VARIATIONS OF THE ELEPHANT RANDOM WALK

ALLAN GUT,* *Uppsala University*
ULRICH STADTMÜLLER ,** *Ulm University*

Abstract

In the classical simple random walk the steps are independent, that is, the walker has no memory. In contrast, in the elephant random walk, which was introduced by Schütz and Trimper [19] in 2004, the next step always depends on the whole path so far. Our main aim is to prove analogous results when the elephant has only a restricted memory, for example remembering only the most remote step(s), the most recent step(s), or both. We also extend the models to cover more general step sizes.

Keywords: Elephant random walk; law of large numbers; asymptotic (non)normality; method of moments; difference equation; Markov chain

2010 Mathematics Subject Classification: Primary 60F05; 60G50
Secondary 60F15; 60J10

1. Introduction

In the classical *simple* random walk the steps are equal to plus or minus one and independent $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = -1)$: the walker has no memory. This random walk is, in particular, Markovian. Motivated by applications, but interesting in its own right, is the case when the walker has some memory. The extreme case is, of course, when the walker has a complete memory, that is, when the 'next step' depends on the whole process so far. This so-called elephant random walk (ERW) was introduced by Schütz and Trimper [19] in 2004, the name inspired by the fact that elephants have a very long memory. The first more mathematically rigorous work on elephant random walks are, to the best of our knowledge, the papers by Bercu [2] and Coletti *et al.* [4], which contain a number of limit theorems. A main point is that the process is subject to a kind of phase transition, which divides the problem into a diffusive regime, a critical regime, and a superdiffusive regime, with somewhat different asymptotics.

Our main interest is the situation in which the elephant has only a limited memory, remembering only some distant past, only a recent past, or a mixture of both. We were motivated by simulations studying the case in which a given fraction of the distant/recent past is remembered [6, 18, 20] but where no theoretical results are provided. This is a first step towards a better understanding of situations with a memory increasing in time, and for finding the breaking point for the phase transitions. This is interesting, for example, in connection with the concept of memory lapse [9].

We begin by studying the cases when the walker only remembers the first (two) step(s) or only the most recent steps. In particular, the latter case involves rather cumbersome computations and we therefore invite the reader(s) to try to push our results further. One such source is Ben-Ari *et al.* [1], where one application of their more general model is the case when the elephant remembers the L most recent steps. The paper by Engländer and Volkov [8] is devoted to a variation in that the next step is not generated by flipping a coin, but rather by turning it over or not. They have a somewhat different focus; in particular, they allow for different p-values in each step. In addition, there is a large literature dealing with so-called correlated random walks, though with different aims. Let us mention Chen and Renshaw [3], for example, who investigated a walk in dimension d and the probability of returns. Menshikov and Volkov [17] considered continuous-time processes generalizing the ERW and questions of transience and recurrence, and Comets *et al.* [5] studied a kind of self-avoiding walk in $\mathbb{R}^d$. See also the literature cited therein for further references.

The cases with limited memory behave very differently mathematically; in particular, there are no phase transitions in these cases.

A second point concerns the extension of (some of) Bercu's results in [2] from the simple random walk to more general sums.

We begin by defining the various models in Section 2. After some preliminaries in Section 3, some results for general ERWs are obtained in Section 4. Sections 5 and 6 are devoted to the distant past and Sections 7 and 8 to the recent past, respectively. These 'one-sided' memories are then followed up in Sections 9 and 10, where we consider mixed cases, that is, when the memory contains some early steps as well as a recent one, after which we briefly discuss some extensions. We close with a section containing some questions and remarks. For easier reading we collect some of the more lengthy (elementary and tedious) computations in the Appendix.

2. Background

The elephant random walk is defined as a simple random walk, where, however, the steps are not i.i.d. but dependent, as follows. The first step X_1 equals 1 with probability $r \in [0, 1]$ and is equal to -1 with probability $1 - r$. After n steps, that is, at position $S_n = \sum_{k=1}^{n} X_k$, one defines

$$X_{n+1} = \begin{cases} +X_K & \text{with probability } p \in [0, 1], \\ -X_K & \text{with probability } 1 - p, \end{cases}$$

where K has a uniform distribution on the integers $1, 2, \ldots, n$. With the σ-algebras $\mathcal{G}_n = \sigma\{X_1, X_2, \ldots, X_n\}$, this means (formula (2.2) of [2]) that

$$\mathbb{E}(X_{n+1} \mid \mathcal{G}_n) = (2p - 1) \cdot \frac{S_n}{n}, \tag{2.1}$$

after which, setting $a_n = \Gamma(n) \cdot \Gamma(2p)/\Gamma(n + 2p - 1)$, it turns out that $\{M_n = a_n S_n, \ n \geq 1\}$ is a martingale; see [2, Section 2].

Our main aim is to extend these results to the case when the elephant has a restricted memory, for example remembering only the most remote step(s) and/or the most recent one(s). A result in Section 4 allows us to conclude that our results also remain true (suitably modified) when the steps of the ERWs follow a general distribution on the integers.

First in line is the case when the elephant only remembers the distant past, the most extreme one being when the memory is reduced to the first step only, that is,

$$X_{n+1} = \begin{cases} +X_1 & \text{with probability } p \in [0, 1], \\ -X_1 & \text{with probability } 1 - p. \end{cases}$$

Somewhat more sophisticated is the case when the memory covers the first two steps, for which

$$X_{n+1} = \begin{cases} +X_K & \text{with probability } p \in [0, 1], \\ -X_K & \text{with probability } 1 - p, \end{cases}$$

where $\mathbb{P}(K = 1) = \mathbb{P}(K = 2) = 1/2$.

Technically more complicated is the case when the elephant only remembers the recent past. Here we focus on the very recent past, which is the last step, that is,

$$X_{n+1} = \begin{cases} +X_n & \text{with probability } p \in [0, 1], \\ -X_n & \text{with probability } 1 - p. \end{cases}$$

This case is also called a *correlated random walk* (CRW).

We begin, throughout, by assuming that $X_1 = 1$, and specialize our findings in this setting (for simplicity) to the case $r = p$. We denote our partial sums by T_n, $n \geq 1$, when the first variable(s) is/are fixed, and let S_n be reserved for the case when they are random.

In order to move from T_n to S_n we also need to discuss the behavior of the walk when the initial value equals -1. However, in that case the evolution of the walk is the same *except* for the fact that the trend of the walk is reversed, that is, the corresponding walk equals the mirrored image in the time axis. This implies that the mean after n steps equals $-\mathbb{E}(T_n)$, but the identical dynamics implies that the variance remains the same ($\text{var}(-Y) = \text{var}(Y)$ for a random variable Y). In fact the second moments of the walk remain the same. The same goes for higher-order moments: odd moments equal the negative of those when $X_1 = 1$, and even moments remain the same. In Sections 6 and 10 we depart from the assumption that X_1 and X_2 are fixed, and then the additional case $X_1 + X_2 = 0$ has to be taken care of.

Finally, in order to avoid special effects we assume throughout that $0 < p < 1$; note that $p = 1$ corresponds to $X_n = X_1$ for all n.

3. Some auxiliary material

For easier access to the arguments below we present some auxiliary results from probability and analysis.

3.1. Disturbed limit distributions

The following (well-known) result (which is a special case of the Cramér–Slutsky theorem) will be used in order to go from a special case to a more general one.

Proposition 3.1. *Let $\{U_n, n \geq 1\}$ be a sequence of random variables and suppose that V is independent of all of them. If $U_n \xrightarrow{d} U$ as $n \to \infty$, then $U_n V \xrightarrow{d} UV$ as $n \to \infty$.*

Proof. Using characteristic functions and bounded convergence we have, as $n \to \infty$,

$$\varphi_{U_n V}(t) = \mathbb{E}\exp\{itU_n V\}$$
$$= \mathbb{E}(\mathbb{E}(\exp\{itU_n V\} \mid V))$$
$$= \mathbb{E}\varphi_{U_n}(tV) \to \mathbb{E}\varphi_U(tV)$$
$$= \mathbb{E}(\mathbb{E}(\exp\{itUV\} \mid V))$$
$$= \mathbb{E}\exp\{itUV\})$$
$$= \varphi_{UV}(t).$$

An application of the continuity theorem for characteristic functions finishes the proof. $\quad\square$

3.2. Conditioning in case of a restricted memory

Let $\{S_n, \ n \geq 1\}$ be an ERW, and let $\mathfrak{M}_n \subset \{1, 2, \ldots, n\}$ be the memory of the elephant from the first n steps, that is, $\mathfrak{M}_n$ *contains the steps from the past (up to step n) on which the elephant bases the next step.* Further, set $\mathcal{F}_n = \sigma\{X_k, k \in \mathfrak{M}_n\}$, $n \geq 1$, and let $\mathcal{G}_n = \sigma\{X_1, X_2, \ldots, X_n\}$, $n \geq 1$, stand for the σ-algebra generated by the complete past up to step n. We already know from (2.1) above that $\mathbb{E}(X_{n+1} \mid \mathcal{G}_n) = (2p - 1)S_n/n$. Our aim is to establish analogs when the elephant has a restricted memory, that is, expressions for $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)$. Then we have

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = p \cdot \sum_{i \in \mathfrak{M}_n} \frac{1}{|\mathfrak{M}_n|} X_i + (1 - p) \cdot \sum_{i \in \mathfrak{M}_n} \frac{1}{|\mathfrak{M}_n|}(-X_i)$$
$$= (2p - 1) \cdot \frac{\sum_{i \in \mathfrak{M}_n} X_i}{|\mathfrak{M}_n|}, \tag{3.1}$$

that is, the conditional mean equals the average of the possible choices multiplied by the expected value of the sign, in analogy with (2.1).

