p} (\log |R^{(j)}|)^+] < \infty$. Then,*

$$\lim_{m \to \infty} d_p(\hat{F}_{k,m}, F_k) = 0 \quad a.s.$$

The moment condition requiring the finiteness of the $2p$ absolute moment also appears in some related (stronger) results for the convergence of the Wasserstein distance between a distribution function and its empirical measure, specifically, concentration inequalities [15] and a central limit theorem [11]. In our case, where we seek only to establish the almost sure convergence of the algorithm, this condition is too strong, so we provide below an improved result under the finer Assumption 2.3.

Theorem 2.9 *Fix $1 \leq p < \infty$ and suppose that Φ satisfies Assumption 2.3. Assume further that $E[|R^{(0)}|^{p+\delta} + Z^{p+\delta}] < \infty$ for some $\delta > 0$, where $Z = \sum_{i=1}^N \varphi(C_i)$. Then, for any fixed $k \in \mathbb{N}$,*

$$\lim_{m \to \infty} d_p(\hat{F}_{k,m}, F_k) = 0 \qquad a.s.$$

Our last result relates the convergence of $d_p(\hat{F}_{k,m}, F_k)$ to the consistency of estimators based on the pool $\mathscr{P}^{(k,m)}$. More precisely, the value of the algorithm lies in the fact that it efficiently produces a sample of identically distributed random variables whose distribution is approximately F_k. A natural estimator for quantities of the form $E[h(R^{(k)})]$ is then given by

$$\frac{1}{m} \sum_{i=1}^m h(\hat{R}_i^{(k,m)}) = \int_{\mathbb{R}} h(x) d\hat{F}_{k,m}(x). \tag{2.2}$$

However, the random variables in $\mathscr{P}^{(k,m)}$ are not independent of each other, and the consistency of such estimators requires proof. In the sequel, the symbol $\stackrel{P}{\to}$ denotes convergence in probability.

Definition 2.10 *We say that Θ_n is a* weakly *consistent estimator for θ if $\Theta_n \stackrel{P}{\to} \theta$ as $n \to \infty$. We say that it is a* strongly *consistent estimator for θ if $\Theta_n \to \theta$ a.s.*

Our last result shows the consistency of estimators of the form in (2.2) for a broad class of functions.

Proposition 2.11 *Fix $1 \leq p < \infty$ and suppose that $h : \mathbb{R} \to \mathbb{R}$ satisfies $|h(x)| \leq C(1 + |x|^p)$ for all $x \in \mathbb{R}$ and some constant $C > 0$. Then, the following hold:*

a.) If $E[d_p(\hat{F}_{k,m}, F_k)^p] \to 0$ as $m \to \infty$, then (2.2) is a weakly consistent estimator for $E[h(R^{(k)})]$ for each fixed $k \in \mathbb{N}$.

b.) If $d_p(\hat{F}_{k,m}, F_k) \to 0$ a.s., as $m \to \infty$, then (2.2) is a strongly consistent estimator for $E[h(R^{(k)})]$ for each fixed $k \in \mathbb{N}$.

We conclude that the population dynamics algorithm can be used to efficiently generate sample pools of random variables having a distribution that closely approximates that of the special endogenous solution to SFPEs of the form in (1.1). Furthermore, these sample pools can be used to produce consistent estimators for a broad class of functions. The gain of efficiency of the algorithm compared to a naive Monte Carlo approach, combined with the consistency guarantees proved in this paper, make it extremely useful for the numerical analysis of many problems where SFPEs appear.

3 Proofs

This section includes the proofs of Theorems 2.6, 2.8, 2.9 and that of Proposition 2.11. The proofs are based on a construction of the pools $\{\mathscr{P}^{(j,m)} : 0 \leq j \leq k\}$ where we carefully couple the random variables $\{\hat{R}_i^{(j,m)}\}$ with i.i.d. observations from their limiting distribution F_j.

To start, for any $k \in \mathbb{N}$ let

$$\mathscr{E}_k = \left\{ \left(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}_{r\geq 1}, \{U_{(i,r)}^{(j)}\}_{r\geq 1} \right) : i \geq 1,\ 0 \leq j \leq k \right\} \tag{3.1}$$

be a collection of i.i.d. random vectors where $\left(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}_{r\geq 1} \right)$ has the same distribution as the generic branching vector $(Q, N, \{C_r\}_{i\geq 1})$ and the $\{U_{(i,r)}^{(j)}\}_{r\geq 1}$ are i.i.d. random variables uniformly distributed in $[0,1]$, independent of $\left(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}_{r\geq 1} \right)$. Next, we recursively construct a sequence of random variables $\{(\hat{R}_i^{(j,m)}, R_i^{(j)}) : 1 \leq i \leq m,\ 0 \leq j \leq k\}$ as follows:

i. Set $\hat{R}_i^{(0)} = F_0^{-1}(U_{(i,1)}^{(0)}) = R_i^{(0,m)}$, for $1 \leq i \leq m$; define

$$\hat{F}_{0,m}(x) = \frac{1}{m} \sum_{i=1}^{m} 1(\hat{R}_i^{(0,m)} \leq x) = F_{0,m}(x).$$

ii. For $1 \leq j \leq k$ and each $1 \leq i \leq m$,

$$\begin{aligned} \hat{R}_i^{(j,m)} &= \Phi \left(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}_{r\geq 1}, \{\hat{F}_{j-1,m}^{-1}(U_{(i,r)}^{(j)})\}_{r\geq 1} \right) \qquad \text{and} \\ R_i^{(j)} &= \Phi \left(Q_i^{(j)}, N_i^{(j)}, \{C_{(i,r)}^{(j)}\}_{r\geq 1}, \{F_{j-1}^{-1}(U_{(i,r)}^{(j)})\}_{r\geq 1} \right); \end{aligned}$$

define

$$\hat{F}_{j,m}(x) = \frac{1}{m} \sum_{i=1}^{m} 1(\hat{R}_i^{(j,m)} \leq x) \qquad \text{and} \qquad F_{j,m}(x) = \frac{1}{m} \sum_{i=1}^{m} 1(R_i^{(j)} \leq x).$$

Note that the random variables $\{R_i^{(j)}\}_{i=1}^m$ are i.i.d. and have distribution F_j, and therefore, $F_{j,m}$ is an empirical distribution function for F_j. The distribution functions $\hat{F}_{j,m}$ are those obtained through the population dynamics algorithm.

Throughout the proofs we will also use repeatedly the sigma-algebra $\mathcal{F}_k = \sigma(\mathscr{E}_k)$ for $k \in \mathbb{N}$. We point out that all the random variables $\{(\hat{R}_i^{(k,m)}, R_i^{(k)}) : i \geq 1\}$ are measurable with respect to $\mathcal{F}_k$ for all $m \geq 1$.

We are now ready to prove Theorem 2.6.

Proof of Theorem 2.6. Let $\{(\hat{R}_i^{(j,m)}, R_i^{(j)}) : 1 \leq i \leq m,\ 0 \leq j \leq k\}$ be a sequence of random vectors constructed as explained above.

Next, note that from the triangle inequality we obtain

$$d_p(\hat{F}_{j,m}, F_j) \leq d_p(\hat{F}_{j,m}, F_{j,m}) + d_p(F_{j,m}, F_j). \tag{3.2}$$

Now let χ be a Uniform(0,1) random variable independent of everything else, and define the random variables

$$\hat{R}^{(j,m)} = \sum_{i=1}^{m} \hat{R}_i^{(j,m)} 1((i-1)/m < \chi \leq i/m) \qquad \text{and} \qquad R^{(j)} = \sum_{i=1}^{m} R_i^{(j)} 1((i-1)/m < \chi \leq i/m),$$

which conditionally on $\mathcal{F}_j$ are distributed according to $\hat{F}_{j,m}$ and $F_{j,m}$, respectively. Then, from the definition of d_p we have

$$\begin{aligned} d_p(\hat{F}_{j,m}, F_{j,m})^p &= \inf_{X \sim \hat{F}_{j,m}, Y \sim F_{j,m}} E\left[|X - Y|^p \middle| \mathcal{F}_j \right] \\ &\leq E\left[\left| \hat{R}^{(j,m)} - R^{(j)} \right|^p \middle| \mathcal{F}_j \right] \\ &= \frac{1}{m} \sum_{i=1}^{m} \left| \hat{R}_i^{(j,m)} - R_i^{(j)} \right|^p. \end{aligned} \tag{3.3}$$

It follows from the observation that the random variables $X_i^{(j)} = \hat{R}_i^{(j,m)} - R_i^{(j)}$ are identically distributed, that

$$E\left[d_p(\hat{F}_{j,m}, F_{j,m})^p \right] \leq E\left[\left| \hat{R}_1^{(j,m)} - R_1^{(j)} \right|^p \right].$$

Next, note that by Assumption 2.2,

$$\begin{aligned} E\left[\left| \hat{R}_1^{(j,m)} - R_1^{(j)} \right|^p \right] &= E\Bigg[\Bigg| \Phi\left(Q_1^{(j)}, N_1^{(j)}, \{C_{(1,r)}^{(j)}\}_{r \geq 1}, \{\hat{F}_{j-1,m}^{-1}(U_{(1,r)}^{(j)})\}_{r \geq 1} \right) \\ &\qquad - \Phi\left(Q_1^{(j)}, N_1^{(j)}, \{C_{(1,r)}^{(j)}\}_{r \geq 1}, \{F_{j-1}^{-1}(U_{(1,r)}^{(j)})\}_{r \geq 1} \right) \Bigg|^p \Bigg] \\ &\leq H_p E\left[\left| \hat{F}_{j-1,m}^{-1}(U_{(1,1)}^{(j)}) - F_{j-1}^{-1}(U_{(1,1)}^{(j)}) \right|^p \right] \\ &= H_p E\left[d_p(\hat{F}_{j-1,m}, F_{j-1})^p \right]. \end{aligned}$$

It follows from (3.2) and Minkowski's inequality, that

$$\begin{aligned} \left(E\left[d_p(\hat{F}_{j,m}, F_j)^p \right] \right)^{1/p} &\leq \left(E\left[\left(d_p(\hat{F}_{j,m}, F_{j,m}) + d_p(F_{j,m}, F_j) \right)^p \right] \right)^{1/p} \\ &\leq \left(E\left[d_p(\hat{F}_{j,m}, F_{j,m})^p \right] \right)^{1/p} + \left(E\left[d_p(F_{j,m}, F_j)^p \right] \right)^{1/p} \\ &\leq \left(H_p E\left[d_p(\hat{F}_{j-1,m}, F_{j-1})^p \right] \right)^{1/p} + \left(E\left[d_p(F_{j,m}, F_j)^p \right] \right)^{1/p}. \end{aligned}$$

Iterating the recursion above we obtain

$$\left(E\left[d_p(\hat{F}_{j,m},F_j)^p\right]\right)^{1/p} \le \sum_{r=1}^{j}(H_p^{1/p})^{j-r}\left(E\left[d_p(F_{r,m},F_r)^p\right]\right)^{1/p} + (H_p^{1/p})^j\left(E\left[d_p(\hat{F}_{0,m},F_0)^p\right]\right)^{1/p}$$
$$= \sum_{r=0}^{j}(H_p^{1/p})^{j-r}\left(E\left[d_p(F_{r,m},F_r)^p\right]\right)^{1/p}.$$

Now let $\lambda_{j,r} = (H_p^{1/p})^{j-r}\left(\sum_{r=0}^{j}(H_p^{1/p})^{j-r}\right)^{-1}$ and use the fact that $g(x) = x^{1/p}$ is concave to obtain

$$\left(\sum_{r=0}^{j}(H_p^{1/p})^{j-r}\right)^{-1}\left(E\left[d_p(\hat{F}_{j,m},F_j)^p\right]\right)^{1/p} \le \sum_{r=0}^{j}\lambda_{j,r}\left(E\left[d_p(F_{r,m},F_r)^p\right]\right)^{1/p}$$
$$\le \left(\sum_{r=0}^{j}\lambda_{j,r}E\left[d_p(F_{r,m},F_r)^p\right]\right)^{1/p},$$

or equivalently,

$$E\left[d_p(\hat{F}_{j,m},F_j)^p\right] \le \left(\sum_{s=0}^{j}(H_p^{1/p})^s\right)^{p-1}\sum_{r=0}^{j}(H_p^{1/p})^{j-r}E\left[d_p(F_{r,m},F_r)^p\right].$$

This completes the first part of the proof.

Next, assume that $\max_{0\le r\le k}E[|R^{(r)}|^q] < \infty$ for $q > p \ge 1$, $q \ne 2p$, and use Theorem 1 in [15] to obtain that

$$E\left[d_p(F_{r,m},F_r)^p\right] \le C(E[|R^{(r)}|^q])^{p/q}\left(m^{-1/2} + m^{-(q-p)/q}\right),$$

where $C = C(p,q)$ is a constant that does not depend on F_r. The second statement of the theorem now follows. ■

We now turn to the proof of Theorem 2.8. To simplify its exposition we first provide a preliminary result for the mean Wasserstein distance between a distribution and its empirical distribution function.

Lemma 3.1 *Let G be a distribution on $\mathbb{R}$ and let $\{X_i\}_{i\ge1}$ be i.i.d. random variables distributed according to G. Suppose $E[|X_1|^q(\log|X_1|)^+] < \infty$ for some $q \ge 2$, and let $G_m(x) = m^{-1}\sum_{i=1}^{m}1(X_i \le x)$ denote the empirical distribution function of the $\{X_i\}$. Then,*

$$\sum_{m=1}^{\infty}\frac{1}{m}E[d_q(G_m,G)^q] < \infty.$$

Proof. Fix $\epsilon > 0$ and define for $x \ge 0$ the functions

$$a(x) = \min\{1/\overline{G}(x),\, x^{q+\epsilon}\} \qquad \text{and} \qquad b(x) = \min\{1/G(-x), x^{q+\epsilon}\}.$$

Next, use Proposition 7.14 in [8] followed by the monotonicity of the L_p norm, to see that

$$\begin{aligned}
\sum_{m=1}^{\infty} \frac{1}{m} E\left[d_q(G_m, G)^q\right] &\le q2^{q-1} \sum_{m=1}^{\infty} \frac{1}{m} \int_{-\infty}^{\infty} |x|^{q-1} E\left[|G_m(x) - G(x)|\right] dx \\
&\le q2^{q-1} \sum_{m=1}^{\infty} \frac{1}{m} \int_{-b^{-1}(m)}^{a^{-1}(m)} |x|^{q-1} \left(E\left[(G_m(x) - G(x))^2\right]\right)^{1/2} dx \\
&\quad + q2^{q-1} \sum_{m=1}^{\infty} \frac{1}{m} \int_{a^{-1}(m)}^{\infty} x^{q-1} E\left[\overline{G}_m(x) + \overline{G}(x)\right] dx \\
&\quad + q2^{q-1} \sum_{m=1}^{\infty} \frac{1}{m} \int_{-\infty}^{-b^{-1}(m)} |x|^{q-1} E\left[G_m(x) + G(x)\right] dx \\
&= q2^{q-1} \sum_{m=1}^{\infty} \frac{1}{m} \int_{-b^{-1}(m)}^{a^{-1}(m)} |x|^{q-1} \sqrt{\frac{G(x)\overline{G}(x)}{m}} dx \qquad (3.4) \\
&\quad + q2^{q} \sum_{m=1}^{\infty} \frac{1}{m} \int_{a^{-1}(m)}^{\infty} x^{q-1} \overline{G}(x)\, dx \qquad (3.5) \\
&\quad + q2^{q} \sum_{m=1}^{\infty} \frac{1}{m} \int_{-\infty}^{-b^{-1}(m)} |x|^{q-1} G(x)\, dx, \qquad (3.6)
\end{aligned}$$

where $g^{-1}(t) = \inf\{x \in \mathbb{R} : g(x) \ge t\}$ is the generalized inverse of function g.

Next, to bound (3.4) note that

$$\begin{aligned}
&\sum_{m=1}^{\infty} \frac{1}{m^{3/2}} \int_{-b^{-1}(m)}^{a^{-1}(m)} |x|^{q-1} \sqrt{G(x)\overline{G}(x)}\, dx \\
&\le \sum_{m=1}^{\infty} \frac{1}{m^{3/2}} \int_{0}^{a^{-1}(m)} x^{q-1} \sqrt{\overline{G}(x)}\, dx + \sum_{m=1}^{\infty} \frac{1}{m^{3/2}} \int_{-b^{-1}(m)}^{0} (-x)^{2p-1} \sqrt{G(x)}\, dx \\
&= \int_{0}^{\infty} \sum_{m=\lfloor a(x) \rfloor + 1}^{\infty} \frac{x^{q-1}}{m^{3/2}} \sqrt{\overline{G}(x)}\, dx + \int_{0}^{\infty} \sum_{m=\lfloor b(x) \rfloor + 1}^{\infty} \frac{x^{q-1}}{m^{3/2}} \sqrt{G(-x)}\, dx,
\end{aligned}$$

where in the last equality we used the observation that $\{x < a^{-1}(m)\} = \{a(x) < m\}$, respectively, $\{x < b^{-1}(m)\} = \{b(x) < m\}$. Now note that for any $n \ge 0$ we have

$$\begin{aligned}
\sum_{m=n+1}^{\infty} \frac{1}{m^{3/2}} &\le \sum_{m=n+1}^{\infty} \left(\frac{m+1}{m}\right)^{3/2} \int_{m}^{m+1} \frac{1}{t^{3/2}}\, dt \le \left(1 + \frac{1}{n+1}\right)^{3/2} \int_{n+1}^{\infty} t^{-3/2}\, dt \\
&\le 2^{5/2} (n+1)^{-1/2}.
\end{aligned}$$