If, for example, $\mathfrak{M}_n = \{n\}$ the elephant only considers the most recent step, and if $\mathfrak{M}_n = \{1, n\}$ only the first and the most recent steps; these are two cases that will be investigated below. In these cases (3.1) states that

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = (2p - 1)X_n \quad \text{and} \quad \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = (2p - 1)\frac{X_1 + X_n}{2},$$

respectively.

The next problem is when we condition on steps that are not contained in some $\mathfrak{M}_n$. The elephant therefore cannot choose among them in a subsequent step. Technically, let $\mathfrak{M} \subset \{1, 2, \ldots, n\}$ be an arbitrary set of indices such that $\mathfrak{M} \cap \mathfrak{M}_n = \emptyset$. Then

$$\mathbb{E}(X_{n+1} \mid \sigma\{\mathfrak{M}_n \cup \mathfrak{M}\}) = \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = (2p - 1)\frac{\sum_{i \in \mathfrak{M}_n} X_i}{|\mathfrak{M}_n|}. \tag{3.2}$$

It follows, in particular, that

$$\mathbb{E}(X_{n+1} \mid \mathcal{G}_n) = (2p - 1)\frac{\sum_{i \in \mathfrak{M}_n} X_i}{|\mathfrak{M}_n|},$$

and that

$$\mathbb{E}(S_n X_{n+1} \mid \mathcal{G}_n) = S_n \mathbb{E}(X_{n+1} \mid \mathcal{G}_n) = S_n(2p - 1)\frac{\sum_{i \in \mathfrak{M}_n} X_i}{|\mathfrak{M}_n|}. \tag{3.3}$$

Exploiting the smoothing lemma (see e.g. [10, Lemma 10.1.1]), according to which $\mathbb{E}(S_n X_{n+1} \mid \mathcal{G}_n) = \mathbb{E}(\mathbb{E}(S_n X_{n+1} \mid \mathcal{G}_n))$, and the fact that $X_{n+1}^2 = 1$, both of which will be useful several times for the computation of second moments, yields

$$\begin{aligned}
\mathbb{E}(S_{n+1}^2) &= \mathbb{E}(S_n + X_{n+1})^2 \\
&= \mathbb{E}(S_n^2) + 2\,\mathbb{E}(S_n X_{n+1}) + \mathbb{E}(X_{n+1}^2) \\
&= \mathbb{E}(S_n^2) + \frac{2(2p-1)}{|\mathfrak{M}_n|}\,\mathbb{E}\left(S_n \sum_{i \in \mathfrak{M}_n} X_i\right) + 1.
\end{aligned} \tag{3.4}$$

3.3. Some notation

We use the standard $\delta_a(x)$ to denote the distribution function with a jump of height one at a and $\mathcal{N}_{\mu, \sigma^2}$ for the normal distribution with mean μ and variance σ^2. The arrows $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and convergence in distribution, respectively. Constants c and C are always numerical constants that may change between appearances.

4. General elephant random walks

Let $\{\widetilde{S}_n,\ n \geq 1\}$ be an ERW, and suppose that R is a random variable with distribution function F_R that is independent of the walk. If $\widetilde{S}_n/a_n \xrightarrow{\text{a.s.}} Z$ as $n \to \infty$ for some normalizing positive sequence $a_n \to \infty$ as $n \to \infty$, and some random variable Z, it follows from Proposition 3.1 that $R\widetilde{S}_n/a_n \xrightarrow{\text{a.s.}} RZ$ as $n \to \infty$. An immediate consequence of this fact is that we can extend Theorems 3.1, 3.4, and (the first half of) Theorem 3.7 of [2] to cover more general step sizes. Namely, consider the ERW for which $\widetilde{X}_1 \equiv 1$, and let the random variables $\widetilde{X}_n,\ n \geq 2$, be constructed as in Section 2 with this special $\widetilde{X}_1$ as starting point. Furthermore, let R be a random variable, independent of $\{\widetilde{X}_n,\ n \geq 1\}$, and consider $X_n = R \cdot \widetilde{X}_n,\ n \geq 1$, and hence $S_n = R \cdot \widetilde{S}_n$.

The following theorem (which reduces to the cited Theorems 3.1, 3.3, and 3.7 of [2], respectively, if R is a coin-tossing random variable) holds for $S_n = R\widetilde{S}_n$.

Theorem 4.1.

(a) *For $0 < p < 3/4$,*

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

(b) *For $p = 3/4$,*

$$\frac{S_n}{\sqrt{n}\log n} \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

(c) *For $3/4 < p < 1$,*

$$\frac{S_n}{n^{2p-1}} \xrightarrow{\text{a.s.}} RL \quad \text{as } n \to \infty,$$

where L is a non-degenerate random variable.

As for convergence in distribution, we have to distinguish more carefully between the three cases.

Theorem 4.2. *For $0 < p < 3/4$, we obtain*

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \int_{\mathbb{R}\setminus\{0\}} \mathcal{N}_{0,1/(3-4p)}(\,\cdot\,/|t|)\,\mathrm{d}F_R(t) + \mathbb{P}(R=0)\cdot\delta_{[0,\infty)}(\,\cdot\,) \quad \text{as } n \to \infty.$$

Moreover, if $\mathbb{E}(R^2) < \infty$, then

$$\mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right) \to 0 \quad and \quad \mathbb{E}\left(\left(\frac{S_n}{\sqrt{n}}\right)^2\right) \to \frac{\mathbb{E}(R^2)}{(3-4p)} \quad as\ n \to \infty.$$

Proof. As R and S_n are independent, we find that

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \le x\right) = \mathbb{E}\left(\mathbb{P}\left(R\frac{\widetilde{S}_n}{\sqrt{n}} \le x \mid R\right)\right)$$

$$= \int_{\mathbb{R}} \mathbb{P}\left(t\frac{\widetilde{S}_n}{\sqrt{n}} \le x\right) dF_R(t)$$

$$= \int_{(-\infty,0)} \mathbb{P}\left(\frac{\widetilde{S}_n}{\sqrt{n}} \ge x/t\right) dF_R(t) + \int_{(0,\infty)} \mathbb{P}\left(\frac{\widetilde{S}_n}{\sqrt{n}} \le x/t\right) dF_R(t)$$

$$+ \mathbb{P}(R=0)\cdot\delta_{[0,\infty)}(x)$$

$$\to \int_{(-\infty,0)} (1 - \mathcal{N}_{0,1/(3-4p)}(x/t)) dF_R(t) + \int_{(0,\infty)} \mathcal{N}_{0,1/(3-4p)}(x/t) dF_R(t)$$

$$+ \mathbb{P}(R=0)\cdot\delta_{[0,\infty)}(x),$$

by dominated convergence, which yields the desired result.

The second part is immediate, since R is independent of everything else. $\qquad\square$

Remark 4.1. If $R = \pm 1$ with probabilities r and $(1-r)$, respectively, the limit distributions of $S_n/\sqrt{n}$ and $\widetilde{S}_n/\sqrt{n}$ are the same, and we rediscover Theorem 3.3 of [2].

Remark 4.2. For the critical case, $p = 3/4$, one similarly obtains, using [2, Theorem 3.6], that

$$\frac{S_n}{\sqrt{n\log n}} \xrightarrow{d} \int_{\mathbb{R}\setminus\{0\}} \mathcal{N}_{0,1}(\cdot/|t|) dF_R(t) + \mathbb{P}(R=0)\cdot\delta_{[0,\infty)}(\cdot) \quad as\ n \to \infty.$$

The supercritical case, $3/4 < p < 1$, has a different evolution and no analogous result exists.

5. Remembering only the distant past 1: $\mathfrak{M}_n = \{1\}$

This is the easiest case. We begin by assuming that the elephant only remembers the first step, i.e. that $\mathcal{F}_n = \sigma\{X_1\}$, and that $X_1 = 1$ (recall that partial sums are denoted by the letter T). The following steps are then either $+1$ or -1 independently of each other. This means that we are faced with a simple random walk with drift $2p - 1$, except for the fact that the first step is always equal to one. It follows immediately that

$$\mathbb{E}(T_{n+1}) = 1 + n(2p-1) \quad and \quad \text{var}(T_{n+1}) = 4p(1-p)n.$$

Hence we obtain the following result.

Proposition 5.1. *The strong law of large numbers, the central limit theorem, and the law of the iterated logarithm all hold for $\{T_n,\ n \ge 1\}$ with the corresponding normalizations.*

If, on the other hand, the first step equals -1, the analogous random walk has drift $-(2p-1)$, and $\mathbb{E}(T_{n+1}) = -1 - n(2p-1)$. The variance remains the same (recall the discussion towards the end of Section 2), and the classical laws hold again.

Assuming that X_1 is a coin-tossing random variable, we are confronted with two random walks, one for each of the two portions of the probability space. In fact, if we imagine the situation that $r = \mathbb{P}(X_1 = +1)$ is close to zero or one, it is rather apparent how the very first step determines along which branch the process will evolve. If $p = 1/2$, the two 'branches' determined by the first step collapse (asymptotically) into one, and we are ultimately faced with a simple symmetric random walk. Combining the two branches, the following theorem emerges.