Hence, (3.4) is bounded from above by a constant times

$$\begin{aligned}
&\int_0^\infty x^{q-1}\sqrt{\overline{G}(x)}(\lfloor a(x)\rfloor+1)^{-1/2}\,dx+\int_0^\infty x^{q-1}\sqrt{G(-x)}(\lfloor b(x)\rfloor+1)^{-1/2}\,dx\\
&\le 2+\int_1^\infty x^{q-1}\sqrt{\frac{\overline{G}(x)}{a(x)}}\,dx+\int_1^\infty x^{q-1}\sqrt{\frac{G(-x)}{b(x)}}\,dx\\
&=2+\int_{\{x\ge 1:1/\overline{G}(x)\le x^{q+\epsilon}\}} x^{q-1}\overline{G}(x)\,dx+\int_{\{x\ge 1:1/\overline{G}(x)> x^{q+\epsilon}\}} x^{q/2-1-\epsilon/2}\sqrt{\overline{G}(x)}\,dx\\
&\quad+\int_{\{x\ge 1:1/G(-x)\le x^{q+\epsilon}\}} x^{q-1}G(-x)\,dx+\int_{\{x\ge 1:1/G(-x)> x^{q+\epsilon}\}} x^{q/2-1-\epsilon/2}\sqrt{G(-x)}\,dx\\
&\le 2+\int_1^\infty x^{q-1}\overline{G}(x)\,dx+\int_1^\infty x^{q-1}G(-x)\,dx+2\int_1^\infty x^{-1-\epsilon}\,dx\\
&\le 2+\frac{1}{q}\int_1^\infty x^qG(dx)+\frac{1}{q}\int_{-\infty}^{-1}(-x)^qG(dx)+\frac{2}{\epsilon}\\
&\le 2+\frac{1}{q}E[|X_1|^q]+\frac{2}{\epsilon}<\infty.
\end{aligned}$$

To analyze (3.5) use the observation that $\{x\ge a^{-1}(m)\}=\{a(x)\ge m\}$ to obtain that

$$\begin{aligned}
\sum_{m=1}^\infty\frac{1}{m}\int_{a^{-1}(m)}^\infty x^{q-1}\overline{G}(x)\,dx&=\int_{a^{-1}(1)}^\infty\sum_{m=1}^{\lfloor a(x)\rfloor}\frac{1}{m}\,x^{q-1}\overline{G}(x)\,dx\\
&\le\int_{a^{-1}(1)}^\infty x^{q-1}\overline{G}(x)\sum_{m=1}^{\lfloor a(x)\rfloor}\frac{m+1}{m}\int_m^{m+1}\frac{1}{t}\,dt\,dx\\
&\le 2\int_{a^{-1}(1)}^\infty x^{q-1}\overline{G}(x)\int_1^{\lfloor a(x)\rfloor+1}t^{-1}\,dt\,dx\\
&\le 2\int_{a^{-1}(1)}^\infty x^{q-1}\overline{G}(x)\log(x^{q+\epsilon}+1)\,dx\\
&\le 2\log 2+2(q+\epsilon)\sup_{t\ge 1}\frac{\log(t+1)}{\log t}\int_1^\infty x^{q-1}(\log x)\overline{G}(x)\,dx.
\end{aligned}$$

Since $\sup_{t\ge1}\log(t+1)/\log t<\infty$ and

$$\begin{aligned}
\int_1^\infty x^{q-1}(\log x)\overline{G}(x)\,dx&=\left.\frac{x^q(\log x-1)}{q}\overline{G}(x)\right|_1^\infty+\int_1^\infty\frac{x^q(\log x-1)}{q}G(dx)\\
&=\frac{\overline{G}(1)}{q}+\frac{E[|X_1|^q(\log X_1-1)1(X_1\ge 1)]}{q}\\
&\le\frac{E[|X_1|^q\log X_1 1(X_1\ge 1)]}{q}<\infty,
\end{aligned}$$

we obtain that (3.5) is finite. Finally, the same steps used to bound (3.5) give that (3.6) is bounded by

$$q2^q\left(2\log 2+2(q+\epsilon)\frac{E[|X_1|^q\log|X_1|1(X_1\le -1)]}{q}\sup_{t\ge 1}\frac{\log(t+1)}{\log t}\right)<\infty.$$

■

Proof of Theorem 2.8. We will start the proof by deriving an upper bound for $d_p(\hat{F}_{k,m}, F_k)$. To this end, we construct the random variables $\{(\hat{R}_i^{(j,m)}, R_i^{(j)}) : 1 \leq i \leq m,\, 0 \leq j \leq k\}$ according to the construction given at the beginning of the section. Recall that $\mathcal{F}_j = \sigma(\mathscr{E}_j)$, where $\mathscr{E}_j$ is given by (3.1), and that Assumption 2.2 holds for both p and $2p$.

We start by noting that the triangle inequality followed by (3.3) give

$$d_p(\hat{F}_{k,m}, F_k) \leq d_p(\hat{F}_{k,m}, F_{k,m}) + d_p(F_{k,m}, F_k) \leq \frac{1}{m}\sum_{i=1}^m \left|\hat{R}_i^{(k,m)} - R_i^{(k)}\right|^p + d_p(F_{k,m}, F_k).$$

Next, define for $j \geq 1$, $X_i^{(j,m)} = \left|\hat{R}_i^{(j,m)} - R_i^{(j)}\right|^p$ and note that by construction, the random variables $\{X_i^{(j,m)}\}_{i\geq 1}$ are identically distributed and conditionally independent given $\mathcal{F}_{j-1}$. Now set $Z_i^{(j,m)} = X_i^{(j,m)} - E[X_1^{(j,m)}|\mathcal{F}_{j-1}]$ and note that

$$\begin{aligned}
E[X_1^{(j,m)}|\mathcal{F}_{j-1}] &= E\left[\left|\Phi\left(Q_1^{(j)}, N_1^{(j)}, \{C_{(1,r)}^{(j)}\}_{r\geq 1}, \{\hat{F}_{j-1,m}^{-1}(U_{(1,r)}^{(j)})\}_{r\geq 1}\right)\right.\right. \\
&\qquad \left.\left. -\Phi\left(Q_1^{(j)}, N_1^{(j)}, \{C_{(1,r)}^{(j)}\}_{r\geq 1}, \{F_{j-1}^{-1}(U_{(1,r)}^{(j)})\}_{r\geq 1}\right)\right|^p \middle| \mathcal{F}_{j-1}\right] \\
&\leq H_p E\left[\left|\hat{F}_{j-1,m}^{-1}(U_{(1,1)}^{(j)}) - F_{j-1}^{-1}(U_{(1,1)}^{(j)})\right|^p \middle| \mathcal{F}_{j-1}\right] \qquad \text{(by Assumption 2.2)} \\
&= H_p d_p(\hat{F}_{j-1,m}, F_{j-1})^p.
\end{aligned}$$

It follows that

$$d_p(\hat{F}_{k,m}, F_{k,m})^p \leq \frac{1}{m}\sum_{i=1}^m Z_i^{(k,m)} + H_p d_p(\hat{F}_{k-1,m}, F_{k-1})^p,$$

which in turn implies that

$$\begin{aligned}
d_p(\hat{F}_{k,m}, F_k) &\leq d_p(F_{k,m}, F_k) + \left(\frac{1}{m}\sum_{i=1}^m Z_i^{(k,m)} + H_p d_p(\hat{F}_{k-1,m}, F_{k-1})^p\right)^{1/p} \\
&\leq d_p(F_{k,m}, F_k) + \left(\frac{1}{m}\sum_{i=1}^m Z_i^{(k,m)}\right)^{1/p} + H_p^{1/p} d_p(\hat{F}_{k-1,m}, F_{k-1}),
\end{aligned}$$

where in the last step we used the inequality $(x+y)^\beta \leq x^\beta + y^\beta$ for $0 < \beta \leq 1$ and $x, y \geq 0$. Iterating $k-1$ more times we obtain

$$\begin{aligned}
d_p(\hat{F}_{k,m}, F_k) &\leq \sum_{j=1}^k \left(d_p(F_{j,m}, F_j) + \left(\frac{1}{m}\sum_{i=1}^m Z_i^{(j,m)}\right)^{1/p}\right)(H_p^{1/p})^{k-j} + (H_p^{1/p})^k d_p(\hat{F}_{0,m}, F_0) \\
&= \sum_{j=0}^k (H_p^{1/p})^{k-j} d_p(F_{j,m}, F_j) + \sum_{j=1}^k (H_p^{1/p})^{k-j}\left(\frac{1}{m}\sum_{i=1}^m Z_i^{(j,m)}\right)^{1/p}.
\end{aligned}$$

Now note that by the Glivenko-Cantelli lemma and the strong law of large numbers,

$$\sup_{x\in\mathbb{R}} |F_{j,m}(x) - F_j(x)| \to 0 \qquad \text{a.s.} \qquad \text{and}$$
$$\frac{1}{m}\sum_{i=1}^{m} |R_i^{(j)}|^p = \int_{-\infty}^{\infty} |x|^p dF_{j,m}(x) \to \int_{-\infty}^{\infty} |x|^p dF_j(x) \quad \text{a.s.},$$

as $m \to \infty$, and therefore, by Theorem 6.8 in [27], $d_p(F_{j,m}, F_j) \to 0$ a.s. for each $j \geq 1$. It suffices then to show that for each $1 \leq j \leq k$ the sums $m^{-1}\sum_{i=1}^m Z_i^{(j,m)} \to 0$ a.s. as well.

To see this note that for any $\epsilon > 0$,

$$\begin{aligned}
\sum_{m=1}^{\infty} P\left(\frac{1}{m}\sum_{i=1}^{m} Z_i^{(j,m)} > \epsilon\right) &\leq \sum_{m=1}^{\infty} \frac{1}{\epsilon^2 m^2} E\left[\left(\sum_{i=1}^{m} Z_i^{(j,m)}\right)^2\right] \\
&= \frac{1}{\epsilon^2}\sum_{m=1}^{\infty}\frac{1}{m}\left(E\left[\left(Z_1^{(j,m)}\right)^2\right] + (m-1)E\left[Z_1^{(j,m)}Z_2^{(j,m)}\right]\right) \\
&= \frac{1}{\epsilon^2}\sum_{m=1}^{\infty}\frac{1}{m}E\left[\mathrm{Var}(X_1^{(j,m)}|\mathcal{F}_{j-1})\right].
\end{aligned}$$

Moreover, using the same arguments we used in the proof of Theorem 2.6, we obtain that

$$\begin{aligned}
&\mathrm{Var}(X_1^{(j,m)}|\mathcal{F}_{j-1}) \\
&\leq E\left[(X_1^{(j,m)})^2 \middle| \mathcal{F}_{j-1}\right] \\
&= E\left[\left(\Phi\left(Q_1^{(j)}, N_1^{(j)}, \{C_{(1,r)}^{(j)}\}_{r\geq 1}, \{\hat{F}_{j-1,m}^{-1}(U_{(1,r)}^{(j)})\}_{r\geq 1}\right)\right.\right. \\
&\qquad\qquad \left.\left. -\Phi\left(Q_1^{(j)}, N_1^{(j)}, \{C_{(1,r)}^{(j)}\}_{r\geq 1}, \{F_{j-1}^{-1}(U_{(1,r)}^{(j)})\}_{r\geq 1}\right)\right)^{2p} \middle| \mathcal{F}_{j-1}\right] \\
&\leq H_{2p}\, E\left[\left(\hat{F}_{j-1,m}^{-1}(U_{(1,1)}^{(j)}) - F_{j-1}^{-1}(U_{(1,1)}^{(j)})\right)^{2p} \middle| \mathcal{F}_{j-1}\right] \qquad \text{(by Assumption 2.2)} \\
&= H_{2p}\, d_{2p}(\hat{F}_{j-1,m}, F_{j-1})^{2p}.
\end{aligned}$$

Next, note that by Theorem 2.6 we have

$$E\left[d_{2p}(\hat{F}_{j-1,m}, F_{j-1})^{2p}\right] \leq \left(\sum_{s=0}^{j-1} H_{2p}^s\right)^{2p-1} \sum_{r=0}^{j-1} H_{2p}^{j-1-r} E\left[d_{2p}(F_{r,m}, F_r)^{2p}\right].$$

It follows that for any $1 \leq j \leq k$,

$$\begin{aligned}
\sum_{m=1}^{\infty} P\left(\frac{1}{m}\sum_{i=1}^{m} Z_i^{(j,m)} > \epsilon\right) &\leq \frac{H_{2p}}{\epsilon^2}\sum_{m=1}^{\infty}\frac{1}{m}E\left[d_{2p}(\hat{F}_{j-1,m}, F_{j-1})^{2p}\right] \\
&\leq \frac{H_{2p}}{\epsilon^2}\left(\sum_{s=0}^{j-1} H_{2p}^s\right)^{2p-1} \sum_{r=0}^{j-1} H_{2p}^{j-1-r} \sum_{m=1}^{\infty}\frac{1}{m}E\left[d_{2p}(F_{r,m}, F_r)^{2p}\right].
\end{aligned}$$

Finally, since by Lemma 3.1 we have that

$$\sum_{m=1}^{\infty} \frac{1}{m} E\left[d_{2p}(F_{r,m}, F_r)^{2p}\right] < \infty$$

for each $0 \le r \le j-1$, the Borel-Cantelli Lemma gives that $\lim_{m\to\infty} m^{-1}\sum_{i=1}^m Z_i^{(j,m)} = 0$ a.s. This completes the proof. ■

We now move on to the proof of Theorem 2.9, for which we first state and prove three preliminary results. The first one provides an upper bound for the generalized inverse of any distribution function having finite q absolute moments.

Lemma 3.2 *Let G be a distribution function on $\mathbb{R}$, and let G^{-1} be its generalized inverse. Suppose that G has finite absolute moments of order $q > 0$. Then, for any $u \in (0,1)$,*

$$|G^{-1}(u)| \le ||X^+||_q (1-u)^{-1/q} + ||X^-||_q u^{-1/q}.$$

Proof. Let X be a random variable having distribution G, and define $G_+(x) = P(X^+ \le x) = G(x)1(x \ge 0)$ and $G_-(x) = P(X^- \le x) = P(X \ge -x)1(x \ge 0)$. Then,

$$G_+^{-1}(u) = \inf\{x \in \mathbb{R} : G_+(x) \ge u\} = \inf\{x \ge 0 : G(x) \ge u\} = G^{-1}(u)^+,$$

while if we define G_-^* to be the right-continuous generalized inverse of G_-, then

$$\begin{aligned} G_-^*(1-u) &= \inf\{x \in \mathbb{R} : G_-(x) > 1-u\} = \inf\{x \ge 0 : 1 - G(-x) + P(X = -x) > 1-u\} \\ &= \inf\{x \ge 0 : G(-x) - P(X = -x) < u\} = -\inf\{x \le 0 : G(x) \ge u\} = G^{-1}(u)^-. \end{aligned}$$

Now use Markov's inequality to obtain that for all $x > 0$,

$$1 - G_+(x) \le \min\{1, E[(X^+)^q]\} x^{-q} \triangleq 1 - H_+(x)$$

and

$$1 - G_-(x) \le \min\{1, E[(X^-)^q]\} x^{-q} \triangleq 1 - H_-(x).$$

These first inequality implies that for any $u \in (0,1)$,

$$G_+^{-1}(u) = \inf\{x \in \mathbb{R} : G_+(x) \ge u\} \le \inf\{x \in \mathbb{R} : H_+(x) \ge u\} = H_+^{-1}(u) = ||X^+||_q (1-u)^{-1/q},$$

while the second one plus the continuity of H_- gives

$$\begin{aligned} G^{-1}(u)^- = G_-^*(1-u) &= \inf\{x \in \mathbb{R} : G_-(x) > 1-u\} \le \inf\{x \in \mathbb{R} : H_-(x) > 1-u\} \\ &= \inf\{x \in \mathbb{R} : H_-(x) \ge 1-u\} = H_-^{-1}(1-u) = ||X^-||_q u^{-1/q}. \end{aligned}$$

It follows that

$$|G^{-1}(u)| = G^{-1}(u)^+ + G^{-1}(u)^- \le ||X^+||_q (1-u)^{-1/q} + ||X^-||_q u^{-1/q}.$$

■

The next two preliminary results provide key steps for the proof of Theorem 2.9.