Theorem 5.1. *Let* $S_n = \sum_{k=1}^n X_k$. *Then*

(a) $\dfrac{S_n}{n} \xrightarrow{d} \begin{cases} 2p - 1 & \text{with probability } p, \\ -(2p - 1) & \text{with probability } 1 - p, \end{cases} \quad \text{as } n \to \infty,$

(b) $\mathbb{E}\left(\dfrac{S_n}{n}\right) \to (2p - 1)^2$ *and* $\text{var}\left(\dfrac{S_n}{n}\right) \to 4p(1 - p)(2p - 1)^2$ *as* $n \to \infty$.

Proof. (a) If $X_1 = \pm 1$, we know from the above that $\mathbb{E}(T_n) = \pm(1 + (n - 1)(2p - 1))$ and that $\text{var}(T_n) = 4p(1 - p)(n - 1)$. This tells us that $T_n/n \xrightarrow{p} \pm(2p - 1)$ as $n \to \infty$. The conclusion follows.

(b) Immediate (bounded convergence). $\qquad\square$

Remark 5.1.

(i) Part (b) tells us that the asymptotic variance is of order n^2. This is due to the fact that the two branches force the elephant to walk in opposite directions.

(ii) An interpretation of the limit in (a) is that the random walk at hand, on average, behaves, asymptotically, like a coin-tossing random variable with values at the points $\pm(2p - 1)$.

(iii) An alternative way of phrasing the conclusion of the theorem is that

$$F_{S_n/n}(x) \to p \cdot \delta_{-(2p-1)}(x) + (1 - p) \cdot \delta_{2p-1}(x) \quad \text{as } n \to \infty.$$

However, if we use a random normalization we obtain the following result.

Theorem 5.2. *Let* $S_n = \sum_{k=1}^n X_k$. *Then*

(a) $\dfrac{S_n - n(2p - 1)X_1}{\sqrt{4np(1 - p)}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty,$

(b) $\dfrac{S_n - n(2p - 1)X_1}{n} \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty,$

(c) $\limsup\limits_{n\to\infty} \left(\liminf\limits_{n\to\infty}\right) \dfrac{S_n - n(2p - 1)X_1}{\sqrt{8np(1 - p)\log\log n}} = 1\,(-1) \quad \text{a.s.}$

Proof. (a) We use the fact that if $\mathfrak{M}_n = X_1$ then $S_n = X_1 T_n$, in order to get

$$\frac{S_n - n(2p - 1)X_1}{\sqrt{4n\,p\,(1 - p)}} = X_1 \frac{T_n - n(2p - 1)}{\sqrt{4np(1 - p)}},$$

together with Theorem 4.2 and its Remark 4.1.

Alternatively, one may condition on the value of X_1. This procedure will be exploited in the proof of Theorem 6.2 in the next section.

(b,c) Define $\Omega_1 = \{\omega \in \Omega : X_1(\omega) = 1\}$ and $\Omega_2 = \Omega_1^c$. After renormalization, the original probability measure will be a probability measure on Ω_1. Based on this measure on Ω_1, we obtain an SLLN and an LIL for $S_n - n(2p - 1)X_1$, and similarly on Ω_2. Combining them yields the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\Box$

Remark 5.2. If X_1 is a general random variable with distribution F having no mass at zero, then

$$\frac{S_n - n(2p - 1)X_1}{\sqrt{4np(1 - p)}} \xrightarrow{d} \int_{-\infty}^{\infty} \mathcal{N}_{0,1}(\cdot / |t|) \, \mathrm{d}F(t) \quad \text{as } n \to \infty.$$

A special case is, once again, $p = 1/2$.

Corollary 5.1. If $p = 1/2$, then

$$\frac{S_n}{n} \xrightarrow{a.s.} 0, \quad \frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty \quad \text{and} \quad \limsup_{n \to \infty} (\liminf_{n \to \infty}) \frac{S_n}{\sqrt{2n \log \log n}} \overset{a.s.}{=} 1 \, (-1).$$

6. Remembering the distant past 2: $\mathfrak{M}_n = \{1, 2\}$

Suppose that the elephant only remembers the first two steps, i.e. $\mathcal{F}_1 = \sigma\{X_1\}$ and $\mathcal{F}_n = \sigma\{X_1, X_2\}$ for $n \geq 2$. This case is slightly more involved, since we are faced with three branches. Extending the arguments from the previous section, it follows that the walk evolves as an ordinary simple random walk beginning at the third step.

Suppose first that $X_1 = X_2 = 1$. Then, for $n \geq 2$,

$$\begin{aligned}
\mathbb{E}(X_{n+1}) &= \mathbb{E}(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) \\
&= \mathbb{E}(\mathbb{E}(X_{n+1} \mid X_1, X_2)) \\
&= \mathbb{E}\left((2p - 1) \cdot \frac{1 + 1}{2}\right) \\
&= 2p - 1,
\end{aligned}$$

and hence

$$\mathbb{E}(T_{n+1}) = 2 + (n - 1)(2p - 1) = n(2p - 1) + 3 - 2p.$$

Since randomness is involved only from the third step and onwards,

$$\operatorname{var}(T_{n+1}) = 4p(1 - p)(n - 1), \quad n \geq 3.$$

By continuing as before, we obtain, after proper centering, limit theorems for these initial X-values, and similarly for the other branches. One can also ascertain that the variance is not linear if we assume random beginnings, except for the case $p = 1/2$, when, as in the previous section, the branches collapse into one and we are faced with a simple symmetric random walk, that is, the three main limit theorems (SLLN, CLT, LIL) hold (as in Corollary 5.1).

The following analog of Theorem 5.1 holds in the general case (as one might expect).

Theorem 6.1. *Let* $S_n = \sum_{k=1}^n X_k$. *Then*

(a) $\dfrac{S_n}{n} \xrightarrow{d} \begin{cases} 2p-1 & \text{with probability } p^2, \\ 0 & \text{with probability } 1-p, \\ -(2p-1) & \text{with probability } p(1-p), \end{cases}$ *as* $n \to \infty$,

(b) $\mathbb{E}\left(\dfrac{S_n}{n}\right) \to p(2p-1)^2$ *and* $\operatorname{var}\left(\dfrac{S_n}{n}\right) \to p(1-p)(2p-1)^2(4p^2+1)$.

Proof. (a) If $X_1 = X_2 = \pm 1$, we know from the above that $\mathbb{E}(T_n) = \pm(n(2p-1) + 3 - 2p)$, and that $\operatorname{var}(T_n) = 4p(1-p)(n-2)$. Moreover, $\mathbb{E}(T_n) = 0$ whenever X_1 and X_2 have different signs. The variance remains the same (with $p = 1/2$). This, together with the fact that

$$\mathbb{P}(X_1 = X_2 = 1) = p^2,$$
$$\mathbb{P}(X_1 = X_2 = -1) = (1-p)p,$$
$$\mathbb{P}(X_1 \neq X_2) = p(1-p) + (1-p)^2 = 1-p,$$

helps us to finish the proof of the first part. Part (b) follows. □

Remark 6.1.

(i) In analogy with Remark 5.1 we have the interpretation that the elephant, asymptotically, on average, performs a random walk on the points $\pm(2p-1)$ and 0.

(ii) Mimicking Remark 5.1, we may rewrite the conclusion of the theorem as

$$F_{S_n/n}(x) \to p(1-p) \cdot \delta_{-(2p-1)}(x) + (1-p) \cdot \delta_0(x) + p^2 \cdot \delta_{2p-1}(x) \quad \text{as } n \to \infty.$$

Once again random normalization produces further limit results.

Theorem 6.2. *Let* $S_n = \sum_{k=1}^n X_k$. *Then*

(a) $\dfrac{S_n - n(2p-1)(X_1 + X_2)/2}{\sqrt{n}} \xrightarrow{d} p \cdot \mathcal{N}_{0,4p(1-p)} + (1-p) \cdot \mathcal{N}_{0,1}$ *as* $n \to \infty$,

(b) $\dfrac{S_n - n(2p-1)(X_1 + X_2)/2}{n} \xrightarrow{a.s.} 0$ *as* $n \to \infty$,

(c) $\limsup\limits_{n\to\infty} (\liminf\limits_{n\to\infty}) \dfrac{S_n - n(2p-1)(X_1 + X_2)/2}{\sqrt{2n \log \log n}} = \sigma_{1,2}\,(-\sigma_{1,2})$ *a.s.*,

where

$$\sigma_{1,2}^2 = \begin{cases} 4p(1-p) & \text{for } \omega \in \{\omega \in \Omega : X_1(\omega) \cdot X_2(\omega) = 1\}, \\ 1 & \text{otherwise.} \end{cases}$$

Proof. (a) Conditioning on the value of $(X_1 + X_2)/2$, we obtain

$$\mathbb{P}\left(\frac{S_n - n(2p-1)(X_1+X_2)/2}{\sqrt{n}} \leq x\right)$$

$$= \mathbb{P}\left(\frac{S_n - n(2p-1)(X_1+X_2)/2}{\sqrt{n}} \leq x \mid X_1 = X_2 = 1\right) \cdot p^2$$

$$+ \mathbb{P}\left(\frac{S_n - n(2p-1)(X_1+X_2)/2}{\sqrt{n}} \leq x \mid X_1 = X_2 = -1\right) \cdot p(1-p)$$

$$+ \mathbb{P}\left(\frac{S_n - n(2p-1)(X_1+X_2)/2}{\sqrt{n}} \leq x \mid X_1 + X_2 = 0\right) \cdot (1-p)$$

$$= \mathbb{P}\left(\frac{T_n - n(2p-1)}{\sqrt{n}} \leq x\right) \cdot p^2 + \mathbb{P}\left(\frac{-T_n + n(2p-1)}{\sqrt{n}} \leq x\right) \cdot p(1-p)$$

$$+ \mathbb{P}\left(\frac{T_n}{\sqrt{n}} \leq x \mid X_1 + X_2 = 0\right) \cdot (1-p)$$

$$\to (p^2 + p(1-p)) \cdot \mathcal{N}_{0,4p(1-p)}(x) + (1-p) \cdot \mathcal{N}_{0,1}(x)$$

$$= p \cdot \mathcal{N}_{0,4p(1-p)}(x) + (1-p) \cdot \mathcal{N}_{0,1}(x) \quad \text{as } n \to \infty.$$

Parts (b) and (c) follow along the lines of the proof of Theorem 5.1. □

We close this section by mentioning that analogous results can be obtained if the elephant only remembers the first m random variables for some $m \in \mathbf{N}$. The following natural extension of the above results emerges by generalizing the above proofs.