Lemma 3.3 *Fix $1 \leq p < \infty$ and $\epsilon > 0$. Suppose Assumption 2.3 holds and $E[|R^{(0)}|^{p+\delta} + Z^{p+\delta}] < \infty$ for some $\delta > 0$, where $Z = \sum_{i=1}^N \varphi(C_i)$. Let $\mathcal{F}_j = \sigma(\mathscr{E}_j)$, where $\mathscr{E}_j$ is defined by (3.1), set $\delta_j = \delta(k-j)/k$, $0 \leq j \leq k$, $\eta = \left(\epsilon^{-1} 4e^{2/\epsilon} \max\{1, E[Z^{p+\delta}]\}\right)^{-(p+\delta_j)/(p+\delta_{j+1})}$, and*

$$Y_i^{(j,m)} = \left(\sum_{r=1}^{N_i^{(j+1)}} \varphi(C_{(i,r)}^{(j+1)}) \left| \hat{F}_{j,m}^{-1}(U_{(i,r)}^{(j+1)}) - F_j^{-1}(U_{(i,r)}^{(j+1)}) \right| \right)^{p+\delta_{j+1}}, \qquad i = 1, \ldots, m. \tag{3.7}$$

Then, on the event $\left\{ \sup_{m \geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta \right\}$, we have

$$P\left(\sup_{m \geq n} \frac{1}{m} \sum_{i=1}^m Y_i^{(j,m)} 1(Y_i^{(j,m)} \leq m/\log m) > \epsilon \,\middle|\, \mathcal{F}_j \right) \leq 2(n-1)^{-1/2}.$$

Proof. We start by noting that

$$\begin{aligned} &P\left(\sup_{m \geq n} \frac{1}{m} \sum_{i=1}^m Y_i^{(j,m)} 1(Y_i^{(j,m)} \leq m/\log m) > \epsilon \,\middle|\, \mathcal{F}_j \right) \\ &\leq \sum_{m=n}^{\infty} P\left(\frac{1}{m} \sum_{i=1}^m Y_i^{(j,m)} 1(Y_i^{(j,m)} \leq m/\log m) > \epsilon \,\middle|\, \mathcal{F}_j \right). \end{aligned} \tag{3.8}$$

To bound each of the probabilities in (3.13) use Chernoff's bound to obtain that

$$\begin{aligned} &P\left(\frac{1}{m} \sum_{i=1}^m Y_i^{(j,m)} 1(Y_i^{(j,m)} \leq m/\log m) > \epsilon \,\middle|\, \mathcal{F}_j \right) \\ &\leq \min_{\theta \geq 0} e^{-\theta\epsilon m} \left(E\left[e^{\theta Y_1^{(j,m)} 1(Y_1^{(j,m)} \leq m/\log m)} \,\middle|\, \mathcal{F}_j \right] \right)^m. \end{aligned}$$

Note that by Remark 2.4(i), we have that on the event $\left\{ \sup_{m \geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta \right\}$,

$$\begin{aligned} E\left[Y_1^{(j,m)} \,\middle|\, \mathcal{F}_j \right] &\leq 2E[Z^{p+\delta_{j+1}}] E\left[\left| \hat{F}_{j,m}^{-1}(U_1) - F_j^{-1}(U_1) \right|^{p+\delta_{j+1}} \,\middle|\, \mathcal{F}_j \right] \\ &= ||Z||_{p+\delta_{j+1}}^{p+\delta_{j+1}} d_{p+\delta_{j+1}}(\hat{F}_{j,m}, F_j)^{p+\delta_{j+1}} \\ &\leq ||Z||_{p+\delta}^{p+\delta_{j+1}} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_{j+1}} \\ &\leq \max\{1, E[Z^{p+\delta}]\} \eta^{(p+\delta_{j+1})/(p+\delta_j)} = \frac{\epsilon}{4e^{2/\epsilon}}. \end{aligned}$$

Next, use the inequality $e^x \leq 1 + xe^x$ for $x \geq 0$ to obtain that

$$\begin{aligned} &E\left[e^{\theta Y_1^{(j,m)} 1(Y_1^{(j,m)} \leq m/\log m)} \,\middle|\, \mathcal{F}_j \right] \\ &\leq 1 + \theta E\left[Y_1^{(j,m)} 1(Y_1^{(j,m)} \leq m/\log m) \, e^{\theta Y_1^{(j,m)} 1(Y_1^{(j,m)} \leq m/\log m)} \,\middle|\, \mathcal{F}_j \right] \\ &\leq 1 + \theta E\left[Y_1^{(j,m)} \,\middle|\, \mathcal{F}_j \right] e^{\theta m/\log m} \\ &\leq 1 + \theta e^{\theta m/\log m} \frac{\epsilon}{4e^{2/\epsilon}}. \end{aligned}$$

Now use the inequality $1 + x \leq e^x$ to see that

$$\left(E \left[e^{\theta Y_1^{(j,m)} 1(Y_1^{(j,m)} \leq m/\log m)} \middle| \mathcal{F}_j \right] \right)^m \leq e^{\theta \epsilon m e^{\theta m/\log m}/(4e^{2/\epsilon})}.$$

It follows that by choosing $\theta = (2/\epsilon)\log m/m$ we obtain

$$\begin{aligned} P\left(\frac{1}{m} \sum_{i=1}^m Y_i^{(j,m)} 1(Y_i^{(j,m)} \leq m/\log m) > \epsilon \middle| \mathcal{F}_j \right) &\leq \min_{\theta \geq 0} e^{-\theta\epsilon m + \theta \epsilon m e^{\theta m/\log m}/(4e^{2/\epsilon})} \\ &= \min_{\theta \geq 0} e^{-\theta \epsilon m \left(1 - \frac{e^{\theta m/\log m}}{4e^{2/\epsilon}}\right)} \\ &\leq e^{-2\log m \left(1 - \frac{1}{4}\right)}, \end{aligned}$$

which in turn implies that (3.13) is bounded from above by

$$\sum_{m=n}^\infty e^{-(3/2)\log m} = \sum_{m=n}^\infty m^{-3/2} \leq \sum_{m=n}^\infty \int_{m-1}^m \frac{1}{x^{3/2}}\, dx = \int_{n-1}^\infty x^{-3/2}\, dx = 2(n-1)^{-1/2}.$$

This completes the proof. ■

Lemma 3.4 *Fix $1 \leq p < \infty$. Suppose Assumption 2.3 holds and $E[Z^{p+\delta}] < \infty$ for some $\delta > 0$, where $Z = \sum_{i=1}^N \varphi(C_i)$. Let $\delta_j = \delta(k-j)/k$ and $q_j = p + \delta_j$ for $0 \leq j < k$, fix $\eta > 0$, and let $Y_1^{(j,m)}$ be defined according to (3.7). Then, for any $q_{j+1} < r_j < q_j$ and all $t \geq n$,*

$$\begin{aligned} &P\left(\sup_{m \geq t} \frac{\log m}{m} Y_1^{(j,m)} > 1, \sup_{m \geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta \right) \\ &\leq 3^{r_j} ||Z||_{r_j}^{r_j} \left\{ \frac{4(\eta^{1/q_j} + ||R^{(j)}||_{q_j})^{r_j}}{1 - r_j/q_j} + 2||R^{(j)}||_{r_j}^{r_j} \right\} \left(\frac{\log t}{t} \right)^{r_j/q_{j+1}}. \end{aligned}$$

Proof. To simplify the notation, let

$$(Q, N, \{C_r\}_{r \geq 1}, \{U_r\}_{r \geq 1}) = \left(Q_1^{(j+1)}, N_1^{(j+1)}, \{C_{(1,r)}^{(j+1)}\}_{r \geq 1}, \{U_{(1,r)}^{(j+1)}\}_{r \geq 1} \right).$$

Next, note that

$$\begin{aligned} \sup_{m \geq t} \frac{\log m}{m} Y_1^{(j,m)} &= \sup_{m \geq t} \frac{\log m}{m} \left(\sum_{r=1}^N \varphi(C_r) \left| \hat{F}_{j,m}^{-1}(U_r) - F_j^{-1}(U_r) \right| \right)^{p+\delta_{j+1}} \\ &\leq \left(\sum_{r=1}^N \varphi(C_r) \sup_{m \geq t} \left(\frac{\log m}{m} \right)^{1/(p+\delta_{j+1})} \left| \hat{F}_{j,m}^{-1}(U_r) - F_j^{-1}(U_r) \right| \right)^{p+\delta_{j+1}} \\ &= \left(\sum_{r=1}^N \varphi(C_r) W_r^{(j,t)} \right)^{p+\delta_{j+1}}, \end{aligned}$$

where

$$W_r^{(j,t)} = \sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/(p+\delta_{j+1})}\left|\hat{F}_{j,m}^{-1}(U_r) - F_j^{-1}(U_r)\right|.$$

Now, let $\mathcal{F}_j = \sigma(\mathscr{E}_j)$, where $\mathscr{E}_j$ is given by (3.1), and note that

$$\begin{aligned}
&P\left(\sup_{m\geq t}\frac{\log m}{m}Y_1^{(j,m)} > 1,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta\right)\\
&\leq P\left(\sum_{r=1}^{N}\varphi(C_r)W_r^{(j,t)} > 1,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta\right)\\
&= E\left[P\left(\sum_{r=1}^{N}\varphi(C_r)W_r^{(j,t)} > 1\middle|\mathcal{F}_j\right)1\left(\sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta\right)\right].
\end{aligned}$$

Moreover, if we let $q_j = p + \delta_j$ and use Lemma 3.2, we obtain that, conditionally on $\mathcal{F}_j$,

$$\begin{aligned}
W_r^{(j,t)} &\leq \sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/q_{j+1}}\left|\hat{F}_{j,m}^{-1}(U_r)\right| + \sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/q_{j+1}}|F_j^{-1}(U_r)|\\
&\leq \sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/q_{j+1}}\left(E\left[|\hat{F}_{j,m}^{-1}(U_r)|^{q_j}\middle|\mathcal{F}_j\right]\right)^{1/q_j}\left\{U_r^{-1/q_j} + (1-U_r)^{-1/q_j}\right\}\\
&\quad + \left(\frac{\log t}{t}\right)^{1/q_{j+1}}|F_j^{-1}(U_r)|.
\end{aligned}$$