Theorem 6.3. *For* $q_k = \mathbb{P}(S_m = m - 2k)$, $r_k = ((m-k)p + k(1-p))/m$, *and* $p_k = (m-2k)$ $(2p-1)/m$, *where* $0 \leq k \leq m$ *and* $m \in \mathbf{N}$,

$$\frac{S_n}{n} \xrightarrow{d} \sum_{k=0}^{m} q_k \delta_{p_k} \quad \text{as } n \to \infty,$$

and

$$\frac{S_n - n(2p-1)S_m/m}{\sqrt{n}} \xrightarrow{d} \sum_{k=0}^{m} q_k \mathcal{N}_{0,4r_k(1-r_k)} \quad \text{as } n \to \infty.$$

7. Remembering only the recent past 1: $\mathfrak{M}_n = \{n\}$

This situation is more complex, because, even though one remembers only recent steps, the path depends on the whole history so far (some remarks will be given in Section 11.3). We begin by assuming that the elephant only remembers the very last step. From Section 2 we recall that this model is also called a correlated random walk (CRW). The setting is reminiscent of [8], where one turns over a coin instead of tossing it. The main focus there, however, is on different p-values at each step and, for example, how this may affect phase transitions and behavior at critical values.

We begin, as always, by assuming that $X_1 = 1$. Then $\mathbb{E}(X_1) = 1$, and

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(X_{n+1} \mid X_n) = (2p-1) \cdot X_n \quad \text{and} \quad \mathbb{E}(X_{n+1}) = (2p-1)\,\mathbb{E}(X_n)$$

for all $n \geq 2$. By iterating this, it follows, for $n \geq 0$, that

$$\mathbb{E}(X_{n+1}) = (2p-1)^n \, \mathbb{E}(X_1) = (2p-1)^n, \tag{7.1}$$

and that

$$\mathbb{E}(T_{n+1}) = \frac{1 - (2p-1)^{n+1}}{2(1-p)}.$$

For the second moment we have, by (3.4) and (3.2),

$$\mathbb{E}(T_{n+1}^2 \mid \mathcal{G}_n) = T_n^2 + 2T_n(2p-1)X_n + 1 \quad \text{and} \quad \mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + 2(2p-1)\,\mathbb{E}(T_nX_n) + 1.$$

For the middle term we obtain by (3.2),

$$\mathbb{E}(T_nX_n) = \mathbb{E}(X_n^2) + \mathbb{E}(T_{n-1}\,\mathbb{E}(X_n \mid \mathcal{G}_{n-1})) = 1 + (2p-1)\,\mathbb{E}(T_{n-1}X_{n-1}),$$

which, after iteration, yields

$$\mathbb{E}(T_nX_n) = 1 + \sum_{k=1}^{n-1}(2p-1)^k = \frac{1-(2p-1)^n}{2(1-p)}.$$

Now we can calculate the second moment:

$$\mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + 2(2p-1) \cdot \frac{1-(2p-1)^n}{2(1-p)} + 1 = \mathbb{E}(T_n^2) + \frac{p}{1-p} - \frac{(2p-1)^{n+1}}{1-p}.$$

By telescoping, we obtain

$$\mathbb{E}(T_{n+1}^2) = \frac{np}{1-p} + \mathrm{O}(1) \quad \text{as } n \to \infty,$$

which implies the following formula for the asymptotic variance:

$$\mathrm{var}(T_{n+1}) = \frac{np}{1-p} + \mathrm{O}(1) \quad \text{as } n \to \infty. \tag{7.2}$$

Noticing that $S_n = X_1T_n$ and that $X_1 = \pm 1$, a glance at (7.1) and (7.2) shows that $T_n/n \xrightarrow{p} 0$ and that $S_n/n \xrightarrow{p} 0$ as $n \to \infty$, suggesting the following result.

Theorem 7.1. *For $X_1 = \pm 1$,*

$$\frac{T_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,p/(1-p)} \quad \text{and} \quad \frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,p/(1-p)} \quad \text{as } n \to \infty.$$

Proof. The sequence $\{X_n, \, n \geq 1\}$ is a stationary recurrent Markov chain with finite state space which, hence, is uniformly ergodic. The asymptotic normality of T_n therefore follows from a CLT for Markov chains; see e.g. Corollary 5 of [14] (see also [13, Theorem 19.1]). The limit result for S_n then follows as in Theorem 4.2. □

The Markov property also provides a strong law.

Theorem 7.2. *We have*

$$\frac{S_n}{n} \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

Proof. The stationary distribution of the ergodic Markov chain $\{X_n, \, n \geq 1\}$ is $(1/2, 1/2)$, which has expectation zero. An application of Theorem 6.1 of [7] yields the conclusion. □

8. Remembering only the recent past 2: $\mathfrak{M}_n = \{n-1, n\}$

In this section we assume that the elephant remembers the two most recent steps. We have, as always, $X_1 = 1$, $\mathbb{E}(X_2 \mid \mathcal{F}_1) = (2p-1)X_1$,

$$\mathbb{E}(X_3 \mid \mathcal{F}_2) = (2p-1)\frac{X_1 + X_2}{2} = (2p-1)\frac{1 + X_2}{2},$$

and, for $n \geq 3$,

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(X_{n+1} \mid X_{n-1}, X_n) = (2p-1)\frac{X_{n-1} + X_n}{2}.$$

Computing the moments, one obtains the following result. For the proof we refer to the Appendix, Section A.2.

Lemma 8.1. *As $n \to \infty$,*

$$\mathbb{E}(X_n) \to 0,$$

$$\mathbb{E}(S_n) \to \frac{(2p-1)(2p+1)}{4(1-p)},$$

$$\mathrm{var}\left(\frac{S_n}{\sqrt{n}}\right) \to 1 + \frac{(2p-1)(5-2p)}{2(1-p)(3-2p)} = \sigma_2^2.$$

The expectation of X_n tends to zero geometrically fast.

Remark 8.1. For $p = 1/2$ the process reduces, as usual, to a simple symmetric random walk.

For the following limit theorems we lean on the Markov property (and invite the reader to try the moment method).

Theorem 8.1. *We have*

$$\frac{S_n}{n} \xrightarrow{a.s.} 0 \quad \text{and} \quad \frac{S_n}{\sigma_2 \sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty.$$

Proof. The sequence $\{X_n, \ n \geq 1\}$ now forms a Markov chain of order two. Theorem 6.1 of [7] yields the strong law, and the results in [11, Section 3] or [12], combined with Corollary 5 of [14], yield the asymptotic normality with the moments as calculated above. $\quad\square$

Remark 8.2. If we suppose that the elephant remembers a fixed but finite number, say k, of the most recent steps, the sequence of steps forms a Markov chain of order k, and we obtain by (basically) the same arguments as above that $S_n/\sqrt{n}$ will be asymptotically normal (a Markov chain of order k can be considered as a k-dimensional Markov chain; now use [12], for example).

9. Remembering the distant as well as the recent past 1: $\mathfrak{M}_n = \{1, n\}$

Imagine a(n old) person who remembers their early childhood and events from the last few days but nothing in between. The most elementary model would be that of the heading.