Furthermore, by Minkowski's inequality, we have that on the event $\{\sup_{m\geq n} d_{q_j}(\hat{F}_{j,m}, F_j)^{q_j} \leq \eta\}$,

$$\begin{aligned}
&\sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/q_{j+1}}\left(E\left[|\hat{F}_{j,m}^{-1}(U_r)|^{q_j}\middle|\mathcal{F}_j\right]\right)^{1/q_j}\\
&\leq \sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/q_{j+1}}\left\{\left(E\left[|\hat{F}_{j,m}^{-1}(U_r) - F_j^{-1}(U_r)|^{q_j}\middle|\mathcal{F}_j\right]\right)^{1/q_j} + ||F_j^{-1}(U_r)||_{q_j}\right\}\\
&= \sup_{m\geq t}\left(\frac{\log m}{m}\right)^{1/q_{j+1}}\left\{d_{q_j}(\hat{F}_{j,m}, F_j) + ||R^{(j)}||_{q_j}\right\}\\
&\leq \left(\frac{\log t}{t}\right)^{1/q_{j+1}}\left\{\eta^{1/q_j} + ||R^{(j)}||_{q_j}\right\}.
\end{aligned}$$

It follows that conditionally on $\mathcal{F}_j$, we have that on the event $\{\sup_{m\geq n} d_{q_j}(\hat{F}_{j,m}, F_j)^{q_j} \leq \eta\}$,

$$W_r^{(j,t)} \leq \left(\frac{\log t}{t}\right)^{1/q_{j+1}}\left\{K_j\left(U_r^{-1/q_j} + (1-U_r)^{-1/q_j}\right) + |F_j^{-1}(U_r)|\right\},$$

where $K_j \triangleq \eta^{1/q_j} + ||R^{(j)}||_{q_j} < \infty$ by Remark 2.4(ii).

Thus, we have that on the event $\{\sup_{m\geq n} d_{q_j}(\hat{F}_{j,m}, F_j)^{q_j} \leq \eta\}$, the union bound and Markov's inequality yield

$$
\begin{aligned}
&P\left(\sum_{r=1}^{N}\varphi(C_r)W_r^{(j,t)} > 1 \,\middle|\, \mathcal{F}_j\right) \\
&\leq P\left(\sum_{r=1}^{N}\varphi(C_r)\left\{K_j\left(U_r^{-1/q_j} + (1-U_r)^{-1/q_j}\right) + |F_j^{-1}(U_r)|\right\} > \left(\frac{t}{\log t}\right)^{1/q_{j+1}}\right) \\
&\leq P\left(\sum_{r=1}^{N}\varphi(C_r)K_jU_r^{-1/q_j} > \frac{1}{3}\left(\frac{t}{\log t}\right)^{1/q_{j+1}}\right) \\
&\quad + P\left(\sum_{r=1}^{N}\varphi(C_r)K_j(1-U_r)^{-1/q_j} > \frac{1}{3}\left(\frac{t}{\log t}\right)^{1/q_{j+1}}\right) \\
&\quad + P\left(\sum_{r=1}^{N}\varphi(C_r)|F_j^{-1}(U_r)| > \frac{1}{3}\left(\frac{t}{\log t}\right)^{1/q_{j+1}}\right) \\
&\leq 3^{r_j}\left(\frac{\log t}{t}\right)^{r_j/q_{j+1}}\left\{2E\left[\left(\sum_{i=1}^{N}\varphi(C_i)K_jU_i^{-1/q_j}\right)^{r_j}\right] + E\left[\left(\sum_{i=1}^{N}\varphi(C_i)R_i^{(j)}\right)^{r_j}\right]\right\},
\end{aligned}
$$

where by assumption $q_{j+1} < r_j < q_j$, and we have used the observation that $U_i \stackrel{\mathcal{D}}{=} 1 - U_i$. Finally, note that by Remark 2.4(i), we have

$$
E\left[\left(\sum_{i=1}^{N}\varphi(C_i)K_jU_i^{-1/q_j}\right)^{r_j}\right] \leq 2E[Z^{r_j}]K_j^{r_j}E[U_1^{-r_j/q_j}] = \frac{2K_j^{r_j}||Z||_{r_j}^{r_j}}{1 - r_j/q_j}
$$

and

$$
E\left[\left(\sum_{i=1}^{N}\varphi(C_i)R_i^{(j)}\right)^{r_j}\right] \leq 2E[Z^{r_j}]E\left[|R^{(j)}|^{r_j}\right] = 2||Z||_{r_j}^{r_j}||R^{(j)}||_{r_j}^{r_j}.
$$

We conclude that

$$
\begin{aligned}
&P\left(\sup_{m\geq t}\frac{\log m}{m}X_1^{(m)} > 1,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \leq \eta\right) \\
&\leq 3^{r_j}\left(\frac{\log t}{t}\right)^{r_j/q_{j+1}}\left\{\frac{4K_j^{r_j}||Z||_{r_j}^{r_j}}{1 - r_j/q_j} + 2||Z||_{r_j}^{r_j}||R^{(j)}||_{r_j}^{r_j}\right\}.
\end{aligned}
$$

■

We are now ready to prove Theorem 2.9.

Proof of Theorem 2.9. Define $\delta_j = \delta(k-j)/k$ for $0 \leq j \leq k$. We will prove by induction in j that

$$
\lim_{m\to\infty} d_{p+\delta_j}(\hat{F}_{j,m}, F_j) = 0 \qquad \text{a.s} \tag{3.9}
$$

for $0 \le j \le k$. Since $\hat{F}_{0,m}(x) \equiv F_{0,m}(x)$ for all $x \in \mathbb{R}$ and $E[|R_0|^{p+\delta}] < \infty$, the Glivenko-Cantelli lemma and the strong law of large numbers yield

$$\sup_{x\in\mathbb{R}} |F_{0,m}(x) - F_0(x)| \to 0 \qquad \text{a.s. as } m \to \infty \qquad \text{and}$$
$$\frac{1}{m}\sum_{i=1}^{m} |R_i^{(0)}|^{p+\delta} = \int_{-\infty}^{\infty} |x|^{p+\delta} dF_{0,m}(x) \to \int_{-\infty}^{\infty} |x|^{p+\delta} dF_0(x) \quad \text{a.s. as } m \to \infty.$$

Therefore, by Theorem 6.8 in [27],

$$\lim_{m\to\infty} d_{p+\delta_0}(\hat{F}_{0,m}, F_0) = \lim_{m\to\infty} d_{p+\delta}(F_{0,m}, F_0) = 0 \quad \text{a.s.}$$

Suppose now that (3.9) holds for $0 \le j < k$. To prove that $d_{p+\delta_{j+1}}(\hat{F}_{j+1,m}, F_{j+1}) \to 0$ a.s. as $m \to \infty$, we start by constructing the random variables $\{(\hat{R}_i^{(t,m)}, R_i^{(t)}) : 1 \le i \le m, 0 \le t \le k\}$ as explained at the beginning of this section. Now note that for any $\epsilon, \eta > 0$,

$$\begin{aligned}
&P\left(\sup_{m\ge n} d_{p+\delta_{j+1}}(\hat{F}_{j+1,m}, F_{j+1})^{p+\delta_{j+1}} > 2^{p+\delta_{j+1}}\epsilon\right) \\
&\le P\left(\sup_{m\ge n} \left\{ d_{p+\delta_{j+1}}(\hat{F}_{j+1,m}, F_{j+1,m}) + d_{p+\delta_{j+1}}(F_{j+1,m}, F_{j+1})\right\} > 2\epsilon^{1/(p+\delta_{j+1})}\right) \\
&\le P\left(\sup_{m\ge n} d_{p+\delta_{j+1}}(\hat{F}_{j+1,m}, F_{j+1,m}) > \epsilon^{1/(p+\delta_{j+1})}\right) \\
&\quad + P\left(\sup_{m\ge n} d_{p+\delta_{j+1}}(F_{j+1,m}, F_{j+1}) > \epsilon^{1/(p+\delta_{j+1})}\right) \\
&\le P\left(\sup_{m\ge n} d_{p+\delta_{j+1}}(\hat{F}_{j+1,m}, F_{j+1,m})^{p+\delta_{j+1}} > \epsilon, \ \sup_{m\ge n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \le \eta\right) &(3.10)\\
&\quad + P\left(\sup_{m\ge n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} > \eta\right) &(3.11)\\
&\quad + P\left(\sup_{m\ge n} d_{p+\delta_{j+1}}(F_{j+1,m}, F_{j+1})^{p+\delta_{j+1}} > \epsilon\right). &(3.12)
\end{aligned}$$

To analyze (3.11) note that its convergence to zero as $n \to \infty$ is equivalent to the a.s. convergence of $d_{p+\delta_j}(\hat{F}_{j,m}, F_j)$ to zero as $m \to \infty$, which corresponds to the induction hypothesis (3.9).