Following the approach of earlier variants, we begin by assuming that $X_1 = 1$. Then, for $n \geq 2$,

$$\mathbb{E}(X_2) = \mathbb{E}(\mathbb{E}(X_2 \mid X_1)) = \mathbb{E}((2p-1)X_1) = 2p-1,$$

and

$$\begin{aligned}
\mathbb{E}(X_{n+1}) &= \mathbb{E}(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) \\
&= \mathbb{E}(\mathbb{E}(X_{n+1} \mid X_1, X_n)) \\
&= (2p-1)\,\mathbb{E}\left(\frac{X_1 + X_n}{2}\right) \\
&= (2p-1)\,\mathbb{E}\left(\frac{1 + X_n}{2}\right) \\
&= \frac{2p-1}{2} \cdot (1 + \mathbb{E}(X_n)).
\end{aligned}$$

Exploiting Proposition A.1(i) we obtain, for $n \geq 1$,

$$\mathbb{E}(X_n) = \frac{2p-1}{3-2p} + \left(\frac{2p-1}{2}\right)^{n-1} \cdot \frac{4(1-p)}{3-2p},$$

and hence that

$$\begin{aligned}
\mathbb{E}(T_n) &= 1 + (2p-1) + (n-2)\frac{2p-1}{3-2p} + \frac{4(1-p)}{3-2p}\sum_{k=1}^{n}\left(\frac{2p-1}{2}\right)^{k-1} \\
&= n \cdot \frac{2p-1}{3-2p} + \frac{8(1-p)}{(3-2p)^2} + \mathrm{o}(1) \quad \text{as } n \to \infty.
\end{aligned} \tag{9.1}$$

Next we note that $\mathbb{E}(T_1^2) = 1$ and, by (3.4), that, for $n \geq 1$,

$$\mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + (2p-1)\,\mathbb{E}(T_n) + (2p-1)\,\mathbb{E}(T_n X_n) + 1.$$

In order to establish a difference equation for the second moment we first have to compute the mixed moment. For the computational details we refer to Appendix A.3 and obtain (formula (A.3)),

$$\mathbb{E}(T_n^2) = \frac{(2p-1)^2}{(3-2p)^2} \cdot n^2 + \left(1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 40p + 35)\right) \cdot n + \mathrm{o}(n).$$

Joining the expressions for the first two moments, finally, tells us that the variance is linear in n:

$$\begin{aligned}
\mathrm{var}(T_n) &= n^2 \cdot \frac{(2p-1)^2}{(3-2p)^2} + n \cdot \left(1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 40p + 35)\right) \\
&\quad - \left(n \cdot \frac{2p-1}{3-2p} + \frac{8(1-p)}{(3-2p)^2}\right)^2 + \mathrm{o}(n) \\
&= n \cdot \sigma_T^2 + \mathrm{o}(n) \quad \text{as } n \to \infty,
\end{aligned}$$

where

$$\sigma_T^2 = 1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 24p + 19). \tag{9.2}$$

Given the expressions for mean and variance, a weak law is immediate:

$$\frac{T_n}{n} \xrightarrow{p} \frac{2p-1}{3-2p} \quad \text{as } n \to \infty.$$

In analogy with our earlier results, this suggests that T_n is asymptotically normal. That this is indeed the case follows from the fact that $\{X_n, \ n \geq 1\}$ is, once again, a uniformly ergodic Markov chain, since the only random piece from the past is the previous step. We may thus apply Corollary 5 of [14] (see also [13, Theorem 19.1]) to conclude that $T_n - \mathbb{E}(T_n)$ is asymptotically normal with mean zero and variance $\sigma_T^2 n$, with σ_T^2 as defined in (9.2), which, in view of (9.1), establishes that

$$\frac{T_n - \frac{2p-1}{3-2p}n}{\sigma_T \sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty. \tag{9.3}$$

An appeal to the discussion at the end of Section 2 concerning the relation between the two branches (the means have opposite signs and the variances are the same) now allows us to conclude that

$$\mathbb{E}(S_n) = p\,\mathbb{E}(T_n) + (1-p)\,\mathbb{E}(-T_n) = (2p-1)\,\mathbb{E}(T_n),$$

$$\mathbb{E}(S_n^2) = \mathbb{E}(T_n^2),$$

$$\text{var}(S_n) = \mathbb{E}(T_n^2) - (2p-1)^2(\mathbb{E}(T_n))^2,$$

which tells us that, as $n \to \infty$,

$$\mathbb{E}\left(\frac{S_n}{n}\right) \to \frac{(2p-1)^2}{3-2p} \quad \text{and} \quad \text{var}\left(\frac{S_n}{n}\right) \to \sigma_S^2 = 4p(1-p)\frac{(2p-1)^2}{(3-2p)^2}.$$

Furthermore, in analogy to Theorem 5.1, we arrive at the following asymptotic distributional behavior of S_n.

Theorem 9.1. *We have*

$$\frac{S_n}{n} \xrightarrow{d} S = \begin{cases} \dfrac{2p-1}{3-2p} & \text{with probability } p, \\[2ex] -\dfrac{2p-1}{3-2p} & \text{with probability } 1-p, \end{cases} \quad \text{as } n \to \infty.$$

Moreover, $\mathbb{E}(S_n/n)^r \to \mathbb{E}(S^r)$ for all $r > 0$, since $|S_n/n| \leq 1$ for all n.

Remark 9.1. Comparing this with Theorem 5.1, we see that the jump points are closer together here. This can be explained by the fact that the current random variables are less dependent than those in Section 5.

Finally, by combining (9.3) with the analog for the case $X_1 = -1$, asymptotic normality follows with a *random* centering.

Theorem 9.2. *We have*

$$\frac{S_n - \frac{(2p-1)X_1}{3-2p}n}{\sigma_T \sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty.$$

Proof. We first note that it follows from the discussion following (9.3) that the CLT there remains true when $X_1 = -1$, with $+$ replacing the $-$ in the numerator. We may thus argue as in the proof of Theorem 5.1, via the fact that

$$\frac{S_n - \frac{(2p-1)X_1}{3-2p}n}{\sigma_T\sqrt{n}} = X_1 \cdot \frac{T_n - \frac{(2p-1)}{3-2p}n}{\sigma_T\sqrt{n}}.$$

Alternatively, condition on the value of X_1 and proceed as in the proof of Theorem 6.2. $\quad\square$

10. Remembering the recent as well as the distant past 2: $\mathfrak{M}_n = \{1, 2, n\}$

Following the approach of earlier variants, we begin by assuming that $X_1 = X_2 = 1$. Then $\mathbb{E}(X_1) = \mathbb{E}(X_2) = 1$, $\mathbb{E}(X_3) = (2p-1)$ and, for $n \geq 3$,

$$\mathbb{E}(X_{n+1}) = \mathbb{E}(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) = \mathbb{E}(\mathbb{E}(X_{n+1} \mid X_1, X_2, X_n)) = \frac{2p-1}{3} \cdot (2 + \mathbb{E}(X_n)).$$

Exploiting Proposition A.1(i) yields

$$\mathbb{E}(X_n) = \frac{2p-1}{2-p} + \left(\frac{2p-1}{3}\right)^{n-1}\left(1 - \frac{2p-1}{2-p}\right) = \frac{2p-1}{2-p} + \frac{3(1-p)}{2-p}\left(\frac{2p-1}{3}\right)^{n-2},$$

and hence

$$\mathbb{E}(T_n) = 1 + 1 + (2p-1) + (n-3) \cdot \frac{2p-1}{2-p} + \frac{3(1-p)}{2-p} \sum_{k=4}^{n}\left(\frac{2p-1}{3}\right)^{k-2}$$

$$= n \cdot \frac{2p-1}{2-p} + \frac{3(1-p)(7-2p)}{2(2-p)^2} + o(1) \quad \text{as } n \to \infty. \tag{10.1}$$

As for second moments, $\mathbb{E}(T_1^2) = 1$, $\mathbb{E}(T_2^2) = 4$, $\mathbb{E}(T_3^2) = \mathbb{E}(1 + 1 + X_3)^2 = 4 + 4\,\mathbb{E}(X_3) + 1 = 4 + 4(2p-1) + 1 = 8p + 1$, and, generally,

$$\mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + 2\,\mathbb{E}(T_n X_{n+1}) + 1. \tag{10.2}$$

Concerning the mixed moments and other details we refer to Appendix A.4, from which we obtain

$$\mathbb{E}(T_n^2) = n^2 \cdot \frac{(2p-1)^2}{(2-p)^2} + n \cdot \left(1 + \frac{(2p-1)}{(2-p)^3} \cdot (5p^2 - 32p + 26)\right) + o(n).$$

The variance, finally, turns out as

$$\mathrm{var}(T_n) = n \cdot \sigma_T^2 + o(n), \quad \text{where } \sigma_T^2 = 1 + \frac{(2p-1)}{(2-p)^3} \cdot (5 - 5p - p^2). \tag{10.3}$$

Following the path of the previous section, we now immediately obtain a weak law:

$$\frac{T_n}{n} \xrightarrow{p} \frac{2p-1}{2-p} \quad \text{as } n \to \infty.$$

It remains to consider the general case with arbitrary X_1 and X_2. There is a slight change here from the previous section. Namely, we first have the case when $X_1 = X_2 = -1$, for which

the arguments from the previous section carry over without change, that is, the mean equals $\mathbb{E}(-T_n)$ and the second moment equals $\mathbb{E}(T_n^2)$. However, now we also have a mixed case which behaves somewhat differently.

Namely, consider the case when the first two summands are not equal: $X_1 + X_2 = 0$, $X_1 X_2 = -1$. Then

$$\mathbb{E}(X_3) = \mathbb{E}(\mathbb{E}(X_3 \mid X_1, X_2)) = \mathbb{E}\left((2p-1)\frac{X_1 + X_2}{2}\right) = (2p-1)\mathbb{E}(0) = 0,$$

and, for $n \geq 3$,

$$
\begin{aligned}
\mathbb{E}(X_{n+1}) &= \mathbb{E}(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) \\
&= \mathbb{E}(\mathbb{E}(X_{n+1} \mid X_1, X_2, X_n)) \\
&= \frac{2p-1}{3} \cdot (0 + \mathbb{E}(X_n)) \\
&= \frac{2p-1}{3} \mathbb{E}(X_n) \\
&= \cdots = C\,\mathbb{E}(X_3) = 0,
\end{aligned}
$$

from which we conclude that, for $n \geq 2$,

$$\mathbb{E}(T_n) = 0.$$

For the calculation of the second moment we refer again to Appendix A.4 and find that

$$\mathbb{E}(T_n^2) = n \cdot \frac{1+p}{2-p} + \mathrm{o}(n) = \mathrm{var}(T_n^2) \quad \text{as } n \to \infty.$$

The weak law now runs slightly differently, in that

$$\frac{T_n}{n} \xrightarrow{p} 0 \quad \text{as } n \to \infty.$$

We note in passing that the mean is linear in n and that the second moment is of order n^2 when the first two summands are equal, whereas the mean is zero and the second moment is linear in n when they are not. However, the variance is linear in n in all cases.

As for central limit theorems, the main arguments are the same as in Section 9, in that

$$\frac{T_n \pm \frac{2p-1}{2-p}n}{\sigma_T \sqrt{n}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty$$

for the cases $X_1 = X_2 = -1$ and $X_1 = X_2 = 1$, respectively, and

$$\frac{T_n}{\sqrt{n \cdot \frac{1+p}{2-p}}} \xrightarrow{d} \mathcal{N}_{0,1} \quad \text{as } n \to \infty$$

when the first two summands are unequal.

Switching to moments of S_n, using T_n^+, T_n^-, and T_n^0 for the three cases, we obtain

$$\mathbb{E}(S_n) = p^2\,\mathbb{E}(T_n^+) + (1-p) \cdot \mathbb{E}(T_n^0) + p(1-p)\,\mathbb{E}(T_n^-) = p(2p-1)\,\mathbb{E}(T_n^+),$$

$$\mathbb{E}(S_n^2) = p^2\,\mathbb{E}((T_n^+)^2) + (1-p) \cdot \mathbb{E}((T_n^0)^2) + p(1-p)\,\mathbb{E}((T_n^-)^2).$$

Collecting the various pieces tell us that, as $n \to \infty$,

$$\mathbb{E}\left(\frac{S_n}{n}\right) \to \frac{p(2p-1)^2}{2-p} \quad \text{and} \quad \text{var}\left(\frac{S_n}{n}\right) \to \frac{p(1-p)(2p-1)^2(4p^2+1)}{(2-p)^2}.$$

Finally, by modifying our earlier results of this kind, one ends up as follows.

Theorem 10.1. *We have*

$$\frac{S_n}{n} \overset{d}{\to} S = \begin{cases} \dfrac{2p-1}{2-p} & \text{with probability } p^2, \\ 0 & \text{with probability } 1-p, \qquad \text{as } n \to \infty. \\ -\dfrac{2p-1}{2-p} & \text{with probability } p(1-p), \end{cases}$$

Moreover, $\mathbb{E}(S_n/n)^r \to \mathbb{E}(S^r)$ for all $r > 0$, since $|S_n/n| \le 1$ for all n.

We finally wish to combine the three different beginnings of the process in order to arrive at a limit theorem for the S-process. This works (in theory) the same way as in Section 9. However, there is a problem with the variance. Namely, in Theorem 9.2 both cases had the same variance, whereas the variance when X_1 and X_2 are equal is not the same as when they are different. Nevertheless, here is the result.

Theorem 10.2. *We have*

$$\frac{S_n - \frac{(2p-1)(X_1+X_2)/2}{2-p}n}{\sqrt{n}} \overset{d}{\to} p \cdot \mathcal{N}_{0,\sigma_T^2} + (1-p) \cdot \mathcal{N}_{0,(1+p)/(2-p)} \quad \text{as } n \to \infty,$$

with σ_T^2 as given in (10.3)*.*

Proof. The conclusion follows by conditioning on the value of $(X_1 + X_2)/2$, and proceeding as in the proof of Theorem 6.2. $\qquad\square$

11. Miscellania

We close by mentioning some further specific models and by describing some problems and challenges for further research.

11.1. More on restricted memories

(i) The next logical step would be to check the case $\mathfrak{M}_n = \{1, n-1, n\}$. By modifying the computations in Appendix A.2, setting $a = (2p-1)/3$ and $d = 3a^2$, we find that

$$\begin{aligned} \mu_{n+1} &= \mathbb{E}(X_{n+1}) \\ &= \mathbb{E}(\mathbb{E}(X_{n+1} \mid X_1, X_{n-1}, X_n)) \\ &= \frac{2p-1}{3}\mathbb{E}(X_1 + X_{n-1} + X_n) \\ &= a((2p-1) + \mu_{n-1} + \mu_n) \\ &= a(\mu_{n-1} + \mu_n) + d, \end{aligned}$$

after which Proposition A.1(iv) and a glance at the computations in Appendix A.2 tell us that

$$\mathbb{E}(X_n) = \frac{(2p-1)^2}{5-4p} + O(q^n) \quad \text{as } n \to \infty,$$

where $q = \max\{|\lambda_1|, |\lambda_2|\} < 1$, with λ_i, $i = 1, 2$, defined in Appendix A.2. It follows that

$$\mathbb{E}(S_n) \sim n \frac{(2p-1)^2}{5-4p} \quad \text{as } n \to \infty.$$

If $\mathfrak{M}_n = \{1, 2, n-1, n\}$, then, with $a = (2p-1)/4$ and $d = 2p(2p-1)^2/4$,

$$\mathbb{E}(S_n) \sim n \frac{p(2p-1)^2}{3-2p} \quad \text{as } n \to \infty.$$

In fact, *theoretically* it is possible to obtain results of the above kind for any fixed number of early and/or late memory steps.

(ii) The case when the number of memory steps depends on n, e.g. $\log n$ or $\sqrt{n}$, is more subtle.

(iii) Another model is when the elephant remembers everything *except* the first step; more generally, the elephant remembers all but the first k steps for some $k \in \mathbb{N}$. Set $p_k = \mathbb{P}(X_{k+1} = 1)$, $V_n = X_{k+1} + \cdots + X_n$, and $\mathcal{H}_n = \sigma\{X_{k+1}, \ldots, X_n\}$, and let $n \geq k+1$. Then

$$\mathbb{E}(X_{n+1} \mid \mathcal{H}_n) = (2p-1)\frac{V_n}{n-k} \quad \text{and} \quad \mathbb{E}(V_{n+1} \mid \mathcal{H}_n) = \tilde{\gamma}_n V_n,$$

where $\gamma_n = (n-1-k+2p_k)/(n-k)$. With

$$\tilde{a}_n = \prod_{v=k+1}^{n-1} \tilde{\gamma}_v^{-1}$$

one can, as in [2], show that $\tilde{a}_n V_n$ is a martingale. From the same paper it follows that, provided $0 < p_k < 3/4$,

$$\frac{V_n}{\sqrt{n-k}} \xrightarrow{d} \mathcal{N}_{0, 1/(4-3p_k)} \quad \text{as } n \to \infty,$$

which implies that

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_{0, 1/(4-3p_k)} \quad \text{as } n \to \infty.$$

The quantity p_k depends on the construction used for the k steps $X_2, \ldots, X_{k+1}$.

Other cases one might think of is when the memory covers everything *except*

- the last j steps,

- the first k steps and the last j steps,

- the first $\alpha \log n$ steps and/or the last $\beta \log n$ steps for some $\alpha, \beta > 0$,

- the first, say, $\alpha \sqrt{n}$ steps and/or the last $\beta \sqrt{n}$ steps for some $\alpha, \beta > 0$,

- the first $\alpha \log n$ steps and/or the last $\beta \log n$ steps for some $\alpha, \beta > 0$,

- and so on, aiming at more general (final) results.

11.2. Phase transition

The results of Bercu [2] show that for the full memory one has a phase transition at $p = 3/4$. There is no such thing in our results. An interesting question would be to find the breaking point. There exist some papers on this topic using simulations (see e.g. [20], [6], and [18], and further papers cited therein), but we are not aware of any theoretical results concerning this matter.

11.3. Remembering the first versus the last step

There is a fundamental difference in behavior in these extreme cases; it is not just a matter of recalling some earlier step. Namely, it is a matter of comparing

$$X_{n+1} = \begin{cases} +X_1 & \text{with probability } p, \\ -X_1 & \text{with probability } 1-p, \end{cases}$$

with

$$X_{n+1} = \begin{cases} +X_n & \text{with probability } p, \\ -X_n & \text{with probability } 1-p. \end{cases}$$

In order to see the difference more clearly, let us imagine that p is close to one.

In the first case every new step most likely equals the *first* one, that is, a typical path will then consist of an overwhelming number of steps equal to the first one, interfoliated by an occasional $-X_1$. In the second case every new step most likely equals the *most recent* one, that is, a typical path will consist of an overwhelming number of steps equal to the first one, followed by an overwhelming number of steps equal to $-X_1$, and so on, that is, alternating long stretches of the same kind.

Moreover, since, in the first case, every new step is a function of just the first one, the independence structure does not come as a surprise, whereas in the second case the next step depends on the previous one, which in turn depends on its previous one, and so on, which implies that the next step in fact depends on the whole past.

11.4. Final remarks

(i) We have seen that the more the elephant remembers, the more cumbersome are the computations. However, once again, *in theory* it would be possible to compute higher-order moments and thus, for example, use the moment method to prove limit theorems.

(ii) By using the device in the first paragraph of Section 4, one can extend all limit theorems for ERWs to the case with general steps.

(iii) There is nothing particularly one-dimensional about our arguments. With a sensible definition in higher dimensions it would be possible to investigate analogous problems.

Appendix A

In this appendix we collect some details on difference equations and the more technical calculations.

A.1. Difference equations

In the proofs we use several difference equations. For convenience and easy reference we will summarize some well-known facts about linear difference equations that are used on and off (consult [15], for example).

Proposition A.1.