To show that (3.12) converges to zero as $n \to \infty$, note that by Remark 2.4(ii) we have $E[|R^{(j+1)}|^{p+\delta}] < \infty$, which implies that $E[|R^{(j+1)}|^{p+\delta_{j+1}}] < \infty$. Hence, the Glivenko-Cantelli lemma, the strong law of large numbers, and Theorem 6.8 in [27] give that $\lim_{m\to\infty} d_{p+\delta_{j+1}}(F_{j+1,m}, F_{j+1}) = 0$ a.s., which is equivalent to

$$\lim_{n\to\infty} P\left(\sup_{m\ge n} d_{p+\delta_{j+1}}(F_{j+1,m}, F_{j+1})^{p+\delta_{j+1}} > \epsilon\right) = 0.$$

Next, to prove that (3.10) converges to zero we first define the random variables $\{Y_i^{(j,m)} : 1 \le i \le m\}$ according to (3.7), and define the events

$$A_{i,n} = \left\{ \sup_{m\ge n\vee i} \frac{\log m}{m} Y_i^{(j,m)} \le 1 \right\}.$$

Now use (2.1) and Assumption 2.3 to obtain

$$
\begin{aligned}
&P\left(\sup_{m\geq n} d_{p+\delta_{j+1}}(\hat{F}_{j+1,m},F_{j+1,m})^{p+\delta_{j+1}}>\epsilon,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right)\\
&\leq P\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}\left|\hat{R}_i^{(j+1,m)}-R_i^{(j+1)}\right|^{p+\delta_{j+1}}>\epsilon,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right)\\
&\leq P\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}Y_i^{(j,m)}>\epsilon,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta,\ \bigcap_{i=1}^{\infty}A_{n,i}\right)\\
&\quad+P\left(\sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta,\ \bigcup_{i=1}^{\infty}A_{n,i}^c\right)\\
&\leq P\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}Y_i^{(j,m)}1(Y_i^{(j,m)}\leq m/\log m)>\epsilon,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right) \qquad (3.13)\\
&\quad+\sum_{i=1}^{\infty}P\left(A_{n,i}^c,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right). \qquad (3.14)
\end{aligned}
$$

To analyze (3.13), choose $\eta=\left(\epsilon^{-1}4e^{2/\epsilon}\max\{1,E[Z^{p+\delta}]\}\right)^{-(p+\delta_j)/(p+\delta_{j+1})}$ and let $\mathcal{F}_j=\sigma(\mathscr{E}_j)$ denote the sigma-algebra generated by $\mathscr{E}_j$, as given by (3.1). Note that

$$
\begin{aligned}
&P\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}Y_i^{(j,m)}1(Y_i^{(j,m)}\leq m/\log m)>\epsilon,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right)\\
&=E\left[P\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}Y_i^{(j,m)}1(Y_i^{(j,m)}\leq m/\log m)>\epsilon\,\middle|\,\mathcal{F}_j\right)1\left(\sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right)\right].
\end{aligned}
$$

By Lemma 3.3, we obtain that on the event $\left\{\sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right\}$, we have

$$
P\left(\sup_{m\geq n}\frac{1}{m}\sum_{i=1}^{m}Y_i^{(j,m)}1(Y_i^{(j,m)}\leq m/\log m)>\epsilon\,\middle|\,\mathcal{F}_j\right)\leq 2(n-1)^{-1/2},
$$

which implies that (3.13) is bounded from above by $2(n-1)^{-1/2}$.

To analyze (3.14) note that

$$
\begin{aligned}
&\sum_{i=1}^{\infty}P\left(A_{n,i}^c,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right)\\
&=nP\left(\sup_{m\geq n}\frac{\log m}{m}Y_1^{(j,m)}>1,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right)\\
&\quad+\sum_{t=n+1}^{\infty}P\left(\sup_{m\geq t}\frac{\log m}{m}Y_1^{(j,m)}>1,\ \sup_{m\geq n} d_{p+\delta_j}(\hat{F}_{j,m},F_j)^{p+\delta_j}\leq \eta\right).
\end{aligned}
$$

Now set $q_j = p + \delta_j$ and $r_j = q_{j+1} + \delta/(2k)$, and note that $q_{j+1} < r_j < q_j \le p + \delta$. Then, by Lemma 3.4,

$$P\left(\sup_{m\ge t} \frac{\log m}{m} Y_1^{(j,m)} > 1,\ \sup_{m\ge n} d_{p+\delta_j}(\hat{F}_{j,m}, F_j)^{p+\delta_j} \le \eta\right) \le \tilde{K}_j \left(\frac{\log t}{t}\right)^{r_j/q_{j+1}}$$

for any $t \ge n$, where

$$\tilde{K}_j = 3^{r_j} \left|\left|\sum_{i=1}^{N} \varphi(C_i)\right|\right|_{r_j}^{r_j} \left\{\frac{4(\eta^{1/q_j} + ||R^{(j)}||_{q_j})^{r_j}}{1 - r_j/q_j} + 2||R^{(j)}||_{r_j}^{r_j}\right\} < \infty$$

by Remark 2.4(ii). It follows that (3.14) is bounded from above by

$$\begin{aligned}
&\tilde{K}_j n \left(\frac{\log n}{n}\right)^{r_j/q_{j+1}} + \tilde{K}_j \sum_{t=n+1}^{\infty} \left(\frac{\log t}{t}\right)^{r_j/q_{j+1}} \\
&\le \tilde{K}_j n \left(\frac{\log n}{n}\right)^{r_j/q_{j+1}} + \tilde{K}_j \sum_{t=n+1}^{\infty} \int_{t-1}^{t} \left(\frac{\log x}{x}\right)^{r_j/q_{j+1}} dx \\
&= \tilde{K}_j n \left(\frac{\log n}{n}\right)^{r_j/q_{j+1}} + \tilde{K}_j \int_{n}^{\infty} \left(\frac{\log x}{x}\right)^{r_j/q_{j+1}} dx
\end{aligned}$$

for all $n \ge 3$. Since $r_j/q_{j+1} > 1$ and

$$\int_n^\infty \left(\frac{\log x}{x}\right)^{r_j/q_{j+1}} = \frac{(\log n)^{r_j/q_{j+1}}}{(r_j/q_{j+1} - 1) n^{r_j/q_{j+1}-1}} (1 + o(1))$$

as $n \to \infty$, we conclude that (3.10) is bounded from above by

$$2(n-1)^{-1/2} + \tilde{K}_j \left(1 + \frac{1}{r_j/q_{j+1} - 1} + o(1)\right) \frac{(\log n)^{r_j/q_{j+1}}}{n^{r_j/q_{j+1}-1}},$$

which converges to zero as $n \to \infty$. This completes the proof. ■

The last proof in the paper is that of Proposition 2.11, which we give below.

Proof of Proposition 2.11. The second statement of the proposition, regarding the almost sure convergence, follows directly from Definition 6.8 and Theorem 6.9 in [27]. The convergence in probability is slightly less direct, so we give a proof below.

Using the same construction described at the beginning of Section 3, we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} h(\hat{R}_i^{(k,m)}) - E[h(R^{(k)})]\right| \le \left|\frac{1}{m}\sum_{i=1}^{m} (h(\hat{R}_i^{(k,m)}) - h(R_i^{(k)}))\right| + \left|\frac{1}{m}\sum_{i=1}^{m} h(R_i^{(k)}) - E[h(R^{(k)})]\right|,$$

where the second term converges to zero almost surely (and hence in probability) by the strong law of large numbers. Moreover, since all random variables in $\mathscr{P}^{(k,m)}$ are identically distributed,

$$E\left[\left|\frac{1}{m}\sum_{i=1}^{m}(h(\hat{R}_i^{(k,m)}) - h(R_i^{(k)}))\right|\right] \leq E\left[\frac{1}{m}\sum_{i=1}^{m}\left|h(\hat{R}_i^{(k,m)}) - h(R_i^{(k)})\right|\right]$$
$$= E\left[\left|h(\hat{R}_1^{(k,m)}) - h(R_1^{(k)})\right|\right].$$