(i) *Consider the first-order equation*

$$x_{n+1} = a\,x_n + b_n \text{ for } n \geq 1, \text{ with } x_1 = x_1^* \text{ as initial value.}$$

Then

$$x_n = a^{n-1}x_1^* + \sum_{\nu=0}^{n-2} a^\nu b_{n-1-\nu}.$$

If, in addition, $|a| < 1$ and $b_n = bn^\gamma$ with $\gamma > -1$, then

$$x_n = \frac{b_{n-1}}{1-a} - \frac{\gamma\,ab_{n-1}}{n(1-a)^2}(1+\mathrm{o}(1)) \quad \text{as } n \to \infty.$$

(ii) *If, in particular, $|a| < 1$ and $x_{n+1} = ax_n + b$, then*

$$x_n = \frac{b}{1-a} + a^{n-1}\Big(x_1^* - \frac{b}{1-a}\Big) = \frac{b}{1-a}(1+\mathrm{o}(1)) \quad \text{as } n \to \infty.$$

(iii) *Next is the homogeneous second-order equation*

$$x_{n+1} = a\,x_n + b\,x_{n-1} \text{ for } n \geq 2, \text{ with } x_1^*,\ x_2^* \text{ given.}$$

Then, with $\lambda_{1/2} = (a \pm \sqrt{a^2 + 4b})/2$, provided $a^2 + 4b \neq 0$,

$$x_n^h = c_1\lambda_1^n + c_2\lambda_2^n, \text{ with } c_1,\ c_2 \text{ chosen such that } x_i^h = x_i^* \text{ for } i = 1,\,2.$$

(iv) *As for the inhomogeneous second-order equation*

$$x_{n+1} = a\,x_n + b\,x_{n-1} + d_n \text{ for } n \geq 2, \text{ with } x_1^*,\ x_2^* \text{ given,}$$

we have $x_n = x_n^h + y_n$, where y_n is some solution of the inhomogeneous equation, where the constants c_1, c_2 in x_n^h are chosen properly. If $d_n \equiv d$ and $a + b \neq 1$, we may choose $y_n = d/(1 - a - b)$.

A.2. Proof of Lemma 8.1

Set $a = p - 1/2 \in (-1/2,\ 1/2)$. Then

$$\mathbb{E}(X_1) = 2a \quad \text{and} \quad \mathbb{E}(X_2) = \mathbb{E}(X_2 \mid X_1) = 2a\,\mathbb{E}(X_1) = 4a^2.$$

For $n \geq 2$ we have

$$\mu_{n+1} = \mathbb{E}(X_{n+1}) = \mathbb{E}(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)) = \mathbb{E}((2p-1)\,(X_{n-1}+X_n)/2) = a\,(\mu_{n-1} + \mu_n).$$

With $\lambda_{1/2} = (a \pm \sqrt{a^2 + 4a})/2$ (note that $|\lambda_{1/2}| < 1$), this difference equation, with the two starting values $2a$ and $4a^2$, has, for $n \geq 1$, the solution

$$\mu_n = \mathbb{E}(X_n) = \frac{a(3a + \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}}\lambda_1^{n-1} - \frac{a(3a - \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}}\lambda_2^{n-1}.$$

For $p < 1/2$ we have $\sqrt{a^2 + 4a} = i\sqrt{|a^2 + 4a|}$, but the solution is still real. Next,

$$\mathbb{E}(S_n) = \sum_{k=1}^{n} \mu_k = \frac{a(3a + \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}} \frac{1 - \lambda_1^n}{1 - (a + \sqrt{a^2 + 4a})/2}$$

$$- \frac{a(3a - \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}} \frac{1 - \lambda_2^n}{1 - (a - \sqrt{a^2 + 4a})/2}$$

$$\to \frac{2a(3a + \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}(2 - a - \sqrt{a^2 + 4a})} - \frac{2a(3a - \sqrt{a^2 + 4a})}{\sqrt{a^2 + 4a}(2 - a + \sqrt{a^2 + 4a})}$$

$$= \frac{2a(a + 1)}{1 - 2a} = \frac{p^2 - 1/4}{1 - p} \quad \text{as } n \to \infty.$$

The second moment is more tedious. We begin with

$$\mathbb{E}(S_{n+1}^2 \mid \mathcal{F}_n) = S_n^2 + (2p - 1)S_n (X_{n-1} + X_n) + 1$$

and obtain

$$v_{n+1} = \mathbb{E}(S_{n+1}^2) = v_n + (2p - 1) \mathbb{E}(S_n X_n + S_n X_{n-1}) + 1. \tag{A.1}$$

As for the mixed moments,

$$\mathbb{E}(S_n X_n \mid \mathcal{F}_{n-1}) = a S_{n-1} X_{n-1} + a(S_{n-2} X_{n-2} + X_{n-1} X_{n-2}) + 1.$$

By the usual trick we find that

$$\mathbb{E}(X_n X_{n-1}) = a + a^2 + \cdots + a^{n-1} \mathbb{E}(X_2 X_1) \to \frac{a}{1 - a} = \frac{2p - 1}{3 - 2p},$$

and with $\zeta_n = \mathbb{E}(S_n X_n)$ that

$$\zeta_n = a(\mu_{n-1} + \mu_{n-2}) + 1 + \frac{4a^2}{1 - a} + \mathrm{O}(a^n),$$

from which it follows that

$$\zeta_n \to \frac{p^2 - 2p + 7/4}{(1 - p)(3 - 2p)} \quad \text{as } n \to \infty,$$

the stationary solution.

Next,

$$\mathbb{E}(S_n X_{n-1}) = \mathbb{E}(S_{n-1} X_{n-1}) + \mathbb{E}(X_n X_{n-1}) = \zeta_{n-1} + \frac{a}{1 - a} + \mathrm{O}(a^n).$$

Finally, recalling (A.1), we arrive at

$$v_{n+1} = v_n + (2p - 1)\left(\frac{2p^2 - 4p + 7/2}{(1 - p)(3 - 2p)} + \frac{2p - 1}{3 - 2p} \right) + 1 + \mathrm{o}(1) \quad \text{as } n \to \infty,$$

and thus, via telescoping, we obtain

$$v_n \sim n\left(1 + \frac{(2p - 1)(5 - 2p)}{2(1 - p)(3 - 2p)} \right) \quad \text{as } n \to \infty.$$

A.3. Calculation of second moments in Section 9

We first note that $\mathbb{E}(T_1^2) = 1$, and, by (3.4), that, for $n \geq 1$,

$$\mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + (2p-1)\,\mathbb{E}(T_n) + (2p-1)\,\mathbb{E}(T_nX_n) + 1. \tag{A.2}$$

At this point we have to pause and compute the mixed moments. We first note that $\mathbb{E}(T_1X_1) = 1$, and that

$$\mathbb{E}(X_2X_1) = \mathbb{E}(X_1\,\mathbb{E}(X_2 \mid X_1)) = \mathbb{E}(X_1(2p-1)X_1) = 2p-1,$$

so that

$$\mathbb{E}(T_2X_2 \mid \mathcal{F}_n) = \mathbb{E}(X_1X_2 + 1) = 2p - 1 + 1 = 2p.$$

For $n \geq 2$ we exploit (3.3), (9.1), and the fact that $X_n^2 = 1$, to obtain

$$\begin{aligned}
\mathbb{E}(T_{n+1}X_{n+1}) &= \frac{2p-1}{2} \cdot \mathbb{E}(T_nX_n) + \frac{2p-1}{2}\,\mathbb{E}(T_n) + 1 \\
&= \frac{2p-1}{2} \cdot \mathbb{E}(T_nX_n) + \frac{2p-1}{2}\left(n \cdot \frac{2p-1}{3-2p} + \frac{8(1-p)}{(3-2p)^2} + o(1)\right) + 1 \\
&= \frac{2p-1}{2} \cdot \mathbb{E}(T_nX_n) + \frac{(2p-1)^2}{2(3-2p)} \cdot n + \frac{4(2p-1)(1-p)}{(3-2p)^2} + 1 + o(1).
\end{aligned}$$

Another application of Proposition A.1(i) then tells us that

$$\begin{aligned}
\mathbb{E}(T_nX_n) &= \left\{\left(\frac{(2p-1)^2}{2(3-2p)} \cdot (n-1) + \frac{4(2p-1)(1-p)}{(3-2p)^2} + 1\right)\Big/\left(1 - \frac{2p-1}{2}\right)\right. \\
&\qquad\left. - \frac{2p-1}{2} \cdot \frac{(2p-1)^2}{2(3-2p)} \cdot (n-1)\Big/n\left(1 - \frac{2p-1}{2}\right)^2\right\} \cdot (1 + o(1)) \\
&= \frac{(2p-1)^2}{(3-2p)^2} \cdot n + \frac{8(1+p-2p^2)}{(3-2p)^3} + o(1).
\end{aligned}$$

Hence, using (A.2), we obtain

$$\begin{aligned}
\mathbb{E}(T_{n+1}^2) &= \mathbb{E}(T_n)^2 + (2p-1)\left(n \cdot \frac{2p-1}{3-2p} + \frac{8(1-p)}{(3-2p)^2}\right) \\
&\qquad + (2p-1)\left(\frac{(2p-1)^2}{(3-2p)^2} \cdot n + \frac{8(1+p-2p^2)}{(3-2p)^3}\right) + 1 + o(1) \\
&= \mathbb{E}(T_n)^2 + \frac{2(2p-1)^2}{(3-2p)^2} \cdot n + \frac{32(2p-1)(1-p)}{(3-2p)^3} + 1 + o(1),
\end{aligned}$$

after which, via telescoping, we obtain

$$\begin{aligned}
\mathbb{E}(T_n^2) &= \frac{(2p-1)^2}{(3-2p)^2} \cdot n^2 - \frac{(2p-1)^2}{(3-2p)^2} \cdot n + (n-1) \cdot \left(1 + \frac{32(2p-1)(1-p)}{(3-2p)^3}\right) + o(n) \\
&= \frac{(2p-1)^2}{(3-2p)^2} \cdot n^2 + \left(1 + \frac{(2p-1)}{(3-2p)^3}(4p^2 - 40p + 35)\right) \cdot n + o(n). \tag{A.3}
\end{aligned}$$

A.4. Calculation of second moments in Section 10

The point of departure in this case is (10.2), that is,

$$\mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + 2\,\mathbb{E}(T_n X_{n+1}) + 1. \tag{A.4}$$

For the mixed moments we use (3.3):

$$\mathbb{E}(T_n X_{n+1} \mid \mathcal{G}_n) = \frac{2}{3}(2p-1)T_n + \frac{2p-1}{3}T_n X_n = \frac{2}{3}(2p-1)T_n + \frac{2p-1}{3}T_{n-1}X_n + \frac{2p-1}{3}.$$

We thus find, using (10.1), that for $n \geq 3$

$$\mathbb{E}(T_n X_{n+1})$$

$$= \frac{2p-1}{3}\,\mathbb{E}(T_{n-1}X_n) + \frac{2p-1}{3} + \frac{2(2p-1)}{3}\left(n \cdot \frac{2p-1}{2-p} + \frac{3(1-p)(7-2p)}{2(2-p)^2} + \mathrm{o}(1)\right)$$

$$= \frac{2p-1}{3}\,\mathbb{E}(T_{n-1}X_n) + \frac{2(2p-1)^2}{3(2-p)^2} \cdot n + \frac{(2p-1)(1-p)(7-2p)}{(2-p)^2} + \frac{2p-1}{3} + \mathrm{o}(1).$$

Invoking Proposition A.1(i) then tells us that

$$\mathbb{E}(T_n X_{n+1})$$

$$= \left\{\left(\frac{2(2p-1)^2}{3(2-p)^2} \cdot n + \frac{(2p-1)(1-p)(7-2p)}{(2-p)^2} + \frac{2p-1}{3}\right)\Big/\left(1 - \frac{2p-1}{3}\right)\right.$$

$$\left. - \frac{2p-1}{3} \cdot \frac{2(2p-1)^2}{3(2-p)^2} \cdot n\Big/\left((n+1)\left(1 - \frac{2p-1}{3}\right)^2\right)\right\} \cdot (1 + \mathrm{o}(1))$$

$$= \frac{(2p-1)^2}{(2-p)^2} \cdot n + \frac{3(2p-1)}{(2-p)^3} \cdot (p^2 - 9p + 8) + \mathrm{o}(1),$$

which, inserted into (A.4), yields

$$\mathbb{E}(T_{n+1}^2) = \mathbb{E}(T_n^2) + \frac{2(2p-1)^2}{(2-p)^2} \cdot n + \frac{3(2p-1)(p^2 - 9p + 8)}{(2-p)^3} + 1 + \mathrm{o}(1),$$

and, after summation,

$$\mathbb{E}(T_n^2)$$

$$= n^2 \cdot \frac{(2p-1)^2}{(2-p)^2} - n \cdot \frac{(2p-1)^2}{(2-p)^2} + (n-1) \cdot \left(1 + \frac{3(2p-1)(p^2 - 9p + 8)}{(2-p)^3}\right) + \mathrm{o}(n)$$

$$= n^2 \cdot \frac{(2p-1)^2}{(2-p)^2} + n \cdot \left(1 + \frac{(2p-1)}{(2-p)^3} \cdot (5p^2 - 32p + 26)\right) + \mathrm{o}(n).$$

Finally we turn our attention to the second moment for the case when $X_1 \cdot X_2 = -1$, where, again, the mixed moment is first in focus. Now $\mathbb{E}(T_1 X_1) = 1$, $\mathbb{E}(T_2 X_2) = \mathbb{E}(X_1 X_2 + X_2^2) = -1 + 1 = 0$, and $\mathbb{E}(T_3 X_3) = \mathbb{E}(T_2 X_3 + X_3)^2 = 0 + 1 = 1$.

For $n \geq 3$ we follow the usual pattern. Due to the fact that the mean is zero, an application of (3.3) now yields

$$\mathbb{E}(T_n X_{n+1}) = \frac{2p-1}{3}\,\mathbb{E}(T_n X_n) = \frac{2p-1}{3}\,\mathbb{E}(T_{n-1}X_n) + \frac{2p-1}{3},$$

which, together with Proposition A.1(i), tells us that

$$\mathbb{E}(T_n X_{n+1}) = \frac{(2p-1)/3}{1-(2p-1)/3} + \mathrm{o}(1) = \frac{2p-1}{2(2-p)} + \mathrm{o}(1) \quad \text{as } n \to \infty. \tag{A.5}$$

Moving into second moments, $\mathbb{E}(T_1^2) = 1$, $\mathbb{E}(T_2^2) = 0$, and $\mathbb{E}(T_3^2) = \mathbb{E}(X_3^2) = 1$.

For $n \geq 3$ we insert our findings in (A.5) into (A.4):

$$\begin{aligned}
\mathbb{E}(T_{n+1}^2) &= \mathbb{E}(T_n^2) + 2\,\mathbb{E}(T_n X_{n+1}) + 1 \\
&= \mathbb{E}(T_n)^2 + \frac{2p-1}{2-p} + 1 + \mathrm{o}(1) \\
&= \mathbb{E}(T_n^2) + \frac{1+p}{2-p} + \mathrm{o}(1) \quad \text{as } n \to \infty,
\end{aligned}$$

so that, via telescoping,

$$\mathbb{E}(T_n^2) = n \cdot \frac{1+p}{2-p} + \mathrm{o}(n) = \mathrm{var}(T_n^2) \quad \text{as } n \to \infty.$$

Acknowledgements

References

[1] BEN-ARI, I., GREEN, J., MEREDITH, T, PANZO H. AND TAN, X. (2019). Finite-memory elephant random walk and the central limit theorem for additive functionals. Available at arXiv:1911.05716.

[2] BERCU, B. (2018). A martingale approach for the elephant random walk. *J. Phys. A* **81**, 015201.

[3] CHEN, A. AND RENSHAW, E.(1992). The Gillis–Domb–Fisher correlated random walk. *J. Appl. Prob.* **29**, 792–813.

[4] COLETTI, C. F., GAVA, R. AND SCHÜTZ, G. M. (2017). Central limit theorem and related results for the elephant random walk. *J. Math. Phys.* **58**, 053303.

[5] COMETS, F., MENSHIKOV, M. V. AND WADE, A. R. (2019). Random walks avoiding their convex hull with a finite memory. Available at arXiv:1902.09812.

[6] CRESSONI, J. C., DA SILVA, M. A. A. AND VISWANATHAN, G. M. (2007). Amnestically induced persistence in random walks. *J. Phys. A* **46**, 505002.

[7] DOOB, J. L.(1953). *Stochastic Processes*. John Wiley, New York.

[8] ENGLÄNDER, J. AND VOLKOV, S. (2018). Turning a coin over instead of tossing it. *J. Theoret. Prob.* **31**, 1097–1118.

[9] GONZÁLEZ-NAVARETTE, M. AND LAMBERT, R. (2018). Non-Markovian random walks with memory lapses. *J. Math. Phys.* **59**, 113301.

[10] GUT, A. (2013). *Probability: A Graduate Course*, 2nd edn. Springer, New York.

[11] HERKENRATH, U. (2003). A new approach to Markov processes of order 2. *Ann. Univ. Craiova, Math. Comp. Sci. Ser.* **30**, 106–115.

[12] HERKENRATH, U., IOSIFESCU, M. AND RUDOLPH, A. (2003). A note on invariance principles for iterated random functions. *J. Appl. Prob.* **40**, 834–837.

[13] IBRAGIMOV, I. A. AND LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters–Noordhof, Groningen.

[14] JONES, G. L. (2004). On the Markov chain central limit theorem. *Prob. Surveys* **1**, 299–320.

[15] KELLEY, W. AND PETERSON, A. (2001). *Difference Equations: An Introduction with Applications*. Harcourt/Academic Press, San Diego.

[16] LYONS, R. (1988). Strong laws of large numbers for weakly correlated random variables. *Michigan Math. J.* **35**, 353–359.

[17] MENSHIKOV, M. AND VOLKOV, S. (2008). Urn-related random walk with drift $\rho x^{\alpha}/t^{\beta}$. *Electron. J. Prob.* **13**, 944–960.

[18] MOURA, TH. R. S., VISWANATHAN, G. M., DA SILVA, M. A. A., CRESSONI, J. C. and DA SILVA, L. R. (2016). Transient superdiffusion in random walks with a q-exponentially decaying memory profile. *Physica A* **453**, 259–263.

[19] SCHÜTZ, G. M. AND TRIMPER, S. (2004). Elephants can always remember: exact long-range memory effects in a non-Markovian random walk. *Phys. Rev. E*, **70**, 045101.

[20] DA SILVA, M. A. A., CRESSONI, J. C. AND VISWANATHAN, G. M. (2006). Discrete-time non-Markovian random walks: the effect of memory limitations on scaling. *Physica A* **364**, 70–78.