Hence, the result will follow once we show that $\lim_{m\to\infty} E\left[\left|h(\hat{R}_1^{(k,m)}) - h(R_1^{(k)})\right|\right] = 0$, since this would imply that $m^{-1}\sum_{i=1}^{m}(h(\hat{R}_i^{(k,m)}) - h(R_i^{(k)})) \to 0$ in L_1, and therefore, in probability.
To this end, fix $M > 0$ and define $h_M(x) = h(x)1(|x| \leq M) + h(M)1(x > M) + h(-M)1(x < -M)$; note that h_M is bounded and continous. Then,

$$\begin{aligned}
E\left[\left|h(\hat{R}_1^{(k,m)}) - h(R_1^{(k)})\right|\right] &\leq E\left[\left|h(\hat{R}_1^{(k,m)}) - h_M(\hat{R}_1^{(k,m)})\right|\right] + E\left[\left|h_M(R_1^{(k)}) - h(R_1^{(k)})\right|\right] \\
&\quad + E\left[\left|h_M(\hat{R}_1^{(k,m)}) - h_M(R_1^{(k)})\right|\right] \\
&\leq E\left[\left(|h(\hat{R}_1^{(k,m)})| + |h(M)| + h(-M)|\right) 1(|\hat{R}_1^{(k,m)}| > M)\right] \\
&\quad + E\left[\left(|h(R_1^{(k)})| + |h(M)| + h(-M)|\right) 1(|R_1^{(k)}| > M)\right] \\
&\quad + E\left[\left|h_M(\hat{R}_1^{(k,m)}) - h_M(R_1^{(k)})\right|\right] \\
&\leq CE\left[(3 + |\hat{R}_1^{(k,m)}|^p + 2M^p)1(|\hat{R}_1^{(k,m)}| > M)\right] \\
&\quad + CE\left[(3 + |R_1^{(k)}|^p + 2M^p)1(|R_1^{(k)}| > M)\right] \\
&\quad + E\left[\left|h_M(\hat{R}_1^{(k,m)}) - h_M(R_1^{(k)})\right|\right].
\end{aligned}$$

Moreover, by Minkowski's inequality followed by the union bound we have

$$\begin{aligned}
&E\left[(3 + |\hat{R}_1^{(k,m)}|^p + 2M^p)1(|\hat{R}_1^{(k,m)}| > M)\right] \\
&= (3 + 2M^p)P\left(|\hat{R}_1^{(k,m)}| > M\right) + \left|\left|\hat{R}_1^{(k,m)}1(|\hat{R}_1^{(k,m)}| > M)\right|\right|_p^p \\
&\leq (3 + 2M^p)P\left(|\hat{R}_1^{(k,m)} - R_1^{(k)}| + |R_1^{(k)}| > M\right) + \left|\left|\hat{R}_1^{(k,m)} - R_1^{(k)}\right|\right|_p^p \\
&\quad + \left|\left||R_1^{(k)}|1(|\hat{R}_1^{(k,m)} - R_1^{(k)}| > M/2) + |R_1^{(k)}|1(|R_1^{(k)}| > M/2)\right|\right|_p^p \\
&\leq (3 + 2M^p)P\left(|\hat{R}_1^{(k,m)} - R_1^{(k)}| > M/2\right) + (3 + 2M^p)P\left(|R_1^{(k)}| > M/2\right) \\
&\quad + E\left[d_p(\hat{F}_{k,m}, F_k)^p\right] + \left(\left|\left|R_1^{(k)}1(|\hat{R}_1^{(k,m)} - R_1^{(k)}| > M/2)\right|\right|_p + \left|\left|R_1^{(k)}1(|R_1^{(k)}| > M/2)\right|\right|_p\right)^p \\
&\leq \left(\frac{2^p(3 + 2M^p)}{M^p} + 1\right) E\left[d_p(\hat{F}_{k,m}, F_k)^p\right] + (3 + 2M^p)P\left(|R_1^{(k)}| > M/2\right) \\
&\quad + \left(\left|\left|R_1^{(k)}1(|\hat{R}_1^{(k,m)} - R_1^{(k)}| > M/2)\right|\right|_p + \left|\left|R_1^{(k)}1(|R_1^{(k)}| > M/2)\right|\right|_p\right)^p.
\end{aligned}$$

Since $P\left(|\hat{R}_1^{(k,m)} - R_1^{(k)}| > \epsilon\right) \leq \epsilon^{-p} E[d_p(\hat{F}_{k,m}, F_k)^p]$ for any $\epsilon > 0$, our assumption implies that $\hat{R}_1^{(k,m)} \xrightarrow{P} R_1^{(k)}$ as $m \to \infty$. By the continuous mapping theorem this also implies that $h_M(\hat{R}_1^{(k,m)}) \xrightarrow{P} h_M(\hat{R}_1^{(k)})$. It follows by Theorems 13.6 and 13.7 in [29] (note that $|R_1^{(k)}|^p 1(|\hat{R}_1^{(k,m)} - R_1^{(k)}| > M/2)$ is uniformly integrable by Theorem 13.3(b) in [29]) that

$$\begin{aligned}
&\lim_{m\to\infty} E\left[\left|h(\hat{R}_1^{(k,m)}) - h(R_1^{(k)})\right|\right] \\
&\leq C(3 + 2M^p)P(|R_1^{(k)}| > M/2) + C\left|\left|R_1^{(k)} 1(|R_1^{(k)}| > M/2)\right|\right|_p^p \\
&\quad + C(3 + 2M^p)P(|R_1^{(k)}| > M) + CE\left[|R_1^{(k)}|^p 1(|R_1^{(k)}| > M)\right] \\
&\leq 2C(3 + 2M^p)P(|R_1^{(k)}| > M/2) + 2CE\left[|R_1^{(k)}|^p 1(|R_1^{(k)}| > M/2)\right].
\end{aligned}$$

Taking the limit as $M \to \infty$ completes the proof since $E[|R_1^{(k)}|^p] < \infty$. ■

References

[1] D.J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equation. *Annals of Applied Probability*, 15(2):1047–1110, 2005.

[2] G. Alsmeyer, J.D. Biggins, and M. Meiners. The functional equation of the smoothing transform. *Ann. Probab.*, 40(5):2069–2105, 2012.

[3] G. Alsmeyer and P. Dyszewski. Thin tails of fixed points of the nonhomogeneous smoothing transform. *Stochastic Processes and their Applications*, 2017.

[4] G. Alsmeyer and M. Meiners. Fixed points of inhomogeneous smoothing transforms. *J. Differ. Equ. Appl.*, 18(8):1287–1304, 2012.

[5] G. Alsmeyer and M. Meiners. Fixed points of the smoothing transform: Two-sided solutions. *Probab. Theory Rel.*, 155(1-2):165–199, 2013.

[6] K.B. Athreya. Discounted branching random walks. *Advances in Applied Probability*, 17:53–66, 1985.

[7] J.D. Biggins. Lindley-type equations in the branching random walk. *Stochastic Process. Appl.*, 75:105–133, 1998.

[8] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *To appear in Memoirs of the American Mathematical Society*, 2017.

[9] N. Chen, N. Litvak, and M. Olvera-Cravioto. Generalized pagerank on directed configuration networks. *Random Structures & Algorithms*, 2016.

[10] N. Chen and M. Olvera-Cravioto. Efficient simulation for branching linear recursions. In *Proceedings of the Winter Simulation Conference 2015*, pages 2716–2727, Huntington Beach, CA, 2015.

[11] E. del Barrio, E. Giné, and C. Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Annals of Probability*, pages 1009–1071, 1999.

[12] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *Ann. Appl. Probab.*, 20(2):565–592, 2010.

[13] L. Devroye. On the probabilistic worst-case time of FIND. *Algorithmica*, 31:291–303, 2001.

[14] J.A. Fill and S. Janson. Approximating the limiting Quicksort distribution. *Random Structures Algorithms*, 19(3-4):376–406, 2001.

[15] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields*, 162:707–738, 2015.

[16] A.M. Iksanov. Elementary fixed points of the BRW smoothing transforms with infinite number of summands. *Stochastic Process. Appl.*, 114:27–50, 2004.

[17] S. Janson. On the tails of the limiting quicksort distribution. *Electron. Commun. Probab.*, 20(81):1–7, 2015.

[18] P.R. Jelenković and M. Olvera-Cravioto. Information ranking and power laws on trees. *Adv. Appl. Prob.*, 42(4):1057–1093, 2010.

[19] P.R. Jelenković and M. Olvera-Cravioto. Implicit renewal theory and power tails on trees. *Adv. Appl. Prob.*, 44(2):528–561, 2012.

[20] P.R. Jelenković and M. Olvera-Cravioto. Maximums on trees. *Stochastic Processes and their Applications*, 125:217–232, 2015.

[21] F.I. Karpelevich, M.Ya. Kelbert, and Yu.M. Suhov. Higher-order Lindley equations. *Stochastic Process. Appl.*, 53:65–96, 1994.

[22] Q. Liu. Fixed points of a generalized smoothing transformation and applications to the branching random walk. *Adv. Appl. Prob.*, 30:85–112, 1998.

[23] M. Mezard and A. Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[24] M. Olvera-Cravioto and O. Ruiz-Lacedelli. Parallel queues with synchronization. *arXiv:1501.00186*, 2014.

[25] U. Rösler. A limit theorem for "Quicksort". *RAIRO Theor. Inform. Appl.*, 25:85–100, 1991.

[26] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1-2):3–33, 2001.

[27] C. Villani. *Optimal transport, old and new.* Springer, New York, 2009.

[28] Y. Volkovich and N. Litvak. Asymptotic analysis for personalized web search. *Adv. Appl. Prob.*, 42(2):577–604, 2010.

[29] D. Williams. *Probability with Martingales.* Cambridge University Press, Cambridge, 2001.